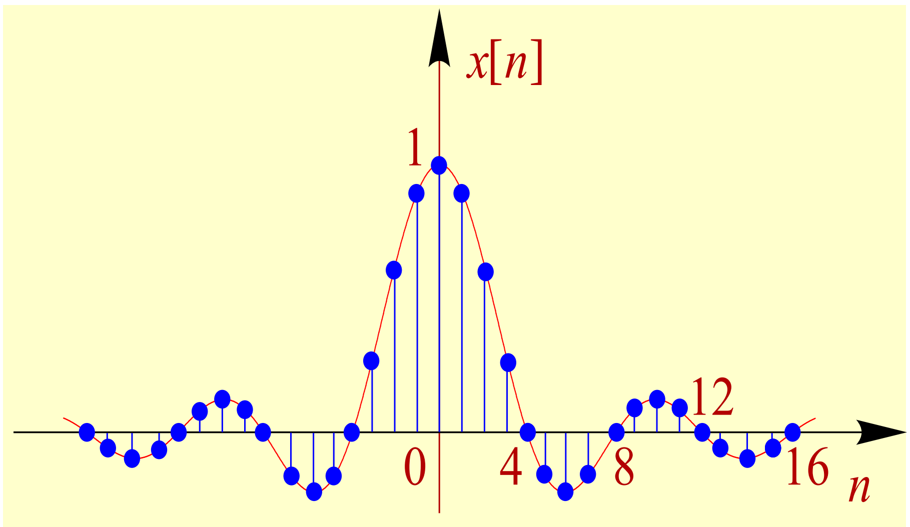
Przetwarzanie sygnałów

Jerzy Szabatin



23 września 2003

Spis treści

Spis rysunków

Rozdział 1

Elementy ogólnej teorii sygnałów

Lekcja 1

Sygnały deterministyczne

W lekcji 1 wprowadzimy pojęcie sygnału i dokonamy podstawowej klasyfikacji sygnałów. Podkreślimy rolę sygnału jako nośnika informacji. Uwagę skupimy na sygnałach deterministycznych. Omówimy zarówno analogowe sygnały deterministyczne jak i dyskretne sygnały deterministyczne. Zdefiniujemy podstawowe parametry charakteryzujące sygnały, w tym energię i moc sygnału. Na tej podstawie dokonamy podziału sygnałów na sygnały o ograniczonej energii i sygnały o ograniczonej mocy. Podamy także przykłady prostych, typowych sygnałów analogowych i dyskretnych należących do obu klas, wraz z ich notacją i wykresami przebiegów czasowych.

1.1. Wprowadzenie

1.1.1. Modele matematyczne

W wielu naukach, badających otaczającą nas rzeczywistość fizyczną, wprowadza się w celu jej opisania odpowiednie *modele matematyczne*. Tak postępuje się w fizyce, gdzie np. ruch masy zawieszonej na nici jest opisywany i analizowany za pomocą modeli matematycznych: *wahadła matematycznego* lub *wahadła fizycznego*. Tak też postępuje się w teorii obwodów, gdzie fizyczne elementy: opornik, cewkę i kondensator opisuje się modelami matematycznymi w postaci elementów obwodowych: *oporu, indukcyjności* i *pojemności*. Tak również postępuje się w teorii sygnałów, gdzie sygnały występujące w rzeczywistości opisuje się za pomocą różnego rodzaju modeli matematycznych.

Posługiwanie się modelami matematycznymi ma szereg istotnych zalet. Opis sygnału za pomocą modelu matematycznego umożliwia przede wszystkim jego formalną, teoretyczną analizę. Ważną zaletą tego opisu jest także to, iż –

w zależności od potrzeb i postawionego celu analizy – temu samemu fizycznemu sygnałowi możemy przyporządkować różne modele o zróżnicowanymi stopniu złożoności. Jeżeli interesują nas jedynie zasadnicze, dominujące cechy sygnału, stosujemy modele prostsze. W przypadku, gdy chcemy w opisie sygnału uwzględnić także jego cechy drugorzędne, wprowadzamy modele bardziej złożone. Operowanie modelami matematycznymi umożliwia ponadto wprowadzenie jednoznacznych kryteriów podziału sygnałów i na tej podstawie dokonanie ich klasyfikacji. I wreszcie, jeśli rozpatrujemy sygnały w kategoriach modeli matematycznych, możemy abstrahować od ich natury fizycznej. W analizie formalnej sygnałów nie jest bowiem istotne jakie jest fizyczne źródło ich pochodzenia.

1.1.2. Pojęcie sygnału

W znaczeniu potocznym pojęcie *sygnału* jest rozumiane jako proces zmian w czasie pewnej wielkości fizycznej lub stanu obiektu fizycznego. Z tego względu za modele matematyczne sygnałów przyjmujemy funkcje, których argumentem jest czas t. Opisują one ewolucję sygnałów w czasie. W najprostszym przypadku są to funkcje tylko jednej zmiennej t. W przypadkach bardziej złożonych, np. w teorii linii długich lub zagadnieniach przetwarzania obrazów, mogą to być funkcje wielu zmiennych: czasu i współrzędnych przestrzennych. Niekiedy stosowane są także funkcje wektorowe.

Jeżeli funkcje opisujące sygnały przyjmują wartości rzeczywiste, to modele takie nazywamy *rzeczywistymi*. Do opisu sygnałów często są stosowane także funkcje o wartościach zespolonych. Modele *zespolone* zwiększają wprawdzie stopień abstrakcji opisu sygnałów, jednak, jak przekonamy się wkrótce, ułatwiają znacznie ich analizę formalną. Jako modele sygnałów wprowadza się również wielkości niefunkcyjne nazywane *dystrybucjami*.

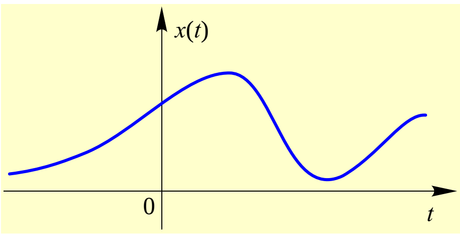
W dalszym ciągu pojęcie sygnału będziemy utożsamiać z jego modelem matematycznym, a stosując termin sygnał będziemy mieli na myśli jego model.

1.1.3. Klasyfikacja sygnałów

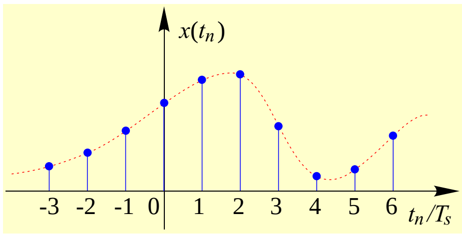
Sygnały można klasyfikować w różny sposób, stosując różne kryteria ich podziału. Pierwsze z tych kryteriów wynika z omówionego wyżej podziału modeli matematycznych sygnałów, według którego sygnały dzielimy na rzeczywiste, zespolone i dystrybucyjne.

Inne kryterium jest związane ze zdolnością do przewidywania wartości sygnału w dowolnej chwili t. Prowadzi ono podziału sygnałów na *deterministyczne* i *stochastyczne* (*losowe*). Jeżeli w każdej chwili potrafimy przewidzieć wartość sygnału, a jego zachowanie opisać jednoznacznie formułą matematyczną, przedstawić za pomocą wykresu lub tablicy jego wartości, to sygnał uważamy za deterministyczny. Jeżeli prognozy takiej nie możemy dokonać, a znamy jedynie ogólne prawa statystyczne, według których sygnał ewoluuje w czasie, to traktujemy go jako losowy. Należy podkreślić, że podział sygnałów na deterministyczne i losowe jest arbitralny. Ten sam fizyczny sygnał możemy opisywać raz modelem deterministycznym, innym zaś razem modelem losowym, w zależności od posiadanej o nim wiedzy *a priori*. W klasie sygnałów losowych wyróżnia się czasami podklasę sygnałów *quasi-deterministycznych*, których niedeterminizm jest związany jedynie z losowym charakterem pewnych parametrów sygnału.

Ważna linia podziału sygnałów dotyczy dziedziny ich określoności. Sygnały określone w zbiorze ciągłym osi czasu są nazywane sygnałami *ciągłymi w czasie* lub krótko sygnałami *ciągłymi* (rys. 1.1). Najczęściej dziedziną takich sygnałów jest cała oś $(-\infty, \infty)$, dodatnia półoś $[0, \infty)$ lub odcinek $[t_1, t_2]$ osi czasu. Sygnały określone w dyskretnym (przeliczalnym lub skończonym) zbiorze punktów osi czasu $(\ldots, t_{-1}, t_0, t_1, t_2, \ldots)$ i nieokreślone w pozostałych punktach są nazywane sygnałami *dyskretnymi w czasie* lub krótko sygnałami *dyskretnymi* (rys. 1.2). Najczęściej dziedziną tych sygnałów jest zbiór chwil $t_n = nT_s$, $n \in \mathscr{A}$, odległych od siebie o stały odstęp T_s nazywany *przedziałem dyskretyzacji* . Zbiór $\mathscr{A}$ wartości zmiennej dyskretnej n jest zwykle zbiorem wszystkich liczb całkowitych, zbiorem liczb naturalnych lub skończonym podzbiorem $[n_1, n_2]$ zbioru liczb całkowitych.

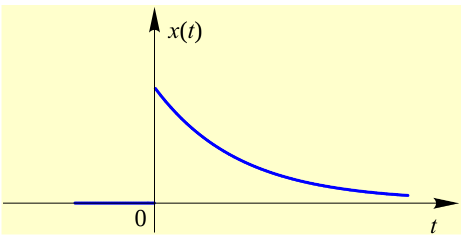


Rys. 1.1. Sygnał ciągły w czasie

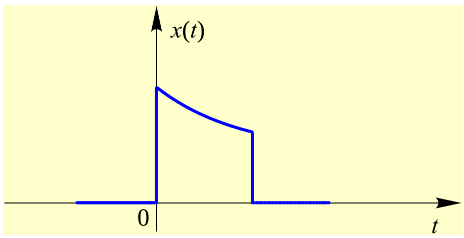


Rys. 1.2. Sygnał dyskretny w czasie

Jeżeli sygnał przybiera wartości różne od zera w przedziale nieskończonym, nazywamy go sygnałem *o nieskończonym czasie trwania* (rys. 1.3). W przypadku, gdy sygnał przybiera wartości różne od zera jedynie w przedziale skończonym nazywamy go sygnałem *o skończonym czasie trwania* lub krótko – sygnałem *impulsowym* (rys. 1.4).



Rys. 1.3. Sygnał o nieskończonym czasie trwania

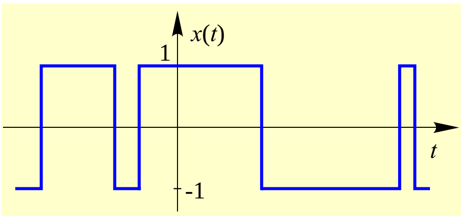


Rys. 1.4. Sygnał impulsowy

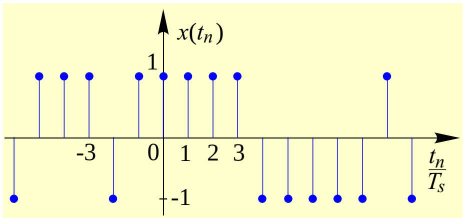
Zarówno sygnały ciągłe, jak i dyskretne dzielimy także w zależności od rodzaju ich przeciwdziedziny. Jeżeli zbiór możliwych wartości sygnału jest zbiorem ciągłym, sygnał nazywamy *ciągłym w amplitudzie*. Jeżeli zbiór ten jest zbiorem dyskretnym (przeliczalnym lub skończonym), sygnał nazywamy *dyskretnym w amplitudzie*. Ze względu na charakter dziedziny i przeciwdziedziny sygnały dzielimy zatem na:
– *ciągłe w czasie i ciągłe w amplitudzie* (nazywane także *analogowymi*),
– *ciągłe w czasie i dyskretne w amplitudzie,*
– *dyskretne w czasie i ciągłe w amplitudzie,*
– *dyskretne w czasie i dyskretne w amplitudzie.*

W tej ostatniej klasie wyróżnia się sygnały *cyfrowe*. Są to sygnały dyskretne w czasie, których zbiór możliwych wartości jest skończony.

Szczególną podklasę sygnałów dyskretnych w amplitudzie stanowią sygnały, które w każdej chwili określoności przybierają tylko dwie wartości utożsamiane ze znakami (cyframi) binarnymi „0" oraz „1". Sygnały takie są nazywane *binarnymi*. Sygnały binarne mogą być zarówno ciągłe (rys. 1.5), jak i dyskretne w czasie (rys. 1.6). Sygnały binarne dyskretne w czasie otrzymuje się w wyniku binarnego kodowania sygnałów cyfrowych. Sygnały binarne odgrywają coraz większą rolę we współczesnych systemach przetwarzania i transmisji sygnałów.



Rys. 1.5. Ciągły sygnał binarny



Rys. 1.6. Dyskretny sygnał binarny

Oprócz wymienionych istnieje jeszcze wiele innych linii podziału sygnałów. Będziemy je omawiać w miarę wprowadzania odpowiednich definicji parametrów i charakterystyk sygnałów, na których są oparte dalsze kryteria klasyfikacji.

1.1.4. Sygnał a informacja

Sygnały służą do przekazywania wiadomości. Wraz z przesłaniem wiadomości dostarczana jest określona informacja. Pojęcie sygnału jest zatem nierozerwalnie związane z pojęciem informacji. Mówi się często, że sygnał jest nośnikiem informacji.

Informacja, obok materii i pola, jest jednym z najważniejszych pojęć przyrodoznawstwa. W świadomości ludzkiej długo funkcjonowała jako pojęcie subiektywne, niemierzalne, tak jak np. dobro, czy piękno. Zobiektywizował je dopiero Claude E. Shannon. Jego praca *A mathematical theory of communication* [1], opublikowana w 1948 r. i określana dziś zgodnie przez uczonych jako

Magna Carta ery informacyjnej, dała początek nowej dziedzinie wiedzy – *teorii informacji*. Problematyka teorii informacji wykracza poza ramy wykładu. Studiując wszakże jego kolejne strony, warto pamiętać, że wraz z przetwarzaniem i przesyłaniem sygnału jest przetwarzana i przekazywana informacja. Dziś, dzięki Shannonowi, wiemy już jak tę informację mierzyć [2].

Zastanówmy się jednak, czy każdy sygnał niesie ze sobą informację. Przekazanie informacji jest aktem wypełnienia naszej niewiedzy. Jeśli sygnał jest deterministyczny, znamy dokładnie jego przebieg w przeszłości, wartość w chwili bieżącej i zachowanie się w przyszłości. Nasza wiedza o nim jest pełna. Nie może on nam zatem dostarczyć informacji. Informację przekazują tylko takie sygnały, które dla odbiorcy są losowe. Tylko wtedy bowiem, kiedy nie jesteśmy w stanie przewidzieć zachowania się sygnału w przyszłości, a jego przebieg możemy prognozować jedynie z pewnym prawdopodobieństwem (dokładniej, przypisać mu pewną miarę prawdopodobieństwa), fakt odebrania sygnału wypełnia naszą niewiedzę.

Sygnałami losowymi są oczywiście sygnały transmitowane w systemach komunikacyjnych powszechnego użytku: telefonicznych, radiowych, telewizyjnych. W przeciwnym przypadku rozmawianie przez telefon, słuchanie radia, czy oglądanie telewizji byłoby pozbawione sensu. Podobnie, gdyby przepływ sygnałów w naszym komputerze osobistym przebiegał według praw deterministycznych, siedzenie przy nim byłoby zajęciem nader nudnym.

Sygnałami losowymi są również sygnały pochodzące z przestrzeni kosmicznej odbierane przez radioteleskopy. Dostarczają nam one informacji np. o strukturze gwiazd. Do sygnałów losowych należą także wszelkie szumy i zakłócenia towarzyszące nieuchronnie sygnałom użytecznym. Jeżeli zastanowimy się głębiej, dojdziemy do przekonania, że w istocie rzeczy wszystkie realne sygnały fizyczne są losowe. To my, w zależności od celów badań oraz posiadanej wiedzy *a priori*, tworzymy ich modele deterministyczne lub stochastyczne. Jeżeli na oscylogram sygnału sinusoidalnego o znanej amplitudzie, częstotliwości i fazie początkowej, otrzymanego na wyjściu fizycznego generatora, spojrzymy gołym okiem, to widzimy regularną deterministyczną krzywą o dobrze nam znanym kształcie. Jeśli jednak ten sam oscylogram będziemy oglądać w dostatecznie dużym powiększeniu, to okaże się, że ta gładka krzywa jest w rzeczywistości postrzępionym, nieregularnym przebiegiem. Na sygnał sinusoidalny nakładają się bowiem szumy wewnętrzne elementów generatora, powodujące jego fluktuacje. W przypadku, gdy sygnał ten wykorzystujemy do celów technicznych (np. jako sygnał nośny lub synchronizujący), przyporządkowujemy mu makroskopowy model deterministyczny. Jeżeli natomiast chcemy pozyskać informację np. o mocy szumu generatora, tworzymy jego dokładniejszy model stochastyczny, uwzględniający także składową szumową.

1.1.5. Reprezentacje sygnałów

Tylko nieliczne proste sygnały można opisać formułami matematycznymi i analizować w naturalnej dziedzinie czasu. Większość sygnałów, z jakimi spotykamy się w praktyce, ma przebieg na tyle złożony i nieregularny, że ich bezpośredni opis w tej dziedzinie jest niemożliwy. Z tego względu w analizie formalnej sygnałów posługujemy się często ich różnego rodzaju reprezentacjami.

Reprezentacja sygnału stanowi pewien rodzaj jego symbolicznego opisu, niekiedy o znacznym stopniu abstrakcji. Jej istotą jest to, że zawiera ona pełną informację o sygnale, choć zwykle wyrażoną w innym języku, niż bezpośredni język opisu czasowego. Oznacza to, że znając sygnał, możemy jednoznacznie wyznaczyć jego reprezentację, znając zaś tę reprezentację – odtworzyć jednoznacznie sygnał.

Istnieje wiele sposobów reprezentacji sygnałów. Kilka najważniejszych poznamy w trakcie wykładu. Niektórymi z nich już posługiwaliśmy się w teorii obwodów. Na przykład, sygnał harmoniczny $A\cos(\omega t + \varphi)$ o ustalonej i znanej pulsacji ω reprezentowaliśmy za pomocą liczby (amplitudy) zespolonej $A\mathrm{e}^{\mathrm{j}\varphi}$. Znając tę liczbę (i pulsację ω), możemy jednoznacznie wyznaczyć przebieg sygnału harmonicznego w czasie. Podobnie, sygnał okresowy o ustalonym okresie T_0 reprezentowaliśmy za pomocą jego trygonometrycznego szeregu Fouriera :

$$a_0 + \sum_{k=1}^{\infty}(a_k \cos k\omega_0 t + b_k \sin k\omega_0 t), \tag{1.1}$$

gdzie $\omega_0 = 2\pi/T_0$. Reprezentację sygnału stanowi w tym przypadku nieskończony przeliczalny zbiór liczb rzeczywistych $\{a_0, a_k, b_k \colon k = 1, 2, \dots\}$ nazywanych współczynnikami Fouriera. W szczególnych przypadkach zbiór ten jest skończony. Znając te liczby (oraz okres T_0), jesteśmy w stanie odtworzyć jednoznacznie przebieg czasowy sygnału. Można zatem powiedzieć, że przy założeniu znanego okresu T_0 w zbiorze współczynników Fouriera jest zakodowana pełna informacja o sygnale.

1.2. Sygnały analogowe

Sygnały deterministyczne nie przenoszą informacji. Mają zatem ograniczone znaczenie w zagadnieniach telekomunikacji, radiokomunikacji, metrologii i innych dziedzinach, w których punkt ciężkości jest położony na problemy pozyskiwania, przetwarzania i przesyłania informacji. Mimo to, znaczna część wykładu będzie poświęcona tej właśnie klasie sygnałów. Ich dokładne omówienie jest ważne z co najmniej kilku powodów.

Bez opanowania podstaw teorii sygnałów deterministycznych trudno byłoby zrozumieć znacznie trudniejszą od strony pojęciowej i analitycznej problematykę sygnałów stochastycznych. Na przykładzie sygnałów deterministycznych można w prosty sposób wyjaśnić podstawowe sposoby opisu i reprezentacji sygnałów oraz zdefiniować szereg fundamentalnych pojęć związanych z sygnałem, takich jak funkcja autokorelacji czy widmo. Znając definicje tych pojęć dla sygnałów deterministycznych, można je potem łatwo uogólnić na sygnały losowe. Ponadto, w kategoriach sygnałów deterministycznych są opisywane podstawowe operacje na sygnałach, a także układy, przez które sygnały te są przetwarzane. I wreszcie, posługiwanie się modelami deterministycznymi jest w wielu zagadnieniach dużo prostsze, wygodniejsze, a jednocześnie wystarczające do przeprowadzenia odpowiedniej analizy. Sytuacja taka występuje np. w teorii modulacji, gdzie podstawowe właściwości różnych systemów modulacji można zbadać stosując modele deterministyczne sygnałów zmodulowanych, podczas gdy w istocie sygnały transmitowane w tych systemach mają charakter losowy.

1.2.1. Notacja

Analogowe sygnały deterministyczne będziemy oznaczać literami łacińskimi $x(t), y(t), z(t), \ldots$, dodając w razie potrzeby odpowiednie indeksy. Niektóre typowe sygnały będziemy oznaczać symbolami specjalnymi, ułatwiającymi zapis operacji formalnych. Dopóki sygnały deterministyczne będziemy rozpatrywać jako modele matematyczne i nie będziemy ich wiązać z konkretnymi sygnałami fizycznymi, dopóty wartości sygnału będziemy traktować jako bezwymiarowe.

1.2.2. Parametry

Najprostszymi charakterystykami sygnałów są ich parametry. Do najczęściej używanych należy: wartość średnia, energia oraz moc. Definicje tych parametrów podamy najpierw dla sygnałów rzeczywistych.

Definicja 1.1. *Wartość średnia* analogowego impulsowego sygnału deterministycznego $x(t)$ określonego w przedziale $[t_1, t_2]$ jest całką z tego sygnału w przedziale $[t_1, t_2]$ odniesioną do szerokości tego przedziału:

$$\langle x \rangle = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t)\,\mathrm{d}t. \tag{1.2}$$

W przypadku sygnałów o nieskończonym czasie trwania wartość średnia jest określona jako wielkość graniczna:

$$\langle x \rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\,\mathrm{d}t. \tag{1.3}$$

W szczególnym przypadku, gdy sygnał o nieskończonym czasie trwania jest sygnałem okresowym o okresie T_0, uśrednianie w czasie nieskończonym jest równoważne uśrednianiu za okres:

$$\langle x \rangle = \frac{1}{T_0} \int\limits_{t_0}^{t_0+T_0} x(t)\,\mathrm{d}t, \tag{1.4}$$

przy czym chwila t_0 jest dowolna.

Definicja 1.2. *Energią* analogowego sygnału deterministycznego $x(t)$ nazywamy wielkość:

$$E_x = \int\limits_{-\infty}^{\infty} x^2(t)\,\mathrm{d}t. \tag{1.5}$$

Definicja 1.3. *Mocą (średnią)* analogowego sygnału deterministycznego $x(t)$ nazywamy wielkość graniczną:

$$P_x = \lim_{T\to\infty} \frac{1}{2T} \int\limits_{-T}^{T} x^2(t)\,\mathrm{d}t. \tag{1.6}$$

W przypadku sygnałów okresowych wzór (1.6) przybiera postać:

$$P_x = \frac{1}{T_0} \int\limits_{t_0}^{t_0+T_0} x^2(t)\,\mathrm{d}t, \tag{1.7}$$

gdzie T_0 jest okresem, a t_0 – dowolną chwilą.

Tak zdefiniowane wielkości energii i mocy sygnału nie mają sensu nadawanego im w fizyce i należy je rozumieć w znaczeniu uogólnionym. Przy przyjętym założeniu bezwymiarowości sygnałów wymiarem energii sygnału jest sekunda, a moc jest bezwymiarowa. Gdyby jednak sygnał był sygnałem napięcia lub prądu, to wydzieliłby na oporze jednostkowym $1\,\Omega$ energię (lub moc) równą liczbowo wielkości (1.5) (lub (1.6)).

Definicja 1.4. *Wartością skuteczną sygnału* jest nazywany pierwiastek z jego mocy $x_{sk} = \sqrt{P_x}$.

Energia i moc charakteryzują właściwości energetyczne sygnału. Na ich podstawie sygnały deterministyczne są dzielone na dwie podstawowe rozłączne klasy.

Definicja 1.5. Sygnał $x(t)$ jest nazywany *sygnałem o ograniczonej energii* , jeśli $0 < E_x < \infty$.

Definicja 1.6. Sygnał $x(t)$ jest nazywany *sygnałem o ograniczonej mocy* , jeśli $0 < P_x < \infty$.

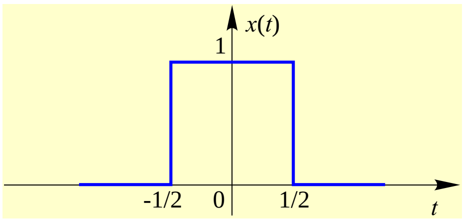
Moc sygnałów o ograniczonej energii jest równa zeru. Energia sygnałów o ograniczonej mocy jest nieskończona. Klasa sygnałów o ograniczonej energii obejmuje oczywiście wszystkie sygnały impulsowe ograniczone w amplitudzie, ale nie tylko. Do klasy tej należą także sygnały o nieskończonym czasie trwania, których wartości maleją dostatecznie szybko w funkcji czasu. Sygnały o ograniczonej mocy i ograniczone w amplitudzie są sygnałami o nieskończonym czasie trwania. Szczególną podklasą tych ostatnich są sygnały okresowe.

1.2.3. Przykłady prostych sygnałów analogowych o ograniczonej energii i skończonym czasie trwania

Impuls prostokątny $\sqcap(t)$

$$x(t) = \sqcap(t) = \begin{cases} 1 & \text{dla } |t| < 1/2, \\ 1/2 & \text{dla } |t| = 1/2, \\ 0 & \text{dla } |t| > 1/2, \end{cases}$$

$$\langle x \rangle = 1, \quad E_x = 1.$$

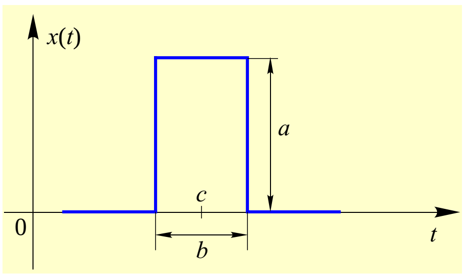


Rys. 1.7. Impuls prostokątny

Symbol specjalny $\sqcap(t)$ oznacza unormowany symetryczny impuls prostokątny o jednostkowym czasie trwania i jednostkowej amplitudzie. Zarówno wartość średnia tego sygnału za czas jego trwania, jak i energia są jednostkowe. Za pomocą tego symbolu można zapisać impuls prostokątny o dowolnej wysokości a, dowolnej szerokości b i przesunięty w czasie o dowolną wartość c (rys. 1.8).

$$x(t) = a \sqcap\left(\frac{t-c}{b}\right),$$

$$\langle x \rangle = a, \quad E_x = a^2 b.$$



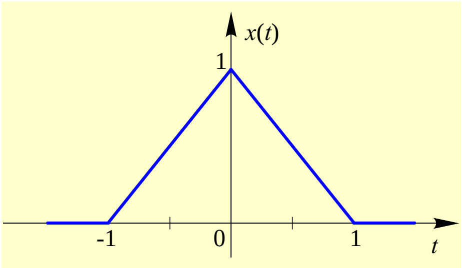
Rys. 1.8. Przesunięty impuls prostokątny

Mnożąc dany sygnał przez $x(t) = \sqcap((t-c)/b)$ można „wyciąć” jego dowolny fragment. Zapis $\sqcap(t/T)$ oznacza symetryczny impuls prostokątny o czasie trwania T.

Impuls trójkątny $\wedge(t)$

$$x(t) = \wedge(t) = \begin{cases} 1-|t| & \text{dla } |t| \leqslant 1, \\ 0 & \text{dla } |t| > 1, \end{cases}$$

$$\langle x \rangle = \frac{1}{2}, \quad E_x = \frac{2}{3}.$$



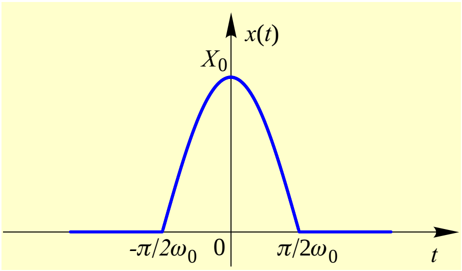
Rys. 1.9. Impuls trójkątny

Symbol specjalny $\wedge(t)$ oznacza unormowany symetryczny impuls trójkątny o czasie trwania równym 2 i wartości w zerze równej 1. Zapis $\wedge(t/T)$ oznacza symetryczny impuls trójkątny o czasie trwania $2T$.

Impuls kosinusoidalny

$$x(t) = X_0 \cos\omega_0 t \sqcap \left(\frac{t}{\pi/\omega_0}\right),$$

$$\langle x \rangle = \frac{2X_0}{\pi}, \quad E_x = \frac{\pi X_0^2}{2\omega_0}.$$

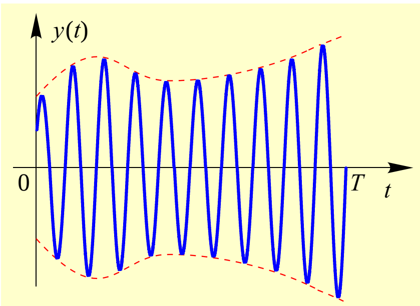


Rys. 1.10. Impuls kosinusoidalny

Sygnał ten jest symetrycznym impulsem obejmującym pół okresu sygnału kosinusoidalnego o amplitudzie X_0 i pulsacji ω_0.

Impuls radiowy

Sygnał ten ma postać $y(t) = x(t)\cos(\omega_0 t + \varphi_0)$, gdzie $x(t)$ jest dowolnym sygnałem impulsowym.



Rys. 1.11. Impuls radiowy

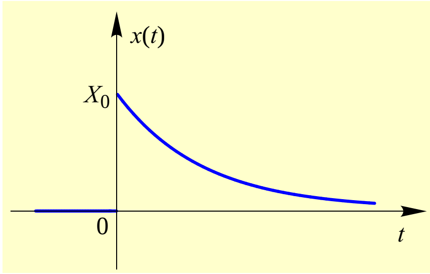
Sygnały takie są wykorzystywane m.in. w systemach radiokomunikacyjnych i radiolokacyjnych. Stąd pochodzi ich nazwa. Sygnał $x(t)$ jest nazywany *obwiednią* impulsu $y(t)$, a funkcja $\cos(\omega_0 t + \varphi_0)$ – jego wypełnieniem. Czas trwania impulsu T jest z reguły wielokrotnie dłuższy od okresu wypełnienia $T_0 = 2\pi/\omega_0$.

1.2.4. Przykłady prostych sygnałów analogowych o ograniczonej energii i nieskończonym czasie trwania

Sygnał wykładniczy malejący

$$x(t) = \begin{cases} X_0\,\mathrm{e}^{-\alpha t} & \text{dla } t \geqslant 0, \\ 0 & \text{dla } t < 0, \end{cases} \quad \alpha > 0,$$

$$\langle x \rangle = 0, \quad E_x = \frac{X_0^2}{2\alpha}.$$



Rys. 1.12. Sygnał wykładniczy malejący

Sygnał sinusoidalny malejący wykładniczo

$$x(t) = \begin{cases} X_0\,\mathrm{e}^{-\alpha t} \sin\omega_0 t & \text{dla } t \geqslant 0, \\ 0 & \text{dla } t < 0, \end{cases} \quad \alpha > 0,$$

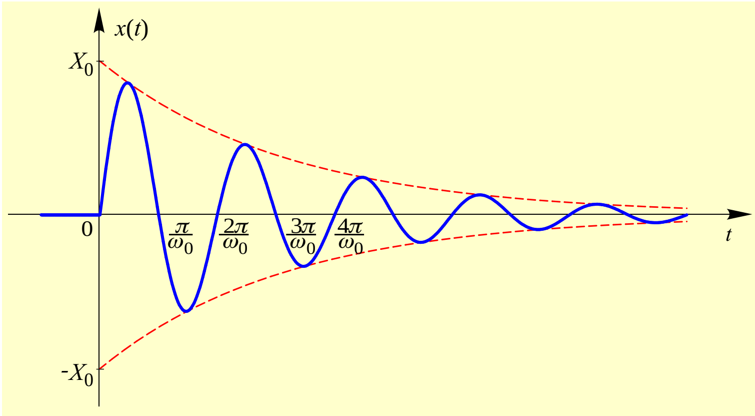
$$\langle x \rangle = 0, \quad E_x = \frac{X_0^2}{4\alpha}\frac{\omega_0^2}{\alpha^2 + \omega_0^2}.$$

Sygnał Sa

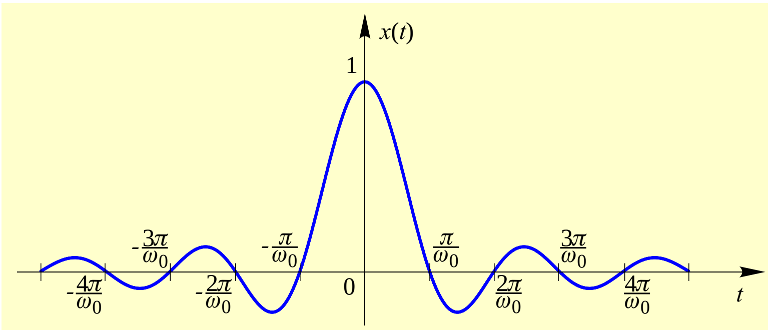
$$x(t) = \mathrm{Sa}\,\omega_0 t = \begin{cases} \dfrac{\sin\omega_0 t}{\omega_0 t} & \text{dla } t \neq 0, \\ 1 & \text{dla } t = 0, \end{cases}$$

$$\langle x \rangle = 0, \quad E_x = \frac{\pi}{\omega_0}.$$

Funkcja $(\sin x)/x$, dobrze znana z analizy matematycznej, nie ma wartości w zerze. W teorii sygnałów przyjmuje się dodatkowo tę wartość za równą 1. Symbol specjalny Sa pochodzi od angielskiego słowa *sampling* (*próbkowanie*). Przyjęcie tego oznaczenia jest uzasadnione szczególną rolą, jaką sygnał Sa odgrywa w zagadnieniach próbkowania sygnałów (por. lekcję 6). W literaturze na oznaczenie tego sygnału jest także stosowany symbol sinc.



Rys. 1.13. Sygnał sinusoidalny malejący wykładniczo

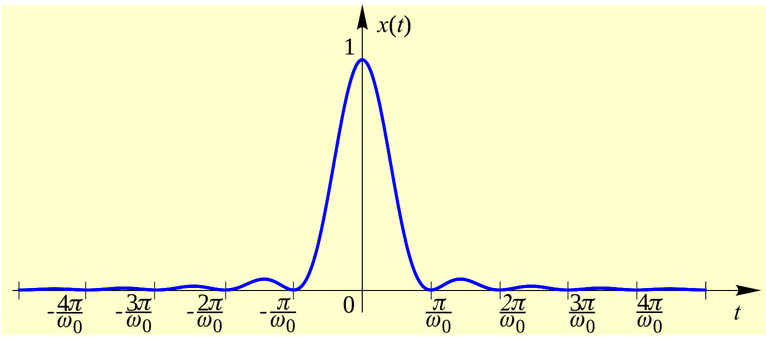


Rys. 1.14. Sygnał Sa

Sygnał Sa^2

$$x(t) = \mathrm{Sa}^2\,\omega_0 t = \begin{cases} \dfrac{\sin^2 \omega_0 t}{(\omega_0 t)^2} & \text{dla } t \neq 0, \\ 1 & \text{dla } t = 0, \end{cases}$$

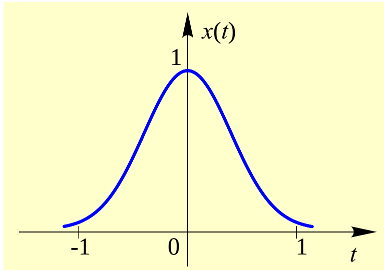
$$\langle x \rangle = 0, \quad E_x = \frac{2}{3}\frac{\pi}{\omega_0}.$$



Rys. 1.15. Sygnał Sa^2

Sygnał Gaussa

$$x(t) = \mathrm{e}^{-\pi t^2},$$
$$\langle x \rangle = 0, \quad E_x = \frac{1}{\sqrt{2}}.$$



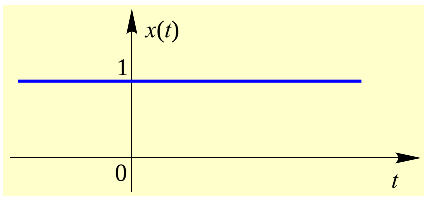
Rys. 1.16. Sygnał Gaussa

Sygnał ten jest opisany funkcją Gaussa o charakterystycznym kształcie dzwonu. Jak pamiętamy, funkcja ta spełnia ważną rolę w teorii prawdopodobieństwa i jest wykorzystywana do charakteryzowania właściwości sygnałów losowych. W teorii sygnałów deterministycznych sygnał Gaussa służy przede wszystkim jako sygnał pomocniczy przy definiowaniu modeli dystrybucyjnych sygnałów (por. p. 1.2.8).

1.2.5. Przykłady prostych nieokresowych sygnałów analogowych o ograniczonej mocy

Sygnał stały

$$x(t) = 1 \qquad \text{dla} \;\; -\infty < t < \infty,$$
$$\langle x \rangle = 1, \quad P_x = 1.$$

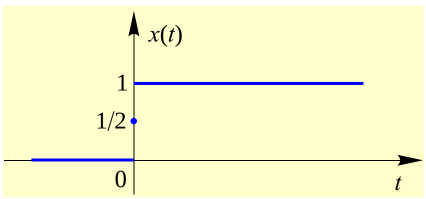


Rys. 1.17. Sygnał stały

Skok jednostkowy $\mathbf{\mathit{1}}(t)$

$$x(t) = \mathbf{\mathit{1}}(t) = \begin{cases} 1 & \text{dla} \;\; t > 0, \\ 1/2 & \text{dla} \;\; t = 0, \\ 0 & \text{dla} \;\; t < 0, \end{cases}$$

$$\langle x \rangle = \frac{1}{2}, \quad P_x = \frac{1}{2}.$$



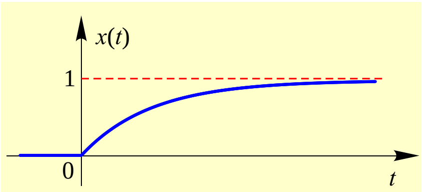
Rys. 1.18. Skok jednostkowy

Zapis $X_0 \mathbf{\mathit{1}}(t - t_0)$ oznacza sygnał skoku o dowolną wartość X_0 przesunięty o dowolny odcinek t_0 osi czasu. Za pomocą symbolu specjalnego $\mathbf{\mathit{1}}(t)$ można w wygodny sposób zapisywać sygnały określone tylko na dodatniej półosi czasu. Na przykład, zapis $\sin\omega_0 t\,\boldsymbol{1}(t)$ oznacza sygnał sinusoidalny rozpoczynający się w chwili $t = 0$.

Sygnał wykładniczy narastający

$$x(t) = (1 - \mathrm{e}^{-\alpha t})\boldsymbol{1}(t), \quad \alpha > 0,$$

$$\langle x\rangle = \frac{1}{2}, \quad P_x = \frac{1}{2}.$$

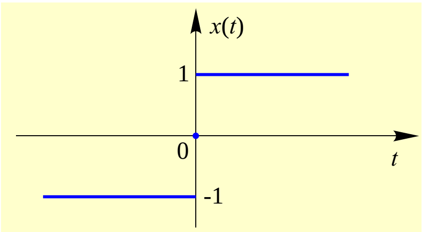


Rys. 1.19. Sygnał wykładniczy narastający

Sygnał sgn *t*

$$x(t) = \operatorname{sgn} t = \begin{cases} 1 & \text{dla } t > 0, \\ 0 & \text{dla } t = 0, \\ -1 & \text{dla } t < 0. \end{cases}$$

$$\langle x\rangle = 0, \quad P_x = 1.$$



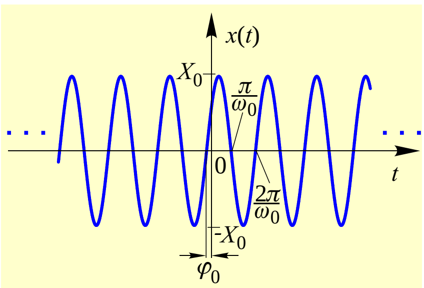
Rys. 1.20. Sygnał znaku $\operatorname{sgn} t$

1.2.6. Przykłady prostych okresowych sygnałów analogowych o ograniczonej mocy

Sygnał harmoniczny

$$x(t) = X_0 \sin(\omega_0 t + \varphi_0),$$
$$-\infty < t < \infty,$$

$$\langle x\rangle = 0, \quad P_x = \frac{1}{2}X_0^2.$$

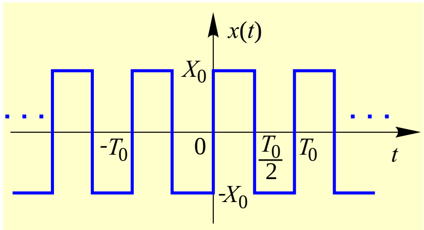


Rys. 1.21. Sygnał harmoniczny

Sygnał ten jest określony przez trzy parametry: amplitudę X_0, pulsację $\omega_0 = 2\pi f_0 = 2\pi/T_0$, gdzie f_0 jest częstotliwością, a T_0 – okresem, oraz fazę początkową φ_0. Spełnia on w teorii sygnałów szczególnie ważną rolę.

Fala prostokątna bipolarna

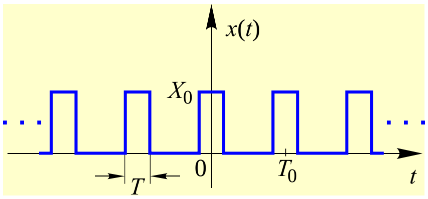
$$\langle x \rangle = 0, \quad P_x = X_0^2.$$



Rys. 1.22. Fala prostokątna bipolarna

Fala prostokątna unipolarna

$$\langle x \rangle = \frac{T}{T_0} X_0, \quad P_x = \frac{T}{T_0} X_0^2.$$



Rys. 1.23. Fala prostokątna unipolarna

Iloraz T/T_0 jest nazywany *współczynnikiem wypełnienia.*

1.2.7. Sygnały zespolone

W teorii sygnałów – obok modeli rzeczywistych – stosowane są często zespolone modele sygnałów. O wyborze modelu rzeczywistego lub zespolonego decydują przede wszystkim względy wygody matematycznej. Modele zespolone, mimo abstrakcyjnego charakteru, umożliwiają znaczne uproszczenie analizy, zwłaszcza w przypadku sygnałów złożonych.

Sygnał zespolony ma postać:

$$z(t) = x(t) + \mathrm{j}\, y(t), \tag{1.8}$$

gdzie $x(t) = \mathrm{Re}\, z(t)$ oraz $y(t) = \mathrm{Im}\, z(t)$ są sygnałami rzeczywistymi. Wzór (1.8) opisuje postać algebraiczną sygnału. Podobnie jak każdą liczbę zespoloną, sygnał zespolony można także zapisać w postaci biegunowej:

$$z(t) = |z(t)|\, \mathrm{e}^{\mathrm{j}\varphi(t)}, \tag{1.9}$$

gdzie $|z(t)| = \sqrt{x^2(t) + y^2(t)}$ jest modułem, a $\varphi(t) = \mathrm{arctg}[y(t)/x(t)]$ – argumentem sygnału. Sygnał:

$$z^*(t) = x(t) - \mathrm{j}\, y(t) = |z(t)|\, \mathrm{e}^{-\mathrm{j}\varphi(t)}\,. \tag{1.10}$$

nazywamy *sygnałem sprzężonym* z sygnałem $z(t)$.

Również sygnały zespolone dzielimy na sygnały o ograniczonej energii i ograniczonej mocy. Energia i moc sygnałów zespolonych są zdefiniowane identycznie jak w przypadku sygnałów rzeczywistych, z tym, że we wzorach definicyjnych (1.5)–(1.7) zamiast kwadratu sygnału $x^2(t)$ należy podstawić kwadrat modułu $|x(t)|^2$. Rozpatrzymy dwa przykłady.

Zespolony sygnał harmoniczny

$$\mathrm{e}^{\mathrm{j}\omega_0 t} = \cos\omega_0 t + \mathrm{j}\sin\omega_0 t, \quad t \in (-\infty, \infty). \tag{1.11}$$

Sygnał ten, nazywany niekiedy *zespoloną sinusoidą*, jest często wykorzystywany jako zespolona reprezentacja rzeczywistego sygnału harmonicznego. Można łatwo sprawdzić, że jest on sygnałem okresowym o okresie $T_0 = 2\pi/\omega_0$:

$$\mathrm{e}^{\mathrm{j}\omega_0(t+T_0)} = \mathrm{e}^{\mathrm{j}\omega_0 t}\,\mathrm{e}^{\mathrm{j}\omega_0 T_0} = \mathrm{e}^{\mathrm{j}\omega_0 t}\,\mathrm{e}^{\mathrm{j}2\pi} = \mathrm{e}^{\mathrm{j}\omega_0 t}$$

i sygnałem o ograniczonej mocy:

$$P_x = \frac{1}{T_0}\int\limits_0^{T_0} \left|\mathrm{e}^{\mathrm{j}\omega_0 t}\right|^2 \mathrm{d}t = \frac{1}{T_0}\int\limits_0^{T_0} 1\,\mathrm{d}t = 1.$$

Sygnał analityczny

Sygnał analityczny stanowi szczególnego rodzaju zespoloną reprezentację sygnału rzeczywistego, często stosowaną w zagadnieniach modulacji sygnałów. Sygnałem analitycznym, reprezentującym rzeczywisty sygnał $x(t)$, nazywamy sygnał zespolony:

$$z_x(t) = x(t) + \mathrm{j}\,\hat{x}(t), \tag{1.12}$$

którego częścią rzeczywistą jest sygnał $x(t)$, a częścią urojoną – transformata Hilberta $\hat{x}(t)$ tego sygnału określona wzorem:

$$\hat{x}(t) = \frac{1}{\pi}\int\limits_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau}\,\mathrm{d}\tau, \tag{1.13}$$

gdzie całka jest rozumiana w sensie wartości głównej Cauchy'ego.

Podobnie jak przekształcenia Fouriera i Laplace'a, przekształcenie Hilberta jest przekształceniem całkowym odwzorowującym dany sygnał w inną wielkość funkcyjną. Cechą charakterystyczną przekształcenia Hilberta jest to, iż w odróżnieniu od przekształceń Fouriera i Laplace'a przekształca ono funkcję zmiennej t (sygnał) w inną funkcję zmiennej t (inny sygnał), a więc w wielkość funkcyjną w tej samej dziedzinie.

Przekształcenie Hilberta i pojęcie sygnału analitycznego omówimy dokładniej w lekcji 10(p. 10.1) poświęconejsygnałom zmodulowanym. W tym miejscu ograniczymy się jedynie do podania przykładu. Można wykazać, że transformata Hilberta sygnału $\cos\omega_0$ jest równa $\sin\omega_0 t$. Tak więc sygnałem analitycznym, reprezentującym rzeczywisty sygnał harmoniczny $\cos\omega_0 t$, jest zespolony sygnał $\mathrm{e}^{\mathrm{j}\omega_0 t}$.

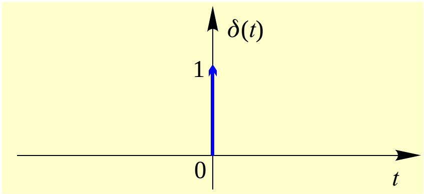
1.2.8. Sygnały dystrybucyjne

W wielu zagadnieniach teorii sygnałów bardzo użytecznymi modelami sygnałów są wielkości matematyczne zwane *dystrybucjami* (*funkcjami uogólnionymi*). Dystrybucje nie są funkcjami w sensie przyjętym w klasycznej analizie matematycznej i są definiowane w sposób odmienny niż zwykłe funkcje. Najważniejszą z nich jest *impuls Diraca* $\delta(t)$ (nazywany także *dystrybucją Diraca* lub *deltą Diraca*).

Impuls Diraca

$$\delta(t) = \begin{cases} 0 & \text{dla } t \neq 0, \\ \infty & \text{dla } t = 0. \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(t)\,\mathrm{d}t = 1.$$



Rys. 1.24. Impuls Diraca

Impuls Diraca jest modelem nierealizowalnego fizycznie nieskończenie wąskiego sygnału, o nieskończenie dużej amplitudzie i polu równym 1. Przyjętym powszechnie symbolem graficznym impulsu Diraca jest prążek zakończony strzałką umieszczony w punkcie $t = 0$, którego wysokość jest równa polu impulsu (rys. 1.24). Symbol $X_0\delta(t - t_0)$ oznacza impuls Diraca o polu X_0 występujący w chwili t_0.

Impuls Diraca opisuje gęstość amplitudy przypadającej na jednostkę czasu. Jej miarą jest pole impulsu. W przeciwieństwie do poprzednio omawianych sygnałów, które przy przyjętej konwencji są bezwymiarowe, impuls Diraca ma wymiar $1/s$, a więc taki sam jak częstotliwość. Należy także podkreślić, że z formalnego punktu widzenia impuls Diraca nie należy do klasy sygnałów o ograniczonej mocy, jego moc jest bowiem nieskończona.

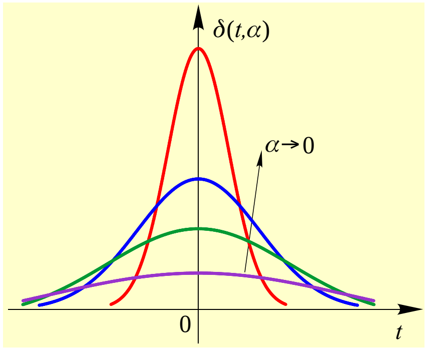
Podana definicja dystrybucji Diraca nie spełnia wymogu ścisłości matematycznej. Dla potrzeb analizy sygnałów konieczne jest przyjęcie definicji bardziej ścisłej, w oparciu o którą można byłoby poprawnie zdefiniować operacje wykonywane na sygnałach z udziałem dystrybucji Diraca. W tzw. elementarnej teorii dystrybucji dystrybucję Diraca rozumie się jako granicę ciągu $\{\delta(t, \alpha)\}$ zwykłych funkcji $\delta(t, \alpha)$, gdzie $\alpha > 0$ jest parametrem, spełniającego warunki:

$$\bigwedge_{\alpha} \int_{-\infty}^{\infty} \delta(t, \alpha)\,\mathrm{d}t = 1, \qquad \lim_{\alpha \to 0} \delta(t, \alpha) = \delta(t). \tag{1.14}$$

Ciąg $\{\delta(t, \alpha)\}$ jest nazywany *ciągiem aproksymującym* dystrybucję $\delta(t)$. Ciągów takich można utworzyć wiele (por. [3], p. 2.3.6). Jako przykład może służyć ciąg funkcji Gaussa (por. rys. 1.16).

Ciąg funkcji Gaussa aproksymujący impuls Diraca

$$\delta(t, \alpha) = \frac{1}{\sqrt{2\pi\alpha^2}} \exp\Big(-\frac{t^2}{2\alpha^2}\Big).$$



Rys. 1.25. Ciąg funkcji Gaussa aproksymujący impuls Diraca

W miarę jak α dąży do zera funkcje te stają się coraz węższe, a ich wartość maksymalna w chwili $t = 0$ rośnie do nieskończoności. W granicy, gdy $\alpha \to 0$, otrzymujemy impuls Diraca.

Właściwości impulsu Diraca

Omówimy podstawowe właściwości impulsu Diraca, na które będziemy się dalej często powoływać.

1. *Właściwość próbkowania*

$$x(t)\delta(t - t_0) = x(t_0)\delta(t - t_0). \tag{1.15}$$

W wyniku mnożenia dowolnego sygnału przez impuls Diraca $\delta(t - t_0)$ występujący w chwili $t = t_0$ otrzymujemy impuls Diraca w tej samej chwili o polu równym wartości (próbce) tego sygnału w chwili $t = t_0$. Wyrażenie $x(t_0)\delta(t-t_0)$ po prawej stronie równości (1.15) można przyjąć za reprezentację dystrybucyjną próbki $x(t_0)$.

2. *Właściwość filtracji*

$$\int\limits_{-\infty}^{\infty} x(t)\delta(t - t_0)\,\mathrm{d}t = x(t_0). \tag{1.16}$$

Właściwość ta jest konsekwencją właściwości próbkowania. Całka iloczynu dowolnego sygnału $x(t)$ i impulsu Diraca $\delta(t - t_0)$ występującego w chwili $t = t_0$ jest równa próbce $x(t_0)$ tego sygnału w tej chwili.

3. *Związki ze skokiem jednostkowym*

$$\int\limits_{-\infty}^{t} \delta(t')\,\mathrm{d}t' = \boldsymbol{1}(t), \tag{1.17}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{1}(t) = \delta(t). \tag{1.18}$$

Różniczkowanie i całkowanie należy tu rozumieć w sensie dystrybucyjnym, tj. jako operacje na odpowiednich ciągach aproksymujących, a otrzymaną równość – jako związek między granicami tych ciągów.
4. *Właściwość splotu*

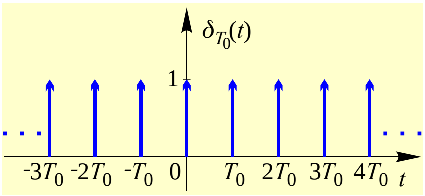
$$x(t) * \delta(t) = x(t), \quad x(t) * \delta(t - t_0) = x(t - t_0). \tag{1.19}$$

Splot sygnału $x(t)$ z dystrybucją Diraca $\delta(t)$ daje w wyniku ten sam sygnał $x(t)$. Wynika stąd, że $\delta(t)$ jest elementem identycznościowym operacji splotu. W przypadku splatania z dystrybucją przesuniętą w czasie o t_0 otrzymujemy kopię sygnału przesuniętą o t_0. Właściwość splotu jest czasami nazywana właściwością powtarzania.

Komentarz. Pojęcie splotu jest jednym z fundamentalnych pojęć teorii sygnałów. Będzie ono dalej często występować w różnych fragmentach tekstu, przy czym operacja splotu będzie rozpatrywana zarówno w dziedzinie czasu, jako operacja na sygnałach, jak i w dziedzinie częstotliwości, jako operacja na widmach. Będziemy przy tym zakładać, że definicja splotu oraz jego podstawowe własności są znane. Obszerne omówienie operacji splotu można znaleźć w [4], p. 9.1.1 i 9.1.2.

Okresowy ciąg impulsów Diraca (dystrybucja grzebieniowa)

$$\delta_{T_0}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0)$$



Rys. 1.26. Dystrybucja grzebieniowa

Dystrybucja grzebieniowa jest – obok impulsu Diraca – modelem dystrybucyjnym najczęściej wykorzystywanym w analizie sygnałów. Dystrybucja ta jest okresowym ciągiem impulsów Diraca powtarzanych z okresem T_0. Jej wykres przypomina nieskończony grzebień, którego „zęby" są równoodległe i mają jednakową wysokość. Uzasadnia to przyjętą nazwę tej dystrybucji. W literaturze dystrybucja grzebieniowa jest nazywana także *dystrybucją sza*, *dystrybucją comb* lub *idealnym sygnałem próbkującym*.

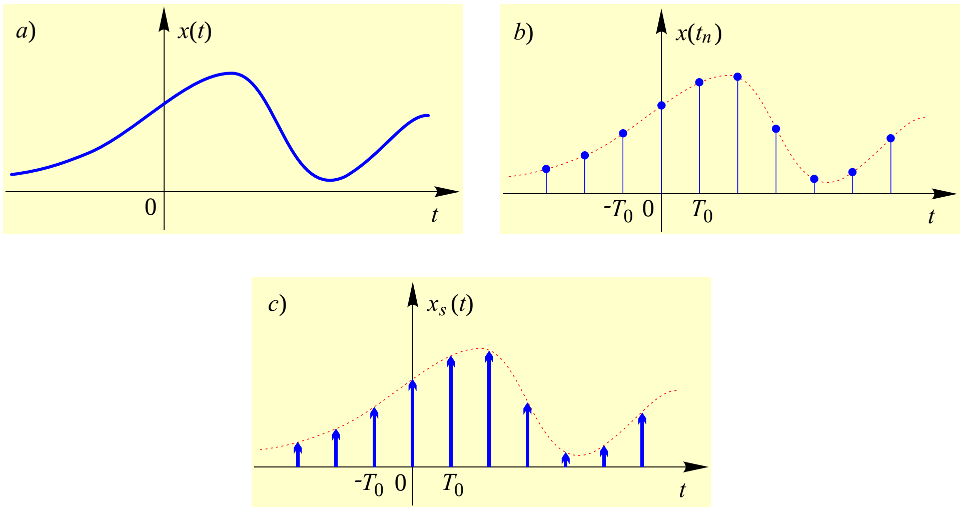
Właściwości dystrybucji grzebieniowej

1. *Właściwość próbkowania*

$$x(t)\delta_{T_0}(t) = \sum_{n=-\infty}^{\infty} x(nT_0)\delta(t - nT_0). \tag{1.20}$$

W wyniku mnożenia dowolnego sygnału $x(t)$ przez dystrybucję grzebieniową $\delta_{T_0}(t)$ otrzymujemy ciąg powtarzanych z okresem T_0 impulsów Diraca o wysokościach określonych przez próbki $x(nT_0)$ sygnału. Obwiednią tych impulsów jest sygnał $x(t)$. Mówiąc obrazowo w wyniku tej operacji otrzymujemy grzebień, którego wierzchołki zębów układają się w kształt sygnału.

Sygnał dystrybucyjny $x_s(t) = \sum_{n=-\infty}^{\infty} x(nT_0)\delta(t - nT_0)$. stanowi dystrybucyjną reprezentację spróbkowanego sygnału $x(t)$. Będziemy go nazywać *impulsowym sygnałem spróbkowanym* (rys. 1.27).

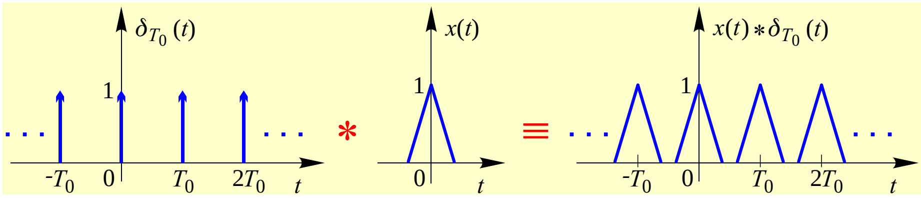


Rys. 1.27. Sygnał (a), jego wersja spróbkowana (b) i reprezentacja za pomocą impulsowego sygnału spróbkowanego (c)

2. *Właściwość powielenia okresowego*

$$x(t) * \delta_{T_0}(t) = \sum_{n=-\infty}^{\infty} x(t - nT_0). \tag{1.21}$$

W wyniku splecenia dowolnego sygnału $x(t)$ z dystrybucją grzebieniową $\delta_{T_0}(t)$ następuje powielenie okresowe tego sygnału z okresem T_0. W przypadku, gdy sygnał $x(t)$ jest sygnałem impulsowym o czasie trwania krótszym niż okres dystrybucji grzebieniowej T_0, sygnał powielony jest ciągiem wiernych kopii sygnału $x(t)$ powtórzonych co przedział T_0 (rys. 1.28). Można obrazowo powiedzieć, że w efekcie splecenia sygnału $x(t)$ z dystrybucją grzebieniową na każdym jej zębie pozostawia on swój wierny ślad.



Rys. 1.28. Powielenie okresowe sygnału impulsowego

1.3. Sygnały dyskretne

Sygnał dyskretny jest określony jedynie w dyskretnym zbiorze chwil $\{\dots t_{-1}, t_0, t_1, t_2, \dots\}$ i zdefiniowany jako ciąg swoich wartości w tych chwilach. Sygnały dyskretne występujące w technice otrzymujemy z reguły w wyniku próbkowania sygnałów analogowych, tj. rejestracji wartości sygnału analogowego w określonych chwilach. Najczęściej rejestracja próbek lub – jak mówimy – ich pobieranie następuje w chwilach jednakowo od siebie odległych. Próbkowanie nazywamy wówczas *równomiernym*.

Zauważmy, że w sensie przytoczonej definicji sygnałami dyskretnymi są także: ciąg ocen uzyskany przez studenta podczas studiów, ciąg notowań kursu złotego względem dolara podawany codziennie w gazecie, ciąg corocznych pomiarów położenia pewnej gwiazdy względem Ziemi, czy też ciąg odczytów wodowskazu mierzącego stan wody w rzece, dokonywanych w określonych godzinach. W pierwszych dwóch przypadkach sygnały dyskretne nie mają swoich „pierwowzorów" analogowych. Natomiast w dwóch ostatnich przypadkach sygnały dyskretne powstają w wyniku próbkowania wielkości analogowych, gdyż zarówno położenie gwiazdy, jak i poziom wody w rzece zmienia się płynnie w czasie.

W dalszym ciągu będziemy rozpatrywać wyłącznie sygnały dyskretne spróbkowane równomiernie, a więc określone w chwilach $t_n = nT_s$, gdzie T_s jest *przedziałem dyskretyzacji*. Przedział ten jest nazywany również *okresem* lub *przedziałem próbkowania*. Odwrotność okresu próbkowania $f_s = 1/T_s$ jest nazywana *częstotliwością próbkowania*. Ponieważ w analizie formalnej sygnałów dyskretnych nie jest istotne w jakiej chwili t_n jest pobierana próbka, a istotny jest jedynie jej kolejny numer w ciągu próbek, za argument sygnałów dyskretnych będziemy przyjmować zmienną bezwymiarową $n = t_n/T_s$, tj. czas unormowany względem okresu próbkowania. Dyskretne sygnały deterministyczne będziemy oznaczać $x[n], y[n], z[n], \dots$, a ich próbki – $x(n), y(n), z(n), \dots$

1.3.1. Parametry

Wzory definiujące podstawowe parametry dyskretnych sygnałów deterministycznych są odpowiednikami wzorów (1.2)–(1.7) dla sygnałów ciągłych i mają postać sum.

Definicja 1.7. Wartość średnia dyskretnego impulsowego sygnału deterministycznego $x[n]$, określonego w skończonym przedziale $[n_1, n_2]$, jest stosunkiem sumy wszystkich próbek w tym przedziale do liczby tych próbek:

$$\langle x \rangle = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} x(n). \tag{1.22}$$

W przypadku sygnałów o nieskończonym czasie trwania wartość średnia jest określona wzorem:

$$\langle x \rangle = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n). \tag{1.23}$$

W szczególnym przypadku dyskretnych sygnałów okresowych o okresie N_0 wzór (1.23) przybiera postać (n_0 jest dowolną liczbą całkowitą):

$$\langle x \rangle = \frac{1}{N_0} \sum_{n=n_0}^{n_0+N_0-1} x(n). \tag{1.24}$$

Definicja 1.8. Energią dyskretnego sygnału deterministycznego $x[n]$ nazywamy wielkość:

$$E_x = \sum_{n=\infty}^{\infty} x^2(n), \tag{1.25}$$

a mocą – wielkość graniczną:

$$P_x = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{-N}^{N} x^2(n). \tag{1.26}$$

Jeśli sygnał jest okresowy, moc można obliczać ze wzoru równoważnego:

$$P_x = \frac{1}{N_0} \sum_{n_0}^{n_0+N_0-1} x^2(n), \tag{1.27}$$

gdzie N_0 jest okresem, a n_0 dowolną liczbą całkowitą. Wielkość $x_{sk} = \sqrt{P_x}$ jest nazywana wartością skuteczną sygnału $x[n]$.

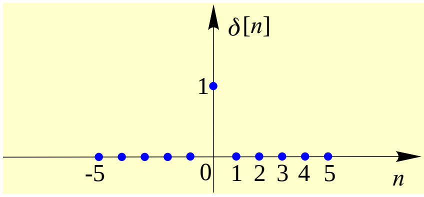
Podobnie jak sygnały ciągłe, dyskretne sygnały deterministyczne dzielimy na dwie podstawowe klasy. Jeśli $0 < E_x < \infty$, sygnał $x[n]$ nazywamy sygnałem o ograniczonej energii. Jeśli $0 < P_x < \infty$, sygnał $x[n]$ nazywamy sygnałem o ograniczonej mocy.

1.3.2. Przykłady prostych sygnałów dyskretnych o ograniczonej energii i skończonym czasie trwania

Impuls (delta) Kroneckera $\delta[n]$

$$x[n] = \delta[n] = \begin{cases} 1 & \text{dla } n = 0, \\ 0 & \text{dla } n \neq 0. \end{cases}$$

$$\langle x \rangle = 1, \quad E_x = 1.$$



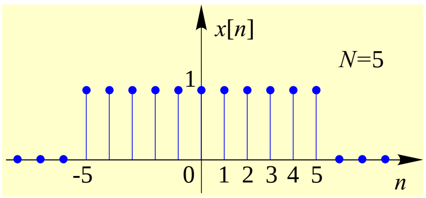
Rys. 1.29. Delta Kroneckera

Delta (impuls) Kroneckera jest dla sygnałów dyskretnych odpowiednikiem delty Diraca dla sygnałów analogowych. W przeciwieństwie do delty Diraca jest to zwykła funkcja. Symbol $X_0\delta[n-n_0]$ oznacza sygnał o wartości X_0 w punkcie $n = n_0$ i pozostałych wartościach równych zeru.

Impuls prostokątny

$$x[n] = \begin{cases} 1 & \text{dla } n \leqslant |N|, \\ 0 & \text{dla } n > |N|. \end{cases}$$

$$\langle x \rangle = 1, \quad E_x = 2N + 1.$$

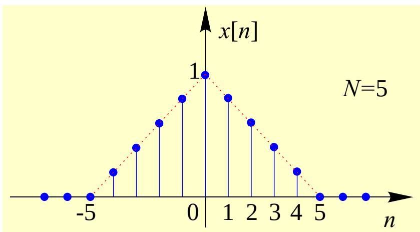


Rys. 1.30. Dyskretny impuls prostokątny

Impuls trójkątny

$$x[n] = \begin{cases} 1 - \dfrac{|n|}{N} & \text{dla } n \leqslant |N|, \\ 0 & \text{dla } n > |N|. \end{cases}$$

$$\langle x \rangle = \frac{N}{2N+1}, \quad E_x = \frac{2N^2+1}{3N}.$$



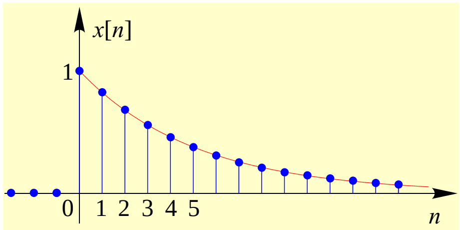
Rys. 1.31. Dyskretny impuls trójkątny

Na rys. 1.31 przedstawiono wykres tego sygnału dla $N = 5$. W tym przypadku $\langle x \rangle = 5/11$ oraz $E_x = 17/5$.

1.3.3. Przykłady prostych sygnałów dyskretnych o ograniczonej energii i nieskończonym czasie trwania

Sygnał wykładniczy

$$x[n] = a^n, \quad n \geqslant 0, \quad 0 < a < 1.$$
$$\langle x \rangle = 0, \quad E_x = \frac{1}{1-a^2}.$$

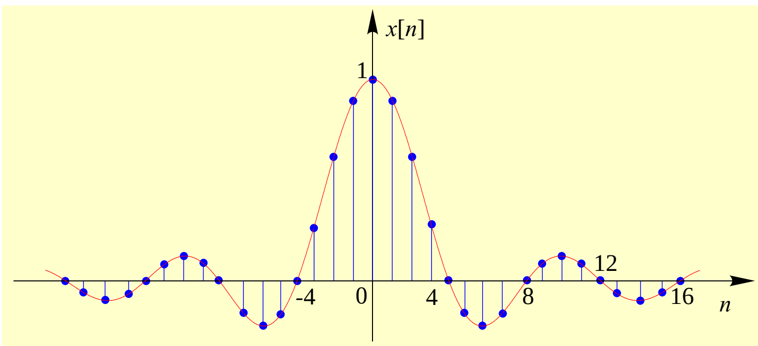


Rys. 1.32. Dyskretny sygnał wykładniczy

Sygnał Sa

$$x[n] = \mathrm{Sa}(n\theta_0) = \begin{cases} \dfrac{\sin n\theta_0}{n\theta_0} & \text{dla } n \neq 0, \\ 1 & \text{dla } n = 0. \end{cases}$$

$$\langle x \rangle = 0, \qquad E_x = \frac{\pi}{\theta_0}.$$



Rys. 1.33. Dyskretny sygnał Sa

Bezwymiarowy parametr θ_0 ma tu znaczenie *pulsacji unormowanej*. Ponieważ parametr ten pojawia się po raz pierwszy, podamy przy tej okazji jego interpretację.

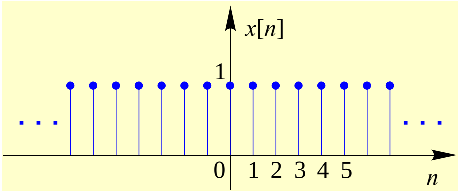
Dyskretny sygnał Sa otrzymuje się jako wynik próbkowania ciągłego sygnału $\mathrm{Sa}(2\pi f_0 t)$ w chwilach $t_n = nT_s$, gdzie T_s jest okresem próbkowania. Po spróbkowaniu sygnału ciągłego otrzymujemy sygnał dyskretny $x[n] = \mathrm{Sa}(n2\pi f_0 T_s)$.

W teorii sygnałów dyskretnych normuje się zwykle nie tylko skalę czasu względem okresu próbkowania T_s, lecz również skalę częstotliwości względem częstotliwości próbkowania $f_s = 1/T_s$. W tym celu wprowadza się pojęcie *częstotliwości unormowanej* $\nu_0 = f_0/f_s = f_0 T_s$. Posługując się tym parametrem, możemy zapisać dyskretny sygnał Sa w postaci $x[n] = \mathrm{Sa}(n2\pi\nu_0)$. Definiując z kolei unormowaną pulsację jako $\theta_0 = 2\pi\nu_0 = 2\pi f_0 T_s = \omega_0 T_s$, otrzymujemy podany wyżej zapis dyskretnego sygnału Sa. Z uwagi na związek $\theta_0 = \omega_0 T_s$ mówimy, że pulsację θ_0 otrzymujemy w wyniku unormowania pulsacji ω_0 względem okresu próbkowania T_s. Wykres dyskretnego sygnału Sa przedstawiony na rys. 1.33 został sporządzony dla $\theta_0 = \pi/4$.

1.3.4. Przykłady prostych sygnałów dyskretnych o ograniczonej mocy

Sygnał stały

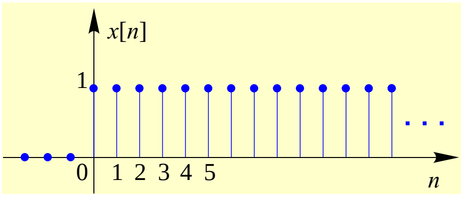
$$x[n] = 1 \quad \text{dla } -\infty < n < \infty,$$
$$\langle x \rangle = 1, \quad P_x = 1.$$



Rys. 1.34. Dyskretny sygnał stały

Skok jednostkowy $\boldsymbol{1}[n]$

$$x[n] = \boldsymbol{1}[n] = \begin{cases} 1 & \text{dla } n \geqslant 0, \\ 0 & \text{dla } n < 0, \end{cases}$$
$$\langle x \rangle = \frac{1}{2}, \quad P_x = \frac{1}{2}.$$

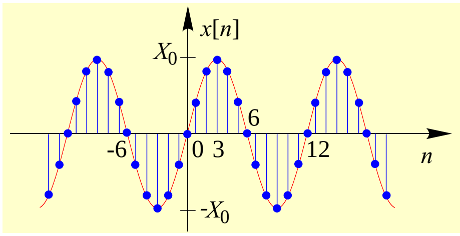


Rys. 1.35. Dyskretny skok jednostkowy

Zapis $X_0\boldsymbol{1}[n - n_0]$ oznacza sygnał skoku o dowolna wartość X_0 w punkcie n_0 osi czasu.

Sygnał harmoniczny

$$x[n] = X_0 \sin(n\theta_0 + \varphi_0),$$
$$-\infty < n < \infty,$$
$$\langle x \rangle = 0, \quad P_x = \frac{1}{2}, X_0^2.$$



Rys. 1.36. Dyskretny sygnał harmoniczny

Dyskretny sygnał harmoniczny, nazywany też *dyskretną sinusoidą*, otrzymujemy w wyniku próbkowania ciągłego sygnału harmonicznego $x(t) = X_0 \sin(\omega_0 t+ +\varphi_0)$. Parametr $\theta_0 = \omega_0 T_s$ jest jego unormowaną pulsacją. Dyskretny sygnał harmoniczny nie zawsze jest sygnałem okresowym zmiennej n. Warunkiem okresowości jest, aby liczba $2\pi/\theta_0$ była liczbą wymierną. Istnieje wtedy taka liczba całkowita k, że $2\pi k/\theta_0$ jest liczbą całkowitą. Okresem (podstawowym) jest wówczas najmniejsza liczba n_0 taka, że $n_0 = 2\pi k/\theta_0$ dla pewnego k. W przypadku, gdy $2\pi/\theta_0$ jest liczbą niewymierną, dyskretny sygnał harmoniczny nie ma okresu (jest tzw. sygnałem *prawie okresowym*). Wykres na rys. 1.36 został sporządzony dla $\theta_0 = \pi/6$ i $\varphi_0 = 0$.

Słownik

częstotliwość unormowana
częstotliwość unormowana względem okresu próbkowania

dwubity
dwuelementowe ciągi znaków binarnych

dystrybucja
wielkość matematyczna nie będącą funkcją, wykorzystywana jako model matematyczny sygnału

dystrybucja grzebieniowa
okresowy ciąg impulsów Diraca

dziedzina czasu
naturalna dziedzina opisu i analizy sygnałów i układów w funkcji zmiennej t – czasu

fala prostokątna bipolarna
: analogowy okresowy sygnał binarny przybierający dwie wartości: X_0 w pierwszej części okresu i $-X_0$ – w drugiej części okresu

fala prostokątna unipolarna
: analogowy okresowy sygnał binarny przybierający dwie wartości 0 lub X_0; okresowy ciąg impulsów prostokątnych o jednakowej amplitudzie i jednakowym czasie trwania, z reguły znacznie mniejszym od okresu powtarzania impulsów

impuls
: sygnał o skończonym czasie trwania

impuls (delta, dystrybucja) Diraca
: wielkość niefunkcyjna, przybierająca wartości zerowe dla chwil $t \neq 0$ i wartość nieskończenie dużą w chwili $t = 0$, której pole (całka) jest równa jedności; przyjmowana jako model matematyczny nieskończenie wąskiego impulsu o nieskończonej amplitudzie

impuls (delta) Kroneckera
: sygnał dyskretny przybierający wartość 1 w chwili $n = 0$ i wartości zerowe w pozostałych chwilach

impuls prostokątny
: sygnał impulsowy o stałej wartości w przedziale jego określoności

impuls radiowy
: sygnał oscylacyjny o wysokiej częstotliwości oscylacji i wolnozmieniającej się obwiedni (amplitudy chwilowej)

impulsowy sygnał spróbkowany
: model matematyczny sygnału spróbkowanego w postaci ciągu impulsów Diraca o polach równych wartościom kolejnych próbek

modulacja QPSK
: modulacja cyfrowa z czterowartościowym kluczowaniem fazy

próbka sygnału
: wartość sygnału w określonej chwili próbkowania

próbkowanie
: operacja na sygnale polegająca na pobraniu (zarejestrowaniu) wartości sygnału analogowego w dyskretnych chwilach, najczęściej równoodległych; w wyniku próbkowania sygnał analogowy jest przetwarzany w sygnał dyskretny

próbkowanie równomierne
: próbkowanie ze stałym odstępem między kolejnymi próbkami

przedział (okres) próbkowania (dyskretyzacji)
długość odcinka czasu między dwiema kolejnymi próbkami przy założeniu próbkowania równomiernego

przekształcenie Hilberta
przekształcenie całkowe przyporządkowujące sygnałowi inny sygnał (transformatę Hilberta)

pulsacja unormowna
pulsacja unormowana względem okresu próbkowania

reprezentacja sygnału
przedstawienie sygnału za pomocą innych wielkości, np. wektora, innego sygnału, sumy ważonej innych sygnałów lub funkcji określonej w innej dziedzinie (częstotliwości, zespolonej)

splot w dziedzinie czasu
operacja na parze sygnałów przyporządkowująca im inny sygnał

sygnał
proces zmian w czasie pewnej wielkości fizycznej lub stanu obiektu fizycznego

sygnał analityczny
reprezentacja sygnału w postaci sygnału zespolonego, którego częścią rzeczywistą jest dany sygnał, a częścią urojoną jego transformata Hilberta

sygnał analogowy
sygnał określony w ciągłym przedziale czasu i przybierający wartości w zbiorze ciągłym

sygnał binarny
sygnał, który w każdej chwili, w której jest określony, przybiera jedną z dwóch wartości

sygnał cyfrowy
sygnał określony w dyskretnych chwilach czasu i przybierający w tych chwilach wartości ze zbioru skończonego

sygnał deterministyczny
sygnał, którego wartości w każdej chwili są określone jednoznacznie

sygnał dyskretny
sygnał określony w dyskretnych chwilach czasu (najczęściej równoodległych) i nieokreślony w pozostałych chwilach

sygnał dystrybucyjny
sygnał, którego przebieg jest opisany dystrybucją

sygnał harmoniczny
sygnał, którego przebieg w czasie jest opisany funkcją sinusoidalną (ciągłą lub dyskretną) o ustalonej amplitudzie, częstotliwości i fazie początkowej

sygnał impulsowy (o skończonym czasie trwania)
sygnał określony w skończonym przedziale czasu (ciągłym w przypadku sygnałów analogowych lub dyskretnym w przypadku sygnałów dyskretnych)

sygnał losowy (stochastyczny)
sygnał, którego wartości w każdej chwili są zmiennymi losowymi i którego ewolucja w czasie następuje według określonych praw probabilistycznych

sygnał o nieskończonym czasie trwania
sygnał określony w nieskończonym przedziale czasu (najczęściej na dodatniej osi czasu bądź całej osi czasu w przypadku sygnałów analogowych lub na zbiorze liczb naturalnych bądź zbiorze wszystkich liczb całkowitych w przypadku sygnałów dyskretnych

sygnał o ograniczonej energii
sygnał, którego energia jest skończona

sygnał o ograniczonej mocy
sygnał, którego moc jest skończona

sygnał quasi-deterministyczny
sygnał o parametrach będących zmiennymi losowymi (np. sygnał harmoniczny o losowej fazie)

sygnał rzeczywisty
sygnał, którego przebieg jest opisany funkcją rzeczywistą czasu

sygnał zespolony
sygnał, którego przebieg jest opisany funkcją zespoloną czasu

szum
sygnał losowy, z reguły szerokopasmowy, zakłócający sygnał użyteczny

transformata Hilberta
sygnał otrzymany w wyniku przekształcenia Hilberta danego sygnału

trygonometryczny szereg Fouriera
przedstawienie sygnału okresowego jako sumy ważonej (w szczególnych przypadkach skończonej) sygnałów harmonicznych o częstotliwościach będących wielokrotnościami częstotliwości podstawowej równej odwrotności okresu sygnału

wartość średnia, energia, moc, wartość skuteczna sygnału
parametry charakteryzujące sygnał

współczynnik wypełnienia
stosunek czasu trwania impulsu do okresu unipolarnej fali prostokątnej

Literatura

[1] Shannon C.E.: A mathematical theory of communication. Bell Syst. Techn. Journal, vol. 27, July 1948.

[2] Szabatin J.: Era informacyjna a teoria Shannona. Przegląd Telekomunikacyjny, r. LXXIII, nr 4, 2000.

[3] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

[4] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom III. WNT, Warszawa, 1995.

Lekcja 2

Przestrzenie sygnałów

W lekcji 2 sygnały będziemy rozpatrywać jako elementy pewnych przestrzeni funkcyjnych, w których — analogicznie jak w zwykłej przestrzeni wektorowej — można formalnie określić wielkość sygnału (jego normę), odległość między sygnałami oraz kąt między nimi. Wprowadzimy ważne pojęcia ortogonalności i ortonormalności sygnałów oraz rozpatrzymy rozwinięcia sygnałów w ortogonalne (ortonormalne) uogólnione szeregi Fouriera. Omówimy przykłady takich szeregów, m.in. trygonometryczne szeregi Fouriera rzeczywisty i zespolony oraz szereg Kotielnikowa-Shannona, odgrywające w teorii sygnałów szczególnie istotną rolę.

2.1. Sygnały jako elementy przestrzeni funkcyjnych

2.1.1. Reprezentacja sygnału za pomocą sygnałów bazowych

W teorii sygnałów analogowych szeroko wykorzystuje się metody, w których sygnały są rozpatrywane jako elementy *przestrzeni funkcyjnych*. Przestrzeń funkcyjna jest zbiorem funkcji, którego wszystkie elementy charakteryzują się pewną wspólną cechą i w którym jest określona pewna struktura algebraiczna, tzn. są zdefiniowane operacje na jego elementach oraz rządzące nimi prawa – *aksjomaty* przestrzeni funkcyjnej.

Ujęcie sygnałów analogowych w kategoriach przestrzeni funkcyjnych ma dobrze ugruntowaną teorię i ułatwia rozwiązanie szeregu teoretycznych i praktycznych problemów teorii sygnałów. Umożliwia ono przede wszystkim określenie formalnej miary odległości między sygnałami w danej przestrzeni, analogicznie jak czyni się to w zwykłej przestrzeni wektorowej, wprowadzając euklidesowską miarę odległości dwóch wektorów. Określenie takiej miary umożliwia obiektywną ocenę stopnia podobieństwa dwóch sygnałów i jednoznaczne rozwiązanie zagadnienia poszukiwania sygnału najbliższego (przy przyjętych założeniach) danemu sygnałowi.

Metody oparte na pojęciu przestrzeni funkcyjnej umożliwiają ponadto reprezentację złożonych sygnałów za pomocą pewnych standardowych prostych sygnałów, nazywanych *sygnałami bazowymi*. Reprezentacja ta polega na przedstawieniu danego sygnału w postaci kombinacji liniowej sygnałów bazowych. Także i tu występuje analogia do przestrzeni wektorowych, gdzie każdy wektor można przedstawić jako kombinację liniową wersorów. Ważną zaletą takiej reprezentacji jest kompresja informacji o sygnale. Aby w pełni scharakteryzować sygnał, wymagana jest znajomość jego wartości we wszystkich chwilach, w których jest on określony, a więc znajomość nieprzeliczalnego w ogólnym przypadku zbioru liczb. Wyznaczenie reprezentacji sygnału w przestrzeni funkcyjnej jest równoznaczne z zapamiętaniem informacji o nim w przeliczalnym, a w szczególnych przypadkach – w skończonym zbiorze współczynników odpowiadającej mu kombinacji liniowej. Dla ustalonego zbioru funkcji bazowych zbiór tych współczynników, czyli współrzędnych sygnału w danej przestrzeni funkcyjnej, stanowi jego jednoznaczną reprezentację.

Metody reprezentacji i analizy sygnałów oparte na pojęciu przestrzeni funkcyjnej mają także inne zalety. Ułatwiają głębsze wniknięcie w często złożoną „strukturę wewnętrzną” sygnałów, umożliwiając tym samym zbadanie ich właściwości, ustalenie wzajemnych relacji między sygnałami, a także prostą i poglądową interpretację wykonywanych na nich operacji. Stanowią także punkt wyjścia do syntezy złożonych sygnałów za pomocą prostych sygnałów, łatwo generowanych z wykorzystaniem typowych układów elektronicznych. Z uwagi na ścisłą analogię z przestrzeniami wektorowymi, noszą one nazwę *metod geometrycznych*.

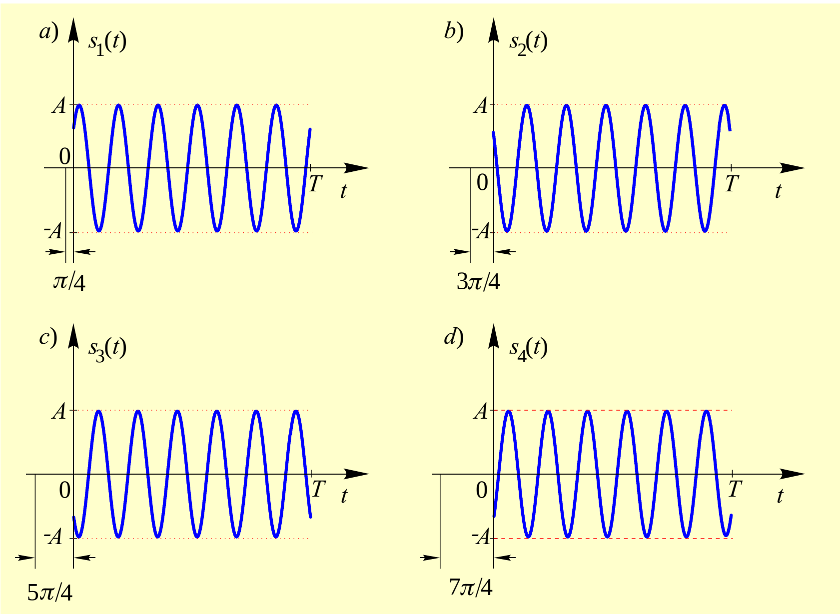
2.1.2. Przykłady przestrzeni sygnałów

Rozważmy dwa proste przykłady, przybliżające pojęcie przestrzeni funkcyjnej i związane z nim pojęcie *bazy* przestrzeni.

Przykład 2.1. We współczesnych cyfrowych systemach transmisji danych informacja źródłowa ma postać ciągów symboli binarnych „0” oraz „1”. W celu transmisji tych symboli stosowane są różne systemy cyfrowej modulacji sygnałów. Jednym z takich systemów jest system QPSK z *czterowartościowym kluczowaniem fazy* (ang. *Quadriphase-Shift Keying*) – por. 13.3.5. W systemie tym w każdym kolejnym odcinku czasu o długości T jest transmitowany jeden z czterech sygnałów reprezentujących pary symboli binarnych (*dwubity*):

$$
\begin{aligned}
\boxed{10} \quad s_1(t) &= A\cos\Big(2\pi f_0 t + \frac{1}{4}\pi\Big), \\
\boxed{00} \quad s_2(t) &= A\cos\Big(2\pi f_0 t + \frac{3}{4}\pi\Big), \\
\boxed{01} \quad s_3(t) &= A\cos\Big(2\pi f_0 t + \frac{5}{4}\pi\Big), \\
\boxed{11} \quad s_4(t) &= A\cos\Big(2\pi f_0 t + \frac{7}{4}\pi\Big).
\end{aligned} \tag{2.1}
$$

Sygnały te są harmonicznymi impulsami radiowymi (por. rys. 1.11) o czasie trwania T, amplitudzie A, częstotliwości oscylacji (*nośnej*) f_0 i różnych fazach początkowych odległych od siebie kolejno o $\pi/2$ (rys. 2.1). Informacja źródłowa jest zakodowana w fazie. Częstotliwość f_0 jest tak dobrana, że na odcinek czasu T przypada całkowita liczba okresów $T_0 = 1/f_0$ oscylacji, przy czym z reguły $T_0 \ll T$.



Rys. 2.1. Sygnały transmitowane w systemie QPSK

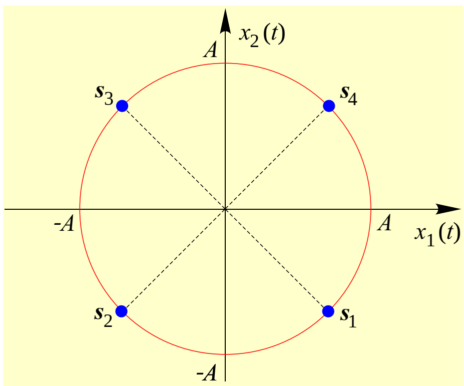
Korzystając z odpowiednich tożsamości trygonometrycznych, sygnały $s_i(t)$ można zapisać w postaci kombinacji liniowych dwóch elementarnych sygnałów

$x_1(t) = \cos(2\pi f_0 t)$ oraz $x_2(t) = \sin(2\pi f_0 t)$:

$$
\begin{aligned}
s_1(t) &= \frac{\sqrt{2}}{2} A\cos(2\pi f_0 t) - \frac{\sqrt{2}}{2} A\sin(2\pi f_0 t),\\
s_2(t) &= -\frac{\sqrt{2}}{2} A\cos(2\pi f_0 t) - \frac{\sqrt{2}}{2} A\sin(2\pi f_0 t),\\
s_3(t) &= -\frac{\sqrt{2}}{2} A\cos(2\pi f_0 t) + \frac{\sqrt{2}}{2} A\sin(2\pi f_0 t),\\
s_4(t) &= \frac{\sqrt{2}}{2} A\cos(2\pi f_0 t) + \frac{\sqrt{2}}{2} A\sin(2\pi f_0 t).
\end{aligned}
\tag{2.2}
$$

Sygnały $s_i(t)$ można zatem traktować jako elementy przestrzeni funkcyjnej $\mathscr{P}$, w której zostały wyróżnione sygnały bazowe $x_1(t)$ oraz $x_2(t)$ i która zawiera wszystkie sygnały o postaci $s(t) = ax_1(t) + bx_2(t)$, $a, b \in \mathscr{R}$. Zbiór $\{x_1(t), x_2(t)\}$ jest nazywany *bazą* tej przestrzeni, a przestrzeń $\mathscr{P}$ – *przestrzenią rozpiętą* na tej bazie. Liczba elementów bazy jest nazywana *wymiarem* przestrzeni. Rozważana przestrzeń $\mathscr{P}$ jest więc przestrzenią dwuwymiarową.

Interpretacja geometryczna przestrzeni $\mathscr{P}$ jest pokazana na rys. 2.2. Każdemu sygnałowi tej przestrzeni odpowiada punkt na płaszczyźnie w układzie współrzędnych, którego osiami są sygnały bazowe.



Rys. 2.2. Interpretacja geometryczna sygnałów QPSK

Na rysunku zaznaczono punkty odpowiadające poszczególnym sygnałom $s_i(t)$. Tworzą one tzw. *konstelację* sygnałów w systemie QPSK. Sygnały transmitowane w systemie QPSK można zatem utożsamiać z wektorami, których współrzędne są równe współczynnikom odpowiadających im kombinacji liniowych:

$$
\boldsymbol{s}_1 = \left[+\frac{\sqrt{2}}{2}A, -\frac{\sqrt{2}}{2}A\right], \quad \boldsymbol{s}_2 = \left[-\frac{\sqrt{2}}{2}A, -\frac{\sqrt{2}}{2}A\right],
$$
$$
\boldsymbol{s}_3 = \left[-\frac{\sqrt{2}}{2}A, +\frac{\sqrt{2}}{2}A\right], \quad \boldsymbol{s}_4 = \left[+\frac{\sqrt{2}}{2}A, +\frac{\sqrt{2}}{2}A\right].
$$

Wektory te stanowią reprezentację sygnałów QPSK w przestrzeni $\mathscr{P}$. Za miarę odległości między sygnałami można przyjąć zwykłą euklidesowską miarę odległości między odpowiadającymi im punktami płaszczyzny.

Przykład 2.2. W pierwszym przykładzie rozpatrzyliśmy przypadek przestrzeni skończenie wymiarowej. Z kolei podamy przykład przestrzeni nieskończenie wymiarowej.

Rozważmy przestrzeń funkcyjną, do której należą wszystkie rzeczywiste lub zespolone sygnały okresowe o ograniczonej mocy mające wspólny okres T_0. Przestrzeń ta jest oznaczana w matematyce przez $L^2_{T_0}$. Nieskończony zbiór sygnałów:

$$\left\{ \mathrm{e}^{\mathrm{j}k\omega_0 t} \colon k = 0, \pm 1, \pm 2, \ldots \right\}, \tag{2.3}$$

gdzie $\omega_0 = 2\pi/T_0$, jest bazą w tej przestrzeni. Każdy sygnał $x(t) \in L^2_{T_0}$ można przedstawić w postaci nieskończonej kombinacji liniowej elementów tej bazy:

$$x(t) = \sum_{k=-\infty}^{\infty} X_k \, \mathrm{e}^{\mathrm{j}k\omega_0 t}. \tag{2.4}$$

Przestrzeń $L^2_{T_0}$ jest więc przestrzenią rozpiętą na nieskończonej bazie (2.3).

Wyrażenie (2.4) jest nazywane *zespolonym szeregiem Fouriera* sygnału $x(t)$, a zbiór liczb zespolonych $\{X_k \colon k = 0, \pm 1, \ldots\}$ – zbiorem *współczynników Fouriera* tego szeregu. Zbiór współczynników Fouriera stanowi reprezentację sygnału $x(t)$ w przestrzeni $L^2_{T_0}$. W szczególnych przypadkach zbiór ten może być skończony.

2.2. Przestrzeń Hilberta sygnałów

W podanych przykładach pojęcia przestrzeni sygnałów oraz jej bazy zostały omówione w sposób poglądowy. Dalsze rozważania dotyczące reprezentacji sygnałów w przestrzeniach funkcyjnych wymagają większej ścisłości matematycznej i wprowadzenia szeregu nowych pojęć. Ich interpretację ułatwi odwoływanie się w każdym punkcie rozważań do analogii ze zwykłą przestrzenią wektorową.

W matematyce definiuje się różne typy przestrzeni funkcyjnych o różnych właściwościach formalnych. Największe znaczenie dla zagadnień teorii sygnałów mają tzw. *ośrodkowe przestrzenie Hilberta*. Przestrzeń Hilberta sygnałów ma dość złożoną strukturę algebraiczną. Postaramy się ją stopniowo przybliżyć. Dla wygody zapisów będziemy niekiedy opuszczać argument t w oznaczeniu sygnałów.

2.2.1. Przestrzeń liniowa

Przestrzeń Hilberta sygnałów jest przede wszystkim przestrzenią liniową. Oznacza to, że jest ona pewnym zbiorem $\mathscr{P}$, w którym określone zostały dwie operacje: dodawania elementów tego zbioru oraz ich mnożenia przez liczby rzeczywiste (w przypadku przestrzeni rzeczywistej) lub zespolone (w przypadku przestrzeni zespolonej). Operacje te nie wyprowadzają poza zbiór $\mathscr{P}$, tzn. jeśli $x \in \mathscr{P}$ i $y \in \mathscr{P}$, to także $x + y \in \mathscr{P}$, oraz jeśli $x \in \mathscr{P}$, to dla każdej liczby α także $\alpha x \in \mathscr{P}$. Operacja dodawania jest łączna i przemienna. Ponadto w zbiorze $\mathscr{P}$ istnieje dodatkowy element $\varnothing$, taki że $x + \varnothing = x$ dla każdego $x \in \mathscr{P}$. Element ten jest nazywany *elementem zerowym* przestrzeni.

Przykład 2.3. Zbiór sygnałów o ograniczonej energii, uzupełniony sygnałem zerowym $x(t) \equiv 0$, jest przestrzenią liniową. Przestrzeń ta jest nazywana w matematyce *przestrzenią sygnałów całkowalnych z kwadratem* i oznaczana L^2.

Przykład 2.4. Przestrzenią liniową jest także przestrzeń $L^2_{T_0}$ sygnałów okresowych o okresie T_0 z dodatkowym elementem zerowym $x(t) \equiv 0$.

Przykład 2.5. Zbiór sygnałów analogowych o ograniczonej mocy nie jest przestrzenią liniową. Suma dwóch sygnałów o ograniczonej mocy nie musi być bowiem sygnałem o ograniczonej mocy. Na przykład, suma sygnału $\boldsymbol{1}(t)$ i sygnału $-\boldsymbol{1}(t - T)$ jest impulsem prostokątnym, a więc sygnałem o ograniczonej energii. Gdybyśmy jednak zbiór sygnałów o ograniczonej mocy podzielili na klasy równoważności, zaliczając do każdej klasy wszystkie sygnały różniące się jedynie o sygnał mocy zerowej, to tak skonstruowany zbiór ilorazowy jest już przestrzenią liniową. Przestrzeń ta jest nazywana w literaturze *przestrzenią* M (Marcinkiewicza).

Przykład 2.6. Zbiór sygnałów dyskretnych $x[n]$ o ograniczonej energii uzupełniony sygnałem zerowym $x[n] \equiv 0$ jest przestrzenią liniową. Przestrzeń ta jest nazywana w matematyce *przestrzenią ciągów sumowalnych z kwadratem* i oznaczana l^2.

2.2.2. Baza przestrzeni

W każdej przestrzeni liniowej $\mathscr{P}$ można wyróżnić zbiór elementów, które tworzą pewien układ współrzędnych w tej przestrzeni pełniący analogiczną funkcję, jak układ współrzędnych zwykłej przestrzeni wektorowej.

Definicja 2.1. Każdy skończony lub przeliczalny podzbiór $\{x_k \colon k \in K\}$ *liniowo niezależnych* elementów przestrzeni liniowej $\mathscr{P}$, tj. elementów spełniających warunek:

$$\sum_{k \in K} \alpha_k x_k = \varnothing \quad \Leftrightarrow \quad \bigwedge_{k \in K} \alpha_k = 0$$

nazywamy *bazą* wyróżnioną w tej przestrzeni.

Jeśli zbiór indeksów K jest skończony, mówimy o bazie skończonej. Jeśli zbiór K jest nieskończony, bazę nazywamy nieskończoną.W zależności od sposobu numeracji jej elementów, zbiór K jest wówczas zbiorem liczb naturalnych $\mathscr{N}$, zbiorem $\mathscr{N} \cup \{0\}$ lub zbiorem liczb całkowitych $\mathscr{C}$. Zbiór wszystkich kombinacji liniowych elementów bazy wyróżnionej w przestrzeni $\mathscr{P}$ nazywamy przestrzenią *rozpiętą* na tej bazie. Przestrzeń ta jest podprzestrzenią przestrzeni $\mathscr{P}$, w szczególnych przypadkach – całą przestrzenią. Jeśli przestrzeń jest rozpięta na bazie skończonej zawierającej n elementów, to liczbę n nazywamy *wymiarem* tej przestrzeni (por. przykład 2.1). Przestrzenie rozpięte na bazach nieskończonych nazywamy *nieskończenie wymiarowymi* (por. przykład 2.2).

Większość interesujących nas przestrzeni sygnałów jest przestrzeniami nieskończenie wymiarowymi. W przestrzeniach tych szczególnie istotną rolę spełniają bazy o pewnych specjalnych właściwościach. Dokładne omówienie tych właściwości będzie możliwe dopiero po pełnym scharakteryzowaniu przestrzeni Hilberta.

2.2.3. Przestrzeń metryczna. Odległość między sygnałami

Przestrzeń Hilberta jest *przestrzenią metryczną*. Oznacza to, że w przestrzeni tej jest określone odwzorowanie $\varrho\colon \mathscr{P} \times \mathscr{P} \to \mathscr{R}^+ \cup \{0\}$, przyporządkowujące każdej parze jej elementów nieujemną liczbę rzeczywistą i spełniające dla każdego $x, y, z \in \mathscr{P}$ następujące warunki:

$$
\begin{array}{lll}
1. & \varrho(x,y) = 0 \;\leftrightarrow\; x = y, & \\
2. & \varrho(x,y) = \varrho(y,x) & \text{– warunek symetrii,} \\
3. & \varrho(x,y) + \varrho(y,z) \geqslant \varrho(x,z) & \text{– nierówność trójkąta.}
\end{array}
\tag{2.5}
$$

Odwzorowanie to jest nazywane *metryką* przestrzeni $\mathscr{P}$. Metryka stanowi miarę odległości między elementami tej przestrzeni. W danej przestrzeni można zdefiniować różne metryki. Ten sam zbiór z dwiema różnymi metrykami stanowi dwie różne przestrzenie metryczne.

Przykład 2.7. Przestrzeń L^2 z metryką

$$\varrho(x,y) = \sqrt{\int_{-\infty}^{\infty} |x(t) - y(t)|^2 \, \mathrm{d}t} \tag{2.6}$$

jest przestrzenią metryczną.

Przykład 2.8. Przestrzeń $L^2_{T_0}$ z metryką

$$\varrho(x,y) = \sqrt{\frac{1}{T_0} \int_{0}^{T_0} |x(t) - y(t)|^2 \, \mathrm{d}t} \tag{2.7}$$

jest przestrzenią metryczną.

Przykład 2.9. Przestrzeń M z metryką

$$\varrho(x,y) = \overline{\lim_{T\to\infty}} \sqrt{\frac{1}{2T}\int_{-T}^{T} |x(t)-y(t)|^2 \,\mathrm{d}t}, \tag{2.8}$$

gdzie $\overline{\lim}$ oznacza granicę górną, jest przestrzenią metryczną.

Przykład 2.10. Przestrzeń l^2 z metryką

$$\varrho(x,y) = \sqrt{\sum_{n=-\infty}^{\infty} |x(n)-y(n)|^2} \tag{2.9}$$

jest przestrzenią metryczną.

Metryki (2.6)–(2.9) zostały zdefiniowane od razu dla przestrzeni zespolonych.

2.2.4. Przestrzeń zupełna

Zgodnie z definicją przestrzeni liniowej każda skończona kombinacja liniowa jej elementów $\sum_{k=1}^{N}\alpha_k x_k(t)$ jest także elementem tej przestrzeni. Jeżeli liczbę wyrazów tej kombinacji będziemy zwiększać, to granica tak utworzonego ciągu sygnałów $x_N(t) = \sum_{k=1}^{N}\alpha_k x_k(t)$ przy $N \to \infty$ może istnieć, bądź też ciąg ten może być rozbieżny. Liniową przestrzeń metryczną $\mathscr{P}$ nazywamy *przestrzenią zupełną*, jeżeli dla każdego ciągu $x_N(t)$, dla którego istnieje granica przy $N \to \infty$, tj. zbieżnego w sensie metryki przestrzeni $\mathscr{P}$ do pewnego sygnału $x(t)$:

$$\lim_{N\to\infty} \varrho(x, x_N) = 0,$$

sygnał graniczny $x(t)$ jest także elementem tej przestrzeni. Każda przestrzeń Hilberta jest przestrzenią zupełną.

Przykład 2.11. Przestrzenie L^2, $L^2_{T_0}$, M oraz l^2 z metrykami, odpowiednio, (2.6)–(2.9) są przestrzeniami zupełnymi.

2.2.5. Przestrzeń unormowana

Kolejnym ważnym pojęciem w teorii przestrzeni funkcyjnych jest pojęcie normy. Jeżeli w przestrzeni liniowej $\mathscr{P}$ istnieje odwzorowanie $\|\cdot\|\colon \mathscr{P} \to \mathscr{R}^{+} \cup \{0\}$, przyporządkowujące każdemu sygnałowi $x(t) \in \mathscr{P}$ nieujemną liczbę rzeczywistą $\|x\|$ i spełniające warunki:

$$\begin{array}{ll} 1. & \|x\| = 0 \;\Leftrightarrow\; x = \varnothing, \\ 2. & \|\alpha x\| = |\alpha|\|x\|, \\ 3. & \|x+y\| \leqslant \|x\| + \|y\|, \end{array} \tag{2.10}$$

to przestrzeń $\mathscr{P}$ nazywamy *unormowaną*, a liczbę $\|x\|$ – *normą* sygnału $x(t)$.

Norma sygnału jest odpowiednikiem długości wektora w zwykłej przestrzeni wektorowej. Wprowadzenie tej miary nadaje ścisłe znaczenie pojęciu „wielkości" sygnału jako elementu danej przestrzeni. Dwa sygnały są równe w sensie normy, jeśli $\|x - y\| = 0$, tzn. jeśli sygnał różnicowy jest sygnałem zerowym.

Przykład 2.12. Przestrzenie L^2, $L^2_{T_0}$, M oraz l^2 można unormować definiując normy:

$$\|x\|_{L^2} = \sqrt{\int_{-\infty}^{\infty} |x(t)|^2 \,\mathrm{d}t}, \tag{2.11}$$

$$\|x\|_{L^2_{T_0}} = \sqrt{\frac{1}{T_0}\int_{0}^{T_0} |x(t)|^2 \,\mathrm{d}t}, \tag{2.12}$$

$$\|x\|_{M} = \overline{\lim_{T\to\infty}} \sqrt{\frac{1}{2T}\int_{-T}^{T} |x(t)|^2 \,\mathrm{d}t}, \tag{2.13}$$

$$\|x\|_{l^2} = \sqrt{\sum_{n=-\infty}^{\infty} |x(n)|^2}. \tag{2.14}$$

Miarą wielkości sygnału w przestrzeniach L^2 i l^2 jest więc pierwiastek z energii sygnału, a w przestrzeniach $L^2_{T_0}$ oraz M – jego wartość skuteczna.

Można łatwo wykazać, że w przestrzeni liniowej unormowanej pewną normą $\|\cdot\|$ wielkość

$$\varrho(x, y) = \|x - y\| \tag{2.15}$$

spełnia wszystkie warunki metryki. Metrykę taką nazywamy *indukowaną* przez normę $\|\cdot\|$. Metryki (2.6)–(2.9) są zatem indukowane przez normy (2.11)–(2.14).

2.2.6. Przestrzeń Banacha

Przestrzeń liniową unormowaną normą $\|\cdot\|$ nazywamy *przestrzenią Banacha*, jeśli jako przestrzeń metryczna z metryką indukowaną przez tę normę jest przestrzenią zupełną.

Przykład 2.13. Przestrzenie L^2, $L^2_{T_0}$, M oraz l^2, unormowane normami (2.11)–(2.14), są przestrzeniami Banacha.

2.2.7. Iloczyn skalarny sygnałów

Wprowadzone dotąd pojęcia nie charakteryzują jeszcze w pełni właściwości geometrycznych przestrzeni Hilberta. Odwołując się do analogii z wektorami, można powiedzieć, że w strukturze przestrzeni Banacha jest określone położenie sygnału, jego długość oraz odległość między sygnałami, nie jest natomiast określona miara kąta między sygnałami. Umożliwia to dopiero wprowadzenie nowej operacji na elementach przestrzeni liniowej – iloczynu skalarnego sygnałów. Pojęcie iloczynu skalarnego zdefiniujemy od razu dla ogólniejszego przypadku przestrzeni zespolonej.

Niech $\mathscr{P}$ będzie przestrzenią liniową, x, y, z – jej elementami oraz $\alpha, \beta \in \mathscr{C}$ liczbami zespolonymi. *Iloczynem skalarnym* określonym w przestrzeni $\mathscr{P}$ nazywamy odwzorowanie $\mathscr{P} \times \mathscr{P} \to \mathscr{C}$, przyporządkowujące parze uporządkowanej elementów $x, y \in \mathscr{P}$ liczbę $(x, y) \in \mathscr{C}$ i spełniające warunki:

$$\begin{array}{ll} 1. & (x,y) = (y,x)^*, \\ 2. & (\alpha x + \beta y, z) = \alpha(x,z) + \beta(y,z), \\ 3. & \begin{cases} (x,x) > 0, & \text{jeśli } x \neq \varnothing, \\ (x,x) = 0, & \text{jeśli } x = \varnothing. \end{cases} \end{array} \tag{2.16}$$

Iloczyn skalarny sygnałów ma wszystkie właściwości formalne zwykłego iloczynu skalarnego w przestrzeniach wektorowych. Spełnione są m.in. następujące właściwości:

– $(\varnothing, x) = 0$ dla każdego $x \in \mathscr{P}$,
– wielkość $\|x\| = \sqrt{(x,x)}$ spełnia warunki normy,
– zachodzi nierówność: $|(x,y)| \leqslant \sqrt{(x,x)(y,y)} = \|x\|\|y\|$ (nierówność Buniakowskiego-Schwarza).

Po wprowadzeniu pojęcia iloczynu skalarnego, możemy już teraz podać formalną definicję przestrzeni Hilberta.

Definicja 2.2. Przestrzeń liniowa unormowana i zupełna (przestrzeń Banacha), w której określony jest iloczyn skalarny (x, y), norma zaś jest zdefiniowana jako $\|x\| = \sqrt{(x,x)}$, jest nazywana *przestrzenią Hilberta*.

Przykład 2.14. Przestrzenie L^2, $L^2_{T_0}$ i l^2, w których iloczyn skalarny jest określony wzorami:

$$(x,y)_{L^2} = \int_{-\infty}^{\infty} x(t) y^*(t)\,\mathrm{d}t, \tag{2.17}$$

$$(x,y)_{L^2_{T_0}} = \frac{1}{T_0}\int_{0}^{T_0} x(t) y^*(t)\,\mathrm{d}t, \tag{2.18}$$

i odpowiednio:

$$(x,y)_{l^2} = \sum_{-\infty}^{\infty} x(n) y^*(n), \tag{2.19}$$

są przestrzeniami Hilberta.

Przykład 2.15. W przypadku przestrzeni M sytuacja jest bardziej złożona. Ściśle rzecz biorąc, przestrzeń ta nie jest przestrzenią Hilberta, nie można w niej bowiem zdefiniować iloczynu skalarnego spełniającego warunki (2.16). W przestrzeni tej wprowadza się pojęcie *pseudoiloczynu skalarnego*:

$$(x, y)_M = \overline{\lim_{T\to\infty}} \frac{1}{2T} \int_{-T}^{T} x(t) y^*(t)\,\mathrm{d}t, \tag{2.20}$$

gdzie $\overline{\lim}$ oznacza granicę górną. Pseudoiloczyn skalarny (2.20) spełnia warunek pierwszy i trzeci iloczynu skalarnego (2.16), natomiast w warunku drugim zamiast równości zachodzi słaba nierówność. W przestrzeni M można jednakże wyróżnić wiele podprzestrzeni (np. przestrzeń $L^2_{T_0}$ sygnałów okresowych), dla których granica górna jest równa zwykłej granicy. Spełniony jest wówczas także drugi z warunków (2.16), a więc w podprzestrzeniach tych jest określony iloczyn skalarny. Będziemy zakładać, że rozpatrywane dalej sygnały o ograniczonej mocy są elementami przestrzeni Hilberta z iloczynem skalarnym (2.20) określonym zwykłą granicą.

2.2.8. Kąt między sygnałami

Korzystając z pojęcia iloczynu skalarnego można zdefiniować kąt między sygnałami $x, y \in \mathscr{P}$. W przypadku przestrzeni rzeczywistych jest on określony wzorem:

$$\cos\varphi_{xy} = \frac{(x, y)}{\|x\|\|y\|}, \tag{2.21}$$

natomiast w przypadku przestrzeni zespolonej można stosować wzór:

$$\cos\varphi_{xy} = \frac{\mathrm{Re}\,(x, y)}{\|x\|\|y\|}. \tag{2.22}$$

2.3. Sygnały ortogonalne. Uogólniony szereg Fouriera

2.3.1. Pojęcie ortogonalności sygnałów

Jak pamiętamy, w zwykłej przestrzeni wektorowej kwadrat długości sumy dwóch wektorów nie jest równy sumie kwadratów ich długości. Właściwość ta jest słuszna dla dowolnej przestrzeni Hilberta. Można ją sformułować w kategoriach ogólnych następująco: kwadrat normy sumy dwóch elementów przestrzeni Hilberta nie jest równy sumie kwadratów ich norm. Zachodzi natomiast następująca równość, wynikająca wprost ze związku między normą a iloczynem skalarnym:

$$\|x + y\|^2 = (x + y)(x + y) = \|x\|^2 + \|y\|^2 + (x, y) + (y, x). \tag{2.23}$$

Na przykład, dla przestrzeni L^2 unormowanej normą (2.11) równość ta przybiera postać:

$$E_{x+y} = E_x + E_y + E_{xy} + E_{yx}, \tag{2.24}$$

gdzie E_{xy} i E_{yx} są *energiami wzajemnymi* określonymi wzorami:

$$E_{xy} = \int_{-\infty}^{\infty} x(t)y^*(t)\,\mathrm{d}t, \qquad E_{yx} = \int_{-\infty}^{\infty} y(t)x^*(t)\,\mathrm{d}t. \tag{2.25}$$

Energia sygnału nie jest zatem wielkością addytywną. Analogiczne zależności są słuszne dla przestrzeni $L^2_{T_0}$ i M , a więc również moc sygnałów nie jest wielkością addytywną.

Biorąc pod uwagę pierwszy warunek iloczynu skalarnego oraz wzór (2.22), możemy przepisać równość (2.24) w postaci:

$$\|x+y\|^2 = \|x\|^2+\|y\|^2+2\,\mathrm{Re}\,(x, y) = \|x\|^2+\|y\|^2+2\|x\|\|y\| \cos\varphi_{xy}, \tag{2.26}$$

będącej odpowiednikiem twierdzenia kosinusów dla zwykłej przestrzeni wektorowej. Jak wiemy, w przestrzeni tej kwadrat długości sumy dwóch wektorów x i y jest równy sumie kwadratów ich długości, jeśli są one prostopadłe, a więc jeśli $\cos\varphi_{xy} = 0$ lub równoważnie, jeśli iloczyn skalarny tych wektorów jest równy zeru. Właściwość ta, uogólniona na przestrzenie sygnałów, prowadzi do pojęcia *ortogonalności* (prostopadłości) sygnałów.

Definicja 2.3. Dwa sygnały x, y należące do przestrzeni Hilberta nazywamy *ortogonalnymi*, jeśli ich iloczyn skalarny $(x, y) = 0$.

Zgodnie ze wzorem (2.23), dla sygnałów ortogonalnych spełniona jest równość (uogólnione twierdzenie Pitagorasa):

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2. \tag{2.27}$$

2.3.2. Baza ortogonalna. Ośrodkowa przestrzeń Hilberta

Powróćmy do pojęcia bazy przestrzeni sygnałów. W podanej ogólnej definicji nie był narzucony warunek, aby baza tworzyła w danej przestrzeni liniowej prostokątny układ współrzędnych. W zagadnieniach teorii sygnałów jesteśmy zainteresowani przede wszystkim w wyodrębnieniu w strukturze przestrzeni Hilberta bazy nieskończonej, wyznaczającej w tej przestrzeni prostokątny układ współrzędnych.

Definicja 2.4. Nieskończony zbiór sygnałów $\{x_k\colon k \in K\}$ należących do przestrzeni Hilberta $\mathscr{P}$ nazywamy *bazą ortogonalną* tej przestrzeni, jeżeli sygnały x_k są ortogonalne oraz jeśli w przestrzeni $\mathscr{P}$ nie istnieje nie należący do tego zbioru niezerowy sygnał, który jest ortogonalny do sygnałów x_k.

Nie w każdej przestrzeni Hilberta istnieje baza ortogonalna. Przestrzenie, w których można wyodrębnić bazę ortogonalną są nazywane *ośrodkowymi przestrzeniami Hilberta*. W ogólnym przypadku w danej ośrodkowej przestrzeni Hilberta może istnieć więcej niż jedna baza ortogonalna.

2.3.3. Baza ortonormalna

W rozważaniach analitycznych wygodnie jest posługiwać się bazami, których wszystkie elementy mają jednostkową normę. Sygnały bazowe pełnią wówczas funkcję analogiczną do wersorów w zwykłej przestrzeni wektorowej.

Definicja 2.5. Jeśli zbiór $\{x_k\colon k \in K\}$ jest bazą ortogonalną w przestrzeni Hilberta, a ponadto $\|x_k\| = 1$ dla każdego $k \in K$, to bazę tę nazywamy *ortonormalną* (*ortogonalną unormowaną*).

Jeśli dana jest baza ortogonalna $\{u_k\colon k \in K\}$, to normując każdy jej element:

$$x_k = \frac{u_k}{\|u_k\|} \tag{2.28}$$

otrzymujemy odpowiadającą jej bazę ortonormalną $\{x_k\colon k \in K\}$. W celu wyznaczania baz ortonormalnych w ośrodkowych przestrzeniach Hilberta na podstawie zbioru elementów liniowo niezależnych stosowana jest iteracyjna procedura ortonormalizacji Grama-Schmidta (por. [1], p. 3.2.3).

2.3.4. Uogólniony szereg Fouriera

Mając daną ortonormalną bazę w przestrzeni Hilberta, możemy wyznaczyć reprezentację dowolnego sygnału tej przestrzeni za pomocą sygnałów bazowych.

Niech $\{x_k\colon k \in K\}$ będzie bazą ortonormalną w ośrodkowej przestrzeni Hilberta $\mathscr{P}$ sygnałów. Każdy sygnał x tej przestrzeni można przedstawić w postaci szeregu:

$$x = \sum_{k\in K} \alpha_k x_k, \tag{2.29}$$

gdzie

$$\alpha_k = (x, x_k), \tag{2.30}$$

a równość we wzorze (2.29) jest rozumiana w sensie normy przestrzeni $\mathscr{P}$, tzn.:

$$\left\| x - \sum_{k \in K} \alpha_k x_k \right\| = 0. \tag{2.31}$$

Szereg (2.29) nazywamy *uogólnionym szeregiem Fouriera* sygnału x, określonym względem bazy ortonormalnej $\{x_k \colon k \in K\}$. Współczynniki α_k tego szeregu są współrzędnymi sygnału x w układzie współrzędnych wyznaczonym przez tę bazę. Wzór (2.30), określający te współczynniki, można otrzymać obliczając iloczyny skalarne obu stron równości (2.29) z sygnałami bazowymi x_k i uwzględniając ich ortonormalność:

$$(x, x_k) = \left(\sum_{l \in K} \alpha_l x_l,\, x_k \right) = \sum_{l \in I} \alpha_l (x_l, x_k) = \begin{cases} 0 & \text{dla } l \neq k, \\ \alpha_k & \text{dla } l = k. \end{cases} \tag{2.32}$$

Zbiór $\{\alpha_k \colon k \in K\}$ nazywamy *rzutem* sygnału x w przestrzeni $\mathscr{P}$ na sygnały bazowe x_k.

Wzór (2.29) opisuje rozwinięcie sygnału x w uogólniony szereg Fouriera. Szereg ten stanowi jednoznaczną reprezentację sygnału x w przestrzeni $\mathscr{P}$ względem bazy $\{x_k \colon k \in K\}$. Należy podkreślić, że dobrze znane trygonometryczne szeregi Fouriera: rzeczywisty i zespolony są przypadkami szczególnymi uogólnionego szeregu Fouriera (2.29). Przykłady innych uogólnionych szeregów ortonormalnych, spełniających ważną rolę w zagadnieniach teorii sygnałów, podamy w punkcie 2.4.

Sygnał x można także rozwinąć w uogólniony szereg Fouriera względem bazy ortogonalnej $\{u_k \colon k \in K\}$:

$$x = \sum_{k \in K} \alpha'_k u_k. \tag{2.33}$$

Jeśli elementy bazy ortogonalnej $\{u_k \colon k \in K\}$ są związane z elementami bazy ortonormalnej $\{x_k \colon k \in K\}$ zależnością (2.28), to współczynniki ortogonalnego szeregu Fouriera (2.33) są związane ze współczynnikami ortonormalnego szeregu Fouriera (2.29) zależnością:

$$\alpha'_k = \frac{\alpha_k}{\|u_k\|}. \tag{2.34}$$

Komentarz. Rozwinięcie sygnału w uogólniony szereg Fouriera względem bazy ortonormalnej wyróżnionej w danej przestrzeni Hilberta $\mathscr{P}$ oznacza w istocie rzeczy określenie wzajemnie jednoznacznego odwzorowania $\chi \colon \mathscr{P} \to l^2$ przestrzeni $\mathscr{P}$ w przestrzeń Hilberta l^2 ciągów liczbowych sumowalnych z kwadratem. Odwzorowanie to jest liniowe, a także, co jest jego bardzo ważną cechą, zachowuje normę, tzn. odpowiednie normy w przestrzeniach $\mathscr{P}$ i l^2 są równe. O odwzorowaniach takich mówimy, że są *izometryczne*. Każda ośrodkowa przestrzeń Hilberta jest więc izometryczna z odpowiadającą jej przestrzenią l^2. Odwzorowanie χ zachowuje ponadto iloczyn skalarny.

Stwierdzenia te są niezwykle ważkie dla teorii sygnałów. Wynika z nich, że analizę sygnałów należących do danej przestrzeni Hilberta, badanie ich właściwości oraz operacji na sygnałach można przenieść do przestrzeni l^2, rozpatrując odpowiednie reprezentacje sygnałów w tej przestrzeni. Operacje na sygnałach i zależności między nimi sprowadzają się wówczas do prostych operacji i zależności algebraicznych między elementami przestrzeni l^2. Takie ujęcie zagadnień analizy sygnałów umożliwia jej znaczne uproszczenie oraz ułatwia interpretację otrzymanych wyników.

2.3.5. Zagadnienie najlepszej aproksymacji

Załóżmy, że w przestrzeni Hilberta $\mathscr{P}$ jest określona nieskończona baza ortonormalna $\{x_k \colon k \in K = \mathscr{N}\}$. W praktyce zachodzi często potrzeba aproksymacji danego sygnału x tej przestrzeni skończoną kombinacją liniową pierwszych N elementów jej bazy:

$$\tilde{x}_N = \sum_{k=1}^{N} \beta_k x_k. \tag{2.35}$$

Żąda się przy tym, aby sygnał $\tilde{x}_N$ najlepiej przybliżał sygnał x, tzn. aby odległość między tymi sygnałami w przestrzeni Hilberta była najmniejsza. Współczynniki β_k kombinacji (2.35) powinny być zatem tak dobrane, aby norma sygnału błędu aproksymacji $\varepsilon_N = x - \tilde{x}_N$ osiągała minimum:

$$\|\varepsilon_N\| = \|x - \tilde{x}_N\| = \left\| x - \sum_{k=1}^{N} \beta_k x_k \right\| = \min . \tag{2.36}$$

Minimalizacja miary błędu (2.36) stanowi treść *zagadnienia najlepszej aproksymacji*. Rozwiązanie tego zagadnienia wynika z tzw. *twierdzenia o rzucie* (por. [1], p. 3.2.4). Zgodnie z tym twierdzeniem miara błędu aproksymacji (2.36) osiąga minimum, gdy współczynniki kombinacji liniowej (2.35) są równe odpowiadającym im współczynnikom uogólnionego szeregu Fouriera (2.29) wyznaczonego względem bazy $\{x_k \colon k = 1, 2, \ldots\}$:

$$\bigwedge_{k=1,\ldots,N} \beta_k = \alpha_k. \tag{2.37}$$

Tak więc, spośród sygnałów $\tilde{x}_N$ o postaci (2.35), dany sygnał x najlepiej aproksymuje N-ta suma częściowa jego uogólnionego szeregu Fouriera (2.29).

Otrzymany rezultat jest konsekwencją ortogonalnej struktury bazy. Z twierdzenia o rzucie wynika ponadto, że także sygnał błędu aproksymacji $\varepsilon_N = x - \tilde{x}_N$ jest ortogonalny do każdego elementu $x_1, \ldots, x_N$ bazy, a więc do całej podprzestrzeni rozpiętej na pierwszych N elementach bazy. Aproksymowany sygnał x jest zatem sumą dwóch ortogonalnych sygnałów $\tilde{x}_N$ oraz ε_N. Biorąc to pod uwagę i uwzględniając ortonormalność sygnałów bazowych, a ponadto wzór (2.28) oraz warunek 2 normy (2.10), możemy obliczyć miarę (normę) błędu aproksymacji:

$$\|x\|^2 = \|\tilde{x}_N + \varepsilon_N\|^2 = \left\|\sum_{k=1}^{N} \alpha_k x_k\right\|^2 + \|\varepsilon_N\|^2 = \sum_{k=1}^{N} \|\alpha_k x_k\|^2 + \|\varepsilon_N\|^2 =$$
$$= \sum_{k=1}^{N} |\alpha_k|^2 \|x_k\|^2 + \|\varepsilon_N\|^2 = \sum_{k=1}^{N} |\alpha_k|^2 + \|\varepsilon_N\|^2,$$

a stąd:

$$\|\varepsilon_N\| = \sqrt{\|x\|^2 - \sum_{k=1}^{N} |\alpha_k|^2}. \tag{2.38}$$

W przypadku przestrzeni L^2, $L^2_{T_0}$, M oraz l^2 posługujemy się zwykle kryteriami energetycznymi miary błędu, tj. za miarę błędu przyjmujemy nie normę, lecz kwadrat normy sygnału błędu, a więc jego energię lub moc.

2.3.6. Twierdzenie Parsevala

Konsekwencją ortonormalności uogólnionego szeregu Fouriera (2.29) jest ważne twierdzenie Parsevala. Stanowi ono, iż norma sygnału może być wyrażona w prosty sposób przez współczynniki tego szeregu:

$$\|x\| = \sqrt{\sum_{k\in K} |\alpha_k|^2}. \tag{2.39}$$

Równość Parsevala (2.39) można wykazać, obliczając iloczyny skalarne obu stron równości (2.29) z sygnałem x i uwzględniając przy tym warunki 1 i 2 iloczynu skalarnego (2.16):

$$\begin{aligned}\|x\|^2 = (x, x) = \left(\textstyle\sum_{k\in K} \alpha_k x_k, x\right) &= \textstyle\sum_{k\in K} \alpha_k (x_k, x) = \\ &= \sum_{k\in K} \alpha_k \alpha_k^* = \sum_{k\in K} |\alpha_k|^2.\end{aligned} \tag{2.40}$$

W przypadku przestrzeni L^2, $L^2_{T_0}$ oraz M twierdzenie Parsevala orzeka, że energia (lub moc) sygnału może być obliczona jako suma kwadratów modułów współczynników jego rozwinięcia w uogólniony szereg Fouriera.

Z twierdzenia Parsevala wynika natychmiast, że wzór (2.38), określający miarę błędu aproksymacji sygnału x szeregiem (2.35) można zapisać w postaci:

$$\|\varepsilon_N\| = \sqrt{\sum_{k=N+1}^{\infty} |\alpha_k|^2}. \tag{2.41}$$

Sumowanie we wzorze (2.41) rozciąga się na wszystkie wyrazy rozwinięcia sygnału x w uogólniony szereg Fouriera, które nie zostały uwzględnione we wzorze aproksymującym (2.35). Stwierdzenie to pozostaje słuszne, gdy zbiór K indeksów uogólnionego szeregu Fouriera jest zbiorem $\mathscr{N} \bigcup \{0\}$ lub zbiorem wszystkich liczb całkowitych $\mathscr{C}$.

2.3.7. Układowa realizacja rozwinięcia w ortogonalny szereg Fouriera

Rozwinięcie sygnału w uogólniony szereg Fouriera (2.29) polega na wyznaczeniu współczynników tego szeregu zgodnie ze wzorem (2.30). Układowa realizacja tego rozwinięcia wymaga zatem zaprojektowania układów obliczających odpowiednie iloczyny skalarne. Układy te mogą mieć bardzo różnorodną strukturę, zależną od przyjętej definicji iloczynu skalarnego w danej przestrzeni Hilberta sygnałów.

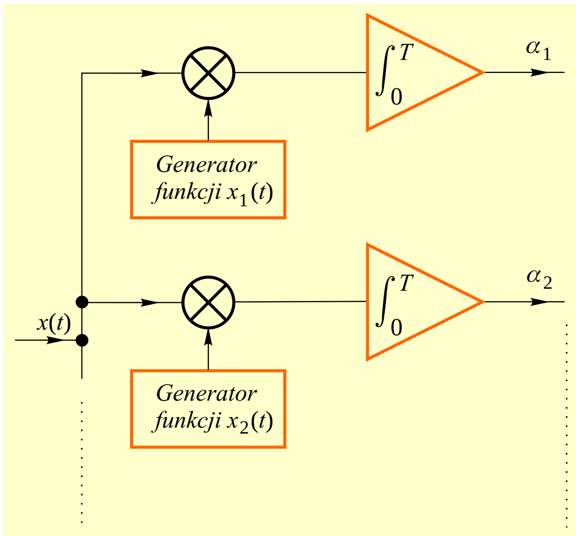
Rozważmy przestrzeń $L^2(0,T)$ impulsowych sygnałów o ograniczonej energii określonych w przedziale czasu $[0,T]$. W przypadku sygnałów rzeczywistych iloczyny skalarne (2.30) są określone wówczas wzorem:

$$\alpha_k = \int_0^T x(t)x_k(t)\,\mathrm{d}t. \tag{2.42}$$

Układ realizujący rozwinięcie sygnału $x(t)$ w szereg Fouriera powinien zatem zawierać generatory sygnałów bazowych $x_k(t)$, układy mnożące sygnał $x(t)$ przez sygnały bazowe oraz integratory. Schemat tego układu jest przedstawiony na rys. 2.3.

Na wyjściach integratorów otrzymujemy w chwili T sygnały o wartościach równych współczynnikom rozwinięcia α_k. W przypadku sygnałów zespolonych poszczególne tory wyznaczania współczynników α_k składają się z dwóch niezależnych gałęzi, w których wyznaczane są ich części rzeczywiste i urojone.

W rozwiązaniach praktycznych możemy oczywiście wykorzystać jedynie skończoną liczbę generatorów sygnałów bazowych. Układ z rys. 2.3 realizuje zatem optymalną aproksymację sygnału $x(t)$ skończonym ortonormalnym szeregiem Fouriera.



Rys. 2.3. Schemat blokowy układu realizującego rozwinięcie sygnału w ortogonalny szereg Fouriera

2.4. Przykłady ortonormalnych uogólnionych szeregów Fouriera

2.4.1. Szereg Haara

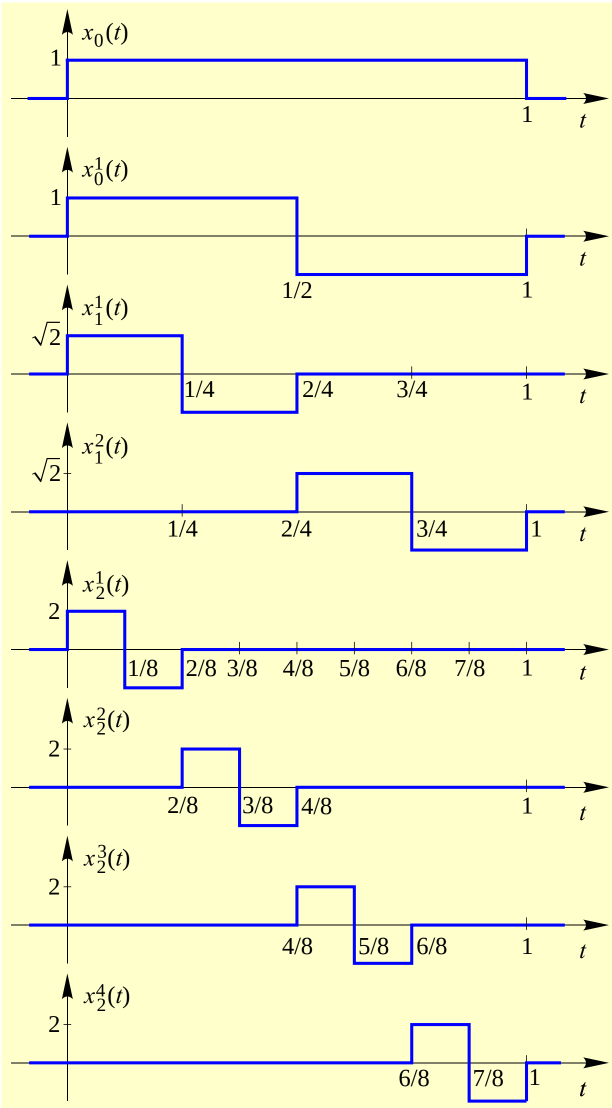
W praktyce często wykorzystywane są reprezentacje sygnałów za pomocą odpowiednio ukształtowanych sygnałów prostokątnych, przyjmujących w określonych przedziałach czasu stałe wartości. Sygnały takie można łatwo generować w układach cyfrowych, a układowa aproksymacja za ich pomocą złożonych sygnałów jest szczególnie prosta do realizacji. Przykładem zbioru takich sygnałów tworzących bazę ortonormalną w przestrzeni $L^2(0,1)$ są *funkcje Haara*.

Funkcje Haara są porządkowane według dwóch wskaźników $m = 0, 1, \ldots$ oraz $i = 1, 2, \ldots, 2^m$ i zdefiniowane wzorem:

$$x_m^i(t) = \begin{cases} \sqrt{2^m} & \text{dla} \quad \dfrac{2i-2}{2^{m+1}} < t < \dfrac{2i-1}{2^{m+1}}, \\ -\sqrt{2^m} & \text{dla} \quad \dfrac{2i-1}{2^{m+1}} < t < \dfrac{2i}{2^{m+1}}. \end{cases} \tag{2.43}$$

Dla pozostałych chwil z przedziału $[0,1]$ funkcje Haara przyjmują wartość zerową.

Ortonormalność zbioru funkcji Haara można sprawdzić bezpośrednio, obliczając iloczyny skalarne dwóch dowolnych jego elementów. Kilka pierwszych funkcji Haara dla $m = 0, 1, 2$ wykreślono na rys. 2.4.



Rys. 2.4. Funkcje Haara

Funkcje Haara można uporządkować także według jednego indeksu, definiując $H_k(t) = x_m^i(t)$, gdzie $k = 2^m + i - 1$ oraz $i = 1, \ldots, 2^m$. Uogólniony szereg Fouriera względem ortonormalnej bazy Haara można wówczas zapisać w postaci:

$$x(t) = \sum_{k=1}^{\infty} \alpha_k H_k(t), \tag{2.44}$$

gdzie

$$\alpha_k = \int\limits_0^1 x(t) H_k(t)\,\mathrm{d}t. \tag{2.45}$$

Układowe wyznaczanie iloczynów skalarnych (2.45) jest w przypadku bazy Haara bardzo proste. Mnożenie można zrealizować za pomocą układów kluczujących, zmieniających polaryzację sygnału w odpowiednich chwilach.

Dokonując odpowiednich transformacji skali czasu i skali amplitud, możemy oczywiście określić bazę ortonormalną funkcji Haara w dowolnym skończonym przedziale czasu.

2.4.2. Szereg Walsha

Bazę ortonormalną o podobnym charakterze tworzą w przestrzeni $L^2(0,1)$ *funkcje Walsha*. Są to funkcje binarne, odcinkami stałe, przyjmujące w każdej chwili $t \in [0,1]$ jedną z dwóch wartości: $+1$ lub -1. Dzięki temu, a także innym właściwościom, zdobyły one dużą popularność i są często stosowane w praktyce.

Podobnie jak funkcje Haara, funkcje Walsha są porządkowane według dwóch wskaźników. Funkcje te definiuje się następująco:

$$\begin{aligned}
x_0(t) &= \quad 1 \quad \text{dla} \quad 0 \leqslant t \leqslant 1, \\
x_1(t) &= \begin{cases} 1 & \text{dla} \quad 0 \leqslant t \leqslant \frac{1}{2}, \\ -1 & \text{dla} \quad \frac{1}{2} < t \leqslant 1, \end{cases} \\
x_2^1(t) &= \begin{cases} 1 & \text{dla} \quad 0 \leqslant t < \frac{1}{4} \quad \text{i} \quad \frac{3}{4} < t \leqslant 1, \\ -1 & \text{dla} \quad \frac{1}{4} < t < \frac{3}{4}, \end{cases} \\
x_2^2(t) &= \begin{cases} 1 & \text{dla} \quad 0 \leqslant t < \frac{1}{4} \quad \text{i} \quad \frac{1}{2} < t < \frac{3}{4}, \\ -1 & \text{dla} \quad \frac{1}{4} < t < \frac{1}{2} \quad \text{i} \quad \frac{3}{4} < t \leqslant 1, \end{cases}
\end{aligned} \tag{2.46a}$$

oraz rekurencyjnie dla $m = 1, 2, \ldots$ oraz $i = 1, \ldots, 2^{m-1}$:

$$\begin{aligned}
x_{m+1}^{2i-1}(t) &= \begin{cases} x_m^i(2t) & \text{dla} \quad 0 < t < \frac{1}{2}, \\ (-1)^{i+1} x_m^i(2t-1) & \text{dla} \quad \frac{1}{2} < t \leqslant 1, \end{cases} \\
x_{m+1}^{2i}(t) &= \begin{cases} x_m^i(2t) & \text{dla} \quad 0 \leqslant t < \frac{1}{2}, \\ (-1)^i x_m^i(2t-1) & \text{dla} \quad \frac{1}{2} < t \leqslant 1. \end{cases}
\end{aligned} \tag{2.46b}$$

Jak widzimy, definicja funkcji Walsha jest dość skomplikowana, ale sposób ich tworzenia jest dobrze widoczny na rys. 2.5. Na rysunku tym wykreślono 16 pierwszych funkcji Walsha. W punktach nieciągłości możemy przyjąć dowolne wartości, np. równe 1/2.

Funkcje Walsha można uporządkować także według jednego indeksu, przyjmując $x_0(t)$ i $x_1(t)$ za dwa pierwsze elementy i definiując $W_k(t) = x_m^i(t)$ dla $k = 2, 3, \ldots$, gdzie $k = 2^{m-1} + i - 1$ oraz $i = 1, \ldots, 2^{m-1}$. Przy takiej numeracji numer k funkcji Walsha jest równy liczbie jej przejść przez zero. Funkcje o numerach $2^k - 1$, $k = 1, 2, \ldots$ są w przedziale $[0, 1]$ zwykłymi bipolarnymi falami prostokątnymi.

Uogólniony szereg Fouriera względem ortonormalnej bazy Walsha ma postać:

$$x(t) = \sum_{k=0}^{\infty} \alpha_k W_k(t), \tag{2.47}$$

gdzie

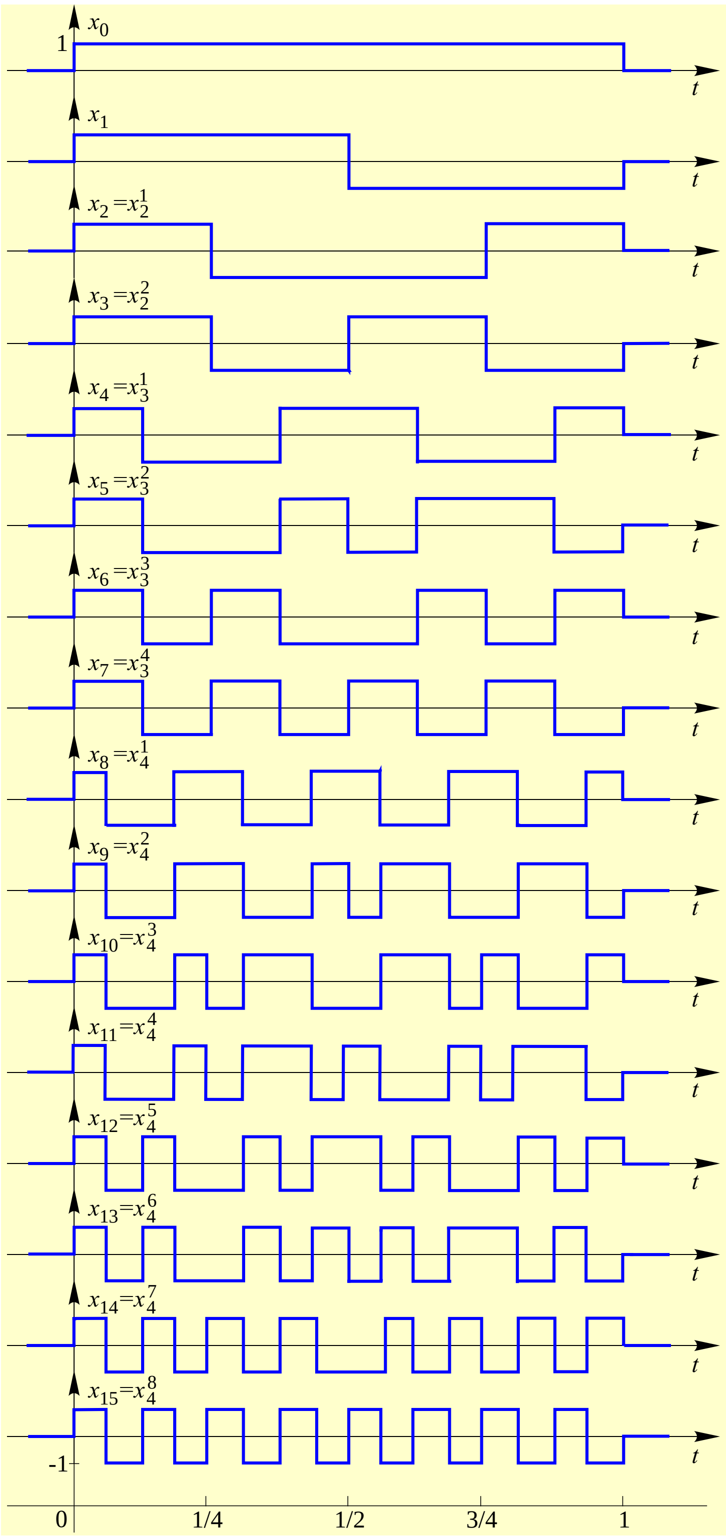
$$\alpha_k = \int_0^1 x(t) W_k(t)\, \mathrm{d}t. \tag{2.48}$$

Przykład 2.16. Wyznaczymy rozwinięcie impulsu trójkątnego

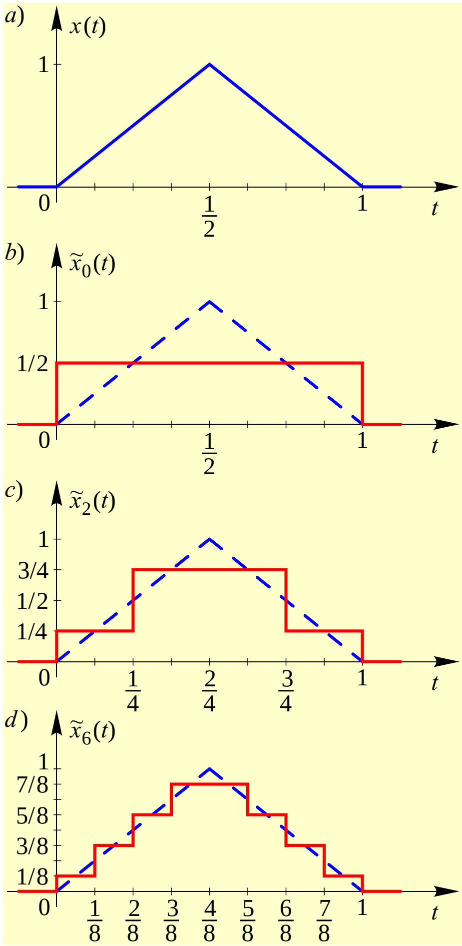
$$x(t) = \wedge\left(\frac{t - 1/2}{1/2}\right)$$

w ortonormalny szereg Walsha. Wykres tego impulsu jest pokazany na rys. 2.6a. Jego energia $E_x = 1/3$.

Ze wzoru (2.48) obliczamy współczynniki rozwinięcia: $\alpha_0 = 1/2$, $\alpha_1 = 0$, $\alpha_2 = -1/4$, $\alpha_3 = 0$, $\alpha_4 = 0$, $\alpha_5 = 0$, $\alpha_6 = -1/8$ itd. Na rys. 2.6b-d są pokazane kolejne przybliżenia impulsu $x(t)$ pierwszym oraz pierwszymi trzema i pięcioma wyrazami szeregu Walsha. Sygnały aproksymujące $\tilde{x}_N(t)$ mają postać funkcji schodkowych, których kształt w miarę zwiększania liczby wyrazów szeregu Walsha coraz bardziej przybliża się do impulsu trójkątnego. Korzystając ze wzoru (2.38) możemy obliczyć energie E_{ε_N} kolejnych sygnałów błędu $\varepsilon_N(t) = x(t) - \tilde{x}_N(t)$. Wynoszą one: $E_{\varepsilon_0} = 1/12$, $E_{\varepsilon_2} = 1/48$, $E_{\varepsilon_6} = 1/192$. Przyjmując miarę energetyczną błędu aproksymacji, widzimy, że już dla pierwszych trzech niezerowych wyrazów szeregu Walsha energia sygnału błędu wynosi $1,56\%$ energii impulsu $x(t)$. Biorąc pod uwagę kryterium energetyczne, aproksymację impulsu trójkątnego pierwszymi pierwszymi trzema niezerowymi wyrazami szeregu Walsha możemy zatem uznać za dostatecznie dokładną.



Rys. 2.5. Funkcje Walsha



Rys. 2.6. Aproksymacja impulsu trójkątnego szeregiem Walsha

2.4.3. Trygonometryczny szereg Fouriera

Zbiór funkcji:

$$\frac{1}{\sqrt{T_0}}, \quad \sqrt{\frac{2}{T_0}}\cos k\frac{2\pi}{T_0}t, \quad \sqrt{\frac{2}{T_0}}\sin k\frac{2\pi}{T_0}t, \qquad k = 1, 2, \ldots \tag{2.49}$$

tworzy bazę ortonormalną w przestrzeni $L^2_{T_0}$ rzeczywistych sygnałów okresowych o okresie T_0. Każdy sygnał $x(t)$ tej przestrzeni można zatem reprezentować szeregiem:

$$x(t) = \alpha_0\frac{1}{\sqrt{T_0}} + \sum_{k=1}^{\infty}\left(\alpha_k\sqrt{\frac{2}{T_0}}\cos k\frac{2\pi}{T_0}t + \beta_k\sqrt{\frac{2}{T_0}}\sin k\frac{2\pi}{T_0}t\right). \tag{2.50}$$

Współczynniki $\alpha_0, \alpha_k, \beta_k$ tego szeregu są określone wzorami:

$$\begin{aligned} \alpha_0 &= \left(x(t), \frac{1}{\sqrt{T_0}}\right) = \frac{1}{\sqrt{T_0}} \int\limits_{t_0}^{t_0+T_0} x(t)\,\mathrm{d}t, \\ \alpha_k &= \left(x(t), \sqrt{\frac{2}{T_0}} \cos k\frac{2\pi}{T_0}t\right) = \sqrt{\frac{2}{T_0}} \int\limits_{t_0}^{t_0+T_0} x(t)\cos k\frac{2\pi}{T_0}t\,\mathrm{d}t, \\ \beta_k &= \left(x(t), \sqrt{\frac{2}{T_0}} \sin k\frac{2\pi}{T_0}t\right) = \sqrt{\frac{2}{T_0}} \int\limits_{t_0}^{t_0+T_0} x(t)\sin k\frac{2\pi}{T_0}t\,\mathrm{d}t, \end{aligned} \tag{2.51}$$

gdzie t_0 jest dowolną chwilą.

Wzór (2.50) opisuje rozwinięcie sygnału $x(t)$ w *trygonometryczny* (*rzeczywisty*) *szereg Fouriera* względem bazy ortonormalnej (2.49). Ze względu na prostotę zapisu, w zagadnieniach teorii sygnałów (a także teorii obwodów) wykorzystuje się zwykle rozwinięcie w ortogonalny szereg funkcji trygonometrycznych, włączając współczynniki $1/\sqrt{T_0}$ oraz $\sqrt{2/T_0}$ normujące zbiór (2.49) do współczynników Fouriera. Szereg (2.50) przyjmuje wówczas dobrze znaną w literaturze i cytowaną już wcześniej postać (por. wzór (1.1)):

$$x(t) = a_0 + \sum_{k=1}^{\infty}(a_k \cos k\omega_0 t + b_k \sin k\omega_0 t), \tag{2.52}$$

gdzie $\omega_0 = 2\pi/T_0$, a współczynniki a_0, a_k, b_k są określone wzorami:

$$\begin{aligned} a_0 &= \frac{1}{\sqrt{T_0}}\alpha_0 = \frac{1}{T_0} \int\limits_{t_0}^{t_0+T_0} x(t)\,\mathrm{d}t, \\ a_k &= \sqrt{\frac{2}{T_0}}\alpha_k = \frac{2}{T_0} \int\limits_{t_0}^{t_0+T_0} x(t)\cos k\omega_0 t\,\mathrm{d}t, \\ b_k &= \sqrt{\frac{2}{T_0}}\beta_k = \frac{2}{T_0} \int\limits_{t_0}^{t_0+T_0} x(t)\sin k\omega_0 t\,\mathrm{d}t. \end{aligned} \tag{2.53}$$

Współczynnik a_0 jest równy składowej stałej sygnału $x(t)$, a współczynniki a_k, b_k są amplitudami jego poszczególnych składowych kosinusoidalnych i odpowiednio sinusoidalnych.

2.4.4. Zespolony szereg Fouriera

W przestrzeni $L^2_{T_0}$ bazę ortonormalną tworzy także zbiór zespolonych funkcji harmonicznych:

$$x_k(t) = \frac{1}{\sqrt{T_0}}\,\mathrm{e}^{\mathrm{j}k\frac{2\pi}{T_0}t}, \quad k = 0, \pm 1, \ldots \tag{2.54}$$

Reprezentacja sygnału $x(t)$ względem tej bazy:

$$x(t) = \sum_{k=-\infty}^{\infty} \alpha_k \frac{1}{\sqrt{T_0}}\, \mathrm{e}^{\mathrm{j}k\frac{2\pi}{T_0}t} \tag{2.55}$$

nosi nazwę *zespolonego szeregu Fouriera*. Jego współczynniki α_k są określone wzorem:

$$\alpha_k = (x(t), x_k^*(t)) = \frac{1}{\sqrt{T_0}} \int\limits_{t_0}^{t_0+T_0} x(t)\, \mathrm{e}^{-\mathrm{j}k\frac{2\pi}{T_0}t}\, \mathrm{d}t, \tag{2.56}$$

gdzie t_0 jest dowolną chwilą.

Podobnie jak w przypadku rzeczywistego trygonometrycznego szeregu Fouriera, w praktyce wygodnie jest posługiwać się omawianą już wcześniej w przykładzie 2.2 ortogonalną bazą zespolonych funkcji harmonicznych (2.3) i włączyć współczynnik $1/\sqrt{T_0}$ normujący zbiór (2.54) do współczynników Fouriera. Szereg (2.55) przyjmuje wówczas postać (por. wzór (2.4)):

$$x(t) = \sum_{k=-\infty}^{\infty} X_k\, \mathrm{e}^{\mathrm{j}k\omega_0 t}, \qquad \omega_0 = \frac{2\pi}{T_0}. \tag{2.57}$$

Współczynniki $X_k = \alpha_k/\sqrt{T_0}$ tego szeregu są określone wzorem:

$$X_k = \frac{1}{T_0} \int\limits_{t_0}^{t_0+T_0} x(t)\, \mathrm{e}^{-\mathrm{j}k\omega_0 t}\, \mathrm{d}t, \tag{2.58}$$

gdzie t_0 jest dowolną chwilą. Współczynnik X_0 jest składową stałą sygnału $x(t)$, a współczynniki $X_k, k = \pm 1, \pm 2, \ldots$, mają interpretację amplitud zespolonych poszczególnych składowych harmonicznych szeregu (2.57).

W reprezentacji sygnału okresowego zespolonym szeregiem Fouriera (2.57) występują składowe harmoniczne zarówno o dodatnich ($k > 0$), jak i ujemnych pulsacjach ($k < 0$). Pojęcie pulsacji ujemnej nie ma oczywiście sensu fizycznego, a występowanie w szeregu (2.57) składowych o ujemnych pulsacjach jest jedynie konsekwencją sposobu matematycznego opisu sygnału za pomocą tego szeregu. Dodajmy, że w przypadku sygnałów rzeczywistych zespolony charakter szeregu Fouriera jest w pewnym sensie pozorny, gdyż jego wartości dla każdej chwili t są rzeczywiste. Istotnie, korzystając ze wzoru (2.58) można łatwo wykazać, że dla sygnałów rzeczywistych współczynniki X_k tego szeregu spełniają równość $X_{-k} = X_k^*$, lub równoważnie $|X| = |X_{-k}|$ oraz $\arg X_k = -\arg X_{-k} \triangleq \varphi_k$. Dla każdej pary wskaźników k i $-k$ w szeregu (2.57) mamy zatem:

$$\begin{aligned} X_k\, \mathrm{e}^{\mathrm{j}k\omega_0 t} &+ X_{-k}\, \mathrm{e}^{-\mathrm{j}k\omega_0 t} = \\ &= |X_k|\, \mathrm{e}^{\mathrm{j}(k\omega_0 t+\varphi_k)} + |X_k|\, \mathrm{e}^{-\mathrm{j}(k\omega_0 t+\varphi_k)} = \\ &= 2|X_k| \cos(k\omega_0 t + \varphi_k), \end{aligned} \tag{2.59}$$

a więc otrzymujemy sygnał będący k-tą rzeczywistą składową harmoniczną sygnału $x(t)$.

Między szeregami Fouriera rzeczywistym (2.52) i zespolonym (2.57) istnieje ścisły związek. Znając jedną z postaci szeregu można wyznaczyć drugą. Przejścia między nimi można dokonać na podstawie związków między ich współczynnikami, wynikającymi z porównania wzorów (2.53) i (2.58):

$$X_0 = a_0, \qquad X_k = \frac{a_k - \mathrm{j}\, b_k}{2}, \tag{2.60}$$

oraz

$$a_0 = X_0, \qquad a_k = X_k + X_{-k}, \qquad b_k = \mathrm{j}\,(X_k - X_{-k}), \tag{2.61}$$

2.4.5. Szereg Kotielnikowa-Shannona

W ostatnim przykładzie omówimy ortonormalny uogólniony szereg Fouriera, który – obok szeregów trygonometrycznych Fouriera – odgrywa w teorii sygnałów bodaj najważniejszą rolę. Podczas prezentacji tego przykładu zmuszeni będziemy wybiec nieco w przód i w niektórych jej fragmentach sięgnąć do pojęć z zakresu analizy częstotliwościowej sygnałów.

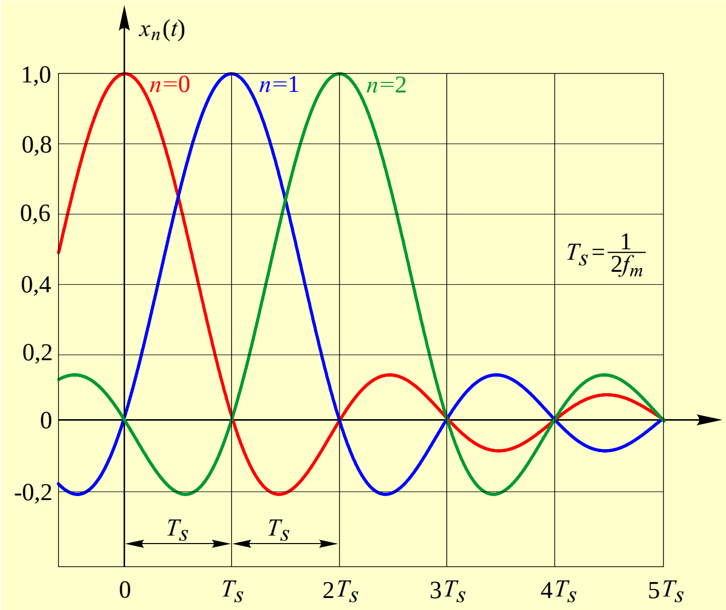
Rozważmy nieskończony zbiór sygnałów Sa (por. rys. 1.14):

$$x_n(t) = \frac{1}{\sqrt{T_s}}\,\mathrm{Sa}\,\frac{\pi}{T_s}(t - nT_s), \quad n = 0, \pm 1, \pm 2, \ldots, \tag{2.62}$$

którego kolejne elementy są swoimi kopiami przesuniętymi względem chwili $t = 0$ o czas nT_s. Wykresy tych sygnałów są pokazane na rys. 2.7.

Sygnały (2.62) należą do przestrzeni Hilberta $L^2(-\infty, \infty)$ sygnałów o ograniczonej energii. Ich energie są równe jedności, a więc są to sygnały o jednostkowych normach. Ponadto są to sygnały ortonormalne w tej przestrzeni. W przeciwieństwie do wcześniej omówionych zbiorów, których właściwość ortonormalności można wykazać wprost obliczając odpowiednie iloczyny skalarne, bezpośrednie obliczenie iloczynów skalarnych sygnałów (2.62) nie jest trywialne i wymaga znajomości elementów analizy częstotliwościowej sygnałów. Dowód ortonormalności tych sygnałów będziemy mogli przeprowadzić dopiero po wprowadzeniu pojęcia widma sygnału i omówieniu jego właściwości (por. p. 6.2.5).

Mimo że zbiór (2.62) sygnałów Sa jest zbiorem ortonormalnym w przestrzeni $L^2(-\infty, \infty)$, nie tworzy on w tej przestrzeni bazy ortonormalnej. Istnieje bowiem wiele innych elementów tej przestrzeni ortonormalnych do elementów tego zbioru (por. definicje (2.4) i (2.5)). Zbiór (2.62) tworzy natomiast bazę ortonormalną w podprzestrzeni przestrzeni $L^2(-\infty, \infty)$ zawierającej wszystkie sygnały, w których widmie nie występują składowe częstotliwościowe o pulsacjach większych od pulsacji granicznej ω_m związanej z parametrem T_s zależnością $\omega_m = \pi/T_s$ (równoważnie – o częstotliwościach większych od częstotliwości granicznej f_m



Rys. 2.7. Ortonormalny zbiór funkcji Sa

związanej z parametrem T_s zależnością $f_m = 1/2T_s$). Sygnały, w których widmie nie występują składowe o pulsacjach większych od pewnej pulsacji granicznej nazywamy sygnałami *o ograniczonym paśmie*. Dla różnych wartości pulsacji granicznej ω_m sygnały te tworzą różne podprzestrzenie przestrzeni $L^2(-\infty, \infty)$.

Tak więc, każdy sygnał należący do przestrzeni sygnałów o paśmie ograniczonym pulsacją $\omega_m = \pi/T_s$ można rozwinąć w uogólniony szereg Fouriera względem ortonormalnej bazy (2.62). Szereg ten nosi nazwę szeregu *Kotielnikowa-Shannona* i ma postać:

$$x(t) = \sum_{n=-\infty}^{\infty} \alpha_n \frac{1}{\sqrt{T_s}} \operatorname{Sa} \frac{\pi}{T_s}(t - nT_s) = \sum_{n=-\infty}^{\infty} \alpha_n' \operatorname{Sa} \frac{\pi}{T_s}(t - nT_s), \qquad (2.63)$$

gdzie w drugiej równości czynnik normujący $1/\sqrt{T_s}$ został już włączony do współczynników α_n'.

Szereg Kotielnikowa-Shannona (2.63) ma bardzo ważną właściwość, a mianowicie jego współczynniki Fouriera α_n' są równe wartościom (*próbkom*) $x(nT_s)$ sygnału $x(t)$ w chwilach nT_s. Można go więc zapisać w postaci:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \operatorname{Sa} \frac{\pi}{T_s}(t - nT_s) = \sum_{n=-\infty}^{\infty} x(nT_s) \operatorname{Sa} \omega_m(t - nT_s). \quad (2.64)$$

Oznacza to, że zbiór $\{x(nT_s\}$ próbek sygnału $x(t)$ o ograniczonym pulsacją ω_m paśmie, pobieranych w chwilach $nT_s = n\pi/\omega_m$, $n = 0, \pm 1, \pm 2, \ldots$, stanowi reprezentację tego sygnału względem bazy (2.62) w przestrzeni sygnałów, których pasmo jest ograniczone pulsacją ω_m. Parametr T_s występujący w zapisie funkcji Sa (2.62) ma zatem znaczenie okresu (przedziału) próbkowania. Uzasadnia to przyjęte we wzorze (2.62) oznaczenie tego parametru symbolem T_s, w którym indeks „s" pochodzi od angielskiego terminu *sampling*.

Szereg Kotielnikowa-Shannona odgrywa kluczową rolę w zagadnieniach próbkowania sygnałów. Dokładne jego omówienie, wraz z dowodem, że jego współczynnikami Fouriera są próbki sygnału, przeprowadzimy w lekcji 6 poświęconej problematyce próbkowania sygnałów (por. p. 6.2.5).

Słownik

aproksymacja sygnału
: przybliżenie sygnału skończoną kombinacją liniową pewnych standardowych sygnałów

baza ortogonalna
: baza przestrzeni sygnałów, której elementy są względem siebie ortogonalne

baza ortonormalna
: baza ortogonalna o unormowanych do jedności elementach

baza przestrzeni sygnałów
: wyróżniony w danej przestrzeni sygnałów zbiór pewnych standardowych sygnałów (skończony lub nieskończony), z reguły ortogonalnych, tworzący w tej przestrzeni układ współrzędnych; każdy sygnał w tej przestrzeni może być przedstawiony w postaci kombinacji liniowej tych sygnałów

dziedzina częstotliwości
: dziedzina opisu i analizy sygnałów i układów w funkcji zmiennej f – częstotliwości

iloczyn skalarny sygnałów
: funkcjonał przyporządkowujący parze sygnałów w danej przestrzeni sygnałów liczbę i mający właściwości formalne iloczynu skalarnego zwykłej przestrzeni wektorowej

metryka przestrzeni
: funkcjonał przyporządkowujący parze sygnałów w danej przestrzeni sygnałów liczbę określającą odległość między tymi sygnałami

norma sygnału
funkcjonał przyporządkowujący sygnałowi w danej przestrzeni sygnałów liczbę określającą jego długość (wielkość)

przestrzeń Hilberta
przestrzeń liniowa unormowana i zupełna, w której jest określona operacja iloczynu skalarnego na jej elementach

przestrzeń metryczna
przestrzeń sygnałów, w której określona jest metryka

przestrzeń rozpięta na bazie
przestrzeń sygnałów zawierająca wszystkie kombinacje liniowe elementów bazy

przestrzeń sygnałów
zbiór sygnałów (najczęściej nieskończony), którego elementy charakteryzują się pewną wspólną cechą i w którym jest określona pewna struktura algebraiczna (są zdefiniowane operacje na elementach tego zbioru oraz rządzące nimi prawa – tzw. aksjomaty przestrzeni)

przestrzeń unormowana
przestrzeń sygnałów, w której określona jest norma

sygnały bazowe (baza przestrzeni sygnałów)
zbiór pewnych standardowych sygnałów, z reguły ortogonalnych, tworzących w przestrzeni sygnałów prostokątny układ współrzędnych

sygnały ortogonalne
sygnały, których iloczyn skalarny w danej przestrzeni sygnałów jest równy zeru

sygnały ortonormalne
sygnały ortogonalne o jednostkowej normie

uogólniony szereg Fouriera
przedstawienie sygnału jako sumy ważonej (najczęściej nieskończonej) innych standardowych sygnałów tworzących bazę ortogonalną w określonej przestrzeni sygnałów

Literatura

[1] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

Lekcja 3

Analiza częstotliwościowa sygnałów analogowych

W lekcji 3 omówimy całkowe przekształcenie Fouriera, stanowiące podstawowe narzędzie matematyczne analizy częstotliwościowej sygnałów analogowych. Na tej podstawie wprowadzimy definicje widma sygnału oraz jego widma amplitudowego i fazowego. Podamy najważniejsze twierdzenia dotyczące całkowego przekształcenia Fouriera oraz przykłady typowych par transformat Fouriera. Omówimy ponadto związki między trygonometrycznym szeregiem Fouriera a przekształceniem Fouriera. Lekcję zakończymy krótkim omówieniem zasady nieoznaczoności w teorii sygnałów.

3.1. Przekształcenie Fouriera

Wartości sygnału zmieniają się w czasie. Zmiany te następują z różną prędkością. W jednych odcinkach czasu zachodzą szybciej, w innych – wolniej. Niektóre prędkości zmian dominują w sygnale, inne występują rzadko lub wcale. Struktura tych prędkości oraz intensywność z jaką występują stanowią ważną charakterystykę sygnału, którą można przeanalizować przenosząc rozważania do *dziedziny częstotliwości* i definiując pojęcie *widma* sygnału.

Dziedzina częstotliwości jest alternatywną dziedziną opisu i analizy sygnałów, ściśle związaną z dziedziną czasu. Metody analizy sygnałów w dziedzinie częstotliwości noszą nazwę *metod częstotliwościowych* lub *metod widmowych*. Stanowią one niejednokrotnie bardziej efektywne i skuteczne narzędzie badania sygnałów, niż metody czasowe. W „języku” częstotliwościowym można w wielu przypadkach w sposób prostszy i bardziej poglądowy niż w „języku” czasowym opisać podstawowe cechy sygnału. W kategoriach częstotliwościowych wygodnie jest także rozpatrywać i interpretować operacje przetwarzania sygnałów przez układy, w tym przede wszystkim operację filtracji. Warto wspomnieć, że proces słyszenia (nawiasem mówiąc, nie do końca rozpoznany) jest rezultatem analizy struktury częstotliwościowej sygnału akustycznego przez ucho ludzkie, które pod względem funkcjonalnym pełni rolę analizatora widma.

Częstotliwościowe metody analizy sygnałów ciągłych są oparte na formalizmie matematycznym przekształceń całkowych Fouriera – prostego i odwrotnego. Ustalają one wzajemne relacje między dziedziną czasu i dziedziną częstotliwości. Mimo że przekształcenie całkowe Fouriera jest znane od przeszło dwustu lat, pozostaje nadal fundamentalnym i najbardziej rozpowszechnionym narzędziem analizy sygnałów.

3.1.1. Proste przekształcenie Fouriera

Rozważmy analogowy sygnał deterministyczny $x(t)$, w ogólnym przypadku zespolony.

Definicja 3.1. *Proste przekształcenie Fouriera* sygnału $x(t)$ jest zdefiniowane całką

$$X(\omega) = \int\limits_{-\infty}^{\infty} x(t)\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t, \tag{3.1}$$

gdzie zmienna $\omega \in (-\infty, \infty)$ jest pulsacją.

Obliczając całkę (3.1) dla poszczególnych wartości pulsacji $\omega_1, \omega_2, \ldots$, otrzymujemy wartości zespolone $X(\omega_1), X(\omega_2), \ldots$. W rezultacie całka ta przyporządkowuje sygnałowi $x(t)$ funkcję zespoloną $X(\omega)$ zmiennej rzeczywistej ω. Przyporządkowanie to będziemy oznaczać $X(\omega) = \mathscr{F}[x(t)]$. Funkcję $X(\omega)$ nazywamy *transformatą Fouriera* (w skrócie *$\mathscr{F}$-transformatą*) sygnału $x(t)$.

Przykład 3.1. Wyznaczymy transformatę Fouriera sygnału wykładniczego malejącego (por. rys. 1.12):

$$\begin{aligned} X(\omega) = \textstyle\int_0^{\infty} X_0\,\mathrm{e}^{-\alpha t}\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t &= X_0 \int\limits_0^{\infty} \mathrm{e}^{-(\alpha+\mathrm{j}\omega)t}\,\mathrm{d}t = \\ &= -\frac{X_0}{\alpha + \mathrm{j}\,\omega}\,\mathrm{e}^{-(\alpha+\mathrm{j}\omega)t}\Big|_0^{\infty} = \frac{X_0}{\alpha + \mathrm{j}\,\omega}. \end{aligned} \tag{3.2}$$

Przykład 3.2. Wyznaczymy transformatę Fouriera impulsu prostokątnego $X_0 \sqcap (t/T)$ (por. rys. 1.7):

$$\begin{aligned} X(\omega) &= X_0 \int_{-\infty}^{\infty} \sqcap\left(\frac{t}{T}\right) \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t = X_0 \int_{-T/2}^{T/2} \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t = \\ &= -\frac{X_0}{\mathrm{j}\,\omega} \mathrm{e}^{-\mathrm{j}\omega t} \Big|_{-T/2}^{T/2} = \frac{X_0}{\mathrm{j}\,\omega}\left(\mathrm{e}^{\mathrm{j}\omega \frac{T}{2}} - \mathrm{e}^{-\mathrm{j}\omega \frac{T}{2}}\right) = \\ &= \frac{2X_0}{\omega} \sin\frac{\omega T}{2} = X_0 T \,\mathrm{Sa}\, \frac{\omega T}{2}. \end{aligned} \tag{3.3}$$

3.1.2. Odwrotne przekształcenie Fouriera

Definicja 3.2. *Odwrotne przekształcenie Fouriera* jest zdefiniowane całką

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)\, \mathrm{e}^{\mathrm{j}\omega t}\, \mathrm{d}\omega. \tag{3.4}$$

Całka (3.4) przyporządkowuje funkcji zespolonej $X(\omega)$ sygnał $x(t)$. Przyporządkowanie to będziemy oznaczać $x(t) = \mathscr{F}^{-1}[X(\omega)]$. Sygnał $x(t)$ nazywamy *odwrotną transformatą Fouriera* lub *retransformatą Fouriera* (w skrócie *$\mathscr{F}^{-1}$-transformatą*) funkcji $X(\omega)$.

3.1.3. $\mathscr{F}$-transformowalność i wzajemna jednoznaczność

Definiując proste przekształcenie Fouriera, nie sformułowaliśmy warunków, jakie powinien spełniać sygnał $x(t)$, aby całka (3.1) istniała. Nie określiliśmy tym samym zakresu stosowalności wzoru (3.1). Sygnały, dla których całka (3.1) istnieje, nazywamy *$\mathscr{F}$-transformowalnymi* w zwykłym sensie.

Ustalenie warunków koniecznych i dostatecznych $\mathscr{F}$-transformowalności w zwykłym sensie jest zagadnieniem złożonym. Z reguły formułowane są jedynie różne postacie warunków dostatecznych. Nie wnikając w szczegóły, podamy warunek najczęściej cytowany w literaturze:

$$\int_{-\infty}^{\infty} |x(t)|\, \mathrm{d}t < \infty. \tag{3.5}$$

Zgodnie z tym warunkiem każdy sygnał bezwzględnie całkowalny jest $\mathscr{F}$-transformowalny w zwykłym sensie. Klasa sygnałów bezwzględnie całkowalnych jest jednak stosunkowo wąska i mało interesująca dla zagadnień teorii sygnałów.

Istotna jest natomiast odpowiedź na pytanie, czy $\mathscr{F}$-transformowalne są sygnały o ograniczonej energii i ograniczonej mocy.

W przypadku sygnałów o ograniczonej mocy całka (3.1) jest rozbieżna, nie są one zatem $\mathscr{F}$-transformowalne w zwykłym sensie. W przypadku sygnałów o ograniczonej energii można natomiast wykazać, że całka (3.1) istnieje *prawie wszędzie*, tzn. jest zbieżna wszędzie z wyjątkiem skończonej lub przeliczalnej liczby wartości pulsacji ω. Należy jednak podkreślić, że dla większości interesujących nas i omówionych wcześniej sygnałów o ograniczonej energii całka (3.1) jest zbieżna wszędzie, a więc wzór (3.1) można stosować bez żadnych ograniczeń.

Z kolei, definiując odwrotne przekształcenie Fouriera, nie przesądzaliśmy, czy obliczając odwrotną transformatę Fouriera względem $\mathscr{F}$-transformaty $X(\omega)$ pewnego sygnału $x(t)$ otrzymamy z powrotem sygnał $x(t)$, tzn. nie rozstrzygaliśmy, czy i dla jakiej klasy sygnałów przekształcenie Fouriera jest wzajemnie jednoznaczne. Również zagadnienie wzajemnej jednoznaczności jest skomplikowane i obfituje w wiele subtelnych szczegółów. Ogólnie można powiedzieć, że jeśli $X(\omega)$ jest $\mathscr{F}$-transformatą bezwzględnie całkowalnego sygnału $x(t)$ i sygnał ten spełnia pewne niezbyt ograniczające warunki regularnościowe (np. jest sygnałem przedziałami regularnym – por. [1], str. 225]), to dla każdej chwili zachodzi równość:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\,\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega = \frac{x(t+)+x(t-)}{2}, \tag{3.6}$$

tzn. odwrotna transformata Fouriera w chwili t jest równa średniej arytmetycznej granic lewo- i prawostronnej sygnału w tej chwili. W szczególności, w każdej chwili t , w której sygnał $x(t)$ jest ciągły, spełniona jest równość (3.4). Dla sygnałów takich przekształcenie Fouriera jest wzajemnie jednoznaczne. Relację między sygnałem i jego $\mathscr{F}$-transformatą będziemy wówczas oznaczać:

$$x(t) \leftrightarrow X(\omega), \tag{3.7}$$

a parę funkcji $x(t)$ i $X(\omega)$ nazywać *parą transformat Fouriera w sensie zwykłym* . Przykłady prostych par transformat Fouriera w sensie zwykłym zostaną podane w p. 3.4.1.

Komentarz. Dyskusja warunków dostatecznych $\mathscr{F}$-transformowalności i wzajemnej jednoznaczności została przeprowadzona przy założeniu, że całki (3.1) i (3.4) są rozumiane w sensie Riemanna. Sformułowanie tych warunków w odniesieniu do najbardziej nas interesującej klasy L^2 sygnałów o ograniczonej energii jest wówczas skomplikowane. Nieco inaczej wygląda sytuacja, jeżeli całki (3.1) i (3.4) są rozumiane w sensie Lebesgue'a (por. [2], str. 430). Przy takim rozumieniu tych całek sygnały klasy L^2 są $\mathscr{F}$-transformowalne, a ponadto, jeśli równość sygnałów będziemy rozumieć w sensie normy przestrzeni L^2, przekształcenie Fouriera jest dla tych sygnałów wzajemnie jednoznaczne. Przekształcenie Fouriera określa wówczas odwzorowanie przestrzeni L^2_t sygnałów w dziedzinie czasu w przestrzeń L^2_ω transformat w dziedzinie częstotliwości.

Sygnały o ograniczonej mocy nie są oczywiście $\mathscr{F}$-transformowalne także wtedy, kiedy całkę (3.1) będziemy rozumieć w sensie Lebesgue'a.

3.1.4. Przekształcenie Fouriera w sensie granicznym

Przekształcenie Fouriera w sensie zwykłym, określone całką (3.1), nie obejmuje wielu podstawowych sygnałów, odgrywających w teorii sygnałów istotną rolę. Należą do nich m.in. sygnał stały, sygnał skoku jednostkowego, sygnały harmoniczne, czy też sygnały dystrybucyjne. W celu rozszerzenia zakresu stosowalności przekształcenia Fouriera, i w konsekwencji rozszerzenia analizy częstotliwościowej na sygnały nie mające $\mathscr{F}$-transformaty w sensie zwykłym, wprowadzane są jego różnego rodzaju uogólnienia. Jednym z takich uogólnień jest *przekształcenie Fouriera w sensie granicznym*. Koncepcja tego przekształcenia polega na konstrukcji odpowiednich ciągów aproksymujących w dziedzinie czasu i dziedzinie częstotliwości i zdefiniowaniu pary transformat Fouriera jako pary granic tych ciągów.

Rozważmy sygnał $x(t)$ nie mający $\mathscr{F}$-transformaty w sensie zwykłym. Utwórzmy ciąg $\{x_\alpha(t)\colon \alpha \in \mathscr{R}^+\}$ sygnałów $\mathscr{F}$-transformowalnych w sensie zwykłym dla każdej wartości parametru α, który aproksymuje sygnał $x(t)$, tj. spełnia dla każdego t warunek:

$$\lim_{\alpha\to 0} x_\alpha(t) = x(t). \tag{3.8}$$

Ponieważ każdy element ciągu $\{x_\alpha(t)\}$ jest $\mathscr{F}$-transformowalny w sensie zwykłym, zatem ciągowi temu odpowiada w dziedzinie częstotliwości ciąg $\{X_\alpha(\omega)\}$ zwykłych transformat Fouriera, takich że $X_\alpha(\omega) = \mathscr{F}[x_\alpha(t)]$. Para $x_\alpha(t) \leftrightarrow X_\alpha(\omega)$ jest parą transformat Fouriera w sensie zwykłym.

Definicja 3.3. Jeżeli $\lim_{\alpha\to 0} x_\alpha(t) = x(t)$ dla każdego t oraz $\lim_{\alpha\to 0} X_\alpha(\omega) = X(\omega)$ dla każdego ω, to parę $x(t) \leftrightarrow X(\omega)$ nazywamy *parą transformat Fouriera w sensie granicznym*.

Ciągi aproksymujące (3.8) konstruuje się zwykle mnożąc sygnał $x(t)$ nie mający $\mathscr{F}$-transformaty w sensie zwykłym przez funkcje typu $\mathrm{e}^{-\alpha t}\,\boldsymbol{1}(t)$, jeśli $t \in [0,\infty)$, oraz $\mathrm{e}^{-\alpha|t|}$ lub $\mathrm{e}^{-\alpha t^2}$, jeśli $t \in (-\infty,\infty)$, dostatecznie szybko malejące do zera, gdy $t \to \pm\infty$. Należy podkreślić, że chcąc wyprowadzić daną parę w sensie granicznym, należy każdorazowo skonstruować ciąg aproksymujący sygnał $x(t)$ i wykazać, że odpowiadający mu ciąg transformat jest zbieżny do transformaty $X(\omega)$ dla każdej wartości ω. Cechą charakterystyczną takich ciągów jest to, że często dążą one w granicy do wielkości dystrybucyjnych w dziedzinie czasu lub dziedzinie częstotliwości. Przykłady par transformat Fouriera w sensie granicznym omówimy w p. 3.4.2.

3.2. Widmo sygnału

3.2.1. Interpretacja widmowa transformaty Fouriera

Całkę (3.4) możemy uważać za postać graniczną zespolonego szeregu Fouriera (2.57). Przejścia graniczne dokonuje się formalnie, zwiększając okres T_0 we wzorze (2.57) do nieskończoności. Pulsacja dyskretna $k\omega_0$ przechodzi wówczas w pulsację ciągłą ω, a czynnik $e^{j\omega t}$, występujący pod całką we wzorze (3.4), możemy traktować jako zespoloną reprezentację sygnału sinusoidalnego o pulsacji ω zmieniającej się ciągle w przedziale $(-\infty, +\infty)$.

Całkowanie względem zmiennej ω we wzorze (3.4) możemy więc uważać za operację graniczną sumowania po wszystkich wartościach ω elementarnych sygnałów harmonicznych $e^{j\omega t}$ ważonych przez wartości funkcji $X(\omega)$. Przez analogię do interpretacji współczynnika X_k zespolonego szeregu Fouriera jako zespolonej amplitudy k-tej składowej harmonicznej tego szeregu (por. p.2.4.4), funkcję $X(\omega)$ możemy zatem interpretować jako gęstość zespolonych amplitud sumowanych sygnałów przypadającą na jednostkę pulsacji. Z tego względu $\mathscr{F}$-transformatę $X(\omega)$ sygnału $x(t)$ nazywamy jego *gęstością widmową* lub krótko jego *widmem*.

Widmo stanowi podstawową charakterystykę sygnału w dziedzinie częstotliwości. Jest to charakterystyka zespolona określona zarówno dla dodatnich, jak i ujemnych, niefizycznych pulsacji. Reprezentacja sygnału w dziedzinie częstotliwości za pomocą widma ma więc aspekt czysto matematyczny, a samo pojęcie widma nie ma bezpośredniej interpretacji fizycznej.

3.2.2. Widmo amplitudowe i widmo fazowe

Widmo sygnału może być zapisane w postaci biegunowej:

$$X(\omega) = |X(\omega)|\, e^{j \arg X(\omega)} \triangleq A(\omega)\, e^{j\varphi(\omega)}\,. \tag{3.9}$$

Funkcję $|X(\omega)| \triangleq A(\omega)$ nazywamy *widmem amplitudowym*, a funkcję $\arg X(\omega) \triangleq \varphi(\omega)$+ – *widmem fazowym* sygnału. Widmo amplitudowe i widmo fazowe są już rzeczywistymi charakterystykami sygnału i dla $\omega \geqslant 0$ mają wyraźną interpretację fizyczną. Opisują one gęstość amplitudy i odpowiednio gęstość fazy elementarnych składowych harmonicznych sygnału w funkcji pulsacji ω.

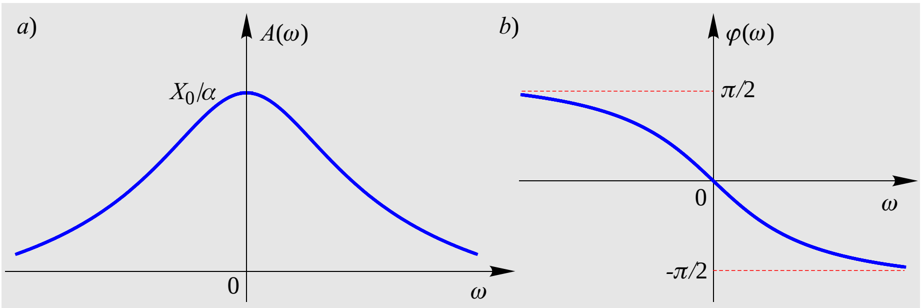
Przykład 3.3. Wyznaczymy widma amplitudowe i widma fazowe sygnałów rozpatrywanych w przykładach 3.1 i 3.2.

Ze wzoru (3.2) wynika bezpośrednio, że dla sygnału wykładniczego malejącego:

$$A(\omega) = \frac{X_0}{\sqrt{\alpha^2 + \omega^2}}, \tag{3.10}$$

$$\varphi(\omega) = -\operatorname{arctg}\frac{\omega}{\alpha}. \tag{3.11}$$

Wykresy widma amplitudowego i widma fazowego tego sygnału są przedstawione na rys. 3.1a,b. Wykres widma amplitudowego jest charakterystyczny dla *sygnałów dolnopasmowych*, dla których gęstość widmowa maleje do zera wraz ze wzrostem pulsacji do nieskończoności. O szybkości z jaką widmo to maleje decyduje wartość parametru α. Dla $\omega = \alpha$ widmo amplitudowe maleje do wartości $\sqrt{2}$ krotnie mniejszej od wartości maksymalnej w punkcie $\omega = 0$ (w mierze decybelowej odpowiada to spadkowi o 3 dB).



Rys. 3.1. Widmo amplitudowe (a) i widmo fazowe (b) sygnału wykładniczego malejącego

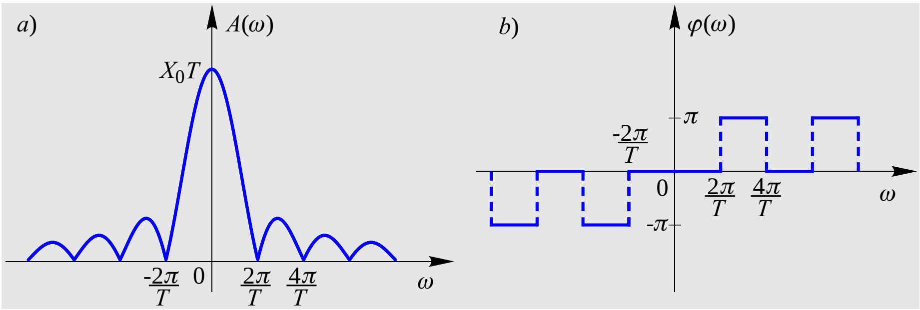
Ze wzoru (3.3) wynika z kolei, że widmo amplitudowe impulsu prostokątnego ma postać:

$$A(\omega) = X_0 T \left|\operatorname{Sa}\frac{\omega T}{2}\right|. \tag{3.12}$$

Jego wykres jest pokazany na rys. 3.2a. Widmo fazowe jest przedziałami stałe i w poszczególnych przedziałach przybiera wartości 0 lub $\pm\pi$. Wykres widma fazowego jest pokazany na rys. 3.2b. Impuls prostokątny jest również sygnałem dolnopasmowym, gdyż jego widmo amplitudowe zanika do zera, gdy $\omega \to \infty$. Widmo to nie maleje jednak monotonicznie, tak jak w przypadku sygnału wykładniczego malejącego. Ma ono charakterystyczną strukturę „listkową". Przedział pulsacji $|\omega| < 2\pi/T$ obejmuje tzw. *listek główny*, a pozostałe przedziały – *listki boczne* widma. W punktach $\omega = n2\pi/T$, $n \neq 0$ widmo impulsu prostokątnego jest równe zeru.

3.2.3. Widmo jako funkcja częstotliwości

W zagadnieniach praktycznych, zwłaszcza w analizie pomiarowej, charakterystyki widmowe sygnału wygodnie jest rozpatrywać nie jako funkcje zmiennej ω –



Rys. 3.2. Widmo amplitudowe (a) i widmo fazowe (b) impulsu prostokątnego

pulsacji, lecz jako funkcję zmiennej $f = \omega/2\pi$ – częstotliwości. Przekształcenia Fouriera proste i odwrotne, wyrażone w funkcji zmiennej f przebierają postać:

$$X(f) = \int_{-\infty}^{\infty} x(t)\,\mathrm{e}^{-\mathrm{j}2\pi f t}\,\mathrm{d}t, \tag{3.13}$$

$$x(t) = \int_{-\infty}^{\infty} X(f)\,\mathrm{e}^{\mathrm{j}2\pi f t}\,\mathrm{d}f. \tag{3.14}$$

We wszystkich wzorach określających widma sygnałów w funkcji zmiennej ω należy wówczas podstawić $\omega = 2\pi f$. W rozważaniach analitycznych wygodniej jest jednak posługiwać się zmienną ω. Z tego względu charakterystyki widmowe będziemy dalej rozpatrywać z reguły jako funkcję pulsacji.

3.2.4. Podstawowe właściwości widm

Założymy najpierw, że sygnały są rzeczywiste. Korzystając wprost z definicji (3.1), można łatwo wykazać, że widmo $X(\omega)$ sygnału rzeczywistego $x(t)$ jest *funkcją hermitowską*, tj. taką funkcją zespoloną, której część rzeczywista jest parzysta, część urojona zaś – nieparzysta. Konsekwencją tej właściwości jest parzystość widma amplitudowego oraz nieparzystość widma fazowego. Dla każdego ω spełnione są zatem równości (por. przykład 3.3):

$$A(\omega) = A(-\omega) \qquad \varphi(\omega) = -\varphi(-\omega). \tag{3.15}$$

Jeśli sygnał jest parzysty, jego widmo jest rzeczywiste i parzyste (por. przykład 3.2). Jeśli sygnał jest nieparzysty, widmo jest urojone i nieparzyste.

Wymienione właściwości przestają być słuszne dla sygnałów zespolonych. Dla dowolnych sygnałów spełnione są natomiast następujące właściwości związane z operacją odbicia sygnału względem osi rzędnych i operacją sprzężenia. Jeśli $x(t) \to X(\omega)$, to:

$$x(-t) \leftrightarrow X(-\omega), \tag{3.16}$$

$$x^*(t) \leftrightarrow X^*(-\omega), \tag{3.17}$$

$$x^*(-t) \leftrightarrow X^*(\omega). \tag{3.18}$$

3.3. Twierdzenia

Omówimy obecnie dalsze właściwości przekształcenia Fouriera. Ze względu na ich znaczenie będą one ujęte w formie twierdzeń. Twierdzenia te dotyczą operacji wykonywanych na sygnale lub na sygnałach w dziedzinie czasu lub dziedzinie częstotliwości i stanowią swojego rodzaju „słownik” tłumaczący efekt wykonania danej operacji w jednej dziedzinie na efekt towarzyszący mu w drugiej dziedzinie. Ułatwiają one wyznaczanie widm sygnałów, bez konieczności często skomplikowanego obliczania tych widm na podstawie wzoru definicyjnego (3.1).

Twierdzenia przytoczymy bez dowodów (niektóre z nich można bez trudu przeprowadzić, korzystając wprost z definicji 3.1). Pominiemy także dyskusję warunków, przy których są one spełnione. Podamy natomiast ich interpretację.

Przyjmiemy założenie, że $x(t) \leftrightarrow X(\omega)$ oraz $y(t) \leftrightarrow Y(\omega)$ są parami transformat Fouriera, przy czym sygnały $x(t)$ i $y(t)$ są rzeczywiste lub zespolone.

Twierdzenie 3.1 (*o liniowości*)

$$ax(t) + by(t) \leftrightarrow aX(\omega) + bY(\omega). \tag{3.19}$$

Przekształcenie Fouriera jest operacją liniową, tzn. widmo sygnału będącego kombinacją liniową pewnych sygnałów jest kombinacją liniową o tych samych współczynnikach widm tych sygnałów. Właściwość liniowości wynika wprost z definicji 3.1 prostego przekształcenia Fouriera.

Twierdzenie 3.2 (*o symetrii*)

$$x(t) \leftrightarrow X(\omega) \Rightarrow X(t) \leftrightarrow 2\pi x(-\omega). \tag{3.20}$$

Z twierdzenia tego wynika, że przy przekształceniu Fouriera kształt sygnału i widma jest cechą wymienną, tzn. jeżeli sygnał o kształcie $x(t)$ ma widmo $X(\omega)$, to sygnał $X(t)$ o kształcie tego widma ma widmo o kształcie sygnału pierwotnego, odbite zwierciadlanie względem osi rzędnych i pomnożone dodatkowo przez 2π. Właściwość ta jest jednym z przejawów dualizmu, jaki występuje między dziedziną czasu a dziedziną częstotliwości.

Twierdzenie 3.3 (*o zmianie skali*)

$$x\left(\frac{t}{a}\right) \leftrightarrow aX(a\omega), \qquad a \in \mathscr{R}^{+}. \tag{3.21}$$

Zmianie skali czasu sygnału, a więc jego „rozciągnięciu” ($a > 1$) lub „ściśnięciu” w czasie ($a < 1$), towarzyszą w dziedzinie częstotliwości dwa efekty. Rozciągnięcie sygnału w czasie oznacza jego spowolnienie. Zmniejszają się tym samym prędkości jego zmian. Widmo skupia się wówczas wokół mniejszych pulsacji i jednocześnie gęstość widmowa w tym zakresie proporcjonalnie wzrasta. W przypadku ściśnięcia sygnału, efekty są przeciwne.

Twierdzenie 3.4 (*o przesunięciu w dziedzinie czasu*)

$$x(t - t_0) \leftrightarrow X(\omega)\,\mathrm{e}^{-\mathrm{j}\omega t_0}\,. \tag{3.22}$$

Opóźnienie sygnału w czasie ($t_0 > 0$) lub jego przyspieszenie ($t_0 < 0$) odpowiada mnożeniu widma przez czynnik $\mathrm{e}^{-\mathrm{j}\omega t_0}$. Widmo amplitudowe nie ulega przy tym zmianie, a widmo fazowe zmienia się o składnik $-\omega t_0$. Widmo fazowe maleje ($t_0 > 0$) lub rośnie ($t_0 < 0$) liniowo w funkcji pulsacji z szybkością proporcjonalną do wartości bezwzględnej przesunięcia t_0. Wnioski te są w pełni zgodne z intuicyjną interpretacją operacji przesunięcia sygnału, która nie zmienia struktury amplitudowej sygnału, a wpływa jedynie na jego strukturę fazową.

Twierdzenie 3.5 (*o przesunięciu w dziedzinie częstotliwości*)

$$x(t)\,\mathrm{e}^{\mathrm{j}\omega_0 t} \leftrightarrow X(\omega - \omega_0). \tag{3.23}$$

Twierdzenie to jest dualne do twierdzenia 3.4. Przesunięcie widma sygnału $x(t)$ wzdłuż osi pulsacji w prawo o wartość $\omega_0 > 0$ uzyskujemy w wyniku mnożenia tego sygnału przez zespolony sygnał harmoniczny $\mathrm{e}^{\mathrm{j}\omega_0 t}$ o pulsacji ω_0. Operacja mnożenia sygnału harmonicznego przez sygnał $x(t)$ jest nazywana operacją *modulacji* sygnału harmonicznego. Z tego względu twierdzenie to nosi także nazwę *twierdzenia o modulacji*.

Przesunięcie widma sygnału w lewo o wartość $\omega_0 > 0$ odpowiada mnożeniu tego sygnału przez sygnał $\mathrm{e}^{-\mathrm{j}\omega_0 t}$:

$$x(t)\,\mathrm{e}^{-\mathrm{j}\omega_0 t} \leftrightarrow X(\omega + \omega_0). \tag{3.24}$$

Dodając i odejmując stronami wzory (3.23) i (3.24) i korzystając ze wzorów Eulera:

$$\cos\omega_0 t = \frac{\mathrm{e}^{\mathrm{j}\omega_0 t} + \mathrm{e}^{-\mathrm{j}\omega_0 t}}{2}, \qquad \sin\omega_0 t = \frac{\mathrm{e}^{\mathrm{j}\omega_0 t} - \mathrm{e}^{-\mathrm{j}\omega_0 t}}{2\,\mathrm{j}}, \tag{3.25}$$

otrzymujemy dwa warianty twierdzenia o modulacji dla przypadku modulacji rzeczywistego sygnału harmonicznego:

$$x(t)\cos\omega_0 t \leftrightarrow \frac{1}{2}[X(\omega-\omega_0) + X(\omega+\omega_0)], \tag{3.26}$$

$$x(t)\sin\omega_0 t \leftrightarrow \frac{1}{2\,\mathrm{j}}[X(\omega-\omega_0) - X(\omega+\omega_0)]. \tag{3.27}$$

Ze wzoru (3.26) wynika, że modulacja rzeczywistego sygnału harmonicznego $\cos\omega_0 t$ sygnałem $x(t)$ powoduje rozszczepienie widma $X(\omega)$ tego sygnału na dwie części o tym samym kształcie, przesunięte po osi pulsacji do punktów $\pm\omega_0$. Gęstość widmowa tych części maleje przy tym dwukrotnie. Podobny efekt uzyskuje się w wyniku modulacji sygnału $\sin\omega_0 t$. Widma amplitudowe są w obu przypadkach identyczne. Różnice występują tylko między widmami fazowymi.

Na twierdzenie o modulacji będziemy się niejednokrotnie powoływać w rozdziale 9, w którym omawiać będziemy analogowe systemy modulacji.

Twierdzenie 3.6 (*o różniczkowaniu w dziedzinie czasu*)

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) \leftrightarrow \mathrm{j}\,\omega X(\omega). \tag{3.28}$$

Różniczkowaniu sygnału w dziedzinie czasu odpowiada mnożenie jego widma przez $\mathrm{j}\,\omega$ w dziedzinie częstotliwości. Operacja ta zwiększa gęstość widmową sygnału w zakresie dużych pulsacji.

Twierdzenie 3.7 (*o całkowaniu w dziedzinie czasu*)

$$\int\limits_{-\infty}^{t} x(\tau)\,\mathrm{d}\tau \leftrightarrow \frac{1}{\mathrm{j}\,\omega}X(\omega). \tag{3.29}$$

Całkowaniu sygnału w dziedzinie czasu odpowiada dzielenie jego widma przez $\mathrm{j}\,\omega$ w dziedzinie częstotliwości. Całkowanie sygnału uwypukla fragmenty jego widma z zakresie małych pulsacji i zmniejsza gęstość widmową w zakresie dużych pulsacji.

Twierdzenie 3.8 (*o splocie w dziedzinie czasu*)

$$x(t) * y(t) \triangleq \int\limits_{-\infty}^{\infty} x(t-\tau)y(\tau)\,\mathrm{d}\tau \leftrightarrow X(\omega)Y(\omega). \tag{3.30}$$

Twierdzenie to odgrywa kluczową rolę w zagadnieniach przetwarzania sygnałów przez układy liniowe. Orzeka ono, że splataniu sygnałów w dziedzinie czasu odpowiada mnożenie ich widm w dziedzinie częstotliwości. Tym samym z reguły złożoną obliczeniowo operację na sygnałach można zastąpić prostą operacją na ich widmach. W konsekwencji umożliwia to algebraizację problemu obliczania sygnału na wyjściu układu (por. p. 7.3.2).

Twierdzenie 3.9 (*o splocie w dziedzinie częstotliwości*)

$$x(t)y(t) \leftrightarrow \frac{1}{2\pi}[X(\omega) * Y(\omega)] \triangleq \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega-\lambda)Y(\lambda)\,\mathrm{d}\lambda. \tag{3.31}$$

Mnożeniu sygnałów w dziedzinie czasu odpowiada splatanie ich widm w dziedzinie częstotliwości. Twierdzenie 3.9 jest dualne względem twierdzenia 3.8 o splocie w dziedzinie czasu. Zauważmy, że korzystając z tego twierdzenia można dowieść twierdzenia o modulacji. Wystarczy w tym celu przyjąć, że jednym z mnożonych sygnałów jest sygnał harmoniczny: $\mathrm{e}^{\mathrm{j}\omega_0 t}$, $\cos\omega_0 t$ lub $\sin\omega_0 t$.

Twierdzenie o splocie w dziedzinie częstotliwości może być także sformułowane w bardziej ogólnej formie (por. wzór (3.17)):

$$x(t)y^*(t) \leftrightarrow \frac{1}{2\pi}[X(\omega) * Y^*(-\omega)]. \tag{3.32}$$

Twierdzenie 3.10 (*uogólnione twierdzenie Rayleigha*)

$$\int_{-\infty}^{\infty} x(t)y^*(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)Y^*(\omega)\,\mathrm{d}\omega. \tag{3.33}$$

Twierdzenie to, bardzo ważne w teorii sygnałów, odzwierciedla w jeszcze innej formie wieloraką strukturę powiązań między dziedziną czasu i dziedziną częstotliwości. Jak pamiętamy (por. komentarz w p. 3.2.3), przekształcenie Fouriera można traktować jako odwzorowanie przestrzeni L_t^2 sygnałów w przestrzeń L_ω^2 transformat. Twierdzenie Rayleigha orzeka, że (z dokładnością do współczynnika 2π) odwzorowanie to zachowuje iloczyn skalarny w tych przestrzeniach, tzn. iloczyn skalarny dwóch sygnałów w przestrzeni L_t^2 jest równy iloczynowi skalarnemu ich widm w przestrzeni L_ω^2 podzielonemu przez 2π.

Twierdzenie 3.11 (*twierdzenie Parsevala*)

$$\int_{-\infty}^{\infty} |x(t)|^2\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2\,\mathrm{d}\omega. \tag{3.34}$$

Twierdzenie Parsevala, nazywane również *twierdzeniem o energii*, wynika z twierdzenia Rayleigha przy podstawieniu $y(t) = x(t)$. Wyraża ono równość (z dokładnością do współczynnika 2π) norm w przestrzeniach L_t^2 oraz L_ω^2. Zgodnie z tym twierdzeniem energię sygnału możemy także obliczać w dziedzinie częstotliwości jako całkę z kwadratu widma amplitudowego $|X(\omega)|^2$ sygnału podzieloną przez 2π. Z tego względu funkcja $|X(\omega)|^2 = A^2(\omega)$ jest nazywana *widmem energii*. Widmo energii jest także oznaczane $\Phi_x(\omega)$. Jest ono jeszcze jedną charakterystyką sygnału w dziedzinie częstotliwości opisującą rozkład (gęstość widmową) jego energii wzdłuż osi pulsacji. Do charakterystyki tej powrócimy jeszcze w p. 5.1.4, gdzie uzasadnimy przyjęte oznaczenie $\Phi_x(\omega)$.

3.4. Przykłady par transformat Fouriera

Tylko nieliczne pary transformat Fouriera można wyprowadzić wprost z definicji 3.1. Większość z nich można natomiast wyznaczyć bez trudu, korzystając z odpowiednich twierdzeń. Kontynuując przykłady 3.1 i 3.2, omówimy obecnie dalsze często wykorzystywane pary transformat. W przypadku par transformat w sensie granicznym podamy konstrukcje odpowiednich ciągów aproksymujących.

3.4.1. Pary transformat w sensie zwykłym

Sygnał wykładniczy malejący

$$X_0\,\mathrm{e}^{-\alpha t} \leftrightarrow \frac{X_0}{\alpha + \mathrm{j}\,\omega} \tag{3.35}$$

Parę tę wyznaczyliśmy na podstawie definicji w przykładzie 3.1. Widmo amplitudowe i fazowe tego sygnału są przedstawione na rys. 3.1.

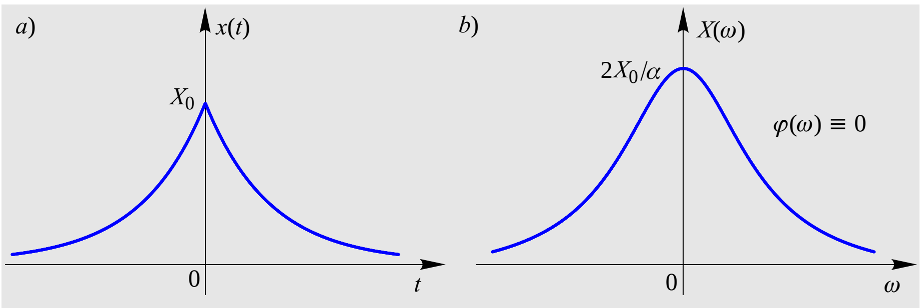
Sygnał wykładniczy dwustronny

$$X_0\,\mathrm{e}^{-\alpha|t|} \leftrightarrow \frac{2X_0\alpha}{\alpha^2 + \omega^2} \tag{3.36}$$

Parę tę wyprowadzimy wprost z definicji:

$$\begin{aligned} X(\omega) &= X_0 \int\limits_{-\infty}^{0} \mathrm{e}^{\alpha t}\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t + X_0 \int\limits_{0}^{\infty} \mathrm{e}^{-\alpha t}\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t = \\ &= X_0 \int\limits_{-\infty}^{0} \mathrm{e}^{(\alpha-\mathrm{j}\omega)t}\, \mathrm{d}t + X_0 \int\limits_{0}^{\infty} \mathrm{e}^{-(\alpha+\mathrm{j}\omega)t}\, \mathrm{d}t = \\ &= \frac{X_0}{\alpha - \mathrm{j}\,\omega}\, \mathrm{e}^{(\alpha-\mathrm{j}\omega)t} \Bigg|_{-\infty}^{0} - \frac{X_0}{\alpha + \mathrm{j}\,\omega}\, \mathrm{e}^{-(\alpha+\mathrm{j}\omega)t} = \\ &= \frac{X_0}{\alpha - \mathrm{j}\,\omega} + \frac{X_0}{\alpha + \mathrm{j}\,\omega} = X_0 \frac{\alpha + \mathrm{j}\,\omega + \alpha - \mathrm{j}\,\omega}{(\alpha - \mathrm{j}\,\omega)(\alpha + \mathrm{j}\,\omega)} = \frac{2X_0\alpha}{\alpha^2 + \omega^2}. \end{aligned}$$

Ponieważ sygnał jest w tym przypadku parzysty, jego widmo jest rzeczywiste i parzyste (rys. 3.3).



Rys. 3.3. Para transformat (3.36)

Impuls prostokątny

$$X_0 \sqcap \left(\frac{t}{T}\right) \leftrightarrow X_0 T \,\mathrm{Sa}\, \frac{\omega T}{2} \tag{3.37}$$

Również ta para została wyprowadzona wprost z definicji w przykładzie 3.2. Widma amplitudowe i fazowe impulsu prostokątnego są pokazane na rys. 3.2.

Sygnał Sa

$$X_0 \,\mathrm{Sa}\, \omega_0 t \leftrightarrow \frac{\pi X_0}{\omega_0} \sqcap \left(\frac{\omega}{2\omega_0}\right). \tag{3.38}$$

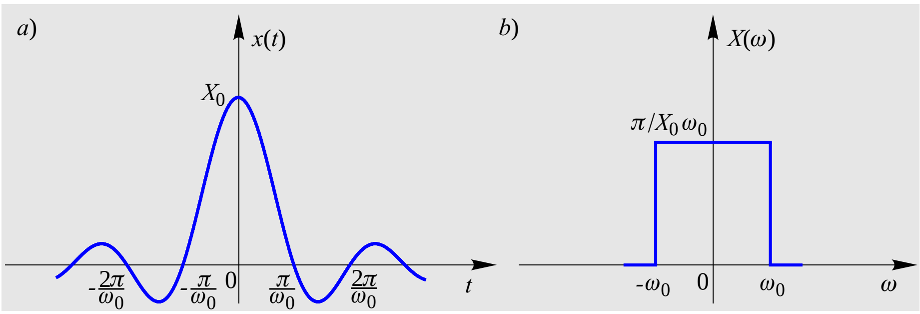
Wyznaczenie tej pary na podstawie definicji jest bardzo skomplikowane. Można ją natomiast łatwo wyprowadzić, korzystając z twierdzenia o symetrii względem pary rozpatrywanej w przykładzie 3.2. Ponieważ widmo impulsu prostokątnego ma kształt funkcji Sa, zatem zgodnie z tym twierdzeniem widmo sygnału

Sa ma kształt prostokątny. Aby wyznaczyć współczynniki tego widma, wystarczy we wzorze (3.3) zamienić zmienną ω na zmienną t oraz w miejsce $T/2$ podstawić ω_0. Otrzymamy wtedy:

$$2\omega_0 X_0 \operatorname{Sa} \omega_0 t \leftrightarrow 2\pi X_0 \sqcap \left(-\frac{\omega}{2\omega_0}\right) = 2\pi X_0 \sqcap \left(\frac{\omega}{2\omega_0}\right),$$

skąd wynika już słuszność pary (3.38).

Ze wzoru (3.38) wynika ważna interpretacja sygnału Sa. Widmo tego sygnału zawiera jedynie składowe częstotliwościowe o pulsacjach $|\omega| < \omega_0$, przy czym gęstość widmowa w tym przedziale jest stała. Z tego względu sygnał Sa jest nazywany *idealnym sygnałem dolnopasmowym*. Pulsacja ω_0 oscylacji sygnału Sa jest *pulsacją graniczną* jego widma.

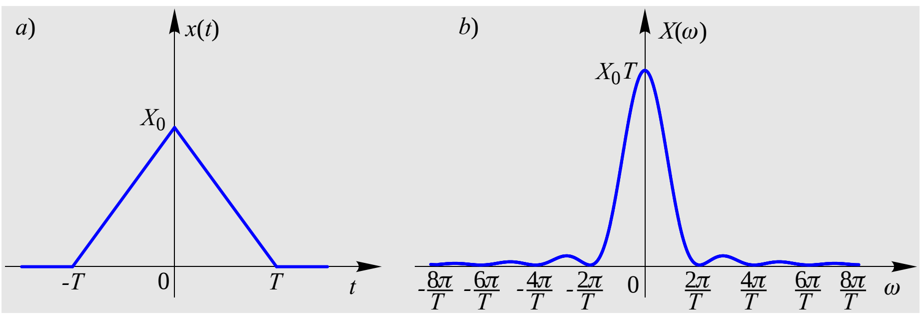


Rys. 3.4. Para transformat (3.38)

Impuls trójkątny

$$X_0 \wedge \left(\frac{t}{T}\right) \leftrightarrow X_0 T \operatorname{Sa}^2 \frac{\omega T}{2}. \tag{3.39}$$

Parę tę możemy wyznaczyć na podstawie twierdzenia o splocie w dziedzinie czasu, przyjmując we wzorze (3.30) $x(t) = y(t) = \sqcap(t/T)$.

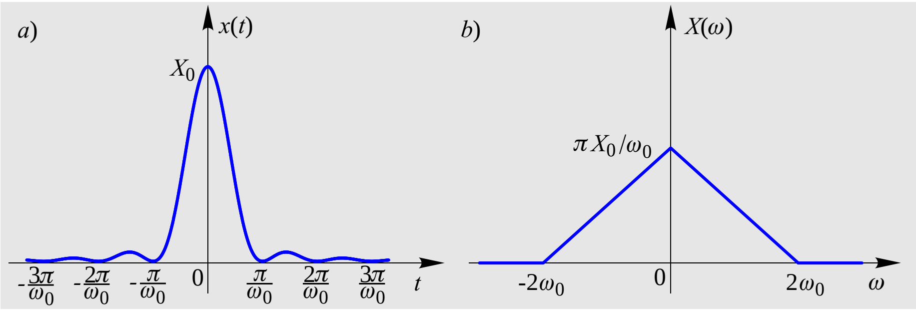


Rys. 3.5. Para transformat (3.39)

Sygnał Sa^2

$$X_0 \,\mathrm{Sa}^2\, \omega_0 t \leftrightarrow \frac{\pi X_0}{\omega_0} \wedge \left(\frac{\omega}{2\omega_0}\right). \tag{3.40}$$

Parę tę możemy wyprowadzić, stosując do pary (3.39) twierdzenie o symetrii.

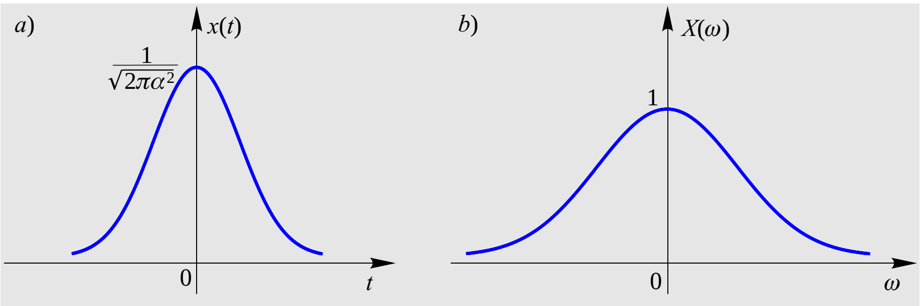


Rys. 3.6. Para transformat (3.40)

Sygnał Gaussa

$$\frac{1}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{t^2}{2\alpha^2}\right) \leftrightarrow \exp\left(-\frac{\alpha^2\omega^2}{2}\right). \tag{3.41}$$

Wyprowadzenie tej pary jest skomplikowane. Jest ona jednym z nielicznych przykładów par, w których zarówno sygnał, jak i widmo, mają ten sam kształt.

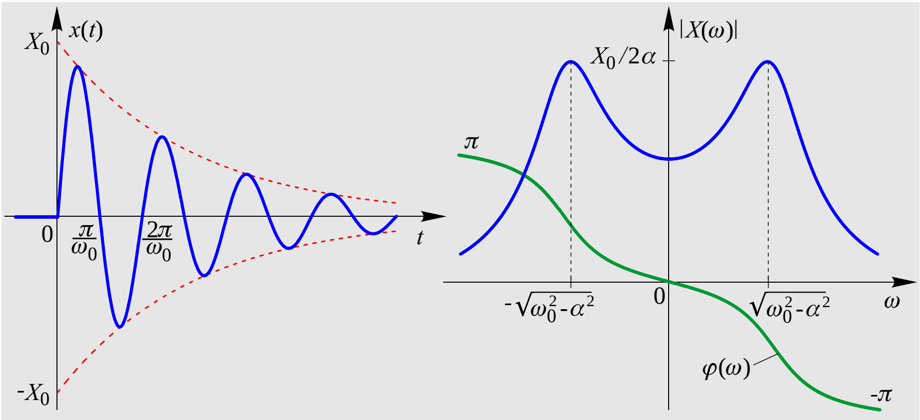


Rys. 3.7. Para transformat (3.41)

Sygnał sinusoidalny malejący wykładniczo

$$X_0 \,\mathrm{e}^{-\alpha t} \sin \omega_0 t \leftrightarrow X_0 \frac{\omega_0}{(\alpha + \mathrm{j}\,\omega)^2 + \omega_0^2}. \tag{3.42}$$

Parę tę można wyprowadzić z twierdzenia o modulacji (3.27). Podobnie jak wszystkie poprzednie sygnały, sygnał sinusoidalny tłumiony wykładniczo jest sygnałem dolnopasmowym. Jeśli $\omega_0 > \alpha$, to w punktach $\pm\sqrt{\omega_0^2 - \alpha^2}$ widmo tego sygnału ma maksima równe $X_0/2\alpha$.



Rys. 3.8. Para transformat (3.42)

Prostokątny impuls radiowy

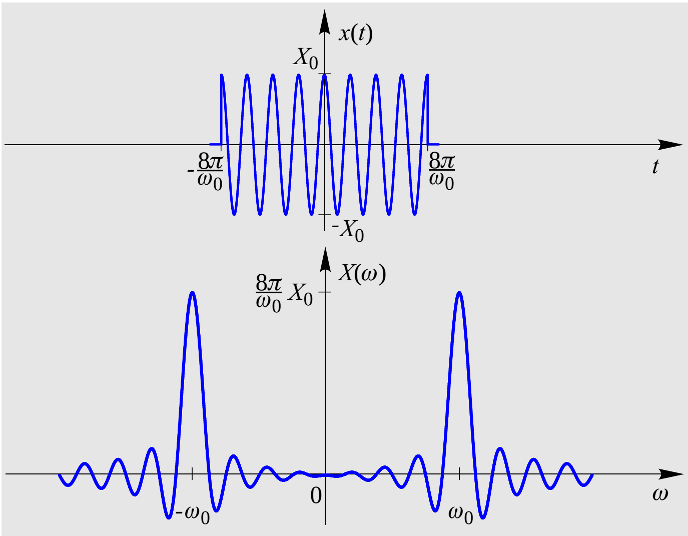
$$X_0 \cos \omega_0 t \sqcap \left(\frac{t}{T}\right) \leftrightarrow \frac{X_0 T}{2} \left[\mathrm{Sa}\,\frac{(\omega - \omega_0)T}{2} + \mathrm{Sa}\,\frac{(\omega + \omega_0)T}{2}\right]. \qquad (3.43)$$

Para ta wynika wprost z twierdzenia o modulacji (3.26). Wykresy na rys. 3.9 zostały sporządzone dla przypadku $T = 16\pi/\omega_0$. Jeśli czas trwania impulsu jest dostatecznie krótki w porównaniu z okresem fali sinusoidalnej, tzn. spełniona jest nierówność $T \ll 2\pi/\omega_0$, to jego widmo jest skupione głównie w wąskich przedziałach pulsacji wokół punktów $\pm\omega_0$. Sygnały takie noszą nazwę *wąskopasmowych*. Sygnałami wąskopasmowymi są sygnały występujące w większości systemów modulacji.

3.4.2. Pary transformat w sensie granicznym

Impuls Diraca

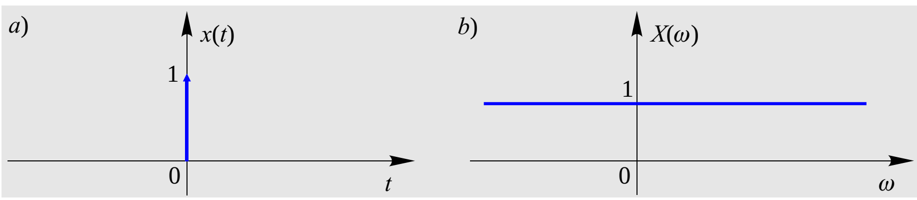
$$\delta(t) \leftrightarrow 1. \qquad (3.44)$$



Rys. 3.9. Para transformat (3.43)

Przykładem ciągu par transformat Fouriera w sensie zwykłym aproksymującym parę (3.44) jest ciąg (3.41). Jeśli parametr α dąży do zera, ciąg lewostronny dąży do impulsu Diraca $\delta(t)$ (por. rys. 1.25). Jednocześnie ciąg prawostronny dąży do funkcji stałej równej 1. Rozkład gęstości widmowej impulsu Diraca jest więc równomierny w całym zakresie pulsacji. Widmo o tej właściwości nazywamy *widmem białym*.

Otrzymany rezultat wydaje się na pozór zaskakujący, oznacza bowiem, że impuls Diraca zawiera składowe o wszystkich częstotliwościach, z których każda wnosi taki sam wkład w jego strukturę widmową. Należy jednak wziąć pod uwagę, że taka struktura widma impulsu Diraca jest uzasadniona nieskończenie szybką zmianą jego wartości w chwili $t = 0$.

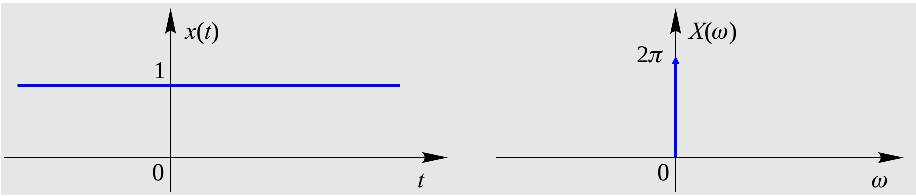


Rys. 3.10. Para transformat (3.44)

Sygnał stały

$$1 \leftrightarrow 2\pi\delta(\omega). \tag{3.45}$$

Para ta jest dualna względem pary (3.44). Można ją wykazać na podstawie twierdzenia o symetrii, uwzględniając, że $\delta(\omega) = \delta(-\omega)$. W sygnale stałym występuje tylko jedna, zerowa prędkość zmian. Efektem tego jest skupienie jego gęstości widmowej jedynie w punkcie $\omega = 0$, przy czym gęstość w tym punkcie jest nieskończona. Gęstość ta jest reprezentowana dystrybucją Diraca w dziedzinie widmowej o polu 2π. Występowanie tego współczynnika jest wynikiem niepełnej symetrii między obu dziedzinami.



Rys. 3.11. Para transformat (3.45)

Sygnał sgn t

$$\operatorname{sgn} t \leftrightarrow \frac{2}{\mathrm{j}\,\omega}. \tag{3.46}$$

Parę tę wyprowadzimy, konstruując odpowiednie ciągi aproksymujące. Jako ciąg aproksymujący sygnał $\operatorname{sgn} t$ wybierzemy ciąg $x_\alpha(t) = \mathrm{e}^{-\alpha|t|}\operatorname{sgn} t$. Elementy tego ciągu są $\mathscr{F}$-transformowalne w sensie zwykłym:

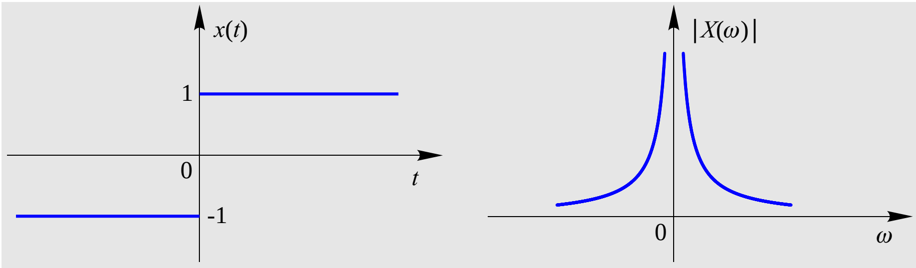
$$\begin{aligned}
X_\alpha(\omega) &= \int_{-\infty}^{\infty} \mathrm{e}^{-\alpha|t|}\operatorname{sgn} t\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \\
&= -\int_{-\infty}^{0} \mathrm{e}^{\alpha t}\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t + \int_{0}^{\infty} \mathrm{e}^{-\alpha t}\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \\
&= -\frac{1}{\alpha - \mathrm{j}\,\omega} + \frac{1}{\alpha + \mathrm{j}\,\omega} = \frac{-\mathrm{j}\,2\omega}{\alpha^2 + \omega^2}.
\end{aligned}$$

Ponieważ dla każdego ω granica ciągu $X_\alpha(\omega)$ przy α dążącym do zera jest równa $2/\mathrm{j}\,\omega$, dowodzi to słuszności pary (3.46).

Sygnał $\dfrac{1}{\pi t}$

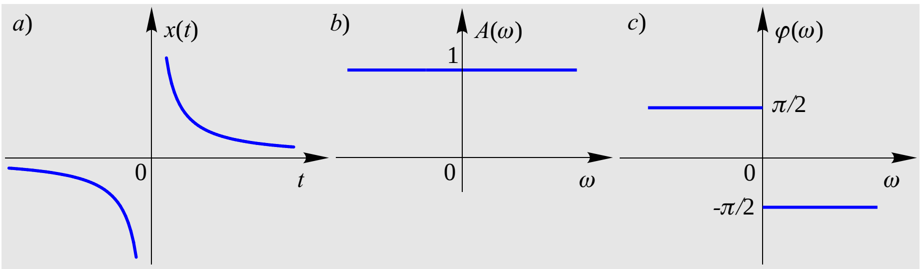
$$\frac{1}{\pi t} \leftrightarrow -\mathrm{j}\operatorname{sgn}\omega \tag{3.47}$$

Sygnał $1/\pi t$ odgrywa istotną rolę w teorii całkowego przekształcenia Hilberta (por. p. 10.1.7).



Rys. 3.12. Para transformat (3.46)

Jest to sygnał o ograniczonej mocy, a więc nie ma $\mathscr{F}$-transformaty w sensie zwykłym. Jego $\mathscr{F}$-transformatę w sensie granicznym można wyprowadzić, stosując twierdzenie o symetrii względem pary dualnej (3.46).



Rys. 3.13. Para transformat (3.47)

Skok jednostkowy $\mathit{1}(t)$

$$\mathbf{1}(t) \leftrightarrow \pi\delta(\omega) + \frac{1}{\mathrm{j}\,\omega}. \tag{3.48}$$

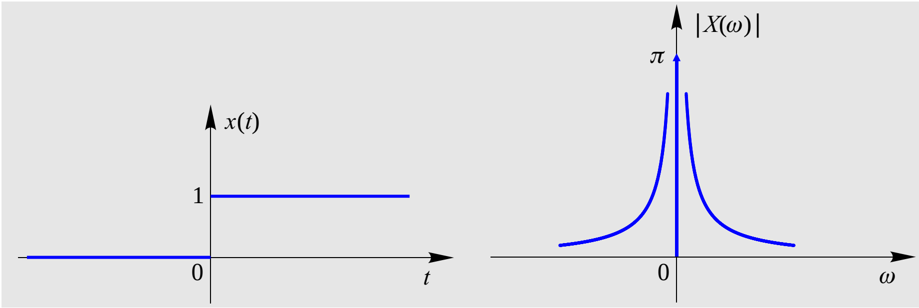
Ponieważ sygnał skoku jednostkowego możemy przedstawić w postaci $\mathbf{1}(t) = \frac{1}{2} + \frac{1}{2}\operatorname{sgn} t$, zatem korzystając z pary (3.46) otrzymujemy:

$$X(\omega) = \mathscr{F}[\mathbf{1}(t)] = \mathscr{F}\left[\frac{1}{2}\right] + \mathscr{F}\left[\frac{1}{2}\operatorname{sgn} t\right] = \pi\delta(\omega) + \frac{1}{\mathrm{j}\,\omega}.$$

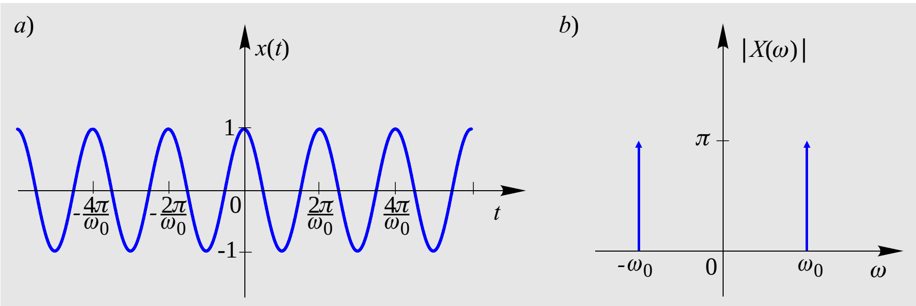
Sygnał $\cos\omega_0 t$

$$\cos\omega_0 t \leftrightarrow \pi[\delta(\omega-\omega_0) + \delta(\omega+\omega_0)]. \tag{3.49}$$

Para ta wynika z pary (3.45) i twierdzenia o modulacji (3.26). Ponieważ w sygnale $\cos\omega_0 t$ występuje tylko jedna pulsacja ω_0, widmo tego sygnału stanowią dwie dystrybucje Diraca występujące w punktach $\pm\omega_0$.



Rys. 3.14. Para transformat (3.48)

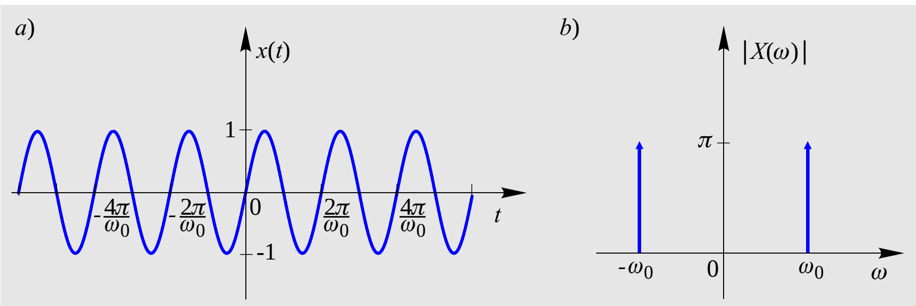


Rys. 3.15. Para transformat (3.49)

Sygnał $\sin \omega_0 t$

$$\sin \omega_0 t \leftrightarrow -\mathrm{j}\,\pi[\delta(\omega - \omega_0) - \delta(\omega + \omega_0)]. \tag{3.50}$$

Podobnie jak poprzednio, para ta wynika z pary (3.45) i twierdzenia o modulacji (3.27).



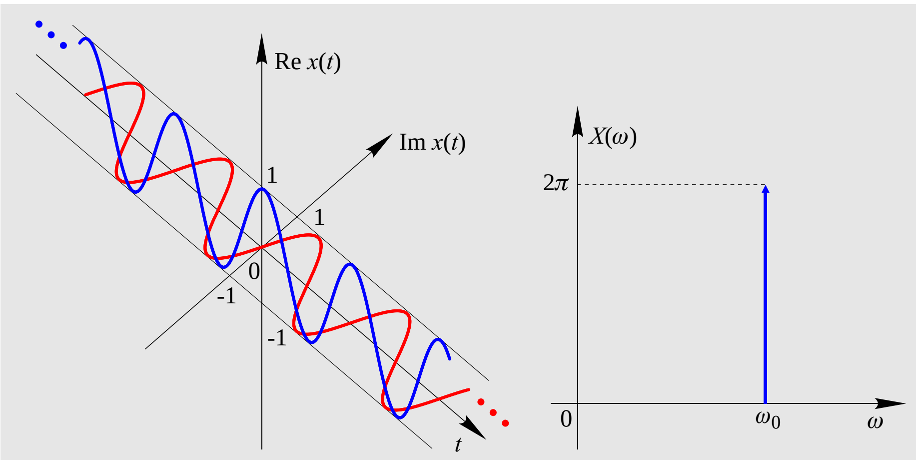
Rys. 3.16. Para transformat (3.50)

Sygnał $\mathrm{e}^{\mathrm{j}\omega_0 t}$

$$\mathrm{e}^{\mathrm{j}\omega_0 t} \leftrightarrow 2\pi\delta(\omega - \omega_0). \tag{3.51}$$

Parę otrzymujemy stosując do pary (3.45) twierdzenie o modulacji (3.25). Na rys. 3.17, ilustrującym tę parę, wykres zespolonego sygnału harmonicznego $\mathrm{e}^{\mathrm{j}\omega_0 t}$ został sporządzony w trójwymiarowym układzie współrzędnych, gdzie w płaszczyźnie rzeczywistej wykreślono część rzeczywistą, a w płaszczyźnie urojonej – część urojoną sygnału $\mathrm{e}^{\mathrm{j}\omega_0 t}$. Zwróćmy uwagę, że widmo tego sygnału jest reprezentowane tylko jedną dystrybucją Diraca występującą w punkcie ω_0. Jest to więc widmo prawostronne. Właściwość ta ma ścisły związek z faktem, że zespolony sygnał harmoniczny $\mathrm{e}^{\mathrm{j}\omega_0 t}$ jest sygnałem analitycznym reprezentującym rzeczywisty sygnał harmoniczny $\cos\omega_0 t$ (por. p. 1.2.7).

Para (3.51) stanowi punkt wyjścia do wyprowadzenia ogólnego wzoru określającego $\mathscr{F}$-transformatę Fouriera w sensie granicznym dowolnego sygnału okresowego reprezentowanego zespolonym szeregiem Fouriera (2.57).



Rys. 3.17. Para transformat (3.51)

Sygnały okresowe

Jeżeli $x(t)$ jest sygnałem okresowym o okresie $T_0 = 2\pi/\omega_0$ reprezentowanym zespolonym szeregiem Fouriera (2.57), to

$$x(t) = \sum_{k=-\infty}^{\infty} X_k \, \mathrm{e}^{\mathrm{j}k\omega_0 t} \leftrightarrow 2\pi \sum_{k=-\infty}^{\infty} X_k \delta(\omega - k\omega_0). \tag{3.52}$$

Para ta wynika z pary (3.51) i twierdzenia o liniowości przekształcenia Fouriera. Widmo w sensie granicznym sygnału $x(t)$ jest zatem ciągiem dystrybucji

Diraca występujących w punktach $k\omega_0$, $k = 0, \pm 1, \ldots$, o zespolonych wartościach pól równych współczynnikom X_k rozwinięcia sygnału $x(t)$ w zespolony szereg Fouriera. Widmo amplitudowe sygnału $x(t)$ jest ciągiem dystrybucji Diraca w punktach $k\omega_0$ o polach $|X_k|$ równych modułom tych współczynników. Ponieważ widmo fazowe nie ma charakteru gęstości widmowej, jest ono ciągiem zwykłych liczb $\arg X_k$, określonych w punktach $k\omega_0$ i przybiera wartości zerowe poza tymi punktami.

Widma sygnałów okresowych są skupione jedynie w dyskretnym zbiorze punktów osi pulsacji. Z tego względu są one nazywane widmami *dyskretnymi* lub *prążkowymi*.

Przykład 3.4. Wyznaczymy widmo unipolarnej fali prostokątnej $x(t)$ przedstawionej na rys. 1.23 i przedyskutujemy jak zmienia się ono w funkcji współczynnika wypełnienia tej fali.

Na podstawie wzoru (2.58) obliczamy współczynniki rozwinięcia fali w zespolony szereg Fouriera (w celu uniknięcia kolizji oznaczeń amplitudę fali oznaczamy A_0):

$$\begin{aligned} X_k &= \frac{1}{T_0}\int\limits_{-T_0/2}^{T_0/2} x(t)\,\mathrm{e}^{-\mathrm{j}k\omega_0 t}\,\mathrm{d}t = \frac{A_0}{T_0}\int\limits_{-T/2}^{T/2} \mathrm{e}^{-\mathrm{j}k\omega_0 t}\,\mathrm{d}t = \\ &= -\frac{A_0}{\mathrm{j}\,k\omega_0 T_0}\,\mathrm{e}^{-\mathrm{j}k\omega_0 t}\bigg|_{-T/2}^{T/2} = \frac{A_0}{\mathrm{j}\,k\omega_0 T_0}\left(\mathrm{e}^{\mathrm{j}k\omega_0 T/2} - \mathrm{e}^{-\mathrm{j}k\omega_0 T/2}\right) = \\ &= \frac{2A_0}{k\omega_0 T_0}\sin k\omega_0\frac{T}{2} = A_0\frac{T}{T_0}\,\mathrm{Sa}\,k\omega_0\frac{T}{2} = A_0\frac{T}{T_0}\,\mathrm{Sa}\,k\pi\frac{T}{T_0}, \end{aligned}$$

gdzie $\omega_0 = 2\pi/T_0$. Zgodnie ze wzorem (3.52) widmo fali $x(t)$ ma postać:

$$X(\omega) = 2\pi\sum_{k=-\infty}^{\infty} A_0\frac{T}{T_0}\,\mathrm{Sa}\left(k\pi\frac{T}{T_0}\right)\delta(\omega - k\omega_0), \qquad \omega_0 = \frac{2\pi}{T_0}. \tag{3.53}$$

Kształt tego widma zależy od współczynnika wypełnienia T/T_0.

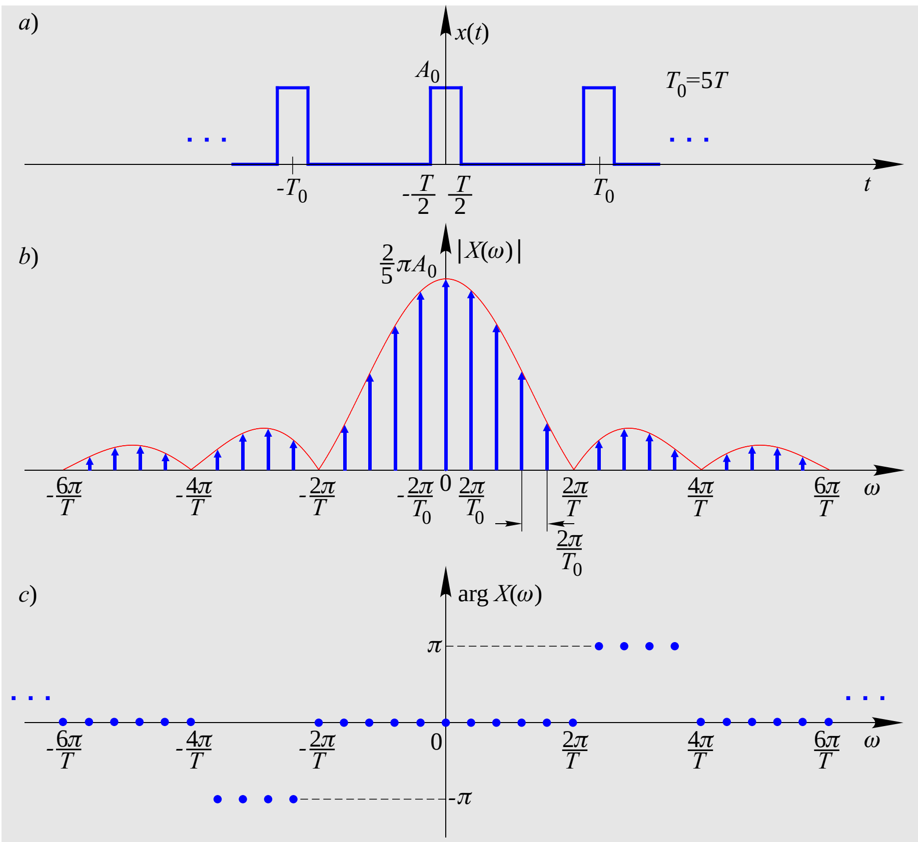
Na rys. 3.18 jest pokazany sygnał unipolarnej fali prostokątnej, jej widmo amplitudowe i widmo fazowe dla przypadku $T/T_0 = 1/5$. Dystrybucje w widmie amplitudowym występują w punktach $\omega = 2\pi k/T_0$, zależnych wyłącznie od okresu fali T_0. Obwiednia widma jest równa:

$$2\pi A_0\frac{T}{T_0}\left|\mathrm{Sa}\,\frac{T\omega}{2}\right|,$$

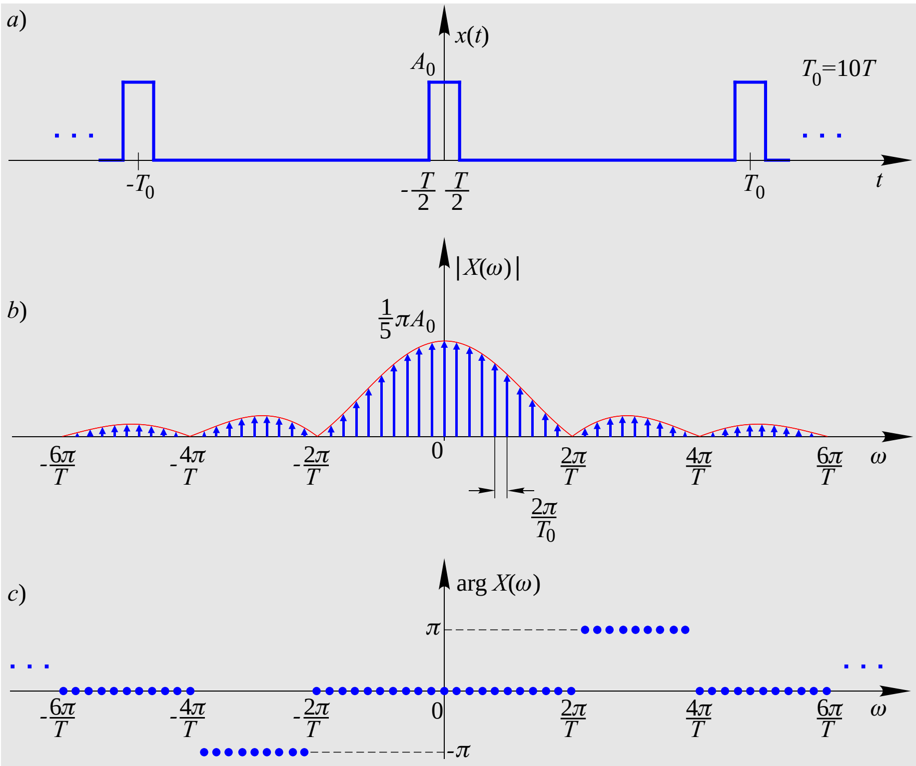
a więc ma kształt modułu funkcji Sa i przyjmuje wartości zerowe w punktach $\omega = 2\pi n/T$, zależnych tylko od szerokości T impulsów fali. Jeżeli współczynnik wypełnienia zmniejszymy do wartości $T/T_0 = 1/10$, zwiększając dwukrotnie okres fali bez zmiany szerokości impulsów, to dystrybucje w widmie amplitudowym zagęszczają się, a ich obwiednia dwukrotnie maleje. Punkty zerowe obwiedni pozostaną bez zmian. Przypadek ten jest zilustrowany na rys. 3.19.

Dalsze zwiększanie okresu T_0 przy stałej szerokości impulsu T powoduje dalsze zagęszczenie dystrybucji w widmie i jednoczesne zmniejszanie się ich obwiedni. W granicy sygnał dąży do impulsu prostokątnego $A_0 \sqcap(t/T)$, dystrybucje wypełniają oś pulsacji w sposób ciągły, a ich obwiednia dąży do zera. Jeśli jednak przed przejściem granicznym wszystkie dystrybucje widmowe pomnożymy przez odwrotność pulsacji ω_0, tj. przez $T_0/2\pi$, a następnie okres T_0 będziemy zwiększać do nieskończoności, to w granicy widmo przejdzie do widma ciągłego pojedynczego impulsu prostokątnego (por. wzór (3.3)):

$$A_0 T \,\mathrm{Sa}\, \frac{\omega T}{2}.$$



Rys. 3.18. Fala prostokątna unipolarna (a), jej widmo amplitudowe (b) i fazowe (c) dla przypadku $T/T_0 = 1/5$



Rys. 3.19. Fala prostokątna unipolarna (a), jej widmo amplitudowe (b) i fazowe (c) dla przypadku $T/T_0 = 1/10$

Dystrybucja grzebieniowa $\delta_{T_0}(t)$

$$\delta_{T_0}(t) \leftrightarrow \omega_0 \delta_{\omega_0}(\omega), \qquad \omega_0 = \frac{2\pi}{T_0}. \tag{3.54}$$

Dystrybucja grzebieniowa w dziedzinie czasu o okresie T_0 ma widmo równe dystrybucji grzebieniowej w dziedzinie częstotliwości o okresie $\omega_0 = 2\pi/T_0$ pomnożonej przez ω_0. Aby wykazać słuszność tej pary, obliczymy współczynniki rozwinięcia dystrybucji $\delta_{T_0}(t)$ w zespolony szereg Fouriera. Uwzględniając, że całkowanie za okres obejmuje tylko jeden środkowy impuls Diraca dystrybucji grzebieniowej oraz korzystając z właściwości próbkowania (1.15) impulsu Diraca, otrzymamy:

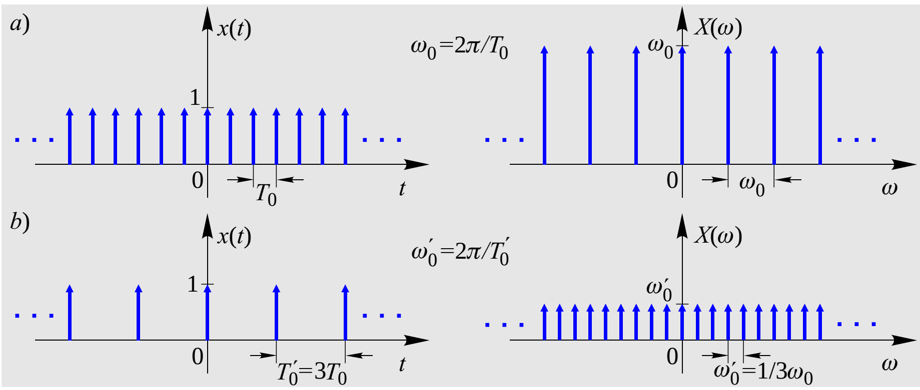
$$X_k = \frac{1}{T_0}\int\limits_{-T_0/2}^{T_0/2} \delta_{T_0}(t)\,\mathrm{e}^{-\mathrm{j}k\omega_0 t}\,\mathrm{d}t = \frac{1}{T_0}\int\limits_{-T_0/2}^{T_0/2} \delta(t)\,\mathrm{e}^{-\mathrm{j}k\omega_0 t}\,\mathrm{d}t =$$

$$= \frac{1}{T_0}\int\limits_{-T_0/2}^{T_0/2} \delta(t)\,\mathrm{d}t = \frac{1}{T_0}.$$

Widzimy więc, że wszystkie współczynniki rozwinięcia w zespolony szereg Fouriera dystrybucji $\delta_{T_0}(t)$ są identyczne i równe $1/T_0$, a więc szereg ten ma postać:

$$\delta_{T_0}(t) = \frac{1}{T_0}\sum_{k=-\infty}^{\infty} \mathrm{e}^{\mathrm{j}k\omega_0 t}, \qquad \omega_0 = \frac{2\pi}{T_0}. \tag{3.55}$$

Po podstawieniu wartości współczynników rozwinięcia dystrybucji $\delta_{T_0}(t)$ w szereg (3.55) do wzoru ogólnego (3.52) otrzymujemy parę (3.54). Para ta została zilustrowana na rys. 3.20 dla dwóch przypadków wartości okresu T_0. Wraz ze zwiększeniem okresu w dziedzinie czasu zmniejsza się okres w dziedzinie częstotliwości i jednocześnie maleje wielkość dystrybucji widmowych. Para ta ma szczególne znaczenie w analizie operacji próbkowania sygnałów (por. p 6.2).



Rys. 3.20. Para transformat (3.54)

3.4.3. Widmo impulsowego sygnału spróbkowanego

Omawiając właściwość próbkowania (1.20) dystrybucji grzebieniowej, wprowadziliśmy dystrybucyjną reprezentację sygnału analogowego $x(t)$ spróbkowanego z okresem T_s w postaci impulsowego sygnału spróbkowanego (rys. 1.27):

$$x_s(t) = x(t)\delta_{T_s}(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s). \tag{3.56}$$

Sygnał ten jest dystrybucyjnym sygnałem analogowym, a więc jest $\mathscr{F}$-transformowalny w sensie granicznym. Ponieważ jest on iloczynem sygnału $x(t)$ i dystrybucji grzebieniowej $\delta_{T_s}(t)$, zatem zgodnie z twierdzeniem 3.9 jego widmo jest splotem widm tych sygnałów:

$$X_s(\omega) = \mathscr{F}[x_s(t)] = \mathscr{F}[x(t)\delta_{T_s}(t)] = \frac{1}{2\pi}\mathscr{F}[x(t)] * \mathscr{F}[\delta_{T_s}(t)]. \tag{3.57}$$

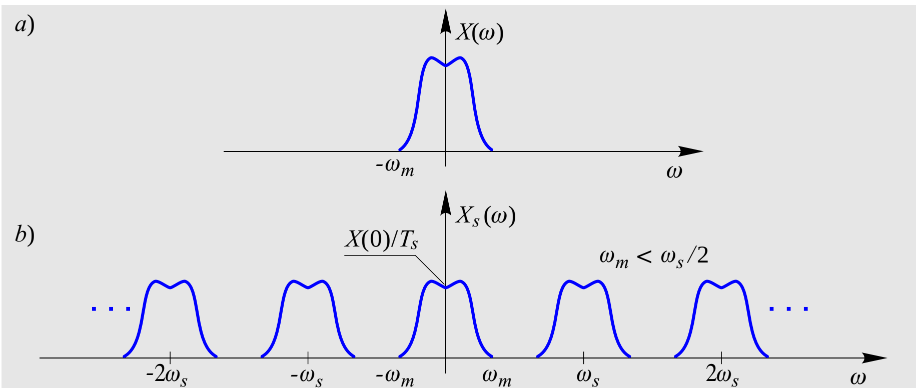
Oznaczając $X(\omega) = \mathscr{F}[x(t)]$ i korzystając z pary (3.54) oraz właściwości (1.21) powielenia okresowego dystrybucji grzebieniowej, otrzymamy:

$$X_s(\omega) = \frac{1}{2\pi}X(\omega) * \omega_s\delta_{\omega_s}(\omega) = \frac{1}{T_s}\sum_{k=-\infty}^{\infty} X(\omega - k\omega_s), \tag{3.58}$$

gdzie $\omega_s = 2\pi/T_s$. Wynika stąd, że widmo impulsowego sygnału spróbkowanego reprezentującego sygnał $x(t)$ o widmie $X(\omega)$ jest powieleniem okresowym z okresem $\omega_s = 2\pi/T_s$ widma $X(\omega)$ podzielonym przez okres próbkowania T_s.

W zagadnieniach próbkowania sygnałów i analizie widmowej sygnałów dyskretnych szczególnie ważny jest przypadek, gdy sygnał $x(t)$ jest sygnałem dolnopasmowym o paśmie ograniczonym częstotliwością f_m i próbkowany z częstotliwością f_s, taką że $f_s \geqslant 2f_m$. Spełniony jest wówczas warunek $\omega_m \leqslant \omega_s/2 = \pi/T_s$, przy którym widmo (3.57) jest (z dokładnością do współczynnika $1/T_s$) ciągiem niezniekształconych kopii widma $X(\omega)$ (rys. 3.21).

Otrzymany rezultat jest jeszcze jednym przejawem dualizmu między dziedziną czasu i dziedziną częstotliwości. Jak pamiętamy, widmo każdego sygnału okresowego jest widmem prążkowym (reprezentowanym odpowiednim ciągiem dystrybucji Diraca w dziedzinie częstotliwości). Ze wzoru (3.7) wynika natomiast, że impulsowy sygnał spróbkowany $x_s(t)$ (reprezentowany odpowiednim ciągiem dystrybucji Diraca w dziedzinie czasu) ma widmo okresowe.



Rys. 3.21. Widmo sygnału dolnopasmowego (a) i widmo reprezentującego go impulsowego sygnału spróbkowanego (b) dla przypadku $\omega_s > 2\omega_m$

3.5. Szereg Fouriera a przekształcenie Fouriera

3.5.1. Widmo sygnału okresowego jako szczególny przypadek transformaty Fouriera w sensie granicznym

Zatrzymajmy się jeszcze przez chwilę na parze (3.52). W większości podręczników poświęconych teorii sygnałów zespolony szereg Fouriera (2.57) sygnałów okresowych jest omawiany przed przekształceniem całkowym Fouriera (3.1). Pojęcie przekształcenia całkowego Fouriera wprowadza się dopiero później, jako uogólnienie zespolonego szeregu Fouriera na sygnały nieokresowe. Uogólnienie to otrzymuje się w wyniku odpowiednio skonstruowanego przejścia granicznego przy okresie T_0 dążącym do nieskończoności. Tym samym pojęcie widma sygnału wprowadza się najpierw dla sygnałów okresowych. Definiuje się je jako zbiór $\{X_k\}$ współczynników jego rozwinięcia w zespolony szereg Fouriera. Widmo amplitudowe sygnału okresowego definiuje się jako zbiór modułów $A_k \triangleq |X_k|$, a widmo fazowe jako zbiór $\varphi_k = \arg X_k$ argumentów tych współczynników.

W ujęciu przedstawionym w poprzednich punktach pojęcie widma zostało zdefiniowane jako ciągła transformata Fouriera rozumiana w sensie zwykłym lub w sensie granicznym. Definicja widma w sensie granicznym obejmuje także sygnały okresowe. Takie podejście umożliwia zatem jednolite ujęcie analizy częstotliwościowej zarówno sygnałów nieokresowych, jak i okresowych w kategoriach formalizmu matematycznego całkowego przekształcenia Fouriera. Różnica między ujęciem opartym na szeregu Fouriera, a ujęciem w kategoriach całkowego przekształcenia Fouriera polega na odmiennym określeniu widma sygnału okresowego. W prezentowanym podejściu widmo sygnału okresowego jest określone nie jako zbiór liczb, lecz – zgodnie ze wzorem (3.52) – jako wielkość dystrybucyjna. Różnica ta jest natury jedynie formalnej, gdyż dla obu tych reprezentacji widmowych pełna informacja o sygnale okresowym jest zawarta z zbiorze współczynników $\{X_k\}$.

3.5.2. Związek między widmem sygnału impulsowego a współczynnikami Fouriera jego przedłużenia okresowego

Niech $x(t)$ będzie sygnałem impulsowym określonym w przedziale $(-T_0/2, T_0/2)$ (lub w dowolnym innym przedziale o długości T_0) oraz $X(\omega)$ –

jego widmem. Rozważmy sygnał $x_{T_0}(t)$, będący przedłużeniem okresowym tego sygnału z okresem T_0. Zgodnie z właściwością (1.21) powielenia okresowego dystrybucji grzebieniowej, sygnał $x_{T_0}(t)$ możemy zapisać w postaci:

$$x_{T_0}(t) = \sum_{n=-\infty}^{\infty} x(t - nT_0) = x(t) * \delta_{T_0}(t).$$

Stosując twierdzenie o splocie w dziedzinie czasu i korzystając z pary (3.54), widmo tego sygnału możemy zapisać jako:

$$X_{T_0}(\omega) = X(\omega)\omega_0\delta_{\omega_0}(\omega), \qquad \omega_0 = \frac{2\pi}{T_0}.$$

Korzystając z kolei z właściwości (1.20) próbkowania dystrybucji grzebieniowej, otrzymamy:

$$X_{T_0}(\omega) = \frac{2\pi}{T_0} \sum_{k=-\infty}^{\infty} X(k\omega_0)\delta(\omega - k\omega_0) = 2\pi \sum_{k=-\infty}^{\infty} \frac{X(k\omega_0)}{T_0}\delta(\omega - k\omega_0).$$

Z porównania tego wzoru z parą (3.52) wynika, że

$$X_k = \frac{X(k\omega_0)}{T_0}, \tag{3.59}$$

tzn. k-ty współczynnik rozwinięcia powielenia okresowego $x_{T_0}(t)$ sygnału impulsowego $x(t)$ w zespolony szereg Fouriera (2.57) jest równy wartości widma sygnału $x(t)$ w punkcie $k\omega_0$ podzielonej przez okres T_0.

Po podstawieniu wzoru (3.59) do wzoru (2.58) otrzymujemy:

$$x_{T_0}(t) = \frac{1}{T_0} \sum_{k=-\infty}^{\infty} X(k\omega_0)\, \mathrm{e}^{\mathrm{j}k\omega_0 t}. \tag{3.60}$$

Rezultat ten określa podstawowy związek między transformatą Fouriera sygnału impulsowego a zespolonym szeregiem Fouriera jego przedłużenia okresowego.

Przykład 3.5. Korzystając ze wzoru (3.59), możemy wyznaczyć widmo powielenia okresowego $x_{T_0}(t)$ sygnału impulsowego $x(t)$ bezpośrednio na podstawie widma sygnału impulsowego. Na przykład, stosując wzór (3.59) w przypadku impulsu prostokątnego o widmie określonym wzorem (3.3), otrzymujemy wprost widmo (3.53) unipolarnej fali prostokątnej, bez konieczności obliczania współczynników X_k na podstawie definicji.

Przykład 3.6. Korzystając ze wzoru (3.59), wykażemy inną drogą słuszność wzoru (3.55). Ponieważ widmo pojedynczego impulsu Diraca jest stałe i równe 1 dla każdej wartości ω (por. parę (3.44)), zatem $X(k\omega_0) = 1$ dla każdego k. Uwzględniając ten wynik we wzorze (3.60), otrzymujemy bezpośrednio wzór (3.55).

3.6. Zasada nieoznaczoności w teorii sygnałów

3.6.1. Związek między czasem trwania sygnału a szerokością jego widma

W komentarzu do twierdzenia 3.3 o zmianie skali stwierdziliśmy, że im szybciej sygnał zmienia się w czasie. tym szersze jest jego widmo. Relację tę można ująć w formie zależności liczbowej, wprowadzając odpowiednie miary czasu trwania sygnału i szerokości widma.

Definicja 3.4. *Równoważnym czasem trwania sygnału* $x(t)$ nazywamy wielkość:

$$\Delta t_x = \frac{\int\limits_{-\infty}^{\infty} x(t)\,\mathrm{d}t}{x(0)}. \tag{3.61}$$

Definicja 3.5. *Szerokością równoważną widma* $X(\omega)$ nazywamy wielkość:

$$\Delta \omega_x = \frac{\int\limits_{-\infty}^{\infty} X(\omega)\,\mathrm{d}\omega}{X(0)}. \tag{3.62}$$

Jeżeli w całce (3.1) podstawimy $\omega = 0$, a w całce (3.4) $t = 0$, otrzymamy:

$$\Delta t_x = \frac{X(0)}{\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} X(\omega)\,\mathrm{d}\omega}, \tag{3.63}$$

skąd wynika, że iloczyn równoważnego czasu trwania sygnału i równoważnej szerokości jego widma jest wielkością stałą:

$$\Delta t_x \Delta \omega_x = 2\pi. \tag{3.64}$$

Jeśli jedna z tych miar wzrasta, druga proporcjonalnie maleje.

3.6.2. Zasada nieoznaczoności

Wzory (3.61) i (3.62) mają ograniczony zakres stosowalności. Miara (3.61) nie jest odpowiednią miarą czasu trwania sygnału wtedy, kiedy całka sygnału jest równa zeru, jak również wtedy, kiedy wartość sygnału w chwili $t = 0$ jest równa bądź bliska zeru. Istotną wadą tej miary jest także jej zależność od położenia sygnału na osi czasu. Podobne wady ma szerokość równoważna widma (3.62).

Z tych względów wprowadzane są również inne miary czasu trwania sygnału i szerokości jego widma, z których najbardziej uniwersalne i pozbawione wymienionych wad są tzw. *szerokości średniokwadratowe* sygnału i odpowiednio widma. Charakteryzują one stopień koncentracji energii sygnału w dziedzinie czasu lub w dziedzinie częstotliwości (por.[3], p. 6.11).

Jednak bez względu na to, jak byśmy nie zdefiniowali miary czasu trwania i szerokości widma sygnału, zależność (3.64) wyraża pewną ogólną zasadę obowiązującą w teorii sygnałów i nazywaną *zasadą nieoznaczoności*. Orzeka ona, że iloczyn miary czasu trwania sygnału i odpowiadającej jej miary szerokości jego widma jest ograniczony od dołu i że nie można tych miar zmniejszać jednocześnie. W przypadku zmniejszania czasu trwania sygnału trzeba się liczyć z tym, że jego widmo będzie się odpowiednio poszerzać. I odwrotnie, chcąc wygenerować sygnał o wąskim widmie, należy uwzględnić, że musi on trwać w czasie dostatecznie długo.

Zasada nieoznaczoności stanowi jeszcze jedną relację matematyczną między dziedziną czasu i dziedziną częstotliwości. Znajduje ona swoje odbicie w wielu zagadnieniach technicznych. Na przykład, w projektowaniu cyfrowych systemów telekomunikacyjnych dąży się do przesyłania impulsów w jak najkrótszym czasie, a zarazem w jak najwęższym paśmie częstotliwości. Krótki czas trwania impulsów umożliwia zwiększenie szybkości transmisji informacji, natomiast wąskie pasmo ich przesyłania – zwiększenie pojemności kanału transmisyjnego. W świetle zasady nieoznaczoności wymagania te są sprzeczne. W praktyce wybór czasu trwania impulsu i szerokości jego pasma dokonuje się na drodze kompromisu, którego konieczność jest konsekwencją zasady nieoznaczoności.

Słownik

metody częstotliwościowe
metody analizy sygnałów w dziedzinie częstotliwości

para transformat Fouriera
para funkcji, którą tworzy sygnał i jego widmo

pasmo sygnału
przedział częstotliwości, w którym widmo sygnału przybiera wartości niezerowe

przekształcenie Fouriera
przekształcenie całkowe przyporządkowujące sygnałowi jego widmo

pulsacja graniczna
największa pulsacja, dla której widmo sygnału przybiera wartość różną od zera

splot w dziedzinie częstotliwości
operacja na parze widm przyporządkowująca im inne widmo

sygnał $\mathscr{F}$-transformowalny
sygnał, dla którego istnieje transformata Fouriera

sygnał dolnopasmowy
sygnał, którego zasadnicza część widma jest skupiona w zakresie małych częstotliwości

sygnał wąskopasmowy
sygnał, którego widmo jest skupione w wąskim przedziale częstotliwości $[f_d, f_g]$, $f_d \neq 0$, gdzie $[f_d$ jest częstotliwością dolną, a $[f_g$ – częstotliwością górną pasma sygnału

trygonometryczny szereg Fouriera
przedstawienie sygnału okresowego jako sumy ważonej (w szczególnych przypadkach skończonej) sygnałów harmonicznych o częstotliwościach będących wielokrotnościami częstotliwości podstawowej równej odwrotności okresu sygnału

widmo sygnału
funkcja zespolona zmiennej rzeczywistej (częstotliwości) opisująca strukturę częstotliwościową sygnału; transformata Fouriera sygnału

widmo amplitudowe
moduł widma sygnału

widmo białe
widmo o stałej gęstości widmowej w całym zakresie częstotliwości

widmo energii
funkcja opisująca rozkład energii sygnału o ograniczonej energii w funkcji częstotliwości; kwadrat widma amplitudowego sygnału

widmo fazowe
argument widma sygnału

zasada nieoznaczoności
zasada orzekająca, że im zmiany sygnału w czasie następują szybciej, tym pasmo tego sygnału jest węższe i odwrotnie

Literatura

[1] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom III. WNT, Warszawa, 1995.

[2] Żakowski W., Leksiński W.: Matematyka, część IV. WNT, Warszawa, wyd. 7, 1984.

[3] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

Lekcja 4

Analiza częstotliwościowa dyskretnych sygnałów deterministycznych

W lekcji 4 omówimy metody analizy częstotliwościowej sygnałów dyskretnych. Wprowadzimy pojęcie przekształcenia Fouriera sygnałów dyskretnych i przedyskutujemy jego podstawowe właściwości, w tym odnoszące się do niego twierdzenia. Podamy definicję widma sygnału dyskretnego i podkreślimy jego najważniejszą cechę — okresowość. Zdefiniujemy także pojęcie dyskretnego szeregu Fouriera. Jako szczególną klasę sygnałów dyskretnych omówimy sygnały N-okresowe i ich widma.

W lekcji przedyskutujemy ponadto zagadnienie numerycznego obliczania widm sygnałów i wprowadzimy ważne pojęcie dyskretnego przekształcenie Fouriera (DPF) jako podstawowego narzędzia matematycznego numerycznej analizy częstotliwościowej sygnałów. Dyskretne przekształcenie Fouriera omówimy najpierw dla przypadku sygnałów impulsowych, a następnie N-okresowych. W tym drugim przypadku wskażemy na jego związek z dyskretnym szeregiem Fouriera. Przedyskutujemy także podstawowe właściwości DPF. Na zakończenie lekcji rozpatrzymy zagadnienie odtwarzania sygnału dyskretnego o nieskończonym czasie trwania na podstawie próbek jego widma.

4.1. Przekształcenie Fouriera sygnałów dyskretnych

Wartości sygnałów dyskretnych zmieniają się w czasie od próbki do próbki w sposób skokowy. Podobnie jak w przypadku sygnałów analogowych, ewolucję sygnału dyskretnego w czasie można rozpatrywać jako efekt superpozycji nieprzeliczalnej liczby elementarnych dyskretnych sygnałów harmonicznych. Badanie struktury tych sygnałów i analiza ich udziału w tworzeniu danego sygnału dyskretnego są dokonywane w dziedzinie częstotliwości.

U podstaw matematycznych analizy częstotliwościowej sygnałów dyskretnych leżą pojęcia przekształceń Fouriera – prostego i odwrotnego. Z uwagi na dyskretny charakter dziedziny sygnałów pojęcia te są definiowanie odmiennie, niż w przypadku sygnałów analogowych.

4.1.1. Proste przekształcenie Fouriera. Widmo sygnału dyskretnego

Rozważmy sygnał dyskretny $x[nT_s]$ otrzymany np. w wyniku okresowego próbkowania w chwilach $t_n = nT_s$ pewnego rzeczywistego lub zespolonego sygnału analogowego $x(t)$ (celowo przyjmujemy tu na razie nieunormowaną skalę czasu). Założymy, że $x[nT_s]$ jest sygnałem o ograniczonej energii:

$$E_x = \sum_{n=-\infty}^{\infty} |x(nT_s)|^2 < \infty. \tag{4.1}$$

Jak pamiętamy (por. p. 2.2.1), sygnały takie należą do przestrzeni Hilberta l^2 sygnałów sumowalnych z kwadratem, w której iloczyn skalarny jest określony wzorem (2.19).

Definicja 4.1. *Proste przekształcenie Fouriera* (*$\mathscr{F}$-przekształcenie*) sygnału dyskretnego $x[nT_s]$ jest zdefiniowane wzorem:

$$X(\mathrm{e}^{\mathrm{j}\omega T_s}) \triangleq \mathscr{F}[x(nT_s)] = \sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{e}^{-\mathrm{j}n\omega T_s}. \tag{4.2}$$

Funkcję $X(\mathrm{e}^{\mathrm{j}\omega T_s})$ nazywamy *$\mathscr{F}$-transformatą* lub *widmem* sygnału dyskretnego $x[nT_s]$.

Wzór (4.2) definiuje przekształcenie Fouriera sygnału dyskretnego w sensie zwykłym. W wyniku tego przekształcenia sygnałowi dyskretnemu $x[nT_s]$ o ograniczonej energii zostaje przyporządkowana funkcja zespolona $X(\mathrm{e}^{\mathrm{j}\omega T_s})$ zmiennej rzeczywistej ω. Z uzasadnionych względów, w teorii sygnałów dyskretnych argument tej funkcji oznacza się zwyczajowo $\mathrm{e}^{\mathrm{j}\omega T_s}$, zamiast naturalnego oznaczenia ω.

Funkcję $|X(\mathrm{e}^{\mathrm{j}\omega T_s})| \triangleq A(\mathrm{e}^{\mathrm{j}\omega T_s})$ nazywamy *widmem amplitudowym*, a funkcję $\arg X(\mathrm{e}^{\mathrm{j}\omega T_s}) \triangleq \varphi(\mathrm{e}^{\mathrm{j}\omega T_s})$ – *widmem fazowym* sygnału $x[nT_s]$.

4.1.2. Okresowość widma sygnału dyskretnego

Można łatwo sprawdzić, że funkcja $X(\mathrm{e}^{\mathrm{j}\omega T_s})$ jest funkcją okresową zmiennej ω o okresie $\omega_s = 2\pi/T_s$. W tym celu wystarczy we wzorze (4.2) uwzględnić tożsamość:

$$\mathrm{e}^{-\mathrm{j}n(\omega+\omega_s)T_s} = \mathrm{e}^{-\mathrm{j}n\left(\omega+\frac{2\pi}{T_s}\right)T_s} = \mathrm{e}^{-\mathrm{j}n\omega T_s}\,\mathrm{e}^{-\mathrm{j}2\pi n} = \mathrm{e}^{-\mathrm{j}n\omega T_s}\,.$$

Wynika stąd ważny wniosek: widma sygnałów dyskretnych są okresowe. Jeżeli widmo rozpatrujemy jako funkcję pulsacji ω, jego okres jest równy pulsacji próbkowania ω_s, jeśli natomiast rozpatrujemy je jako funkcję częstotliwości f, okres jest równy częstotliwości próbkowania f_s. Okresowość widm jest podstawową cechą sygnałów dyskretnych.

Ponieważ funkcja $X(\mathrm{e}^{\mathrm{j}\omega T_s})$ jest okresowa, może być rozwinięta w zespolony szereg Fouriera (w tym wypadku w dziedzinie częstotliwości, tj. względem zmiennej ω). Z porównania wzorów (4.2) i (2.57) (przy podstawieniu $t \to \omega$, $\omega_0 \to T_s$, $X_k \to x(nT_s)$) wynika, że szeregiem tym jest właśnie suma występująca po prawej stronie wzoru (4.2). Podkreślmy, że współczynnikami Fouriera tego szeregu są próbki $x(nT_s)$ sygnału.

4.1.3. Porównanie z widmem impulsowego sygnału spróbkowanego

Powróćmy na chwilę do wzoru (3.56) opisującego dystrybucyjną reprezentację $x_s(t)$ sygnału analogowego $x(t)$ spróbkowanego z okresem T_s. W punkcie 3.4.3 wykazaliśmy, że widmo $X_s(\omega)$ analogowego sygnału dystrybucyjnego (3.56) jest okresowe (por. wzór (3.58)). Nie podaliśmy natomiast jego jawnej postaci. Pokażemy obecnie, że widmo to można wyrazić przez próbki $x(nT_s)$ sygnału $x(t)$.

Widmo $X_s(\omega)$, jako funkcja okresowa zmiennej ω, może być rozwinięte w zespolony szereg Fouriera w dziedzinie częstotliwości. Szereg ten możemy otrzymać, wychodząc z definicji (3.1) widma sygnału analogowego i korzystając przy tym ze wzoru (3.56) oraz właściwości filtracji impulsu Diraca (1.16):

$$X_s(\omega) = \int_{-\infty}^{\infty} x_s(t)\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} x(nT_s)\delta(t-nT_s)\right] \mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t =$$
$$= \sum_{n=-\infty}^{\infty} x(nT_s) \int_{-\infty}^{\infty} \delta(t-nT_s)\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{e}^{-\mathrm{j}\omega nT_s}. \quad (4.3)$$

Jak widzimy, przyjęta definicja (4.2) widma sygnału dyskretnego jest w pełni zgodna z postacią widma, która wynika z analogowego opisu sygnału spróbkowanego sygnałem dystrybucyjnym (3.56). Niezależnie więc od tego, czy sygnał spróbkowany będziemy traktować jako dystrybucyjny sygnał analogowy (3.56) (ciąg impulsów Diraca) i pozostaniemy w obrębie metod analizy częstotliwościowej sygnałów analogowych, czy też jako sygnał dyskretny (ciąg próbek) i analizę częstotliwościową sygnałów dyskretnych oprzemy na niezależnej definicji (4.2), otrzymane rezultaty będą identyczne.

4.1.4. Widmo jako funkcja pulsacji unormowanej

Jak już wspominaliśmy (por. p. 1.3), w analizie sygnałów dyskretnych normuje się zarówno czas, jak i pulsację. Widmo sygnału dyskretnego $x[n]$ rozpatruje się zwyczajowo nie jako funkcję zmiennej ω – pulsacji, tak jak we wzorze (4.2), lecz jako funkcję zmiennej $\theta = \omega T_s$, tj. pulsacji unormowanej względem okresu próbkowania T_s. Widmo to definiuje się zatem jako funkcję:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{n=-\infty}^{\infty} x(n)\,\mathrm{e}^{-\mathrm{j}n\theta}, \quad (4.4)$$

do której jest zbieżny szereg Fouriera określony prawą stroną tej równości. Widmo sygnału dyskretnego wyrażone w funkcji unormowanej pulsacji θ jest funkcją okresową zmiennej θ o okresie 2π. Widmo amplitudowe $|X(\mathrm{e}^{\mathrm{j}\theta})| \triangleq A(\mathrm{e}^{\mathrm{j}\theta})$ i widmo fazowe $\arg X(\mathrm{e}^{\mathrm{j}\theta}) \triangleq \varphi(\mathrm{e}^{\mathrm{j}\theta})$ są zatem również funkcjami okresowymi o okresie 2π. Wykresy tych widm wystarczy zatem sporządzać w przedziale $[-\pi, \pi]$. Jeśli sygnał dyskretny jest rzeczywisty, jego widmo amplitudowe jest funkcją okresową parzystą, a widmo fazowe – funkcją okresową nieparzystą zmiennej θ. W takim przypadku wykresy tych widm wystarczy sporządzać w przedziale $[0, \pi]$.

W dalszym ciągu widmo sygnału dyskretnego będziemy definiować wzorem (4.4). Należy jednak podkreślić, że w zagadnieniach praktycznych jest ono niekiedy rozpatrywane jako funkcja unormowanej częstotliwości $\nu = \theta/2\pi$. Jego okres jest wówczas równy 1.

Komentarz. Z formalnego punktu widzenia $\mathscr{F}$-przekształcenie sygnałów dyskretnych określone wzorem (4.4) możemy traktować jako odwzorowanie przestrzeni Hilberta l^2

ciągów sumowalnych z kwadratem w dziedzinie czasu w przestrzeń $L^2_{2\pi}$ ciągłych funkcji okresowych zmiennej θ o okresie 2π w dziedzinie częstotliwości.

Przykład 4.1. Wyznaczymy widmo impulsu Kroneckera $\delta[n]$ (por. rys. 1.29). Zgodnie ze wzorem (4.4):

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{n=-\infty}^{\infty} \delta[n]\,\mathrm{e}^{-\mathrm{j}n\theta} = 1 \cdot \mathrm{e}^{\mathrm{j}0} = 1, \tag{4.5}$$

a więc jest to widmo stałe w przedziale $-\pi \leqslant \theta \leqslant \pi$ (a zarazem na całej osi θ).

Przykład 4.2. Wyznaczymy widmo dyskretnego impulsu prostokątnego (por. rys. 1.30):

$$x[n] = \begin{cases} 1 & \text{dla} \quad n \leqslant |N|, \\ 0 & \text{dla} \quad n > |N|. \end{cases} \tag{4.6}$$

Przytoczymy w miarę pełne obliczenia, są one bowiem typowe dla analitycznego wyznaczania widm sygnałów dyskretnych. Podstawiając wartości tego sygnału do wzoru (4.4), otrzymujemy:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{n=-N}^{N} \mathrm{e}^{-\mathrm{j}n\theta}\,.$$

Korzystając ze wzoru na sumę częściową szeregu geometrycznego:

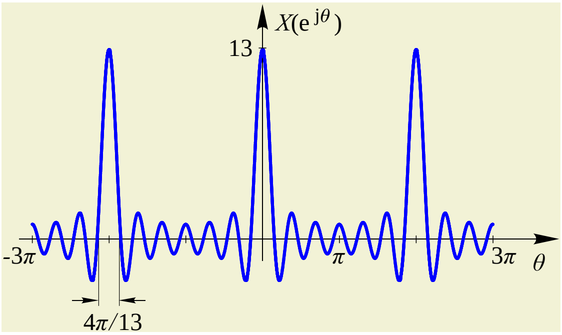
$$\sum_{i=1}^{n} a_1 q^{i-1} = a_1 \frac{1-q^n}{1-q},$$

gdzie $n = 2N+1$, $a_1 = \mathrm{e}^{\mathrm{j}N\theta}$ i $q = \mathrm{e}^{-\mathrm{j}\theta}$, otrzymujemy:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \mathrm{e}^{\mathrm{j}N\theta}\,\frac{1-\mathrm{e}^{-\mathrm{j}(2N+1)\theta}}{1-\mathrm{e}^{-\mathrm{j}\theta}} =$$

$$= \mathrm{e}^{\mathrm{j}N\theta}\,\frac{\mathrm{e}^{-\mathrm{j}(2N+1)\theta/2}}{\mathrm{e}^{-\mathrm{j}\theta/2}}\,\frac{\mathrm{e}^{\mathrm{j}(2N+1)\theta/2}-\mathrm{e}^{-\mathrm{j}(2N+1)\theta/2}}{\mathrm{e}^{\mathrm{j}\theta/2}-\mathrm{e}^{-\mathrm{j}\theta/2}} = \frac{\sin(2N+1)\theta/2}{\sin\theta/2}. \tag{4.7}$$

Ponieważ sygnał jest rzeczywisty i parzysty, widmo (4.7) jest również rzeczywiste i parzyste. W punkcie $\theta = 0$ osiąga ono wartość $2N+1$ równą liczbie próbek impulsu (można ją wyznaczyć stosując regułę l'Hospitala), a w punktach $\theta = 2\pi m/(2N+1)$ przyjmuje wartości zerowe. Na rys. 4.1 pokazano widmo impulsu (4.6) wykreślone w przedziale $-3\pi \leqslant \theta \leqslant 3\pi$ dla $N = 6$. Ma ono charakterystyczną strukturę listkową. Wysokość listka głównego, równa $2N+1$, rośnie proporcjonalnie do liczby próbek, zaś jego szerokość wynosi $4\pi/(2N+1)$ i maleje proporcjonalnie do liczby próbek. Szerokości listków bocznych wynoszą $2\pi/(2N+1)$, a wartości bezwzględne ich ekstremów maleją w przedziale $[-\pi, \pi]$ w miarę oddalania się od punktu $\theta = 0$ w przybliżeniu jak funkcja $1/|\sin(\theta/2)|$.



Rys. 4.1. Widmo dyskretnego impulsu prostokątnego (4.6) dla $N = 6$

Przykład 4.3. W ostatnim przykładzie wyznaczymy widmo dyskretnego sygnału wykładniczego $x[n] = a^n \boldsymbol{1}[n], |a| < 1$ (por. rys. 1.32). Korzystając ze wzoru na sumę nieskończonego szeregu geometrycznego $\sum_0^\infty a_1 q^n = a_1/(1-q)$, otrzymujemy:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{n=0}^{\infty} a^n \, \mathrm{e}^{-\mathrm{j}n\theta} = \sum_{n=0}^{\infty} (a \, \mathrm{e}^{-\mathrm{j}\theta})^n = \frac{1}{1 - a \, \mathrm{e}^{-\mathrm{j}\theta}}. \tag{4.8}$$

W tym przypadku widmo jest zespolone. Widmo amplitudowe i widmo fazowe są opisane wzorami:

$$A(\mathrm{e}^{\mathrm{j}\theta}) = \left| \frac{1}{1 - (a\cos\theta - \mathrm{j}\, a \sin\theta)} \right| =$$

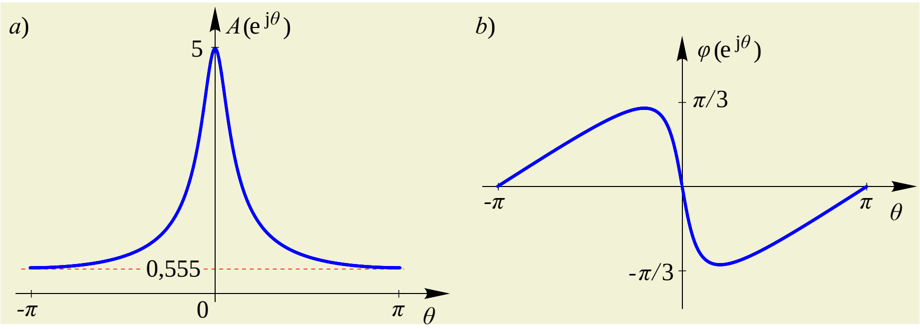
$$= \frac{1}{\sqrt{(1 - a\cos\theta)^2 + (asin\theta)^2}} = \frac{1}{\sqrt{1 + a^2 - 2a\cos\theta}}, \tag{4.9a}$$

$$\varphi(\mathrm{e}^{\mathrm{j}\theta}) = -\operatorname{arctg} \frac{a \sin\theta}{1 - a\cos\theta}. \tag{4.9b}$$

Wykresy tych widm w przedziale $[-\pi, \pi]$ dla przypadku $a = 0,8$ pokazano na rys. 4.2.

4.1.5. Odwrotne przekształcenie Fouriera

Współczynnikami zespolonego szeregu Fouriera (4.4) definiującego widmo $X(\mathrm{e}^{\mathrm{j}\theta})$ sygnału dyskretnego $x[n]$ są próbki $x(n)$. Zgodnie z ogólnym wzorem określającym współczynniki zespolonego szeregu (por. wzór (2.58)), próbki te można wyznaczyć na podstawie widma $X(\mathrm{e}^{\mathrm{j}\theta})$.



Rys. 4.2. Widmo amplitudowe (a) i widmo fazowe (b) dyskretnego sygnału wykładniczego

Definicja 4.2. *Odwrotne przekształcenie Fouriera* dla przypadku sygnałów dyskretnych jest zdefiniowane całką

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\theta})\, e^{jn\theta}\, d\theta, \qquad n = 0, \pm 1, \ldots \tag{4.10}$$

przyporządkowującą funkcji okresowej $X(e^{j\theta})$ o okresie 2π sygnał $x[n]$.

Całka (4.10) umożliwia wyznaczenie sygnału dyskretnego na podstawie jego widma. Określa ona zatem odwzorowanie $x[n] = \mathscr{F}^{-1}[X(e^{j\theta})]$, którego wynik $x[n]$ nazywamy *odwrotną transformatą Fouriera* ($\mathscr{F}^{-1}$*-transformatą*) dla przypadku sygnałów dyskretnych. Interpretacja całki (4.10) jest analogiczna jak całki (3.4). Dyskretny sygnał $x[n]$ o ograniczonej energii może być traktowany jako superpozycja nieprzeliczalnej liczby elementarnych dyskretnych zespolonych sygnałów harmonicznych $e^{jn\theta}$, o pulsacjach unormowanych należących do ciągłego przedziału $\theta \in [-\pi, \pi]$, których udział w tworzeniu sygnału $x[n]$ jest określony gęstością ich zespolonej amplitudy, tj. widmem $X(e^{j\theta})$.

4.1.6. Przekształcenie Fouriera w sensie granicznym

Sygnały dyskretne o ograniczonej mocy nie mają widm w sensie definicji (4.4), szereg występujący po prawej stronie równości (4.4) jest bowiem dla tych sygnałów rozbieżny. Podobnie jak w przypadku sygnałów analogowych o ograniczonej mocy, dla sygnałów dyskretnych o ograniczonej mocy można wprowadzić pojęcie widma w sensie granicznym. W tym celu sygnał $x[n]$ o ograniczonej mocy aproksymuje się ciągiem sygnałów o ograniczonej energii i konstruuje się odpowiednie przejście graniczne.

Konstrukcję graniczną można przeprowadzić w różny sposób. Podamy przykład takiej konstrukcji, umożliwiającej wyznaczenie widma w sensie granicznym dyskretnego sygnału stałego $x[n] = 1$ dla $n = 0, \pm 1, \ldots$

Przykład 4.4. Dyskretny sygnał stały można uważać za granicę ciągu dyskretnych impulsów prostokątnych (4.6), gdy $N \to \infty$. Widmo dyskretnego impulsu prostokątnego (4.6) opisuje wzór (4.7). Przeanalizujemy jak zmienia się to widmo, gdy $N \to \infty$. Wysokość listka głównego, równa $2N+1$, rośnie wówczas do nieskończoności, a jego szerokość $4\pi/(2N+1)$ maleje do zera. Listki boczne zagęszczają się, a ich szerokości $2\pi/(2N+1)$ również maleją do zera. W efekcie widmo dąży do ciągu dystrybucji Diraca powtarzanych okresowo co 2π. Można pokazać, że całka widma (4.7) w przedziale $[-\pi, \pi]$ jest równa 2π, zatem w granicy otrzymujemy:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \mathscr{F}[1] = 2\pi\delta_{2\pi}(\theta). \tag{4.11}$$

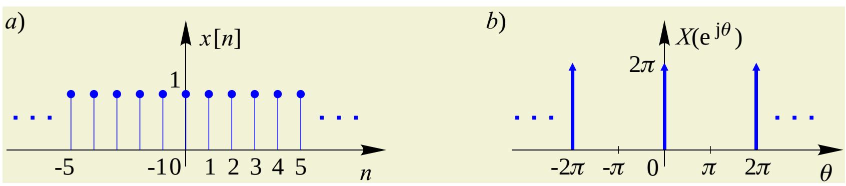
Rezultat ten potwierdzimy, obliczając widmo dyskretnego sygnału stałego wprost z definicji (4.4). Podstawiając wartości tego sygnału do wzoru (4.4), otrzymujemy:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{n=-\infty}^{\infty} x[n]\,\mathrm{e}^{-\mathrm{j}n\theta} = \sum_{n=-\infty}^{\infty} \mathrm{e}^{-\mathrm{j}n\theta} = 2\pi\left(\frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \mathrm{e}^{\mathrm{j}n\theta}\right).$$

Porównajmy wyrażenie ujęte w nawias w ostatniej równości z wyrażeniem $\delta_{T_0}(t) = 1/T_0 \sum_{k=-\infty}^{\infty} \mathrm{e}^{\mathrm{j}k\omega_0 t}$ opisującym rozwinięcie dystrybucji grzebieniowej $\delta_{T_0}(t)$ o okresie T_0 w zespolony szereg Fouriera w dziedzinie czasu (por. wzór (3.55)). Z porównania wynika, że wyrażenie w nawiasie jest rozwinięciem w zespolony szereg Fouriera w dziedzinie częstotliwości (względem zmiennej θ) dystrybucji grzebieniowej $\delta_{2\pi}(\theta)$ o okresie 2π:

$$\delta_{2\pi}(\theta) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \mathrm{e}^{\mathrm{j}n\theta}. \tag{4.12}$$

Stąd wynika już słuszność wzoru (4.11). Para transformat $1 \leftrightarrow 2\pi\delta_{2\pi}(\theta)$ jest przedstawiona na rys. 4.3.



Rys. 4.3. Dyskretny sygnał stały (a) i jego widmo (b)

Innym sposobem wyznaczania widm w sensie granicznym dyskretnych sygnałów o ograniczonej mocy jest ich aproksymacja ciągami typu $r^{-n}x[n]$ (w przypadku sygnałów określonych dla $n = 0, 1, \ldots$) lub $r^{-|n|}x[n]$ (w przypadku sygnałów określonych dla $n = 0, \pm 1, \ldots$), gdzie parametr $r > 1$ jest tak dobrany, aby elementy tych ciągów były sygnałami o ograniczonej energii, tzn. były

$\mathscr{F}$-transformowalne w sensie definicji (4.4). Jeżeli ciąg zwykłych widm tych sygnałów oznaczymy $X_r(\mathrm{e}^{\mathrm{j}\theta})$, to za definicję widma sygnału $x[n]$ w sensie granicznym możemy przyjąć granicę prawostronną $X(\mathrm{e}^{\mathrm{j}\theta}) = \lim_{r\to 1+} X_r(\mathrm{e}^{\mathrm{j}\theta})$. Naszkicowany sposób postępowania nosi nazwę *metody mnożników potęgowych* lub *metody Abela* ([1], p. 2.3.2-A).

Przeprowadzenie odpowiedniego przejścia granicznego w przypadku sygnałów dyskretnych o ograniczonej mocy napotyka często poważne trudności analityczne. Z tego względu nie będziemy dalej dokładnie omawiać tego zagadnienia. Ma ono zresztą ograniczone znaczenie dla zagadnień praktycznych, gdyż w rzeczywistości sygnał możemy obserwować w skończonym przedziale czasu i pobrać w tym przedziale jedynie skończoną liczbę próbek. Wyjątkiem jest ważna podklasa okresowych sygnałów dyskretnych. Widmo sygnału okresowego może być definiowane w sensie granicznym i traktowane jako wielkość dystrybucyjna, jednak zwykle jest ono definiowane w sposób odmienny. Przypadek okresowych sygnałów dyskretnych omówimy dokładnie w punkcie 4.3.

4.2. Twierdzenia

Podamy obecnie najważniejsze twierdzenia dotyczące przekształcenia Fouriera sygnałów dyskretnych. Są one odpowiednikami twierdzeń omówionych w p. 3.3 dla przypadku sygnałów analogowych. Przyjmiemy założenie, że $x[n] \leftrightarrow X(\mathrm{e}^{\mathrm{j}\theta})$ oraz $y[n] \leftrightarrow Y(\mathrm{e}^{\mathrm{j}\theta})$ są parami transformat Fouriera, przy czym sygnały $x[n]$ i $y[n]$ są w ogólnym przypadku zespolone.

Twierdzenie 4.1 (*o liniowości*)

$$ax[n] + by[n] \leftrightarrow aX(\mathrm{e}^{\mathrm{j}\theta}) + bY(\mathrm{e}^{\mathrm{j}\theta}). \tag{4.13}$$

Właściwość liniowości wynika bezpośrednio z definicji (4.4).

Twierdzenie 4.2 (*o przesunięciu w dziedzinie czasu*)

$$x[n-n_0] \leftrightarrow X(\mathrm{e}^{\mathrm{j}\theta})\,\mathrm{e}^{-\mathrm{j}\theta n_0}\,. \tag{4.14}$$

W wyniku opóźnienia ($n_0 > 0$) lub przyspieszenia sygnału ($n_0 < 0$) widmo amplitudowe nie ulega zmianie, a widmo fazowe zmienia się o składnik $-\theta n_0$.

Przykład 4.5. Twierdzenie o przesunięciu w dziedzinie czasu zilustrujemy na przykładzie impulsu prostokątnego:

$$x[n] = \begin{cases} 1 & \text{dla} \quad n = 0, \ldots, L-1, \\ 0 & \text{dla} \quad n < 0 \quad \text{i} \quad n \geqslant L. \end{cases} \tag{4.15}$$

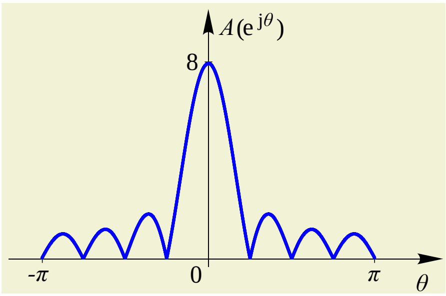
Impuls ten można otrzymać z impulsu (4.6), przesuwając go o $n_0 = N$ próbek w prawo i przyjmując oznaczenie $2N+1 = L$. Widmo impulsu (4.6) zostało wyznaczone w przykładzie 4.2 i jest opisane wzorem (4.7). Korzystając z twierdzenia o przesunięciu w dziedzinie czasu i podstawiając we wzorze (4.7) $N = (L-1)/2$, otrzymujemy:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \mathrm{e}^{-\mathrm{j}(L-1)\theta/2}\,\frac{\sin(L\theta/2)}{\sin(\theta/2)}. \tag{4.16}$$

Na rys. 4.4 jest przedstawione widmo amplitudowe impulsu (4.15):

$$A(\mathrm{e}^{\mathrm{j}\theta}) = |X(\mathrm{e}^{\mathrm{j}\theta})| = \left|\frac{\sin(L\theta/2)}{\sin(\theta/2)}\right|. \tag{4.17}$$

Wykres został sporządzony w przedziale $-\pi \leqslant \theta \leqslant \pi$ dla przypadku $L = 8$.



Rys. 4.4. Widmo amplitudowe impulsu prostokątnego (4.15) dla $L = 8$

Twierdzenie 4.3 (*o przesunięciu w dziedzinie częstotliwości – o modulacji)*

$$x[n]\,\mathrm{e}^{\mathrm{j}n\theta_0} \leftrightarrow X(\mathrm{e}^{\mathrm{j}(\theta-\theta_0)}), \tag{4.18a}$$

$$x[n]\cos n\theta_0 \leftrightarrow \frac{1}{2}\left[X(\mathrm{e}^{\mathrm{j}(\theta-\theta_0)}) + X(\mathrm{e}^{\mathrm{j}(\theta+\theta_0)})\right], \tag{4.18b}$$

$$x[n]\sin n\theta_0 \leftrightarrow \frac{1}{2\,\mathrm{j}}\left[X(\mathrm{e}^{\mathrm{j}(\theta-\theta_0)}) - X(\mathrm{e}^{\mathrm{j}(\theta+\theta_0)})\right]. \tag{4.18c}$$

Interpretacja wszystkich trzech wersji tego twierdzenia jest identyczna jak w przypadku sygnałów analogowych.

Twierdzenie 4.4 (*o splocie w dziedzinie czasu)*

$$x[n] * y[n] \triangleq \sum_{i=-\infty}^{\infty} x(n-i)y(i) \leftrightarrow X(\mathrm{e}^{\mathrm{j}\theta})Y(\mathrm{e}^{\mathrm{j}\theta}). \tag{4.19}$$

Widmo splotu dwóch sygnałów dyskretnych jest iloczynem ich widm. W twierdzeniu tym splot jest rozumiany jako *splot dyskretny liniowy* ([2]. p. 5.9 i 5.10).

Twierdzenie 4.5 (*o splocie w dziedzinie częstotliwości*)

$$x[n]y[n] \leftrightarrow \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\mathrm{e}^{\mathrm{j}(\theta-\lambda)})Y(\mathrm{e}^{\mathrm{j}\lambda})\,\mathrm{d}\lambda. \tag{4.20}$$

Widmo iloczynu dwóch sygnałów dyskretnych jest splotem ich widm. W wyniku operacji splotu dwóch funkcji okresowych o jednakowym okresie otrzymujemy funkcję okresową o tym samym okresie.

Twierdzenie 4.6 (*uogólnione twierdzenie Rayleigha*)

$$\sum_{n=-\infty}^{\infty} x(n)y^*(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\mathrm{e}^{\mathrm{j}\theta})Y^*(\mathrm{e}^{\mathrm{j}\theta})\,\mathrm{d}\theta. \tag{4.21}$$

Twierdzenie to orzeka równość (z dokładnością do współczynnika 2π) iloczynów skalarnych w przestrzeniach l^2 i $L^2_{2\pi}$ (por. komentarz w p. 4.1.4).

Twierdzenie 4.7 (*twierdzenie Parsevala – o energii*)

$$\sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\mathrm{e}^{\mathrm{j}\theta})|^2\,\mathrm{d}\theta. \tag{4.22}$$

Twierdzenie Parsevala dla sygnałów dyskretnych wynika z twierdzenia Rayleigha (4.21) przy podstawieniu $y[n] = x[n]$. Zgodnie z tym twierdzeniem energię sygnału dyskretnego możemy obliczać w dziedzinie częstotliwości jako całkę za okres z kwadratu widma amplitudowego sygnału podzieloną przez 2π. Funkcję $|X(\mathrm{e}^{\mathrm{j}\theta})|^2 \triangleq A^2(\mathrm{e}^{\mathrm{j}\theta})$ nazywamy *widmem energii* sygnału dyskretnego. Widmo energii sygnału $x[n]$ będziemy także oznaczać $\Phi_x(\mathrm{e}^{\mathrm{j}\theta})$ (por. p. 5.5.2).

4.3. Dyskretny szereg Fouriera. Widmo dyskretnego sygnału okresowego

4.3.1. Sygnały N-okresowe

Jak podkreśliliśmy w p. 1.3.3, próbkowanie okresowego sygnału analogowego nie zawsze daje w wyniku okresowy sygnał dyskretny. Załóżmy jednak, że okresowy sygnał analogowy $x(t)$ o okresie T_0 jest próbkowany dokładnie N razy w każdym okresie sygnału $x(t)$ w chwilach $t_n = nT_s = nT_0/N$. W takim przypadku sygnał spróbkowany jest sygnałem okresowym spełniającym warunek $\bar{x}[nT_s] = \bar{x}[(n+N)T_s]$. Wprowadzając czas unormowany względem okresu próbkowania T_s, otrzymujemy dyskretny sygnał okresowy $\bar{x}[n] = \bar{x}[n+N]$, którego próbki $\bar{x}(n)$, $n = 0, \ldots, N-1$, pobrane w przedziale $[0, T_0]$, powtarzają się co N próbek. Sygnały takie nazywamy sygnałami (ciągami) N-*okresowymi*.

W celu podkreślenia cechy okresowości i wyróżnienia sygnałów N-okresowych, będziemy je oznaczać kreską u góry. Sygnały N-okresowe są oczywiście sygnałami o ograniczonej mocy, tzn. (por. wzór (1.27)):

$$\frac{1}{N}\sum_{n=0}^{N-1} |\bar{x}(n)|^2 < \infty.$$

4.3.2. Dyskretny szereg Fouriera sygnału N-okresowego

Załóżmy, że okresowy sygnał analogowy $x(t)$ o okresie T_0, reprezentowany swoim zespolonym szeregiem Fouriera:

$$x(t) = \sum_{k=-\infty}^{\infty} X_k \, \mathrm{e}^{\mathrm{j}k\frac{2\pi}{T_0}t}, \tag{4.23}$$

jest próbkowany w chwilach $t_n = nT_s = nT_0/N$. W chwilach tych spełniona jest równość:

$$\bar{x}(nT_s) = \sum_{k=-\infty}^{\infty} X_k \, \mathrm{e}^{\mathrm{j}k\frac{2\pi}{T_0}\frac{nT_0}{N}}.$$

W wyniku próbkowania sygnału $x(t)$ powstaje zatem N-okresowy sygnał $\bar{x}[n]$ o postaci:

$$\bar{x}(n) = \sum_{k=-\infty}^{\infty} X_k \, \mathrm{e}^{\mathrm{j}2\pi kn/N}. \tag{4.24}$$

Wyrażenie (4.24) stanowi reprezentację sygnału $\bar{x}[n]$ w postaci nieskończonej sumy dyskretnych zespolonych sygnałów harmonicznych $\mathrm{e}^{\mathrm{j}2\pi kn/N}$, $k = 0, \pm 1, \ldots$, ważonych współczynnikami X_k. Rozpatrzymy tę sumę dokładniej. Dla wygody rozważań wprowadzimy wielkość pomocniczą

$$W = \mathrm{e}^{\mathrm{j}2\pi/N} \tag{4.25}$$

równą pierwiastkowi N-tego stopnia z jedności. Wielkość W jest liczbą zespoloną, której moduł jest jednostkowy, a argument równy $2\pi/N$. Jej N-ta potęga jest równa jedności: $W^N = (\mathrm{e}^{\mathrm{j}2\pi/N})^N = \mathrm{e}^{\mathrm{j}2\pi} = 1$. Wielkość ta odgrywa w teorii sygnałów dyskretnych bardzo ważną rolę. Jej znaczenie wynika z faktu, że w wyniku próbkowania N razy w okresie w chwilach nT_s zespolonego analogowego sygnału harmonicznego $e^{j\omega_0 t}$ otrzymujemy próbki będące kolejnymi potęgami liczby W. Istotnie

$$e^{j\omega_0 n T_s} = e^{j2\pi n T_s/T_0} = (e^{j2\pi/N})^n = W^n, \qquad n = 0, \ldots, N-1.$$

Za pomocą wielkości W zespolone sygnały harmoniczne występujące we wzorze (4.24) można zapisać w skróconej postaci:

$$e^{j2\pi kn/N} = (e^{j2\pi/N})^{(kn)} = W^{kn}. \tag{4.26}$$

Zbadamy właściwości tych sygnałów. Zauważmy przede wszystkim, że są one N-okresowe:

$$W^{k(n+N)} = W^{kn} W^{kN} = W^{kn} e^{j2\pi k} = W^{kn}. \tag{4.27}$$

Wykażemy ponadto, że dla $k = 0, \ldots, N-1$ sygnały (4.26) tworzą układ ortonormalny w przestrzeni Hilberta zespolonych N-okresowych ciągów liczbowych. W tym celu obliczymy ich iloczyny skalarne w tej przestrzeni $(m, k = 0, \ldots, N-1)$:

$$\frac{1}{N}\sum_{n=0}^{N-1}(W^{kn})(W^{mn})^* = \frac{1}{N}\sum_{n=0}^{N-1} W^{(k-m)n}. \tag{4.28}$$

Dla $m = k$ mamy $W^{(k-m)n} = W^0 = 1$. Suma we wzorze (4.28) jest więc w tym przypadku równa N, a iloczyn skalarny równy 1. Dla $m \neq k$ korzystamy ze wzoru na sumę częściową szeregu geometrycznego:

$$\frac{1}{N}\sum_{n=0}^{N-1} W^{(k-m)n} = \frac{1}{N}\frac{1 - W^{(k-m)N}}{1 - W^{(k-m)}}.$$

Ponieważ dla dowolnych k, m, $(m \neq k)$, mamy $W^{(k-m)N} = (W^N)^{(k-m)} = 1$ i jednocześnie $W^{(k-m)} \neq 1$, zatem ostatecznie otrzymujemy:

$$\frac{1}{N}\sum_{n=0}^{N-1}(W^{kn})(W^{mn})^* = \begin{cases} 0 & \text{dla} \quad m \neq k, \\ 1 & \text{dla} \quad m = k, \end{cases} \tag{4.29}$$

co kończy dowód ortonormalności sygnałów W^{kn}.

Wprowadzając oznaczenie (4.25) do wzoru (4.24), możemy przepisać go w zwartej postaci:

$$\bar{x}(n) = \sum_{k=-\infty}^{\infty} X_k W^{kn}. \tag{4.30}$$

Z właściwości N-okresowości sygnałów W^{kn} wynika, że co każdy N-ty współczynnik X_k w sumie (4.30) jest mnożony przez ten sam sygnał W^{kn}, tzn. współczynniki o indeksach iN, $i = 0, \pm 1, \ldots$, będą mnożone przez sygnał W^0, współczynniki o indeksach $1 + iN$ będą mnożone przez sygnał W^n, współczynniki o indeksach $2+iN$ będą mnożone przez sygnał W^{2n} itd., aż do współczynników o indeksach $(N-1)+iN$, które będą mnożone przez sygnał $W^{(N-1)n}$. Sumę we wzorze (4.30) można zatem rozbić na N sum odpowiadającym kolejnym wartościom wskaźnika $k = 0, \ldots, N-1$:

$$\bar{x}(n) = \left(\sum_{i=-\infty}^{\infty} X_{iN}\right) W^0 + \left(\sum_{i=-\infty}^{\infty} X_{1+iN}\right) W^n + \cdots + \\ + \left(\sum_{i=-\infty}^{\infty} X_{k+iN}\right) W^{kn} + \cdots + \left(\sum_{i=-\infty}^{\infty} X_{N-1+iN}\right) W^{(N-1)n}. \quad (4.31)$$

Zdefiniujemy teraz nowe współczynniki (nazywane niekiedy w literaturze *współczynnikami nałożonymi* względem współczynników X_k):

$$X(k) \triangleq N \sum_{i=-\infty}^{\infty} X_{k+iN}, \qquad k = 0, \ldots, N-1. \quad (4.32)$$

Wzór (4.30) możemy wówczas zapisać w postaci:

$$\bar{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W^{kn}. \quad (4.33)$$

Definicja 4.3. Wyrażenie (4.33) nazywamy *dyskretnym szeregiem Fouriera* N-okresowego sygnału $\bar{x}[n]$. Wprowadzając ponownie jawną postać wielkości W, szereg ten możemy zapisać w postaci:

$$\bar{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)\, \mathrm{e}^{\mathrm{j}2\pi kn/N}. \quad (4.34)$$

Dyskretny szereg Fouriera (4.34) możemy rozpatrywać jako reprezentację N-okresowego sygnału $\bar{x}[n]$ w przestrzeni Hilberta N-okresowych sygnałów, wyznaczoną względem ortonormalnej skończonej N -elementowej bazy $\{\mathrm{e}^{\mathrm{j}2\pi kn/N} \colon k = 0, \ldots, N-1\}$ dyskretnych sygnałów harmonicznych o pulsacjach unormowanych $\theta_k = 2\pi k/N$. Współczynnikami tej reprezentacji (amplitudami zespolonymi sygnałów bazowych) są liczby zespolone $X(k)/N$.

4.3.3. Widmo sygnału N-okresowego

Pojęcie widma dyskretnego sygnału N-okresowego jest oparte na jego rozwinięciu w dyskretny szereg Fouriera. *Widmem* sygnału $\bar{x}[n]$ nazywamy zbiór $\{X(k) : k = 0, \ldots, N-1\}$ współczynników jego rozwinięcia w szereg (4.34). *Widmem amplitudowym* nazywamy zbiór $\{|X(k)| \triangleq A(k)\}$ modułów tych współczynników, zaś *widmem fazowym* – zbiór $\{\arg X(k) \triangleq \varphi(k)\}$ ich argumentów.

W przeciwieństwie do widm nieokresowych sygnałów dyskretnych, które są funkcjami ciągłymi pulsacji unormowanej θ, widma sygnałów N-okresowych są funkcjami dyskretnymi. Można je rozpatrywać jako funkcje zmiennej dyskretnej $\theta_k = 2\pi k/N$ lub zmiennej całkowitej k.

4.3.4. Właściwości widma

Z określenia współczynników $X(k)$ (wzór (4.32)) wynika, że są one N-okresowe, tzn. $X(k) = X(k+N)$. Tak więc, zarówno próbki $\bar{x}(n)$ sygnału, jak i wartości jego widma $X(k)$ tworzą ciągi N-okresowe: $\bar{x}[n] = \bar{x}[n+N]$ i odpowiednio $X[k] = X[k+N]$.

Ze wzoru (4.32) wynika także, że właściwości współczynników X_k szeregu (4.23) przenoszą się na właściwości współczynników $X(k)$ szeregu (4.34). Ponieważ dla sygnałów rzeczywistych $X_k = X^*_{-k}$, zatem dla tej klasy sygnałów współczynniki $X(k)$ spełniają równość:

$$X(k) = X^*(-k), \tag{4.35}$$

z której wynika, że

$$A(k) = A(-k), \qquad \varphi(k) = -\varphi(-k). \tag{4.36}$$

Widmo amplitudowe rzeczywistego sygnału N-okresowego jest więc funkcją parzystą, a widmo fazowe – funkcją nieparzystą zmiennej k.

Rozpatrzymy jeszcze wartości $X(k)$ widma w jednym jego okresie obejmującym punkty $k = 0, \ldots, N-1$ przy założeniu, że N-okresowy sygnał $\bar{x}[n]$ jest rzeczywisty, a liczba N jest parzysta (w praktyce liczbę tę wybiera się z reguły jako liczbę parzystą). W takim przypadku wartości $X(k)$ o indeksach położonych symetrycznie względem $N/2$ są parami sprzężone:

$$X(k) = X^*(N-k). \tag{4.37}$$

Właściwość ta wynika natychmiast z N-okresowości widma $X[k]$ i właściwości (4.35). Mamy bowiem $X^*(N-k) = X^*(N-k-N) = X^*(-k) = X(k)$. Dochodzimy zatem do wniosku, że jeśli sygnał N-okresowy jest rzeczywisty i liczba jego próbek jest parzysta, to jego widmo jest w pełni określone $N/2+1$ liczbami: dwiema liczbami rzeczywistymi $X(0)$ i $X(N/2)$ oraz $N/2-1$ liczbami zespolonymi $X(1), \ldots, X(N/2-1)$.

Znając wartości widma sygnału N-okresowego dla $k = 0, \ldots, N-1$ (a w przypadku sygnałów rzeczywistych i parzystej liczby N – dla $k = 0, \ldots, N/2$), możemy, zgodnie ze wzorem (4.34), wyznaczyć próbki tego sygnału. Dyskusję problemu odwrotnego, tj. sposobu wyznaczania widma sygnału N-okresowego na podstawie jego próbek odłożymy do p. 4.4.7, jest on bowiem ściśle związany z kolejnym ważnym pojęciem teorii sygnałów dyskretnych, a mianowicie dyskretnym przekształceniem Fouriera.

4.3.5. Dystrybucyjna reprezentacja widma sygnału okresowego

Jak wspomnieliśmy w p. 4.1.6, widma dyskretnych sygnałów okresowych mogą być, podobnie jak w przypadku analogowych sygnałów okresowych, określone w sensie granicznym jako wielkości dystrybucyjne (w tym wypadku zmiennej ciągłej θ). Na przykład, korzystając z pary (4.11) oraz twierdzenia o modulacji (4.18b), możemy przedstawić widmo dyskretnego sygnału harmonicznego $X_0 \cos n\theta_0$ w postaci:

$$X(e^{j\theta}) = \mathscr{F}[X_0 \cos n\theta_0] = \pi X_0[\delta_{2\pi}(\theta - \theta_0) + \delta_{2\pi}(\theta + \theta_0)]. \tag{4.38}$$

W teorii sygnałów dyskretnych widma sygnałów okresowych nie są zwykle rozpatrywane jako wielkości dystrybucyjne, a ich definicję opiera się na rozwinięciu sygnału okresowego w dyskretny szereg Fouriera. Niekiedy jednak ze względów analitycznych wygodniej jest posłużyć się ich opisem dystrybucyjnym.

4.4. Dyskretne przekształcenie Fouriera (DPF)

4.4.1. Zagadnienie numerycznego obliczania widma sygnału

W poprzednich punktach scharakteryzowaliśmy podstawowe narzędzia teoretyczne analizy częstotliwościowej sygnałów dyskretnych. Opierając się na wprowadzonych pojęciach i korzystając z odpowiednich twierdzeń i właściwości, możemy wyznaczać widma sygnałów dyskretnych metodami analitycznymi. Stosowanie metod analitycznych wymaga jednak znajomości jawnej postaci formuły matematycznej opisującej sygnał. Nawet jeśli formuła ta jest znana, jak np. w przypadku omówionych wcześniej prostych sygnałów: dyskretnego impulsu prostokątnego (przykład 4.2), czy dyskretnego sygnału wykładniczego (przykład 4.3), wyznaczenie ich widm jest uciążliwe i związane ze stosunkowo skomplikowanymi obliczeniami. W przypadku bardziej złożonych sygnałów uzyskanie dokładnych postaci analitycznych widma jest z reguły niemożliwe. W zastosowaniach praktycznych spotykamy się obecnie z sygnałami na tyle złożonymi, że nie można ich opisać zamkniętymi formułami zarówno w dziedzinie czasu, jak i dziedzinie częstotliwości. Z tego względu współczesna analiza częstotliwościowa sygnałów dyskretnych jest dokonywana głównie metodami numerycznymi z wykorzystaniem komputerów i specjalizowanych układów mikroprocesorowych (tzw. procesorów sygnałowych).

Spójrzmy na zagadnienie analizy widmowej sygnału oczami badacza-praktyka i rozpatrzmy typową sytuację jaka występuje w praktycznych zagadnieniach przetwarzania sygnałów. Załóżmy, że obserwujemy analogowy sygnał $x(t)$, np. sygnał otrzymany na wyjściu czujnika pomiarowego, sygnał fizjologiczny człowieka, czy też sygnał odebrany przez antenę stacji radiolokacyjnej. Sygnał ten możemy obserwować jedynie w pewnym przedziale czasu $[0, T]$, zwanym *przedziałem obserwacji*. Może być on krótszy lub dłuższy, ale zawsze jest skończony. Poza tym przedziałem informacja o sygnale jest niedostępna. W celu przeprowadzenia analizy widmowej obserwowanego sygnału próbkujemy go ze stałym przedziałem próbkowania T_s w chwilach $t_0 = 0$, $t_1 = T_s, \ldots, t_{N-1} = (N-1)T_s$, pobierając w przedziale obserwacji $[0, T]$ skończoną liczbę $N = T/T_s$ próbek. Wprowadzając czas bezwymiarowy n, unormowany względem okresu próbkowania, tworzymy sygnał dyskretny $x[n]$ o skończonym czasie trwania N. Postępując w ten sposób, przyjmujemy milczące założenie, że wartości próbek poza przedziałem $[0, N-1]$ są równe zeru. Próbki (liczby) $x(n)$, $n = 0, \ldots, N-1$ stanowią jedyną informację, na podstawie której powinniśmy przeprowadzić analizę właściwości widmowych sygnału analogowego $x(t)$ metodami komputerowymi.

Komputerowe metody analizy widmowej sygnałów opierają się na koncepcji *dyskretnego przekształcenia Fouriera*. Pojęcie to odgrywa kluczową rolę w teorii sygnałów dyskretnych. Stanowi ono nie tylko punkt wyjścia do opracowania odpowiednich numerycznych metod wyznaczania widm, ale jest również silnym narzędziem teoretycznym wykorzystywanym w zagadnieniach analizy, syntezy i przetwarzania sygnałów dyskretnych.

Dyskusję dyskretnego przekształcenia Fouriera przeprowadzimy najpierw dla przypadku dyskretnych sygnałów impulsowych o skończonym czasie trwania N. Następnie rozpatrzymy przypadek sygnałów N-okresowych i pokażemy związek między dyskretnym przekształceniem Fouriera a dyskretnym szeregiem Fouriera sygnału N-okresowego. Na końcu omówimy przypadek sygnałów nieokresowych o nieskończonym czasie trwania. Taki podział rozważań jest uwarunkowany specyfiką każdego z tych trzech przypadków.

4.4.2. Proste dyskretne przekształcenie Fouriera sygnału impulsowego

Załóżmy, że dyskretny sygnał $x[n]$ o skończonym czasie trwania N jest określony swoimi próbkami $x(n)$, $n = 0, 1, \ldots, N-1$. Zgodnie ze wzorem ogólnym (4.4), jego widmo jest opisane wzorem:

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{n=0}^{N-1} x(n)\,\mathrm{e}^{-\mathrm{j}n\theta} \tag{4.39}$$

i jest funkcją okresową o okresie 2π ciągłej zmiennej θ.

Metody numeryczne obliczania widma (4.39) polegają na wyznaczeniu jego wartości $X(\mathrm{e}^{\mathrm{j}\theta_k})$ w skończonej liczbie równoodległych punktów $\theta_k \in [-\pi, \pi]$. Liczba tych punktów może być różna, ale sensownie jest wybrać ją w taki sposób, aby na podstawie zbioru wartości widma $\{X(\mathrm{e}^{\mathrm{j}\theta_k})\}$ można było jednoznacznie odtworzyć sygnał, tzn. obliczyć jego próbki $x(n)$ w chwilach $n = 0, 1, \ldots, N-1$. Otrzymalibyśmy wówczas wzajemnie jednoznaczny związek między sygnałem a dyskretną reprezentacją jego widma. Najmniejsza niezbędna w tym celu liczność zbioru $\{X(\mathrm{e}^{\mathrm{j}\theta_k})\}$ jest równa liczbie N próbek sygnału. Otrzymujemy wtedy bowiem układ N równań z N niewiadomymi. Wynika stąd, że wartości $X(\mathrm{e}^{\mathrm{j}\theta_k})$ widma $X(\mathrm{e}^{\mathrm{j}\theta})$ powinniśmy obliczać w punktach $\theta_k = 2\pi k/N$, $k = 0, 1, \ldots, N-1$, odległych od siebie o $\Delta\theta = 2\pi/N$. Przeprowadzone rozumowanie prowadzi bezpośrednio do pojęcia dyskretnego przekształcenia Fouriera.

Definicja 4.4. *Dyskretnym przekształceniem Fouriera* (*DPF*) sygnału impulsowego $x[n]$ określonego w chwilach $n = 0, 1, \ldots, N-1$ nazywamy odwzorowanie

$$X(\mathrm{e}^{\mathrm{j}2\pi k/N}) = \sum_{n=0}^{N-1} x(n)\,\mathrm{e}^{-\mathrm{j}2\pi kn/N} \tag{4.40}$$

przyporządkowujące temu sygnałowi funkcję okresową $X[\mathrm{e}^{\mathrm{j}2\pi k/N}]$ zmiennej k o okresie N. Funkcję tę będziemy oznaczać w skrócie $X[k] = \mathscr{F}_D[x(n)]$ i nazywać *dyskretną transformatą Fouriera* (*DTF*) lub *widmem dyskretnym* sygnału $x[n]$. Funkcję $A[k] \triangleq |X[k]|$ będziemy nazywać *dyskretnym widmem amplitudowym*, a funkcję $\varphi[k] \triangleq \arg X[k]$ – *dyskretnym widmem fazowym* sygnału $x[n]$.

Przez analogię do próbek sygnału, wartości $X(k)$ dyskretnej transformaty Fouriera $X[k]$ nazywamy *próbkami widma*. Dla podkreślenia faktu, że DTF (4.40) jest wyznaczona na podstawie N próbek sygnału, nazywamy ją *N-punktową DTF*, a liczbę N – jej *wymiarem*.

Ponieważ z definicji $X(k) = X(\mathrm{e}^{\mathrm{j}\theta})|_{\theta=2\pi k/N}$, zatem okresowość DTF jest konsekwencją okresowości widma (4.39). Tak więc, w wyniku dyskretnego przekształcenia Fouriera N próbkom $x(n)$ sygnału zostaje przyporządkowanych N okresowo powtarzanych próbek $X(k)$ jego widma. DTF (4.40) możemy zatem traktować jako N-okresowy ciąg liczbowy $X[k] = X[k+N]$ w dziedzinie częstotliwości.

Wprowadzając wielkość W określoną wzorem (4.25), wzór definicyjny (4.40) możemy zapisać w prostszej postaci:

$$X(k) = \sum_{n=0}^{N-1} x(n) W^{-kn}. \tag{4.41}$$

Przykład 4.6. Wyznaczymy DTF dyskretnego impulsu prostokątnego (4.15). Podstawiając wartości tego impulsu do wzoru (4.41), otrzymujemy:

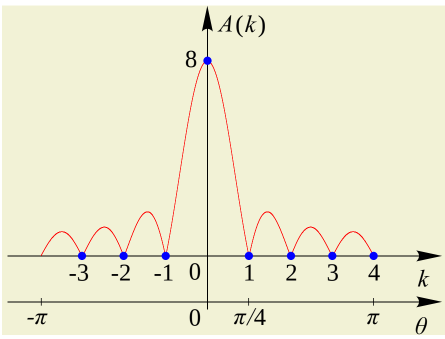
$$\begin{aligned} X(k) &= \sum_{n=0}^{L-1} W^{-kn} = \frac{1-W^{-kL}}{1-W^{-k}} = \frac{1-\mathrm{e}^{-\mathrm{j}\frac{2\pi}{L}kL}}{1-\mathrm{e}^{-\mathrm{j}\frac{2\pi}{L}k}} = \\ &= \frac{\mathrm{e}^{-\mathrm{j}\pi k}}{\mathrm{e}^{-\mathrm{j}\pi k/L}} \frac{\mathrm{e}^{\mathrm{j}\pi k} - \mathrm{e}^{-\mathrm{j}\pi k}}{\mathrm{e}^{\mathrm{j}\pi k/L} - \mathrm{e}^{-\mathrm{j}\pi k/L}} = \mathrm{e}^{\mathrm{j}\pi k(L-1)/L} \frac{\sin(\pi k)}{\sin(\pi k/L)}. \end{aligned} \tag{4.42}$$

Obliczenia są tu podobne jak w przykładzie 4.5. Przeprowadziliśmy je opierając się na definicji (4.41), choć DTF impulsu (4.15) mogliśmy w tym przypadku wyznaczyć bezpośrednio na podstawie wzoru (4.16) opisującego jego widmo ciągłe $X(\mathrm{e}^{\mathrm{j}\theta})$. Wystarczyłoby w tym celu obliczyć wartości tego widma w punktach $\theta_k = 2\pi k/L$. Podkreślmy, że w rozpatrywanym przypadku okres DTF (4.42) jest równy L.

Na rys. 4.5 wykreślono moduł DTF $|X(k)|$, tj. dyskretne widmo amplitudowe $A(k)$ impulsu (4.15):

$$A(k) = \left| \frac{\sin(\pi k)}{\sin(\pi k/L)} \right|. \tag{4.43}$$

Wykres sporządzono dla przypadku $L = 8$ w przedziale $(-L/2, L/2]$ obejmującym jedynie centralny okres widma $A[k]$ o długości $L = 8$. Okres ten odpowiada przedziałowi $[-\pi, \pi]$ na skali ciągłej zmiennej θ. Na tym samym rysunku zamieszczono wykres ciągłego widma amplitudowego $A(\mathrm{e}^{\mathrm{j}\theta})$ dla tej samej wartości $L = 8$ (por. rys. 4.4). Próbki $A(k)$ są oczywiście położone na krzywej $A(\mathrm{e}^{\mathrm{j}\theta})$, gdyż $X(k) = X(\mathrm{e}^{\mathrm{j}\theta})|_{\theta=2\pi k/L}$. Widzimy jednak, że tylko próbka centralna dla $k = 0$ jest różna od zera (i równa $L = 8$). Pozostałe próbki wypadają w punktach zerowych widma $A(\mathrm{e}^{\mathrm{j}\theta})$. Można zatem uznać, że DTF $X[k]$ niezbyt dokładnie oddaje w tym przypadku charakter widma ciągłego $X(\mathrm{e}^{\mathrm{j}\theta})$. Stanowi to pewną wadę DTF związaną z jej skończoną rozdzielczością.



Rys. 4.5. Dyskretne widmo amplitudowe impulsu prostokątnego (4.15) dla $L = 8$

4.4.3. Odwrotne dyskretne przekształcenie Fouriera sygnału impulsowego

Rozpatrzymy teraz problem odwrotny, tzn. wyznaczymy impulsowy sygnał $x[n]$ określony w chwilach $n = 0, \ldots, N-1$ na podstawie jego N-punktowej DTF $X[k]$. Inaczej mówiąc, na podstawie próbek $X(k)$ widma, $k = 0, \ldots, N-1$, wyznaczymy próbki $x(n)$, sygnału $n = 0, \ldots, N-1$, . W tym celu zmienimy wskaźnik sumowania we wzorze (4.41) z n na i, a następnie pomnożymy obie strony równości (4.41) przez W^{kn} kolejno dla $k = 0, \ldots, N-1$. Otrzymamy N równości:

$$X(k)W^{kn} = \sum_{i=0}^{N-1} x(i)W^{-ki}W^{kn}, \qquad k = 0, \ldots, N-1.$$

Z kolei zsumujemy te równości względem k:

$$\sum_{k=0}^{N-1} X(k)W^{kn} = \sum_{k=0}^{N-1}\sum_{i=0}^{N-1} x(i)W^{-ki}W^{kn}.$$

Biorąc pod uwagę ortonormalność funkcji W^{-ki} dla $i = 0, \ldots, N-1$ (którą dowodzi się analogicznie jak w p. 4.3.2 dla przypadku sygnałów (4.26) – por. wzór (4.29)), widzimy, że dla każdego k w sumie wewnętrznej po prawej stronie ostatniej równości tylko jeden składnik dla $i = n$ jest różny od zera i równy próbce $x(n)$. Ponieważ składników tych jest dokładnie N, otrzymujemy:

$$\sum_{k=0}^{N-1} X(k)W^{kn} = Nx(n),$$

skąd wynika ostateczne rozwiązanie postawionego problemu:

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)W^{kn} = \frac{1}{N}\sum_{k=0}^{N-1} X(k)\,\mathrm{e}^{\mathrm{j}2\pi kn/N}. \tag{4.44}$$

Definicja 4.5. *Odwrotnym dyskretnym przekształceniem Fouriera (ODPF)* nazywamy odwzorowanie określone wzorem (4.44), przyporządkowujące widmu dyskretnemu $X[k]$, $k = 0, \ldots, N-1$, sygnał $x[n]$, $n = 0, \ldots, N-1$. Odwzorowanie to będziemy oznaczać $x[n] = \mathscr{F}_D^{-1}\{X[k]\}$, a sygnał $x[n]$ – nazywać *odwrotną dyskretną transformatą Fouriera* (*ODTF*) lub krótko $\mathscr{F}_D^{-1}$*-transformatą*.

Wzory (4.41) i (4.44) określają zatem parę dyskretnych transformat Fouriera:

$$X(k) = \sum_{n=0}^{N-1} x(n)W^{-kn}, \qquad k = 0, \ldots, N-1, \tag{4.45}$$

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)W^{kn}, \qquad n = 0, \ldots, N-1. \tag{4.46}$$

Parę tę będziemy oznaczać symbolicznie $x[n] \underset{N}{\leftrightarrow} X[k]$.

4.4.4. Interpretacja $\mathscr{F}_D^{-1}$-transformaty

W punkcie 4.1.5 $\mathscr{F}^{-1}$-transformatę określoną wzorem (4.10) interpretowaliśmy jako reprezentację dyskretnego sygnału $x[n]$ w postaci nieskończonej sumy elementarnych zespolonych dyskretnych sygnałów harmonicznych $\mathrm{e}^{\mathrm{j}n\theta}$ o pulsacjach należących do ciągłego przedziału $[-\pi, \pi]$, których gęstość amplitudy zespolonej jest określona widmem ciągłym $X(\mathrm{e}^{\mathrm{j}\theta})$ podzielonym przez 2π. W podobny sposób możemy interpretować $\mathscr{F}_D^{-1}$-transformatę określoną wzorem (4.46). Zgodnie z tym wzorem, dyskretny sygnał impulsowy $x[n]$, określony w chwilach $n = 0, \ldots, N-1$, może być zdekomponowany na N zespolonych dyskretnych sygnałów (składowych) harmonicznych $W^{kn} = \mathrm{e}^{\mathrm{j}2\pi kn/N}$ o pulsacjach $\theta_k = 2\pi k/N$, $k = 0, \ldots, N-1$, i amplitudach zespolonych równych $X(k)/N$, gdzie $X(k)$ są wartościami widma dyskretnego $X[k]$ określonego wzorem (4.45). Wyrażenie (4.46) można zatem uważać za rozwinięcie sygnału $x[n]$ w skończony dyskretny zespolony szereg Fouriera względem N-elementowej ortonormalnej bazy $\{W^{kn} : k = 0, \ldots, N-1\}$, którego współczynnikami są próbki widma $X(k)$ podzielone przez N.

4.4.5. Okresowość $\mathscr{F}_D^{-1}$-transformaty

Przeanalizujmy poszczególne składowe harmoniczne szeregu (4.44). Dla $k = 0$ otrzymujemy składową stałą $X(0)/N$. Dla $k = 1$ otrzymujemy pierwszą składową harmoniczną $[X(1)\,\mathrm{e}^{\mathrm{j}2\pi n/N}]/N$ o pulsacji $\theta_1 = 2\pi/N$. Jest to sygnał N-okresowy. Kolejne składowe harmoniczne są również okresowe o okresach N/k, będących podwielokrotnościami okresu N. Suma tych składowych jest zatem sygnałem N-okresowym, co dowodzi, że $\mathscr{F}_D^{-1}$-transformata (4.46) jest N-okresowa. Właściwość N-okresowości $\mathscr{F}_D^{-1}$-transformaty można także wykazać bezpośrednio, podstawiając we wzorze (4.46) $n+N$ w miejsce n. Wynika stąd ważny wniosek: wyrażenie (4.46) opisuje rozwinięcie w zespolony szereg Fouriera względem N-elementowej ortonormalnej bazy $\{W^{kn} : k = 0, \ldots, N-1\}$ zarówno dyskretnego sygnału impulsowego $x[n]$ określonego w chwilach $n = 0, \ldots, N-1$, jak i jego przedłużenia okresowego z okresem N:

$$\bar{x}[n] = \sum_{i=-\infty}^{\infty} x[n + iN] \tag{4.47}$$

utworzonego w wyniku nieskończonego zsumowania sygnałów $x[n]$ poprzesuwanych o kolejne wielokrotności liczby N. Wzór (4.45) definiuje zatem z jednej strony DTF impulsowego sygnału dyskretnego $x[n]$ określonego w skończonym przedziale $n = 0, \ldots, N-1$, z drugiej zaś strony DTF jego N-okresowego przedłużenia $\bar{x}[n]$. Z uwagi na N-okresowość $\mathscr{F}_D$-transformaty (4.45), jest ona w obu przypadkach jednoznacznie określona skończonym zbiorem N liczb zespolonych $X(k)$, $k = 0, \ldots, N-1$.

Komentarz. Sytuacja jest tu analogiczna do rozwinięcia nieokresowego impulsowego sygnału analogowego określonego w skończonym przedziale $[0, T]$ w zespolony szereg Fouriera (2.57), który jest zarazem zespolonym szeregiem Fouriera okresowego przedłużenia tego sygnału z okresem T (por. [3], przykład 13 w p. 3.2.7, a także [4], p. 5.1.1-A, B).

4.4.6. Para dyskretnych transformat Fouriera sygnału N-okresowego

Rozważmy sygnał N-okresowy $\bar{x}[n]$ określony w jednym okresie próbkami $x(n)$, $n = 0, \ldots, N-1$. Z dyskusji przeprowadzonej w p. 4.4.5 wynika, że wzory (4.45), (4.46) można przyjąć za definicje prostego i odpowiednio odwrotnego dyskretnego przekształcenia Fouriera tego sygnału:

$$X(k) = \sum_{n=0}^{N-1} \bar{x}(n) W^{-kn}, \qquad k = 0, \ldots, N-1, \tag{4.48}$$

$$\bar{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W^{kn}, \qquad n = 0, \ldots, N-1. \tag{4.49}$$

Para dyskretnych przekształceń Fouriera (4.48) i (4.49) ustala wzajemnie jednoznaczny związek między dwoma N -okresowymi ciągami liczb $\bar{x}[n]$ i $X[k]$.

Komentarz. Pod względem formalnym DPF możemy traktować jako wzajemnie jednoznaczne odwzorowanie rzeczywistej lub zespolonej przestrzeni Hilberta l_N^2 ciągów N-okresowych w dziedzinie czasu w zespoloną przestrzeń Hilberta l_N^2 ciągów N-okresowych w dziedzinie częstotliwości.

4.4.7. Związek z dyskretnym szeregiem Fouriera

Z podanych wyżej definicji wynika, że $\mathscr{F}_D^{-1}$-transformatę sygnału N-okresowego (4.49) można utożsamiać z jego dyskretnym zespolonym szeregiem Fouriera (4.34). Stąd z kolei wynika, że definicję widma $X[k]$ sygnału N-okresowego, sformułowaną w p. 4.3.3 w oparciu o jego rozwinięcie w szereg (4.34), można wprowadzić równoważnie na podstawie jego $\mathscr{F}_D^{-1}$-transformaty (4.49). Ponieważ między $\mathscr{F}_D^{-1}$-transformatą a $\mathscr{F}_D$-transformatą zachodzi wzajemnie jednoznaczny związek, zatem widmo to jest określone wzorem (4.48). Wzór ten stanowi zarazem rozwiązanie problemu odwrotnego dla sygnałów N-okresowych, o którym wspomnieliśmy w p. 4.3.4. Określa on bowiem sposób wyznaczania widma $X[k]$ sygnału N-okresowego $\bar{x}[n]$ na podstawie jego próbek $x(n)$, pobranych w jednym okresie $n = 0, \ldots, N-1$.

4.4.8. Porównanie DTF sygnału impulsowego i DTF jego przedłużenia okresowego

Z formalnego punktu widzenia DTF sygnału impulsowego $x[n]$ określonego w chwilach $n = 0, \ldots, N-1$ i DTF jego przedłużenia okresowego $\bar{x}[n]$ (4.47) są zdefiniowane tym samym wyrażeniem. Zwróćmy jednak uwagę na zasadniczą różnicę w interpretacji tych transformat. Widmo sygnału impulsowego $x[n]$ jest określone wzorem (4.39) i jest funkcją okresową $X(\mathrm{e}^{\mathrm{j}\theta})$ zmiennej ciągłej θ. DTF (4.45) tego sygnału określa jedynie próbki widma $X(\mathrm{e}^{\mathrm{j}\theta})$ dla dyskretnych wartości $\theta_k = 2\pi k/N$ tej zmiennej. DTF sygnału impulsowego $x[n]$ stanowi więc dyskretną N-punktową reprezentację jego ciągłego widma $X(\mathrm{e}^{\mathrm{j}\theta})$. Natomiast DTF (4.48) przedłużenia okresowego $\bar{x}[n]$ sygnału $x[n]$ określa dokładne widmo sygnału okresowego $\bar{x}[n]$, które z natury rzeczy jest dyskretne. Dyskretny charakter tego widma jest konsekwencją ogólnej zasady obowiązującej w teorii sygnałów, zgodnie z którą cecha dyskretności w jednej dziedzinie jest nierozerwalnie związana z cechą okresowości w drugiej dziedzinie. Sygnały dyskretne mają widma okresowe. Widma dyskretne są widmami sygnałów okresowych. A ponieważ sygnał $\bar{x}[n]$ jest dyskretny i okresowy, jego widmo $X[k]$ określone wzorem (4.48) jest okresowe i zarazem dyskretne.

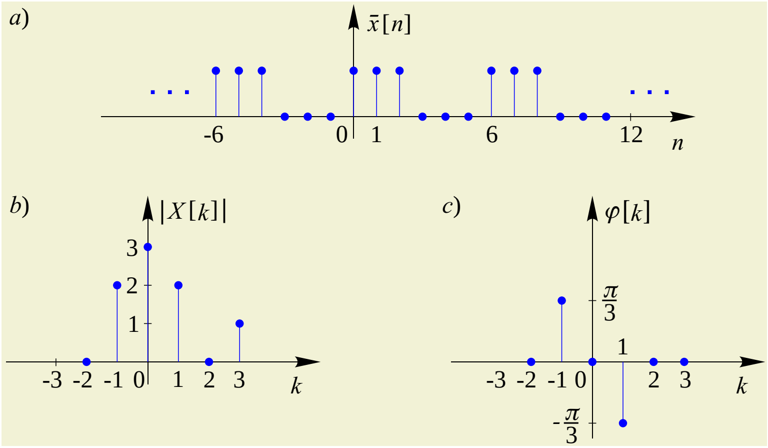
Przykład 4.7. Dyskretny sygnał okresowy $\bar{x}[n]$ przedstawiony na rys. 4.6a jest określony w jednym okresie $N = 6$ próbkami: $x(0) = x(1) = x(2) = 1$, $x(3) = x(4) = x(5) = 0$. Zgodnie ze wzorem (4.48) obliczamy:

$$
\begin{aligned}
&X(0) = 3,\\
&X(1) = 1 + 1 \cdot \mathrm{e}^{-\mathrm{j}\pi/3} + 1 \cdot \mathrm{e}^{-\mathrm{j}2\pi/3} = 1 - \mathrm{j}\sqrt{3},\\
&X(2) = 1 + 1 \cdot \mathrm{e}^{-\mathrm{j}2\pi/3} + 1 \cdot \mathrm{e}^{-\mathrm{j}4\pi/3} = 0,\\
&X(3) = 1 + 1 \cdot \mathrm{e}^{-\mathrm{j}\pi} + 1 \cdot \mathrm{e}^{-\mathrm{j}2\pi} = 1,\\
&X(4) = 1 + 1 \cdot \mathrm{e}^{-\mathrm{j}4\pi/3} + 1 \cdot \mathrm{e}^{-\mathrm{j}4\pi/3} = 0,\\
&X(5) = 1 + 1 \cdot \mathrm{e}^{-\mathrm{j}5\pi/3} + 1 \cdot \mathrm{e}^{-\mathrm{j}5\pi/3} = 1 + \mathrm{j}\sqrt{3}.
\end{aligned}
$$

Widmo amplitudowe i widmo fazowe tego sygnału wykreślono w okresie $-3 < k \leqslant 3$ na rys. 4.6b i odpowiednio 4.6c.

4.5. Właściwości i twierdzenia DPF

Oprócz właściwości N-okresowości dyskretnej transformaty Fouriera, dyskretne przekształcenie Fouriera ma szereg innych interesujących właściwości.



Rys. 4.6. Dyskretny sygnał okresowy z przykładu 4.7 (a) oraz jego widmo amplitudowe (b) i fazowe (c)

Ważniejsze z nich przytoczymy poniżej. Będziemy zakładać, że $x[n] \underset{N}{\leftrightarrow} X[k]$ jest parą dyskretnych transformat Fouriera.

1. *Liniowość.* DPF jest przekształceniem liniowym, tzn. przekształcenie kombinacji liniowej sygnałów jest taką samą kombinacją liniową ich widm dyskretnych.
2. *Wartość w punkcie* $k = 0$. Próbka $X(0)$ widma jest równa sumie próbek sygnału w przedziale $n = 0, \ldots, N-1$:

$$X(0) = \sum_{n=0}^{N-1} x(n). \tag{4.50}$$

Jeśli sygnał jest rzeczywisty, próbka $X(0)$ jest rzeczywista.
3. *Wartość w punkcie* $k = N/2$. Jeśli sygnał jest rzeczywisty oraz N jest liczbą parzystą, to próbka $X(N/2)$ widma jest liczbą rzeczywistą:

$$X(N/2) = \sum_{n=0}^{N-1} (-1)^n x(n). \tag{4.51}$$

4. *Właściwość symetrii.* Jeśli sygnał $x[n]$ jest rzeczywisty i N jest liczbą parzystą, to próbki widma położone symetrycznie względem $N/2$ są parami sprzężone. Właściwość tę można łatwo wykazać, obliczając:

$$\begin{aligned} X(N-k) &= \textstyle\sum_{n=0}^{N-1} x(n)\, \mathrm{e}^{-\mathrm{j}2\pi(N-k)n/N} = \\ &= \textstyle\sum_{n=0}^{N-1} x(n)\, \mathrm{e}^{-\mathrm{j}2\pi n}\, \mathrm{e}^{\mathrm{j}2\pi kn/N} = \\ &= \textstyle\sum_{n=0}^{N-1} x(n)\, \mathrm{e}^{\mathrm{j}2\pi kn/N} = X^*(k). \end{aligned} \tag{4.52}$$

Widmo $X[k]$ sygnałów rzeczywistych ma więc właściwość symetrii, z której wynika, że w celu jego pełnego określenia wystarczy znajomość jedynie

$N/2+1$ pierwszych próbek $X(0), \ldots, X(N/2)$ (dwóch liczb rzeczywistych $x(0)$ i $x(N/2)$ oraz $N/2-1$ liczb zespolonych $X(1), \ldots, X(N/2-1)$). Próbki $X(N/2+1), \ldots, X(N-1)$ nie wnoszą dodatkowej informacji o widmie. Ponieważ próbki te są równe próbkom $X(-N/2+1), \ldots, X(-1)$ przesuniętym o okres N w lewo, można przyjąć, że odpowiadają one pulsacjom ujemnym. Właściwość ta jest identyczna jak w przypadku dyskretnego szeregu Fouriera (4.34) N-okresowego sygnału rzeczywistego (por. wzory (4.35) i (4.37)). Ma ona istotne znaczenie przy numerycznym obliczaniu DTF sygnałów rzeczywistych.

Z właściwości symetrii wynika bezpośrednio, że dla sygnałów rzeczywistych spełnione są dla każdego k równości:

$$|X(k)| = |X(-k)|, \qquad \arg X(k) = \arg X(-k). \tag{4.53}$$

Oznacza to, że dyskretne widmo amplitudowe jest funkcją parzystą, a dyskretne widmo fazowe – funkcją nieparzystą zmiennej k.

Twierdzenie 4.8 (*o przesunięciu w dziedzinie czasu*)

$$x[n-m] \underset{N}{\leftrightarrow} X[k]W^{-km}. \tag{4.54}$$

W wyniku przesunięcia sygnału o czas m jego dyskretne widmo amplitudowe nie ulega zmianie, natomiast wartości widma fazowego zmieniają się o $-2\pi km/N$.

Twierdzenie 4.9 (*o przesunięciu w dziedzinie częstotliwości – o modulacji*)

$$x[n]W^{nm} \underset{N}{\leftrightarrow} X[k-m]. \tag{4.55}$$

W wyniku mnożenia sygnału przez dyskretny zespolony sygnał harmoniczny o pulsacji unormowanej $2\pi m/N$ widmo dyskretne tego sygnału ulega przesunięciu o wartość m.

Twierdzenie 4.10 (*twierdzenie Parsevala – o energii*)

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2. \tag{4.56}$$

Zgodnie z tym twierdzeniem energia $E_x(N)$ sygnału $x[n]$ zawarta w przedziale $[0, N-1]$ może być obliczona jako średnia arytmetyczna za ten przedział kwadratów modułów próbek widma tego sygnału.

Przykład 4.8. Sprawdzimy, że sygnał z rys. 4.6a rozpatrywany w przykładzie 4.7 spełnia właściwości 2, 3 i 4. Istotnie:

$$\sum_{n=0}^{N-1} x(n) = 1 + 1 + 1 = 3 = X(0),$$

$$\sum_{n=0}^{N-1} (-1)^n x(n) = 1 \cdot 1 + (-1) \cdot 1 + 1 \cdot 1 = 1 = X(3),$$

a ponadto zachodzi właściwość symetrii (4.52), gdyż $X(4) = X^*(2)$ oraz $X(5) = X^*(1)$. Uwzględniając te właściwości, sprawdzimy także słuszność twierdzenia Parsevala. Z jednej strony:

$$E_x(6) = \sum_{n=0}^{6} |x(n)|^2 = 3,$$

z drugiej zaś:

$$E_x(6) = \frac{1}{6}\sum_{k=0}^{6} |X(k)|^2 = \frac{1}{6}[X^2(0) + 2|X(1)|^2 + 2|X(2)|^2 + X^2(3)] =$$
$$= \frac{1}{6}[3^2 + 2|1 - \mathrm{j}\sqrt{3}|^2 + 1] = \frac{18}{6} = 3.$$

Przykład 4.9. Harmoniczny sygnał analogowy $x(t) = X_0 \cos 2\pi f_0 t$ o częstotliwości $f_0 = 1\,\text{kHz}$ jest próbkowany z częstotliwością $f_s = 6\,\text{kHz}$. W wyniku próbkowania otrzymujemy dyskretny sygnał harmoniczny $\bar{x}(n) = X_0 \cos n\theta_0$ o pulsacji unormowanej $\theta_0 = \omega_0 T_s = 2\pi f_0/f_s = \pi/3$. Jeden jego okres obejmuje $N = 6$ próbek. Wyznaczymy widmo sygnału $\bar{x}[n]$. Zgodnie z właściwością symetrii (4.52), wystarczy obliczyć próbki widma $X(0), X(1), X(2)$ oraz $X(3)$. Podstawiając próbki sygnału $\bar{x}(n)$ do wzoru (4.48), otrzymujemy:

dla $k = 0$:

$$X(0) = X_0 \sum_{n=0}^{5} \cos(n\pi/3) =$$
$$= X_0[\cos 0 + \cos(\pi/3) + \cos(2\pi/3) + \cos\pi + \cos(4\pi/3) + \cos(5\pi/3)] =$$
$$= X_0\left[1 + \frac{1}{2} - \frac{1}{2} - 1 - \frac{1}{2} + \frac{1}{2}\right] = 0,$$

dla $k = 1$:

$$X(1) = X_0 \sum_{n=0}^{5} \cos(n\pi/3)\,\mathrm{e}^{\mathrm{j}n\pi/3} = X_0[1 + \cos(\pi/3)\,\mathrm{e}^{\mathrm{j}\pi/3} +$$
$$\cos(2\pi/3)\,\mathrm{e}^{\mathrm{j}2\pi/3} + \cos\pi\,\mathrm{e}^{\mathrm{j}\pi} + + \cos(4\pi/3)\,\mathrm{e}^{\mathrm{j}4\pi/3} + \cos(5\pi/3)\,\mathrm{e}^{\mathrm{j}5\pi/3}] =$$
$$= X_0\left[1 + \frac{1}{2}\left(\frac{1}{2} + \mathrm{j}\frac{\sqrt{3}}{2}\right) - \frac{1}{2}\left(-\frac{1}{2} + \mathrm{j}\frac{\sqrt{3}}{2}\right) + (-1)\cdot(-1) -\right.$$
$$\left. -\frac{1}{2}\left(-\frac{1}{2} - \mathrm{j}\frac{\sqrt{3}}{2}\right) + \frac{1}{2}\left(\frac{1}{2} - \mathrm{j}\frac{\sqrt{3}}{2}\right)\right] = 3X_0 = X^*(5),$$

dla $k = 2$:

$$X(2) = X_0 \sum_{n=0}^{5} \cos(\pi n/3)\, \mathrm{e}^{\mathrm{j}2\pi n/3} = X_0[1 + \cos(\pi/3)\, \mathrm{e}^{\mathrm{j}2\pi/3} +$$

$$+ \cos(2\pi/3)\, \mathrm{e}^{\mathrm{j}4\pi/3} + \cos\pi\, \mathrm{e}^{\mathrm{j}2\pi} + \cos(4\pi/3)\, \mathrm{e}^{\mathrm{j}8\pi/3} + \cos(5\pi/3)\, \mathrm{e}^{\mathrm{j}10\pi/3}] =$$

$$X_0 \left[1 + \frac{1}{2}\left(-\frac{1}{2} + \mathrm{j}\,\frac{\sqrt{3}}{2}\right) - \frac{1}{2}\left(-\frac{1}{2} - \mathrm{j}\,\frac{\sqrt{3}}{2}\right) + (-1)\cdot(+1) - \right.$$

$$\left. -\frac{1}{2}\left(-\frac{1}{2} + \mathrm{j}\,\frac{\sqrt{3}}{2}\right) + \frac{1}{2}\left(-\frac{1}{2} - \mathrm{j}\,\frac{\sqrt{3}}{2}\right)\right] = 0 = X^*(4),$$

dla $k = 3$:

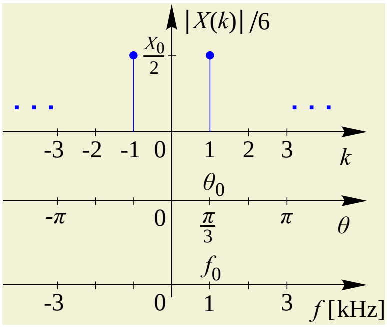
$$X(3) = X_0 \sum_{n=0}^{5} (-1)^n \cos(n\pi/3) = X_0[\cos 0 - \cos(\pi/3) + \cos(2\pi/3) - \cos\pi +$$

$$+ \cos(4\pi/3) - \cos(5\pi/3)] = X_0\left[1 - \frac{1}{2} - \frac{1}{2} + 1 - \frac{1}{2} - \frac{1}{2}\right] = 0.$$

Jak można było spodziewać się, tylko dwie próbki widma w jednym jego okresie w punktach $k = 1$ i $k = 5$ są różne od zera. Wartość $k = 1$ odpowiada na skali pulsacji unormowanej wartości $\theta = \pi/3$, równej pulsacji sygnału harmonicznego $\theta_0 = \pi/3$. Próbka widma w punkcie $k = 5$ powtarza się w punkcie $k = -1$ przesuniętym o okres $N = 6$ w lewo, odpowiadającym pulsacji unormowanej $\theta = -\pi/3$. W centralnym okresie $-N/2 < k \leqslant N/2$ $(-\pi < \theta \leqslant \pi)$ widmo $X[k]$ dyskretnego sygnału harmonicznego jest zatem reprezentowane parą próbek (prążków) występujących w punktach $k = \pm 1$ $(\theta = \pm\theta_0)$. Występuje tu pełna analogia do widma analogowego sygnału harmonicznego.

Zwróćmy uwagę, że próbki widma można było obliczyć w prostszy sposób, korzystając ze wzoru Eulera $\cos n\pi/3 = (\mathrm{e}^{\mathrm{j}n\pi/3} + \mathrm{e}^{-\mathrm{j}n\pi/3})/2$ i z właściwości ortonormalności sygnałów $\mathrm{e}^{\mathrm{j}\pi nk/3}$ dla $k = 0, \ldots, N-1$ w przestrzeni l_N^2 N-okresowych ciągów liczbowych. Celowo przeprowadziliśmy tu bezpośrednie obliczenia, aby zademonstrować złożoność obliczeniową DTF.

Z uwagi na parzystość sygnału widmo $X[k]$ jest w tym przypadku rzeczywiste. Widmo fazowe przybiera wartości zerowe. Zwróćmy jeszcze uwagę na wartość próbek $X(1)$ i $X(-1)$. Wynosi ona $3X_0$. Jeżeli pomnożymy ją przez czynnik skalujący $1/N$ we wzorze (4.49), otrzymamy wartość $X_0/2$, identyczną jak wartość prążków widma analogowego sygnału harmonicznego (rozpatrywanego jako funkcja częstotliwości f). W ogólnym przypadku próbki $|X(k)|$ widma amplitudowego pochodzące od dyskretnego sygnału harmonicznego o amplitudzie X_0 są równe $NX_0/2$. Widmo $X(k)$ rozpatrywanego w przykładzie dyskretnego sygnału harmonicznego pokazano na rys. 4.7. Widmo to zostało unormowane względem liczby $N = 6$ i wykreślone w centralnym okresie $-3 < k \leqslant 3$. W celu ułatwienia interpretacji otrzymanych wyników na wykresie zamieszczono dodatkowo skalę unormowanej pulsacji θ i nieunormowanej częstotliwości f.



Rys. 4.7. Widmo dyskretnego sygnału harmonicznego z przykładu 4.9

4.6. Odtwarzanie sygnału dyskretnego o nieskończonym czasie trwania na podstawie próbek jego widma

4.6.1. Sformułowanie i rozwiązanie problemu

Jeżeli sygnał dyskretny jest sygnałem impulsowym $x[n]$ o czasie trwania N lub sygnałem N-okresowym $\bar{x}[n]$, to znając N próbek $X(k)$ jego widma (jego N-punktową DTF), możemy go na tej podstawie odtworzyć w sposób jednoznaczny. Nie jest to natomiast możliwe w przypadku sygnału o nieskończonym czasie trwania, jak również w przypadku, gdy sygnał impulsowy o czasie trwania N_0 jest obserwowany w oknie czasowym o długości $N < N_0$.

Rozważmy sygnał $x[n]$ o nieskończonym, w ogólnym przypadku, czasie trwania i ograniczonej energii. Widmo tego sygnału

$$X(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{n=-\infty}^{\infty} x(n)\,\mathrm{e}^{-\mathrm{j}n\theta} \tag{4.57}$$

jest funkcją okresową zmiennej ciągłej θ. Załóżmy, że nie znamy sygnału $x(n)$, znamy natomiast N próbek jego widma pobranych w jednym okresie w punktach

$\theta_k = 2\pi k/N$:

$$X(k) = \sum_{n=-\infty}^{\infty} x(n)\,\mathrm{e}^{-\mathrm{j}2\pi kn/N} = \sum_{n=-\infty}^{\infty} x(n)W^{-kn}, \qquad k = 0, \ldots, N-1. \tag{4.58}$$

Na podstawie tych próbek chcemy odtworzyć sygnał. W tym celu wyznaczymy $\mathscr{F}_D^{-1}$-transformatę ciągu $X(k)$. W wyniku otrzymamy pewien sygnał okresowy:

$$\bar{x}[n] = \mathscr{F}^{-1}[X(k)] = \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{m=-\infty}^{\infty} x(m)W^{-km}\right)W^{kn}. \tag{4.59}$$

Z uwagi na N-okresowość funkcji $W^{-k(m+iN)} = W^{-km}W^{-kiN} = W^{km}$, próbki $x(m+iN)$ w sumie wewnętrznej, odległe od siebie o N, są mnożone przez tę samą funkcję W^{-km}. Podobnie jak we wzorze (4.31), sumę wewnętrzną można zatem rozbić na N sum:

$$\sum_{m=0}^{N-1}\left(\sum_{i=-\infty}^{\infty} x(m+iN)W^{-km}\right).$$

Po podstawieniu do wzoru (4.59) otrzymujemy:

$$\bar{x}[n] = \frac{1}{N}\sum_{k=0}^{N-1}\left[\sum_{m=0}^{N-1}\left(\sum_{i=-\infty}^{\infty} x(m+iN)W^{-km}W^{kn}\right)\right].$$

Z ortonormalności funkcji W^{-km} dla $m = 0, \ldots, N-1$ wynika, że dla każdego $k = 0, \ldots, N-1$ suma w nawiasie kwadratowym jest równa $\sum_{i=-\infty}^{\infty} x(n+iN)$, a ponieważ sum takich jest N, otrzymujemy ostatecznie:

$$\bar{x}[n] = \sum_{i=-\infty}^{\infty} x(n+iN). \tag{4.60}$$

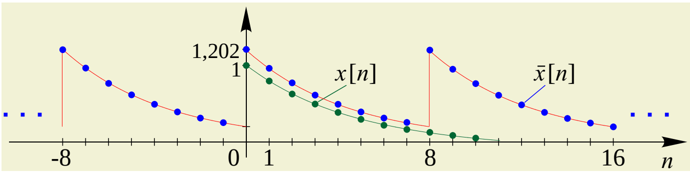
Sygnał odtworzony z N próbek widma nieznanego sygnału $x[n]$ jest więc nieskończoną sumą nałożonych na siebie kopii tego sygnału poprzesuwanych o kolejne krotności okresu N. Kolejne, powtarzające się okresowo próbki sygnału odtworzonego są superpozycją próbek pochodzących od wszystkich przesuniętych kopii. Wynika stąd, że jeśli czas trwania N_0 sygnału $x[n]$ jest większy od N (w szczególności nieskończony), nie jest możliwe dokładne jego odtworzenie na podstawie N próbek widma.

4.6.2. Błąd aliasingu

Efekt nakładania się kolejnych kopii sygnału, nazywany *aliasingiem w dziedzinie czasu*, został zilustrowany na rys. 4.8. Odtwarzając z N próbek widma sygnał o czasie trwania dłuższym niż N, popełniamy błąd

$$\varepsilon(n) = x(n) - \sum_{i=-\infty}^{\infty} x(n + iN), \tag{4.61}$$

nazywany *błędem aliasingu*. W przypadku sygnału $x[n]$ o nieskończonym czasie trwania błąd ten jest tym mniejszy im większa jest liczba N oraz im szybciej wartości próbek sygnału maleją do zera, gdy $n \to \pm\infty$. W przypadku sygnału o skończonym czasie trwania błąd aliasingu jest tym mniejszy, im mniejsza jest różnica $N_0 - N$.



Rys. 4.8. Ilustracja aliasingu w dziedzinie czasu

Słownik

aliasing (w dziedzinie czasu)
: efekt nakładania się powielonych okresowo kopii sygnału

dyskretna transformata Fouriera DTF
: N-okresowy zbiór próbek widma sygnału otrzymany w wyniku dyskretnego przekształcenia Fouriera sygnału

dyskretne przekształcenie Fouriera DPF
: przekształcenie przyporządkowujące sygnałowi dyskretnemu określonemu w N chwilach N próbek jego widma

metoda uzupełniania zerami
: metoda zwiększania rozdzielczości DTF przez dodanie do próbek sygnału próbek zerowych

okienkowanie
: operacja mnożenia sygnału przez okno czasowe

okno czasowe
: dyskretna funkcja czasu o skończonym czasie trwania, przez którą mnoży

się sygnał dyskretny w celu zmniejszenia zniekształceń jego widma obliczanego za pomocą DPF

próbka widma
wartość widma sygnału w określonym punkcie osi częstotliwości

przeciek DTF
efekt występowania w dyskretnym widmie obliczonym za pomocą DPF próbek widmowych w fałszywych punktach osi częstotliwości unormowanej; można go interpretować jako przeciekanie energii (lub mocy) sygnału zawartej w jego faktycznych składowych częstotliwościowych do składowych o częstotliwościach fałszywych

rozdzielczość dyskretnej transformaty Fouriera
odwrotność odległości między kolejnymi próbkami widma sygnału obliczonego za pomocą DTF

rozróżnialność dyskretnej transformaty Fouriera
zdolność do odróżnienia w sygnale dyskretnym dwóch składowych harmonicznych na podstawie jego DTF

sygnał N-okresowy
sygnał dyskretny, którego wartości powtarzają się okresowo co N próbek

szybkie przekształcenie Fouriera SPF
numeryczna metoda obliczania dyskretnego widma sygnału o zredukowanym nakładzie obliczeniowym (?!)

widmo dyskretne
dyskretna transformata Fouriera DTF

Literatura

[1] Wojtkiewicz A.: Elementy syntezy filtrów cyfrowych. WNT, Warszawa, 1984.

[2] Lyons G.R.: Wprowadzenie do cyfrowego przetwarzania sygnałów. WKŁ, Warszawa, 1999.

[3] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

[4] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom II. WNT, Warszawa, wyd. 3, 1998.

Lekcja 5

Analiza korelacyjna sygnałów

Przedmiotem rozważań lekcji 5 jest analiza korelacyjna sygnałów. Opis korelacyjny stanowi jeszcze jeden sposób charakteryzowania sygnałów, ściśle związany z opisem widmowym. W lekcji zdefiniujemy pojęcie funkcji autokorelacji sygnału, zarówno dla sygnałów analogowych, jak i dyskretnych. Omówimy także ich najważniejsze właściwości. Z uwagi na różnice w definicjach formalnych funkcji autokorelacji, pojęcie to wprowadzimy osobno dla klasy sygnałów o ograniczonej energii oraz klasy sygnałów o ograniczonej mocy. Przedyskutujemy związek między funkcją autokorelacji a widmem energii lub, odpowiednio, widmem mocy sygnału. Podamy także definicje funkcji korelacji wzajemnych oraz widm wzajemnych energii (lub, odpowiednio, mocy) i omówimy związki między tymi charakterystykami dla poszczególnych klas sygnałów.

5.1. Funkcja autokorelacji sygnału analogowego o ograniczonej energii

W wielu praktycznych zastosowaniach teorii sygnałów zachodzi konieczność porównywania analizowanego sygnału z innym sygnałem, w szczególności – ze swoją własną, przesuniętą w czasie kopią. Porównania tego należy przy tym dokonać dla różnych wzajemnych położeń obu sygnałów na osi czasu, tj. różnych wartości opóźnienia jednego sygnału względem drugiego. Obiektywną i jednoznaczną miarę podobieństwa dwóch sygnałów możemy wprowadzić, rozpatrując je jako elementy odpowiedniej przestrzeni Hilberta, w której – jak pamiętamy (por. p. 2.2.7) – jest określony iloczyn skalarny. Na podstawie iloczynu skalarnego możemy wyznaczyć zarówno odległość sygnałów w danej przestrzeni, jak i kąt między nimi.

Określenie miary podobieństwa sygnałów dla różnych wartości przesunięcia w czasie w oparciu o ich iloczyn skalarny prowadzi do pojęcia *funkcji korelacyjnych*. Opis korelacyjny sygnałów stanowi jeszcze jeden alternatywny sposób charakteryzowania sygnałów w dziedzinie czasu.

5.1.1. Definicja

Rozważmy sygnał $x(t)$, w ogólnym przypadku zespolony, należący do przestrzeni Hilberta L^2 sygnałów o ograniczonej energii, oraz jego kopię $x_\tau(t) \triangleq x(t-\tau)$ przesuniętą w czasie o wartość $\tau \in \mathscr{R}$. Sygnał przesunięty jest również elementem przestrzeni L^2, gdyż operacja przesunięcia nie wyprowadza sygnału poza daną przestrzeń. Za miarę podobieństwa sygnału i jego przesuniętej kopii przyjmiemy ich iloczyn skalarny w przestrzeni L^2 (por. wzór (2.17)):

$$(x, x_\tau) = \int\limits_{-\infty}^{\infty} x(t)x^*(t-\tau)\,\mathrm{d}t. \tag{5.1}$$

Dla ustalonego przesunięcia τ miara (5.1) jest w ogólnym przypadku liczbą zespoloną. Dla różnych przesunięć τ wartości tej miary są różne. W konsekwencji otrzymujemy zależność funkcyjną iloczynu skalarnego (5.1) od zmiennej τ. Na podstawie tej zależności możemy porównywać sygnał $x(t)$ z jego kopiami dla różnych wartości przesunięć.

Definicja 5.1. *Funkcją autokorelacji* $\varphi_x(\tau)$ sygnału $x(t)$ o ograniczonej energii nazywamy zależność iloczynu skalarnego (5.1) od przesunięcia τ:

$$\varphi_x(\tau) = \int\limits_{-\infty}^{\infty} x(t)x^*(t-\tau)\,\mathrm{d}t. \tag{5.2}$$

Funkcja autokorelacji sygnału zespolonego jest zespolona. W przypadku sygnałów rzeczywistych funkcja autokorelacji przybiera wartości rzeczywiste.

Komentarz. Termin *korelacja* kojarzy się zwykle z korelacją w znaczeniu statystycznym i jest odnoszony do związku statystycznego, jaki występuje między dwoma wielkościami (zmiennymi) losowymi. W rachunku prawdopodobieństwa korelację dwóch zmiennych losowych definiuje się jako wartość oczekiwaną ich iloczynu. Ma więc ona znaczenie wielkości średniej w zbiorze (uśrednionej po wszystkich możliwych realizacjach zmiennych losowych). W przypadku sygnałów deterministycznych korelacja ma natomiast znaczenie wielkości uśrednionej w czasie. Należy jednak podkreślić, że geneza pojęcia „korelacja" jest w obu przypadkach identyczna. Z formalnego punktu widzenia korelacja zmiennych losowych jest bowiem niczym innym, jak iloczynem skalarnym określonym w przestrzeni Hilberta tych zmiennych (por. [1], p. 4.1). Oba pojęcia mają więc taki sam sens formalny iloczynów skalarnych określonych w odpowiednich przestrzeniach Hilberta. Stosowanie terminu „korelacja" w odniesieniu do sygnałów deterministycznych jest zatem w pełni uzasadnione.

5.1.2. Właściwości

Opierając się na definicji (5.2), można łatwo wykazać następujące właściwości funkcji autokorelacji sygnałów o ograniczonej energii.

Właściwość 1. Funkcja autokorelacji $\varphi_x(\tau)$ jest *funkcją hermitowską*:

$$\varphi_x(\tau) = \varphi_x^*(-\tau), \tag{5.3}$$

a więc jej część rzeczywista jest funkcją parzystą zmiennej τ: $\operatorname{Re}\varphi_x(\tau) = \operatorname{Re}\varphi_x(-\tau)$, a część urojona – funkcją nieparzystą tej zmiennej: $\operatorname{Im}\varphi_x(\tau) = -\operatorname{Im}\varphi_x(-\tau)$. W przypadku sygnałów rzeczywistych funkcja autokorelacji jest rzeczywista i parzysta: $\varphi_x(\tau) = \varphi_x(-\tau)$.

Właściwość 2. Wartość funkcji autokorelacji $\varphi_x(\tau)$ w punkcie $\tau = 0$ jest rzeczywista i równa energii sygnału $x(t)$:

$$\varphi_x(0) = \int\limits_{-\infty}^{\infty} |x(t)|^2 \, \mathrm{d}t = E_x. \tag{5.4}$$

Właściwość ta wynika wprost ze wzoru definicyjnego (5.2) przy podstawieniu w nim $\tau = 0$.

Właściwość 3. Wartości funkcji autokorelacji $\varphi_x(\tau)$ nie przekraczają co do modułu jej wartości w punkcie $\tau = 0$, tzn. dla każdego τ spełniona jest nierówność:

$$|\varphi_x(\tau)| \leqslant \varphi_x(0). \tag{5.5}$$

Funkcja autokorelacji ma więc dla $\tau = 0$ zawsze dodatnie maksimum. Właściwość ta wynika z nierówności Buniakowskiego-Schwarza dla iloczynu skalarnego (por. p. 2.2.7).

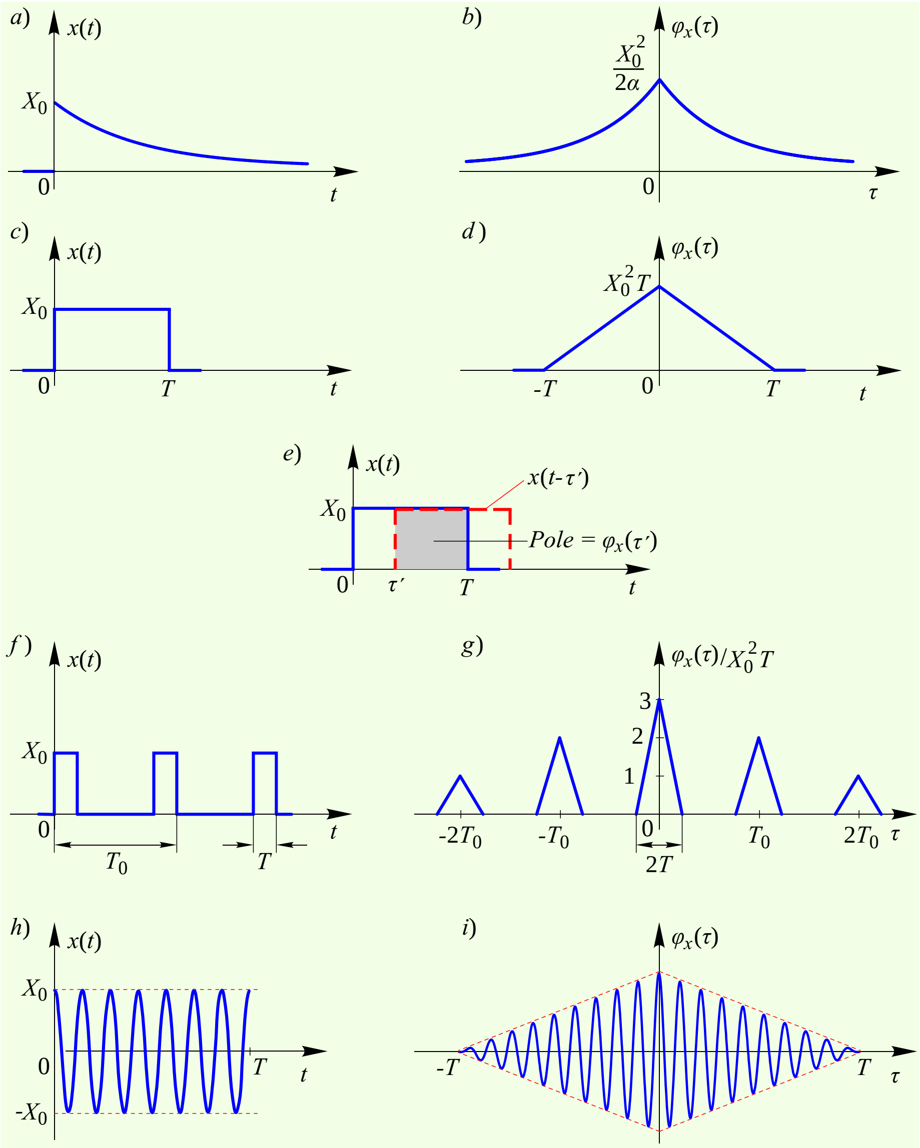
Właściwość 4. Jeśli $\varphi_x(\tau_0) = 0$ dla pewnego przesunięcia τ_0, to sygnały $x(t)$ i $x(t - \tau_0)$ są ortogonalne. Właściwość ta jest konsekwencją określenia funkcji autokorelacji jako iloczynu skalarnego sygnałów.

Właściwość 5. Funkcja autokorelacji jest niezmiennicza względem położenia sygnału na osi czasu, tzn. funkcja autokorelacji sygnału $x(t)$ i funkcja autokorelacji sygnału przesuniętego $x(t-t_0)$ są identyczne dla każdej wartości przesunięcia t_0. Właściwość tę można wykazać, dokonując odpowiedniej zamiany zmiennych we wzorze (5.2).

Właściwość 6. Funkcja autokorelacji sygnału o ograniczonej energii jest także funkcją o ograniczonej energii, a więc $\mathscr{F}$-transformowalną w zwykłym sensie. W p. 5.1.4 odpowiemy na pytanie czemu jest równa $\mathscr{F}$-transformata (widmo) funkcji autokorelacji.

Rozpatrzymy obecnie kilka przykładów wyznaczania funkcji autokorelacji sygnałów o ograniczonej energii.

Przykład 5.1. Wyznaczymy funkcję autokorelacji sygnału wykładniczego malejącego $x(t) = X_0\,\mathrm{e}^{-\alpha t}\boldsymbol{1}(t)$ (rys. 5.1a). Zgodnie ze wzorem (5.2) sygnał $x(t)$ należy w tym celu przesuwać wzdłuż osi czasu i obliczać pola pod iloczynem sygnału i jego przesuniętej kopii dla różnych wartości przesunięcia τ. Dla $\tau \geqslant 0$



Rys. 5.1. Przykłady funkcji autokorelacji sygnałów o ograniczonej energii

mamy:

$$\begin{aligned}\varphi_x(\tau) &= X_0^2 \int\limits_{-\infty}^{\infty} \mathrm{e}^{-\alpha t}\, \mathbf{1}(t)\, \mathrm{e}^{-\alpha(t-\tau)}\, \mathbf{1}(t-\tau)\, \mathrm{d}t = \\ &= X_0^2\, \mathrm{e}^{\alpha\tau} \int\limits_{\tau}^{\infty} \mathrm{e}^{-2\alpha t}\, \mathrm{d}\tau = \frac{X_0^2}{2\alpha}\, \mathrm{e}^{-\alpha\tau}\,.\end{aligned}$$

Dla $\tau < 0$ nie musimy powtarzać obliczeń, wystarczy bowiem uwzględnić właściwość parzystości funkcji autokorelacji sygnałów rzeczywistych. Ostatecznie otrzymujemy (rys. 5.1b):

$$\varphi_x(\tau) = \frac{X_0^2}{2\alpha}\, \mathrm{e}^{-\alpha|\tau|}\,. \tag{5.6}$$

Podstawiając w tym wzorze $\tau = 0$, otrzymujemy, zgodnie z właściwością (5.4), energię sygnału $E_x = \varphi_x(0) = X_0^2/2\alpha$ (por. przykład w p. 1.2.3).

Przykład 5.2. Funkcja autokorelacji impulsu prostokątnego $x(t) = X_0 \sqcap [(t - T/2)/T]$ (rys. 5.1c) ma kształt trójkątny $\varphi_x(\tau) = X_0^2 T \wedge [(\tau/T)]$ (rys. 5.1d). Graficzny sposób jej wyznaczania został zilustrowany na rys. 5.1e. Zauważmy, że odcinek czasu, w którym funkcja autokorelacji przybiera wartości różne od zera, jest dwa razy dłuższy, niż czas T trwania impulsu. Reguła ta jest słuszna dla funkcji autokorelacji dowolnego sygnału impulsowego. Dopóki $|\tau| < T$, dopóty przesunięta kopia pozostaje w oknie czasowym sygnału i oba sygnały korelują się. Dla $|\tau| \geqslant T$ kopia wychodzi poza granice okna i sygnały przestają być skorelowane.

Przykład 5.3. Przeprowadzając podobną konstrukcję graficzną, możemy wyznaczyć funkcję autokorelacji ciągu (paczki) impulsów prostokątnych (rys. 5.1f). Funkcja autokorelacji dla przypadku paczki trzech impulsów prostokątnych jest pokazana na rys. 5.1g.

Przykład 5.4. Wyznaczenie funkcji autokorelacji prostokątnego impulsu radiowego $x(t) = X_0 \cos\omega_0 t \sqcap (t/\tau)$ (rys. 5.1h) jest bardziej skomplikowane. Dla

$|\tau| \geqslant T$ funkcja autokorelacji jest równa zeru. Dla $0 < \tau < T$ mamy:

$$\begin{aligned}
\varphi_x(\tau) &= X_0^2 \int\limits_{-T/2+\tau}^{T/2} \cos\omega_0 t \cos\omega_0(t-\tau)\,\mathrm{d}t = \\
&= \frac{X_0^2}{2}\left[\int\limits_{-T/2+\tau}^{T/2} \cos\omega_0\tau\,\mathrm{d}t + \int\limits_{-T/2+\tau}^{T/2} \cos\omega_0(2t-\tau)\,\mathrm{d}t\right] = \\
&= \frac{X_0^2}{2}\left[(T-\tau)\cos\omega_0\tau + \frac{1}{2\omega_0}\sin\omega_0 t'\,\Big|_{-(T-\tau)}^{T-\tau}\right] = \\
&= \frac{X_0^2}{2}(T-\tau)\left[\cos\omega_0\tau + \frac{\sin\omega_0(T-\tau)}{\omega_0(T-\tau)}\right] = \\
&= \frac{X_0^2}{2}(T-\tau)\left[\cos\omega_0\tau + \mathrm{Sa}\,\omega_0(T-\tau)\right].
\end{aligned}$$

Uwzględniając właściwość parzystości, otrzymujemy ostatecznie:

$$\varphi_x(\tau) = \frac{X_0^2}{2}(T-|\tau|)\left[\cos\omega_0\tau + \mathrm{Sa}\,\omega_0(T-|\tau|)\right]. \tag{5.7}$$

Wykres funkcji autokorelacji prostokątnego impulsu radiowego jest pokazany na rys. 5.1i. Ma ona charakter oscylacyjny, typowy dla funkcji autokorelacji sygnałów wąskopasmowych. Zwróćmy uwagę, że pulsacja oscylacji jest taka sama jak pulsacja sygnału. Energia impulsu wynosi:

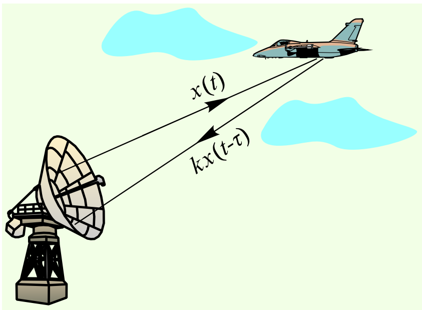
$$E_x = \varphi_x(0) = \frac{1}{2}X_0^2 T(1 + \mathrm{Sa}\,\omega_0 T).$$

Jeśli długość T impulsu jest znacznie większa od okresu oscylacji $T_0 = 2\pi/\omega_0$, tzn. $\omega_0 T \gg 1$, wówczas $E_x \approx X_0^2/2T$.

5.1.3. Przykład zastosowania w praktyce

Jednym z przykładów zastosowań, w których wykorzystuje się rezultaty porównywania sygnałów przesuniętych w czasie jest pomiar odległości od celu za pomocą radaru impulsowego (rys. 5.2).

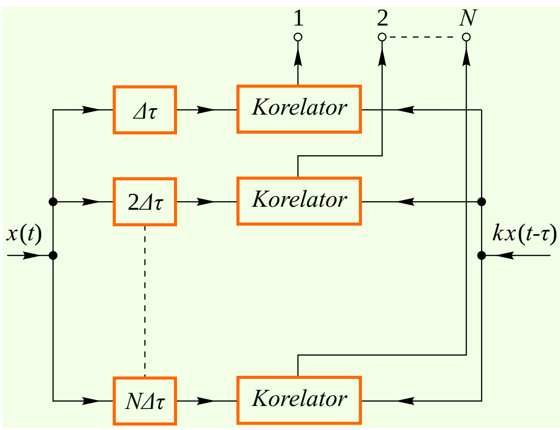
Przedstawimy krótko bardzo uproszczoną zasadę działania takiego radaru. Radar wysyła ciąg krótkich impulsów sondujących, które w najprostszym przypadku są prostokątnymi impulsami radiowymi z rys. 5.1h o bardzo dużej częstotliwości wypełnienia. Impulsy te po odbiciu się od celu powracają do anteny nadawczo-odbiorczej z pewnym opóźnieniem τ. Pomiar odległości jest dokonywany na podstawie pomiaru tego opóźnienia. Schemat ideowy układu do pomiaru odległości od celu jest przedstawiony na rys. 5.3. Zawiera on N kanałów odpowiadających tzw. *bramkom odległościowym*, na które jest podzielony zasięg radaru.



Rys. 5.2. Radar impulsowy

Rozpatrzmy pojedynczy impuls sondujący $x(t)$ i przyjmijmy idealizowane założenie, że sygnał powracający nie jest zniekształcany na drodze transmisji przez szumy i zakłócenia, a jedynie tłumiony amplitudowo. Sygnał odebrany $kx(t-\tau)$ jest wówczas opóźnioną kopią sygnału sondującego o identycznym kształcie.

Sygnał sondujący $x(t)$ jest podawany w każdym z kanałów układu odbiorczego na wejścia elementów opóźniających o wzrastających ustalonych opóźnieniach $n\Delta\tau$. Na wyjściach tych elementów występują zatem kopie sygnału sondującego opóźnione o kolejne odcinki czasu $n\Delta\tau$. Opóźnione kopie są podawane na pierwsze wejścia korelatorów. Na drugie wejścia *korelatorów* jest podawany sygnał odebrany $kx(t-\tau)$. Korelatory obliczają poszczególne funkcje autokorelacji. Funkcje te są próbkowane na wyjściach korelatorów w chwilach $t_n = n\Delta\tau$. Gdy czas trwania impulsu sondującego jest dostatecznie mały w porównaniu z opóźnieniem $\Delta\tau$, wówczas tylko w jednym z kanałów funkcja autokorelacji przybiera w chwili próbkowania wartość różną od zera. Numer kanału, w którym wystąpi niezerowa próbka funkcji autokorelacji jest więc zarazem numerem bramki odległościowej, w której znajduje się cel.



Rys. 5.3. Schemat układu do pomiaru odległości w radarze impulsowym

W rzeczywistości, na skutek szumu kanału transmisyjnego i różnego rodzaju zakłóceń, spowodowanych np. odbiciami od obiektów naziemnych, chmur, opadów atmosferycznych itp., sygnał powracający do anteny jest silnie rozmyty w czasie i zniekształcony w amplitudzie. W efekcie, niezerowe próbki pojawiają się jednocześnie w kilku sąsiadujących kanałach. W celu uniknięcia niejednoznaczności i uzyskania prawidłowego pomiaru należy zatem zapewnić, aby próbka we właściwym kanale miała wartość dominującą. Można to osiągnąć przez odpowiedni dobór impulsu sondującego. Impuls ten powinien być tak ukształtowany, aby jego funkcja autokorelacji przybierała dużą wartość maksymalną w punkcie $\tau = 0$ i szybko malała do zera w miarę oddalania się od tego punktu. Takie cechy funkcji autokorelacji zapewniają wysoką rozdzielczość radaru w odległości. Funkcja autokorelacji prostokątnego impulsu radiowego analizowanego w przykładzie 4.4 nie spełnia tych wymagań w stopniu zadowalającym. Jej obwiednia maleje do zera liniowo (rys. 5.1i), a więc stosunkowo wolno. W praktyce stosowane są impulsy majace znacznie lepsze właściwości korelacyjne zapewniające wysoką rozdzielczość radaru w odległości (np. impuls LFM z liniową modulacją częstotliwości – por. [2], p.4.3).

5.1.4. Związek funkcji autokorelacji z widmem energii

Omawiając w p. 3.3 twierdzenie Parsevala dla sygnałów o ograniczonej energii, wprowadziliśmy pojęcie widma energii $\Phi_x(\omega)$. Zostało ono określone jako kwadrat widma amplitudowego sygnału: $\Phi_x(\omega) = A_x^2(\omega) = |X(\omega)|^2$. Pokażemy, że widmo energii sygnału jest bezpośrednio związane z jego funkcją autokorelacji.

Przypomnijmy, że zgodnie z właściwością 6, funkcja autokorelacji sygnału o ograniczonej energii jest $\mathscr{F}$-transformowalna w zwykłym sensie. W uzupełnieniu twierdzeń podanych w p. 3.3 sformułujemy twierdzenie, które orzeka, że $\mathscr{F}$-transformata funkcji autokorelacji sygnału $x(t)$ jest równa jego widmu energii.

Twierdzenie 5.1 (*o funkcji autokorelacji*)

Funkcja autokorelacji $\varphi_x(\tau)$ oraz widmo energii $\Phi_x(\omega)$ sygnału $x(t)$ o ograniczonej energii tworzą parę zwykłych transformat Fouriera:

$$\Phi_x(\omega) = \int\limits_{-\infty}^{\infty} \varphi_x(\tau)\, \mathrm{e}^{-\mathrm{j}\omega\tau}\, \mathrm{d}\tau, \tag{5.8}$$

$$\varphi_x(\tau) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \Phi_x(\omega)\, \mathrm{e}^{\mathrm{j}\omega\tau}\, \mathrm{d}\omega. \tag{5.9}$$

Twierdzenie o funkcji autokorelacji można dowieść na podstawie uogólnionego twierdzenia Rayleigha (3.33) oraz twierdzenia o przesunięciu w dziedzinie czasu (3.22). Ponieważ zgodnie z tym drugim twierdzeniem $\mathscr{F}[x(t-\tau)] =$

$X(\omega)\,\mathrm{e}^{-\mathrm{j}\omega\tau}$, zatem przyjmując we wzorze (3.33) $y(t) = x(t-\tau)$, otrzymujemy:

$$\int\limits_{-\infty}^{\infty} x(t)x^*(t-\tau)\,\mathrm{d}t = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} X(\omega)X^*(\omega)\,\mathrm{e}^{\mathrm{j}\omega\tau}\,\mathrm{d}\omega = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} |X(\omega)|^2\,\mathrm{e}^{\mathrm{j}\omega\tau}\,\mathrm{d}\omega,$$

co dowodzi słuszności związków (5.8) i (5.9).

Jeżeli we wzorze (5.9) podstawimy $\tau = 0$, otrzymamy równość:

$$E_x = \varphi_x(0) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \Phi_x(\omega)\,\mathrm{d}\omega, \tag{5.10}$$

wyrażającą w nieco odmiennej formie równość Parsevala (3.34). Z równości (5.10) wynika, że energię sygnału można wyznaczyć trzema sposobami:

– w dziedzinie czasu, obliczając całkę z kwadratu modułu sygnału,
– w dziedzinie korelacyjnej, obliczając wartość funkcji autokorelacji sygnału w zerze,
– w dziedzinie częstotliwości, obliczając całkę z jego widma energii i dzieląc ją przez 2π.

Widmo energii jest nieujemną funkcją rzeczywistą zmiennej ω. Jeżeli sygnał jest rzeczywisty, jego funkcja autokorelacji jest rzeczywista i parzysta, a więc widmo energii sygnału rzeczywistego jest funkcją parzystą. Jego energię możemy wówczas obliczać ze wzoru:

$$E_x = \frac{1}{\pi}\int\limits_{0}^{\infty} \Phi_x(\omega)\,\mathrm{d}\omega. \tag{5.11}$$

Energia zawarta w skończonym przedziale pulsacji $[\omega_1, \omega_2]$ jest określona wzorem:

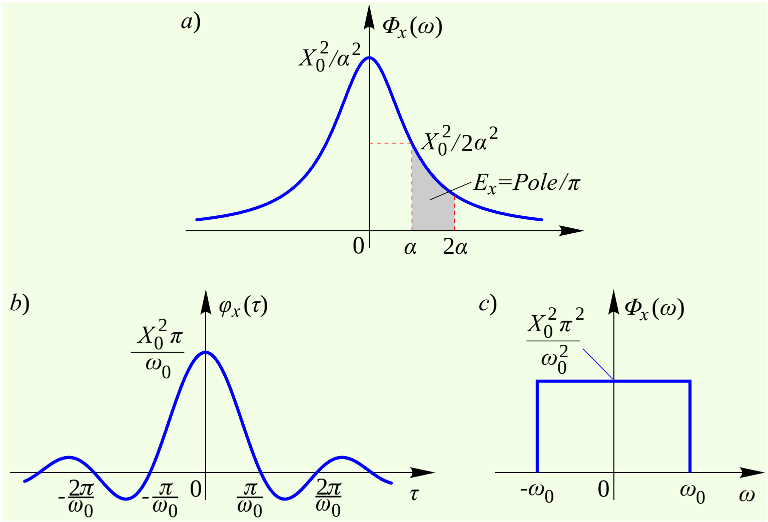
$$E_x(\omega_1, \omega_2) = \frac{1}{\pi}\int\limits_{\omega_1}^{\omega_2} \Phi_x(\omega)\,\mathrm{d}\omega. \tag{5.12}$$

Przykład 5.5. Wyznaczymy widmo energii sygnału wykładniczego malejącego $x(t) = X_0\,\mathrm{e}^{-\alpha t}\,\boldsymbol{1}(t)$ i obliczymy jego energię zawartą w przedziale pulsacji $[\alpha, 2\alpha]$. Funkcja autokorelacji tego sygnału jest określona wzorem (5.6). Zatem, zgodnie z parą transformat Fouriera (3.36), jego widmo energii:

$$\Phi_x(\omega) = \frac{X_0^2}{\alpha^2 + \omega^2}. \tag{5.13}$$

Wykres tego widma jest pokazany na rys. 5.4a. Energię zawartą w przedziale $[\alpha, 2\alpha]$ obliczamy ze wzoru (5.12):

$$E_x(\alpha, 2\alpha) = \frac{1}{\pi}\int\limits_{\alpha}^{2\alpha} \frac{X_0}{\alpha^2 + \omega^2} = \frac{X_0}{\pi\alpha}\,\mathrm{arctg}\,\frac{\omega}{\alpha}\Big|_{\alpha}^{2\alpha} \approx 0{,}2E_x \tag{5.14}$$



Rys. 5.4. Widmo energii sygnału wykładniczego malejącego (a) oraz funkcja autokorelacji (b) i widmo energii (c) idealnego sygnału dolnopasmowego

Przykład 5.6. Rozważmy idealny sygnał dolnopasmowy $x(t) = X_0 \operatorname{Sa} \omega_0 t$ (por. p. 3.4.1, wzór (3.38)). Wyznaczenie jego funkcji autokorelacji na podstawie wzoru definicyjnego (5.2) wymaga obliczenia całki:

$$\varphi_x(\tau) = X_0^2 \int_{-\infty}^{\infty} \operatorname{Sa} \omega_0 t \operatorname{Sa} \omega_0 (t - \tau)\, \mathrm{d}t.$$

Bezpośrednie obliczenie tej całki w dziedzinie czasu jest skomplikowane. Można ją natomiast bez trudu obliczyć, wyznaczając najpierw widmo energii sygnału $x(t)$:

$$\Phi_x(\omega) = |X(\omega)|^2 = \frac{X_0^2 \pi^2}{\omega_0^2} \sqcap \left(\frac{\omega}{2\omega_0} \right), \tag{5.15}$$

a następnie jego odwrotną transformatę Fouriera

$$\varphi_x(\tau) = \mathscr{F}^{-1}[\Phi_x(\omega)] = \frac{X_0^2 \pi}{\omega_0} \operatorname{Sa} \omega_0 \tau. \tag{5.16}$$

Funkcja autokorelacji idealnego sygnału dolnopasmowego ma więc również kształt funkcji Sa. Wykresy funkcji autokorelacji i widma energii tego sygnału są przedstawione na rys. 5.4b, c.

5.1.5. Ilustracja związków między opisami sygnałów w dziedzinie czasu, korelacyjnej i częstotliwości

Przekształcenie całkowe Fouriera określa dla sygnałów analogowych relację między dziedziną czasu i dziedziną częstotliwości. Relacja ta wyraża się bezpośrednio, jako związek między sygnałem i jego widmem, oraz pośrednio, jako związek między funkcją autokorelacji sygnału i jego widmem energii. Oba te związki są wzajemnie jednoznaczne, tzn. znając jedną z wielkości, możemy jednoznacznie wyznaczyć drugą.

Z twierdzenia o funkcji autokorelacji wynika, że stanowi ona jedynie częściowy opis sygnału. Znając funkcję autokorelacji, możemy wyznaczyć tylko widmo amplitudowe sygnału. Funkcja autokorelacji nie zawiera natomiast informacji o widmie fazowym. Na podstawie znajomości funkcji autokorelacji nie możemy zatem odtworzyć sygnału.

Związki między sygnałem i jego charakterystykami w dziedzinie korelacyjnej i dziedzinie częstotliwości ilustruje diagram przedstawiony na rys. 5.5. Na diagramie tym strzałki podwójne oznaczają wzajemnie jednoznaczny związek między odpowiednimi wielkościami. Strzałki pojedyncze oznaczają jednoznaczne przejście tylko w jednym kierunku.

$$\begin{tikzcd}[column sep=large, row sep=large]
x(t) \arrow[r, shift left, "\mathscr{F}"] \arrow[d, "\int_{-\infty}^{\infty} x(t)x^*(t-\tau)\,\mathrm{d}t"'] & X(\omega) \arrow[l, shift left, "\mathscr{F}^{-1}"] \arrow[d, "|X(\omega)|^2"] \\
\varphi_x(\tau) \arrow[r, shift left, "\mathscr{F}"] & \Phi_x(\omega) \arrow[l, shift left, "\mathscr{F}^{-1}"]
\end{tikzcd}$$

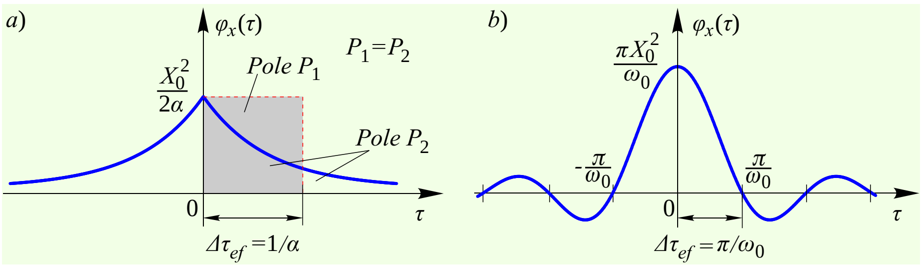
Rys. 5.5. Ilustracja związków między sygnałem a jego charakterystykami

5.1.6. Efektywny czas korelacji i efektywna szerokość pasma sygnału

Funkcje autokorelacji sygnałów o ograniczonej energii przybierają wartość maksymalną dla $\tau = 0$ i maleją do zera, gdy $\tau \to \pm\infty$. Dla różnych sygnałów opadanie funkcji autokorelacji do zera może następować po różnym czasie i z różną prędkością. W zagadnieniach praktycznych często wprowadza się wygodny parametr liczbowy będący miarą *efektywnego czasu korelacji*. Dla sygnałów rzeczywistych, których funkcje autokorelacji dążą monotonicznie do zera, efektywny czas korelacji definiuje się jako:

$$\Delta\tau_{\text{ef}} = \frac{\int\limits_0^{\infty} \varphi_x(\tau)\,\mathrm{d}\tau}{\varphi_x(0)}. \tag{5.17}$$

Interpretacja miary (5.17) została przedstawiona na rys. 5.6a na przykładzie funkcji autokorelacji (5.6) sygnału wykładniczego malejącego. W tym przypadku $\Delta\tau_{\text{ef}} = 1/\alpha$.



Rys. 5.6. Interpretacja efektywnego czasu korelacji dla przypadku monotonicznie malejącej (a) oraz oscylacyjnej funkcji autokorelacji (b)

Jeśli funkcja autokorelacji maleje oscylacyjnie do zera, miara (5.17) nie jest odpowiednią miarą efektywnego czasu korelacji. Efektywny czas korelacyjny definiuje się wówczas jako połowa szerokości listka głównego funkcji autokorelacji. Na przykład, w przypadku funkcji autokorelacji (5.16) idealnego sygnału dolnopasmowego efektywny czas korelacji wynosi $\Delta\tau_{\text{ef}} = \pi/\omega_0$ (rys. 5.6b).

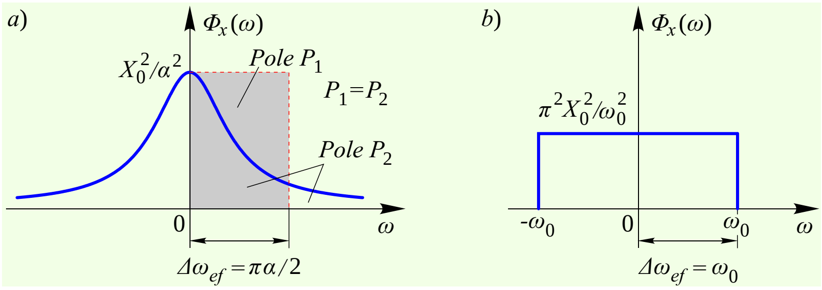
W p. 3.6.1 wprowadziliśmy miarę szerokości widma sygnału. Za miarę tę przyjęliśmy szerokość równoważną widma określoną wzorem (3.62). Miarę szerokości widma można także odnieść do widma energii sygnału, wprowadzając pojęcie *efektywnej szerokości widma*:

$$\Delta\omega_{\text{ef}} = \frac{\int\limits_{-\infty}^{\infty} \Phi_x(\omega)\,\mathrm{d}\omega}{\Phi_{x\,\max}}. \tag{5.18}$$

Interpretację efektywnej szerokości widma dla obu rozważonych wyżej sygnałów przedstawiono na rys. 5.7. W przypadku sygnału wykładniczego malejącego, którego widmo energii jest określone wzorem (5.13), mamy $\Delta\omega_{\text{ef}} = \pi\alpha/2$, natomiast w przypadku idealnego sygnału dolnopasmowego, którego widmo energii jest opisane wzorem (5.15), mamy $\Delta\omega_{\text{ef}} = \omega_0$. W obu przypadkach

$$\Delta\tau_{\text{ef}}\Delta\omega_{\text{ef}} = \text{const}\,. \tag{5.19}$$

Wynika stąd, że im krótszy jest efektywny czas korelacji sygnału, tym większa jest jego efektywna szerokość widma i odwrotnie. Związek (5.19) można przyjąć jako alternatywne sformułowanie zasady nieoznaczoności omówionej w p. 3.6.



Rys. 5.7. Interpretacja efektywnej szerokości widma dla przypadku sygnału wykładniczego malejącego (a) oraz idealnego sygnału dolnopasmowego (b)

5.2. Funkcje korelacji wzajemnej sygnałów analogowych o ograniczonej energii

5.2.1. Definicja

Przedrostek „auto” w nazwie funkcji autokorelacji podkreśla fakt, że iloczyn skalarny (5.2) jest obliczany między wersją nieprzesuniętą i wersją przesuniętą tego samego sygnału. Rozszerzenie tej definicji na dwa różne sygnały prowadzi do pojęcia *funkcji korelacji wzajemnej*. W praktyce potrzeba badania korelacji między dwoma różnymi sygnałami zachodzi na przykład w sytuacji, gdy jeden z nich jest sygnałem występującym na wejściu, a drugi – na wyjściu pewnego układu.

Definicja 5.2. Niech $x(t)$ i $y(t)$ będą sygnałami o ograniczonej energii. Funkcja korelacji wzajemnej między sygnałem $x(t)$ a sygnałem $y(t)$ jest określona wzorem:

$$\varphi_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y^*(t-\tau)\,\mathrm{d}t. \tag{5.20}$$

Analogicznie definiuje się funkcję korelacji wzajemnej między sygnałem $y(t)$ a sygnałem $x(t)$:

$$\varphi_{yx}(\tau) = \int_{-\infty}^{\infty} y(t)x^*(t-\tau)\,\mathrm{d}t. \tag{5.21}$$

Tak jak w przypadku funkcji autokorelacji, funkcje korelacji wzajemnej (5.20) i (5.21) można traktować jako miary wzajemnego położenia dwóch sygnałów w przestrzeni sygnałów o ograniczonej energii rozpatrywane w funkcji przesunięcia τ. Na ich podstawie można w szczególności rozstrzygać, dla jakich wartości przesunięcia τ sygnały są ortogonalne.

Przez analogię do właściwości (5.4), wartości funkcji (5.20) i (5.21) w punkcie $\tau = 0$:

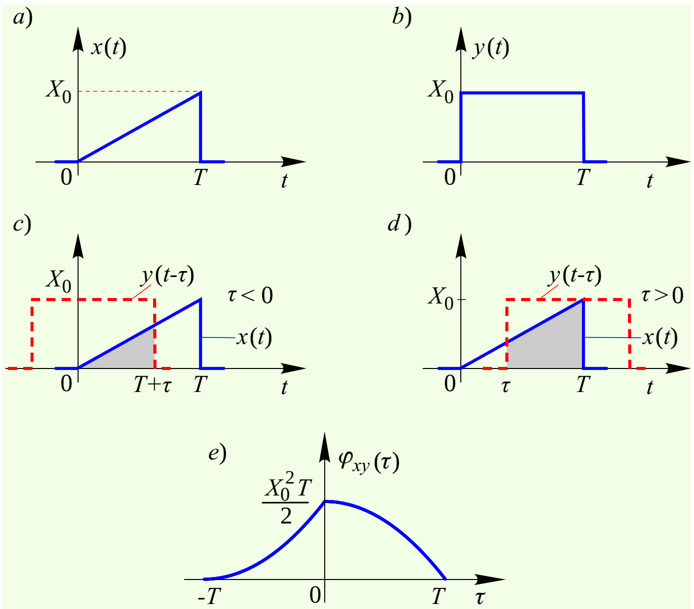
$$E_{xy} = \int\limits_{-\infty}^{\infty} x(t)y^*(t)\,\mathrm{d}t, \qquad E_{yx} = \int\limits_{-\infty}^{\infty} y(t)x^*(t)\,\mathrm{d}t. \tag{5.22}$$

są nazywane *energiami wzajemnymi*. Pojęcie energii wzajemnej ma interesującą interpretację fizyczną. Jeżeli sygnały $x(t)$ i $y(t)$ są rzeczywistymi sygnałami napięcia i odpowiednio prądu na zaciskach pewnego dwójnika elektrycznego, to wielkość $\varphi_{xy}(0) = E_{xy}$ jest energią pobraną przez ten dwójnik.

Przykład 5.7. Funkcja korelacji wzajemnej między sygnałami $x(t)$ i $y(t)$ pokazanymi na rys. 5.8a, b ma postać:

$$\varphi_{xy}(\tau) = \begin{cases} \frac{X_0^2}{2T}(T^2 - \tau^2), & \text{dla} \quad 0 \leqslant \tau \leqslant T, \\ \frac{X_0^2}{2T}(T + \tau)^2, & \text{dla} \quad -T \leqslant t < 0, \\ 0, & \text{dla} \quad |\tau| > T. \end{cases}$$

Sposób obliczania tej funkcji oraz jej wykres są przedstawione na rys. 5.8c-e.



Rys. 5.8. Przykład wyznaczania funkcji korelacji wzajemnej

5.2.2. Właściwości

Funkcja korelacji wzajemnej $\varphi_{xy}(\tau)$ między sygnałem $x(t)$ a sygnałem $y(t)$ jest różna od funkcji korelacji wzajemnej $\varphi_{yx}(\tau)$ między sygnałem $y(t)$ a sygnałem $x(t)$. Dokonując zamiany zmiennych w definicji (5.20), można wykazać, że między tymi funkcjami zachodzi związek:

$$\varphi_{xy}(\tau) = \varphi_{yx}^{*}(-\tau). \tag{5.23}$$

W przypadku sygnałów rzeczywistych równość (5.23) przybiera postać $\varphi_{xy}(\tau) = \varphi_{yx}(-\tau)$, odzwierciedlającą fakt, że dla sygnałów tych taką samą wartość iloczynu skalarnego otrzymujemy przy przesunięciu sygnału $y(t)$ w kierunku opóźnienia o czas τ, jak i przy przesunięciu sygnału $x(t)$ o ten sam czas w kierunku przyspieszenia.

W przeciwieństwie do funkcji autokorelacji, funkcje korelacji wzajemnej sygnałów rzeczywistych nie są funkcjami parzystymi zmiennej τ. Nie muszą one także przybierać wartości maksymalnych dla $\tau = 0$. Mają natomiast następujące ogólne właściwości.

Właściwość 7. Wartości funkcji korelacji wzajemnej $\varphi_{xy}(\tau)$ oraz $\varphi_{yx}(\tau)$ nie przekraczają co do modułu pierwiastka z iloczynu energii sygnałów:

$$|\varphi_{xy}(\tau)| \leqslant \sqrt{E_x E_y}, \qquad |\varphi_{yx}(\tau)| \leqslant \sqrt{E_x E_y}. \tag{5.24}$$

Właściwość tę można dowieść, korzystając z nierówności Buniakowskiego-Schwarza (p. 2.2.7) i uwzględniając, że normą w przestrzeni L^2 sygnałów o ograniczonej energii jest pierwiastek z energii (por. wzór (2.11)).

Właściwość 8. Jeśli $\varphi_{xy}(\tau_0) = 0$ dla pewnego przesunięcia τ_0, to sygnały $x(t)$ i $y(t-\tau_0)$ są ortogonalne.

Właściwość 9. Funkcje korelacji wzajemnej sygnałów o ograniczonej energii są funkcjami całkowalnymi z kwadratem, a więc $\mathscr{F}$-transformowalnymi w zwykłym sensie.

5.2.3. Widma energii wzajemnej

Funkcje korelacji wzajemnej można powiązać z charakterystykami widmowymi sygnałów. W tym celu należy skorzystać ponownie z uogólnionego twierdzenia Rayleigha. Oznaczając $\mathscr{F}[x(t)] = X(\omega)$, $\mathscr{F}[y(t)] = Y(\omega)$ i uwzględniając, że $\mathscr{F}[y^*(t-\tau)] = [Y(\omega)\,\mathrm{e}^{-\mathrm{j}\omega\tau}]^* = Y^*(\omega)\,\mathrm{e}^{\mathrm{j}\omega\tau}$, możemy na podstawie wzoru (3.33) napisać:

$$\int\limits_{-\infty}^{\infty} x(t)y^*(t-\tau)\,\mathrm{d}t = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} X(\omega)Y^*(\omega)\,\mathrm{e}^{\mathrm{j}\omega\tau}\,\mathrm{d}\omega. \tag{5.25}$$

Wielkość

$$\Phi_{xy}(\omega) = X(\omega)Y^*(\omega) \tag{5.26}$$

występująca pod całką po prawej stronie równości (5.25) jest nazywana *widmem energii wzajemnej* między sygnałem $x(t)$ a sygnałem $y(t)$. Ze wzoru (5.25) wynika, że $\varphi_{xy}(\tau) = \mathscr{F}^{-1}[\Phi_{xy}(\omega)]$, a więc funkcja korelacji wzajemnej $\varphi_{xy}(\tau)$ i widmo energii wzajemnej $\Phi_{xy}(\omega)$ tworzą parę transformat Fouriera:

$$\varphi_{xy}(\tau) \;\leftrightarrow\; \Phi_{xy}(\omega). \tag{5.27}$$

Analogicznie, definiując widmo energii wzajemnej między sygnałem $y(t)$ a sygnałem $x(t)$:

$$\Phi_{yx}(\omega) = Y(\omega)X^*(\omega), \tag{5.28}$$

otrzymujemy:

$$\varphi_{yx}(\tau) \;\leftrightarrow\; \Phi_{yx}(\omega). \tag{5.29}$$

Z równości (5.23) i właściwości (3.18) wynika, że widma energii wzajemnej są związane ze sobą zależnością:

$$\Phi_{xy}(\omega) \leftrightarrow \Phi^*_{yx}(\omega). \tag{5.30}$$

Warto odnotować, że w przeciwieństwie do widma energii $\Phi_x(\omega)$ sygnału, które nie zawiera informacji o jego fazie, widma energii wzajemnej zawierają częściową informację o fazach sygnałów, a dokładniej o ich różnicy. Zgodnie ze wzorem (5.26) mamy bowiem:

$$\Phi_{xy}(\omega) = |X(\omega)||Y(\omega)|\,\mathrm{e}^{\mathrm{j}[\arg X(\omega) - \arg Y(\omega)]}\,.$$

5.3. Funkcja autokorelacji sygnału analogowego o ograniczonej mocy

5.3.1. Definicja

Dla sygnałów o ograniczonej mocy całka (5.2) jest rozbieżna. Rozszerzenie pojęcia funkcji autokorelacji na tę klasę sygnałów wymaga zatem odpowiedniej modyfikacji definicji (5.2). Sposób postępowania będzie tu analogiczny do stosowanego wielokrotnie w lekcji 1 przy definiowaniu innych wielkości (mocy, metryki, iloczynu skalarnego) odnoszących się do sygnałów o ograniczonej mocy. Najpierw, korzystając z definicji (5.2), wyznaczamy funkcję autokorelacji w skończonym prostokątnym oknie czasowym, potem odnosimy ją do szerokości tego okna, a następnie przechodzimy do granicy, gdy szerokość okna rośnie do nieskończoności.

Definicja 5.3. *Funkcją autokorelacji* $\psi_x(\tau)$ sygnału $x(t)$ o ograniczonej mocy nazywamy wielkość graniczną:

$$\psi_x(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int\limits_{-T}^{T} x(t)x^*(t-\tau)\,\mathrm{d}t. \tag{5.31}$$

Jeśli sygnał $x(t)$ jest okresowy o okresie T_0, definicja ta jest równoważna definicji:

$$\psi_x(\tau) = \frac{1}{T_0} \int\limits_{t_0}^{t_0+T_0} x(t)x^*(t-\tau)\,\mathrm{d}t, \tag{5.32}$$

gdzie t_0 jest dowolną chwilą.

5.3.2. Właściwości

Właściwości funkcji autokorelacji sygnałów o ograniczonej mocy są odpowiednikami właściwości funkcji autokorelacji sygnałów o ograniczonej energii.

Właściwość 10. Funkcja autokorelacji $\psi_x(\tau)$ jest *funkcją hermitowską*:

$$\psi_x(\tau) = \psi_x^*(-\tau). \tag{5.33}$$

W przypadku sygnałów rzeczywistych jest to funkcja rzeczywista i parzysta: $\psi_x(\tau) = \psi_x(-\tau)$.

Właściwość 11. Wartość funkcji autokorelacji $\psi_x(\tau)$ w punkcie $\tau = 0$ jest rzeczywista i równa jego mocy:

$$\psi_x(0) = \lim_{T\to\infty} \frac{1}{2T} \int\limits_{-T}^{T} |x(t)|^2\,\mathrm{d}t = P_x. \tag{5.34}$$

Właściwość 12. Wartości funkcji autokorelacji $\psi_x(\tau)$ nie przekraczają co do modułu jej wartości w punkcie $\tau = 0$, tzn. dla każdego τ spełniona jest nierówność:

$$|\psi_x(\tau)| \leqslant \psi_x(0). \tag{5.35}$$

Właściwość 13. Jeśli $\psi_x(\tau_0) = 0$ dla pewnego przesunięcia τ_0, to sygnały $x(t)$ i $x(t-\tau_0)$ są ortogonalne.

Właściwość 14. Funkcja autokorelacji $\psi_x(\tau)$ jest niezmiennicza względem położenia sygnału na osi czasu, tzn. funkcje autokorelacji sygnału $x(t)$ i sygnału przesuniętego $x(t-t_0)$ są identyczne dla każdej wartości przesunięcia t_0.

Właściwość 15. Funkcja autokorelacji sygnału o ograniczonej mocy jest $\mathscr{F}$-transformowalna w sensie granicznym.

Właściwość 16. Funkcja autokorelacji sygnału okresowego o okresie T_0 jest również funkcją okresową o tym samym okresie. Właściwość ta wynika wprost ze wzoru (5.32).

Przykład 5.8. Wyznaczymy funkcję autokorelacji skoku jednostkowego $x(t) = \mathbf{1}(t)$. Dla $\tau \geqslant 0$ mamy:

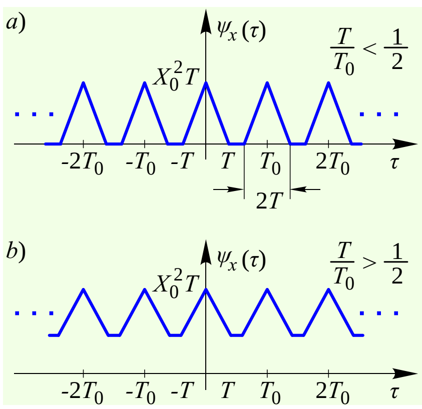
$$\psi_x(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{\tau}^{T} \mathrm{d}t = \lim_{T\to\infty} \frac{1}{2T}(T-\tau) = \frac{1}{2}.$$

Podobnie, dla $\tau < 0$:

$$\psi_x(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{0}^{T} \mathrm{d}t = \lim_{T\to\infty} \frac{1}{2T} T = \frac{1}{2}.$$

Tak więc funkcja autokorelacji skoku jednostkowego jest funkcją stałą o wartości $1/2$.

Przykład 5.9. Funkcja autokorelacji unipolarnej fali prostokątnej (rys. 1.23) jest pokazana na rys. 5.9. Jeśli współczynnik wypełnienia fali $T/T_0 \leqslant 1/2$, funkcja ta jest ciągiem trójkątów o szerokości $2T$ powtarzanych co okres T_0 (rys. 5.9a). Jeśli długość impulsów T fali wzrośnie przy tym samym okresie T_0, tak że $T/T_0 > 1/2$, to trójkąty te nakładają się na siebie, w efekcie czego ich szerokość jest mniejsza niż $2T$ i są one przesunięte do góry wzdłuż osi rzędnych (rys. 5.9b).



Rys. 5.9. Funkcja autokorelacji unipolarnej fali prostokątnej

Przykład 5.10. Wyznaczymy funkcję autokorelacji sygnału harmonicznego $x(t) = X_0 \sin(\omega_0 t + \varphi_0)$. Zgodnie ze wzorem (5.32) mamy:

$$
\begin{aligned}
\psi_x(\tau) &= \frac{X_0^2}{T_0} \int_0^{T_0} \sin(\omega_0 t + \varphi_0) \sin[\omega_0 (t - \tau) + \varphi_0] \, \mathrm{d}t = \\
&= \frac{X_0^2}{T_0} \int_0^{T_0} \sin(\omega_0 t + \varphi_0)[\sin(\omega_0 t + \varphi_0) \cos \omega_0 \tau - \cos(\omega_0 t + \varphi_0) \sin \omega_0 \tau] \, \mathrm{d}t = \\
&= \frac{X_0^2}{T_0} \cos \omega_0 \tau \int_0^{T_0} \sin^2(\omega_0 t + \varphi_0) \, \mathrm{d}t - \frac{X_0^2}{2T_0} \sin \omega_0 \tau \int_0^{T_0} \sin 2(\omega_0 t + \varphi_0) \, \mathrm{d}\tau = \\
&= \frac{X_0^2}{T_0} \cos \omega_0 \tau \int_0^{T_0} \frac{1}{2} \, \mathrm{d}t = \frac{X_0^2}{2} \cos \omega_0 \tau.
\end{aligned}
\tag{5.36}
$$

Funkcja ta jest również harmoniczna o tej samej co sygnał pulsacji ω_0 i nie zależy od fazy początkowej φ_0.

5.3.3. Widmo mocy i jego związek z funkcją autokorelacji

Właściwości energetyczne sygnałów o ograniczonej energii opisuje w dziedzinie częstotliwości widmo energii. Analogiczną charakterystykę wprowadza się dla sygnałów o ograniczonej mocy. Podobnie jak inne wielkości charakteryzujące tę klasę sygnałów, ma ona sens wielkości granicznej.

Rozważmy sygnał $x(t)$ o ograniczonej mocy. Oznaczmy przez $x_T(t) = x(t)\sqcap(t/T)$ jego centralny wycinek o szerokości T oraz przez $\Phi_T(\omega) = |X_T(\omega)|^2$ widmo energii tego wycinka.

Definicja 5.4. *Widmem mocy* sygnału $x(t)$ o ograniczonej mocy nazywamy granicę:

$$\Psi_x(\omega) = \lim_{T \to \infty} \frac{1}{T} \Phi_T(\omega). \tag{5.37}$$

Konstruując ciąg funkcji autokorelacji $\varphi_T(\tau)$ sygnału $x_T(t)$ i dokonując odpowiedniego przejścia granicznego, można wykazać, że funkcja autokorelacji $\psi_x(\tau)$ sygnału $x(t)$ i jego widmo mocy $\Psi_x(\omega)$ tworzą parę transformat Fouriera w sensie granicznym:

$$\Psi_x(\omega) = \int_{-\infty}^{\infty} \psi_x(\tau) \, \mathrm{e}^{-\mathrm{j}\omega\tau} \, \mathrm{d}\tau, \tag{5.38}$$

$$\psi_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_x(\omega) \, \mathrm{e}^{\mathrm{j}\omega\tau} \, \mathrm{d}\omega. \tag{5.39}$$

Podstawiając we wzorze (5.39) $\tau = 0$, otrzymujemy:

$$\psi_x(0) = P_x = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \Psi_x(\omega)\,\mathrm{d}\omega, \tag{5.40}$$

skąd wynika, że moc sygnału $x(t)$ jest równa polu pod wykresem jego widma mocy podzielonemu przez 2π. Uzasadnia to przyjętą nazwę widma $\Psi_x(\omega)$. Podobnie jak widmo energii, widmo mocy nie zawiera informacji o strukturze fazowej sygnału.

Z definicji 5.4 wynika, że widmo mocy jest nieujemną rzeczywistą funkcją zmiennej ω. Jeśli sygnał $x(t)$ jest rzeczywisty, widmo mocy jest funkcją parzystą i moc sygnału możemy wówczas obliczać ze wzoru:

$$P_x = \frac{1}{\pi} \int\limits_{0}^{\infty} \Psi_x(\omega)\,\mathrm{d}\omega. \tag{5.41}$$

5.3.4. Widmo mocy sygnałów okresowych

Zgodnie z właściwością 7, jeśli sygnał $x(t)$ o ograniczonej mocy jest sygnałem okresowym o okresie T_0, jego funkcja autokorelacji $\psi_x(\tau)$ jest również okresowa o tym samym okresie. Można wykazać, że współczynnikami rozwinięcia tej funkcji w zespolony szereg Fouriera są kwadraty modułów $|X_k|^2$ współczynników X_k rozwinięcia sygnału $x(t)$ w zespolony szereg Fouriera. Zespolony szereg Fouriera funkcji autokorelacji sygnału okresowego o okresie T_0 ma zatem postać:

$$\psi_x(\tau) = \sum_{k=-\infty}^{\infty} |X_k|^2\,\mathrm{e}^{\mathrm{j}k\omega_0\tau}, \tag{5.42}$$

gdzie $\omega_0 = 2\pi/T_0$. Wynika stąd, że widmo mocy sygnałów okresowych jest określone wzorem (por. wzór (3.52)):

$$\Psi_x(\omega) = 2\pi \sum_{k=-\infty}^{\infty} |X_k|^2 \delta(\omega - k\omega_0). \tag{5.43}$$

Przykład 5.11. Ze wzorów (5.36) i (3.49) wynika, że widmo mocy sygnału harmonicznego $x(t) = X_0 \sin(\omega_0 t + \varphi_0)$ ma postać:

$$\Psi_x(\omega) = \frac{\pi}{2} X_0^2 [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]. \tag{5.44}$$

5.4. Funkcje korelacji wzajemnej sygnałów analogowych o ograniczonej mocy

5.4.1. Definicja

Definicja funkcji korelacji wzajemnej dwóch sygnałów o ograniczonej mocy jest bezpośrednim uogólnieniem definicji (5.20) i (5.21).

Definicja 5.5. Funkcje korelacji wzajemnej sygnałów $x(t)$ i $y(t)$ o ograniczonej mocy są określone wzorami:

$$\psi_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)y^*(t-\tau)\,\mathrm{d}t, \tag{5.45}$$

$$\psi_{yx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t)x^*(t-\tau)\,\mathrm{d}t. \tag{5.46}$$

Jeśli sygnały $x(t)$ i $y(t)$ są okresowe o tym samym okresie T_0, wzory (5.45) i (5.46) przybierają postać równoważną:

$$\psi_{xy}(\tau) = \frac{1}{T_0} \int_{t_0}^{t_0+T_0} x(t)y^*(t-\tau)\,\mathrm{d}t, ; \tag{5.47}$$

$$\psi_{yx}(\tau) = \frac{1}{T_0} \int_{t_0}^{t_0+T_0} y(t)x^*(t-\tau)\,\mathrm{d}t. \tag{5.48}$$

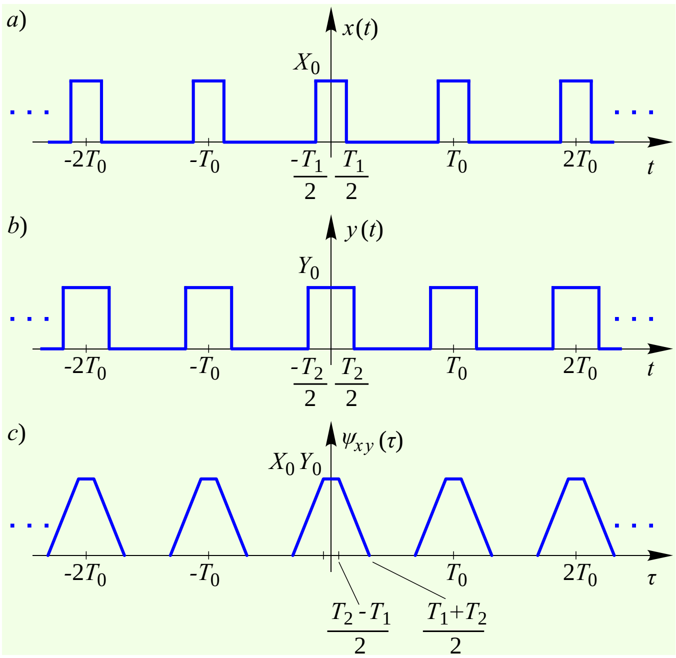
Wartości funkcji (5.46) i (5.47) w punkcie $\tau = 0$:

$$P_{xy} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)y^*(t)\,\mathrm{d}t, \qquad P_{yx} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)y^*(t)\,\mathrm{d}t. \tag{5.49a, b}$$

są nazywane *mocami wzajemnymi*. Jeżeli sygnały $x(t)$ i $y(t)$ są rzeczywistymi okresowymi sygnałami napięcia i odpowiednio prądu na zaciskach pewnego dwójnika elektrycznego, to moc wzajemna $\psi_{xy}(0) = P_{xy}$ ma interpretację mocy czynnej pobranej przez ten dwójnik.

Właściwości funkcji korelacji wzajemnej sygnałów o ograniczonej mocy są analogiczne do właściwości funkcji korelacji wzajemnej sygnałów o ograniczonej energii i w związku z tym nie będziemy ich powtarzać. Podkreślimy jedynie, że funkcje te są $\mathscr{F}$-transformowalne w sensie granicznym.

Przykład 5.12. Funkcja korelacji wzajemnej między dwoma unipolarnymi falami prostokątnymi o tym samym okresie T_0 i różnych współczynnikach wypełnienia T_1/T_0 oraz T_2/T_0 (rys. 5.10a, b) jest pokazana dla przypadku $T_1 < T_2$ oraz $T_2 < T_0/2$ na rys. 5.10c. W celu wyznaczenia tej funkcji wystarczy obliczyć całkę (5.47) jedynie w jednym okresie i dokonać przedłużenia okresowego otrzymanej funkcji.



Rys. 5.10. Funkcja korelacji wzajemnej dwóch fal unipolarnych

5.4.2. Widma mocy wzajemnej

Jeśli $\mathscr{F}[x(t)] = X(\omega)$ oraz $\mathscr{F}[y(t)] = Y(\omega)$ są widmami (w sensie granicznym) sygnałów $x(t)$ i $y(t)$ o ograniczonej mocy, to funkcje

$$\Psi_{xy}(\omega) = X(\omega)Y^*(\omega), \qquad \Psi_{yx}(\omega) = Y(\omega)X^*(\omega) \tag{5.50}$$

nazywamy *widmami mocy wzajemnej*.

Funkcje korelacji wzajemnej i widma mocy wzajemnej sygnałów o ograniczonej mocy tworzą pary transformat Fouriera w sensie granicznym:

$$\psi_{xy}(\tau) \leftrightarrow \Psi_{xy}(\omega), \qquad \psi_{yx}(\tau) \leftrightarrow \Psi_{yx}(\omega). \tag{5.51}$$

5.5. Funkcje korelacyjne sygnałów dyskretnych o ograniczonej energii

Analiza korelacyjna sygnałów dyskretnych przebiega podobnie jak w przypadku sygnałów analogowych. Z tego względu nie będziemy podawać wszystkich definicji i omawiać szczegółowo właściwości definiowanych wielkości. Ograniczymy się do najważniejszych z nich i wskażemy na występujące podobieństwa i różnice. Zasadniczą różnicą jest dyskretny charakter funkcji korelacyjnych sygnałów dyskretnych.

5.5.1. Funkcja autokorelacji

Tak jak w przypadku sygnałów analogowych, definicja funkcji autokorelacji sygnału dyskretnego jest oparta na pojęciu iloczynu skalarnego. W przestrzeni sygnałów dyskretnych o ograniczonej energii (przestrzeni l^2) iloczyn skalarny jest określony wzorem (2.19).

Definicja 5.6. *Funkcją autokorelacji* $\varphi_x[m]$ dyskretnego sygnału $x[n]$ o ograniczonej energii nazywamy funkcję całkowitego argumentu m (przesunięcia) określoną wzorem:

$$\varphi_x(m) = \sum_{n=-\infty}^{\infty} x(n)x^*(n-m). \tag{5.52}$$

Funkcja autokorelacji sygnału dyskretnego jest więc dyskretną funkcją zmiennej m równą iloczynowi skalarnemu tego sygnału i jego przesuniętej w czasie kopii $x[n-m]$ w przestrzeni l^2. W ogólnym przypadku jest to funkcja hermitowska $\varphi_x[m] = \varphi_x^*[-m]$. W przypadku sygnałów rzeczywistych funkcja autokorelacji jest funkcją rzeczywistą parzystą $\varphi_x[m] = \varphi_x[-m]$.

W punkcie $m = 0$ moduł funkcji $\varphi_x[m]$ przybiera wartość maksymalną równą energii sygnału:

$$\varphi_x(0) = \sum_{n=-\infty}^{\infty} |x(n)|^2 = E_x. \tag{5.53}$$

Jeśli $\varphi_x(m_0) = 0$ dla pewnego przesunięcia m_0, to sygnały $x[n]$ i $x[n-m_0]$ są ortogonalne.

Przykład 5.13. Wyznaczymy funkcję autokorelacji binarnego sygnału dyskretnego $x[n]$ o skończonym czasie trwania przybierającego wartości: $x(0) = 1$, $x(1) = 1$, $x(2) = -1$ oraz $x(n) = 0$ dla $m \neq 0, 1, 2$. (rys. 5.11a).

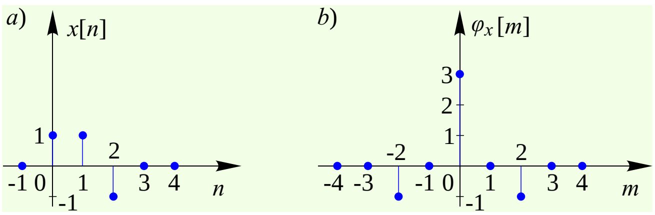
W celu obliczania funkcji autokorelacji sygnałów dyskretnych wygodnie jest zapisać ciąg określający sygnał, a pod nim ciągi określające kolejne jego kopie przesunięte o $m = 1, 2$ itd. pozycji:

$$\begin{array}{rcccccccc} x[n]: & \ldots 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0\ldots \\ x[n-1]: & \ldots 0 & 0 & 0 & 1 & 1 & -1 & 0 & 0\ldots \\ x[n-2]: & \ldots 0 & 0 & 0 & 0 & 1 & 1 & -1 & 0\ldots \\ x[n-3]: & \ldots 0 & 0 & 0 & 0 & 0 & 1 & 1 & -1\ldots \end{array}$$

Obliczając sumy iloczynów wartości próbek sygnału $x[n]$ i występujących w tych samych kolumnach wartości jego kolejnych przesuniętych kopii, otrzymujemy:

$$\begin{aligned} \varphi_x(0) &= 1 + 1 + 1 = 3, \\ \varphi_x(1) &= \varphi_x(-1) = 1 - 1 = 0, \\ \varphi_x(2) &= \varphi_x(-2) = -1, \end{aligned}$$

oraz $\varphi_x(m) = 0$ dla $|m| \geqslant 3$. Wykres funkcji $\varphi_x[m]$ jest pokazany na rys. 5.11b.



Rys. 5.11. Sygnał dyskretny (a) i jego funkcja autokorelacji (b)

Przykład 5.14. Wyznaczymy funkcję autokorelacji dyskretnego sygnału wykładniczego $x[n] = a^n \mathbf{1}[n]$, $|a\ | < 1$ (rys. 5.12a, por. także przykład 4.3). W tym przypadku sygnał jest opisany formułą matematyczną, zatem funkcję autokorelacji obliczamy na podstawie wzoru definicyjnego (5.52). Dla $m \geqslant 0$ mamy:

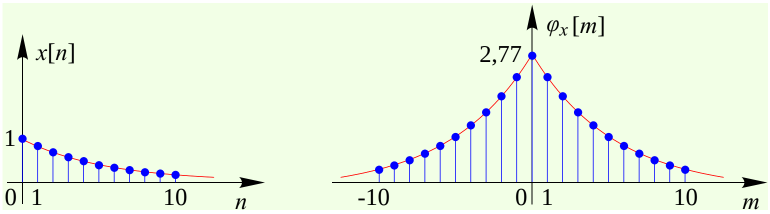
$$\varphi_x(m) = \sum_{n=m}^{\infty} a^n a^{n-m} = a^{-m} \sum_{n=m}^{\infty} a^{2n} = \frac{a^m}{1-a^2},$$

a uwzględniając parzystość funkcji autokorelacji otrzymujemy:

$$\varphi_x(m) = \frac{a^{|m|}}{1-a^2}. \tag{5.54}$$

Wykres tej funkcji dla $a = 0{,}8$ przedstawiono na rys. 5.12b.

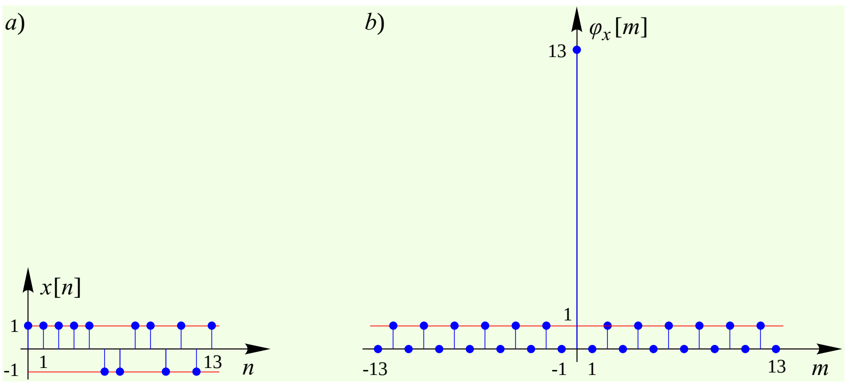
Przykład 5.15. Sygnał rozpatrywany w przykładzie 5.13 jest tzw. 3-pozycyjnym *sygnałem Barkera*. Sygnały Barkera zyskały w praktyce dużą popularność ze względu na bardzo korzystne właściwości ich funkcji autokorelacji. Są to



Rys. 5.12. Dyskretny sygnał wykładniczy (a) i jego funkcja autokorelacji (b)

M-pozycyjne sygnały binarne przybierające na każdej pozycji wartość $+1$ lub -1. Funkcje autokorelacji sygnałów Barkera mają tę właściwość, że $\varphi_x(0) = M$ oraz $|\varphi_x(m)| \leqslant 1$ dla $m \neq 0$, tzn. w punkcie $m = 0$ funkcje te przybierają wartość równą liczbie pozycji, a dla pozostałych m ich wartości nie przekraczają co do modułu jedności. Z uwagi na dużą wartość w zerze (duży poziom „listka głównego") i małe pozostałe wartości funkcji autokorelacji (mały poziom „listków bocznych"), sygnały Barkera, podobnie jak impuls LFM, znalazły szerokie zastosowanie w radiolokacji. W praktyce wartości binarne sygnałów Barkera $+1$ oraz -1 są reprezentowane odpowiednio zmodulowanymi impulsami radiowymi.

Znane są funkcje Barkera o liczbie pozycji $M = 2, 3, 4, 5, 7, 11$ i 13. Ciekawostką jest przy tym, że nie podano jak dotąd dowodu rozstrzygającego istnienie lub nieistnienie sygnału Barkera o liczbie pozycji większej niż 13. Sygnał Barkera i jego funkcja autokorelacji dla $M = 13$ przedstawiono na rys. 5.13.



Rys. 5.13. 13-pozycyjny sygnał Barkera (a) i jego funkcja autokorelacji (b)

5.5.2. Związek z widmem energii

Rozważmy dyskretny sygnał $x[n]$ o ograniczonej energii i widmie $X(\mathrm{e}^{\mathrm{j}\theta})$. Widmo energii tego sygnału określiliśmy w punkcie 4.2 jako kwadrat jego widma amplitudowego $\Phi_x(\mathrm{e}^{\mathrm{j}\theta}) = A_x^2(\mathrm{e}^{\mathrm{j}\theta}) = |X(\mathrm{e}^{\mathrm{j}\theta})|^2$ (por. komentarz do twierdzenia 4.7). Z uogólnionego twierdzenia Rayleigha dla sygnałów dyskretnych (4.21) wynika, że funkcja autokorelacji sygnału dyskretnego o ograniczonej energii i jego widmo energii tworzą parę transformat Fouriera:

$$\Phi_x(\mathrm{e}^{\mathrm{j}\theta}) = \sum_{m=-\infty}^{\infty} \varphi_x(m)\,\mathrm{e}^{-\mathrm{j}m\theta}, \tag{5.55}$$

$$\varphi_x(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_x(\mathrm{e}^{\mathrm{j}\theta})\,\mathrm{e}^{\mathrm{j}m\theta}\,\mathrm{d}\theta. \tag{5.56}$$

Podstawiając we wzorze (5.56) $m = 0$, otrzymujemy:

$$E_x = \varphi_x(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_x(\mathrm{e}^{\mathrm{j}\theta})\,\mathrm{d}\theta, \tag{5.57}$$

a więc energię sygnału $x[n]$ można obliczać jako pole pod wykresem widma energii w przedziale $[-\pi, \pi]$ podzielone przez 2π.

Przykład 5.16. Widmo energii dyskretnego sygnału wykładniczego o funkcji autokorelacji (5.54) ma postać:

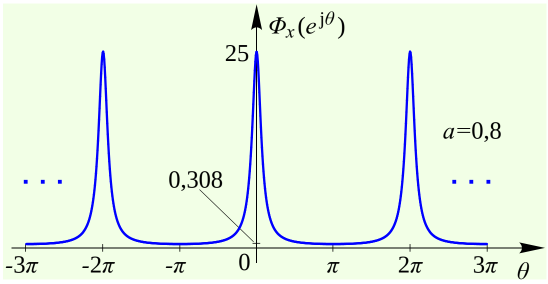
$$\begin{aligned}
\Phi_x(\mathrm{e}^{\mathrm{j}\theta}) &= \frac{1}{1-a^2}\sum_{m=-\infty}^{\infty} a^{|m|}\,\mathrm{e}^{-\mathrm{j}m\theta} = \\
&= \frac{1}{1-a^2}\left[\sum_{m=0}^{\infty}(a\,\mathrm{e}^{-\mathrm{j}\theta})^m + \sum_{m=-\infty}^{-1}(a\,\mathrm{e}^{\mathrm{j}\theta})^{-m}\right] = \\
&= \frac{1}{1-a^2}\left[\sum_{m=0}^{\infty}(a\,\mathrm{e}^{-\mathrm{j}\theta})^m + \sum_{m=1}^{\infty}(a\,\mathrm{e}^{\mathrm{j}\theta})^{m}\right] = \\
&= \frac{1}{1-a^2}\left[\frac{1}{1-a\,\mathrm{e}^{-\mathrm{j}\theta}} + \frac{a\,\mathrm{e}^{\mathrm{j}\theta}}{1-a\,\mathrm{e}^{\mathrm{j}\theta}}\right] = \\
&= \frac{1}{1-a^2}\left[\frac{1-a^2}{1+a^2-2a\cos\theta}\right] = \frac{1}{1+a^2-2a\cos\theta}.
\end{aligned} \tag{5.58}$$

Wynik ten jest oczywiście zgodny ze wzorem (4.9a). Widmo (5.58) dla przypadku $a = 0{,}8$ wykreślono na rys. 5.14.

5.5.3. Funkcje korelacji wzajemnej i widma energii wzajemnej

W przypadku, gdy rozpatrywane są dwa sygnały $x(t)$ i $y(t)$ o ograniczonej energii, wprowadza się ich funkcje korelacji wzajemnej:

$$\varphi_{xy}(m) = \sum_{n=-\infty}^{\infty} x(n)y^*(n-m), \qquad \varphi_{yx}(m) = \sum_{n=-\infty}^{\infty} y(n)x^*(n-m) \tag{5.59a, b}$$



Rys. 5.14. Widmo energii dyskretnego sygnału wykładniczego

oraz widma energii wzajemnej:

$$\Phi_{xy}(\mathrm{e}^{\mathrm{j}\theta}) = X(\mathrm{e}^{\mathrm{j}\theta})Y^*(\mathrm{e}^{\mathrm{j}\theta}), \qquad \Phi_{yx}(\mathrm{e}^{\mathrm{j}\theta}) = Y(\mathrm{e}^{\mathrm{j}\theta})X^*(\mathrm{e}^{\mathrm{j}\theta}). \tag{5.60a, b}$$

Charakterystyki te tworzą pary transformat Fouriera:

$$\varphi_{xy}[m] \leftrightarrow \Phi_{xy}(\mathrm{e}^{\mathrm{j}\theta}), \qquad \varphi_{yx}[m]) \;\leftrightarrow\; \Phi_{yx}(\mathrm{e}^{\mathrm{j}\theta}). \tag{5.61a, b}$$

Ich właściwości są analogiczne jak w przypadku sygnałów analogowych.

5.6. Funkcje korelacyjne sygnałów dyskretnych o ograniczonej mocy

5.6.1. Funkcja autokorelacji

W przypadku sygnałów dyskretnych o ograniczonej mocy suma (5.52) jest nieskończona. Definicję funkcji autokorelacji należy zatem zmodyfikować, dokonując odpowiedniego przejścia granicznego. Sposób postępowania będzie analogiczny jak w przypadku innych wielkości definiowanych w sensie granicznym dla tej klasy sygnałów.

Definicja 5.7. *Funkcją autokorelacji* $\psi_x[m]$ dyskretnego sygnału $x[n]$ o ograniczonej mocy nazywamy funkcję:

$$\psi_x(m) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n)x^*(n-m). \tag{5.62}$$

W szczególnym przypadku sygnałów N-okresowych funkcja $\psi_x[m]$ jest określona wzorem równoważnym:

$$\psi_x(m) = \frac{1}{N} \sum_{n=n_0}^{n_0+N} x(n)x^*(n-m), \tag{5.63}$$

gdzie n_0 jest dowolną liczbą całkowitą.

Funkcja (5.62) jest hermitowska $\psi_x[m] = \psi_x^*[-m]$. W przypadku sygnałów rzeczywistych jest to funkcja rzeczywista parzysta $\psi_x[m] = \psi_x[-m]$. W punkcie $m = 0$ funkcja $\psi_x[m]$ przybiera wartość maksymalną równą mocy sygnału:

$$\psi_x(0) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x(n)|^2 = P_x. \tag{5.64}$$

Jeśli $\psi_x(m_0) = 0$ dla pewnego przesunięcia m_0, to sygnały $x[n]$ i $x[n-m_0]$ są ortogonalne.

W podobny sposób można zdefiniować dla sygnałów o ograniczonej mocy funkcje korelacji wzajemnej. Ich definicje będą uogólnieniami wzorów (5.59).

Przykład 5.17. Funkcja autokorelacji dyskretnego sygnału harmonicznego $x[n] = X_0 \sin(n\theta_0 + \varphi_0)$ ma postać:

$$\psi_x(m) = \frac{1}{2} X_0^2 \cos(m\theta_0). \tag{5.65}$$

Funkcja ta jest także funkcją harmoniczną o tej samej pulsacji unormowanej θ_0 i nie zależy od fazy początkowej φ_0. Dla $m = 0$ jej wartość jest równa mocy sygnału $\psi_x(0) = X_0^2/2$.

5.6.2. Widmo mocy

Pojęcie widma mocy sygnału o ograniczonej mocy wprowadza się jako granicę odpowiednio skonstruowanego ciągu widm energii.

Definicja 5.8. *Widmem mocy* dyskretnego sygnału $x[n]$ o ograniczonej mocy nazywamy granicę:

$$\Psi_x(e^{j\theta}) = \lim_{N\to\infty} \frac{1}{2N+1} \Phi_N(e^{j\theta}), \tag{5.66}$$

gdzie $\Phi_N(e^{j\theta})$ jest widmem energii wycinka sygnału $x[n]$ o skończonym czasie trwania $[-N, N]$.

Analogicznie można wprowadzić pojęcie widma mocy wzajemnej dla sygnałów dyskretnych. W celu wyznaczenia widm mocy należy każdorazowo dokonać odpowiedniego przejścia granicznego. Ze względów omówionych wcześniej w p. 4.1.6, w praktyce przejście takie jest rzadko stosowane. Ograniczymy się zatem do podania jedynie widma mocy dyskretnego sygnału harmonicznego.

Przykład 5.18. Widmo mocy dyskretnego sygnału harmonicznego $x[n] = X_0 \sin(n\theta_0 + \varphi_0)$ ma postać (por. wzór (4.38)):

$$\Psi_x(e^{j\theta}) = \frac{\pi}{2} X_0^2 \left[\delta_{2\pi}(\theta - \theta_0) + \delta_{2\pi}(\theta + \theta_0)\right]. \tag{5.67}$$

5.6.3. Przypadek sygnałów N-okresowych

Jeżeli sygnał o ograniczonej mocy jest sygnałem N-okresowym $\bar{x}[n]$ (por. p. 4.3.1), jego funkcja autokorelacji $\bar{\psi}_x(m)$ jest również N-okresowa. Wynika to wprost ze wzoru (5.63). Podobnie jak sygnał $\bar{x}[n]$ (por. wzór (4.34)), można ją rozwinąć w dyskretny szereg Fouriera:

$$\bar{\psi}_{\bar{x}}(m) = \frac{1}{N}\sum_{k=0}^{N-1}\Psi(k)\,\mathrm{e}^{\mathrm{j}2\pi km/N}\,. \tag{5.68}$$

Definicja 5.9. *Widmem mocy* N-okresowego sygnału $\bar{x}[n]$ nazywamy zbiór współczynników $\{\Psi(k) : k = 0, \ldots, N-1\}$ rozwinięcia funkcji autokorelacji tego sygnału w dyskretny szereg Fouriera (5.68).

Widmo mocy sygnału N-okresowego jest oczywiście również N-okresowe. Można wykazać, że $\Psi(k) = |X(k)|^2$, tj. współczynniki $\Psi(k)$ są kwadratami modułów współczynników rozwinięcia sygnału w dyskretny szereg Fouriera (4.34). Dla $m = 0$ otrzymujemy:

$$P_{\bar{x}} = \bar{\psi}_{\bar{x}}(0) = \sum_{k=0}^{N-1}\Psi(k). \tag{5.69}$$

Słownik

funkcja autokorelacji sygnału
: iloczyn skalarny sygnału i jego kopii przesuniętej w czasie rozpatrywany jako funkcja przesunięcia; charakteryzuje wzajemne rozmieszczenie energii w czasie między sygnałem i sygnałem przesuniętym

funkcja korelacji wzajemnej
: iloczyn skalarny dwóch różnych sygnałów rozpatrywany w funkcji przesunięcia jednego sygnału względem drugiego; charakteryzuje wzajemne rozmieszczenie energii w czasie między tymi sygnałami

korelator
: układ wyznaczający iloczyn skalarny dwóch sygnałów

modulacja częstotliwości liniowa LFM
: modulacja, w której częstotliwość fali nośnej wzrasta liniowo w czasie trwania impulsu

paczka impulsów
: skończony ciąg identycznych impulsów

widmo mocy

funkcja opisująca rozkład mocy sygnału o ograniczonej mocy w funkcji częstotliwości

Literatura

[1] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

[2] Baskakow Ś.: Sygnały i układy radiotechniczne. PWN, Warszawa, 1991 (tłum. z ros.)

Rozdział 2

Podstawy teoretyczne przetwarzania sygnałów

Lekcja 6

Próbkowanie sygnałów

Lekcja 6 jest poświęcona omówieniu operacji próbkowania sygnałów. W wyniku próbkowania sygnał analogowy zostaje przekształcony w sygnał dyskretny. Można zatem powiedzieć, że operacja próbkowania stanowi swojego rodzaju „pomost” między dziedziną sygnałów analogowych a dziedziną sygnałów dyskretnych.

Zagadnienie próbkowania scharakteryzowaliśmy już pokrótce w p. 1.3. W tym rozdziale rozpatrzymy je dokładniej. Omówimy m.in. teoretyczne podstawy operacji próbkowania i przedyskutujemy niektóre aspekty jej praktycznej realizacji. Rozstrzygniemy także, przy jakich założeniach operacja próbkowania jest odwracalna. Odpowiemy tym samym na pytanie o znaczeniu zasadniczym dla zastosowań praktycznych: jakie warunki muszą być spełnione, aby przejście z dziedziny sygnałów analogowych do dziedziny sygnałów dyskretnych było wzajemnie jednoznaczne. Lekcję rozpoczniemy od omówienia ogólnych zasad przetwarzania analogowo-cyfrowego sygnałów.

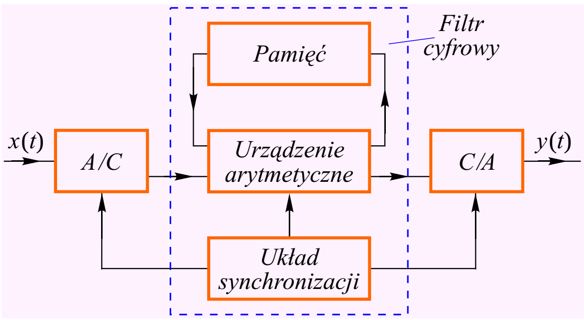
6.1. Ogólne zasady przetwarzania analogowo-cyfrowego sygnałów

Operacja próbkowania stanowi pierwszy etap na drodze przetworzenia sygnału analogowego w sygnał cyfrowy. Dlatego przed przystąpieniem do dokładnego omówienia tej operacji celowe jest krótkie scharakteryzowanie ogólnych zasad przetwarzania analogowo-cyfrowego sygnałów.

6.1.1. Koncepcja cyfrowego przetwarzania sygnałów

Cechą charakterystyczną współczesnego przetwarzania sygnałów jest postępująca dominacja metod cyfrowych. Metody te wkroczyły szerokim frontem do wielu dziedzin techniki, wypierając klasyczne metody analogowe i otwierając przed praktykami nowe obszary zastosowań.

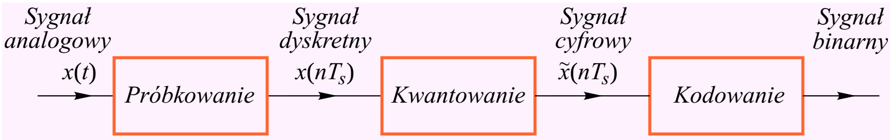
Cyfrowe metody przetwarzania sygnałów polegają na przetworzeniu badanego sygnału analogowego w sygnał cyfrowy reprezentowany ciągiem słów binarnych o ustalonej długości słowa, a następnie dokonywaniu wszelkich operacji na sygnale jako operacji na ciągach binarnych reprezentujących ten sygnał. Schemat blokowy ilustrujący ideę cyfrowego przetwarzania sygnałów jest pokazany na rys. 6.1. Układ przetwarzania składa się z: *przetwornika analogowo-cyfrowego* (*przetwornika A/C*), którego zadaniem jest zamiana postaci analogowej sygnału wejściowego na postać binarną, *filtru cyfrowego*, który realizuje zadane operacje na wejściowym sygnale binarnym, przetwarzając go w inny sygnał binarny, oraz z *przetwornika cyfrowo-analogowego* (*przetwornika C/A*), który zamienia postać binarną sygnału wyjściowego na pożądaną z reguły postać analogową. Filtr cyfrowy zawiera urządzenie arytmetyczne oraz pamięć. Całość jest sterowana i synchronizowana specjalnym układem zewnętrznym.



Rys. 6.1. Schemat blokowy układu cyfrowego przetwarzania sygnałów

6.1.2. Przetwornik analogowo-cyfrowy

Aby przetworzyć sygnał analogowy w sygnał binarny, należy wykonać na nim trzy fundamentalne operacje: próbkowanie, kwantowanie i kodowanie (rys. 6.2). Wszystkie te operacje są realizowane przez przetwornik analogowo-cyfrowy. Jego działanie, podobnie jak innych bloków funkcjonalnych układu z rys. 6.1, jest sterowane generatorem impulsów synchronizujących, których częstotliwość powtarzania określa zarazem częstotliwość próbkowania sygnału analogowego. Na wyjściu przetwornika A/C występuje sygnał reprezentowany ciągiem *słów binarnych* kodujących kolejne próbki sygnału. Każde z tych słów jest ciągiem znaków binarnych „1” oraz „0” o ustalonej dla danego przetwornika długości.



Rys. 6.2. Podstawowe operacje przetwarzające sygnał analogowy w sygnał binarny

6.1.3. Filtr cyfrowy

Słowa binarne z wyjścia przetwornika A/C są przesyłane do układu nazywanego filtrem cyfrowym. Mogą być one przesyłane znak po znaku (transmisja szeregowa), bądź też wszystkie znaki są przesyłane jednocześnie odrębnymi torami (transmisja równoległa). W filtrze cyfrowym następuje przetwarzanie słów. Pojęcie filtru cyfrowego jest przy tym rozumiane bardzo szeroko. Obejmuje ono nie tylko cały zespół środków sprzętowych i programowych przetwarzania, ale także algorytm, według którego jest ono dokonywane. W tym znaczeniu filtrem cyfrowym może być zarówno urządzenie fizyczne, jak i program obliczeniowy. W większości przypadków filtr cyfrowy łączy w sobie obie te funkcje. W wyniku przetwarzania sygnału przez filtr cyfrowy na jego wyjściu otrzymujemy inny sygnał, również reprezentowany ciągiem słów binarnych. Zwykle interesuje nas postać analogowa sygnału wyjściowego, a więc sygnał na wyjściu filtru cyfrowego musi być jeszcze przetworzony na sygnał analogowy. Operację tę realizuje przetwornik C/A.

O jakości i zakresie zastosowań filtru cyfrowego decyduje przede wszystkim jego szybkość działania. Określa ona graniczną częstotliwość sygnałów jakie mogą być przez niego przetwarzane. W wyniku dynamicznego rozwoju współczesnej mikroelektroniki zakres częstotliwości stale rozszerza się. W chwili obecnej dysponujemy już układami cyfrowego przetwarzania sygnałów umożliwiającymi przetwarzanie sygnałów w czasie rzeczywistym z częstotliwościami rzędu kilkuset megaherców.

Filtry cyfrowe są obecnie coraz częściej realizowane z wykorzystaniem specjalizowanych układów mikroprocesorowych nazywanych *procesorami sygnałowymi*. Procesor sygnałowy jest uniwersalnym, programowalnym, cyfrowym układem arytmetycznym, wyposażonym w pamięć i przeznaczonym do szybkiego i sprawnego wykonywania różnorodnych operacji arytmetycznych na sygnałach binarnych, takich jak dodawanie, mnożenie, mnożenie skalarne, opóźnianie sygnałów w czasie itd. W pamięci procesora są ponadto rejestrowane wszelkie dane niezbędne do wykonania algorytmu przetwarzania. Procesory sygnałowe stanowią coraz częściej standardowe wyposażenie współczesnej aparatury elektronicznej. W dniu dzisiejszym jest dostępna na rynku szeroka gama różnego typu procesorów sygnałowych realizowanych w technologii o wielkiej skali integracji, pracujących w arytmetyce stało- lub zmiennoprzecinkowej i różniących się pod względem mocy obliczeniowej i szybkości działania. W bardziej złożonych systemach przetwarzania stosowane są całe sieci współdziałających ze sobą procesorów sygnałowych, umożliwiających jednoczesną realizację wielu skomplikowanych procedur przetwarzania sygnałów.

6.1.4. Próbkowanie

Operacja próbkowania polega na pobieraniu próbek $x(t_n)$ analogowego sygnału $x(t)$ w dyskretnych chwilach t_n. W większości zastosowań praktycznych sygnały są próbkowanie równomiernie, choć znane są przykłady zastosowań, w których próbki sygnału są pobierane nierównomiernie w chwilach rozłożonych na osi czasu według ustalonej reguły, a w niektórych zastosowaniach – w chwilach losowych. Zgodnie z umową przyjętą w p. 1.3, rozważania nasze ograniczymy do przypadku próbkowania równomiernego.

W wyniku próbkowania równomiernego z okresem T_s (częstotliwością $f_s = 1/T_s$) sygnał analogowy $x(t)$ jest przetworzony w sygnał dyskretny $x[nT_s]$. Sygnał spróbkowany jest sygnałem dyskretnym w czasie, ale w ogólnym przypadku nadal analogowym w amplitudzie, tzn. jego wartości chwilowe (próbki) należą do zbioru ciągłego. Aby sygnał spróbkowany można było dalej przetwarzać cyfrowo, zbiór wartości próbek musi być zbiorem skończonym.

6.1.5. Kwantowanie

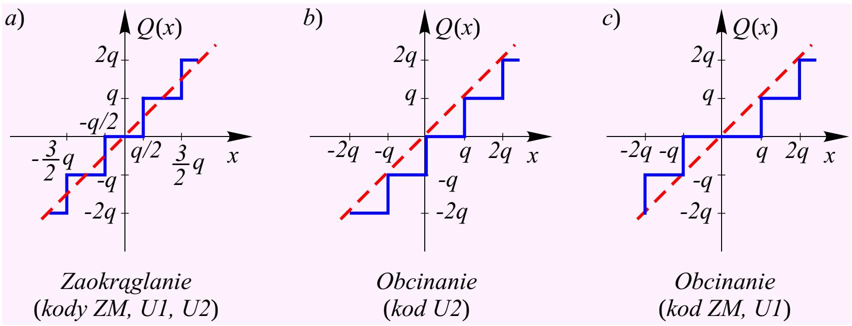
Operacja, która przetwarza sygnał spróbkowany w sygnał o dyskretnej strukturze amplitudowej jest nazywana *kwantowaniem*. Polega ona na podzieleniu zakresu zmian wartości sygnału na skończoną liczbę M *przedziałów kwantyzacji* i przybliżeniu wartości chwilowych próbek wartościami przyporządkowanymi poszczególnym przedziałom. Najczęściej przedziały kwantyzacji mają jednakową szerokość q, nazywaną *kwantem* lub *krokiem kwantowania*. Liczbę M wybiera się z reguły jako naturalną potęgę liczby 2, tj. $M = 2^b$, gdzie $b \in \mathscr{N}$.

W wyniku kwantowania sygnał dyskretny $x[nT_s]$ zostaje przybliżony sygnałem cyfrowym $\tilde{x}[nT_s]$ przybierającym skończoną liczbę wartości. Operację kwantowania można zapisać formalnie w postaci:

$$\tilde{x}[nT_s] = Q(x[nT_s]), \tag{6.1}$$

gdzie Q jest funkcją przyporządkowującą próbce $x(nT_s)$ jej wartość skwantowaną $\tilde{x}(nT_s)$. Dobór funkcji Q określa sposób kwantowania.

W praktyce stosowane są różne rodzaje kwantowania zależne od sposobu cyfrowej reprezentacji liczb ujemnych. Liczby ujemne w arytmetyce stałoprzecinkowej przedstawia się w komputerze za pomocą znaku i modułu (kod ZM), uzupełnienia do jedności (kod U1) lub uzupełnienia do dwóch (kod U2) (por. np. [1], p. 3.4.1 lub [2], rozdz. 9). Na rys. 6.3 przedstawiono trzy najczęściej stosowane sposoby kwantowania. Metoda kwantowania przedstawiona na rys. 6.3a jest nazywana *zaokrąglaniem*. W tym przypadku wartości próbek z przedziału $-q/2 < x \leqslant q/2$ są przybliżane wartością zerową, wartości próbek z przedziału $q/2 < x \leqslant 3q/2$ – wartością q itd. Zaokrąglenie może być stosowane zarówno w przypadku kodu ZM, jak i kodów U1 oraz U2. Na rys. 6.3b,c przedstawiono dwie wersje kwantowania nazywanego *obcinaniem*. Metoda zilustrowana na rys. 6.3b jest stosowana w przypadku kodu U2, natomiast metoda zilustrowana na rys. 6.3c – w przypadku kodów ZM oraz U1. Należy podkreślić, że bez względu na sposób kwantowania, operacja ta jest operacją nieliniową, tzn. sygnał cyfrowy otrzymany w wyniku skwantowania sumy dwóch sygnałów dyskretnych nie jest równy sumie sygnałów cyfrowych otrzymanych w wyniku kwantowania sygnałów składowych.



Rys. 6.3. Podstawowe sposoby kwantowania sygnałów

Operacja kwantowania wprowadza specyficzny błąd do procesu przetwarzania sygnału nazywany *błędem kwantowania*:

$$\varepsilon[nT_s] = \tilde{x}[nT_s] - x[nT_s]. \tag{6.2}$$

Błąd kwantowania jest sygnałem dyskretnym określonym w chwilach próbkowania nT_s i przybierającym losowe wartości w skończonym przedziale o szerokości równej kwantowi q. Dla trzech sposobów kwantowania przedstawionych na rys. 6.3 przedział ten jest równy odpowiednio: $[-q/2, q/2]$ w przypadku zaokrąglania, $[-q, 0]$ w przypadku pierwszego sposobu ucinania oraz $[-q, 0]$, jeśli $x[nT_s] > 0$, i $[0, q]$, jeśli $x[nT_s] < 0$, w przypadku drugiego sposobu ucinania. Ze względu na nieliniowy charakter operacji kwantowania analiza błędu kwantowania jest trudna, a jego właściwości silnie zależą od typu sygnału kwantowanego. Analizę błędu kwantowania przeprowadza się zwykle metodami probabilistycznymi, tzn. sygnał błędu traktuje się jako sygnał losowy. Ponieważ wykazuje on właściwości zbliżone do typowych sygnałów szumowych, jest nazywany *szumem kwantowania*. Szum kwantowania scharakteryzujemy nieco dokładniej w p. 6.7.3.

6.1.6. Kodowanie

W wyniku kwantowania sygnału dyskretnego przedział zmian jego wartości zostaje podzielony na $M = 2^b$ przedziałów kwantyzacji. Przedziały te można wówczas zakodować słowami binarnymi o długości b. Jeżeli sygnał zmienia się w zakresie od $-X_m$ do X_m, to wielkość kwantu (różnica między sąsiednimi poziomami kwantyzacji) jest określona wzorem:

$$q = \frac{2X_m}{2^b} = \frac{X_m}{2^{b-1}}. \tag{6.3}$$

Na przykład, dla typowych przetworników A/C zakres zmian sygnału napięcia na ich wejściu wynosi od -1 V do $+1$ V. Jeżeli sygnał zostanie skwantowany na $2^8 = 256$ przedziałów ($b = 8$), to wielkość kwantu wynosi $7,81$ mV. Wielkość ta jest określona przez najmniej znaczący znak binarny (bit) słowa kodowego.

Podkreślmy, że operacja kwantowania nie jest realizowana w przetworniku A/C przez specjalny układ, a dokonuje się niejako automatycznie w wyniku kodowania kolejnych próbek. W zależności od wartości próbki układ formowania słów kodowych generuje ciąg znaków binarnych słowa kodowego, ucinając go na najmniej znaczącym znaku. Jeśli wartość próbki przekroczy określony poziom, na ostatniej pozycji słowa kodowego występuje „1”. W przeciwnym przypadku na pozycji tej wystąpi „0”

Sposób przyporządkowania słów kodowych poszczególnym przedziałom kwantyzacji określa przyjętą metodę reprezentacji danych numerycznych w układzie przetwarzania. W cyfrowym przetwarzaniu sygnałów stosowane są różne reprezentacje danych liczbowych. Ogólnie można je podzielić na *stałoprzecinkowe* i *zmiennoprzecinkowe*. Wybór danej reprezentacji ma bardzo istotny wpływ na dokładność obliczeń i złożoność implementacji programowej algorytmów przetwarzania. Problematyka ta jest bardzo obszerna i wykracza poza główny nurt naszych rozważań. Można się z nią zapoznać w [2], rozdz. 9.

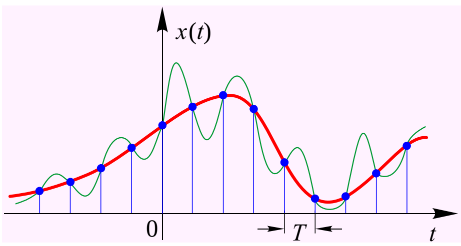
Osobnym zagadnieniem jest fizyczny sposób reprezentacji znaków binarnych „1” oraz „0” w postaci tzw. *kodu impulsowego*. Znak binarny „1” może być np. kodowany krótkim impulsem, a znak binarny „0” – brakiem impulsu. Innym sposobem jest kodowanie znaków binarnych „1” i „0” dwoma różnymi poziomami napięcia. Stosowane są również metody kodowania znaków binarnych za pomocą przejść, tj. zmian poziomów napięcia. Najczęściej stosowane reprezentacje znaków binarnych „1” i „0” za pomocą impulsów elektrycznych zostaną omówione w p. 12.3.5.

6.2. Twierdzenie o próbkowaniu. Wersja podstawowa

Próbkowanie jest operacją, w wyniku której pozyskujemy informację o wartościach sygnału analogowego w chwilach próbkowania. Nie zachowujemy przy tym bezpośredniej informacji o zachowaniu się sygnału między tymi chwilami. Pojawia się jednak pytanie o znaczeniu zasadniczym: czy, a jeśli tak, to przy spełnieniu jakich warunków, możemy odtworzyć sygnał analogowy na podstawie informacji zachowanej w jego próbkach? Mówiąc dokładniej, powstaje problem: czy, i przy jakich założeniach, znajomość próbek wystarcza do wyznaczenia z pełną dokładnością wszystkich pozostałych wartości sygnału między chwilami próbkowania? Problem ten rozstrzyga twierdzenie o próbkowaniu, będące jednym z fundamentalnych twierdzeń teorii sygnałów.

6.2.1. Przypadek ogólny

Intuicja podpowiada nam, że odpowiedź na postawione wyżej pytanie jest w ogólnym przypadku negatywna. Gdyby na podstawie zbioru próbek można było zawsze z pełną dokładnością odtworzyć sygnał pierwotny, oznaczałoby to, że przez próbki (rys. 6.4) można przeprowadzić jego wykres tylko w jeden możliwy sposób. Jest jednak oczywiste, że – przy braku jakichkolwiek założeń ograniczających – próbki sygnału można połączyć na nieprzeliczalnie wiele sposobów. Intuicja nas tutaj nie zawodzi. W ogólnym przypadku, bez dodatkowych założeń dotyczących sygnału i sposobu jego próbkowania, nie jest możliwe jego odtworzenie z próbek z pełną dokładnością.



Rys. 6.4. Niejednoznaczność odtworzenia sygnału na podstawie próbek

6.2.2. Sygnały o ograniczonym paśmie

Inaczej przedstawia się sytuacja, gdy spełnione są pewne warunki ograniczające. Pierwszy z nich dotyczy struktury widmowej sygnału, drugi zaś, związany ściśle z pierwszym – częstotliwości z jaką jest on próbkowany.

Jeżeli sygnał nie zmienia się między próbkami zbyt szybko, to – mówiąc obrazowo – na odcinki łączące dwie sąsiednie próbki są narzucone pewne więzy. Na więzy te wpływają również wartości dalszych próbek. Są one tym silniejsze, im bliżej siebie są położone kolejne próbki. Jeśli próbki są pobierane dostatecznie często, więzy te okazują się tak silne, że określają jednoznacznie przebieg sygnału między próbkami. Warunek ograniczonej szybkości zmian sygnału oznacza, że jego widmo nie zawiera składowych, których częstotliwości przekraczają pewną wartość graniczną. Narzucenie tego warunku prowadzi do wydzielenia klasy sygnałów o ograniczonym paśmie.

Definicja 6.1. Sygnał $x(t)$ o widmie $X(\omega)$ nazywamy sygnałem *o ograniczonym paśmie*, jeżeli istnieje skończona wartość pulsacji ω_m, taka że $X(\omega) \equiv 0$ dla $|\omega| > \omega_m$. Pulsację ω_m (częstotliwość $f_m = \omega_m/2\pi$) nazywamy *pulsacją* (*częstotliwością*) *graniczną* pasma tego sygnału.

Sygnałami o ograniczonym paśmie mogą być zarówno sygnały o ograniczonej energii, jak i ograniczonej mocy średniej. Do sygnałów o ograniczonym paśmie należy m.in. idealny sygnał dolnopasmowy (3.38), idealny sygnał wąskopasmowy, sygnały harmoniczne (3.49)-(3.51), a ogólniej wszystkie sygnały okresowe mające skończone rozwinięcia w szereg Fouriera (2.57).

6.2.3. Sformułowanie twierdzenia

Rozważmy dowolny sygnał $x(t)$ o paśmie ograniczonym pulsacją ω_m. W podstawowej wersji twierdzenia przyjmiemy założenie, że widmo $X(\omega)$ sygnału $x(t)$ nie zawiera składowych częstotliwościowych dla $|\omega| \geqslant \omega_m$.

Twierdzenie 6.1 (*wersja podstawowa twierdzenia o próbkowaniu*)
Jeżeli widmo $X(\omega)$ sygnału $x(t)$ spełnia warunek $X(\omega) \equiv 0$ dla $|\omega| \geqslant \omega_m$, to sygnał ten można odtworzyć z pełną dokładnością na podstawie jego próbek pobieranych z okresem $T_s \leqslant \pi/\omega_m$.

Twierdzenie o próbkowaniu sformułował i dowiódł po raz pierwszy Kotielnikow w 1933 r. W jego ujęciu stanowiło ono jeden z matematycznych rezultatów teorii całkowego przekształcenia Fouriera. Sformułowanie przybliżające to twierdzenie do zastosowań praktycznych podał w 1949 r. Shannon. Z tego względu w literaturze jest ono nazywane *twierdzeniem Kotielnikowa-Shannona*.

Z twierdzenia o próbkowaniu wynika, że jeśli tylko sygnał o ograniczonym paśmie jest próbkowany dostatecznie często, to w zbiorze pobranych próbek zostaje zachowana o nim pełna informacja. Warunek $T_s \leqslant \pi/\omega_m$ jest nazywany *warunkiem Nyquista*, a największy okres próbkowania $T_s = \pi/\omega_m$, przy którym twierdzenie o próbkowaniu jest spełnione – *przedziałem Nyquista*. Warunek Nyquista można także zapisać w postaci $f_s \geqslant \omega_{m/}\pi$, gdzie $f_s = 1/T_s$ jest częstotliwością próbkowania, lub w najczęściej cytowanej w literaturze postaci:

$$f_s \geqslant 2f_m. \tag{6.4}$$

Ten ostatni warunek oznacza, że jeśli próbki sygnału są pobierane z częstotliwością co najmniej dwa razy większą niż częstotliwość graniczna pasma sygnału, to na ich podstawie można odtworzyć sygnał z pełną dokładnością. Najmniejsza częstotliwość $f_s = 2f_m$, przy której jest to możliwe, jest nazywana *częstotliwością Nyquista*.

Dowód twierdzenia 6.1 podawany najpierw dla przypadku próbkowania sygnału z częstotliwością Nyquista $f_s = 2f_m$. Dowód tego twierdzenia w przypadku ogólnym, gdy $f_s \geqslant 2f_m$, przeprowadzimy nieco później.

6.2.4. Wzór interpolacyjny Kotielnikowa-Shannona

W przypadku, gdy $f_s = 2f_m$ słuszne jest następujące twierdzenie.

Twierdzenie 6.2 (*twierdzenie o próbkowaniu z częstotliwością Nyquista*)
Niech $x(t)$ będzie dowolnym sygnałem, którego widmo spełnia warunek $X(\omega) \equiv 0$ dla $|\omega| \geqslant \omega_m$. Jeżeli sygnał ten jest próbkowany z okresem $T_s = \pi/\omega_m$ (częstotliwością $f_s = 2f_m$), to jego wartości między chwilami próbkowania można odtworzyć na podstawie próbek $x(nT_s)$ zgodnie ze wzorem:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{Sa}\,\omega_m(t - nT_s). \tag{6.5}$$

Wzór (6.5) jest nazywany *wzorem interpolacyjnym Kotielnikowa-Shannona*. Wynika z niego, że znając próbki $x(nT_s)$ sygnału analogowego $x(t)$, można obliczyć jego dokładną wartość w dowolnym punkcie t' między węzłami interpolacji, tj. między próbkami określonymi w punktach $t_n = nT_s$. W tym celu należy pomnożyć wartości próbek przez wartości w punkcie t' poprzesuwanych o kolejne odcinki czasu nT_s kopii funkcji Sa i otrzymane wyniki zsumować dla wszystkich n.

6.2.5. Dowód twierdzenia 6.2 dla klasy sygnałów o ograniczonej energii

W przypadku sygnałów o ograniczonej energii, należących do przestrzeni Hilberta $L^2(-\infty,\infty)$ sygnałów o paśmie ograniczonym pulsacją ω_m, dowód wynika wprost z rozwinięcia sygnału $x(t)$ w szereg Kotielnikowa-Shannona (por. wzór (2.64)):

$$x(t) = \sum_{n=-\infty}^{\infty} \alpha_n \frac{1}{\sqrt{T_s}}\,\mathrm{Sa}\,\frac{\pi}{T_s}(t - nT_s) \tag{6.6}$$

względem ortonormalnej bazy funkcji Sa:

$$x_n(t) = \frac{1}{\sqrt{T_s}}\,\mathrm{Sa}\,\frac{\pi}{T_s}(t-nT_s) = \sqrt{\frac{\omega_m}{\pi}}\,\mathrm{Sa}\,\omega_m(t-nT_s) \qquad n = 0, \pm 1, \ldots \tag{6.7}$$

Szereg (6.6) był rozpatrywany w p. 2.4.5 jako jeden z przykładów uogólnionego szeregu Fouriera. Aby wykazać słuszność twierdzenia 6.2, wystarczy uzupełnić rozważania tego punktu, dowodząc, że elementy bazy (6.7) są ortonormalne oraz że współczynniki funkcji Sa w rozwinięciu w szereg Kotielnikowa-Shannona są równe próbkom sygnału.

Zauważmy przede wszystkim, że elementy bazy (6.7) mają jednakowe normy, gdyż norma sygnału nie ulega zmianie w wyniku jego przesunięcia w czasie. Aby wykazać, że baza (6.7) jest ortonormalna, wystarczy zatem pokazać, że element $x_0(t) = \sqrt{\omega_m/\pi}\,\mathrm{Sa}\,\omega_m t$ jest ortonormalny do każdego z pozostałych jej elementów $x_n(t) = \sqrt{\omega_m/\pi}\,\mathrm{Sa}\,\omega_m(t-nT_s)$, $n = \pm 1, \pm 2, \ldots$. Dowód przeprowadzimy, korzystając z uogólnionego wzoru Rayleigha (3.33). Zgodnie z parą transformat (3.38) oraz twierdzeniem o przesunięciu w dziedzinie czasu (3.22) mamy:

$$X_0(\omega) = \mathscr{F}[x_0(t)] = \sqrt{\frac{\pi}{\omega_m}}\,\sqcap\left(\frac{\omega}{2\omega_m}\right), \tag{6.8}$$

$$X_n(\omega) = \mathscr{F}[x_n(t)] = X_0(\omega)\,\mathrm{e}^{-\mathrm{j}nT_s\omega}, \tag{6.9}$$

Uwzględniając wzory (6.8) i (6.2.5) we wzorze (3.33), otrzymujemy:

$$\begin{aligned}(x_0, x_n)_{L_t^2} &= \tfrac{1}{2\pi}(X_0, X_n)_{L_\omega^2} = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} X_0(\omega)X_n^*(\omega)\,\mathrm{d}\omega = \\ &= \tfrac{1}{2\pi}\int_{-\omega_m}^{\omega_m} \tfrac{\pi}{\omega_m}\,\mathrm{e}^{\mathrm{j}nT_s\omega}\,\mathrm{d}\omega = \tfrac{1}{2\mathrm{j}nT_s\omega_m}\,\mathrm{e}^{\mathrm{j}nT_s\omega}\Big|_{-\omega_m}^{\omega_m} = \\ &= \mathrm{Sa}(nT_s\omega_m) = \mathrm{Sa}(n\pi) = \begin{cases} 0 & \text{dla} \quad n \neq 0, \\ 1 & \text{dla} \quad n = 0, \end{cases}\end{aligned} \tag{6.10}$$

co dowodzi ortonormalności bazy (6.7).

Współczynniki α_n szeregu (6.6) są określone przez iloczyny skalarne (por. wzór (2.30)):

$$\alpha_n = (x, x_n) = \left(x(t), \sqrt{\omega_m/\pi}\,\mathrm{Sa}\,\omega_m(t - nT_s)\right). \tag{6.11}$$

W celu ich obliczenia skorzystamy ponownie z uogólnionego wzoru Rayleigha (3.33). Uwzględniając w tym wzorze wzory (6.8) i (6.2.5), otrzymujemy:

$$\begin{aligned}\alpha_n = (x, x_n)_{L_t^2} = \frac{1}{2\pi}(X, X_n)_{L_\omega^2} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)X_n^*(\omega)\,\mathrm{d}\omega = \\ &= \sqrt{\frac{\pi}{\omega_m}}\left[\frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} X(\omega)\,\mathrm{e}^{\mathrm{j}nT_s\omega}\,\mathrm{d}\omega\right].\end{aligned}$$

Zauważmy, że zgodnie ze wzorem (3.4), określającym odwrotną transformatę Fouriera, wyrażenie w nawiasie kwadratowym jest równe wartości sygnału $x(t)$ w punkcie nT_s, a więc jego próbce w tym punkcie. Otrzymujemy zatem równość:

$$\alpha_n = \sqrt{\frac{\pi}{\omega_m}} x(nT_s) = \sqrt{T_s} x(nT_s), \tag{6.12}$$

która po uwzględnieniu w szeregu (6.6) dowodzi słuszności wzoru (6.5).

Podany dowód jest słuszny tylko dla sygnałów o ograniczonej energii. W przypadku dowolnych sygnałów, a więc także sygnałów o ograniczonej mocy, dowód jest oparty na odmiennym rozumowaniu, w którym wykorzystuje się dystrybucyjną reprezentację spróbkowanego sygnału $x(t)$ za pomocą impulsowego sygnału spróbkowanego $x_s(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s)$ (por. rys. 1.27) oraz pojęcie *idealnego analogowego filtru dolnoprzepustowego*. Formalna definicja idealnego dolnoprzepustowego filtru analogowego oraz sposoby jego matematycznego opisu zostaną omówione dokładnie w 7.5.4. Dla celów dowodu twierdzenia o próbkowaniu ograniczymy się w tym miejscu jedynie do scharakteryzowania jego elementarnych właściwości.

6.2.6. Idealny filtr dolnoprzepustowy

Idealnym filtrem dolnoprzepustowym nazywamy układ, który przetwarza sygnał wejściowy $x(t)$ w sygnał wyjściowy $y(t)$ w taki sposób, że przenosi bez zmian wszystkie składowe częstotliwościowe sygnału wejściowego do pewnej pulsacji granicznej ω_g i usuwa z tego sygnału składowe o pulsacjach większych od ω_g. Widmo $Y(\omega)$ sygnału na wyjściu idealnego filtru dolnoprzepustowego jest więc związane z widmem $X(\omega)$ sygnału wejściowego zależnością:

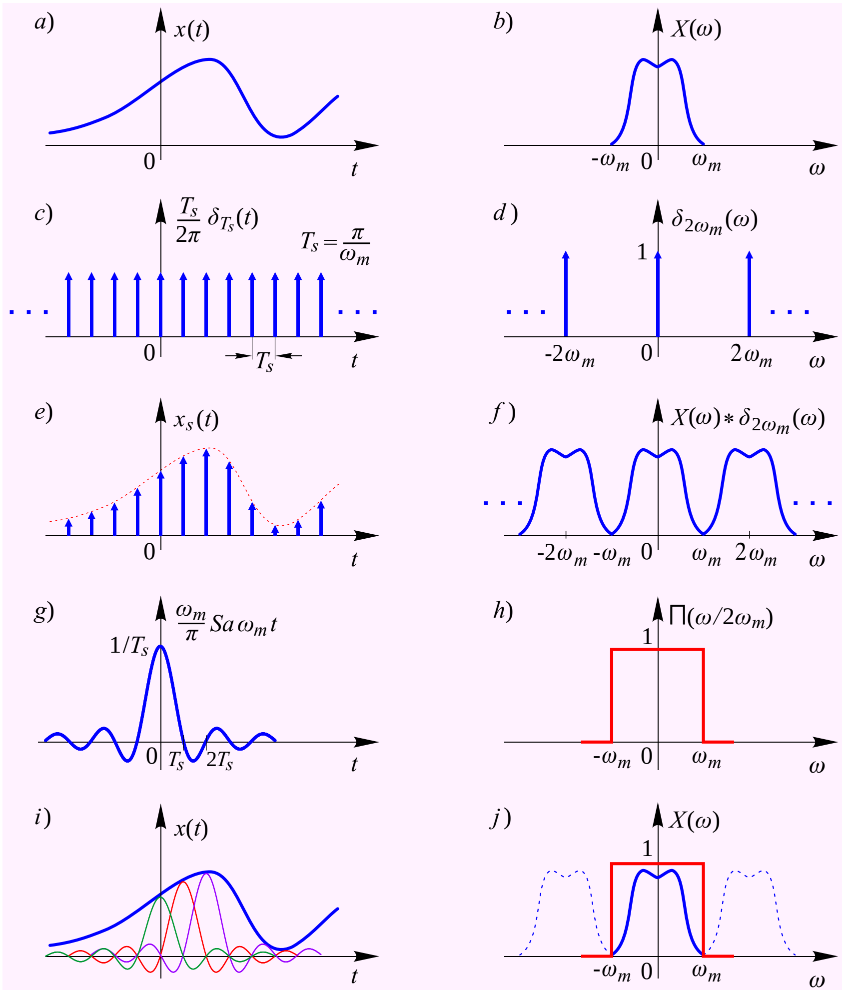
$$Y(\omega) = X(\omega) \sqcap \left(\frac{\omega}{2\omega_g}\right). \tag{6.13}$$

Ponieważ mnożeniu w dziedzinie częstotliwości odpowiada splot w dziedzinie czasu, sygnał na wyjściu filtru jest związany z sygnałem wejściowym zależnością splotową (por. parę transformat (3.38)):

$$y(t) = x(t) * \mathscr{F}^{-1}\left[\sqcap\left(\frac{\omega}{2\omega_g}\right)\right] = x(t) * \frac{\omega_g}{\pi} \operatorname{Sa} \omega_g t. \tag{6.14}$$

6.2.7. Dowód twierdzenia 6.2 dla dowolnych sygnałów

Kolejne kroki dowodu są zilustrowane na rys. 6.5. Po lewej stronie są wykreślone odpowiednie sygnały, po prawej zaś – ich widma. Dla wygody ilustracji zakładamy, że widma są rzeczywiste. Wykresy umieszczone naprzeciwko siebie tworzą pary transformat Fouriera.



Rys. 6.5. Ilustracja dowodu twierdzenia o próbkowaniu z częstotliwością Nyquista $f_s = 2f_m$

Zauważmy, że jeżeli widmo $X(\omega)$ sygnału $x(t)$ o paśmie ograniczonym pulsacją ω_m (rys. 6.5b) powielimy okresowo z okresem $2\omega_m$ (rys. 6.5f), a następnie z powielonego widma odfiltrujemy za pomocą idealnego filtru dolnoprzepustowego o pulsacji granicznej $\omega_g = \omega_m$ powielone kopie (rys. 6.5h), to otrzymamy ponownie widmo pierwotne $X(\omega)$ (rys. 6.5j). Wynika stąd, że operacje powielenia okresowego widma i idealnej filtracji dolnoprzepustowej są operacjami odwrotnymi względem siebie. Powielenie okresowe widma $X(\omega)$ uzyskujemy w wyniku splecenia go z dystrybucją grzebieniową $\delta_{2\omega_m}(\omega)$ o okresie $2\omega_m$ (rys. 6.5d), a więc operacje te możemy zapisać wzorem:

$$X(\omega) = [X(\omega) * \delta_{2\omega_m}(\omega)] \sqcap \left(\frac{\omega}{2\omega_m}\right). \tag{6.15}$$

Dokonamy teraz odwrotnego przekształcenia Fouriera obu stron równości (6.15), uwzględniając przy tym kolejno: twierdzenie o splocie w dziedzinie czasu (3.30), twierdzenie o splocie w dziedzinie częstotliwości (3.31), pary transformat Fouriera (3.38) i (3.54) oraz właściwość splotu dystrybucji Diraca. Otrzymamy:

$$\begin{aligned} x(t) &= \left[2\pi x(t)\frac{1}{2\omega_m}\delta_{T_s}(t)\right] * \frac{\omega_m}{\pi}\operatorname{Sa}\omega_m t = [x(t)\delta_{T_s}(t)] * \operatorname{Sa}\omega_m t = \\ &= \left[\sum_{n=-\infty}^{\infty} x(nT_s)\delta(t-nT_s)\right] * \operatorname{Sa}\omega_m t = \sum_{n=-\infty}^{\infty} x(nT_s)\operatorname{Sa}\omega_m(t-nT_s). \end{aligned} \tag{6.16}$$

Ostatnia równość dowodzi tezę twierdzenia 6.2.

6.3. Próbkowanie sygnału z dowolną częstotliwością

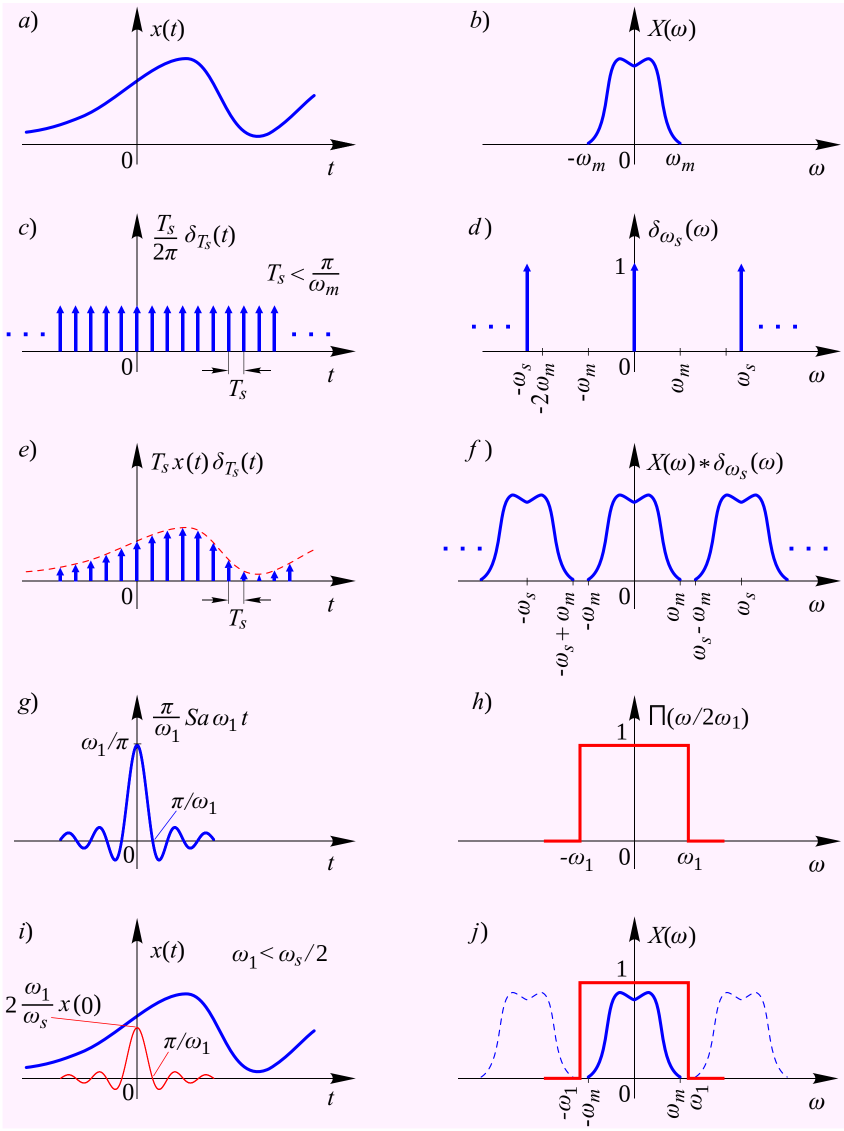
6.3.1. Przypadek $f_s \geqslant 2f_m$. Dowód twierdzenia 6.1

Stosując analogiczne rozumowanie jak w p. 6.2.7, wyprowadzimy obecnie wzór interpolacyjny określający sposób odtwarzania sygnału z próbek pobieranych z dowolną częstotliwością większą od częstotliwości Nyquista. Udowodnimy tym samym tezę twierdzenia 6.1. Dowód jest zilustrowany na rys. 6.6.

Próbkowanie sygnału z częstotliwością większą od częstotliwości Nyquista oznacza, że próbki są pobierane z okresem $T_s < \pi/\omega_m$ (rys. 6.6c,e). W wyniku tego widmo sygnału zostaje powielone okresowo z okresem $\omega_s = 2\pi/T_s > 2\omega_m$ (rys. 6.6d,f). Pomiędzy powielonymi kopiami widma występują wówczas pasma puste o szerokości $\omega_s - 2\omega_m$ (rys. 6.6f). Widmo pierwotne można zatem odtworzyć bez zniekształceń z widma powielonego, stosując idealny filtr dolnoprzepustowy o częstotliwości granicznej ω_1, która może przybierać dowolną wartość z przedziału $\omega_m \leqslant \omega_1 \leqslant \omega_s - \omega_m$ (rys. 6.6h,j). Operacje wykonane w dziedzinie częstotliwości na widmie sygnału można w tym przypadku zapisać wzorem:

$$X(\omega) = [X(\omega) * \delta_{\omega_s}(\omega)] \sqcap\left(\frac{\omega}{2\omega_1}\right), \tag{6.17}$$

skąd po przejściu do dziedziny czasu otrzymujemy:



Rys. 6.6. Ilustracja dowodu twierdzenia o próbkowaniu w przypadku $f_s \geqslant 2f_m$

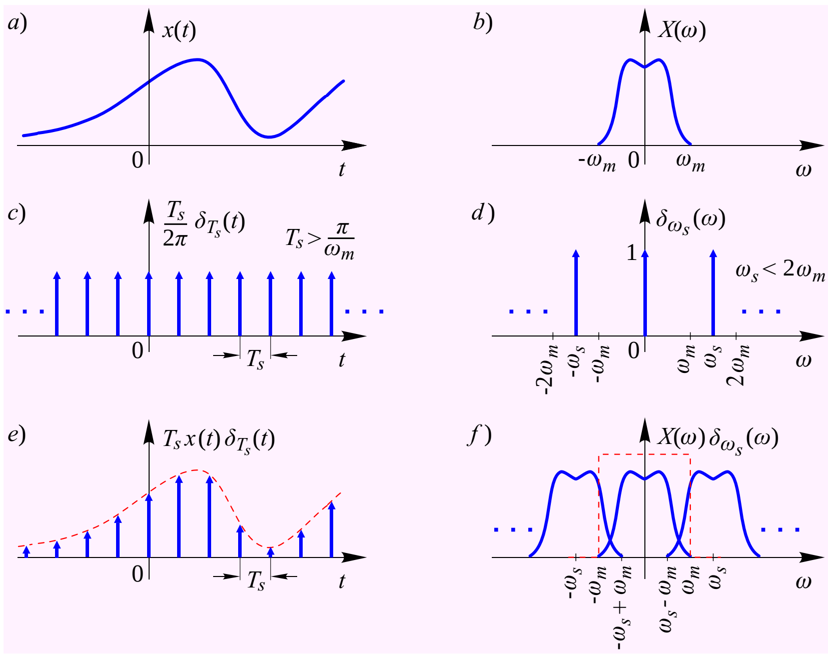
$$x(t) = \left[2\pi x(t)\frac{1}{\omega_s}\delta_{T_s}(t)\right] * \frac{\omega_1}{\pi}\,\mathrm{Sa}\,\omega_1 t = \frac{2\omega_1}{\omega_s}\,[x(t)\delta_{T_s}(t)] * \mathrm{Sa}\,\omega_1 t =$$

$$= \frac{2\omega_1}{\omega_s}\sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{Sa}\,\omega_1(t - nT_s). \quad (6.18)$$

Ze wzoru (6.18) wynika, że w celu odtworzenia sygnału należy przemnożyć przeskalowane w stosunku $2\omega_1/\omega_s$ próbki $x(nT_s)$ przez kolejne przesunięte o nT_s kopie sygnału Sa o pulsacji oscylacji określonej pulsacją graniczną filtru ω_1.

6.3.2. Przypadek $f_s < 2f_m$. Aliasing

Jeśli sygnał jest próbkowany z częstotliwością mniejszą od częstotliwości Nyquista (rys. 6.7c,e), powielone okresowo kopie widma nakładają się na siebie (rys. 6.7f). Odzyskanie niezniekształconego widma $X(\omega)$ jest w takim przypadku niemożliwe. Efekt nakładania się sąsiadujących kopii widma jest nazywany *aliasingiem*. Aliasing powoduje, że w wyniku filtracji idealnym filtrem dolnoprzepustowym o pulsacji granicznej ω_m (jego charakterystyka jest wykreślona na rys. 6.7f linią przerywaną) odzyskujemy sygnał $\tilde{x}(t)$, zniekształcony w porównaniu z sygnałem pierwotnym $x(t)$. Widmo tego sygnału zawiera także części widma pochodzące z sąsiednich powielonych kopii. Popełniany przy tym błąd jest nazywany *błędem aliasingu*. Błąd ten jest tym większy, im mniejsza jest częstotliwość próbkowania. Do kwestii tej powrócimy jeszcze w p. 6.5.2.



Rys. 6.7. Ilustracja twierdzenia o próbkowaniu w przypadku $f_s < 2f_m$

6.4. Odtwarzanie sygnału z próbek

Z teoretycznego punktu widzenia sposób odtwarzania sygnału z próbek określają wzory interpolacyjne: wzór Kotielnikowa-Shannona (6.5) – jeśli próbki są pobierane z częstotliwością Nyquista $f_s = 2f_m$ lub jego ogólniejsza wersja (6.18) – jeśli próbki są pobierane z dowolną częstotliwością $f_s \geqslant 2f_m$. Posługując się tablicą wartości funkcji Sa, można na podstawie tych wzorów wyznaczać numerycznie wartość sygnału w dowolnej chwili między chwilami próbkowania.

Wzory interpolacyjne są silnymi wynikami teoretycznymi, jednak ich znaczenie praktyczne jest niewielkie. Ze względu na nieskończoną liczbę sumowań, jakie należy wykonać, oraz skończoną, z konieczności, rozdzielczość tablicy wartości funkcji Sa, odtworzenie sygnału tą metodą wymaga znacznego nakładu obliczeniowego i może być przeprowadzone jedynie z pewnym przybliżeniem.

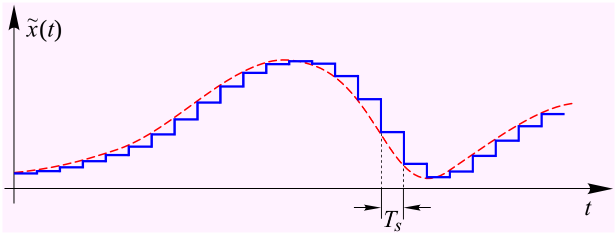
W praktyce operacja odtwarzania sygnału z próbek jest dokonywana układowo. W zależności od zastosowania jest ona realizowana w różny sposób. W zagadnieniach cyfrowego przetwarzania sygnałów odtwarzanie sygnału z próbek jest realizowane za pomocą przetwornika C/A oraz, w razie potrzeby, korekcyjnego filtru dolnoprzepustowego. W tzw. impulsowych systemach transmisji sygnałów (por. p.12), odtwarzanie odbywa się drogą filtracji dolnoprzepustowej.

6.4.1. Odtwarzanie sygnału przez przetwornik C/A. Metoda schodkowa

W systemach cyfrowego przetwarzania sygnałów sygnał analogowy $x(t)$ jest reprezentowany słowami binarnymi otrzymanymi np. na wyjściu procesora sygnałowego. W celu odtworzenia sygnału $x(t)$ na podstawie reprezentujących go słów są one podawane na wejście przetwornika C/A, którego zadaniem jest ich przetworzenie na sygnał $\tilde{x}(t)$, możliwie wiernie przybliżający sygnał faktyczny $x(t)$.

W przetworniku C/A na podstawie kolejnych słów binarnych są formowane próbki. Każdej próbce odpowiada określony poziom napięcia. Z próbek jest następnie formowany sygnał analogowy na wyjściu przetwornika. Najprostszym sposobem wytworzenia tego sygnału jest podtrzymanie napięcia w danym przedziale próbkowania na stałym poziomie, odpowiadającym wartości bieżącej próbki. Poziom ten jest utrzymywany do chwili uformowania nowej próbki. W chwili tej następuje skokowa zmiana napięcia do poziomu odpowiadającego wartości nowej próbki. Na wyjściu przetwornika zostaje tym samym wytworzony przedziałami stały sygnał napięciowy $\tilde{x}(t)$, nazywany *sygnałem schodkowym* (rys. 6.8).

Opisana metoda odtwarzania sygnału z próbek jest nazywana *ekstrapolacją zerowego rzędu* lub prościej *metodą schodkową*. Metoda schodkowa nie jest zbyt dokładna, ale łatwa do realizacji technicznej i z tego względu często stosowana w typowych rozwiązaniach przetworników C/A. Błąd odtwarzania sygnału



Rys. 6.8. Metoda schodkowa odtwarzania sygnału z próbek

metodą schodkową jest tym mniejszy, im mniejszy jest przedział dyskretyzacji sygnału T_s. Schodki są wówczas krótsze i sygnał schodkowy $\tilde{x}(t)$ wierniej przybliża sygnał odtwarzany $x(t)$. Małe błędy odtwarzania możemy zatem uzyskać, próbkując sygnały z częstotliwością znacznie większą od częstotliwości Nyquista. Błąd odtwarzania sygnału pociąga oczywiście za sobą zniekształcenia jego widma. Z analizy tych zniekształceń (przeprowadzimy ją w punkcie następnym) wynikać będzie sposób korekcji błędów odtwarzania.

Znacznie większą dokładność odtwarzania sygnału z próbek możemy uzyskać, stosując ekstrapolację pierwszego lub wyższych rzędów. W układzie ekstrapolatora pierwszego rzędu próbki sygnału są łączone odcinkami linii prostych, drugiego rzędu – odcinkami parabol itd. Zastosowanie ekstrapolatorów wyższych rzędów pociąga jednak za sobą znaczny wzrost złożoności układowej przetwornika C/A.

6.4.2. Zniekształcenia widma sygnału przy odtwarzaniu metodą schodkową

Sygnał schodkowy można traktować jako sumę przesuniętych o odcinki czasu nT_s impulsów prostokątnych o czasie trwania T_s i amplitudach równych wartościom kolejnych próbek. Jeżeli impuls prostokątny o jednostkowej amplitudzie i czasie trwania $0 \leqslant t < T_s$ oznaczymy:

$$p(t) = \sqcap\left(\frac{t - T_s/2}{T_s}\right) = \begin{cases} 1 & \text{dla} \quad 0 \leqslant t < T_s, \\ 0 & \text{dla pozostałych } t, \end{cases} \tag{6.19}$$

to sygnał schodkowy utworzony na podstawie ciągu próbek $x(nT_s)$ można zapisać w postaci:

$$\tilde{x}(t) = \sum_{n=-\infty}^{\infty} x(nT_s)p(t - nT_s). \tag{6.20}$$

Zauważmy, że sygnał (6.20) można przedstawić jako splot impulsowego sygnału spróbkowanego (3.56) z impulsem prostokątnym (6.19). Istotnie, z właściwości splotu dystrybucji Diraca (1.19) wynika, że:

$$p(t) * \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s) = \sum_{n=-\infty}^{\infty} x(nT_s)[p(t) * \delta(t - nT_s)] =$$
$$= \sum_{n=-\infty}^{\infty} x(nT_s)p(t - nT_s) = \tilde{x}(t).$$

Zgodnie z twierdzeniem o splocie w dziedzinie czasu widmo sygnału schodkowego (6.20) jest zatem iloczynem

$$\tilde{X}(\omega) = P(\omega)X_s(\omega) \tag{6.21}$$

widma $P(\omega)$ impulsu prostokątnego przesuniętego o $T_s/2$ (por. wzory (3.3) i (3.22)):

$$P(\omega) = T_s \,\mathrm{Sa}(\omega T_s/2)\,\mathrm{e}^{-\mathrm{j}(\omega T_s/2)}$$

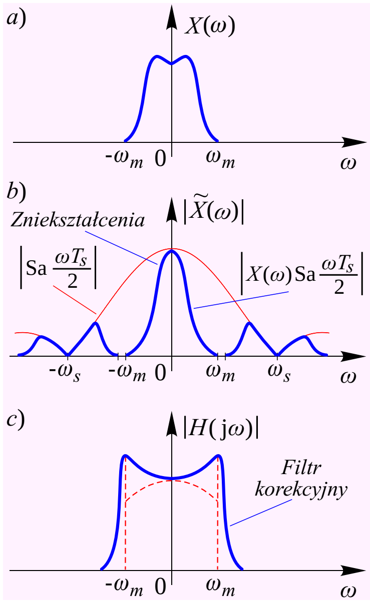
i widma $X_s(\omega)$ impulsowego sygnału spróbkowanego $x_s(t)$. Ponieważ widmo $X_s(\omega)$ jest (z dokładnością do współczynnika $1/T_s$) powieleniem okresowym z okresem $\omega_s = 2\pi/T_s$ widma $X(\omega)$ sygnału odtwarzanego $x(t)$ (por. wzór (3.58)):

$$X_s(\omega) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} X(\omega - n\omega_s), \qquad \omega_s = 2\pi/T_s,$$

zatem sygnału schodkowego przybiera postać:

$$\tilde{X}(\omega) = \mathrm{Sa}(\omega T_s/2)\,\mathrm{e}^{-\mathrm{j}(\omega T_s/2)} \sum_{n=-\infty}^{\infty} X(\omega - n\omega_s). \tag{6.22}$$

Na rys. 6.9b przedstawiono wykres widma amplitudowego $|\tilde{X}(\omega)|$ sygnału schodkowego w przypadku, gdy widmo amplitudowe $|X(\omega)|$ sygnału odtwarzanego ma kształt pokazany na rys. 6.9a. Z wykresu tego wynika charakter zniekształceń widma $|X(\omega)|$. Oprócz pożądanego segmentu centralnego, w widmie $|\tilde{X}(\omega)|$ występują segmenty boczne skupione wokół pulsacji $k\omega_s$. Całość jest zniekształcona multiplikatywnie obwiednią o kształcie funkcji Sa. W przypadku, gdy częstotliwość próbkowania f_s niewiele tylko przekracza częstotliwość Nyquista $2f_m$, zniekształcenie centralnego segmentu widma $|\tilde{X}(\omega)|$ w stosunku do widma $|X(\omega)|$ jest znaczne, zwłaszcza na krańcach pasma sygnału, tj. w pobliżu punktów $\pm\omega_m$. Przy zwiększaniu częstotliwości próbkowania, tj. krótszych schodkach, obwiednia Sa przybiera w przedziale $|\omega| \leqslant \omega_m$ bardziej płaski kształt i zniekształcenia segmentu centralnego są mniejsze. Jednocześnie segmenty boczne oddalają się od segmentu centralnego, a ich poziom maleje. Z reguły jednak częstotliwość próbkowania jest ustalona i zmniejszanie tą drogą zniekształceń towarzyszących odtwarzaniu sygnału metodą schodkową jest niemożliwe.



Rys. 6.9. Dokładne widmo amplitudowe sygnału (a), widmo amplitudowe sygnału odtworzonego metodą schodkową (b) oraz charakterystyka filtru wygładzającego (c)

Praktyczny sposób korekcji błędów powstających przy odtwarzaniu sygnału z próbek metodą schodkową wynika wprost z wykresu widma amplitudowego sygnału schodkowego. Jak widać, w celu korekcji widma sygnału odtworzonego, na wyjściu przetwornika C/A należy zastosować filtr dolnoprzepustowy o charakterystyce filtracji ukształtowanej w taki sposób, aby w przedziale $|\omega| < \omega_m$ przybierała ona kształt odwrotności funkcji Sa (rys. 6.9c). Filtr ten usuwa boczne segmenty widma $|\tilde{X}(\omega)|$ i zarazem koryguje zniekształcenia segmentu centralnego. W dziedzinie czasu efektem działania filtru korekcyjnego jest „wygładzenie" schodków sygnału odtworzonego. Z tego względu jest on nazywany niekiedy *filtrem wygładzającym*.

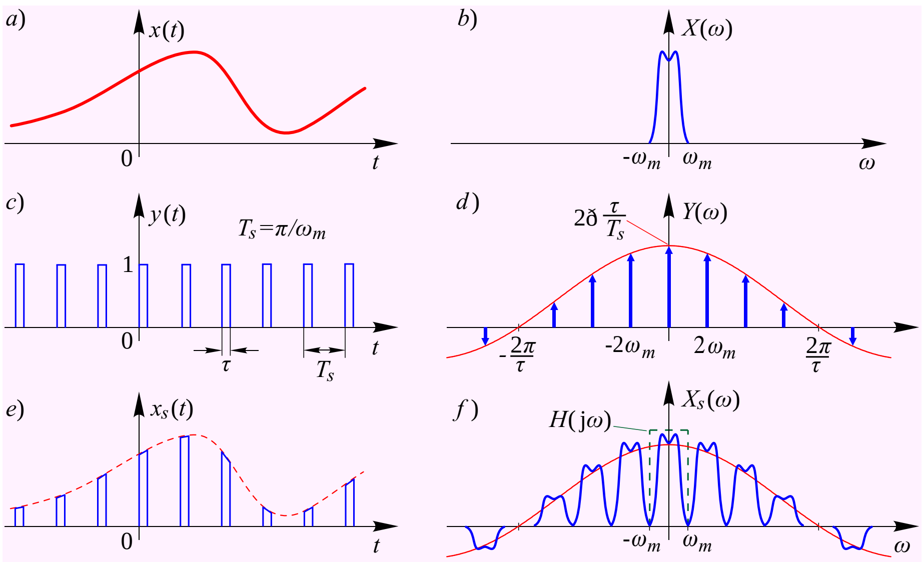
6.4.3. Odtwarzanie sygnału w impulsowych systemach transmisji sygnałów

Celem systemów transmisji sygnałów jest przesłanie na odległość możliwie dokładnej informacji o sygnale. Jednym z trzech podstawowych rodzajów systemów transmisyjnych są – obok systemów analogowych i systemów cyfrowych – systemy impulsowe transmisji sygnałów. W impulsowym systemie amplitudowym PAM (omówimy go dokładniej w p. 12.1) informacja o źródłowym sygnale analogowym (sygnale informacyjnym) $x(t)$ jest zakodowana w nadajniku w amplitudach odpowiednio ukształtowanych krótkich impulsów o stałym czasie trwania, które są następnie transmitowane do odbiornika co przedział próbkowania T_s. Teoretycznym modelem takiego ciągu impulsów jest impulsowy sygnał spróbkowany (3.56) powstały w wyniku mnożenia sygnału $x(t)$ przez dystrybucję grzebieniową o okresie T_s. Taki sposób tworzenia sygnału spróbkowanego jest nazywany *próbkowaniem idealnym*.

Ze względu na nierealizowalność fizyczną dystrybucji Diraca próbkowanie idealne ma znaczenie jedynie teoretyczne. W praktyce ciąg impulsów Diraca jest aproksymowany unipolarną falą prostokątną (rys. 6.10c) o okresie T_s, jednostkowej amplitudzie impulsów i czasie ich trwania $\tau \ll T_s$. Sygnał spróbkowany przybiera wówczas postać (rys. 6.10e):

$$x_s(t) = x(t) \sum_{n=-\infty}^{\infty} \sqcap\left(\frac{t - nT_s}{\tau}\right). \tag{6.23}$$

Zauważmy, że impulsy transmitowane w tym systemie nie są prostokątne, lecz w czasie trwania przybierają kształt sygnału $x(t)$. Taki sposób formowania próbek jest nazywany *próbkowaniem naturalnym*.



Rys. 6.10. Próbkowanie naturalne

6.4.4. Próbkowanie naturalne

Przeanalizujmy sposób odtwarzania po stronie odbiorczej sygnału $x(t)$ z sygnału $x_s(t)$ spróbkowanego naturalnie. W tym celu rozpatrzmy widmo $X_s(\omega)$ sygnału spróbkowanego, przy założeniu próbkowania z przedziałem Nyquista

$T_s = \pi/\omega_m$, gdzie ω_m jest maksymalną pulsacją widma $X(\omega)$ sygnału $x(t)$. Ponieważ mnożeniu sygnałów odpowiada splatanie ich widm, zatem

$$X_s(\omega) = \frac{1}{2\pi} X(\omega) * Y(\omega), \tag{6.24}$$

gdzie $Y(\omega)$ jest widmem unipolarnej fali prostokątnej. Zgodnie ze wzorem (3.53) widmo to ma postać:

$$Y(\omega) = 2\pi \sum_{n=-\infty}^{\infty} \frac{\tau}{T_s} \operatorname{Sa}\left(n\pi\frac{\tau}{T_s}\right) \delta(\omega - 2n\omega_m), \quad \omega_m = \frac{\pi}{T_s}. \tag{6.25}$$

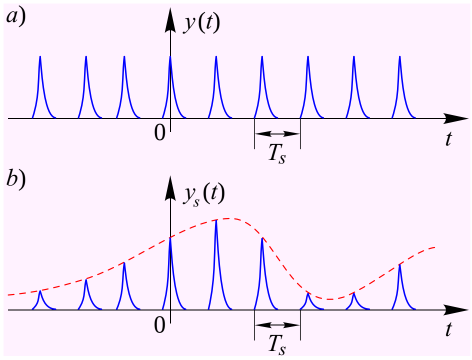
Tak więc:

$$\begin{aligned} X_s(\omega) &= X(\omega) * \sum_{n=-\infty}^{\infty} \frac{\tau}{T_s} \operatorname{Sa}\left(n\pi\frac{\tau}{T_s}\right) \delta(\omega - 2n\omega_m) = \\ &= \frac{\tau}{T_s} \sum_{n=-\infty}^{\infty} \operatorname{Sa}\left(n\pi\frac{\tau}{T_s}\right) [X(\omega) * \delta(\omega - 2n\omega_m)] = \\ &= \frac{\tau}{T_s} \sum_{n=-\infty}^{\infty} \operatorname{Sa}\left(n\pi\frac{\tau}{T_s}\right) X(\omega - 2n\omega_m). \end{aligned} \tag{6.26}$$

Widmo $X_s(\omega)$ jest pokazane na rys. 6.10f. Jest ono ciągiem niezniekształconych kopii widma $X(\omega)$ powtarzanych co $2\omega_m$. W przeciwieństwie do próbkowania idealnego ich wysokości nie są jednakowe (por. wzór (3.58)), lecz określone przez wartości funkcji $(\tau/T_s)\operatorname{Sa}(\tau\omega/2)$ w punktach $2n\omega_m$. Ponieważ kształt środkowego segmentu widma $X_s(\omega)$ jest identyczny jak widma $X(\omega)$, sygnał $x(t)$ można odtworzyć w odbiorniku bez zniekształceń za pomocą idealnego filtru dolnoprzepustowego o charakterystyce:

$$\frac{T_s}{\tau} \sqcap \left(\frac{\omega}{2\omega_m}\right). \tag{6.27}$$

Przeprowadzona analiza pozostanie słuszna, jeśli falę prostokątną zastąpimy jakimkolwiek innym okresowym ciągiem krótkich impulsów próbkujących o okresie $T_s \leqslant \pi/\omega_m$ (rys. 6.11). Mnożeniu sygnału informacyjnego $x(t)$ przez ten ciąg odpowiada zawsze splatanie widma $X(\omega)$ z dystrybucyjnym widmem tego ciągu. Widmo sygnału spróbkowanego będzie zatem ciągiem niezniekształconych kopii widma $X(\omega)$. W szczególności kształt widma $X(\omega)$ zostanie zachowany w środkowym segmencie widma $X_s(\omega)$ sygnału spróbkowanego. Wynika stąd, że w przypadku próbkowania naturalnego, bez względu na kształt stosowanych impulsów próbkujących, można zawsze odzyskać po stronie odbiorczej niezniekształcony sygnał informacyjny, za pomocą filtru dolnoprzepustowego.



Rys. 6.11. Ciąg próbkujących impulsów szpilkowych (a) i sygnał spróbkowany tym ciągiem (b)

6.4.5. Próbkowanie chwilowe

W nieco odmienny sposób jest formowany sygnał spróbkowany w systemie *próbkowania chwilowego* (rys. 6.12). W systemie tym co przedział czasu T_s wytwarzany jest w nadajniku krótki impuls prostokątny o czasie trwania $\tau \ll T_s$, którego amplituda jest proporcjonalna do wartości bieżącej próbki $x(nT_s)$ sygnału informacyjnego $x(t)$. Otrzymany w ten sposób sygnał spróbkowany jest pokazany na rys. 6.12e.

Próbkowanie chwilowe jest nazywane próbkowaniem typu *sample and hold* (spróbkuj i podtrzymaj). W przeciwieństwie do próbkowania naturalnego, wytworzenie sygnału spróbkowanego chwilowo nie wymaga stosowania układu mnożącego. W praktyce stosowane są w tym celu proste układy *próbkująco-pamiętające*.

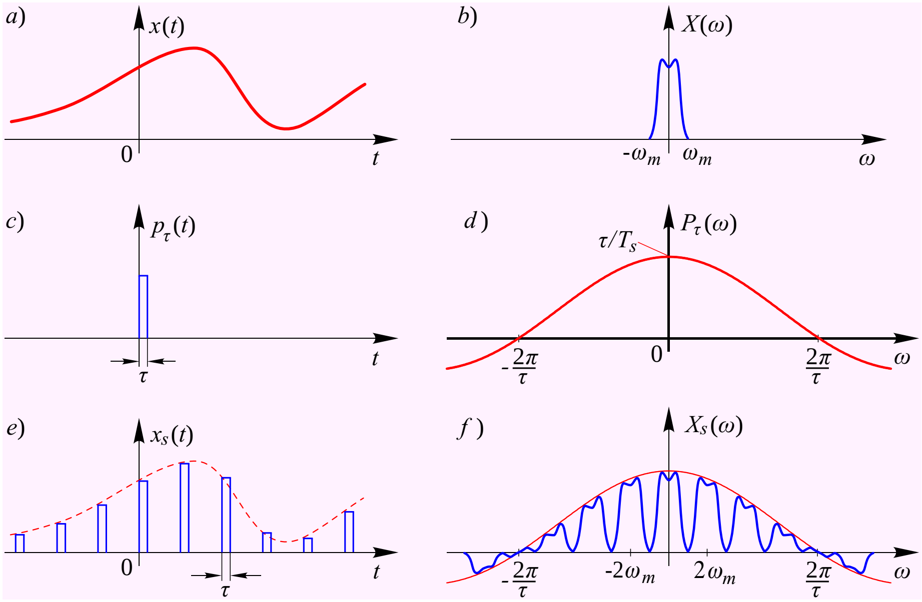
Jeżeli prostokątny impuls próbkujący o czasie trwania τ (rys. 6.12c) oznaczymy:

$$p_\tau(t) = \sqcap\left(\frac{t-\tau/2}{\tau}\right), \tag{6.28}$$

to sygnał spróbkowany w systemie próbkowania chwilowego można w zapisać w postaci:

$$x_s(t) = \sum_{n=-\infty}^{\infty} x(nT_s)p_\tau(t-nT_s). \tag{6.29}$$

Mechanizm generacji tego sygnału jest podobny jak w przypadku sygnału schodkowego (6.20). Sygnał (6.29) różni się od sygnału (6.20) jedynie czasem trwania schodków. W sygnale schodkowym czas trwania poszczególnych schodków jest równy całemu przedziałowi próbkowania T_s, natomiast w przypadku sygnału spróbkowanego chwilowo jest on małym ułamkiem tego przedziału. Aby wyznaczyć widmo $X_s(\omega)$ sygnału (6.29), wystarczy zatem powtórzyć rozumowanie z p.



Rys. 6.12. Próbkowanie chwilowe

6.4.2, uwzględniając we wzorze (6.22) różnicę w czasie trwania impulsów prostokątnych. Przy założeniu granicznego przedziału próbkowania $T_s = \pi/\omega_m$ widmo to przybiera postać (rys. 6.12f):

$$X_s(\omega) = \frac{\tau}{T_s}\,\mathrm{Sa}(\omega\tau/2)\,\mathrm{e}^{-\mathrm{j}(\omega\tau/2)} \sum_{n=-\infty}^{\infty} X(\omega - 2n\omega_m). \tag{6.30}$$

Jak wynika z tego wzoru, widmo sygnału spróbkowanego chwilowo jest ciągiem kopii widma $X(\omega)$ sygnału informacyjnego zniekształconych obwiednią o kształcie funkcji Sa. Z uwagi na krótki czas τ impulsów próbkujących obwiednia ta jest znacznie bardziej rozciągnięta w czasie i tym samym bardziej płaska w przedziale $|\omega| \leqslant \omega_m$ w porównaniu z obwiednią widma sygnału schodkowego pokazaną na rys. 6.9b. Zniekształcenia środkowego segmentu widma $X_s(\omega)$ są w tym przypadku znacznie mniejsze, tym mniejsze, im mniejszy jest czas τ trwania impulsów w porównaniu z okresem próbkowania T_s. Sygnał informacyjny w systemie próbkowania chwilowego może być więc odtworzony w odbiorniku za pomocą idealnego filtru dolnoprzepustowego, z ewentualną korekcją jego charakterystyki analogiczną jak w przypadku odtwarzania sygnału z próbek metodą schodkową.

6.5. Odstępstwa od idealnych założeń twierdzenia o próbkowaniu

W przeprowadzonych dotąd rozważaniach zakładaliśmy, że warunki, przy których słuszne jest twierdzenie o próbkowaniu, są spełnione w sposób ścisły. Zakładaliśmy również idealny charakter operacji próbkowania i odtwarzania sygnału z próbek. W rzeczywistości założenia te mogą być spełnione jedynie z pewnym przybliżeniem. W tym punkcie przedyskutujemy odstępstwa od założeń idealnych oraz wynikające stąd błędy w praktycznej realizacji operacji próbkowania i odtwarzania sygnałów. Analiza tych odstępstw ma istotne znaczenie dla prawidłowego stosowania w praktyce cyfrowych metod przetwarzania sygnałów.

6.5.1. Konsekwencje warunku ograniczoności pasma

Podstawowym założeniem twierdzenia o próbkowaniu jest warunek ograniczonego pasma sygnału. Ograniczone pasmo mogą mieć jednak tylko takie sygnały, których czas trwania jest nieskończony. Stwierdzenie to można uzasadnić, przeprowadzając następujące proste rozumowanie.

Jeśli sygnał $x(t)$ ma widmo $X(\omega)$ ograniczone pulsacją ω_m, to widmo to można przedstawić w postaci tożsamościowej równości:

$$X(\omega) \;\equiv\; X(\omega)\sqcap\left(\frac{\omega}{2\omega_m}\right). \tag{6.31}$$

Z równości tej wynika z kolei równoważna tożsamość w dziedzinie czasu:

$$x(t) \;\equiv\; x(t) * \frac{\omega_m}{\pi}\,\mathrm{Sa}\,\omega_m t. \tag{6.32}$$

Ponieważ sygnał Sa przybiera niezerowe wartości w nieskończonym przedziale czasu, więc jego splot z dowolnym sygnałem $x(t)$ daje w wyniku sygnał przybierający niezerowe wartości również w przedziale nieskończonym. Tak więc, przyjmując we wzorze (6.32) skończony czas trwania sygnału $x(t)$, dochodzimy do sprzeczności.

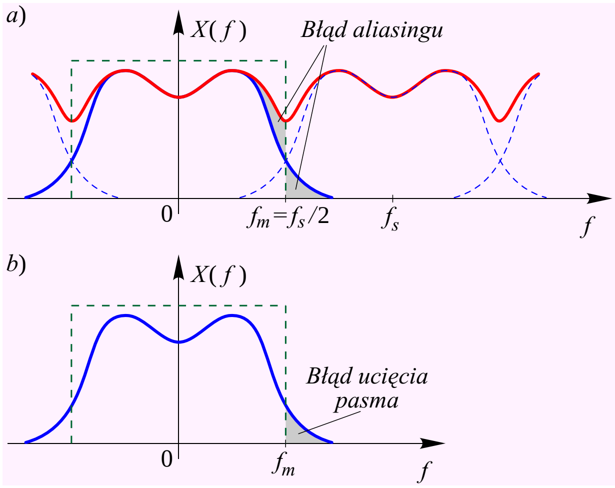
Z przeprowadzonego dowodu wynika, że dokładna informacja o sygnale może być zapamiętana w jego próbkach jedynie wtedy, kiedy jest on sygnałem o nieskończonym czasie trwania. Niezależnie od podanego dowodu, wniosek ten jest konsekwencją ogólnego, dobrze znanego w teorii obwodów, *twierdzenia Paleya-Wienera* (por. np. [3], p. 9.4.3-C), zgodnie z którym każdy sygnał określony w skończonym przedziale czasu ma widmo o paśmie nieograniczonym.

6.5.2. Błąd aliasingu

Sygnały fizyczne występujące w rzeczywistości mogą być obserwowane i badane jedynie w skończonym przedziale czasu. Są one zatem zawsze sygnałami o skończonym czasie trwania i nie spełniają, ściśle rzecz biorąc, warunków twierdzenia o próbkowaniu. Warunki te mogą spełniać jedynie ich modele teoretyczne. Sygnały fizyczne są więc sygnałami o nieograniczonym paśmie.

Z reguły jednak widmo każdego sygnału fizycznego przybiera powyżej pewnej częstotliwości progowej wartości pomijalnie małe. Dla różnych sygnałów wartość częstotliwości progowej jest różna, ale dla danego typu sygnałów można ją zawsze względnie dokładnie określić. Na przykład, maksymalna częstotliwość pasma sygnałów akustycznych jest określona progiem słyszalności ucha ludzkiego, który w typowych przypadkach wynosi około 20 kHz. Sygnały akustyczne wystarczy zatem próbkować z częstotliwością 40 kHz, popełniając przy tym błąd całkowicie pomijalny z punktu widzenia jakości ich przetwarzania. Co więcej, badania wykazały, że dominująca część widma sygnału mowy jest zawarta w paśmie do 3,3 kHz, a więc sygnał mowy można byłoby próbkować już z częstotliwością 6,6 kHz, bez popełnienia istotnego błędu. W rzeczywistości w cyfrowych systemach telefonicznych sygnały są próbkowane ze standardową częstotliwością 8 kHz, zapewniającą nie tylko dostateczną jakość sygnału, ale również separację widmową transmitowanych sygnałów. Podobnie, w systemach radiofonicznych, czy telewizyjnych sygnały są transmitowane w ściśle określonych pasmach, poza którymi ich gęstość widmowa, skupiona w tzw. „ogonach” widmowych, jest szczątkowa. Nawet w przypadku sygnałów szerokopasmowych, np. szumowych, zawsze można wyróżnić częstotliwość progową, choć być może bardzo dużą, powyżej której ich gęstość widmowa jest pomijalna.

Jeżeli jednak na próbkowanie realnych sygnałów o skończonym czasie trwania spojrzymy w pełni ściśle z teoretycznego punktu widzenia, to przyjmując za częstotliwość progową sygnału $x(t)$ pewną częstotliwość f_m i próbkując go z częstotliwością $2f_m$, popełniamy przy odtwarzaniu sygnału z tak pobranych próbek pewien błąd aliasingu (por. p. 6.3.2). Powielone w wyniku próbkowania kopie widma $X(\omega)$ sygnału $x(t)$ nakładają się bowiem wtedy na siebie swoimi „ogonami” i stosując idealny filtr dolnoprzepustowy o pulsacji odcięcia $\omega_m = 2\pi f_m$, nie możemy odzyskać niezniekształconego widma $X(\omega)$ (rys. 6.13a). Zniekształcenie widma na wyjściu takiego filtru występuje zarówno w zakresie pulsacji $\omega \leqslant \omega_m$, w którym następuje nakładanie się kopii widmowych i w rezultacie zwiększenie gęstości widmowej, jak i w zakresie $\omega > \omega_m$, w którym filtr usuwa z sygnału występujące w jego widmie składowe. Za miarę błędu odtwarzania spowodowanego aliasingiem można zatem przyjąć w przybliżeniu sumę pól zaznaczonych na rys. 6.13a.



Rys. 6.13. Porównanie błędu aliasingu i błędu ucięcia pasma

6.5.3. Filtr ochronny. Błąd ucięcia pasma

Jeśli powyżej ustalonej częstotliwości progowej widmo sygnału fizycznego jest pomijalne, to próbkując go z częstotliwością co najmniej dwukrotnie większą od częstotliwości progowej, popełniamy błąd aliasingu na tyle mały, że nie ma on istotnego wpływu na dokładność dalszego przetwarzania sygnału. Często jednak, z uwagi na ograniczoną szybkością działania układów próbkujących jakimi dysponujemy, warunek ten nie może być w praktyce spełniony. W takich przypadkach błąd aliasingu może być znaczący.

W celu złagodzenia skutków aliasingu stosowany jest zwykle dolnoprzepustowy *filtr ochronny* (*antyaliasingowy*), którego zadaniem jest ucięcie pasma sygnału na pewnej częstotliwości f_m przed jego próbkowaniem (rys. 6.13b). Próbkując sygnał otrzymany na wyjściu filtru ochronnego z częstotliwością $2f_m$ i odtwarzając z tak pobranych próbek sygnał pierwotny, popełniamy błąd nazywany *błędem ucięcia pasma*. Za miarę tego błędu można przyjąć pole zaznaczone na rys. 6.13b.

Z porównania rys. 6.13a i b wynika, że przy założeniu tej samej częstotliwości f_m i tej samej częstotliwości próbkowania $2f_m$ błąd aliasingu jest zawsze mniejszy od błędu ucięcia pasma. Dowodzi to celowości stosowania filtru ochronnego.

6.5.4. Miara błędu ucięcia pasma

Ucięcie pasma sygnału $x(t)$ filtrem ochronnym o pulsacji odcięcia ω_m jest równoważne aproksymacji tego sygnału sygnałem

$$x(t;\omega_m) = \frac{1}{2\pi}\int\limits_{-\omega_m}^{\omega_m} X(\omega)\,\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega, \tag{6.33}$$

którego widmo przybiera wartości niezerowe tylko w przedziale $|\omega| \leqslant \omega_m$. Sygnał

$$\varepsilon(t;\omega_m) = x(t) - x(t;\omega_m) \tag{6.34}$$

jest sygnałem błędu aproksymacji. Widmo tego sygnału ma składowe zawarte jedynie w paśmie $|\omega| > \omega_m$.

W przypadku sygnałów o ograniczonej energii za miarę błędu względnego ucięcia pasma można przyjąć stosunek energii E_ε sygnału błędu aproksymacji $\varepsilon(t;\omega_m)$ do całkowitej energii E_x sygnału $x(t)$:

$$\Delta_{\omega_m} = \frac{E_\varepsilon}{E_x}. \tag{6.35}$$

Ponieważ sygnał aproksymujący $x(t;\omega_m)$ i sygnał błędu $\varepsilon(t;\omega_m)$ mają widma rozłączne, więc są sygnałami ortogonalnymi. Oznacza to, że energia sygnału błędu jest różnicą energii sygnału $x(t)$ i sygnału aproksymującego $x(t;\omega_m)$:

$$E_\varepsilon = E_x - E_{x-\varepsilon} = \frac{1}{\pi}\int\limits_0^\infty |X(\omega)|^2\,\mathrm{d}\omega - \frac{1}{\pi}\int\limits_0^{\omega_m} |X(\omega)|^2\,\mathrm{d}\omega = \frac{1}{\pi}\int\limits_{\omega_m}^\infty |X(\omega)|^2\,\mathrm{d}\omega. \tag{6.36}$$

Tak więc:

$$\Delta_{\omega_m} = \frac{E_\varepsilon}{E_x} = \frac{\int\limits_{\omega_m}^\infty |X(\omega)|^2\,\mathrm{d}\omega}{\int\limits_0^\infty |X(\omega)|^2\,\mathrm{d}\omega}. \tag{6.37}$$

W podobny sposób można określić miarę błędu względnego ucięcia pasma w przypadku sygnałów o ograniczonej mocy.

Przykład 6.1. Rozważmy sygnał wykładniczy malejący $x(t) = X_0\,\mathrm{e}^{-\alpha t}\,\boldsymbol{1}(t)$ (rys. 1.12). Jego widmo energii $\Phi_x(\omega) = X_0^2/(\alpha^2+\omega_2)$, a energia $E_x = X_0^2/2\alpha$ (por. przykład 5.5).

Przyjmijmy za częstotliwość progową f_m tego sygnału częstotliwość, dla której widmo energii maleje do 0,1 maksymalnej wartości. Ze wzoru (5.13) obliczamy wartość tej częstotliwości $f_m = 3\alpha/2\pi = 0,477\alpha$. Próbkując sygnał $x(t)$ z częstotliwością $f_s = 2f_m = 0,954\alpha$, pobieramy próbki co przedział $T_s = 1,047/\alpha$, co oznacza, że na odcinku czasu równym równoważnemu czasowi trwania tego sygnału $\Delta t_x = 1/\alpha$ (por. wzór (3.61)), jest pobierana zaledwie jedna próbka. Jest to zatem próbkowanie bardzo wolne. Błąd względny ucięcia pasma w tym przypadku wynosi:

$$\Delta_{\omega_m} = \frac{E_\varepsilon}{E_x} = \frac{\dfrac{X_0^2}{\pi}\displaystyle\int_{3\alpha}^{\infty}\frac{1}{\alpha^2+\omega^2}\,\mathrm{d}\omega}{\dfrac{X_0^2}{2\alpha}} = \frac{2}{\pi}\left(\frac{\pi}{2} - \operatorname{arctg} 3\right) = 20,48\%.$$

Widzimy więc, że sygnał odtworzony z tak pobieranych próbek bardzo niedokładnie aproksymuje sygnał $x(t)$.

Jeśli przyjmiemy próg odcięcia pasma na poziomie 0,01 maksymalnej wartości widma energii, to częstotliwość progowa wyniesie $1,584\alpha$. Przy próbkowaniu z częstotliwością dwukrotnie większą $3,168\alpha$ (z okresem $0,316/\alpha$) w przedziale czasu Δt_x są pobierane nieco więcej niż 3 próbki. W tym przypadku błąd względny ucięcia pasma maleje do wartości 6,38%. Aby uzyskać błąd względny mniejszy niż 1%, należy pobrać w przedziale Δt_x ponad 20 próbek.

6.5.5. Błąd ucięcia w czasie

Rozważmy z kolei przypadek sygnału $x(t)$ o nieograniczonym czasie trwania, którego pasmo jest ograniczone częstotliwością f_m. Zgodnie ze wzorem Kotielnikowa-Shannona (6.5) sygnał ten jest reprezentowany zbiorem swoich próbek pobieranych z częstotliwością Nyquista $f_s = 1/T_s = 2f_m$. Przyjmijmy, że sygnał ten jest obserwowany w skończonym oknie czasowym $[-T_0, T_0]$. Oznacza to, że dysponujemy wówczas $2N+1$ jego próbkami, gdzie N jest największą liczbą naturalną, taką, że $NT_s \leqslant T_0$. Odtworzenie sygnału $x(t)$ na podstawie tych próbek odpowiada przybliżeniu nieskończonego szeregu (6.5) sumą częściową tego szeregu:

$$x_N(t) = \sum_{n=-N}^{N} x(nT_s)\,\mathrm{Sa}\,2\pi f_m(t - nT_s). \tag{6.38}$$

Błąd tego przybliżenia $\varepsilon_N(t) = x(t) - x_N(t)$ jest nazywany *błędem ucięcia w czasie*.

Zwracamy uwagę na różnicę w sformułowaniu zagadnień. W przypadku omawianym w p. 6.5.2-6.5.4 zakładaliśmy, że czas trwania sygnału jest skończony i że jest on obserwowany w całym czasie jego trwania. Jest to równoważne założeniu zerowych próbek poza przedziałem obserwacji. W tym przypadku zakładamy, że sygnał trwa w czasie nieskończenie długo, ale obserwujemy go jedynie w przedziale skończonym. Próbki leżące poza przedziałem obserwacji mają wartości niezerowe.

Zgodnie ze wzorem ogólnym (2.41) oraz wzorem (6.12) miara energetyczna błędu ucięcia w czasie przybiera postać:

$$\|\varepsilon_N\|^2 = T_s \sum_{|n|>N} x^2(nT_s) \tag{6.39}$$

i jest określona przez próbki sygnału leżące poza przedziałem obserwacji. Ponieważ próbki te nie są znane, wzór (6.39) nie może być wykorzystany w praktyce do oszacowania miary błędu ucięcia w czasie. Oszacowanie tej miary jest zagadnieniem trudnym. W literaturze podawane są różne wzory określające kres górny tej miary. Ogólnie rzecz biorąc, skończone okno czasowe obserwacji $[-T_0, T_0]$ powoduje zniekształcenia widma $X(\omega)$ sygnału $x(t)$ wynikające ze splatania tego widma w dziedzinie częstotliwości z funkcją typu Sa (odpowiada to mnożeniu sygnału przez funkcję prostokątną okna):

$$X_{T_0}(\omega) = \frac{1}{2\pi} X(\omega) * 2T_0 \,\mathrm{Sa}\, \omega T_0. \tag{6.40}$$

Jest oczywiste, że im okno czasowe jest szersze, tym listek główny funkcji Sa jest węższy i wyższy, a więc zniekształcenia są mniejsze. W granicy, gdy T_0 dąży do nieskończoności, funkcja $(T_0/\pi)\,\mathrm{Sa}\,\omega T_0$ dąży do delty Diraca $\delta(\omega)$ i, jak wynika ze wzoru (6.40), widmo $X_{T_0}(\omega)$ dąży do widma $X(\omega)$.

6.6. Próbkowanie z częstotliwością mniejszą od częstotliwości Nyquista

6.6.1. Próbkowanie sygnałów wąskopasmowych

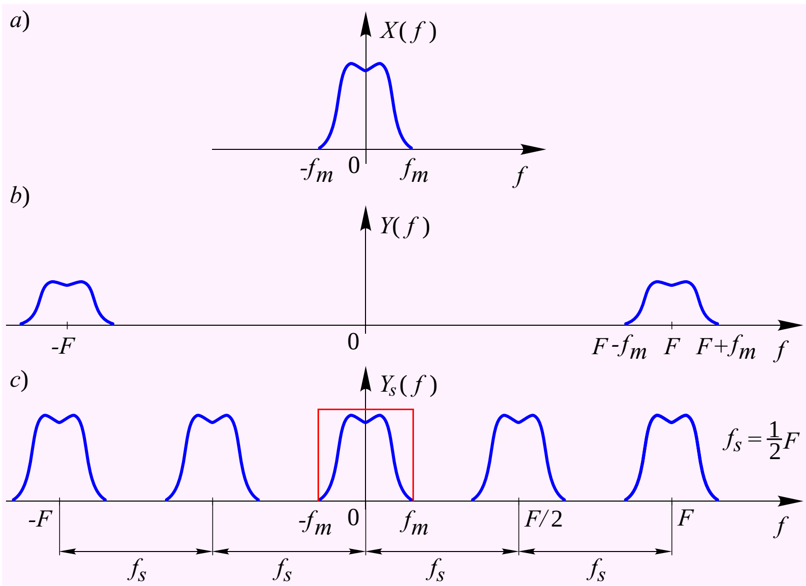
Pobieranie próbek z częstotliwością mniejszą od częstotliwości Nyquista prowadzi zawsze do utraty części informacji zawartej w sygnale. Niekiedy jednak pozostała część informacji, zachowana w tak pobranych próbkach, umożliwia odtworzenie istotnych charakterystyk sygnału. Sytuacja taka ma na przykład miejsce w przypadku próbkowania sygnałów wąskopasmowych.

Sygnałami wąskopasmowymi są m.in. sygnały transmitowane w większości systemów modulacji sygnałów. Rozważmy najprostszy system modulacji amplitudy AM-SC (system dwuwstęgowy bez fali nośnej – por. p. 10.2). W systemie tym sygnał transmitowany ma postać:

$$y(t) = x(t) \cos 2\pi F t, \tag{6.41}$$

gdzie $x(t)$ jest dolnopasmowym sygnałem modulującym (informacyjnym) o widmie $X(f)$ ograniczonym częstotliwością f_m (rys. 6.14a). Częstotliwość F harmonicznego sygnału modulowanego (fali nośnej) jest znacznie większa od częstotliwości f_m. Zgodnie z twierdzeniem o modulacji (3.26), widmo $Y(f)$ sygnału (6.41) otrzymujemy przez rozszczepienie widma $X(f)$ sygnału modulującego na dwie części przesunięte do częstotliwości $\pm F$ (rys. 6.14b). Obie części zachowują przy tym kształt widma $X(f)$, mają dwukrotnie mniejszą gęstość widmową i są położone symetrycznie wokół częstotliwości $\pm F$ w wąskich pasmach o szerokości $2f_m$.

Największa częstotliwość w widmie sygnału zmodulowanego jest równa $F + f_m$. Aby zachować w próbkach pełną informację o sygnale zmodulowanym, należałoby go zatem próbkować z częstotliwością $2(F + f_m)$. W zagadnieniach transmisji sygnałów nie chodzi jednak o dokładne odtworzenie na podstawie próbek samego sygnału zmodulowanego, lecz jego dolnopasmowej obwiedni. W obwiedni jest bowiem zawarta informacja o sygnale użytecznym (np. o sygnale mowy). Wykorzystując fakt, że widmo $Y(f)$ ma szerokie „puste" pasmo częstotliwości $(-F + f_m, F - f_m)$, możemy próbkować go z częstotliwością znacznie mniejszą niż częstotliwość Nyquista $2(F + f_m)$ i mimo to odtworzyć interesującą nas informację zawartą w sygnale modulującym $x(t)$.



Rys. 6.14. Próbkowanie sygnału wąskopasmowego z częstotliwością mniejszą od częstotliwości Nyquista

Na rys. 6.14c zilustrowano przypadek próbkowania sygnału zmodulowanego z częstotliwością $f_s = F/2$, a więc ponad czterokrotnie mniejszą od częstotliwości Nyquista $2(F + f_m)$. W wyniku próbkowania obie części prawostronna i lewostronna widma $Y(f)$ zostają powielone okresowo z okresem $F/2$. Powielone kopie nakładają się na siebie wokół punktów $nF/2$, w efekcie czego widmo $Y_s(f)$ sygnału spróbkowanego jest ciągiem niezniekształconych kopii widma $X(f)$ sygnału modulującego. Ponieważ centralna kopia występuje dokładnie w początku osi częstotliwości, zatem sygnał modulujący można odtworzyć bez zniekształceń za pomocą idealnego filtru dolnoprzepustowego o częstotliwości granicznej równej f_m. Jak można łatwo sprawdzić, odtworzenie niezniekształconego sygnału informacyjnego za pomocą tego filtru jest możliwe również przy wolniejszym próbkowaniu sygnału zmodulowanego z częstotliwościami F/k, $k = 3, \ldots, k_{\max}$, gdzie $k_{\max}$ jest największą liczbą całkowitą spełniającą nierówność $k_{\max} \leqslant F/(2f_m)$. Nierówność ta jest warunkiem nie nakładania się sąsiadujących ze sobą kopii widmowych w widmie $Y_s(f)$. Dla $k > F/(2f_m)$ odtworzenie niezniekształconej obwiedni sygnału zmodulowanego nie jest już możliwe.

6.6.2. Efekt stroboskopowy

Innym przypadkiem, kiedy w próbkach pobieranych z częstotliwością mniejszą od częstotliwości Nyquista zostaje zachowana użyteczna, aczkolwiek tylko częściowa informacja o sygnale, jest *efekt stroboskopowy*. Efekt ten występuje przy zbyt wolnym próbkowaniu sygnałów okresowych. Przyczynę jego powstawania wyjaśnimy najpierw na przykładzie sygnału harmonicznego.

Załóżmy, że z tą samą częstotliwością próbkowania $f_s = 1/T_s$ są próbkowane dwa sygnały harmoniczne: sygnał $x_1(t) = \sin 2\pi f_0 t$, o małej częstotliwości f_0, oraz sygnał $x_2(t) = \sin 2\pi(f_0 + mf_s)t$, o dużo większej częstotliwości $f_0 + mf_s$, gdzie m jest pewną liczbą naturalną. Załóżmy ponadto, że $f_s \gg 2f_0$. Oznacza to, że pierwszy sygnał harmoniczny jest próbkowany z częstotliwością znacznie większą od jego częstotliwości Nyquista $2f_0$, natomiast drugi – z częstotliwością mniejszą (dla dużych m – znacznie mniejszą) od jego częstotliwości Nyquista $2(f_0 + mf_s)$. W wyniku próbkowania otrzymujemy sygnały dyskretne

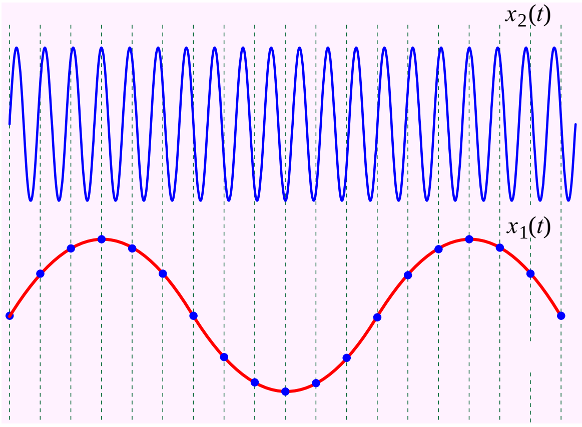
$$x_1(nT_s) = \sin 2\pi f_0 nT_s \tag{6.42}$$

i odpowiednio

$$\begin{aligned} x_2(nT_s) &= \sin 2\pi(f_0 + mf_s)nT_s = \\ &= \sin(2\pi f_0 nT_s + 2\pi mn) = \sin 2\pi f_0 nT_s. \end{aligned} \tag{6.43}$$

W obu przypadkach otrzymujemy zatem identyczny ciąg próbek. Na jego podstawie nie możemy rozstrzygnąć, czy pochodzi on od pierwszego, czy też od drugiego sygnału. Jeżeli faktycznie próbkujemy sygnał $x_2(t)$ o dużej częstotliwości, to jest on „widziany” jako sygnał $x_1(t)$ o małej częstotliwości. Niejednoznaczność jest oczywiście spowodowana zbyt wolnym próbkowaniem sygnału $x_2(t)$. Efekt ten, nazywany *efektem stroboskopowym*, jest zilustrowany dla przypadku $m = 1$ na rys. 6.15. Próbkując wolno sygnał harmoniczny $x_2(t)$ o dużej częstotliwości (w przypadku przedstawionym rys. 6.15 około raz w okresie), odtwarzamy sygnał $x_1(t)$, także harmoniczny, ale o znacznie mniejszej częstotliwości. Dla $m = 1$ częstotliwość ta jest równa różnicy częstotliwości sygnału $x_2(t)$ i częstotliwości próbkowania.

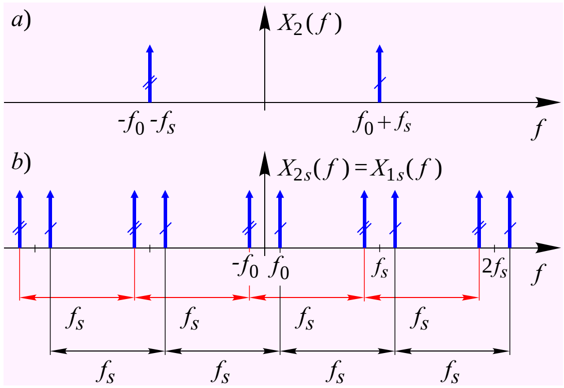
Efekt stroboskopowy można zaobserwować, na przykład, oglądając stare filmy. Jego wizualnym przejawem są nienaturalnie wolno obracające się koła pojazdów. Przyczyną tego jest zbyt wolny przesuw taśmy filmowej (z częstotliwością 24 klatek na sekundę), niedostateczny w stosunku do rzeczywistej prędkości kątowej obrotu kół.



Rys. 6.15. Efekt stroboskopowy w przypadku sygnału harmonicznego

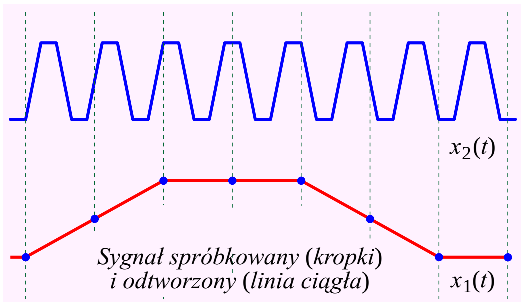
Efekt stroboskopowy można także wyjaśnić, analizując widma obu rozpatrywanych sygnałów harmonicznych. Na rys. 6.16a pokazano widmo sygnału $x_2(t)$ dla $m = 1$. Zawiera ono dwa prążki występujące w punktach $f_0 + f_s$ oraz $-f_0 - f_s$. Próbkowanie sygnału $x_2(t)$ z częstotliwością f_s powoduje powielenie okresowe tego widma z okresem f_s. W wyniku otrzymujemy widmo sygnału spróbkowanego pokazane na rys. 6.16b. Widmo to jest identyczne jak widmo sygnału $x_1(t)$ spróbkowanego z tą samą częstotliwością f_s (para prążków w punktach $\pm f_0$ powielona okresowo z okresem f_s). Widzimy więc, że po spróbkowaniu oba sygnały są nierozróżnialne w dziedzinie częstotliwości. Podobna sytuacja ma miejsce dla $m = 2, 3, \ldots$.

Efekt stroboskopowy występuje również w przypadku zbyt wolnego próbkowania dowolnego sygnału okresowego o okresie T_0 (częstotliwości podstawowej $f_0 = 1/T_0$), który nie zawiera składowych harmonicznych o częstotliwościach większych niż Kf_0, gdzie K jest pewną liczbą naturalną. Aby możliwe było jego wierne odtworzenie z próbek, należałoby je pobierać z częstotliwością co najmniej równą $2Kf_0$. Jeśli pasmo sygnału jest szerokie, tzn. K jest duże, warunek ten może okazać się w praktyce niemożliwy do spełnienia. Próbkując sygnał dużo wolniej z odpowiednio dobraną częstotliwością, można jednak odtworzyć jego spowolnioną kopię, która zachowuje kształt sygnału.



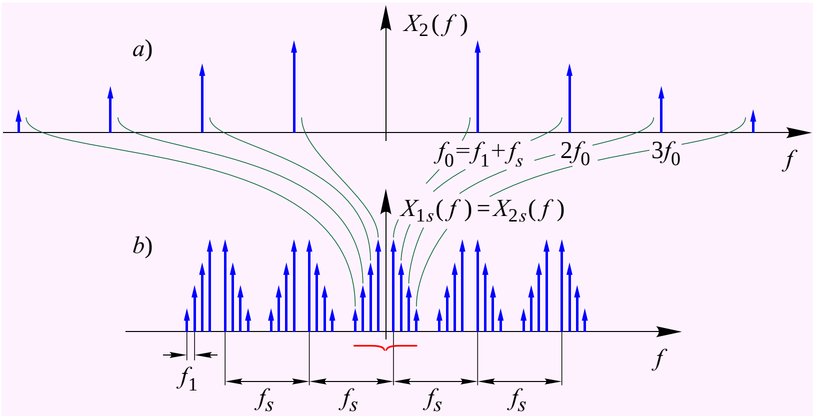
Rys. 6.16. Efekt stroboskopowy w dziedzinie częstotliwości przypadku sygnału harmonicznego

Na rys. 6.17 zilustrowano efekt stroboskopowy w przypadku próbkowania sygnału trapezowego $x_2(t)$ o częstotliwości podstawowej f_0. Dla wygody zakładamy, że jego składowe harmoniczne o częstotliwościach powyżej $4f_0$ są pomijalne. Sygnał ten jest próbkowany nieco ponad raz w okresie z częstotliwością $f_s = \frac{11}{12}f_0$, a więc ponad ośmiokrotnie mniejszą od częstotliwości Nyquista (rys. 6.17a). Taki sam ciąg próbek otrzymujemy, próbkując z tą samą częstotliwością wolny sygnał trapezowy $x_1(t)$ o częstotliwości podstawowej $f_1 = \frac{1}{12}f_0$ (rys. 6.17b).



Rys. 6.17. Efekt stroboskopowy w przypadku sygnału okresowego

Interpretację efektu stroboskopowego w dziedzinie częstotliwości w przypadku próbkowania sygnału trapezowego przedstawiono na rys. 6.18. Widmo tego sygnału zawiera cztery pary prążków w punktach $\pm kf_0$, $k = 1, 2, 3, 4$ (rys. 6.18a). W wyniku próbkowania sygnału z częstotliwością $f_s = \frac{11}{12}f_0$ każdy z tych prążków ulega okresowemu rozmnożeniu z okresem f_s. W rezultacie otrzymujemy układ prążków pokazany na rys. 6.18b. Identyczny układ prążków otrzymalibyśmy, próbkując z tą samą częstotliwością f_s wolny sygnał trapezowy $x_1(t)$ o częstotliwości podstawowej będącej różnicą $f_1 = f_0 - f_s = \frac{1}{12}f_0$. Podstawowy zespół widmowy tego sygnału (cztery pary prążków w punktach kf_1, $k = 1, 2, 3, 4$) jest zaznaczony na rys. 6.18b klamrą. Dodatkowymi liniami zaznaczono „wędrówkę" prążków. Wskazują one, z powielenia których prążków widma sygnału $x_2(t)$ powstają poszczególne prążki widma sygnału $x_1(t)$. Wolnozmienną kopię szybkiego sygnału trapezowego możemy oczywiście odzyskać z próbek, stosując filtr dolnoprzepustowy o częstotliwości odcięcia $4f_1$.



Rys. 6.18. Efekt stroboskopowy w dziedzinie częstotliwości w przypadku sygnału okresowego

Widzimy więc, że efekt stroboskopowy umożliwia zachowanie w próbkach pobieranych z częstotliwością znacznie mniejszą od częstotliwości Nyquista istotnej informacji o sygnale okresowym. W odtworzonym na podstawie próbek sygnale wolnozmiennym zachowany zostaje kształt sygnału szybkozmiennego. Znając relację między częstotliwością próbkowania i częstotliwością podstawową odzyskanego sygnału wolnozmiennego, możemy odtworzyć częstotliwość podstawową sygnału szybkozmiennego. W praktyce efekt stroboskopowy jest wykorzystywany m.in. w lampach stroboskopowych (np. do regulacji zapłonu w silnikach samochodowych), a także w oscyloskopach próbujących pracujących w bardzo szerokim zakresie częstotliwości rzędu GHz. Uzyskanie efektu stroboskopowego wymaga jednak dużej dokładności w zapewnieniu stabilizacji częstotliwości próbkowania i jej synchronizacji z częstotliwością analizowanych sygnałów.

Na zakończenie dyskusji operacji próbkowania sygnałów z częstotliwością mniejszą od częstotliwości Nyquista podkreślmy zasadniczą różnicę między próbkowaniem stroboskopowym i próbkowaniem sygnałów wąskopasmowych. Wprawdzie w obu przypadkach przyczyna szczególnego ukształtowania się widma sygnału w wyniku jego zbyt wolnego próbkowania jest ta sama (powielenie okresowe widma sygnału próbkowanego), to mechanizm powstawania tego widma jest odmienny. W przypadku próbkowania stroboskopowego poszczególne powtarzające się zespoły widmowe prążków są tworzone w efekcie powielenia okresowego różnych fragmentów (prążków) widma sygnału próbkowanego (rys. 6.18b), podczas gdy w przypadku wolnego próbkowania sygnałów wąskopasmowych poszczególne kopie widmowe powstają w rezultacie przesunięcia okresowego tego samego fragmentu widma sygnału próbkowanego (rys. 6.14c).

6.7. Inne odstępstwa

6.7.1. Nierealizowalność idealnego filtru dolnoprzepustowego

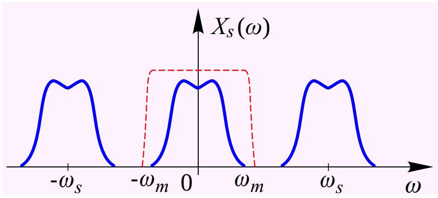
Jeśli na wejście dowolnego liniowego stacjonarnego filtru (por. p. 7.1) zostanie podany impuls Diraca $\delta(t)$, to na wyjściu filtru pojawi się charakterystyczny dla niego sygnał $h(t)$ nazywany odpowiedzią impulsową filtru (por. p. 7.1.9). W punkcie 6.2.6 pokazaliśmy, że jeśli na wejście idealnego filtru dolnoprzepustowego podamy dowolny sygnał $x(t)$, to na jego wyjściu wystąpi sygnał $y(t)$ określony ogólną zależnością (6.14). Z zależności tej i z właściwości splotu dystrybucji Diraca (1.19) wynika zatem, że odpowiedź impulsowa idealnego filtru dolnoprzepustowego ma postać:

$$h(t) = \frac{\omega_g}{\pi} \operatorname{Sa} \omega_g t \tag{6.44}$$

i przybiera wartości niezerowe w przedziale czasu $(-\infty, \infty)$. Ponieważ impuls Diraca jest doprowadzony do wejścia filtru w chwili $t = 0$, oznacza to, że sygnał na wyjściu filtru pojawia się zanim do jego wejścia zostanie doprowadzony sygnał wejściowy. Taki filtr nie może być oczywiście zrealizowany w praktyce. Nie spełnia on bowiem podstawowej zasady przyczynowości obowiązującej w świecie układów fizycznie realizowalnych, zgodnie z którą skutek nie może poprzedzać przyczyny. Tak więc, wykorzystywana w dowodzie twierdzenia o próbkowaniu operacja odtworzenia sygnału za pomocą idealnego filtru dolnoprzepustowego ma znaczenie wyłącznie teoretyczne.

Problem związany z nieistnieniem idealnego filtru dolnoprzepustowego można łatwo rozwiązać, zastępując go filtrem realizowalnym fizycznie o stałej w przybliżeniu charakterystyce w przedziale pulsacji $|\omega| < \omega_m$, gdzie ω_m jest graniczną pulsacją widma sygnału, i zboczach opadających poza tym przedziałem do zera w sposób ciągły (rys. 6.19). W celu uniknięcia zniekształceń odtwarzanego widma należy jednak wówczas próbkować sygnał z częstotliwością odpowiednio większą od częstotliwości Nyquista, tak aby między poszczególnymi powielonymi okresowo kopiami widma sygnału spróbkowanego występowały pasma puste (tzw. *pasma ochronne*). Na przykład, w telefonii cyfrowej za graniczną częstotliwość sygnału mowy przyjmuje się 3,3 kHz. Sygnał telefoniczny wystarczyłoby zatem próbkować z częstotliwością 6,6 kHz. W praktyce stosuje się próbkowanie ze standardową częstotliwością 8 kHz (co 125 μs), co umożliwia dostatecznie dokładne odtworzenie sygnału mowy z próbek po stronie odbiorczej za pomocą rzeczywistego filtru dolnoprzepustowego. Pasmo ochronne stanowi w tym przypadku 17,5% szerokości pasma sygnału.

Należy podkreślić, że w teorii filtrów analogowych znanych jest szereg metod projektowania filtrów dolnoprzepustowych o płaskiej charakterystyce filtracji w zadanym zakresie częstotliwości i zboczach opadających do zera w sposób gładki, lecz na tyle szybki, aby uniknąć zniekształceń sygnału odtwarzanego. Warunki realizowalności tych filtrów są oczywiście tym łagodniejsze, im szersze jest pasma ochronne.



Rys. 6.19. Charakterystyka rzeczywistego filtru dolnoprzepustowego

6.7.2. Jitter

W idealnych warunkach twierdzenia o próbkowaniu zakłada się, że przy równomiernym próbkowaniu sygnału z okresem T_s próbki są pobierane dokładnie w chwilach nT_s. W rzeczywistości, na skutek niedokładności układów próbkujących, faktyczne chwile pobierania próbek odchylają się losowo od założonych chwil teoretycznych. Rozrzut rzeczywistych chwil pobierania próbek wokół chwil nT_s jest nazywany *jitterem* (wahaniem chwil próbkowania).

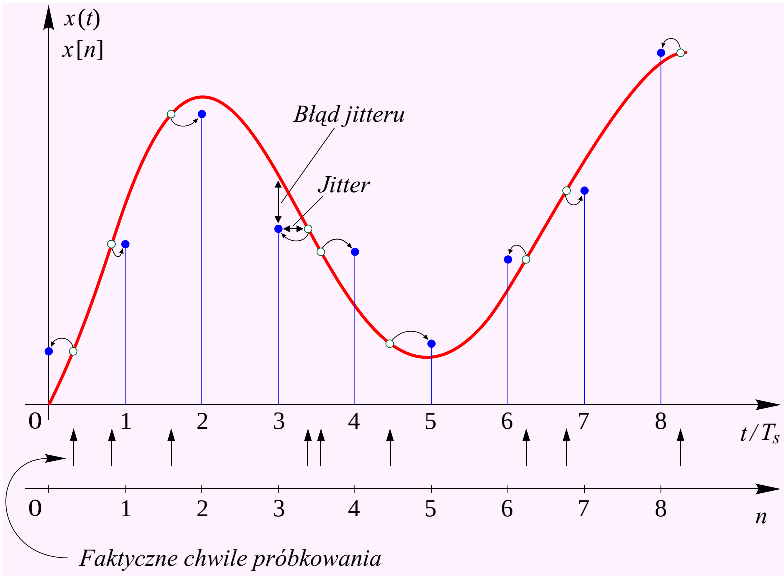
W następstwie jitteru układ próbkujący pobiera próbki sygnału $x(t)$ w losowych chwilach $nT_s+\zeta$, gdzie ζ jest pewną zmienną losową. Z reguły zakłada się, że zmienna ζ nie zależy od numeru n próbki, tzn. jej funkcja gęstości prawdopodobieństwa $f_\zeta(z)$ jest jednakowa dla wszystkich chwil próbkowania. Ponieważ próbka o numerze n nie może wyprzedzać próbki o numerze $n+1$, funkcja $f_\zeta(z)$ jest określona w przedziale $(-T_s/2, T_s/2)$. Zwykle przyjmuje się, że jest ona w tym przedziale symetryczna, tzn. ζ jest zmienną losową o zerowej wartości oczekiwanej i pewnej wariancji σ_ζ^2.

Błąd chwili pobrania próbki powoduje losowy błąd wartości samej próbki nazywany *błędem jitteru*. Jitter i błąd jitteru został zilustrowany na rys. 6.20.

Błąd jitteru możemy zapisać w postaci:

$$\varepsilon_j(nT_s) = x(nT_s+\zeta) - x(nT_s). \tag{6.45}$$

Jest intuicyjnie oczywiste, że wartość błędu jitteru w chwili nT_s zależy od lokalnego zachowania się sygnału w otoczeniu tej chwili. W celu przeanalizowania tego błędu rozwiniemy sygnał $x(t)$ w szereg Taylora wokół punktu nT_s. Jeżeli jitter jest mały, tzn. faktyczna chwila próbkowania jest położona dostatecznie blisko chwili nT_s, to w rozwinięciu w szereg Taylora możemy pominąć składniki



Rys. 6.20. Ilustracja jitteru i błędu jitteru

kwadratowy i wyższe. Uwzględniając tylko dwa pierwsze wyrazy rozwinięcia we wzorze (6.45), otrzymujemy:

$$\varepsilon_j(nT_s) \approx \zeta \frac{\mathrm{d}x(t)}{\mathrm{d}t}\bigg|_{t=nT_s} . \tag{6.46}$$

Wynika stąd, że przy założeniu małego jitteru błąd jitteru w chwili nT_s jest proporcjonalny do wielkości losowej ζ oraz do wartości pochodnej sygnału $x(t)$ w tej chwili. Oznacza to, że błąd jitteru jest tym większy, im szybciej zmienia się sygnał w otoczeniu chwil próbkowania, a więc im szersze jest jego widmo. W praktyce błąd jitteru w przypadku próbkowania sygnałów o niezbyt szerokim paśmie (np. paśmie akustycznym) jest pomijalny i nie ma istotnego wpływu na jakość przetwarzania sygnału. Zjawisko jitteru jest natomiast poważnym źródłem błędów przy przetwarzaniu analogowo-cyfrowym szybkozmiennych sygnałów zawierających składowe częstotliwościowe rzędu megaherców. Wymagane jest wówczas stosowanie wysokostabilnych układów próbkujących lub specjalnych metod kompensacji jitteru

6.7.3. Skończona dokładność reprezentacji danych w cyfrowych układach przetwarzania sygnałów

Jak wspomnieliśmy w p. 6.1.5, niezależnie od błędów występujących podczas próbkowania sygnału, w trakcie jego dalszego przetwarzania w sygnał cyfrowy wprowadzony zostaje błąd $\varepsilon[nT_s]$ nazywany szumem kwantowania. Błąd ten jest związany ze skończoną dokładnością reprezentacji liczb w układach przetwarzania. W zastosowaniach praktycznych istotne jest oszacowanie tego błędu.

Powszechnie przyjmowaną miarą zniekształceń sygnału spowodowanych jego kwantowaniem jest wielkość

$$\mathrm{SNR} = 10 \log \frac{S}{N} \text{ [dB]}, \tag{6.47}$$

nazywana *stosunkiem sygnał-szum* (z ang. *Signal-to-Noise-Ratio*). Miara ta jest miarą decybelową (por. Dodatek 1) stosunku mocy S przetwarzanego sygnału na wejściu przetwornika A/C do mocy N szumu kwantowania $\varepsilon[nT_s]$. W przypadku kwantowania sygnału z zaokrąglaniem (por. p. 6.1.5) wartości szumu kwantowania są rozłożone w przedziale $[-q/2,\ q/2]$, gdzie q jest przedziałem kwantyzacji. Dla bipolarnego przetwornika b-bitowego o zakresie wartości sygnału wejściowego $[-X_m, X_m]$ przedział ten jest określony wzorem:

$$q = \frac{2X_m}{2^b}. \tag{6.48}$$

Moc S zależy oczywiście od typu sygnału wejściowego. W celu jednoznacznego określenia tej mocy za jej średnią miarę przyjmuje się moc wzorcowego sygnału harmonicznego przybierającego wartości w pełnym zakresie wejściowym przetwornika, a więc o amplitudzie X_m równej połowie tego zakresu:

$$S = \frac{X_m^2}{2}. \tag{6.49}$$

Moc N szumu kwantowania wyznacza się przy założeniu, że jest on *dyskretnym szumem białym* (tj. szumem o stałej widmowej gęstości mocy w przedziale $[-\pi, \pi]$), którego wartości $\varepsilon[nT_s]$ mają rozkład równomierny w przedziale $[-q/2, q/2]$ opisany funkcją gęstości prawdopodobieństwa $f_\varepsilon(e) = q^{-1} \sqcap (e/q)$. Ponieważ wartość oczekiwana szumu kwantowania jest wówczas równa zeru, jego moc jest równa wariancji zmiennych losowych $\varepsilon[nT_s]$:

$$N = \sigma_\varepsilon^2 = \int\limits_{-\infty}^{\infty} e^2 \frac{1}{q} \sqcap \left(\frac{e}{q}\right) \mathrm{d}e = \frac{1}{q} \int\limits_{-q/2}^{q/2} e^2 \,\mathrm{d}e = \frac{q^2}{12}. \tag{6.50}$$

Jeśli we wzorze (6.50) uwzględnimy zależność (6.48), moc szumu kwantowania wyrazimy przez zakres przetwornika A/C i jego długość słowa:

$$N = \frac{4X_m^2}{12 \cdot 2^{2b}} = \frac{X_m^2}{3 \cdot 2^{2b}}. \tag{6.51}$$

Podstawiając wzory (6.49) i (6.51) do (6.47), otrzymujemy:

$$\mathrm{SNR} = 10 \log \frac{\frac{1}{2} X_m^2}{\frac{X_m^2}{3 \cdot 2^{2b}}} = 10 \log \left(3 \cdot 2^{2b-1}\right) \text{ [dB]}. \tag{6.52}$$

Wzór (6.52) określa miarę zniekształceń sygnału wprowadzanych w wyniku jego skwantowania przez b -bitowy przetwornik A/C.

W przypadku przetwarzania 8-bitowego, stosowanego np. w telefonii cyfrowej, miara (6.52) wynosi około 50 dB, co w zwykłej skali odpowiada wartości $S/N = 10^5$. Można zatem uznać, że szum kwantowania jest w tym przypadku całkowicie pomijalny i nie ma wpływu na jakość sygnału telefonicznego. Ze wzoru (6.52) wynika ponadto, że każde wydłużenie (skrócenie) słów binarnych o 1 bit powoduje zwiększenie (zmniejszenie) stosunku SNR o około 6 dB. Jeśli wartość SNR jest zadana, ze wzoru (6.52) można wyznaczyć wymaganą liczbę bitów przetwornika. Na przykład, jeśli założona wartość SNR wynosi 30 dB (w skali liniowej 10^3), wystarczy zastosować przetwornik 5-bitowy, jeśli zaś wartość ta wynosi 90 dB, należy stosować przetwornik 15-bitowy.

Szum powstający w wyniku kwantowania próbek przez przetwornik A/C nie jest jedynym źródłem błędów towarzyszących cyfrowemu przetwarzaniu sygnałów i wynikających ze skończonej dokładności reprezentacji liczb w układach przetwarzania. W rzeczywistości także parametry filtru cyfrowego są reprezentowane słowami binarnymi o skończonej długości, a więc ich dokładne wartości mogą być jedynie przybliżone z dokładnością określoną przez przyjętą długość słów. Ponadto wszelkie operacje arytmetyczne realizowane w filtrze cyfrowym, w szczególności sumowanie sygnałów i ich mnożenie przez liczby, mogą być dokonywane ze skończoną dokładnością uwarunkowaną długością rejestrów filtrów. Wpływ powstających przy tym błędów na dokładność przetwarzania zależy przede wszystkim od przyjętej długości słów binarnych kodujących w filtrze dane liczbowe, ale również w znacznym stopniu od struktury filtru, algorytmu jego działania, organizacji obliczeń oraz zastosowanej arytmetyki. Należy podkreślić, że przy wykonywaniu kolejnych kroków obliczeniowych algorytmu przetwarzania błędy te mogą się kumulować. Może to prowadzić do niepożądanych efektów, a nawet całkowitej degradacji obliczeń. Dokładna analiza tych zjawisk jest bardzo złożona. Trudności wynikają głównie z faktu, że poszczególne źródła szumów kwantowania są ze sobą statystycznie skorelowane, a przetwarzanie danych w filtrze cyfrowym ma w rzeczywistości charakter nieliniowy. Obszerną dyskusję tych zagadnień można znaleźć np. w [1], p. 3.4 i 3.5.

Słownik

aliasing (w dziedzinie częstotliwości)
: efekt nakładania się powielonych okresowo kopii widma sygnału w wyniku próbkowania sygnału z częstotliwością mniejszą od częstotliwości Nyquista

błąd aliasingu
błąd wynikający z próbkowania sygnału z częstotliwością mniejszą od częstotliwości Nyquista

błąd (szum) kwantowania
błąd powstający podczas operacji kwantowania sygnału spróbkowanego, wynikający z przybliżenia dokładnych wartości próbek wartościami skwantowanymi

błąd ucięcia pasma
błąd wynikający z odfiltrowania z widma sygnału za pomocą filtru dolnoprzepustowego składowych o częstotliwościach większych od pewnej częstotliwości progowej

błąd ucięcia w czasie
błąd wynikający z rozpatrywania sygnału w przedziale czasu mniejszym od czasu jego trwania

częstotliwość graniczna
największa częstotliwość, dla której widmo sygnału przybiera wartość różną od zera

częstotliwość Nyquista
najmniejsza częstotliwość z jaką należy próbkować sygnał o ograniczonym paśmie, aby w jego próbkach została zachowana pełna informacja o sygnale

częstotliwość próbkowania
liczba próbek pobieranych w czasie 1 sekundy

dziedzina zespolona
dziedzina opisu i analizy sygnałów i układów w funkcji zmiennej zespolonej

efekt stroboskopowy
efekt występujący w przypadku, gdy sygnał okresowy jest próbkowany z częstotliwością mniejszą od częstotliwości Nyquista; polega na pobieraniu próbek pozornego sygnału o tym samym kształcie, ale spowolnionego w czasie

filtr
układ przetwarzający sygnał wejściowy w celu przekształcenia jego widma

filtr cyfrowy
urządzenie fizyczne lub program obliczeniowy przeznaczony do przetwarzania sygnałów cyfrowych; pojęcie to obejmuje cały zespół środków sprzętowych i programowych przetwarzania, a także algorytm, według którego jest ono dokonywane

filtr ochronny (antyaliasingowy)
filtr dolnoprzepustowy stosowany do odcięcia części pasma sygnału analogowego powyżej pewnej częstotliwości w celu uniknięcia aliasingu przy jego dalszym przetwarzaniu

filtr wygładzający
filtr korygujący zniekształcenia sygnału odtwarzanego z próbek metodą schodkową

jitter
zjawisko drżenia chwil próbkowania (losowego rozrzutu faktycznych chwil pobierania próbek wokół nominalnych chwil próbkowania)

kod sygnałowy impulsowy
sposób fizycznej reprezentacji znaków binarnych „1” oraz „0” za pomocą impulsów elektrycznych

kodowanie
operacja polegająca na przyporządkowaniu poszczególnym wartościom sygnału skwantowanego słów binarnych, najczęściej o ustalonej stałej długości

kwantowanie
operacja wykonywana na sygnale spróbkowanym w celu przetworzenia go w sygnał o dyskretnej strukturze amplitudowej (sygnał cyfrowy); polega na podzieleniu zakresu zmian wartości sygnału na skończoną liczbę przedziałów kwantyzacji i przybliżeniu wartości chwilowych próbek wartościami przyporządkowanymi poszczególnym przedziałom

metoda schodkowa
metoda odtwarzania sygnału analogowego z próbek polegająca na wytworzeniu w danym przedziale próbkowania stałego poziomu sygnału odpowiadającego wartości bieżącej próbki i utrzymaniu go do chwili przyjścia nowej próbki

pasma ochronne
pasma puste między pasmami poszczególnych sygnałów w systemie transmisji sygnałów ze zwielokrotnieniem częstotliwościowym

procesor sygnałowy
specjalizowany, uniwersalny, programowalny układ mikroprocesorowy, zawierający urządzenie arytmetyczne i wyposażony w pamięć, przeznaczony do szybkiego i sprawnego wykonywania różnorodnych operacji na sygnałach cyfrowych (takich jak dodawanie, mnożenie, mnożenie skalarne, opóźnianie sygnałów itd.)

próbkowanie
operacja na sygnale polegająca na pobraniu (zarejestrowaniu) wartości sygnału analogowego w dyskretnych chwilach, najczęściej równoodległych;

w wyniku próbkowania sygnał analogowy jest przetwarzany w sygnał dyskretny

próbkowanie równomierne
próbkowanie ze stałym odstępem między kolejnymi próbkami

przedział kwantyzacji (krok kwantowania)
odległość między sąsiednimi wartościami sygnału skwantowanego

przedział Nyquista
największy okres z jakim należy próbkować sygnał o ograniczonym paśmie, aby w jego próbkach została zachowana pełna informacja o tym sygnale; odwrotność częstotliwości Nyquista

przetwarzanie analogowo-cyfrowe sygnałów
zbiór operacji przetwarzających sygnał analogowy w sygnał cyfrowy

przetwarzanie cyfrowe sygnałów
zespół narzędzi teoretycznych, algorytmicznych, sprzętowych i programowych przeznaczonych do analizy i przetwarzania sygnałów z wykorzystaniem technik cyfrowych

przetwornik analogowo-cyfrowy
układ przetwarzający sygnał analogowy w sygnał cyfrowy reprezentowany ciągiem słów binarnych

przetwornik cyfrowo-analogowy
układ odtwarzający z sygnału cyfrowego, reprezentowanego ciągiem słów binarnych, postać analogową sygnału

słowo binarne
ciąg znaków binarnych o określonej długości

stosunek sygnał-szum
miara decybelowa stosunku mocy sygnału użytecznego do mocy szumu zakłócającego

sygnał informacyjny (modulujący)
sygnał, w którym zawarta jest użyteczna informacja przesyłana od nadajnika do odbiornika

sygnał o ograniczonym paśmie
sygnał, którego widmo nie zawiera składowych o częstotliwościach większych od pewnej częstotliwości granicznej

sygnał użyteczny
składowa sygnału zakłóconego, w której zawarta jest informacja użyteczna (sygnał niezakłócony)

szum biały
sygnał losowy, którego gęstość widmowa jest stała w całym paśmie częstotliwości (lub równoważnie – którego wartości są nieskorelowane)

warunek Nyquista
warunek określający częstotliwość, z jaką należy próbkować sygnał analogowy, aby w jego próbkach została zachowana pełna informacja o sygnale, tj. możliwe było dokładne odtworzenie sygnału na podstawie jego próbek; zgodnie z tym warunkiem częstotliwość próbkowania powinna być co najmniej dwa razy większa niż częstotliwość graniczna pasma sygnału

znak binarny (bit)
symbol dwójkowy, najczęściej przedstawiany umownie jako „1” lub „0”

Literatura

[1] Wojtkiewicz A.: Elementy syntezy filtrów cyfrowych. WNT, Warszawa, 1984.

[2] Lyons G.R.: Wprowadzenie do cyfrowego przetwarzania sygnałów. WKŁ, Warszawa, 1999.

[3] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom III. WNT, Warszawa, 1995.

Lekcja 7

Przetwarzanie sygnałów analogowych przez układy LS

Tematem lekcji 7 jest zagadnienie przetwarzania sygnałów analogowych przez układy LS. W zastosowaniach praktycznych sygnały są przetwarzane przez różnego rodzaju urządzenia fizyczne, których zadaniem jest przekształcenie sygnału w inny sygnał o pożądanych cechach. W celu opisu działania tych urządzeń i scharakteryzowania ich właściwości wprowadza się odpowiednie modele matematyczne nazywane *układami*.

W znaczeniu potocznym termin „układ" jest używany zwykle na określenie konkretnego urządzenia fizycznego, złożonego z pewnych elementów i realizującego założoną funkcję. W lekcji 7 pojęcie układu będziemy rozumieć jako model matematyczny urządzenia przeznaczonego do przetwarzania sygnałów. Szczególną uwagę poświęcimy układom liniowym stacjonarnym (układom LS).

Poznamy opis układu w dziedzinie czasu, dziedzinie zespolonej i dziedzinie częstotliwości. Wprowadzimy pojęcia podstawowych charakterystyk opisujących układ w każdej z tych dziedzin, tj. odpowiedzi impulsowej, transmitancji i charakterystyki amplitudowo-fazowej układu. Podkreślimy rolę układu LS jako filtru kształtującego widmo sygnału wyjściowego. Przedyskutujemy zagadnienie wyznaczania sygnału na wyjściu układu dla różnych klas sygnałów wejściowych. Omówimy także przykłady typowych prostych układów.

7.1. Pojęcie układu. Podstawowe definicje

Pojęcie układu jest szeroko wykorzystywane m.in. w teorii obwodów i teorii systemów. W teorii obwodów układ jest rozumiany jako model matematyczny obwodu elektrycznego, a ogólniej – sieci elektrycznej. W teorii systemów pod pojęciem układu rozumie się z reguły model matematyczny urządzenia rozpatrywanego w konwencji „wejście-wyjście". Konwencję tę stosuje się także w teorii sygnałów. Układ traktuje się jako „czarną skrzynkę" (rys. 7.1), na której wejściu występuje *sygnał wejściowy* x, nazywany *pobudzeniem*, na wyjściu zaś – *sygnał wyjściowy* y, nazywany *odpowiedzią* układu. Konwencja ta jest nazywana w literaturze *ujęciem transmisyjnym* (por. np. [1], p. 2.1.5-H oraz [2], p. 9.2.1).

Rys. 7.1. Ogólny schemat układu

W ogólnym przypadku można rozpatrywać układy o wielu wejściach i wielu wyjściach i przyjmować modele wektorowe sygnałów wejściowych i wyjściowych. W lekcji ograniczymy się do najprostszego przypadku układów o jednym wejściu i jednym wyjściu, na których występują sygnały skalarne.

7.1.1. Układ jako operator. Układy analogowe i dyskretne

Istota ujęcia transmisyjnego polega na przypisaniu układowi roli pewnej operacji i opisaniu jego działania za pomocą operatora T odwzorowującego według określonej reguły sygnał wejściowy x w sygnał wyjściowy y:

$$y = T[x]. \tag{7.1}$$

Będziemy zakładać, że operator T jest określony na ustalonym zbiorze sygnałów wejściowych X (zbiorze tzw. *dopuszczalnych sygnałów wejściowych*) i odwzorowuje go w zbiór sygnałów wyjściowych Y (zbiór *dopuszczalnych sygnałów wyjściowych*). Odwzorowanie to można zapisać symbolicznie w postaci $Y = T[X]$.

Symbole x i y we wzorze (7.1) mają charakter uniwersalny i mogą odnosić się do różnych klas sygnałów. Jeśli dziedzina X i przeciwdziedzina Y operatora T są pewnymi zbiorami sygnałów ciągłych w czasie, układ nazywamy *analogowym*. Jeśli X i Y są zbiorami sygnałów dyskretnych w czasie, układ nazywamy *dyskretnym*.

7.1.2. Układy liniowe

Operatory opisujące układy mogą mieć różne właściwości formalne. Właściwości te określają poszczególne klasy układów. Podstawową klasą układów, spełniającą w zagadnieniach przetwarzania sygnałów szczególnie ważną rolę, jest klasa układów opisywanych przez *operatory liniowe stacjonarne* (*operatory LS*).

Definicja 7.1. Operator T nazywamy *liniowym*, jeśli dla dowolnych sygnałów $x_1, x_2 \in X$ i dowolnych liczb rzeczywistych $\alpha_1, \alpha_2 \in \mathscr{R}$ zachodzi równość:

$$T[\alpha_1 x_1 + \alpha_2 x_2] = \alpha_1 T[x_1] + \alpha_2 T[x_2]. \tag{7.2}$$

Układ opisany przez operator liniowy jest nazywany *układem liniowym*. Jeśli operator T nie spełnia równości (7.2), opisany przez niego układ jest *układem nieliniowym*.

Przykład 7.1. Operatory różniczkowania $y(t) = \mathrm{d}x(t)/\,\mathrm{d}t$, całkowania $y(t) = \int x(t)\,\mathrm{d}t$ i opóźnienia $y(t) = x(t - t_0)$ o czas $t_0 > 0$ w dziedzinie sygnałów analogowych są operatorami liniowymi. Układy różniczkujący, całkujący i opóźniający są zatem układami liniowymi.

Przykład 7.2. Operatory mnożenia skalarnego $y[n] = \alpha x[n]$ i opóźnienia sygnału o jedną próbkę $y[n] = x[n-1]$ w dziedzinie sygnałów dyskretnych są operatorami liniowymi. Układy dyskretne realizujące operacje mnożenia skalarnego i opóźnienia są zatem układami liniowymi.

Przykład 7.3. Opierając się na definicji (7.1) można łatwo sprawdzić, że operatory: podnoszenia do kwadratu $y = x^2$, pierwiastkowania $y = \sqrt{x}$ oraz logarytmowania $y = \log x$, określone w dziedzinie sygnałów analogowych lub sygnałów dyskretnych, są operatorami nieliniowymi. Układy: podnoszący do kwadratu, pierwiastkujący i logarytmujący są zatem układami nieliniowymi.

Ściśle rzecz biorąc, wszystkie istniejące w rzeczywistości układy fizyczne są nieliniowe. Charakterystyki realnych układów nie mogą być bowiem dokładnie liniowe w całym zakresie zmian sygnałów wejściowych i wyjściowych od $-\infty$ do ∞. Jednak jeśli sygnały zmieniają się w ograniczonym zakresie, wiele praktycznych układów można opisywać dostatecznie dokładnie modelami liniowymi. Nieliniowe układy przetwarzania sygnałów zawierają zwykle elementy nieliniowe, takie jak diody, tranzystory itp. Odgrywają one w przetwarzaniu sygnałów niemniejszą rolę jak układy liniowe i są wykorzystywane m.in. do modulacji i demodulacji sygnałów, a także wielu innych ważnych operacji na sygnałach. Teoria układów nieliniowych jest bardzo złożona i nie wszystkie rozpatrywane w niej zagadnienia mogą być rozwiązane metodami analitycznymi. Problematyce układów nieliniowych są poświęcone osobne opracowania (np. [3], rozdz. 11). Dalsze nasze rozważania ograniczymy do klasy układów liniowych.

7.1.3. Układy stacjonarne

W celu zdefiniowania pojęcia stacjonarności układu wprowadzimy najpierw definicję operatora przesunięcia w czasie.

Definicja 7.2. *Operatorem przesunięcia* sygnału o czas $t_0 \in \mathscr{R}$ w dziedzinie sygnałów analogowych nazywamy operator:

$$P_{t_0}[x(t)] = x(t - t_0). \tag{7.3}$$

Operatorem przesunięcia sygnału o $n_0 \in \mathscr{I}$ próbek w dziedzinie sygnałów dyskretnych nazywamy operator:

$$P_{n_0}\{x[n]\} = x[n - n_0]. \tag{7.4}$$

Definicja 7.3. Operator T określony w dziedzinie sygnałów analogowych nazywamy *stacjonarnym*, jeśli dla każdego sygnału $x(t)$ i każdego przesunięcia t_0 jest on przemienny z operatorem przesunięcia P_{t_0}:

$$T\{P_{t_0}[x(t)]\} = P_{t_0}\{T[x(t)]\}. \tag{7.5}$$

Podobnie, operator T określony w dziedzinie sygnałów dyskretnych nazywamy *stacjonarnym*, jeśli dla każdego sygnału $x[n]$ i każdego przesunięcia n_0 jest on przemienny z operatorem przesunięcia P_{n_0}:

$$T\{P_{n_0}[x[n]]\} = P_{n_0}\{T[x[n]]\}. \tag{7.6}$$

Układ opisany przez operator stacjonarny jest nazywany *układem stacjonarnym*. Jeśli operator T nie spełnia równości (7.5) lub (7.6), to opisany przez niego układ jest *układem niestacjonarnym*.

Z definicji stacjonarności wynika, że dla układu stacjonarnego kolejność wykonania operacji T i operacji przesunięcia P jest dowolna. Ten sam rezultat uzyskamy w przypadku, gdy operację T wykonamy na sygnale przesuniętym w czasie, jak i w przypadku, gdy operację T wykonamy najpierw na sygnale nieprzesuniętym, a następnie otrzymany sygnał przesuniemy w czasie. Zależności (7.5) i (7.6) oznaczają także, że jeśli $T[x(t)] = y(t)$, to $T[x(t - t_0)] = y(t - t_0)$ oraz jeśli $T\{x[n]\} = y[n]$, to $T\{x[n - n_0]\} = y[n - n_0]$.

Przykład 7.4. Można łatwo sprawdzić, że operatory różniczkowania, całkowania i opóźnienia w dziedzinie sygnałów analogowych są operatorami stacjonarnymi, a więc układy różniczkujący, całkujący i opóźniający są stacjonarne.

Przykład 7.5. Operatory mnożenia skalarnego $y[n] = \alpha x[n]$ i opóźnienia sygnału o jedną próbkę $y[n] = x[n - 1]$ w dziedzinie sygnałów dyskretnych są operatorami stacjonarnymi, a więc układy dyskretne mnożenia skalarnego i opóźniający są stacjonarne.

Przykład 7.6. Korzystając z podanych definicji, można pokazać, że układy: podnoszący do kwadratu, pierwiastkujący i logarytmujący są układami stacjonarnymi. Natomiast układ analogowy opisany np. równaniem $y(t) = tx(t)$ i układ dyskretny opisany odpowiednikiem tego równania $y[n] = nx[n]$ są układami niestacjonarnymi.

7.1.4. Układy LS

Podane wyżej definicje umożliwiają wydzielenie najbardziej podstawowej klasy układów, jaką jest klasa układów LS.

Definicja 7.4. Układy opisane przez operatory liniowe i stacjonarne nazywamy *układami liniowymi stacjonarnymi* lub krótko *układami LS*.

Właściwości liniowości i stacjonarności umożliwiają, jak pokażemy dalej, wprowadzenie jednolitego opisu matematycznego układów LS. Dzięki temu większość zagadnień teorii tych układów może być rozwiązana metodami analitycznymi.

Komentarz. Podane wyżej definicje liniowości i stacjonarności układu są w pełni analogiczne do znanych z teorii obwodów definicji liniowości i stacjonarności obwodu elektrycznego. Pojęcie stacjonarności układu należy jednak odróżniać od pojęcia stacjonarności sygnału wprowadzanego w teorii sygnałów losowych. W języku angielskim w odniesieniu do układu stacjonarnego stosowany jest termin *time-invariant system*, natomiast w odniesieniu do sygnału stacjonarnego – *stationary signal*. W nazewnictwie polskim oba pojęcia stacjonarności są określane tym samym terminem i nie należy ich mylić.

7.1.5. Układy przyczynowe

W każdym fizycznym urządzeniu przetwarzającym sygnały pojawienie się na wyjściu sygnału wyjściowego nie może nastąpić przed podaniem na jego wejście sygnału wejściowego. Jeśli zatem układ, jako model matematyczny urządzenia fizycznego, ma w sposób adekwatny opisywać jego działanie, to powinien spełniać *zasadę przyczynowości*, która orzeka, że skutek nie może poprzedzać przyczyny, tzn. odpowiedź na wyjściu układu nie może pojawić się zanim na jego wejście nie zostanie podane pobudzenie. Formalną definicję pojęcia przyczynowości układu można sformułować w następującym brzmieniu.

Definicja 7.5. Załóżmy, że układ jest opisany operatorem T odwzorowującym zbiór sygnałów X na zbiór sygnałów Y oraz $y_1 = T[x_1]$ i $y_2 = T[x_2]$. Operator T określony w zbiorze sygnałów analogowych X nazywamy *operatorem przyczynowym*, jeśli dla każdych $x_1(t), x_2(t) \in X$ i każdego $t_0 \in \mathscr{R}$ z równości

$$x_1(t) = x_2(t) \quad \text{dla} \quad t < t_0$$

wynika równość
$$y_1(t) = y_2(t) \quad \text{dla} \quad t < t_0.$$

Operator T określony w zbiorze sygnałów dyskretnych X nazywamy *operatorem przyczynowym*, jeśli dla każdych $x_1[n], x_2[n] \in X$ i każdego $n_0 \in \mathscr{I}$ z równości
$$x_1[n] = x_2[n] \quad \text{dla} \quad n < n_0$$
wynika równość
$$y_1[n] = y_2[n] \quad \text{dla} \quad n < n_0.$$

Definicja 7.6. Układ opisany operatorem przyczynowym jest nazywany *układem przyczynowym*. W przeciwnym przypadku jest układem *nieprzyczynowym*.

Z podanych definicji wynika, że dla układu przyczynowego z równości $x(t) \equiv 0$ dla $t < t_0$ ($x[n] \equiv 0$ dla $n < n_0$) wynika równość $y(t) \equiv 0$ dla $t < t_0$ ($y[n] \equiv 0$ dla $n < n_0$) dla każdego $t_0 \in \mathbb{R}$ (każdego $n_0 \in \mathbb{I}$). Oznacza to, iż – zgodnie z podaną wcześniej definicją słowną – odpowiedź układu przyczynowego nie może poprzedzać pobudzenia.

Przykład 7.7. Układ analogowy opóźniający sygnał o czas $t_0 > 0$ o równaniu $y(t) = x(t - t_0)$ oraz układ dyskretny opóźniającyukład!opóźniający!dysktretny sygnał o $n_0 > 0$ próbek o równaniu $y[n] = x[n - n_0]$ są układami przyczynowymi. Układy przyspieszające sygnał opisane równaniami $y(t) = x(t + t_0)$, $t_0 > 0$, ($y[n] = x[n + n_0]$, $n_0 > 0$) są układami nieprzyczynowymi.

Mimo że układy nieprzyczynowe nie opisują urządzeń fizycznie realizowalnych, są niekiedy wykorzystywane jako modele matematyczne procesu przetwarzania sygnału. Sytuacja taka ma miejsce np. w przypadku filtrów cyfrowych pracujących w trybie *off line*, kiedy przetwarzanie sygnału nie odbywa się w czasie rzeczywistym. Sposób organizacji obliczeń realizowanych przez filtr dopuszcza wówczas korzystanie przy wyznaczaniu bieżącej próbki sygnału wyjściowego filtru w chwili n zarówno próbek sygnału wejściowego przed chwilą n, jak i zarejestrowanych wcześniej próbek tego sygnału po chwili n.

7.1.6. Układy skupione i układy o stałych rozłożonych

Liniowość, stacjonarność i przyczynowość są immanentnymi cechami układu, zależnymi jedynie od jego struktury wewnętrznej oraz elementów z jakich jest on zbudowany i niezależnymi od klasy przetwarzanych przez niego sygnałów. Inne ważne kryterium podziału układów, nie odnoszące się bezpośrednio do właściwości operatora układu, wynika z porównania wymiarów geometrycznych modelowanego przez niego układu fizycznego z długością fali odpowiadającą największej częstotliwości widma przetwarzanego sygnału. Kryterium to jest związane ze znanym z teorii obwodów warunkiem *quasi-stacjonarności* (por. [1], p. 1.4) i prowadzi do podziału układów na układy skupione i układy o stałych rozłożonych.

Definicja 7.7. Jeśli największy wymiar geometryczny układu fizycznego jest dużo mniejszy od najmniejszej długości rozchodzącej się w nim fali, to opisujący go model nazywamy *układem skupionym* (*o stałych skupionych*). Jeśli warunek ten nie jest spełniony, opisujący go model nazywamy *układem o stałych rozłożonych*.

W zakresie częstotliwości przetwarzanych sygnałów do setek megaherców większość stosowanych w praktyce układów fizycznych jest w sposób adekwatny opisywana modelami skupionymi. Zakłada się wówczas, że czas propagacji sygnału wewnątrz układu jest nieskończenie mały. W przypadku częstotliwości rzędu gigaherców wymiary geometryczne układów przestają być dużo mniejsze od długości fali i metody opisu takich układów powinny uwzględniać zjawiska falowe związane ze skończonym czasem propagacji sygnałów. W zależności od pasma częstotliwości przetwarzanych sygnałów ten sam układ fizyczny może być raz traktowany jako układ skupiony, innym zaś razem jako układ o stałych rozłożonych.

Przykład 7.8. Analogowy układ opóźniającyukład!opóźniający!analogowy o równaniu $y(t) = x(t - t_0)$, $t_0 > 0$, jest układem liniowym, stacjonarnym i przyczynowym (por. przykłady 7.1, 7.4 i 7.7). Jak pokażemy w p. 7.5.1, układ ten nie jest jednak układem skupionym.

Do układów o stałych rozłożonych należą również linie długie oraz urządzenia mikrofalowe, takie jak rezonatory i falowody pracujące w zakresie bardzo wielkich częstotliwości. Problematyka przetwarzania sygnałów przez te układy jest na ogół złożona i nie będziemy się nią zajmować. W dalszych rozważaniach skoncentrujemy się zatem na dobrze znanej z podstawowego kursu teorii obwodów klasie układów skupionych, liniowych i stacjonarnych, nazywanymi *układami SLS*.

7.1.7. Układ jako filtr

Wraz z przetworzeniem sygnału przez układ następuje przekształcenie jego widma. Niektóre składowe częstotliwościowe widma sygnału mogą być przez układ uwypuklone, inne stłumione bądź całkowicie usunięte, a więc, jak mówimy – odfiltrowane. Bardzo często od projektowanego układu żądamy, aby przekształcał on widmo sygnału wejściowego w widmo sygnału na wyjściu o pożądanym kształcie.

Operacja przetworzenia widma sygnału jest nazywana *filtracją*. Z tego względu w teorii sygnałów układ jest często utożsamiany z *filtrem*, a układy nazywane filtrami. W szczególności układy LS są nazywane *filtrami LS*. W dalszym ciągu nazw tych będziemy używać wymiennie.

Komentarz. Pojęcie filtracji jest rozumiane w teorii sygnałów także w innym znaczeniu. W teorii sygnałów losowych filtracją jest nazywana procedura prognozowania (estymacji) bieżącej wartości sygnału w danej chwili na podstawie jego obserwacji przed tą chwilą. W dalszych rozważaniach filtrację będziemy zawsze rozumieć jako operację przekształcenia widma sygnału.

7.1.8. Opis układu w dziedzinie czasu

Przystąpimy obecnie do omówienia sposobów matematycznego opisu układów analogowych. Przyjmiemy założenie, że rozpatrywane układy są skupionymi przyczynowymi układami LS (będziemy je dalej nazywać krótko układami). Odstępstwa od tego założenia będziemy każdorazowo wyraźnie zaznaczać.

Rozpatrzymy trzy rodzaje opisu układów analogowych: w dziedzinie czasu, w dziedzinie zespolonej oraz w dziedzinie częstotliwości. Podamy definicje podstawowych charakterystyk układu w tych dziedzinach. Omówimy także metody wyznaczania sygnału na wyjściu układu na podstawie znajomości jego odpowiedniej charakterystyki oraz sygnału wejściowego. Przypomnimy w tym celu ważne pojęcie splotu. Pokażemy, że sposób wyznaczania odpowiedzi układu na pobudzenie sygnałem wejściowym zależy od klasy rozpatrywanych układów.

7.1.9. Odpowiedź impulsowa

Koncepcja opisu układu w dziedzinie czasu opiera się na następującym rozumowaniu. Przypuśćmy, że w celu zbadania właściwości układów i porównania ich działania pobudzamy je na wejściu w chwili $t = 0$ pewnym ustalonym standardowym sygnałem i obserwujemy na wyjściu odpowiedzi na to pobudzenie. W ogólnym przypadku odpowiedzi te zależą nie tylko od sygnału pobudzającego, lecz również od warunków początkowych układu (por. np. [1], p. 2.2.1-B, C). Aby możliwe było jednoznaczne określenie reakcji układu na pobudzenie zewnętrzne w zależności jedynie od jego struktury wewnętrznej, a więc niezależnie od tego w jakim stanie znajdował się on przed tym pobudzeniem, przyjmuje się, że przed chwilą $t = 0$ był on w stanie spoczynku, tzn. zakłada się, że w chwili tej w układzie nie była zgromadzona energia. Założenie to jest równoważne przyjęciu zerowych warunków początkowych w chwili $t = 0$.

Przy założeniu zerowych warunków początkowych każdy układ odpowiada na ustalone testowe pobudzenie w charakterystyczny dla siebie sposób. Odpowiedź na to pobudzenie możemy zatem przyjąć jako pewną charakterystykę układu opisującą w dziedzinie czasu jego działanie z punktu widzenia relacji „wejście-wyjście”. Jako testowe sygnały pobudzające układ przyjmuje się dwa standardowe sygnały: impuls Diraca $\delta(t)$ oraz skok jednostkowy $\boldsymbol{1}(t)$.

Definicja 7.8. *Odpowiedzią impulsową* $h(t)$ układu nazywamy sygnał, jaki wystąpi na jego wyjściu, jeśli będzie on pobudzany impulsem Diraca $\delta(t)$ przy zerowych warunkach początkowych w chwili $t = 0$ (rys. 7.2a).

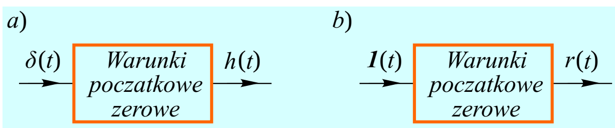
Znaczenie odpowiedzi impulsowej jako charakterystyki układu wynika z faktu, że znając tę charakterystykę można wyznaczyć odpowiedź układu na dowolne pobudzenie wejściowe. Zagadnienie wyznaczania sygnału na wyjściu układu jest jednym z podstawowych zagadnień teorii sygnałów związanym ściśle z pojęciem splotu. Omówimy je dokładniej w p. 7.1.14.

7.1.10. Odpowiedź jednostkowa

Odpowiedź impulsowa układu, podobnie jak impuls Diraca, jest pojęciem wyidealizowanym. Tak jak impuls Diraca, ma ona sens gęstości sygnału przypadającej na jednostkę czasu, a więc jej wymiarem jest $[\mathrm{V/s}]$, $[\mathrm{A/s}]$ itd. Ponieważ impulsu Diraca nie można wytworzyć w układzie fizycznym, odpowiedzi impulsowej nie można zaobserwować na wyjściu układu rzeczywistego. Jest ona pewnym abstraktem, którym posługujemy się w celu matematycznego opisu układu. Z tego względu, jako alternatywny standardowy sygnał pobudzający układ przyjmuje się skok jednostkowy $\boldsymbol{1}(t)$ i wprowadza się pojęcie odpowiedzi jednostkowej układu.

Definicja 7.9. *Odpowiedzią jednostkową* $r(t)$ układu nazywamy sygnał, jaki wystąpi na jego wyjściu, jeśli będzie on pobudzany skokiem jednostkowym $\boldsymbol{1}(t)$ przy zerowych warunkach początkowych w chwili $t = 0$ (rys. 7.2b).

W przeciwieństwie do impulsu Diraca skok jednostkowy można z dobrym przybliżeniem wytworzyć fizycznie. Odpowiedź jednostkowa może więc być obserwowana i badana na wyjściu układu fizycznego. Podobnie jak w przypadku odpowiedzi impulsowej, jej znajomość jest dostateczna, aby wyznaczyć odpowiedź układu na dowolne pobudzenie wejściowe. W tym sensie odpowiedź jednostkowa jest równoważną do odpowiedzi impulsowej charakterystyką układu opisującą go w dziedzinie czasu.



Rys. 7.2. Interpretacja odpowiedzi impulsowej i odpowiedzi jednostkowej układu

7.1.11. Odpowiedź impulsowa układu przyczynowego

Impuls Diraca jest sygnałem przybierającym wartości zerowe dla $t < 0$. Jeśli układ jest przyczynowy, to z warunku przyczynowości opisującego go operatora, określonego definicją 6.5, wynika, że również:

$$h(t) = 0 \quad \text{dla} \quad t < 0. \tag{7.7}$$

Ponieważ każdy realnie istniejący układ jest układem przyczynowym, warunek (7.7) jest nazywany w literaturze *podstawowym warunkiem realizowalności* układu wyrażonym w kategoriach jego odpowiedzi impulsowej. Warunek ten można sformułować także w równoważnej postaci w odniesieniu do odpowiedzi jednostkowej:

$$r(t) = 0 \quad \text{dla} \quad t < 0. \tag{7.8}$$

Komentarz. Dokładnie rzecz biorąc, warunki (7.7) i (7.8) są warunkami koniecznymi przyczynowości układu. Można jednak pokazać, że przy stosunkowo łagodnych założeniach dotyczących klasy sygnałów wejściowych, które są spełnione przez większość interesujących w praktyce sygnałów (ograniczoność sygnału w każdym skończonym przedziale $[0, T]$, $T > 0$ – por. [2], p. 9.2.3-A), warunki te są zarazem dostateczne. W takich przypadkach można je przyjąć za definicje przyczynowości układu równoważne definicjom 7.5 i 7.6.

7.1.12. Związki między odpowiedzią impulsową a odpowiedzią jednostkową

Jeżeli T jest operatorem układu, to zgodnie z definicjami odpowiedzi impulsowej i odpowiedzi jednostkowej $h(t) = T[\delta(t)]$ oraz $r(t) = T[\boldsymbol{1}(t)]$. Z drugiej strony, ponieważ impuls Diraca jest pochodną skoku jednostkowego (por. wzór (1.18)), możemy napisać:

$$h(t) = T\left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{1}(t)\right]. \tag{7.9}$$

Uwzględniając liniowość i stacjonarność operatora T, możemy zamienić kolejność wykonywania operacji T i operacji różniczkowania. Dokonując tej zamiany, otrzymujemy:

$$h(t) = \frac{\mathrm{d}}{\mathrm{d}t}T[\boldsymbol{1}(t)] = \frac{\mathrm{d}}{\mathrm{d}t}r(t). \tag{7.10}$$

Wynika stąd, że odpowiedź impulsowa układu jest pochodną odpowiedzi jednostkowej. Odpowiedź jednostkowa jest zatem całką odpowiedzi impulsowej:

$$r(t) = \int_0^t h(t')\,\mathrm{d}t'. \tag{7.11}$$

Pokazaliśmy tym samym, że odpowiedzi impulsowa i jednostkowa układu są związane zależnościami (7.10) i (7.11). Zależności te wymagają jednak dodatkowego omówienia. Cytując związki (1.17) i (1.18) między skokiem jednostkowym a dystrybucją Diraca podkreślaliśmy, że należy je rozumieć w sensie dystrybucyjnym. Zwykła pochodna skoku jednostkowego jest bowiem równa wszędzie zeru, poza punktem $t = 0$, w którym nie istnieje. Pochodna skoku jednostkowego istnieje natomiast w sensie dystrybucyjnym (por. [2], p. 9.1.2-D) i jest równa dystrybucji Diraca. Z tych samych względów pochodną we wzorze (7.10) należy w ogólnym przypadku rozumieć w sensie dystrybucyjnym. Pochodną tę można traktować jako zwykłą pochodną, jeśli odpowiedź jednostkowa $r(t)$ jest funkcją ciągłą i prawie wszędzie różniczkowalną dla $t \geqslant 0$. Jeśli natomiast $r(t)$ jest funkcją ciągłą i prawie wszędzie różniczkowalną dla $t > 0$ (co jest z reguły spełnione dla realnie istniejących układów), to jedyny możliwy jej skok występuje w punkcie $t = 0$. Zależność (7.10) przybiera wówczas postać:

$$h(t) = r'(t) + r(0+)\delta(t), \tag{7.12}$$

gdzie $r(0+)$ jest granicą prawostronną funkcji $r(t)$, gdy $t \to 0$, a $r'(t)$ oznacza zwykłą pochodną.

Podobnie, w zależności (7.11) całka ma zwykły sens, jeśli funkcja $h(t)$ nie zawiera składników dystrybucyjnych. W przeciwnym razie całkę we wzorze (7.11) należy rozumieć jako granicę:

$$\int_0^t h(t')\,\mathrm{d}t' \triangleq \lim_{\varepsilon \to 0+} \int_{-\varepsilon}^t h(t')\,\mathrm{d}t'. \tag{7.13}$$

Tak więc, jeśli $r(0+) = 0$ i w odpowiedzi impulsowej nie występuje składnik dystrybucyjny, to przy wyznaczaniu odpowiedzi jednostkowej na podstawie odpowiedzi impulsowej (lub odwrotnie) możemy skorzystać z zależności (7.10) i (7.11). W przypadku, gdy odpowiedź impulsowa zawiera składnik dystrybucyjny, w obliczeniach należy zachować ostrożność i przeprowadzać je w oparciu o zależności (7.12) i (7.13).

7.1.13. Ogólna postać odpowiedzi impulsowej

Można pokazać (uzasadnimy to w p. 7.2.4), że dla klasy przyczynowych układów SLS, do której ograniczyliśmy nasze rozważania, odpowiedź impulsowa ma ogólną postać:

$$h(t) = a_n\delta^{(n)}(t) + \ldots + a_1\delta'(t) + a_0 + h_0(t), \tag{7.14}$$

gdzie $\delta^{(k)}(t)$ jest k-tą pochodną (dystrybucyjną) impulsu Diraca, a $h_0(t)$ jest zwykłą funkcją spełniającą warunek $h_0(t) = 0$ dla $t < 0$.

Układy, których odpowiedź impulsowa zawiera jako składniki pochodne impulsu Diraca nie mają jednak swoich odpowiedników wśród układów realnych, a więc ich znaczenie jest wyłącznie teoretyczne. W dalszym ciągu możemy zatem ograniczyć się do przypadku, gdy

$$h(t) = a_0\delta(t) + h_0(t). \tag{7.15}$$

Dla przyczynowych układów SLS, będących adekwatnymi modelami układów rzeczywistych, odpowiedź impulsowa jest zatem sumą składnika dystrybucyjnego $a_0\delta(t)$ i zwykłej funkcji $h_0(t)$. Przypadek ten odpowiada postaci odpowiedzi impulsowej określonej wzorem (7.12). Z porównania wzorów (7.12) i (7.14) wynika, że $a_0 = r(0+)$ oraz $h_0(t) = r'(t)$.

7.1.14. Związek między sygnałami na wejściu i wyjściu układu. Zależność splotowa

Jak podkreśliliśmy w p. 7.1.9, wyjątkowo ważna rola odpowiedzi impulsowej $h(t)$ jako charakterystyki opisującej układ polega na tym, iż znając tę funkcję oraz sygnał wejściowy $x(t)$, możemy wyznaczyć sygnał $y(t)$ na wyjściu układu. Pokażemy teraz jaką operację należy wykonać na sygnałach $x(t)$ i $h(t)$, aby wyznaczyć sygnał wyjściowy $y(t)$.

Punktem wyjścia jest właściwość splotu dystrybucji Diraca (por. wzór (1.19)), zgodnie z którą dystrybucja $\delta(t)$ jest elementem identycznościowym operacji splotu, tj. $\delta(t) * x(t) = x(t)$. Zapisując tę równość w jawnej postaci, otrzymujemy:

$$x(t) = \int\limits_{-\infty}^{\infty} \delta(t-\tau)x(\tau)\,\mathrm{d}\tau. \tag{7.16}$$

Odpowiedź układu opisanego operatorem T na sygnał $x(t)$ można zatem przedstawić w postaci:

$$y(t) = T[x(t)] = T\left[\int\limits_{-\infty}^{\infty} \delta(t-\tau)x(\tau)\,\mathrm{d}\tau\right].$$

Traktując całkę jako granicę sumy i uwzględniając liniowość układu, możemy wprowadzić symbol operatora T pod znak całki:

$$y(t) = \int\limits_{-\infty}^{\infty} T\left[\delta(t-\tau)x(\tau)\right]\,\mathrm{d}\tau.$$

Ponieważ operator T działa na wielkość zależną od zmiennej t, a nie oddziaływuje na wielkości zależne od zmiennej całkowania τ, możemy dalej napisać:

$$y(t) = \int\limits_{-\infty}^{\infty} T\left[\delta(t-\tau)\right]x(\tau)\,\mathrm{d}\tau.$$

Uwzględniając z kolei, że ze stacjonarności układu wynika, iż równość $T[\delta(t)] = h(t)$ implikuje równość $T[\delta(t-\tau)] = h(t-\tau)$, otrzymujemy ostatecznie:

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)\,\mathrm{d}\tau = h(t) * x(t). \tag{7.17}$$

Tak więc, sygnał na wyjściu układu jest splotem odpowiedzi impulsowej i sygnału wejściowego. Ponieważ operacja splotu jest przemienna, zależność (7.17) możemy także zapisać w postaci równoważnej:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,\mathrm{d}\tau. \tag{7.18}$$

Zależność splotowa (7.17) lub (7.18) określająca związek między sygnałem wyjściowym a sygnałem wejściowym, jest konsekwencją liniowości i stacjonarności układu. Wzory (7.17) i (7.18) stanowią podstawę do obliczania w dziedzinie czasu odpowiedzi układów LS na dowolne pobudzenie wejściowe.

Należy podkreślić, że nie dla wszystkich sygnałów całka splotowa jest zbieżna. Aby splot istniał dla każdego t (lub prawie każdego t), sygnały splatane muszą spełniać określone warunki. Warunków tych nie będziemy tu omawiać (ich dyskusję można znaleźć np. w [2] p. 9.1.1-A lub [4] p. 2.8.2), dodamy jedynie, że jeśli odpowiedź impulsowa układu ma postać określoną wzorem (7.15), to zgodnie z liniowością całki splotowej i właściwością splotu dystrybucji Diraca, zależności (7.17) i (7.18) przybierają postać:

$$y(t) = a_0x(t) + \int_{-\infty}^{\infty} h_0(t-\tau)x(\tau)\,\mathrm{d}\tau = a_0x(t) + \int_{-\infty}^{\infty} h_0(\tau)x(t-\tau)\,\mathrm{d}\tau, \tag{7.19}$$

gdzie $h_0(t)$ jest częścią funkcyjną odpowiedzi impulsowej.

7.1.15. Zależność splotowa w przypadku układu przyczynowego

Ze względu na to, że w zależnościach (7.17) i (7.18) całkowanie przebiega w najszerszych możliwych granicach od $-\infty$ do ∞, noszą one nazwę *splotu dwustronnego*. Odnoszą się one zatem do przypadku, gdy obie splatane funkcje, a więc także odpowiedź impulsowa $h(t)$, przybierają wartości niezerowe na całej osi czasu. Korzystając z tych zależności, można zatem obliczać sygnał na wyjściu układów nieprzyczynowych. W przypadku układów przyczynowych odpowiedź impulsowa $h(t)$ dla ujemnych wartości argumentu przybiera wartości zerowe. Granice całkowania we wzorach (7.17) i (7.18) ulegają wówczas odpowiedniemu zawężeniu. Jak można łatwo sprawdzić (por. p. H), dla układów przyczynowych wzory te przybierają postać:

$$y(t) = \int_{-\infty}^{t} h(t-\tau)x(\tau)\,\mathrm{d}\tau = \int_{0}^{\infty} h(\tau)x(t-\tau)\,\mathrm{d}\tau. \tag{7.20}$$

Jeśli również sygnał wejściowy jest równy zeru dla $t < 0$, sygnał na wyjściu układu jest określony *splotem jednostronnym*:

$$y(t) = \int_{0}^{t} h(t-\tau)x(\tau)\,\mathrm{d}\tau = \int_{0}^{t} h(\tau)x(t-\tau)\,\mathrm{d}\tau. \tag{7.21}$$

Podkreślmy, że związek między sygnałem wyjściowym układu a sygnałem wejściowym może być także wyrażony za pomocą odpowiedzi jednostkowej $r(t)$. Prowadzi to do zależności nazywanej *całką Duhamela* (por. np. [2], p. 9.2.3-C). Zależność ta jest jednak mniej użyteczna do wyznaczania sygnału na wyjściu układu.

7.1.16. Obliczanie całki splotowej

Jak wynika z przytoczonych wzorów, w celu obliczenia całki splotowej należy wykonać następujące operacje na splatanych sygnałach:

- zmienić argument obu sygnałów z t na τ, tj. na zmienną całkowania,
- odbić zwierciadlanie względem osi rzędnych jeden ze splatanych sygnałów, tzn. zmienić jego argument z τ na $-\tau$,
- przesunąć sygnał odbity o czas t, tzn. wziąć go od argumentu $t-\tau$,
- pomnożyć odbity i przesunięty sygnał przez drugi sygnał i obliczać pola pod ich iloczynem dla kolejnych wartości przesunięcia t.

Otrzymany w wyniku sygnał jest w ogólnym przypadku funkcją zmiennej t zarówno przez zmienną podcałkową, jak i przez granice całkowania. Bardzo często, zwłaszcza w przypadku sygnałów impulsowych, granice całkowania zmieniają się dla różnych wartości argumentu t i całkę splotową należy wówczas rozbić na sumę całek w poszczególnych przedziałach zmienności tego argumentu. Niekiedy pomocne jest w takich przypadkach poparcie obliczeń odpowiednią konstrukcją graficzną, bądź wręcz obliczenie splotu metodą graficzną.

Przykład 7.9. Wyznaczymy sygnał $y(t)$ na wyjściu układu opisanego odpowiedzią impulsową $h(t) = \delta(t) - \alpha\,\mathrm{e}^{-\alpha t}\,\boldsymbol{1}(t)$ i pobudzanego impulsem prostokątnym $x(t) = \sqcap(\frac{t-T/2}{T})$.

Odpowiedź impulsowa ma w tym przypadku postać (7.15), a więc sygnał wyjściowy jest określony ogólnymi wzorami (7.19). Ponieważ oba splatane sygnały są równe zeru dla $t < 0$, splot sygnału wejściowego $x(t)$ z częścią funkcyjną odpowiedzi impulsowej $h_0(t) = -\alpha\,\mathrm{e}^{-\alpha t}\,\boldsymbol{1}(t)$ obliczamy w granicach od 0 do t:

$$y(t) = x(t) + \int_0^t h_0(t-\tau)x(\tau)\,\mathrm{d}\tau = x(t) - \alpha\int_0^t \mathrm{e}^{-\alpha(t-\tau)}\sqcap\left(\frac{\tau - T/2}{T}\right)\,\mathrm{d}\tau.$$

Sposób obliczenia występującej tu całki jest zilustrowany na rys. 7.3. Dla przesunięć $0 \leqslant t \leqslant T$ splatanie następuje w przedziale $[0,t]$ (rys. 7.3a), a więc:

$$\alpha\int_0^t \mathrm{e}^{-\alpha(t-\tau)}\sqcap\left(\frac{\tau - T/2}{T}\right)\,\mathrm{d}\tau = \alpha\,\mathrm{e}^{-\alpha t}\int_0^t \mathrm{e}^{\alpha\tau}\,\mathrm{d}\tau = 1 - \mathrm{e}^{-\alpha t}.$$

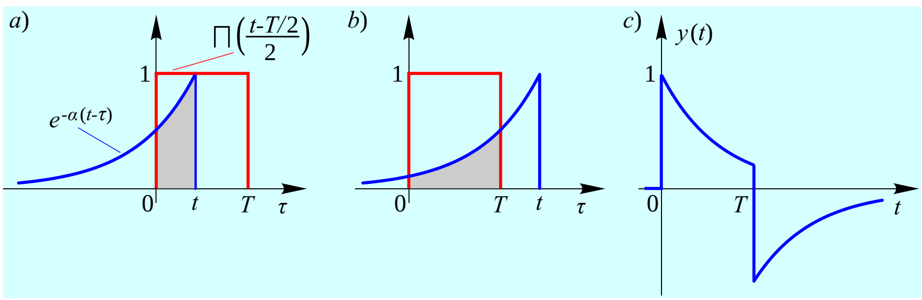
Dla $t > T$ sygnały splatają się w przedziale $[0,T]$ (rys. 7.3b) a więc:

$$\alpha\int_0^t \mathrm{e}^{-\alpha(t-\tau)}\sqcap\left(\frac{\tau - T/2}{T}\right)\,\mathrm{d}\tau = \alpha\,\mathrm{e}^{-\alpha t}\int_0^T \mathrm{e}^{\alpha\tau}\,\mathrm{d}\tau = \mathrm{e}^{-\alpha(t-T)} - \mathrm{e}^{-\alpha t}.$$

Łącząc otrzymane wyniki, otrzymujemy ostatecznie:

$$y(t) = \left\{\begin{array}{lll} \mathrm{e}^{-\alpha t} & \text{dla} & 0 \leqslant t \leqslant T \\ \mathrm{e}^{-\alpha t} - \mathrm{e}^{-\alpha(t-T)} & \text{dla} & t > T. \end{array}\right\} = \mathrm{e}^{-\alpha t}\,\boldsymbol{1}(t) - \mathrm{e}^{-\alpha(t-T)}\,\boldsymbol{1}(t-T).$$

Wykres sygnału wyjściowego $y(t)$ jest pokazany na rys. 7.3c. Zauważmy, że w chwili $t = T$ występuje skokowa zmiana sygnału o wartość -1.



Rys. 7.3. Ilustracja obliczania całki splotowej dla sygnałów z przykładu 7.9

Przykład 7.10. Wyznaczymy sygnał $y(t)$ na wyjściu układu o odpowiedzi impulsowej $h(t) = \alpha\,\mathrm{e}^{-\alpha t}\,\boldsymbol{1}(t)$, $\alpha > 0$, pobudzanego sygnałem sinusoidalnym $x(t) = \sin(\alpha t)$, $-\infty < t < +\infty$.

W tym przypadku jeden z sygnałów splatanych przybiera wartości niezerowe na całej osi czasu. Skorzystamy zatem z pierwszego ze wzorów (7.20):

$$\begin{aligned}
y(t) &= \int\limits_{-\infty}^{t} \alpha\,\mathrm{e}^{-\alpha(t-\tau)} \sin(\alpha\tau)\,\mathrm{d}\tau = \alpha\,\mathrm{e}^{-\alpha t} \int\limits_{-\infty}^{t} \frac{\mathrm{e}^{\mathrm{j}\alpha\tau} - \mathrm{e}^{-\mathrm{j}\alpha\tau}}{2\,\mathrm{j}}\,\mathrm{e}^{\alpha\tau}\,\mathrm{d}\tau = \\
&= \frac{\alpha\,\mathrm{e}^{-\alpha t}}{2\,\mathrm{j}} \int\limits_{-\infty}^{t} \left[\mathrm{e}^{(1+\mathrm{j})\alpha\tau} - \mathrm{e}^{(1-\mathrm{j})\alpha\tau}\right] \mathrm{d}\tau = \frac{\alpha\,\mathrm{e}^{-\alpha t}}{2\,\mathrm{j}} \left[\frac{\mathrm{e}^{(1+\mathrm{j})\alpha t}}{(1+\mathrm{j})\alpha} - \frac{\mathrm{e}^{(1-\mathrm{j})\alpha t}}{(1-\mathrm{j})\alpha}\right] = \\
&= \frac{1}{2\,\mathrm{j}} \left[\frac{(1-\mathrm{j})\,\mathrm{e}^{\mathrm{j}\alpha t} - (1+\mathrm{j})\,\mathrm{e}^{-\mathrm{j}\alpha t}}{2}\right] = \frac{1}{2}\left[\frac{\mathrm{e}^{\mathrm{j}\alpha t} - \mathrm{e}^{-\mathrm{j}\alpha t}}{2\,\mathrm{j}} - \frac{\mathrm{e}^{\mathrm{j}\alpha t} + \mathrm{e}^{-\mathrm{j}\alpha t}}{2}\right] = \\
&= \frac{1}{2}(\sin\alpha t - \cos\alpha t) = \frac{1}{\sqrt{2}} \sin\left(\alpha t - \frac{\pi}{4}\right).
\end{aligned}$$

W wyniku otrzymujemy także sygnał sinusoidalny o tej samej pulsacji α, amplitudzie $\sqrt{2}$ razy mniejszej i opóźniony o $\pi/4$.

7.2. Opis układu w dziedzinie zespolonej

7.2.1. Przekształcenie Laplace'a

Opis układu w dziedzinie zespolonej jest oparty na matematycznym formalizmie całkowego przekształcenia Laplace'a. Definicja i najważniejsze właściwości przekształcenia Laplace'a są zwykle omawiane w podstawowym kursie teorii obwodów (por. np. [5], p. 6.1.4). Przypomnimy tu jedynie definicję.

Definicja 7.10. Niech $f(t)$ będzie funkcją rzeczywistą lub zespoloną zmiennej rzeczywistej t, równą tożsamościowo zeru dla $t < 0$. *Przekształceniem Laplace'a* nazywamy odwzorowanie $\mathscr{L}$ określone całką:

$$F(s) = \mathscr{L}[f(t)] \triangleq \int\limits_{0}^{\infty} f(t)\,\mathrm{e}^{-st}\,\mathrm{d}t, \tag{7.22}$$

które funkcji $f(t)$ przyporządkowuje funkcję $F(s)$ zmiennej zespolonej $s = \alpha + \mathrm{j}\,\omega$.

Funkcje $f(t)$, $f(t) \equiv 0$ dla $t < 0$, dla których całka (7.22) istnieje nazywamy funkcjami *$\mathscr{L}$-transformowalnymi* lub *oryginałami* przekształcenia Laplace'a, a funkcję $F(s)$ – *transformatą Laplace'a*.

Załóżmy, że sygnał $x(t) \equiv 0$ dla $t < 0$, jest $\mathscr{L}$-transformowalny. Jeśli jest to sygnał ciągły, przekształcenie Laplace'a jest wzajemnie jednoznaczne, tzn. znając transformatę $X(s) = \mathscr{L}[x(t)]$ sygnału $x(t)$ możemy, stosując odwrotne przekształcenie Laplace'a $\mathscr{L}^{-1}[X(s)]$, jednoznacznie odtworzyć ten sygnał. Oznacza to, że dla ciągłych sygnałów $x(t)$ równość $x(t) = \mathscr{L}^{-1}[X(s)]$ zachodzi w każdym punkcie t. Inaczej mówiąc, transformata $X(s)$ zawiera w tym przypadku pełną informację o sygnale $x(t)$, z tym że wyrażoną w innym, bardziej abstrakcyjnym języku matematycznym.

Jeśli sygnał $x(t)$ ma punkty nieciągłości, równość $x(t) = \mathscr{L}^{-1}[X(s)]$ zachodzi prawie wszędzie, z dokładnością do wartości w punktach nieciągłości. Ponieważ w praktycznej analizie sygnałów nie jest z reguły istotne, jakie wartości przybierają one w punktach nieciągłości, będziemy zakładać, że sygnał $x(t)$ i jego transformata Laplace'a $X(s)$ są związane wzajemnie jednoznacznie. Parę funkcji $x(t)$ i $X(s) = \mathscr{L}[x(t)]$ będziemy nazywać *parą transformat Laplace'a* i oznaczać symbolicznie $x(t) \leftrightarrow X(s)$.

7.2.2. Transmitancja

Rozważmy przyczynowy układ SLS pobudzany sygnałem $x(t)$ równym tożsamościowo zeru dla $t < 0$ (rys. 7.4a). Przy tych założeniach sygnał $y(t)$ na wyjściu tego układu jest również równy tożsamościowo zeru dla $t < 0$. Oznaczmy transformaty Laplace'a tych sygnałów $X(s) = \mathscr{L}[x(t)]$ oraz $Y(s) = \mathscr{L}[y(t)]$.

Definicja 7.11. *Transmitancją* układu nazywamy iloraz transformaty Laplace'a sygnału na jego wyjściu do transformaty Laplace'a sygnału na jego wejściu:

$$H(s) \triangleq \frac{Y(s)}{X(s)}. \tag{7.23}$$

Z określenia transmitancji wynika równanie

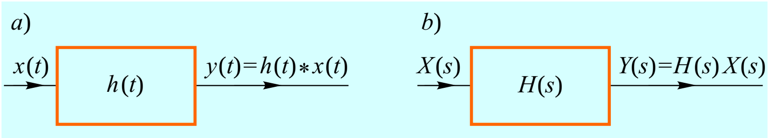
$$Y(s) = H(s)X(s) \tag{7.24}$$

opisujące w dziedzinie zespolonej zależność między sygnałem wejściowym a sygnałem wyjściowym (rys. 7.4b). Równanie to jest nazywane *równaniem transmisyjnym* układu. Transformata Laplace'a sygnału na wyjściu układu jest więc iloczynem transmitancji i transformaty Laplace'a sygnału na jego wejściu. Wynika stąd, że

$$y(t) = \mathscr{L}^{-1}[H(s)X(s)] \tag{7.25}$$

Załóżmy teraz, że układ jest pobudzany impulsem Diraca $\delta(t)$. Uwzględniając w równaniu (7.25), że $\mathscr{L}[\delta(t)] = 1$ oraz fakt, że na wyjściu układu wystąpi wówczas sygnał $h(t)$, otrzymujemy równość:

$$h(t) = \mathscr{L}^{-1}[H(s)], \tag{7.26}$$

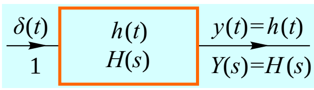


Rys. 7.4. Opis układu w dziedzinie czasu (a) i dziedzinie zespolonej (b)

z której wynika, że transmitancja układu jest transformatą Laplace'a jego odpowiedzi impulsowej (rys. 7.5):

$$H(s) = \mathscr{L}[h(t)]. \tag{7.27}$$

Odpowiedź impulsowa i transmitancja układu są więc ze sobą związane wzajemnie jednoznacznie całkowym przekształceniem Laplace'a. Znając odpowiedź impulsową, możemy jednoznacznie wyznaczyć transmitancję i odwrotnie. Transmitancję $H(s)$ możemy zatem traktować jako pewną charakterystykę układu, równoważną odpowiedzi impulsowej $h(t)$, opisującą ten układ w dziedzinie zespolonej. Zgodnie ze wzorem (7.25), znajomość transmitancji wystarcza, aby wyznaczyć sygnał $y(t)$ na wyjściu układu pobudzanego dowolnym sygnałem $\mathscr{L}$-transformowalnym.



Rys. 7.5. Zależność między transmitancją i odpowiedzią impulsową układu

Zauważmy, że równanie transmisyjne (7.24) jest konsekwencją zależności splotowej (7.17). Istotnie, dokonując przekształcenia Laplace'a obu stron równości (7.17) i uwzględniając, że operacji splatania sygnałów w dziedzinie czasu odpowiada mnożenie ich transformat Laplace'a w dziedzinie zespolonej, otrzymujemy:

$$Y(s) = \mathscr{L}[y(t)] = \mathscr{L}[h(t) * x(t)] = \mathscr{L}[h(t)]\mathscr{L}[x(t)] = H(s)X(s). \tag{7.28}$$

7.2.3. Wyznaczanie sygnału na wyjściu układu w dziedzinie zespolonej

Z przeprowadzonych rozważań wynika, że w przypadku klasy sygnałów $\mathscr{L}$-transformowalnych problem wyznaczania sygnału na wyjściu układu możemy przenieść do dziedziny zespolonej. Zgodnie ze wzorem (7.25) należy w tym celu obliczyć transformatę Laplace'a $X(s)$ sygnału wyjściowego $x(t)$, pomnożyć ją przez transmitancję $H(s)$, a następnie wyznaczyć odwrotną transformatę Laplace'a iloczynu. Ten sposób postępowania nosi nazwę *metody operatorowej*. Stosując metodę operatorową do wyznaczania sygnału wyjściowego unikamy skomplikowanych często obliczeń wymaganych przy bezpośrednim obliczeniu całki splotowej (7.21).

Przykład 7.11. Rozpatrzymy ponownie przykład 7.9 i wyznaczymy sygnał $y(t)$ na wyjściu rozpatrywanego w nim układu, tym razem przeprowadzając obliczenia w dziedzinie zespolonej. Korzystać będziemy przy tym z podstawowych znanych par transformat Laplace'a. Najpierw obliczamy transmitancję:

$$H(s) = \mathscr{L}[h(t)] = \mathscr{L}[\delta(t) - \alpha\, \mathrm{e}^{-\alpha t}\, \boldsymbol{1}(t)] = 1 - \frac{\alpha}{s+\alpha} = \frac{s}{s+\alpha},$$

a następnie transformatę sygnału wejściowego:

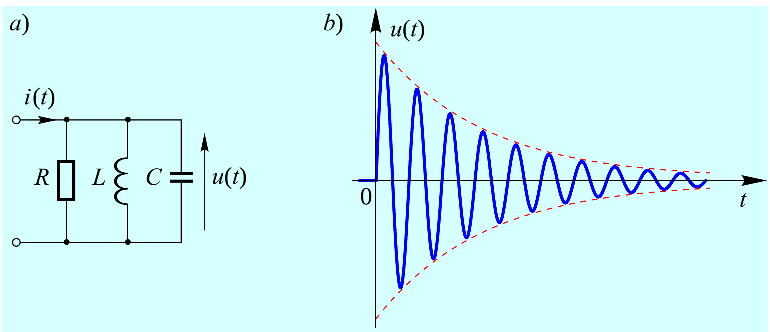
$$X(s) = \mathscr{L}[x(t)] = \mathscr{L}[\boldsymbol{1}(t) - \boldsymbol{1}(t-T)] = \frac{1}{s} - \frac{1}{s}\,\mathrm{e}^{-sT}.$$

Z kolei obliczamy transformatę sygnału wyjściowego:

$$Y(s) = H(s)X(s) = \frac{1}{s+\alpha} - \frac{1}{s+\alpha}\,\mathrm{e}^{-sT}.$$

Wyznaczając odwrotną transformatę $y(t) = \mathscr{L}^{-1}[Y(s)]$, otrzymujemy ten sam wynik, jak w przykładzie 7.9, ale znacznie mniejszym nakładem obliczeniowym.

Przykład 7.12. Wyznaczymy napięciową odpowiedź $u(t)$ równoległego obwodu rezonansowego RLC pokazanego na rys. 7.6a na pobudzenie prądowym skokiem jednostkowym $i(t) = I_0\boldsymbol{1}(t)$ przy założeniu, że dobroć obwodu $Q > 1/2$ $(R^2 > (L/C)/4)$.



Rys. 7.6. Równoległy obwód rezonansowy RLC (a) i jego odpowiedź jednostkowa (b)

Równanie tego obwodu w dziedzinie zespolonej, wiążące transformaty $U(s) = \mathscr{L}[u(t)]$ oraz $I(s) = \mathscr{L}[i(t)]$, ma postać $U(s) = Z(s)I(s)$, gdzie $Z(s)$

jest impedancją obwodu. Wynika stąd, że w tym przypadku transmitancja jest równa impedancji $Z(s)$:

$$H(s) = \frac{1}{\frac{1}{R} + sC + \frac{1}{sL}} = \frac{RLs}{RLCs^2 + Ls + R} = i\frac{1}{C}\frac{s}{(s+\alpha)^2 + \omega_s^2},$$

gdzie oznaczono:

$$\alpha = \frac{1}{2RC}, \qquad \omega_s^2 = \omega_0^2 - \alpha^2 > 0, \qquad \omega_0 = \frac{1}{LC}.$$

Transformata napięcia na zaciskach obwodu ma zatem postać:

$$U(s) = \frac{I_0}{C}\frac{1}{(s+\alpha)^2 + \omega_s^2}.$$

Korzystając z tablic transformat Laplace'a, otrzymujemy:

$$u(t) = \frac{I_0}{\omega_s C}\,\mathrm{e}^{-\alpha t}\sin\omega_s t`\boldsymbol{1}(t).$$

Przebieg napięcia $u(t)$ jest pokazany na rys. 7.6b.

7.2.4. Ogólna postać transmitancji

W obu rozpatrywanych wyżej przykładach transmitancje były *funkcjami wymiernymi rzeczywistymi* (*funkcjami wr*) zmiennej zespolonej s. Dowodzi się (por. np. [2], p. 9.2.1-B), że jest to ogólna właściwość przyczynowych układów SLS. Dla tej klasy układów transmitancja ma ogólną postać:

$$H(s) = \frac{L(s)}{M(s)} = \frac{b_l s^l + b_{l-1}s^{l-1} + \ldots + b_1 s + b_0}{c_m s^m + c_{m-1}s^{m-1} + \ldots + c_1 s + c_0}, \tag{7.29}$$

gdzie współczynniki b_i, $i = 1, \ldots, l$, oraz c_j, $j = 1, \ldots, m$, są liczbami rzeczywistymi, $l = \mathrm{st}L(s)$ jest stopniem licznika, a $m = \mathrm{st}M(s)$ – stopniem mianownika.

Jeśli $l - m > 0$, to dzieląc licznik transmitancji (7.29) przez mianownik, możemy przedstawić ją w postaci:

$$H(s) = a_n s^n + \ldots + a_1 s + a_0 + \frac{L_0(s)}{M(s)} = a_n s^n + \ldots + a_1 s + a_0 + H_0(s), \tag{7.30}$$

gdzie $H_0(s)$ jest *funkcją wymierną właściwą* (o stopniu licznika mniejszym niż stopień mianownika) oraz $n = l - m > 0$. Ze wzoru (7.30) wynika cytowana wcześniej ogólna postać odpowiedzi impulsowej przyczynowych układów SLS, określona wzorem (7.14). Zachodzi przy tym zależność $H_0(s) = \mathscr{L}[h_0(t)]$. Układy, dla których $l - m > 0$, nie mają odpowiedników fizycznie realizowalnych, a więc w praktyce interesujący jest jedynie przypadek $l - m \leqslant 0$. Transmitancja jest wówczas określona wzorem:

$$H(s) = a_0 + H_0(s), \tag{7.31}$$

co odpowiada postaci odpowiedzi impulsowej określonej wzorem (7.15). Zauważmy, że w obu rozpatrywanych przykładach 7.11 i 7.12 mieliśmy do czynienia z przypadkiem $l - m = 0$.

Wydzielając stały czynnik $H_0 = b_l / c_m$, transmitancję (7.29) możemy przedstawić w jeszcze innej postaci:

$$H(s) = H_0 \frac{(s - z_1)(s - z_2)\ldots(s - z_l)}{(s - s_1)(s - s_2)\ldots(s - s_m)}, \tag{7.32}$$

gdzie z_i są *zerami* transmitancji, a s_j – jej *biegunami*, tzn. miejscami zerowymi jej mianownika. Bieguny transmitancji są pierwiastkami równania:

$$c_m s^m + c_{m-1} s^{m-1} + \ldots + c_1 s + c_0 = 0 \tag{7.33}$$

nazywanego *równaniem charakterystycznym* układu. Ponieważ współczynniki obu wielomianów we wzorze (7.29) są rzeczywiste, wszystkie zera i bieguny transmitancji są albo rzeczywiste albo tworzą pary liczb zespolonych sprzężonych. Mogą być one przy tym jednokrotne lub wielokrotne.

7.2.5. Związki między położeniem zer i biegunów transmitancji a postacią odpowiedzi impulsowej

Ze wzoru (7.32) wynika, że z dokładnością do współczynnika skali H_0 transmitancja układu jest jednoznacznie określona przez rozkład jej zer i biegunów na płaszczyźnie zmiennej zespolonej s. Na podstawie znajomości położenia zer i biegunów na płaszczyźnie s oraz współczynnika H_0 można wyznaczyć jednoznacznie przebieg odpowiedzi impulsowej układu. W języku zer i biegunów można zatem wyrazić pełną informację o właściwościach transmisyjnych układu. Jest to wprawdzie język abstrakcyjny, ale posługiwanie się nim bardzo ułatwia analizę wielu istotnych właściwości formalnych układu, m.in. określenie typu układu, czy badanie jego stabilności (por. p. 6.2.6-B,C).

Aby ustalić związek między położeniem zer i biegunów a postacią odpowiedzi impulsowej układu, wystarczy przedstawić transmitancję (7.32) w postaci (7.31) i rozłożyć funkcję $H_0(s) = \mathscr{L}[h_0(t)]$ na ułamki proste. Przy założeniu, że funkcja $H_0(s)$ jest nieskracalna (jej licznik i mianownik nie mają wspólnych czynników), bieguny obu funkcji $H_0(s)$ i $H(s)$ są identyczne. Znając rozkład funkcji $H_0(s)$ na ułamki proste, można na podstawie tablic par transformat Laplace'a wyznaczyć poszczególne składniki funkcji $h_0(t)$.

Jeśli bieguny transmitancji są jednokrotne, w rozkładzie na ułamki proste funkcji $H_0(s)$ wystąpią składniki, których mianowniki są:

– dwumianami $s + \alpha_i$ (co odpowiada przypadkowi biegunów rzeczywistych $s_i = -\alpha_i$); odpowiedź impulsowa $h_0(t)$ będzie wówczas zawierać składowe o postaci $A_i\,\mathrm{e}^{-\alpha_i t}\,\boldsymbol{1}(t)$,
– trójmianami kwadratowymi $s^2 + p_j s + q_j$ (co odpowiada przypadkowi biegunów zespolonych sprzężonych $s_{j1,2} = -\alpha_j \pm \mathrm{j}\,\omega_j$); odpowiedź impulsowa $h_0(t)$ będzie wówczas zawierać składowe o postaci $\mathrm{e}^{-\alpha_j t}(B_j \cos\omega_j t + C_j \sin\omega_j t)\boldsymbol{1}(t)$, gdzie parametry α_j, ω_j są zależne tylko od współczynników p_j, q_j.

Jeśli bieguny są wielokrotne, wymienione składowe odpowiedzi impulsowej są dodatkowo mnożone przez odpowiednie wielomiany zmiennej t. Odpowiedź impulsowa $h(t)$ układu będzie zatem w ogólnym przypadku sumą impulsu Diraca $a_0\delta(t)$ oraz składowych zmiennych o podanych wyżej postaciach. Współczynniki poszczególnych składowych są określone przez współrzędne zer i biegunów. Szczegółowe omówienie metody wyznaczania odwrotnej transformaty Laplace'a funkcji wymiernej właściwej z wykorzystaniem rozkładu na ułamki proste można znaleźć w [5], p. 6.2.3-B.

7.3. Opis układu w dziedzinie częstotliwości

7.3.1. Charakterystyka amplitudowo-fazowa układu

Załóżmy, że odpowiedź impulsowa $h(t)$ układu SLS jest $\mathscr{F}$-transformowalna w sensie zwykłym lub w sensie granicznym.

Definicja 7.12. *Charakterystyką amplitudowo-fazową* układu nazywamy transformatę Fouriera jego odpowiedzi impulsowej:

$$H(\mathrm{j}\,\omega) = \int\limits_{-\infty}^{\infty} h(t)\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t. \tag{7.34}$$

Podana definicja określa pojęcie charakterystyki amplitudowo-fazowej zarówno w odniesieniu do układów przyczynowych, jak i nieprzyczynowych. W przypadku układów przyczynowych całkowanie we wzorze (7.34) przebiega w granicach $[0, +\infty)$. Z definicji tej wynika, że charakterystyka amplitudowo-fazowa jest widmem odpowiedzi impulsowej. Podkreślmy, że w przeciwieństwie do standardowej notacji widma sygnału, argument funkcji $H(\mathrm{j}\,\omega)$ jest zwyczajowo oznaczany $\mathrm{j}\,\omega$. Notację taka jest uzasadniona faktem, iż w przypadku układów przyczynowych między charakterystyką amplitudowo-fazową a transmitancją układu zachodzi zależność:

$$H(\mathrm{j}\,\omega) = H(s)|_{s=\mathrm{j}\omega}, \tag{7.35}$$

tzn. charakterystyka amplitudowo-fazowa $H(\mathrm{j}\,\omega)$ jest równa transmitancji $H(s)$ przy podstawieniu $s = \mathrm{j}\,\omega$.

Przykład 7.13. Wyznaczymy charakterystykę amplitudowo-fazową układu o odpowiedzi impulsowej $h(t) = \delta(t) - \alpha\,\mathrm{e}^{-\alpha t}\boldsymbol{1}(t)$, rozpatrywanego w przykładach 7.9 i 7.11. W przykładzie 7.11 obliczyliśmy transmitancję tego układu $H(s) = s/(s+\alpha)$. Podstawiając w tym wzorze $s = \mathrm{j}\,\omega$, otrzymujemy:

$$H(\mathrm{j}\,\omega) = \frac{\mathrm{j}\,\omega}{\alpha + \mathrm{j}\,\omega}. \tag{7.36}$$

Przykład 7.14. W analogiczny sposób wyznaczamy charakterystyki amplitudowo-fazowe układów rozpatrywanych w przykładzie 7.10:

$$H(\mathrm{j}\,\omega) = \left.\frac{\alpha}{s+\alpha}\right|_{s=\mathrm{j}\omega} = \frac{\alpha}{\alpha + \mathrm{j}\,\omega} \tag{7.37}$$

oraz w przykładzie 7.12.

$$H(\mathrm{j}\,\omega) = \frac{1}{C}\left.\frac{s}{(s+\alpha)^2 + \omega_s^2}\right|_{s=\mathrm{j}\omega} = \frac{1}{C}\frac{\mathrm{j}\,\omega}{(\alpha + \mathrm{j}\,\omega)^2 + \omega_s^2} = \frac{1}{C}\frac{\mathrm{j}\,\omega}{\omega_0^2 - \omega^2 + \mathrm{j}2\alpha\omega}. \tag{7.38}$$

7.3.2. Równanie transmisyjne w dziedzinie częstotliwości

Rozważmy układ pobudzany sygnałem $x(t)$ określonym w ogólnym przypadku w przedziale $t \in (-\infty, +\infty)$. Załóżmy, że na wyjściu układu występuje wówczas sygnał $y(t)$ oraz że oba sygnały $x(t)$ i $y(t)$ są $\mathscr{F}$-transformowalne. Oznaczmy ich widma $X(\omega) = \mathscr{F}[x(t)]$ oraz $Y(\omega) = \mathscr{F}[y(t)]$.

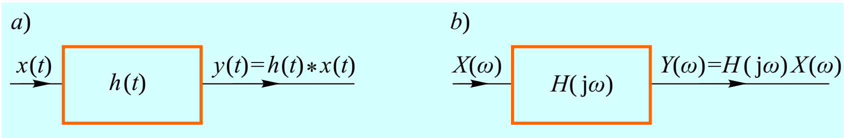
Zależność między sygnałami $x(t)$ i $y(t)$ w dziedzinie czasu jest określona splotem (7.18) (rys. 7.7a). Ponieważ, zgodnie z twierdzeniem 3.8, operacji splotu odpowiada mnożenie transformat Fouriera, zatem ze wzoru (7.18) wynika że

$$Y(\omega) = H(\mathrm{j}\,\omega)X(\omega). \tag{7.39}$$

Zależność (7.39) jest nazywana *równaniem transmisyjnym* układu w dziedzinie częstotliwości. Zgodnie z tą zależnością widmo sygnału na wyjściu układu jest iloczynem charakterystyki amplitudowo-fazowej i widma sygnału na jego wejściu (rys. 7.7b). Obliczając odwrotną transformatę Fouriera tego iloczynu, otrzymujemy:

$$y(t) = \mathscr{F}^{-1}[H(\mathrm{j}\,\omega)X(\omega)]. \tag{7.40}$$

Widzimy więc, że znając charakterystykę amplitudowo-fazową $H(\mathrm{j}\,\omega)$ układu, możemy obliczyć jego odpowiedź na dowolny $\mathscr{F}$-transformowalny sygnał wejściowy. Funkcję $H(\mathrm{j}\,\omega)$ możemy zatem przyjąć za podstawową charakterystykę układu równoważną odpowiedzi impulsowej i opisującą jego właściwości transmisyjne w dziedzinie częstotliwości.



Rys. 7.7. Opis układu w dziedzinie częstotliwości

7.3.3. Charakterystyka amplitudowa i charakterystyka fazowa

Charakterystykę amplitudowo-fazową można przedstawić w postaci biegunowej:

$$H(\mathrm{j}\,\omega) = |H(\mathrm{j}\,\omega)|\,\mathrm{e}^{\mathrm{j}\arg H(\mathrm{j}\omega)} \triangleq A(\omega)\,\mathrm{e}^{\mathrm{j}\varphi(\omega)} \tag{7.41}$$

lub w postaci algebraicznej:

$$H(\mathrm{j}\,\omega) = \mathrm{Re}\; H(\mathrm{j}\,\omega) + \mathrm{j}\,\mathrm{Im}\; H(\mathrm{j}\,\omega) \triangleq P(\omega) + \mathrm{j}\,Q(\omega), \tag{7.42}$$

gdzie $A(\omega)$, $\varphi(\omega)$, $P(\omega)$, $Q(\omega)$ są funkcjami rzeczywistymi zmiennej $\omega \in (-\infty, +\infty)$. Funkcje te są nazywane odpowiednio: *charakterystyką amplitudową*, *charakterystyką fazową*, *charakterystyką rzeczywistą* i *charakterystyką urojoną* układu. Wraz z charakterystyką amplitudowo-fazową tworzą one zbiór *charakterystyk częstotliwościowych* układu opisujących go w dziedzinie częstotliwosci.

Ponieważ charakterystyka amplitudowo-fazowa jest transformatą Fouriera sygnału rzeczywistego, więc jest funkcją hermitowską, tzn. spełnia równość (por. p. 2.1.1-D):

$$H(\mathrm{j}\,\omega) = H^*(-\mathrm{j}\,\omega). \tag{7.43}$$

Oznacza to, że charakterystyka amplitudowa i charakterystyka rzeczywista są funkcjami parzystymi:

$$A(-\omega) = A(\omega), \qquad P(-\omega) = P(\omega), \tag{7.44}$$

a charakterystyka fazowa i charakterystyka urojona – funkcjami nieparzystymi zmiennej ω:

$$\varphi(-\omega) = -\varphi(\omega), \qquad Q(-\omega) = -Q(\omega). \tag{7.45}$$

Z właściwości tych wynika, że charakterystyki częstotliwościowe wystarczy wyznaczać dla fizycznych wartości pulsacji $\omega \geqslant 0$, a następnie dokonać ich odpowiedniego przedłużenia dla $\omega < 0$.

W praktyce jesteśmy zainteresowani przede wszystkim wyznaczeniem charakterystyki amplitudowej i charakterystyki fazowej układu. Opisują one bowiem w sposób przejrzysty właściwości transmisyjne układu i mają wyraźną interpretację fizyczną (por. p. E). Charakterystyki rzeczywista i urojona odgrywają w analizie częstotliwościowej układów rolę pomocniczą.

Przykład 7.15. Ze wzoru (7.36) wynika, że charakterystyka amplitudowa i charakterystyka fazowa układu rozpatrywanego w przykładzie 6.13 mają postacie:

$$A(\omega) = \frac{|\omega|}{\sqrt{\alpha^2+\omega^2}}, \qquad \varphi(\omega) = \frac{\pi}{2}\operatorname{sgn}\omega - \operatorname{arctg}\frac{\omega}{\alpha}. \tag{7.46}$$

Wykresy tych charakterystyk są przedstawione na rys. 7.8a, b. Z wykresu charakterystyki amplitudowej wynika, że układ ma w tym przypadku cechy filtru górnoprzepustowego.

Przykład 7.16. Ze wzoru (7.37) wyznaczamy wprost charakterystykę amplitudową i charakterystykę fazową układu rozpatrywanego w przykładzie 7.10:

$$A(\omega) = \frac{\alpha}{\sqrt{\alpha^2+\omega^2}}, \qquad \varphi(\omega) = -\operatorname{arctg}\frac{\omega}{\alpha}. \tag{7.47}$$

Charakterystyki te są wykreślone na rys. 7.8c, d. Jak widzimy, układ ma w tym przypadku charakter filtru dolnoprzepustowego.

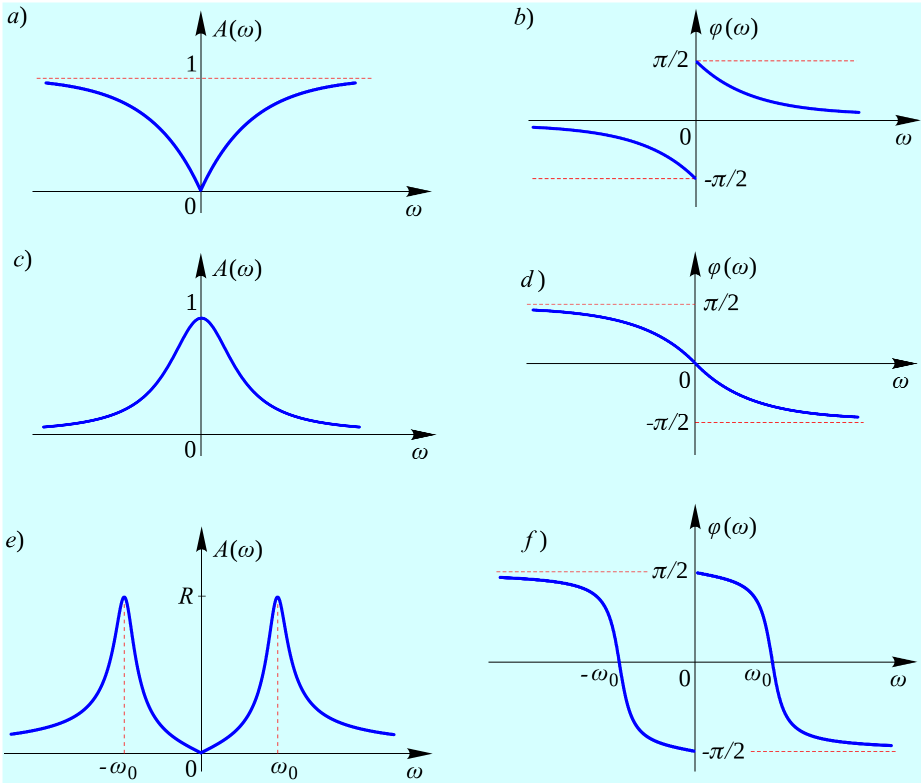
Przykład 7.17. Zgodnie ze wzorem (7.38) charakterystyka amplitudowa układu rozpatrywanego w przykładzie 7.12 ma postać:

$$A(\omega) = \frac{1}{C}\frac{|\omega|}{\sqrt{\left(\omega_0^2-\omega^2\right)^2+4\alpha^2\omega^2}}. \tag{7.48}$$

Nieco więcej biegłości wymaga w tym przypadku wyznaczenie charakterystyki fazowej. Przeprowadzając analizę osobno dla $0 \leqslant \omega \leqslant \omega_0$ oraz dla $\omega \geqslant \omega_0$, otrzymujemy:

$$\varphi(\omega) = \begin{cases} \dfrac{\pi}{2}\operatorname{sgn}\omega - \operatorname{arctg}\dfrac{2\alpha\omega}{\omega_0^2-\omega^2} & \text{dla} \quad |\omega| \leqslant \omega_0, \\[2ex] -\dfrac{\pi}{2}\operatorname{sgn}\omega + \operatorname{arctg}\dfrac{2\alpha\omega}{\omega^2-\omega_0^2} & \text{dla} \quad |\omega| > \omega_0. \end{cases} \tag{7.49}$$

Charakterystyki te są wykreślone na rys. 7.8e, f. Układ ma w tym przypadku charakter filtru środkowoprzepustowego.

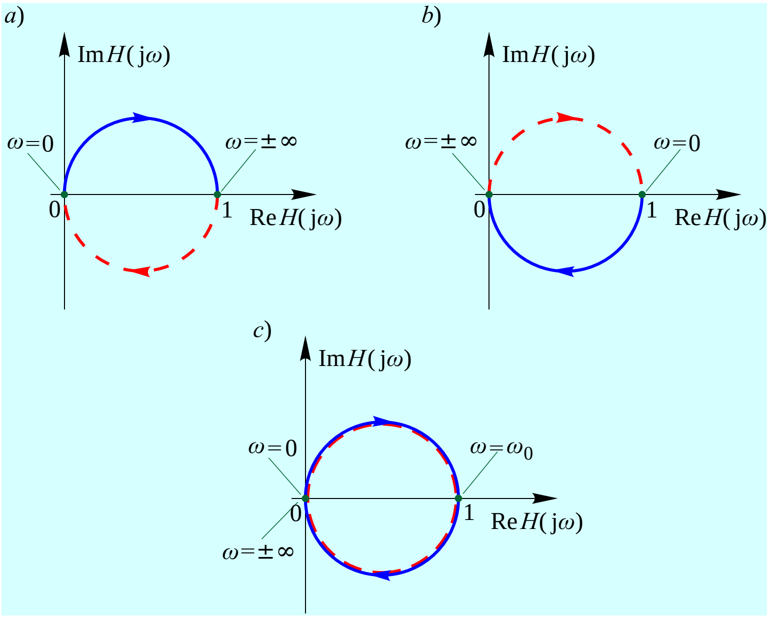


Rys. 7.8. Charakterystyki amplitudowe i charakterystyki fazowe prostych układów

7.3.4. Krzywa Nyquista

Obok wykresów charakterystyki amplitudowej i charakterystyki fazowej, w analizie częstotliwościowej układów posługujemy się często wykresem charakterystyki amplitudowo-fazowej $H(\mathrm{j}\,\omega)$ sporządzonym we współrzędnych $\mathrm{Re}\ H(\mathrm{j}\,\omega), \mathrm{Im}\ H(\mathrm{j}\,\omega)$ płaszczyzny zespolonej. Wykres taki, nazywany *krzywą Nyquista*, jest wykresem parametrycznym, którego parametrem jest pulsacja ω. Zasada sporządzania takiego wykresu polega na tym, że ustalonej wartości pulsacji $\omega = \omega'$ przyporządkowuje się punkt na płaszczyźnie $\mathrm{Re}\ H(\mathrm{j}\,\omega), \mathrm{Im}\ H(\mathrm{j}\,\omega)$ wyznaczony przez koniec wskazu $H(\mathrm{j}\,\omega')$ o długości $A(\omega')$ i kącie $\varphi(\omega')$. Krzywa Nyquista jest zatem miejscem geometrycznym tych wskazów określonym przy zmianach pulsacji w przedziale $(-\infty, +\infty)$.

Na rys. 7.9 są przedstawione krzywe Nyquista układów rozpatrywanych w przykładach: 7.15 (rys. 7.9a), 7.16 (rys. 7.9b) i 7.17 (rys. 7.9c). Strzałką jest zaznaczony kierunek poruszania się wskazu $H(\mathrm{j}\,\omega)$ w miarę jak pulsacja wzrasta od $-\infty$ do $+\infty$. Części krzywych odpowiadające zakresowi pulsacji $-\infty < \omega < 0$ są zaznaczone linią przerywaną. Zauważmy, że są one zwierciadlanymi odbiciami względem osi odciętych $\mathrm{Re}\, H(\mathrm{j}\,\omega)$ części krzywych wyznaczonych dla zakresu pulsacji $0 \leqslant \omega < +\infty$.



Rys. 7.9. Krzywe Nyquista prostych układów

7.3.5. Interpretacja charakterystyki amplitudowo-fazowej

Rozpatrzmy przypadek, gdy układ jest pobudzany na wejściu zespolonym sygnałem harmonicznym $x(t) = X_0\, \mathrm{e}^{\mathrm{j}(\omega_0 t + \varphi_0)}$, $t \in (-\infty, +\infty)$, o ustalonej pulsacji ω_0, amplitudzie X_0 i fazie początkowej φ_0. Korzystając z zależności splotowej (7.18), wyznaczymy sygnał na wyjściu układu:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau) X_0\, \mathrm{e}^{\mathrm{j}[\omega_0(t-\tau)+\varphi_0]}\, \mathrm{d}\tau = \\ &= \left[\int_{-\infty}^{\infty} h(\tau)\, \mathrm{e}^{-\mathrm{j}\omega_0\tau}\, \mathrm{d}\tau \right] X_0\, \mathrm{e}^{\mathrm{j}(\omega_0 t + \varphi_0)} \,. \end{aligned} \tag{7.50}$$

Zauważmy, że wyrażenie w nawiasie kwadratowym jest wartością charakterystyki amplitudowo-fazowej układu dla pulsacji $\omega = \omega_0$. Zatem:

$$y(t) = H(\mathrm{j}\,\omega_0) X_0\, \mathrm{e}^{\mathrm{j}(\omega_0 t + \varphi_0)} = A(\omega_0) X_0\, \mathrm{e}^{\mathrm{j}[\omega_0 t + \varphi_0 + \varphi(\omega_0)]} \,. \tag{7.51}$$

Wynika stąd, że sygnał wyjściowy otrzymujemy w wyniku przemnożenia sygnału wejściowego przez zespolony współczynnik $H(\mathrm{j}\,\omega_0) = A(\omega_0)\,\mathrm{e}^{\mathrm{j}\varphi(\omega_0)}$ równy wartości charakterystyki amplitudowo-fazowej układu w punkcie ω_0. Współczynnik ten jest nazywany *współczynnikiem przenoszenia* układu określonym dla pulsacji ω_0. Wnioskujemy zatem, że sygnał na wyjściu układu pobudzanego zespolonym sygnałem harmonicznym jest również zespolonym sygnałem harmonicznym o tej samej pulsacji i amplitudzie równej amplitudzie sygnału wejściowego pomnożonej przez wartość charakterystyki amplitudowej układu $A(\omega_0)$ w punkcie ω_0 oraz fazie równej fazie sygnału wejściowego przesuniętej o wartość charakterystyki fazowej układu $\varphi(\omega_0)$ w punkcie ω_0.

Komentarz. Spójrzmy na uzyskany rezultat z nieco innego punktu widzenia. Jeśli T jest operatorem układu LS i zachodzi równość:

$$y(t) = T[x(t)] = \lambda x(t), \tag{7.52}$$

tzn. układ powtarza na wyjściu sygnał wejściowy z pewnym współczynnikiem λ, to sygnał $x(t)$ nazywamy *funkcją własną* operatora T, a współczynnik λ – w ogólnym przypadku zespolony – jego *wartością własną*. Z przeprowadzonego rozumowania wynika zatem, że zespolony sygnał harmoniczny $X_0\,\mathrm{e}^{\mathrm{j}(\omega_0 t+\varphi_0)}$ jest funkcją własną operatora LS, a wartość charakterystyki amplitudowo-fazowej $H(\mathrm{j}\,\omega_0)$ – wartością własną tego operatora.

7.4. Zastosowanie charakterystyk układu do analizy sygnału wyjściowego

7.4.1. Odpowiedź układu dla różnych klas pobudzeń

W celu wyznaczenia odpowiedzi układu na pobudzenie wejściowe można wykorzystać jeden z omówionych wyżej trzech rodzajów jego opisu. Obliczenia można przeprowadzić w dziedzinie czasu, w dziedzinie zespolonej lub w dziedzinie częstotliwości, posługując się przy tym albo odpowiedzią impulsową $h(t)$, albo transmitancją $H(s)$ albo charakterystyką amplitudowo-fazową $H(\mathrm{j}\,\omega)$ układu. Zarówno wybór charakterystyki, jak i sposób prowadzenia obliczeń zależy od tego, do jakiej klasy sygnałów należy sygnał pobudzający.

Najbardziej uniwersalną charakterystyką, którą możemy posługiwać się dla najszerszej klasy pobudzeń jest odpowiedź impulsowa. Znając odpowiedź impulsową układu, możemy wyznaczyć sygnał na jego wyjściu na podstawie zależności splotowej (7.17) lub (7.18). Klasa pobudzeń, dla której możemy tak postąpić jest ograniczona jedynie warunkiem zbieżności całki splotowej. Jak wynika z przykładów omówionych w p. 7.1.16, obliczenia w dziedzinie czasu, nawet dla stosunkowo prostych układów i sygnałów pobudzających, są jednak żmudne, a w przypadku sygnałów bardziej złożonych napotykają poważne trudności natury analitycznej.

Znaczne uproszczenie obliczeń uzyskuje się, jeśli korzysta się z transmitancji układu i przeprowadza się je w dziedzinie zespolonej. O zaletach obliczeniowych tej metody świadczą przykłady rozpatrzone w p. 7.2.3. Przy sprawnym posługiwaniu się tablicami par transformat Laplace'a jest ona bardzo efektywna. Może być jednak zastosowana jedynie dla $\mathscr{L}$-transformowalnych sygnałów pobudzających określonych na dodatniej półosi czasu $0 \leqslant t < +\infty$.

Jeśli sygnał pobudzający należy do klasy sygnałów $\mathscr{F}$-transformowalnych w sensie zwykłym lub granicznym i przybierających wartości nieujemne na całej osi czasu $-\infty \leqslant t < +\infty$, sygnał wyjściowy można wyznaczać w dziedzinie częstotliwości na podstawie charakterystyki amplitudowo-fazowej układu. Obliczenia ułatwia w tym przypadku umiejętne korzystanie z twierdzeń dotyczących całkowego przekształcenia Fouriera i słownika par transformat Fouriera.

Szczególnym przypadkiem, wymagającym osobnego omówienia, jest pobudzanie układu w stanie ustalonym sygnałem harmonicznym lub szerzej – sygnałem okresowym. Jak wynika ze wzorów Eulera (3.25) i dyskusji przeprowadzonej w p. 7.3.5, odpowiedź układu LS pobudzanego sygnałem harmonicznym

$$x(t) = X_0 \cos(\omega_0 t + \varphi_{0x})$$

jest również sygnałem harmonicznym

$$y(t) = Y_0 \cos(\omega_0 t + \varphi_{0y})$$

o tej pulsacji ω_0 oraz amplitudzie Y_0 i fazie początkowej φ_{0y} związanej z amplitudą X_0 i fazą początkową φ_{0x} sygnału wejściowego zależnościami:

$$Y_0 = A(\omega_0) X_0 \tag{7.53}$$

$$\varphi_{0y} = \varphi_{0x} + \varphi(\omega_0). \tag{7.54}$$

Do wyznaczenia sygnału na wyjściu filtru wystarcza zatem znajomość wartości charakterystyki amplitudowo-fazowej (współczynnika przenoszenia) układu $H(\mathrm{j}\,\omega_0) = A(\omega_0)\,\mathrm{e}^{\mathrm{j}\varphi(\omega_0)}$ w punkcie ω_0.

Jeśli układ jest pobudzany sygnałem okresowym o okresie T_0 reprezentowanym zespolonym szeregiem Fouriera

$$x(t) = \sum_{k=-\infty}^{\infty} X_k\,\mathrm{e}^{\mathrm{j}k\omega_0 t}, \qquad \omega_0 = \frac{2\pi}{T_0},$$

to z liniowości układu oraz z równości (7.51) wynika, że sygnał na wyjściu układu jest również sygnałem okresowym o tym samym okresie T_0, reprezentowanym zespolonym szeregiem Fouriera

$$y(t) = \sum_{k=-\infty}^{\infty} Y_k \, \mathrm{e}^{\mathrm{j}k\omega_0 t}, \qquad \omega_0 = \frac{2\pi}{T_0},$$

którego współczynniki Y_k są związane ze współczynnikami X_k szeregu Fouriera sygnału pobudzającego zależnością:

$$Y_k = H(\mathrm{j}\, k\omega_0) X_k \tag{7.55}$$

lub równoważnie

$$|Y_k| = A(k\omega_0)|X_k|, \qquad \arg Y_k = \arg X_k + \varphi(k\omega_0). \tag{7.56}$$

Przykład 7.18. Wyznaczymy ponownie sygnał $y(t)$ na wyjściu układu o odpowiedzi impulsowej $h(t) = \alpha \, \mathrm{e}^{\alpha t} \boldsymbol{1}(t)$ $\alpha > 0$ pobudzanego sygnałem sinusoidalnym $x(t) = \sin(\alpha t)$ $\infty t + \infty$.

Problem ten rozwiązaliśmy już wcześniej w przykładzie 7.10. Obliczenia przeprowadziliśmy wówczas w dziedzinie czasu, korzystając z zależności splotowej (7.20). Tym razem wykorzystamy opis częstotliwościowy układu za pomocą charakterystyki amplitudowo-fazowej. Zgodnie ze wzorem (7.37) dla pulsacji sygnału pobudzającego $\omega = \alpha$ mamy:

$$H(\mathrm{j}\, \alpha) = \frac{\alpha}{\alpha + \mathrm{j}\, \alpha} = \frac{1}{\sqrt{2}} \, \mathrm{e}^{-\mathrm{j}\pi/4} \, .$$

Z zależności (7.53) i (7.54) wynika zatem, że

$$y(t) = \frac{1}{\sqrt{2}} \sin\left(\alpha t - \frac{\pi}{4}\right).$$

Otrzymaliśmy zatem ten sam rezultat co przykładzie 7.10, bez konieczności żmudnego obliczania całki splotowej.

7.4.2. Funkcje korelacyjne i widma energetyczne sygnału na wyjściu układu

W wielu zagadnieniach praktycznych jesteśmy zainteresowani jak zmienia się funkcja autokorelacji i widmo energii (lub mocy) sygnału po jego przetworzeniu przez układ. Rozpatrzymy ten problem najpierw dla przypadku sygnałów o ograniczonej energii.

Z równania transmisyjnego układu w dziedzinie częstotliwości (7.39) wynika wprost, że:

$$|Y(\omega)|^2 = |H(\mathrm{j}\, \omega)|^2 |X(\omega)|^2 \tag{7.57}$$

czyli

$$\Phi_y(\omega) = |H(\mathrm{j}\,\omega)|^2\Phi_x(\omega) = A^2(\omega)\Phi_x(\omega). \tag{7.58}$$

Widmo energii sygnału na wyjściu jest więc iloczynem widma energii sygnału na wejściu i kwadratu charakterystyki amplitudowej układu (tj. jego charakterystyki energetycznej).

W celu wyznaczenia funkcji autokorelacji sygnału wyjściowego zapiszmy równość (7.58) w postaci:

$$\Phi_y(\omega) = H(\mathrm{j}\,\omega)H^*(\mathrm{j}\,\omega)\Phi_x(\omega) \tag{7.59}$$

i wyznaczmy odwrotną transformatę Fouriera jej obu stron. Zgodnie z twierdzeniem o splocie w dziedzinie czasu (3.30)

$$\mathscr{F}^{-1}[\Phi_y(\omega)] = \mathscr{F}^{-1}[H(\mathrm{j}\,\omega)] * \mathscr{F}^{-1}[H^*(\mathrm{j}\,\omega)] * \mathscr{F}^{-1}[\Phi_x(\omega)]. \tag{7.60}$$

Biorąc pod uwagę zależność (5.9) oraz właściwość (3.18) przekształcenia Fouriera, wnioskujemy stąd, iż

$$\varphi_y(\tau) = h(\tau) * h(-\tau) * \varphi_x(\tau). \tag{7.61}$$

Tak więc funkcja autokorelacji sygnału na wyjściu układu jest równa podwójnemu splotowi funkcji autokorelacji sygnału wejściowego z jego odpowiedzią impulsową i odpowiedzią impulsową odbitą zwierciadlanie względem osi rzędnych.

W podobny sposób można wykazać słuszność następujących związków między funkcjami korelacji wzajemnej i widmami wzajemnymi energii sygnałów na wejściu i wyjściu układu:

$$\varphi_{xy}(\tau) = h(-\tau) * \varphi_x(\tau), \qquad \varphi_{yx}(\tau) = h(\tau) * \varphi_x(\tau), \tag{7.62}$$

$$\varphi_y(\tau) = h(\tau) * \varphi_{xy}(\tau) = h(-\tau) * \varphi_{yx}(\tau), \tag{7.63}$$

$$\Phi_{xy}(\omega) = H^*(\mathrm{j}\,\omega)\Phi_x(\omega), \qquad \Phi_{yx}(\omega) = H(\mathrm{j}\,\omega)\Phi_x(\omega), \tag{7.64}$$

$$\Phi_y(\omega) = H(\mathrm{j}\,\omega)\Phi_{xy}(\omega) = H^*(\mathrm{j}\,\omega)\Phi_{yx}(\omega). \tag{7.65}$$

Korzystając z definicji (5.3 i 5.4) oraz dokonując odpowiednich przejść granicznych, można wykazać, że analogiczne związki zachodzą między funkcjami korelacyjnymi i widmami mocy sygnałów o ograniczonej mocy. Wystarczy we wzorach (7.58) oraz (7.61)–(7.65) zmienić symbole $\varphi \to \psi$ oraz $\Phi \to \Psi$.

7.5. Wybrane układy

7.5.1. Układ opóźniający

Układem opóźniającym (rys. 7.10a) nazywamy układ o odpowiedzi impulsowej:

$$h(t) = \delta(t - t_0), \qquad t_0 > 0. \tag{7.66}$$

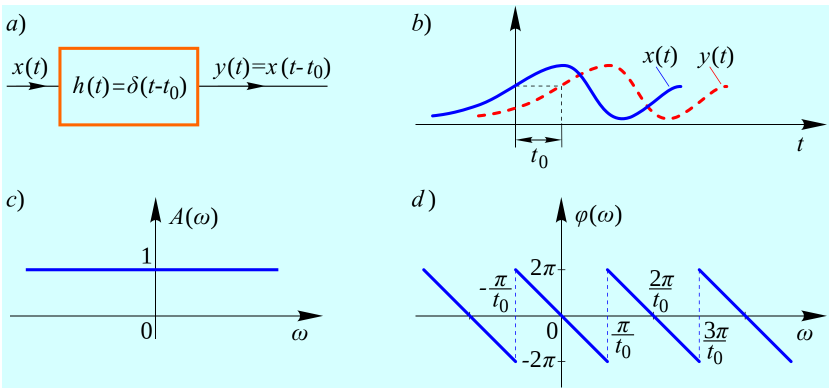
Sygnał na wyjściu tak zdefiniowanego układu jest opóźnioną o t_0 kopią sygnału wejściowego (rys. 7.10b). Istotnie, z właściwości (1.19) impulsu Diraca wynika, że

$$y(t) = h(t) * x(t) = \delta(t - t_0) * x(t) = x(t - t_0). \tag{7.67}$$

Ponieważ $\mathscr{L}[\delta(t)] = 1$, zatem z twierdzenia o przesunięciu w dziedzinie czasu dla przekształcenia Laplace'a wynika, że transmitancja układu opóźniającego jest określona wzorem:

$$H(s) = \mathrm{e}^{-st_0}, \qquad t_0 > 0. \tag{7.68}$$

Jak widzimy, nie jest to funkcja wymierna. Układ opóźniający nie jest zatem układem skupionym. Można jednak wykazać, że jest to układ przyczynowy, liniowy i stacjonarny. W praktyce układ opóźniający jest realizowany jako układ o stałych rozłożonych w postaci różnego rodzaju linii opóźniających.



Rys. 7.10. Układ opóźniający (a), sygnały na wejściu i wyjściu (b), charakterystyka amplitudowa (c) i charakterystyka fazowa modulo 2π (d)

Charakterystyka amplitudowo-fazowa układu opóźniającego ma postać $H(\mathrm{j}\,\omega) = \mathrm{e}^{-\mathrm{j}\omega t_0}$. Zatem $A(\omega) \equiv 1$ oraz $\varphi(\omega) = -\omega t_0$. Charakterystyka amplitudowa układu opóźniającego jest więc stałą funkcją pulsacji (rys. 7.10c), a charakterystyka fazowa maleje liniowo z nachyleniem t_0 (rys. 7.10d).

7.5.2. Idealny układ różniczkujący

Idealnym układem różniczkującym nazywamy układ o odpowiedzi impulsowej:

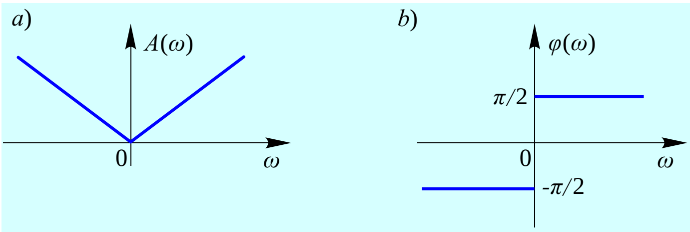
$$h(t) = \delta'(t), \tag{7.69}$$

transmitancji

$$H(s) = s \tag{7.70}$$

i charakterystyce amplitudowo-fazowej (rys. 7.11)

$$H(\mathrm{j}\,\omega) = \mathrm{j}\,\omega = |\omega|\,\mathrm{e}^{\mathrm{j}\frac{\pi}{2}\operatorname{sgn}\omega}\,. \tag{7.71}$$



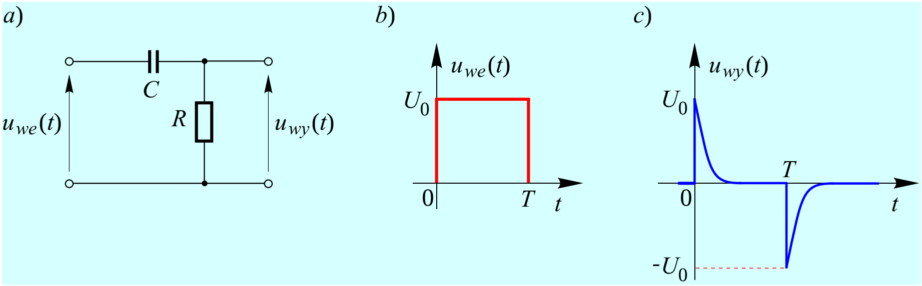
Rys. 7.11. Charakterystyka amplitudowa (a) i charakterystyka fazowa (b) idealnego układu różniczkującego

Funkcję idealnego układu różniczkującego spełnia np. indukcyjność L. Element ten jest opisany w dziedzinie czasu równaniem $u(t) = L\,\mathrm{d}i\,(t)/\,\mathrm{d}t$, a więc sygnał napięcia na indukcyjności jest proporcjonalny do pochodnej pobudzającego ją sygnału prądu. Transmitancja prądowo-napięciowa indukcyjności ma postać $H(s) = u(s)/i(s) = sL$.

W praktyce jako układ realizujący w przybliżeniu operację różniczkowania jest często stosowany czwórnik RC (rys. 7.12a). Rozpatrzmy dokładniej warunki jakie powinny być spełnione, aby czwórnik ten jak najlepiej pełnił funkcję układu różniczkującego. Charakterystyka amplitudowo-fazowa czwórnika ma postać (por. przykład 7.13):

$$H(\mathrm{j}\,\omega) = \frac{U_{wy}(\omega)}{U_{we}(\omega)} = \frac{\mathrm{j}\,\omega\tau_{RC}}{1 + \mathrm{j}\,\omega\tau_{RC}}, \tag{7.72}$$

gdzie $\tau_{RC} = RC$ jest stałą czasową. Charakterystyka ta będzie przybliżać charakterystykę idealnego układu różniczkującego (7.71) tym lepiej, im mniejszy od jedności będzie iloczyn $\omega\tau_{RC}$. Spełnienie równości $\omega\tau_{RC} \ll 1$ zależy jednak od zakresu zmiennej ω, a więc od struktury widmowej sygnału wejściowego. Jeśli różniczkowanym sygnałem wejściowym jest np. impuls prostokątny o czasie trwania T (rys. 7.12b), to jego widmo jest skoncentrowane w głównym listku widmowym, a więc w przedziale $|\omega| \leqslant 2\pi/T$ (por. przykład 3.3). Wynika stąd, że rozpatrywany czwórnik będzie dostatecznie dokładnie różniczkował impuls prostokątny, gdy jego stała czasowa spełnia nierówność $\tau_{RC} \ll T/2\pi$ (rys. 7.12c). Ogólnie można powiedzieć, że czwórnik ten będzie przybliżać idealny układ różniczkujący tym lepiej, im mniejsza jest jego stała czasowa.



Rys. 7.12. Układ różniczkujący RC (a) oraz sygnały na jego wejściu (b) i wyjściu (c)

7.5.3. Idealny układ całkujący

Idealnym układem całkującym nazywamy układ o odpowiedzi impulsowej:

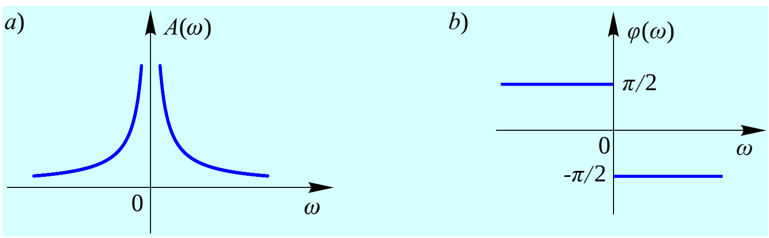
$$h(t) = \boldsymbol{1}(t), \tag{7.73}$$

transmitancji

$$H(s) = \frac{1}{s} \tag{7.74}$$

i charakterystyce amplitudowo-fazowej (rys. 7.13)

$$H(\mathrm{j}\,\omega) = \frac{1}{\mathrm{j}\,\omega} = \frac{1}{|\omega|}\,\mathrm{e}^{-\mathrm{j}\frac{\pi}{2}\,\mathrm{sgn}\,\omega}\,. \tag{7.75}$$



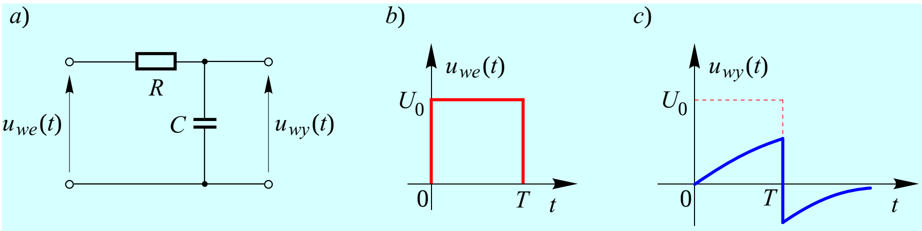
Rys. 7.13. Charakterystyka amplitudowa (a) i charakterystyka fazowa (b) idealnego układu całkującego

Jako idealny układ całkujący można wykorzystać np. pojemność C. Element ten jest opisany w dziedzinie czasu równaniem $i(t) = (1/C)\int_{-\infty}^{t} u(t)\,\mathrm{d}t$, a więc sygnał prądu na pojemności jest proporcjonalny do całki pobudzającego ją sygnału napięcia. Transmitancja napięciowo-prądowa pojemności ma postać $H(s) = u(s)/i(s) = 1/sC$.

Operację całkowania realizuje w przybliżeniu czwórnik RC pokazany na rys. 7.14a. W porównaniu z czwórnikiem różniczkującym RC różni się on innym usytuowaniem elementów R i C. Charakterystyka amplitudowo-fazowa tego czwórnika ma postać (por. przykład 7.14):

$$H(\mathrm{j}\,\omega) = \frac{U_{\mathrm{wy}}(\omega)}{U_{\mathrm{we}}(\omega)} = \frac{1}{1 + \mathrm{j}\,\omega\tau_{RC}}. \tag{7.76}$$

Aby przybliżała ona w zadowalającym stopniu charakterystykę amplitudowo-fazową (7.75) idealnego układu całkującego, musi zachodzić nierówność $\omega\tau_{RC} \gg 1$. Czwórnik całkujący RC będzie zatem przybliżać idealny układ całkujący tym lepiej, im większa jest jego stała czasowa.



Rys. 7.14. Układ całkujący RC (a) oraz sygnały na jego wejściu (b) i wyjściu (c)

7.5.4. Filtry idealne

Układy o charakterystykach amplitudowo-fazowych:

$$H(\mathrm{j}\,\omega) = \sqcap\left(\frac{\omega}{2\omega_g}\right), \tag{7.77}$$

$$H(\mathrm{j}\,\omega) = 1 - \sqcap\left(\frac{\omega}{2\omega_d}\right), \tag{7.78}$$

$$H(\mathrm{j}\,\omega) = \sqcap\left(\frac{|\omega| - \Omega}{2\omega_0}\right) = \begin{cases} 1 & \text{dla} \quad \Omega - \omega_0 \leqslant |\omega| \leqslant \Omega + \omega_0, \\ 0 & \text{przeciwnie}, \end{cases} \tag{7.79}$$

$$H(\mathrm{j}\,\omega) = 1 - \sqcap\left(\frac{|\omega| - \Omega}{2\omega_0}\right) = \begin{cases} 0 & \text{dla} \quad \Omega - \omega_0 \leqslant |\omega| \leqslant \Omega + \omega_0, \\ 1 & \text{przeciwnie}, \end{cases} \tag{7.80}$$

nazywamy: *idealnym filtrem dolnoprzepustowym* (por. p. 6.2.6), *idealnym filtrem górnoprzepustowym*, *idealnym filtrem środkowoprzepustowym* oraz *idealnym filtrem środkowozaporowym*. Filtry te są także nazywane odpowiednio: idealnym *filtrem LP* (od ang. *low-pass*), idealnym *filtrem HP* (od ang. *high-pass*), idealnym *filtrem BP* (od ang. *bandpass*) i idealnym *filtrem SB* (od ang. *stop-band*). Filtr środkowozaporowy jest również nazywany *filtrem wycinającym* lub *filtrem typu notch* (od ang. *notch filter*).

Wykresy charakterystyk (7.77)–(7.80) filtrów idealnych przedstawiono na rys. 7.15. Pulsacje ω_g i ω_d oznaczają odpowiednio: górną pulsację graniczną pasma przepustowego filtru LP oraz dolną pulsację graniczną pasma przepustowego filtru HP. Pulsacja Ω jest pulsacją środkową pasma przepustowego (zaporowego) filtru BP (filtru SB), a $2\omega_0$ – szerokością pasma przepustowego (zaporowego) tych filtrów.

Odpowiedzi impulsowe filtrów idealnych:

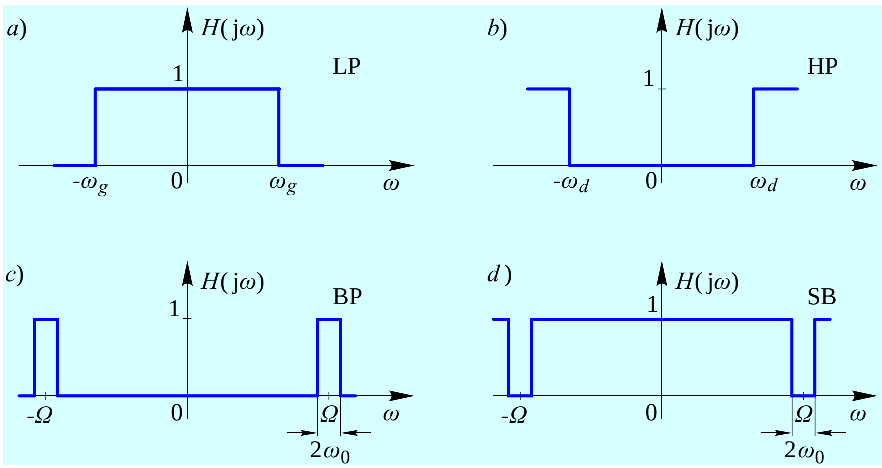
$$h(t) = \frac{\omega_g}{\pi}\,\mathrm{Sa}\,\omega_g t, \tag{7.81}$$

$$h(t) = \delta(t) - \frac{\omega_d}{\pi}\,\mathrm{Sa}\,\omega_d t, \tag{7.82}$$

$$h(t) = \frac{2\omega_0}{\pi}\,\mathrm{Sa}(\omega_0 t)\cos\Omega t, \tag{7.83}$$

$$h(t) = \delta(t) - \frac{2\omega_0}{\pi}\,\mathrm{Sa}(\omega_0 t)\cos\Omega t \tag{7.84}$$

przybierają wartości różne od zera dla $t < 0$. W przeciwieństwie do idealnych układów różniczkującego i całkującego są to zatem filtry nieprzyczynowe, nie spełniające podstawowego warunku realizowalności.



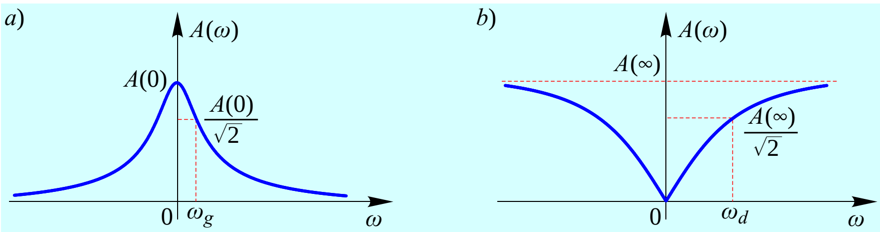
Rys. 7.15. Charakterystyki filtrów idealnych: dolnoprzepustowego LP (a), górnoprzepustowego HP (b), środkowoprzepustowego BP (c) i środkowozaporowego SB (d)

7.5.5. Filtry rzędu pierwszego

W praktyce filtry idealne zastępuje się filtrami fizycznie realizowalnymi, których charakterystyki amplitudowo-fazowe aproksymują według określonego kryterium charakterystyki amplitudowo-fazowe filtrów idealnych. Przykładem najprostszego rzeczywistego filtru dolnoprzepustowego jest czwórnik całkujący RC pokazany na rys. 7.14a o charakterystyce amplitudowo-fazowej (7.76). Jego charakterystyki amplitudowa i fazowa są określone wzorami (por. wzory (7.47)):

$$A(\omega) = \frac{1}{\sqrt{1+\omega^2\tau_{RC}^2}}, \qquad \varphi(\omega) = -\operatorname{arctg}\omega\tau_{RC}. \tag{7.85}$$

Wykres charakterystyki amplitudowej tego filtru jest pokazany na rys. 7.16a. Jak widzimy, w przeciwieństwie do charakterystyki idealnego filtru LP, dąży ona w sposób gładki do zera, gdy pulsacja rośnie do nieskończoności. Za górną pulsację graniczną ω_g pasma przepustowego rozpatrywanego filtru przyjmuje się z reguły pulsację, dla której krzywa $A(\omega)$ maleje o 3 dB ($\sqrt{2}$-krotnie), lub pulsację, dla której zmaleje ona do 0,1 wartości maksymalnej $A(0)$.



Rys. 7.16. Charakterystyki amplitudowe filtrów rzędu pierwszego: dolnoprzepustowego RC (a) i górnoprzepustowego RC (b)

Z kolei przykładem najprostszego rzeczywistego filtru górnoprzepustowego jest czwórnik różniczkujący RC z rys. 7.12a o charakterystyce amplitudowo-fazowej (7.72). Charakterystyki amplitudowa i fazowa tego filtru są określone wzorami (por. wzory (7.46)):

$$A(\omega) = \frac{|\omega|\tau_{RC}}{\sqrt{1+\omega^2\tau_{RC}^2}} \qquad \varphi(\omega) = \frac{\pi}{2}\operatorname{sgn}\omega - \operatorname{arctg}\omega\tau_{RC}. \tag{7.86}$$

Charakterystyka amplitudowa filtru jest pokazana na rys. 7.16b. Podobnie jak w przypadku dolnoprzepustowego filtru RC, za dolną pulsację graniczną ω_d pasma górnoprzepustowego filtru RC przyjmuje się pulsację, dla której charakterystyka $A(\omega)$ jest mniejsza o 3 dB od swojej wartości maksymalnej $A(\infty)$ lub jest równa 0,1 tej wartości.

Podkreślmy, że mianowniki transmitancji obu rozpatrywanych wyżej filtrów są wielomianami stopnia pierwszego (por. przykłady 7.13 i 7.14). Transmitancje te mają zatem tylko jeden pojedynczy biegun rzeczywisty. Z tego względu filtry te są nazywane *filtrami rzędu pierwszego*.

7.5.6. Filtry rzędu drugiego

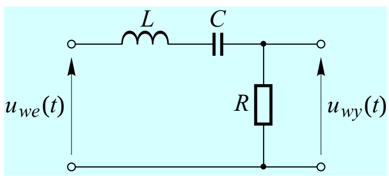
Za pomocą filtrów rzędu pierwszego mogą być aproksymowane jedynie charakterystyki idealnego filtru LP lub idealnego filtru HP. W celu aproksymacji charakterystyk idealnych filtrów BP i SB wymagane jest zastosowanie filtrów co najmniej rzędu drugiego, których transmitancje mają bieguny zespolone.

Najprostszym filtrem fizycznie realizowalnym, za pomocą którego możemy aproksymować charakterystykę idealnego filtru BP, jest obwód rezonansowy RLC, pracujący w strukturze czwórnikowej przedstawionej na rys. 7.17. Transmitancja tego obwodu ma postać:

$$H(s) = \frac{U_{\text{wy}}(s)}{U_{\text{we}}(s)} = \frac{\frac{R}{L}s}{s^2 + \frac{R}{L}s + \frac{1}{LC}}. \tag{7.87}$$

Wprowadzając parametry: pulsację rezonansową $\omega_0 = 1/\sqrt{LC}$ oraz dobroć $Q = \omega_0 L/R$, transmitancję tę możemy zapisać w postaci:

$$H(s) = \frac{\frac{1}{Q\omega_0}s}{\frac{s^2}{\omega_0^2} + \frac{s}{Q\omega_0} + 1}. \tag{7.88}$$



Rys. 7.17. Obwód rezonansowy RLC jako filtr środkowoprzepustowy

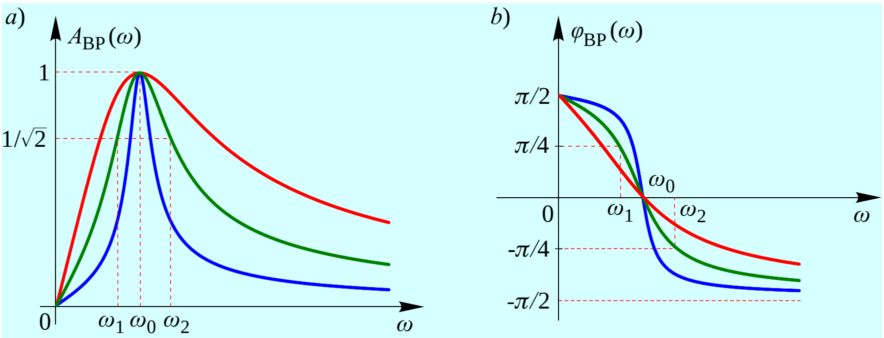
Jeśli dobroć filtru $Q > 1/2$, wyróżnik trójmianu kwadratowego w mianowniku transmitancji (7.88) jest ujemny. Ma ona wówczas dwa bieguny zespolone sprzężone:

$$s_{1,2} = \frac{\omega_0}{2Q}\left(-1 + \mathrm{j}\sqrt{4Q^2 - 1}\right) \tag{7.89}$$

położone w lewej półpłaszczyźnie $\mathrm{Re}\, s < 0$ zmiennej zespolonej s. Ich odległość od osi urojonej $\mathrm{Re}\ s = 0$ maleje wraz ze wzrostem dobroci Q. Charakterystyka amplitudowo-fazowa rozpatrywanego filtru ma postać:

$$H(\mathrm{j}\,\omega) = \frac{\mathrm{j}\,\dfrac{1}{Q}\dfrac{\omega}{\omega_0}}{1 - \dfrac{\omega^2}{\omega_0^2} + \mathrm{j}\,\dfrac{1}{Q}\dfrac{\omega}{\omega_0}}. \tag{7.90}$$

Wykresy prawostronne charakterystyk amplitudowych $A(\omega)$ i fazowych $\varphi(\omega)$ dla kilku wartości dobroci Q są przedstawione na rys. 7.18. Jak widzimy, cha-



Rys. 7.18. Charakterystyki amplitudowe (a) i fazowe (b) filtru środkowoprzepustowego RLC

rakterystyki amplitudowe są skupione wokół pulsacji rezonansowej $\omega = \omega_0$, dla której przybierają wartość maksymalną równą 1 (impedancja gałęzi poziomej czwórnika RLC jest dla tej pulsacji równa zeru). Są one tym węższe, im większa jest dobroć Q. Ze wzrostem dobroci filtru wzrasta więc jego selektywność. Jako pasmo przepustowe filtru przyjmuje się z reguły pasmo 3-decybelowe, tzn. taki przedział pulsacji $\omega \in (\omega_1, \omega_2)$, w którym charakterystyka amplitudowa nie maleje bardziej niż $\sqrt{2}$-krotnie w stosunku do wartości maksymalnej $A(\omega_0) = 1$ (dla którego $A(\omega) \geqslant 1/\sqrt{2}$). Dokonując odpowiednich obliczeń, można wykazać, że dla dostatecznie dużych wartości dobroci (arbitralnie przyjmuje się $Q > 5$) szerokość 3-decybelowego pasma przepustowego jest określona wzorem:

$$B|_{3\,\mathrm{dB}} \cong \frac{\omega_0}{Q} \tag{7.91}$$

a więc jest Q-krotnie mniejsza od pulsacji rezonansowej.

Wadą filtru środkowoprzepustowego RLC jest zbyt wolne opadanie do zera charakterystyki amplitudowej poza pasmem przepustowym, a w konsekwencji niedostateczna dokładność aproksymacji charakterystyki idealnego filtru środkowoprzepustowego. Dokładniejszą aproksymację można uzyskać, stosując bardziej złożone układy selektywne, np. obwody rezonansowe sprzężone (por. np. [6], p. 9.1).

7.5.7. Filtry wszechprzepustowe

W licznych zagadnieniach praktycznych szerokie zastosowanie znajdują filtry, których charakterystyka amplitudowa $A(\omega)$ jest stałą funkcją pulsacji w całym zakresie jej zmian. Filtry takie noszą nazwę *wszechprzepustowych*.

Filtrem wszechprzepustowym jest omówiony w p. 7.5.1 układ opóźniający o transmitancji (7.68). Jego charakterystyki amplitudowa i fazowa zostały przedstawione na rys. 7.10c, d. Przypominamy, że nie jest to jednak układ skupiony.

Przykładem filtru wszechprzepustowego skupionego jest *przesuwnik fazy* zrealizowany w układzie RC przedstawionym na rys. 7.19a. Z analizy układu wynika, że jego transmitancja jest określona wzorem:

$$H(s) = \frac{U_{\text{wy}}(s)}{U_{\text{we}}(s)} = \frac{1 - sRC}{1 + sRC}. \tag{7.92}$$

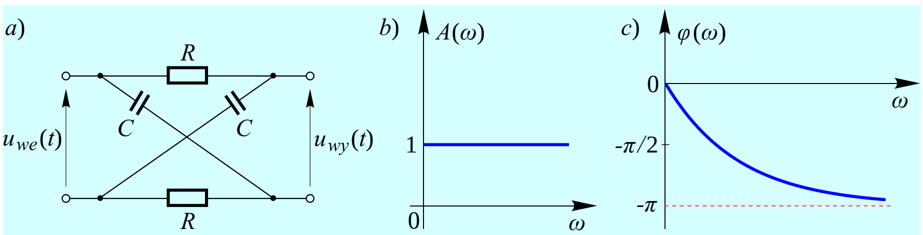
Stąd wyznaczamy charakterystykę amplitudowo-fazową

$$H(\mathrm{j}\,\omega) = \frac{1 - \mathrm{j}\,\omega RC}{1 + \mathrm{j}\,\omega RC} \tag{7.93}$$

oraz charakterystyki amplitudową i fazową

$$A(\omega) \equiv 1, \qquad \varphi(\omega) = -2\mathrm{arctg}\omega RC. \tag{7.94}$$

Charakterystyki te wykreślono dla $\omega \geqslant 0$ na rys. 7.19b i c. Charakterystyka fazowa dla małych pulsacji zmienia się w przybliżeniu liniowo. Zakres tych zmian może być regulowany przez dobór stałej RC filtru.



Rys. 7.19. Przesuwnik fazy RC (a) oraz jego charakterystyki: amplitudowa (b) i fazowa (c)

Słownik

charakterystyka amplitudowa
moduł charakterystyki amplitudowo-fazowej układu

charakterystyka amplitudowo-fazowa
funkcja zespolona zmiennej rzeczywistej ω (pulsacji) lub f (częstotliwości) opisująca układ w dziedzinie częstotliwości; transformata Fouriera odpowiedzi impulsowej układu

charakterystyka fazowa
argument charakterystyki amplitudowo-fazowej układu

filtr dolnoprzepustowy
filtr tłumiący składowe widma sygnału o wysokich częstotliwościach i przepuszczający składowe o małych częstotliwościach

filtr górnoprzepustowy
filtr przepuszczający składowe widma sygnału o wysokich częstotliwościach i tłumiący składowe o małych częstotliwościach

filtr środkowoprzepustowy
filtr przepuszczający składowe widma sygnału położone w ograniczonym wąskim paśmie

filtr środkowozaporowy
filtr tłumiący składowe widma w wąskim paśmie częstotliwości i przepuszczający pozostałe składowe

filtr wszechprzepustowy
filtr o stałej charakterystyce amplitudowej w całym zakresie częstotliwości

krzywa Nyquista
wykres charakterystyki amplitudowo-fazowej na płaszczyźnie zmiennej zespolonej

metoda operatorowa
metoda opisu i analizy sygnałów i układów w dziedzinie zespolonej

odpowiedź impulsowa układu
sygnał, jaki wystąpi na wyjściu układu przy pobudzeniu go impulsem Diraca (w przypadku układów analogowych) lub impulsem Kroneckera (w przypadku układów dyskretnych)

odpowiedź jednostkowa układu
sygnał, jaki wystąpi na wyjściu układu przy pobudzeniu go skokiem jednostkowym (analogowym w przypadku układów analogowych lub dyskretnym w przypadku układów dyskretnych)

odpowiedź układu
sygnał, jaki wystąpi na wyjściu układu przy pobudzeniu go pewnym sygnałem wejściowym

para transformat Laplace'a
para funkcji, którą tworzy sygnał i jego transformata Laplace'a

pobudzenie układu
sygnał występujący na wejściu układu

podstawowy warunek realizowalności układu
warunek, zgodnie z którym odpowiedź impulsowa każdego realnie istniejącego układu musi przybierać wartości zerowe dla $t < 0$

przekształcenie Laplace'a
przekształcenie całkowe przyporządkowujące sygnałowi funkcję zespoloną zmiennej zespolonej s (transformatę Laplace'a)

przesuwnik fazy
układ, którego zadaniem jest odpowiednie przesunięcie fazy sygnału bez zmiany jego amplitudy

równanie transmisyjne układu
równanie wiążące widma sygnału wejściowego i sygnału wyjściowego układu w dziedzinie częstotliwości lub wiążące transformaty Laplace'a tych (?!) sygnałów w dziedzinie zespolonej

transformata Laplace'a
funkcja zespolona zmiennej zespolonej s, otrzymywana w wyniku przekształcenia całkowego Laplace'a sygnału

transmitancja układu
funkcja zespolona zmiennej zespolonej opisująca układ w dziedzinie zespolonej (transformata Laplace'a odpowiedzi impulsowej układu w przypadku układów analogowych)

układ
model matematyczny urządzenia fizycznego przetwarzającego sygnał wejściowy na sygnał wyjściowy

układ analogowy
układ przetwarzający sygnał analogowy w sygnał analogowy

układ dyskretny
układ przetwarzający sygnał dyskretny (ciąg próbek) w inny sygnał dyskretny (inny ciąg próbek)

układ stabilny
układ, który na ograniczone pobudzenie odpowiada ograniczoną odpowiedzią

zasada przyczynowości
zasada orzekająca, iż sygnał na wyjściu układu realizowalnego fizycznie nie może pojawić się przed podaniem na jego wejście sygnału pobudzającego (tj. orzekająca, iż efekt nie może poprzedzać przyczyny)

Literatura

[1] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom I. WNT, Warszawa, wyd. 4, 2000.

[2] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom III. WNT, Warszawa, 1995.

[3] Filipkowski A.: Układy elektroniczne analogowe i cyfrowe. WNT, Warszawa, 1993.

[4] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

[5] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom II. WNT, Warszawa, wyd. 3, 1999?.

[6] Baskakow Ś.: Sygnały i układy radiotechniczne. PWN, Warszawa, 1991 (tłum. z ros.)

Lekcja 8

Repetytorium

Rozdział 3

Modulacja sygnałów

Lekcja 9

Ogólna charakterystyka operacji modulacji

Problemy pozyskiwania, przetwarzania, gromadzenia, a przede wszystkim przesyłania informacji odgrywają we współczesnym świecie coraz bardziej doniosłą rolę. Informacja stała się dobrem powszechnym, o które zabiegają politycy, przedstawiciele świata biznesu, nauki i innych dziedzin życia. Stale rosnące zapotrzebowanie na informację spowodowało, że w ostatnich latach nastąpił niezwykle dynamiczny rozwój *systemów telekomunikacyjnych* przeznaczonych do przesyłania informacji na odległość. Dzięki postępom w tej dziedzinie możliwe jest w chwili obecnej przesyłanie informacji w sposób coraz sprawniejszy, szybszy, bardziej niezawodny i na coraz dalsze odległości.

Wzrost wymagań stawianych przed projektantami systemów telekomunikacyjnych doprowadził do znacznego zwiększenia stopnia ich złożoności. We współczesnych systemach wykorzystuje się wyrafinowane metody przetwarzania sygnałów oparte na zaawansowanym aparacie matematycznym. Wprowadza się między innymi coraz to nowsze, doskonalsze systemy modulacji sygnałów – podstawowej operacji dokonywanej na sygnałach w celu ich efektywnego przesłania na dalekie odległości. Rozdział niniejszy jest poświęcony szczegółowemu omówieniu tej operacji. Rozpatrzymy trzy podstawowe rodzaje modulacji sygnałów: modulacje analogowe, modulacje impulsowe oraz modulacje cyfrowe. Przeanalizujemy właściwości sygnałów zmodulowanych, w tym przede wszystkim ich właściwości widmowe. Omówimy także metody generacji sygnałów zmodulowanych i sposoby ich demodulacji.

W lekcji 9, rozpoczynającej omawianie problematyki modulacji sygnałów, podamy ogólną charakterystykę operacji modulacji. Omówimy schemat blokowy systemu telekomunikacyjnego, wskażemy na cele modulacji oraz przeprowadzimy klasyfikację systemów modulacji.

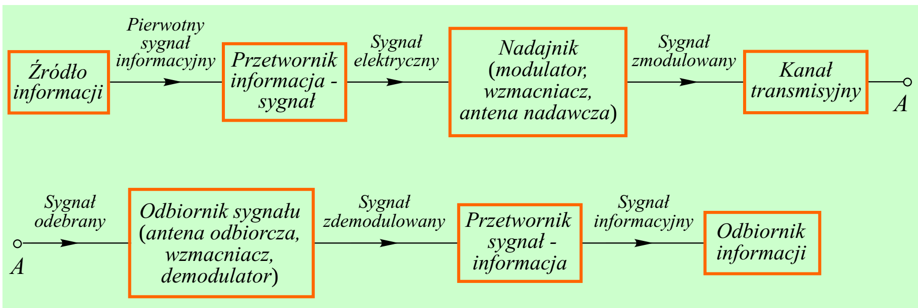
9.1. Schemat systemu telekomunikacyjnego

W istniejących systemach telekomunikacyjnych jest stosowanych wiele różnych rodzajów modulacji sygnałów. Rodzaj zastosowanej modulacji decyduje w zasadniczy sposób o typie, strukturze i właściwościach systemu. W konsekwencji między poszczególnymi systemami telekomunikacyjnymi występują znaczne różnice. Mimo tych różnic, w każdym systemie można wyróżnić pewne standardowe bloki, spełniające identyczną rolę funkcjonalną. Prezentację problematyki modulacji sygnałów rozpoczniemy od krótkiego omówienia schematu blokowego systemu telekomunikacyjnego i zasad działania jego poszczególnych bloków.

9.1.1. Bloki funkcjonalne systemu

Ogólny schemat systemu telekomunikacyjnego, przeznaczonego do przesyłania informacji od nadawcy do odbiorcy, jest przedstawiony na rys. 9.1. Po stronie nadawczej systemu występuje *źródło informacji*, generujące pierwotny *sygnał informacyjny*, *przetwornik informacja-sygnał* oraz *nadajnik*, którego zadaniem jest wytworzenie *sygnału zmodulowanego*. Sygnał otrzymany na wyjściu nadajnika jest transmitowany przez *kanał transmisyjny* do *odbiornika*.

Po stronie odbiorczej występuje *odbiornik sygnału*, *przetwornik sygnał-informacja* oraz *odbiornik informacji*. Zasadniczym zadaniem odbiornika sygnału jest dokonanie *demodulacji* sygnału – operacji odwrotnej do modulacji. Sygnał zdemodulowany jest następnie przetwarzany przez przetwornik sygnał-informacja, na którego wyjściu występuje odzyskany sygnał informacyjny. Sygnał ten jest już bezpośrednio odbierany przez odbiornik informacji, którym w większości przypadków jest człowiek.



Rys. 9.1. Schemat systemu telekomunikacyjnego

9.1.2. Źródło informacji i przetwornik informacja-sygnał

Efektywne przekazywanie informacji na odległość jest realizowane za pomocą fal elektromagnetycznych, tj. sygnałów elektrycznych lub – w przypadku systemów światłowodowych – optycznych. Zazwyczaj sygnał informacyjny generowany w źródle informacji nie jest w swojej pierwotnej postaci sygnałem elektrycznym. Na przykład, w systemach telefonicznych sygnałem informacyjnym jest sygnał akustyczny mowy, w systemach radiowych – sygnał fonii (mowy lub muzyki) emitowany bezpośrednio przez spikera bądź odtwarzany z taśmy magnetycznej, płyt gramofonowych lub płyt CD. W systemach telewizyjnych sygnałem informacyjnym jest – obok sygnału fonii – sygnał wizji, w systemach transmisji danych (modemowych, internetowych) – ciągi danych zapisane na odpowiednim nośniku, itd. System telekomunikacyjny musi zatem zawierać przetwornik przetwarzający sygnał pierwotny w sygnał elektryczny. Rolę takiego przetwornika w systemach telefonicznych pełni mikrofon, w systemach radiowych – mikrofon lub urządzenia odtwarzające, w systemach telewizyjnych – mikrofon i kamera telewizyjna, a w systemach transmisji danych – urządzenia przetwarzające informację binarną w ciągi impulsów elektrycznych.

9.1.3. Nadajnik

W nadajniku jest realizowany cały zespół operacji mających na celu odpowiednie ukształtowanie sygnału przekazywanego do odbiornika przez kanał transmisyjny. Zestaw tych operacji w różnych systemach jest na ogół różny. W każdym nadajniku występują jednak dwa podstawowe układy: *modulator* oraz *wzmacniacz*. Zadaniem modulatora jest wytworzenie sygnału zmodulowanego, tj. przetworzenie sygnału elektrycznego do postaci, która umożliwia jego transmisję na odległość. Zadaniem układów wzmacniających jest natomiast zapewnienie odpowiedniej mocy tego sygnału. W przypadku systemów bezprzewodowych (radiowych, telewizyjnych, radiolokacyjnych, telemetrycznych) w nadajniku występuje ponadto *antena nadawcza*, za pomocą której sygnał zmodulowany jest emitowany w wolną przestrzeń.

9.1.4. Kanał transmisyjny

Mówiąc ogólnie, kanałem transmisyjnym nazywamy środowisko, w którym sygnał jest przekazywany od nadajnika do odbiornika. Można wyróżnić trzy zasadnicze typy kanałów transmisyjnych: *kanały przewodowe* (kablowe), *kanały bezprzewodowe* (powietrzne) i *kanały optyczne* (światłowodowe). W kanałach przewodowych i optycznych, stosowanych na przykład w systemach telefonicznych, modemowych, internetowych, telewizji przemysłowej i kablowej, sygnał jest prowadzony wzdłuż określonej drogi w ośrodku stałym (kablu lub światłowodzie). W kanałach bezprzewodowych, a więc np. w systemach radiowych, telewizyjnych, radiolokacyjnych, telemetrycznych i telefonii komórkowej, sygnał jest emitowany w postaci fali elektromagnetycznej w wolną przestrzeń. W wielu współczesnych systemach kanały transmisyjne mają charakter mieszany. Niektóre odcinki toru transmisyjnego są realizowane w postaci kanałów przewodowych, bądź światłowodowych, inne zaś – powietrznych.

Należy podkreślić, że przesyłaniu sygnałów przez kanał transmisyjny zawsze towarzyszą dwa negatywne zjawiska. Po pierwsze, sygnał na drodze propagacji od nadajnika do odbiornika ulega tłumieniu. Wynika stąd konieczność stosowania różnego rodzaju przekaźników i retlanslatorów, których zadaniem jest m.in. wzmocnienie sygnału po przejściu przez kolejne odcinki toru transmisyjnego. Po drugie zaś, na transmitowany sygnał oddziaływuje w kanale różnego rodzaju szumy i zakłócenia, zniekształcające pierwotny sygnał informacyjny. Walka z tymi zakłóceniami jest jednym z bardziej istotnych problemów telekomunikacji.

9.1.5. Odbiornik sygnału

W odbiorniku sygnału są realizowane dwie podstawowe operacje: wzmocnienie sygnału odebranego i jego demodulacja. Wzmocnienie sygnału jest konieczne, aby możliwe było dalsze jego przetwarzanie. Demodulacja natomiast przywraca pierwotną postać sygnału informacyjnego. W niektórych przypadkach, jak na przykład w systemach radiolokacyjnych, czy systemach łączności z sondami międzyplanetarnymi, moc sygnału użytecznego na wejściu odbiornika może być dużo mniejsza od mocy sygnału zakłócającego. Mówimy wówczas, że sygnał jest ukryty w szumie. Zadaniem odbiornika jest w takim przypadku, po pierwsze, wykrycie sygnału w szumie, a po wtóre – wyeliminowanie towarzyszących mu zakłóceń i tym samym możliwie najbardziej wierne jego odtworzenie.

9.1.6. Przetwornik sygnał-informacja i odbiornik informacji

Sygnał elektryczny, otrzymany na wyjściu demodulatora, musi być jeszcze przetworzony do postaci dostosowanej do odbiorcy informacji, a więc zamieniony np. w sygnał akustyczny lub wizyjny. Zadanie to wykonuje przetwornik sygnał-informacja. W łączności telefonicznej jest nim membrana słuchawki telefonicznej, w radiowej – głośnik, w telewizyjnej – głośnik i ekran telewizyjny, a w systemach transmisji danych – drukarka lub monitor komputera. Sygnał z wyjścia tego przetwornika jest już odbierany bezpośrednio przez odbiorcę informacji.

9.2. Cele i rodzaje modulacji

9.2.1. Cele modulacji

Bez modulacji przesyłanie sygnałów na dalekie odległości byłoby niemożliwe. Modulacja ma przede wszystkim na celu dopasowanie właściwości widmowych sygnału do charakterystyk częstotliwościowych kanału transmisyjnego. Sygnał informacyjny jest z reguły sygnałem dolnopasmowym, o znaczącym udziale składowych niskoczęstotliwościowych, podczas gdy istniejące kanały są zawsze środkowoprzepustowe. Bezpośrednie przesyłanie sygnału w jego naturalnym paśmie, nazywanym *pasmem podstawowym*, powodowałoby zatem niedopuszczalne zniekształcenia tych składowych. W wyniku modulacji widmo sygnału jest przenoszone w zakres wyższych częstotliwości, co umożliwia dopasowanie widmowe sygnału do kanału.

W przypadku kanałów bezprzewodowych modulacja umożliwia efektywne wyemitowanie sygnału w postaci fali elektromagnetycznej w wolną przestrzeń. Z teorii pola elektromagnetycznego wiadomo, że skuteczne wypromieniowanie fali elektromagnetycznej wymaga użycia anteny nadawczej o rozmiarach niemniejszych niż $1/10$ długości tej fali, a ponadto, w celu uzyskania odpowiedniego zasięgu, bardzo dużych mocy nadajnika. Gdyby np. sygnał akustyczny, którego widmo jest zawarte w przedziale częstotliwości 200 Hz – 20 kHz, był transmitowany w swoim paśmie podstawowym, to – jak można łatwo obliczyć – wymagane rozmiary anteny sięgałyby kilku kilometrów przy niewielkim zasięgu. Przeniesienie widma sygnału informacyjnego w zakres wyższych częstotliwości umożliwia zatem zwiększenie zasięgu systemu telekomunikacyjnego lub – przy ustalonym zasięgu – zmniejszenie mocy nadajnika. Im widmo sygnału zmodulowanego jest bardziej przesunięte w zakres dużych częstotliwości, tym mniejsze mogą być rozmiary anteny i mniejsza moc nadajnika.

Modulacja umożliwia ponadto efektywne wykorzystanie pasma przepustowego kanału. Bez modulacji przez jeden kanał można przesłać tylko jeden sygnał. Jednoczesna transmisja dwóch niezmodulowanych sygnałów, o pokrywających się pasmach podstawowych, spowodowałaby nałożenie się w kanale ich widm, co uniemożliwiłoby odseparowanie tych sygnałów po stronie odbiorczej.

Stosując modulacje analogowe, widma sygnałów można przenieść w różne, nie zachodzące na siebie, pasma częstotliwości i tym samym odseparować je w odbiorniku za pomocą dostrojonych do tych pasm filtrów środkowoprzepustowych. Modulacja umożliwia zatem przekazywanie wielu sygnałów przez jeden kanał transmisyjny w systemach *zwielokrotnienia częstotliwościowego*, nazywanych także systemami *z podziałem częstotliwościowym* (*systemami FDM*, od ang. *Frequency-Division Multiplexing*). W ten sposób jest organizowana transmisja sygnałów np. w systemach telefonicznych (zarówno klasycznych analogowych, jak i współczesnych cyfrowych), a także w systemach radiowych i telewizyjnych powszechnego użytku.

Z drugiej strony, stosując modulacje impulsowe lub cyfrowe, można organizować transmisję wielu sygnałów przez jeden kanał w systemach *zwielokrotnienia czasowego*, nazywanych systemami *z podziałem czasowym* (*systemami TDM*, od ang. *Time-Division Multiplexing*). W systemach tych transmitowane sygnały są przemieszane w częstotliwości, ale odseparowane od siebie w czasie, tzn. próbki sygnałów (niezakodowane w systemach impulsowych lub zakodowane w systemach cyfrowych) są przekazywane w różnych, ściśle określonych przedziałach czasu. Umożliwia to rozdzielenie próbek po stronie odbiorczej za pomocą układów separujących odpowiednio synchronizowanych w czasie.

W ostatnich latach jest rozwijana intensywnie jeszcze inna koncepcja cyfrowych systemów transmisji wielu sygnałów przez wspólny kanał. Systemy oparte na tej koncepcji są nazywane systemami *z podziałem kodowym* (*systemami CDMA*, od ang. *Code-Division Multiple Access*). Sygnały w tych systemach są przemieszane zarówno w częstotliwości, jak i w czasie, ale każdy z nich posiada swój własny identyfikator w postaci odpowiedniej *sekwencji kodującej*. Dzięki temu możliwa jest ich separacja po stronie odbiorczej. Systemy CDMA umożliwiają jeszcze bardziej efektywne wykorzystanie możliwości transmisyjnych kanału. Są one coraz częściej stosowane w praktyce i – zdaniem specjalistów – w ich kierunku będzie następował rozwój systemów telekomunikacyjnych.

9.2.2. Fala nośna

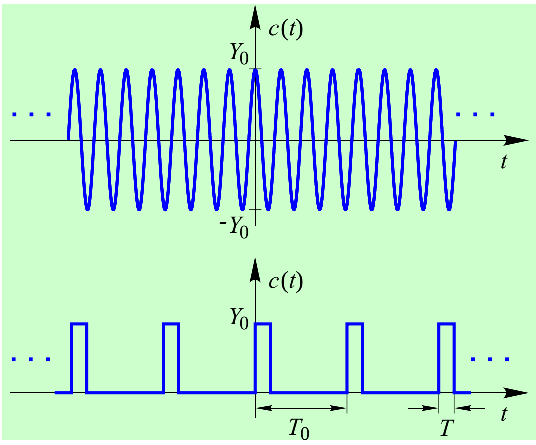
Modulację określa się często jako proces uzmienniania parametrów ustalonego standardowego sygnału $c(t)$, nazywanego *sygnałem nośnym* lub *falą nośną*. Parametry te są uzmienniane w zależności od bieżących wartości sygnału informacyjnego. Sygnał informacyjny jest nazywany *sygnałem modulującym*, a fala nośna – *sygnałem modulowanym*. Sygnał otrzymany w wyniku operacji modulacji jest nazywany *sygnałem zmodulowanym*.

W analogowych systemach modulacji jako falę nośną wykorzystuje się sygnał harmoniczny (rys. 9.2a):

$$c(t) = Y_0 \cos(\Omega t + \varphi_0). \tag{9.1}$$

Sygnał ten jest określony przez trzy parametry: amplitudę Y_0, pulsację Ω (częstotliwość $F = \Omega/2\pi$) oraz fazę początkową φ_0. Pulsacja Ω (częstotliwość F) jest nazywana *pulsacją* (*częstotliwością*) *nośną*. W wyniku modulacji jeden z tych parametrów jest uzmienniany w takt zmian sygnału informacyjnego, przy czym zmiany te następują w sposób ciągły w czasie. Z tego względu modulacje analogowe są niekiedy nazywane *modulacjami ciągłymi*.

W impulsowych systemach modulacji funkcję fali nośnej spełnia unipolarna fala prostokątna (rys. 9.2b), z reguły o małym współczynniku wypełnienia. Również w tym przypadku fala nośna jest określona trzema parametrami: amplitudą impulsów Y_0, czasem ich trwania T oraz okresem T_0. Modulacja polega w tym przypadku na zmianach wartości jednego z tych parametrów w zależności od bieżących wartości próbek sygnału informacyjnego, przy czym uzmiennianie okresu T_0 polega na zmianach odległości między kolejnymi impulsami fali nośnej. Zmiany wartości parametrów następują od impulsu do impulsu (od próbki do próbki), a więc zachodzą w czasie w sposób skokowy. Jeśli uzmienniany parametr przybiera wartości w zbiorze ciągłym, modulację impulsową nazywamy analogową. Jeśli natomiast w wyniku kwantowania uzmienniany parametr przybiera wartości w zbiorze skończonym, modulację impulsową nazywamy cyfrową. W tym drugim przypadku modulacja jest nazywana także modulacją *impulsowo-kodową* .

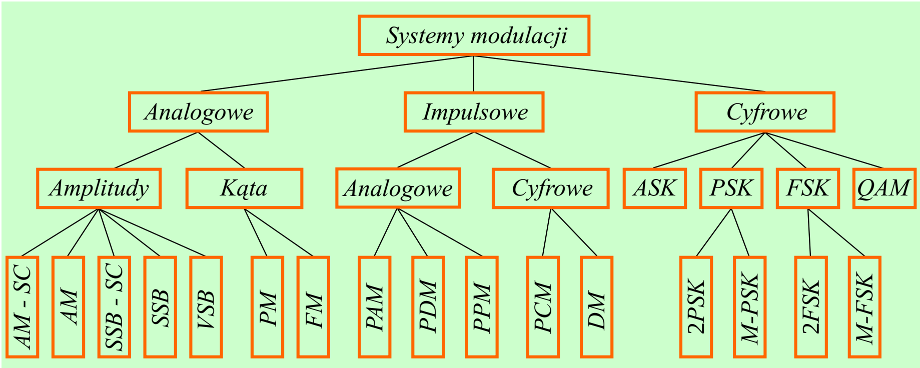


Rys. 9.2. Fale nośne: harmoniczna (a) i unipolarna fala prostokątna (b)

W cyfrowych systemach modulacji, podobnie jak w systemach analogowych, fala nośna jest sygnałem harmonicznym o postaci (9.1). Wartości parametrów tej fali są zmieniane skokowo w kolejnych odcinkach czasu, nazywanych *przedziałami sygnałowymi* (ang. *signaling interval*). Zmiany te następują w zależności od aktualnie transmitowanego znaku binarnego (bitu) lub – ogólniej – w zależności od aktualnie transmitowanego ciągu znaków binarnych o ustalonej długości (transmitowanego symbolu). Zmianom może podlegać amplituda, faza, częstotliwość lub jednocześnie amplituda i faza. Istotną cechą modulacji cyfrowych jest to, iż w każdym przedziale sygnałowym parametry fali nośnej mogą przybierać wartości jedynie ze zbioru skończonego.

9.2.3. Klasyfikacja systemów modulacji

Podział najczęściej spotykanych systemów modulacji jest przedstawiony na rys. 9.3. Systemy te można generalnie podzielić na analogowe (ciągłe), impulsowe oraz cyfrowe.



Rys. 9.3. Klasyfikacja systemów modulacji

Modulacje analogowe, w których w zależności od sygnału informacyjnego jest uzmienniana amplituda fali nośnej, są nazywane *modulacjami amplitudy*. Do systemów tych należą modulacje:

- *dwuwstęgowa bez fali nośnej*, nazywana *modulacją AM-SC* (od ang. *Amplitude Modulation with Suppressed Carrier*) lub *modulacją DSB-SC* (od ang. *Double Sideband with Suppressed Carrier*),
- *dwuwstęgowa z falą nośną – modulacja AM* lub *DSB*,
- *jednowstęgowa bez fali nośnej – SSB-SC* (od ang. *Single Sideband with Suppressed Carrier*),
- *jednowstęgowa z falą nośną – SSB*,
- *z częściowo stłumioną* (*szczątkową*) *wstęgą boczną – VSB* (od ang. *Vestigial Sideband*).

Modulacje analogowe, w których zmianom podlega kąt fali nośnej, są nazywane *modulacjami kąta*. Obejmują one modulacje:

- *fazy* (*modulację PM*, od ang. *Phase Modulation*),
- *częstotliwości* (*modulację FM*, od ang. *Frequency Modulation*).

Do najczęściej stosowanych w praktyce impulsowych systemów modulacji należą modulacje:

- *amplitudy impulsów* (*modulacja PAM*, od ang. *Pulse Amplitude Modulation*),
- *szerokości impulsów* (*modulacja PDM*, od ang. *Pulse Duration Modulation*),
- *położenia impulsów* (*modulacja PPM*, od ang. *Pulse Position Modulation*),
- *modulacja impulsowo-kodowa* (*modulacja PCM*, od ang. *Pulse-Code Modulation*),
- warianty modulacji PCM: *modulacja przyrostową PCM*, nazywana także *modulacją delta DM* oraz *modulacja sigma-delta SDM*.

Z kolei do najczęściej stosowanych w praktyce cyfrowych systemów modulacji zaliczamy modulacje z *kluczowaniem* (*manipulacją*):

- *amplitudy – ASK* (od ang. *Amplitude Shift Keying*),
- *fazy – PSK* (od ang. *Phase Shift Keying*),
- *częstotliwości – FSK* (od ang. *Frequency Shift Keying*),

– *modulację QAM* (od ang. *Quadrature Amplitude Modulation*), z jednoczesnym kluczowaniem amplitudy i fazy.

Najprostszymi modulacjami cyfrowymi są *modulacje dwuwartościowe* (*binarne*): 2ASK, 2PSK i 2FSK. W systemach tych jeden z parametrów fali nośnej, tj. amplituda, faza lub częstotliwość, przybiera w każdym przedziale sygnałowym (nazywanym w tym przypadku *przedziałem bitowym*) tylko jedną z dwóch wartości. Dążenie do coraz bardziej efektywnego i szybkiego przesyłania informacji spowodowało, że ostatnio opracowano i wdrożono w praktyce bardziej złożone systemy modulacji cyfrowej, takie jak wielowartościowe modulacje PSK i FSK, czy modulacja QAM. W tej ostatniej, stosowanej m.in. w najnowszych standardach modemowych, skokowym zmianom w poszczególnych przedziałach sygnałowych (nazywanych w tym przypadku *przedziałami symbolowymi*) podlegają jednocześnie dwa parametry fali nośnej: amplituda i faza.

Należy podkreślić, że przedstawiona klasyfikacja obejmuje jedynie najważniejsze ze znanych i stosowanych w praktyce systemów modulacji sygnałów. Pominięto w niej wiele rodzajów modulacji specjalnego przeznaczenia, wśród których można wymienić np. *liniową modulację częstotliwości LFM* (od ang. *Linear Frequency Modulation*), stosowaną w radiolokacji i technice sonarowej, systemy modulacji *z widmem rozproszonym*, a także wspomniane wyżej systemy CDMA z podziałem kodowym.

Słownik

częstotliwość nośna
: częstotliwość harmonicznej fali nośnej

demodulacja
: operacja odwrotna do operacji modulacji polegająca na odtworzeniu sygnału informacyjnego z sygnału zmodulowanego

demodulator
: układ odtwarzający sygnał informacyjny z sygnału zmodulowanego

fala nośna (sygnał nośny)
: standardowy sygnał (z reguły harmoniczny w analogowych i cyfrowych systemach modulacji lub unipolarna fala prostokątna w impulsowych systemach modulacji), którego parametry są uzmienniane w procesie modulacji zgodnie z bieżącymi wartościami sygnału modulującego

modulacja analogowa
: modulacja, w której parametry harmonicznej fali nośnej są uzmienniane

w czasie w sposób ciągły zgodnie z bieżącymi wartościami sygnału modulującego

modulacja cyfrowa
modulacja, w której w każdym takcie (przedziale bitowym lub przedziale symbolowym) transmitowana jest informacja binarna w postaci znaku binarnego lub ciągu znaków binarnych; znakom tym (lub ciągom) są przyporządkowywane sygnały będące odcinkami harmonicznej fali nośnej, której parametry są uzmienniane skokowo w czasie zgodnie z aktualnie transmitowanym znakiem (ciągiem)

modulator
układ wytwarzający sygnał zmodulowany w danym systemie modulacji

pasmo podstawowe sygnału
naturalne pasmo sygnału, jakie zajmuje on przed operacją modulacji

pulsacja nośna
pulsacja harmonicznej fali nośnej

sygnał informacyjny (modulujący)
sygnał, w którym zawarta jest użyteczna informacja przesyłana od nadajnika do odbiornika

sygnał zmodulowany
sygnał otrzymywany w wyniku modulacji fali nośnej sygnałem modulującym

sygnał zmodulowany amplitudowo
sygnał zmodulowany, którego amplituda chwilowa jest uzmienniana w zależności od sygnału modulującego

sygnał zmodulowany kątowo
sygnał zmodulowany, którego kąt chwilowy jest uzmienniany w zależności od sygnału modulującego

system modulacji z podziałem czasowym TDM
system telekomunikacyjny, w którym w jednym kanale są transmitowane próbki wielu sygnałów w ściśle określonych rozłącznych przedziałach czasu

system modulacji z podziałem częstotliwościowym FDM
system telekomunikacyjny, w którym w jednym kanale jest transmitowanych jednocześnie wiele sygnałów o rozłącznych pasmach

zwielokrotnienie czasowe
system jednoczesnej transmisji wielu sygnałów dyskretnych przez jeden

kanał transmisyjny, w którym poszczególne sygnały są transmitowane w rozłącznych przedziałach czasu

zwielokrotnienie częstotliwościowe
system jednoczesnej transmisji wielu sygnałów analogowych przez jeden kanał transmisyjny, w którym poszczególne sygnały są transmitowane w rozłącznych pasmach częstotliwości

Literatura

[1] Haykin S.: Systemy telekomunikacyjne. WKiŁ, Warszawa, 1999.

[2] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

Lekcja 10

Modulacje analogowe amplitudy

W lekcji 10 omówimy analogowe modulacje amplitudy. Rozpoczniemy od zdefiniowania pojęcia sygnału analitycznego i pojęć pokrewnych: amplitudy chwilowej (obwiedni), pulsacji (częstotliwości) chwilowej, fazy chwilowej, składowej kwadraturowej i składowej synfazowej sygnału oraz jego obwiedni zespolonej. Zdefiniujemy także pojęcie funkcji modulującej. Wprowadzenie tych pojęć ułatwia analizę sygnałów zmodulowanych amplitudowo, a ponadto umożliwia jej przeprowadzenie w ramach jednolitego ujęcia formalnego. Rozpatrzymy dwuwstęgowe systemy modulacji amplitudy AM-SC bez fali nośnej oraz AM z falą nośną, systemy jednowstęgowe SSB-SC i SSB oraz system z częściowo stłumioną wstęgą boczną VSB. W każdym z tych przypadków omówimy krótko zasady generacji sygnałów zmodulowanych i ich demodulacji. Dokonamy także porównania amplitudowych systemów modulacji pod kątem szerokości pasma i sprawności energetycznej.

10.1. Reprezentacja sygnału za pomocą sygnału analitycznego

W analizie sygnałów zmodulowanych analogowo wygodnie jest posługiwać się reprezentacją sygnału za pomocą jego *sygnału analitycznego*. Reprezentację tę można wprowadzić dla bardzo szerokiej klasy sygnałów, jest ona jednak szczególnie użyteczna w odniesieniu do klasy sygnałów wąskopasmowych, a więc klasy obejmującej wszystkie sygnały zmodulowane analogowo.

O pojęciu sygnału analitycznego wspomnieliśmy już w p. 1.2.7 podczas omawiania przykładów modeli zespolonych sygnału. Obecnie, przed przystąpieniem do prezentacji analogowych systemów modulacji, pojęcie to przybliżymy nieco dokładniej. Omówimy m.in. jego najważniejsze właściwości. Reprezentację sygnału za pomocą sygnału analitycznego wykorzystamy zarówno w analizie sygnałów zmodulowanych amplitudowo, jak i kątowo.

10.1.1. Sygnał analityczny

Jak pamiętamy z podstawowego kursu teorii obwodów, analizę obwodów prądu sinusoidalnie zmiennego ułatwia radykalnie posługiwanie się reprezentacją sygnałów harmonicznych za pomocą amplitud zespolonych. Rozszerzenie tej reprezentacji na sygnały nieharmoniczne prowadzi do pojęcia sygnału analitycznego.

Definicja 10.1. Niech $x(t)$ będzie sygnałem rzeczywistym, dla którego istnieje transformata Hilberta (por. wzór (1.13)):

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} \, \mathrm{d}\tau. \tag{10.1}$$

Sygnałem analitycznym sygnału $x(t)$ nazywamy sygnał zespolony:

$$z_x(t) = x(t) + \mathrm{j}\,\hat{x}(t), \tag{10.2}$$

którego częścią rzeczywistą jest sygnał $x(t)$, a częścią urojoną jego transformata Hilberta $\hat{x}(t) = \mathscr{H}[x(t)]$. Wzór (10.2) nazywamy także *postacią analityczną* sygnału $x(t)$.

Reprezentację analityczną (10.2) można przyporządkować wszystkim sygnałom, dla których całka (10.1) jest zbieżna. Zbiór tych sygnałów jest bardzo obszerny i obejmuje m.in. klasę sygnałów o ograniczonej energii oraz sygnały o ograniczonej mocy przedziałami ciągłe o zerowej składowej stałej. Znając sygnał analityczny $z_x(t)$ sygnału $x(t)$, można oczywiście wyznaczyć sygnał oryginalny ze wzoru:

$$x(t) = \mathrm{Re}\, z_x(t). \tag{10.3}$$

Natomiast znając transformatę Hilberta $\hat{x}(t)$, można odtworzyć sygnał oryginalny $x(t)$ obliczając odwrotną transformatę Hilberta:

$$x(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{x}(\tau)}{t-\tau} \, \mathrm{d}\tau. \tag{10.4}$$

Przykład 10.1. Transformatą Hilberta sygnału harmonicznego

$$x(t) = X_0 \cos(\omega_0 t + \varphi_0) \tag{10.5}$$

jest sygnał $\hat{x}(t) = X_0 \sin(\omega_0 t + \varphi_0)$ (por. przykład 10.5). Wynika stąd, że sygnał analityczny rzeczywistego sygnału harmonicznego (10.5) jest zespolonym sygnałem harmonicznym

$$z_x(t) = X_0 \, \mathrm{e}^{\mathrm{j}(\omega_0 t + \varphi_0)} \, . \tag{10.6}$$

10.1.2. Amplituda chwilowa i pulsacja chwilowa

Pojęcie sygnału analitycznego umożliwia uogólnienie na sygnały nieharmoniczne zwykłego pojęcia amplitudy i fazy sygnału harmonicznego. Przedstawmy w tym celu sygnał analityczny w postaci biegunowej:

$$\begin{aligned} z_x(t) = x(t) + \mathrm{j}\,\hat{x}(t) &= \sqrt{x^2(t) + \hat{x}^2(t)}\, \mathrm{e}^{\mathrm{j}\,\mathrm{arctg}[\hat{x}(t)/x(t)]} = \\ &= |z_x(t)|\, \mathrm{e}^{\mathrm{j}\arg z(t)} \triangleq X(t)\, \mathrm{e}^{\mathrm{j}\psi(t)} \, . \end{aligned} \tag{10.7}$$

Definicja 10.2. *Amplitudą chwilową* (*obwiednią*) $X(t)$ rzeczywistego sygnału $x(t)$ nazywamy moduł jego sygnału analitycznego:

$$X(t) \triangleq |z_x(t)| = \sqrt{x^2(t) + \hat{x}^2(t)}. \tag{10.8}$$

Definicja 10.3. *Pulsacją chwilową* $\omega(t)$ sygnału $x(t)$ nazywamy pochodną argumentu $\psi(t)$ jego sygnału analitycznego:

$$\omega(t) \triangleq \frac{\mathrm{d}\psi(t)}{\mathrm{d}t} = \mathrm{Im}\,[\ln z_x(t)]' = \mathrm{Im}\, \frac{z_x'(t)}{z_x(t)} = \frac{x(t)\hat{x}'(t) - x'(t)\hat{x}(t)}{x^2(t) + \hat{x}^2(t)}. \tag{10.9}$$

Przykład 10.2. W przypadku sygnału harmonicznego (10.5) mamy $X(t) = X_0$ oraz $\omega(t) = \mathrm{d}(\omega_0 t + \varphi_0)/\,\mathrm{d}t = \omega_0$, a więc amplituda chwilowa i pulsacja chwilowa tego sygnału są stałymi funkcjami czasu.

10.1.3. Faza chwilowa

Również trzeci parametr charakteryzujący sygnał harmoniczny, tj. faza, może być uogólniony na sygnały nieharmoniczne. W przeciwieństwie do amplitudy chwilowej i pulsacji chwilowej, uogólnienie to nie jest jednoznaczne. Pojęcie fazy chwilowej określa się względem ustalonej arbitralnie pulsacji ω_0.

Definicja 10.4. *Fazą chwilową* rzeczywistego sygnału $x(t)$, określoną względem pulsacji ω_0, nazywamy funkcję $\varphi(t)$, taką że:

$$\psi(t) = \omega_0 t + \varphi(t), \tag{10.10}$$

gdzie $\psi(t)$ jest argumentem sygnału analitycznego sygnału $x(t)$.

Dla różnych wartości pulsacji ω_0 faza chwilowa jest różną funkcją czasu. Wybór pulsacji ω_0 jest w zasadzie dowolny. W przypadku sygnałów wąskopasmowych jako pulsację ω_0 wybiera się jednak zwykle pulsację środkową ich widma prawostronnego. Obwiednia $X(t)$ i faza chwilowa $\varphi(t)$ sygnału wąskopasmowego $x(t)$ zmieniają się wówczas w czasie wolno w porównaniu z tym sygnałem.

Przykład 10.3. Faza chwilowa sygnału harmonicznego (10.5) określona względem jego pulsacji ω_0 jest stałą funkcją czasu $\varphi(t) = \varphi_0$.

10.1.4. Drganie uogólnione

Pojęcie sygnału analitycznego umożliwia wprowadzenie jeszcze jednej reprezentacji sygnałów nieharmonicznych użytecznej w analizie sygnałów wąskopasmowych. Zgodnie ze wzorami (10.3), (10.7) i (10.10), dla ustalonej pulsacji ω_0 sygnał $x(t)$ można przedstawić w postaci:

$$\begin{aligned} x(t) = \operatorname{Re}\, z_x(t) &= \operatorname{Re}\left[X(t)\,\mathrm{e}^{\mathrm{j}\psi(t)}\right] = \operatorname{Re}\left[X(t)\,\mathrm{e}^{\mathrm{j}[\omega_0 t+\varphi(t)]}\right] = \\ &= X(t)\cos[\omega_0 t + \varphi(t)]. \end{aligned} \tag{10.11}$$

Z postaci tej wynika, że jeśli pulsacja ω_0 jest pulsacją środkową widma wąskopasmowego sygnału $x(t)$, to sygnał ten można traktować jako oscylacyjny sygnał nieharmoniczny, którego pulsacja chwilowa oscyluje wokół pulsacji ω_0, zaś amplituda i faza zmieniają się w czasie zgodnie z funkcjami $X(t)$ i $\varphi(t)$. W podobny sposób można przedstawić transformatę Hilberta $\hat{x}(t)$:

$$\begin{aligned} \hat{x}(t) = \operatorname{Im}\, z_x(t) &= \operatorname{Im}\left[X(t)\,\mathrm{e}^{\mathrm{j}\psi(t)}\right] = \operatorname{Im}\left[X(t)\,\mathrm{e}^{\mathrm{j}[\omega_0 t+\varphi(t)]}\right] = \\ &= X(t)\sin[\omega_0 t + \varphi(t)]. \end{aligned} \tag{10.12}$$

Definicja 10.5. Przedstawienie sygnału $x(t)$ w postaci:

$$x(t) = X(t)\cos[\omega_0 t + \varphi(t)] \tag{10.13}$$

nosi nazwę *drgania uogólnionego*.

10.1.5. Składowa synfazowa i składowa kwadraturowa

Korzystając z tożsamości trygonometrycznych, wzory (10.11) i (10.12) można przekształcić do postaci:

$$\begin{aligned} x(t) &= X(t)\cos\varphi(t)\cos\omega_0 t - X(t)\sin\varphi(t)\sin\omega_0 t \triangleq \\ &\triangleq x_I(t)\cos\omega_0 t - x_Q(t)\sin\omega_0 t, \end{aligned} \tag{10.14}$$

$$\begin{aligned}\hat{x}(t) &= X(t)\cos\varphi(t)\sin\omega_0 t + X(t)\sin\varphi(t)\cos\omega_0 t \triangleq \\ &\triangleq x_I(t)\sin\omega_0 t + x_Q(t)\cos\omega_0 t,\end{aligned} \tag{10.15}$$

nazywanych *zapisem Rice'a*. Sygnały $x_I(t) = X(t)\cos\varphi(t)$ oraz $x_Q(t) = X(t)\sin\varphi(t)$ są nazywane *składową synfazową* (ang. *in-phase component*) i *składową kwadraturową* (ang. *quadrature component*) sygnału $x(t)$. Składowe te są sygnałami wolnozmiennymi w porównaniu z sygnałem $x(t)$. O sygnałach $x_I(t)\cos\omega_0 t$ oraz $-x_Q(t)\sin\omega_0 t$ (lub $x_I(t)\sin\omega_0 t$ oraz $x_Q(t)\cos\omega_0 t$) mówimy, że są względem siebie w *kwadraturze*.

10.1.6. Obwiednia zespolona

Sygnał $x(t)$ reprezentowany drganiem uogólnionym (10.12) można zapisać w jeszcze innej postaci:

$$\begin{aligned}x(t) = \operatorname{Re}\left[X(t)\,\mathrm{e}^{\mathrm{j}[\omega_0 t+\varphi(t)]}\right] &= \operatorname{Re}\left[X(t)\,\mathrm{e}^{\mathrm{j}\varphi(t)}\,\mathrm{e}^{\mathrm{j}\omega_0 t}\right] = \\ &= \operatorname{Re}\left[\boldsymbol{X}(t)\,\mathrm{e}^{\mathrm{j}\omega_0 t}\right].\end{aligned} \tag{10.16}$$

Definicja 10.6. Funkcję zespoloną

$$\boldsymbol{X}(t) = X(t)\,\mathrm{e}^{\mathrm{j}\varphi(t)} \tag{10.17}$$

nazywamy *obwiednią zespoloną* sygnału $x(t)$.

Przykład 10.4. W przypadku sygnału harmonicznego (10.5) reprezentacja (10.16) ma postać:

$$x(t) = \operatorname{Re}\left[X_0\,\mathrm{e}^{\mathrm{j}[\omega_0 t+\varphi_0]}\right] = \operatorname{Re}\left[X_0\,\mathrm{e}^{\mathrm{j}\varphi_0}\,\mathrm{e}^{\mathrm{j}\omega_0 t}\right] = \operatorname{Re}\left[\boldsymbol{X}\,\mathrm{e}^{\mathrm{j}\omega_0 t}\right], \tag{10.18}$$

gdzie liczba zespolona

$$\boldsymbol{X} = X_0\,\mathrm{e}^{\mathrm{j}\varphi_0} \tag{10.19}$$

jest amplitudą zespoloną tego sygnału. Dla sygnału harmonicznego pojęcie obwiedni zespolonej jest zatem tożsame z pojęciem jego amplitudy zespolonej.

Z porównania wzorów (10.16) i (10.17) ze wzorami (10.18) i (10.19) wynika, że pojęcie obwiedni zespolonej jest uogólnieniem pojęcia amplitudy zespolonej na sygnały nieharmoniczne. Z tego względu obwiednia zespolona $\boldsymbol{X}(t)$ jest nazywana również *chwilową amplitudą zespoloną* nieharmonicznego sygnału $x(t)$.

Zauważmy, że obwiednię zespoloną można przedstawić jako:

$$\boldsymbol{X}(t) = X(t)\,\mathrm{e}^{\mathrm{j}\varphi(t)} = X(t)\cos\varphi(t) + \mathrm{j}\,X(t)\sin\varphi(t) = x_I(t) + \mathrm{j}\,x_Q(t) \tag{10.20}$$

oraz że zwykła obwiednia sygnału jest modułem jego obwiedni zespolonej

$$X(t) = |\boldsymbol{X}(t)|, \tag{10.21}$$

przy czym:

$$X(t) = |\boldsymbol{X}(t)| = \sqrt{x^2(t) + \hat{x}^2(t)} = \sqrt{x_I^2(t) + x_Q^2(t)}. \tag{10.22}$$

Odnotujmy ponadto, że sygnał analityczny sygnału $x(t)$ można przedstawić względem ustalonej pulsacji ω_0 w postaci:

$$z_x(t) = \boldsymbol{X}(t)\,\mathrm{e}^{\mathrm{j}\omega_0 t} = [x_I(t) + \mathrm{j}\,x_Q(t)]\,\mathrm{e}^{\mathrm{j}\omega_0 t}\,. \tag{10.23}$$

10.1.7. Filtr Hilberta

Rozpatrzymy obecnie związek między widmem $X(\omega)$ sygnału $x(t)$ a widmem $\hat{X}(\omega)$ jego transformaty Hilberta $\hat{x}(t)$. Związek ten wykorzystamy w następnym punkcie do zbadania widma sygnału analitycznego.

Zauważmy, że całki (10.1) i (10.2) definiujące proste i odwrotne przekształcenie Hilberta można zapisać w postaci splotów:

$$\hat{x}(t) = \frac{1}{\pi t} * x(t), \tag{10.24}$$

$$x(t) = -\frac{1}{\pi t} * \hat{x}(t). \tag{10.25}$$

Z twierdzenia o splocie i pary transformat Fouriera (3.47) wynikają odpowiedniki związków (10.24) i (10.25) w dziedzinie częstotliwości:

$$\hat{X}(\omega) = (-\mathrm{j}\,\mathrm{sgn}\,\omega)X(\omega), \tag{10.26}$$

$$X(\omega) = (\mathrm{j}\,\mathrm{sgn}\,\omega)\hat{X}(\omega). \tag{10.27}$$

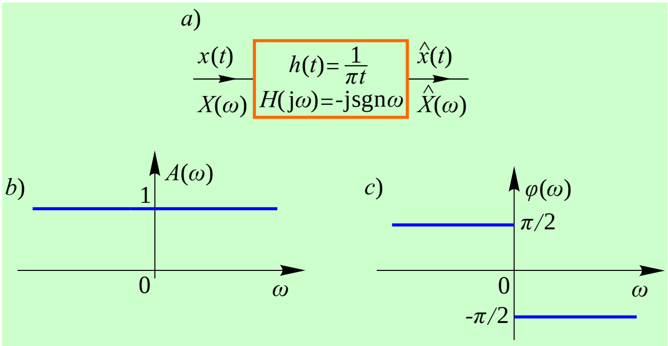
Z zależności (10.24) i (10.26) wynika, że transformatę Hilberta $\hat{x}(t)$ można traktować jako odpowiedź na pobudzenie sygnałem $x(t)$ filtru (rys. 10.1a) o odpowiedzi impulsowej

$$h(t) = \frac{1}{\pi t} \tag{10.28}$$

i charakterystyce amplitudowo-fazowej

$$H(\mathrm{j}\,\omega) = -\mathrm{j}\,\mathrm{sgn}\,\omega = \mathrm{e}^{\mathrm{j}[\pi/2 - \pi \boldsymbol{1}(\omega)]}\,. \tag{10.29}$$

Filtr ten jest nazywany *filtrem kwadraturowym* lub *filtrem Hilberta*. Jego charakterystyki amplitudowa i fazowa są przedstawione na rys. 10.1b,c. Filtr Hilberta jest



Rys. 10.1. Filtr Hilberta (a) oraz jego charakterystyki: amplitudowa (b) i fazowa (c)

filtrem wszechprzepustowym (por. p. 7.5.7), opóźniającym fazę każdej składowej sygnału o $\pi/2$. Jest to filtr nieprzyczynowy, a więc nierealizowalny fizycznie.

Przykład 10.5. Z wykresów charakterystyk filtru Hilberta (rys. 10.1b,c) wynika wprost, że transformatą Hilberta sygnału harmonicznego $x(t) = X_0 \cos(\omega_0 t + \varphi_0)$ jest sygnał $\hat{x}(t) = X_0 \sin(\omega_0 t + \varphi_0)$ (por. przykład 10.1).

Korzystając z pojęcia filtru Hilberta, wyznaczymy transformatę Hilberta bardziej ogólnego sygnału:

$$y(t) = x(t) \cos \omega_0 t, \tag{10.30}$$

gdzie $x(t)$ jest sygnałem dolnopasmowym o paśmie ograniczonym pulsacją $\omega_m \leqslant \omega_0$. Zgodnie ze wzorem (10.24) transformatę Hilberta sygnału (10.30) możemy zapisać w postaci:

$$\hat{y}(t) = \frac{1}{\pi t} * x(t) \cos \omega_0 t.$$

Stąd, a także z twierdzenia o modulacji (3.26), wynika, że widmo $\hat{Y}(\omega)$ sygnału $\hat{y}(t)$ jest określone wzorem:

$$\begin{aligned} \hat{Y}(\omega) &= \frac{1}{2} [X(\omega - \omega_0) + X(\omega + \omega_0)] (-\mathrm{j} \operatorname{sgn} \omega) = \\ &= \frac{1}{\mathrm{j}\,2} [X(\omega - \omega_0) + X(\omega + \omega_0)] (\operatorname{sgn} \omega). \end{aligned}$$

Ponieważ warunek $\omega_m \leqslant \omega_0$ zapewnia, że widma $X(\omega - \omega_0)$ oraz $X(\omega + \omega_0)$ są rozłączne, a ponadto pierwsze z nich przybiera wartości niezerowe jedynie dla pulsacji dodatnich, drugie zaś – dla pulsacji ujemnych, możemy dalej napisać:

$$\hat{Y}(\omega) = \frac{1}{\mathrm{j}\,2} [X(\omega - \omega_0) - X(\omega + \omega_0)] .$$

Zatem, zgodnie ze wzorem (3.27), otrzymujemy:

$$\hat{y}(t) = x(t) \sin \omega_0 t, \tag{10.31}$$

skąd wynika, że sygnał analityczny sygnału (10.30) ma postać:

$$z_y(t) = x(t)\,\mathrm{e}^{\mathrm{j}\omega_0 t}\,. \tag{10.32}$$

Zauważmy, że sygnały (10.30) i (10.31) są względem siebie w *kwadraturze* (por. p. 10.1.5).

Postępując podobnie można wykazać, że sygnałem analitycznym sygnału

$$y(t) = x(t)\cos(\omega_0 t + \varphi_0),$$

gdzie $x(t)$ jest sygnałem o paśmie ograniczonym pulsacją $\omega_m \leqslant \omega_0$, jest sygnał

$$z_y(t) = x(t)\,\mathrm{e}^{\mathrm{j}(\omega_0 t+\varphi_0)}\,. \tag{10.33}$$

10.1.8. Widmo sygnału analitycznego

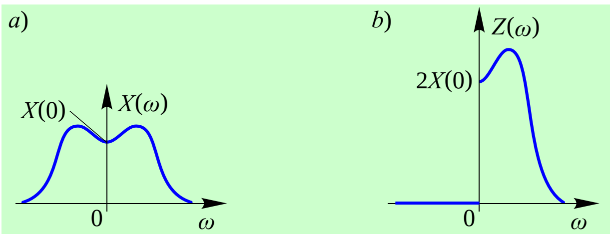
Ze wzoru (10.26) wynika, że widmo $Z_x(\omega)$ sygnału analitycznego $z_x(t) = x(t) + \mathrm{j}\,\hat{x}(t)$ ma postać:

$$Z_x(\omega) = X(\omega) + \mathrm{j}\,\hat{X}(\omega) = X(\omega) + \mathrm{j}\,(-\mathrm{j}\,\mathrm{sgn}\,\omega)X(\omega) = 2X(\omega)\boldsymbol{1}(\omega). \tag{10.34}$$

Widzimy więc, że widmo to jest równe zeru dla $\omega < 0$, natomiast dla $\omega > 0$ zachowuje kształt widma $X(\omega)$, przy czym jego gęstość w tym przedziale jest dwukrotnie większa (rys. 10.2). Prawostronny charakter widma sygnału analitycznego jest jego ważną cechą, ułatwiającą analizę widmową sygnałów zmodulowanych.

Znając widmo sygnału analitycznego, możemy wyznaczyć widmo sygnału oryginalnego zgodnie ze wzorem:

$$X(\omega) = \frac{1}{2}\left[Z_x(\omega) + Z^*(-\omega)\right]. \tag{10.35}$$



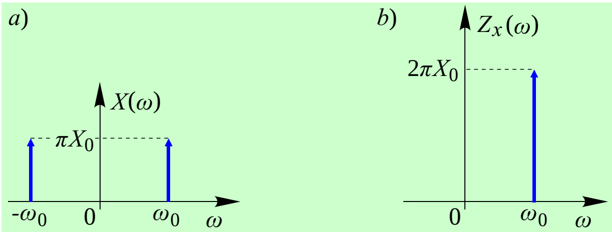
Rys. 10.2. Widmo sygnału (a) i widmo jego sygnału analitycznego (b)

Przykład 10.6. Widmo sygnału harmonicznego $x(t) = X_0\cos\omega_0 t$ ma postać (por. wzór (3.49)):

$$X(\omega) = \pi X_0[\delta(\omega-\omega_0) + \delta(\omega+\omega_0)],$$

zatem widmo jego sygnału analitycznego jest określone wzorem (rys. 10.3):

$$Z_x(\omega) = 2\pi X_0 \delta(\omega - \omega_0). \tag{10.36}$$



Rys. 10.3. Widmo sygnału harmonicznego (a) i widmo jego sygnału analitycznego (b)

10.1.9. Funkcja modulująca

Operowanie postacią analityczną sygnałów występujących w procesie modulacji umożliwia jednolity i zwarty zapis sygnałów zmodulowanych w analogowych systemach modulacji. Dla każdego analogowego systemu modulacji sygnał analityczny sygnału zmodulowanego $y(t)$ można zapisać jako iloczyn

$$z_y(t) = m(t) z_c(t) \tag{10.37}$$

pewnej funkcji $m(t)$, zwanej *funkcją modulującą,* i sygnału analitycznego $z_c(t)$ fali nośnej $c(t)$. Funkcja modulująca jest w ogólnym przypadku funkcją zespoloną $m(t) = |m(t)| \,\mathrm{e}^{\mathrm{j} \arg m(t)}$, zależną od sygnału modulującego $x(t)$, a jej postać określa rodzaj modulacji. Za pomocą tej funkcji można opisać wszystkie uzmiennienia amplitudy i kąta fali nośnej.

Przyjmijmy dla wygody i bez straty ogólności rozważań zerową wartość fazy początkowej φ_0 fali nośnej we wzorze (9.1). Mamy wówczas $z_c(t) = Y_0 \,\mathrm{e}^{\mathrm{j}\Omega t}$ i zgodnie ze wzorami (10.7) i (10.23) sygnał analityczny sygnału zmodulowanego można przedstawić w postaci:

$$z_y(t) = |z_y(t)| \,\mathrm{e}^{\mathrm{j}\psi(t)} = Y_0 m(t) \,\mathrm{e}^{\mathrm{j}\Omega t} = \boldsymbol{Y}(t) \,\mathrm{e}^{\mathrm{j}\Omega t}\,. \tag{10.38}$$

Obwiednia zespolona, obwiednia rzeczywista, kąt sygnału zmodulowanego, jego faza chwilowa (określona względem pulsacji Ω) oraz pulsacja (częstotliwość) chwilowa wyrażają się zatem wzorami:

$$\boldsymbol{Y}(t) = Y_0 m(t) = Y_0 |m(t)| \,\mathrm{e}^{\mathrm{j} \arg m(t)}, \tag{10.39}$$

$$|z_y(t)| = Y_0 |m(t)|, \tag{10.40}$$

$$\psi(t) = \Omega t + \arg m(t), \tag{10.41}$$

$$\varphi(t) = \arg m(t), \tag{10.42}$$

$$\omega(t) = \Omega + \frac{\mathrm{d}}{\mathrm{d}t} \arg m(t), \qquad f(t) = F + \frac{1}{2\pi} \frac{\mathrm{d}}{\mathrm{d}t} \arg m(t). \tag{10.43}$$

Funkcje modulujące dla podstawowych rodzajów modulacji analogowych przedstawiono w tablicy 10.1.

Tablica 10.1. Funkcje modulujące dla podstawowych modulacji analogowych

Rodzaje modulacji	**Funkcja modulująca**
Amplitudy	
AM-SC	$x(t)$
AM	$1 + kx(t)$
SSB-SC (wstęga górna)	$x(t) + \mathrm{j}\,\hat{x}(t)$
SSB-SC (wstęga dolna)	$x(t) - \mathrm{j}\,\hat{x}(t)$
SSB (wstęga górna)	$1 + x(t) + \mathrm{j}\,\hat{x}(t)$
SSB (wstęga dolna)	$1 + x(t) - \mathrm{j}\,\hat{x}(t)$
VSB	$x(t) + \mathrm{j}\,[x(t) * h_Q(t)]$
kąta	
PM	$\exp[\mathrm{j}\,k_p x(t)]$
FM	$\exp[\mathrm{j}\,k_f \int x(t)\,\mathrm{d}t]$
k, k_p, k_f – parametry modulacji	
$h_Q(t)$ – odpowiedź impulsowa filtru (por. p. 10.5)	

10.2. Modulacja AM-SC

10.2.1. Sygnał AM-SC

Najprostszym rodzajem modulacji jest modulacja dwuwstęgowa bez fali nośnej AM-SC. W przypadku tej modulacji $m(t) = x(t)$, tj. funkcja modulująca jest równa sygnałowi modulującemu. Reprezentacja analityczna (10.38) sygnału zmodulowanego oraz sygnał zmodulowany AM-SC są zatem określone wzorami:

$$z_{\text{AM-SC}}(t) = Y_0 x(t)\,\mathrm{e}^{\mathrm{j}\Omega t}, \tag{10.44}$$

$$y_{\text{AM-SC}}(t) = \operatorname{Re} z_{\text{AM-SC}}(t) = Y_0 x(t) \cos \Omega t. \tag{10.45}$$

Wynika stąd, że $Y(t) = Y_0|x(t)|$ oraz $\omega(t) = \Omega$, tzn. obwiednia sygnału AM-SC jest proporcjonalna do modułu sygnału modulującego, a jego pulsacja chwilowa jest stała i równa pulsacji fali nośnej. Sygnał modulujący $x(t)$ jest z założenia sygnałem dolnopasmowym o paśmie ograniczonym pewną pulsacją $\omega_m \ll \Omega$. Będziemy zakładać ponadto, że jest to sygnał o ograniczonej mocy.

Zgodnie z twierdzeniem o modulacji (3.26) widmo (w sensie granicznym) sygnału AM-SC ma postać:

$$Y_{\text{AM-SC}}(\omega) = \frac{Y_0}{2}\left[X(\omega - \Omega) + X(\omega + \Omega)\right], \tag{10.46}$$

gdzie $X(\omega)$ jest widmem sygnału modulującego. Sygnały modulujący i zmodulowany oraz ich widma są pokazane na rys. 10.4.

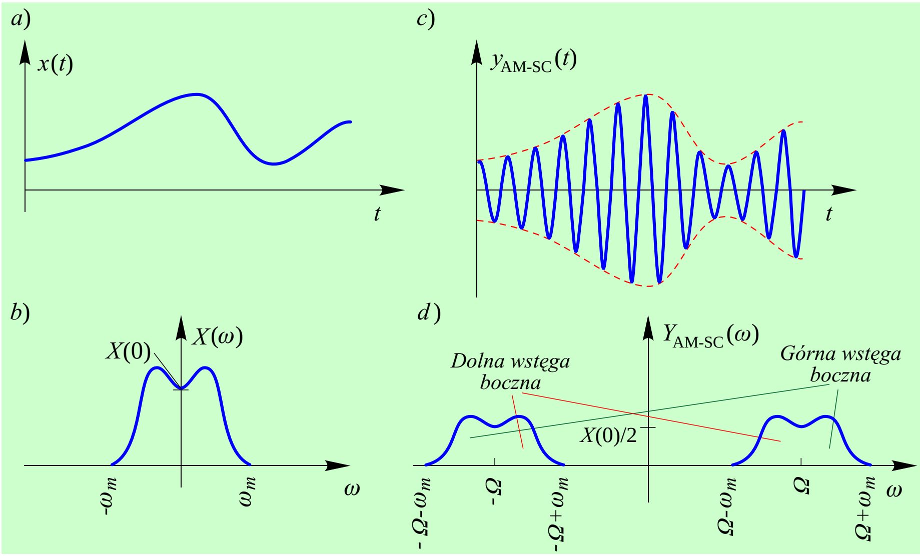
W wyniku modulacji widmo $X(\omega)$ zostaje rozszczepione na dwie części, przesunięte do punktów $\pm\Omega$ osi pulsacji i zachowujące kształt widma $X(\omega)$. Części widma sygnału zmodulowanego położone w przedziale pulsacji $|\omega| > \Omega$ noszą nazwę *górnej wstęgi bocznej*, a części widma położone w przedziale $|\omega| < \Omega$ – *dolnej wstęgi bocznej*. Górna i dolna wstęga boczna zostały zaznaczone na rys. 10.4.

Jeżeli przez P_x oznaczymy moc sygnału modulującego, to zgodnie ze wzorem definicyjnym (1.6) moc P_y sygnału AM-SC jest określona wzorem:

$$P_y = \frac{Y_0^2}{2} P_x. \tag{10.47}$$

10.2.2. Szerokość pasma sygnału AM-SC

Bardzo ważnym parametrem, stanowiącym jedno z kryteriów porównawczych różnych systemów modulacji, jest *szerokość pasma* sygnału zmodulowanego. Parametr ten decyduje o tym jak szerokie pasmo należy przeznaczyć w kanale



Rys. 10.4. Sygnał modulujący (a), jego widmo (b) oraz sygnał zmodulowany AM-SC (c) i jego widmo (d)

telekomunikacyjnym na transmisję sygnału zmodulowanego, a w konsekwencji określa liczbę sygnałów jakie w danym systemie z podziałem częstotliwościowym można przesłać jednocześnie przez kanał w zadanym z góry paśmie.

Z wykresu widma sygnału AM-SC (rys. 10.4) wynika, że szerokość pasma tego sygnału (wyrażona w hercach) jest określona wzorem:

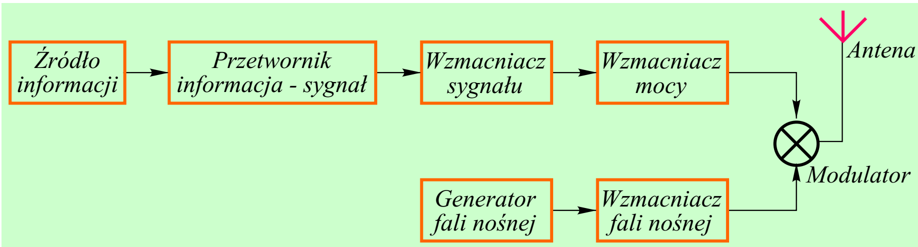
$$B_{\text{AM-SC}} = 2f_m, \tag{10.48}$$

gdzie f_m jest maksymalną częstotliwością widma sygnału modulującego $x(t)$.

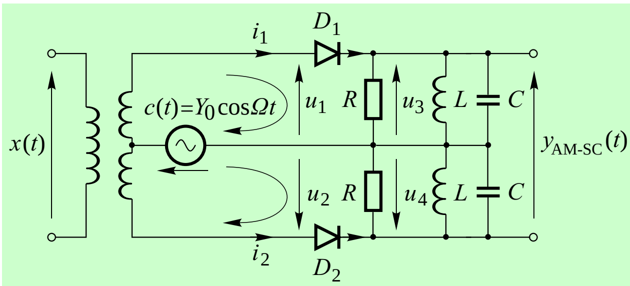
10.2.3. Generacja sygnału AM-SC

Schemat blokowy nadajnika AM-SC jest przedstawiony na rys. 10.5. Aby wytworzyć sygnał AM-SC, sygnał modulujący $x(t)$ musi być przemnożony w nadajniku przez falę nośną $c(t) = Y_0 \cos \Omega t$. Operację tę wykonuje układ nazywany *modulatorem iloczynowym*.

Modulator iloczynowy można zrealizować za pomocą różnych układów. Ponieważ operacja mnożenia jest operacją nieliniową, w układach tych muszą wystąpić elementy nieliniowe. Na rys. 10.6 pokazano schemat tzw. *modulatora zrównoważonego* zawierającego dwie diody półprzewodnikowe. Układ ten składa się z dwóch identycznych obwodów połączonych przeciwsobnie, zawierających, oprócz diod, równoległe obwody rezonansowe RLC. Do obu obwodów są doprowadzone sygnały napięciowe: modulujący oraz fali nośnej, w sposób pokazany na schemacie.



Rys. 10.5. Schemat blokowy nadajnika AM-SC



Rys. 10.6. Schemat modulatora zrównoważonego

Pokażemy, że układ modulatora zrównoważonego z rys. 10.6 istotnie realizuje operację mnożenia. W celu ułatwienia analizy pominiemy na razie elementy LC. Z napięciowych równań Kirchhoffa dla obwodów zaznaczonych na schemacie obliczamy napięcia $u_1(t)$ i $u_2(t)$ na szeregowych połączeniach diod i oporów R:

$$u_1(t) = c(t) + x(t), \qquad u_2(t) = c(t) - x(t).$$

Przyjmiemy założenie, że charakterystyka diody jest tak dobrana, iż zależność między prądem a napięciem w gałęzi utworzonej przez szeregowe połączenie diody i oporu jest opisana równaniem $i = au^2 + bu$, gdy $u \geqslant 0$ (zakres przewodzenia diody) oraz równaniem $i = 0$, gdy $u < 0$ (zakres zaporowy). Przy tym założeniu w przedziałach czasu, w których przewodzi dioda D_1 w górnym obwodzie, na oporze R tego obwodu występuje napięcie

$$u_3(t) = Ri_1(t) = Ra[c^2(t) + 2x(t)c(t) + x^2(t)] + Rb[c(t) + x(t)],$$

natomiast w przedziałach czasu, w których dioda D_1 znajduje się w stanie zaporowym, napięcie $u_3(t)$ jest równe zeru. Sytuacja jest dokładnie odwrotna w przypadku dolnego obwodu. W przedziałach czasu, w których dioda D_1 jest w stanie zaporowym, dioda D_2 znajduje się w stanie przewodzenia i na oporze R dolnego obwodu występuje napięcie

$$u_4(t) = Ri_2(t) = Ra[c^2(t) - 2x(t)c(t) + x^2(t)] + Rb[c(t) - x(t)],$$

natomiast gdy dioda D_1 przewodzi, dioda D_2 znajduje się w stanie zaporowym i napięcie $u_4(t)$ jest równe zeru. Na zaciskach wyjściowych układu występuje zatem sygnał napięcia

$$u(t) = u_3(t) - u_4(t) = 4Rax(t)c(t) + 2Rbx(t).$$

Sygnał ten zawiera składową iloczynową o widmie skupionym wokół pulsacji Ω fali nośnej i składową niskoczęstotliwościową proporcjonalną do sygnału modulującego. Zadaniem obwodów rezonansowych LC jest odfiltrowanie sygnału o niskiej częstotliwości. W tym celu wystarczy dostroić pulsację rezonansową tych obwodów do pulsacji nośnej Ω i odpowiednio dobrać ich dobroć. W efekcie na wyjściu układu otrzymujemy sygnał zmodulowany $y_{\text{AM-SC}}(t)$ proporcjonalny do iloczynu sygnału modulującego i fali nośnej.

10.2.4. Demodulacja sygnału AM-SC. Detektor koherentny

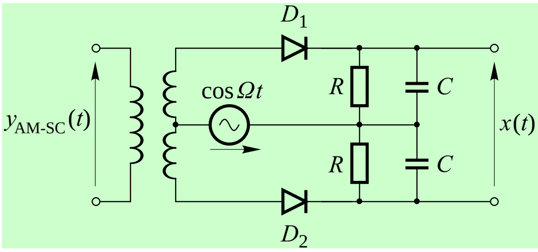
Operacja demodulacji polega na odtworzeniu po stronie odbiorczej sygnału modulującego z sygnału zmodulowanego. Aby zdemodulować sygnał AM-SC wystarczy pomnożyć go przez sygnał $\cos \Omega t$, o pulsacji równej pulsacji nośnej Ω, wytworzony przez lokalny generator umieszczony w odbiorniku. Przy założeniu, że faza początkowa tego sygnału jest zsynchronizowana z fazą fali nośnej generowanej w nadajniku, na wyjściu układu mnożącego otrzymujemy sygnał

$$y_{\text{AM-SC}}(t) \cos \Omega t = Y_0 x(t) \cos^2 \Omega t = \frac{Y_0}{2} x(t) + \frac{Y_0}{2} x(t) \cos 2\Omega t. \tag{10.49}$$

Sygnał ten zawiera składową dolnopasmową proporcjonalną do sygnału informacyjnego i składową wąskopasmową o widmie skupionym wokół dużej pulsacji 2Ω. Demodulator sygnału może być zatem zrealizowany w układzie przedstawionym na rys. 10.7, którego lewa część, mająca za zadanie realizację operacji mnożenia sygnałów, jest identyczna jak w układzie modulatora sygnału AM-SC, natomiast filtry dolnoprzepustowe RC służą do odfiltrowania składowej o dużej częstotliwości. Układ demodulatora z rys. 10.7 jest nazywany *detektorem koherentnym* lub *detektorem synchronicznym*.

10.2.5. Błędy częstotliwości i fazy

Aby detektor koherentny sygnału AM-SC działał prawidłowo, musi być zapewniona wysoka stabilność pulsacji lokalnego generatora fali nośnej w odbiorniku oraz równość faz fali nośnej w odbiorniku i nadajniku. Równość tych faz jest warunkiem tzw. *odbioru koherentnego* sygnału. Rozpatrzmy zniekształcenia



Rys. 10.7. Detektor koherentny sygnału AM-SC

sygnału zdemodulowanego, jakie powstają w przypadku, gdy warunki te nie są spełnione. Załóżmy, że generator lokalny generuje sygnał $\cos[(\Omega + \Delta\Omega)t + \Delta\varphi]$, gdzie $\Delta\Omega$ jest błędem pulsacji, a $\Delta\varphi$ błędem fazy. W wyniku mnożenia sygnałów w detektorze koherentnym otrzymujemy wówczas sygnał

$$y_{\text{AM-SC}}(t)\cos[(\Omega + \Delta\Omega)t + \Delta\varphi] = \frac{Y_0}{2}x(t)\cos(\Delta\Omega t + \Delta\varphi) + \\ + \frac{Y_0}{2}x(t)\cos[(2\Omega + \Delta\Omega)t + \Delta\varphi],$$

a po odfiltrowaniu drugiej składowej – sygnał zdemodulowany:

$$y_d(t) = \frac{Y_0}{2}x(t)\cos(\Delta\Omega t + \Delta\varphi).$$

Na wyjściu detektora otrzymujemy zatem sygnał proporcjonalny do sygnału informacyjnego $x(t)$, ale zniekształconego multiplikatywnie sygnałem $\cos(\Delta\Omega t+\Delta\varphi]$. Jeśli $\Delta\Omega = 0$, tzn. błąd pulsacji nie występuje, sygnał wyjściowy ma postać $(Y_0/2)x(t)\cos\Delta\varphi$. W przypadku, gdy błąd fazy $\Delta\varphi$ jest stały i jego wartość jest mała, sygnał odbierany jest niezniekształcony, a jedynie następuje jego tłumienie w stosunku $\cos\Delta\varphi$. W rzeczywistości, na skutek zmieniających się warunków propagacji, błąd fazy zmienia się losowo w czasie, powodując niepożądany efekt zniekształcający. Nie jest on jednak tak krytyczny dla jakości odbieranego sygnału, jak zniekształcenia wywołane błędem pulsacji. Przy zerowym błędzie fazy błąd pulsacji powoduje, iż na wyjściu demodulatora zamiast sygnału niezniekształconego $(Y_0/2)x(t)$ występuje sygnał zmodulowany $(Y_0/2)x(t)\cos\Delta\Omega t$ o pulsacji nośnej równej błędowi pulsacji $\Delta\Omega$. W wyniku nawet niewielkiego procentowo odstrojenia dużej pulsacji nośnej powstaje błąd pulsacji współmierny z pulsacjami pasma sygnału $x(t)$, w efekcie czego sygnał odbierany jest silnie zniekształcany. Należy jednak podkreślić, że w praktyce łatwiej jest zapewnić synchronizację pulsacji fal nośnych nadajnika i odbiornika, niż synchronizację ich faz.

Wrażliwość systemu AM-SC na błędy częstotliwości i fazy jest jego istotną wadą. W celu wyeliminowania zniekształceń spowodowanych tymi błędami, w odbiorniku systemu AM-SC są stosowane złożone układy synchronizujące i kompensacyjne. Jednym z możliwych rozwiązań jest zastosowanie *detektora Costasa*.

Detektor ten składa się z dwóch detektorów koherentnych zawierających generatory fal nośnych pozostających względem siebie w kwadraturze, tzn. przesuniętych o $\pi/2$, i wspólnym wejściowym sygnale zmodulowanym. Pulsacja i faza lokalnego generatora są automatycznie korygowane w układzie pętli ujemnego sprzężenia zwrotnego ([1], p. 3.4).

Innym rozwiązaniem problemu synchronizacji pulsacji i fazy jest przesyłanie wraz z sygnałem AM-SC dodatkowego *sygnału pilotującego* fali nośnej o niewielkim poziomie w stosunku do sygnału AM-SC. Sygnał ten jest w odbiorniku odfiltrowany i po odpowiednim wzmocnieniu wykorzystany do synchronizacji pulsacji i fazy lokalnej fali nośnej generowanej w odbiorniku. Wymienione rozwiązania komplikują jednak odbiornik i podrażają jego koszty, co w systemach powszechnego zastosowania o wielu punktach odbiorczych, takich jak telefoniczne, radiowe czy telewizyjne, wyklucza stosowanie modulacji AM-SC.

10.3. Modulacja AM

10.3.1. Sygnał AM

W przypadku modulacji AM funkcja modulująca ma postać $m(t) = 1+kx(t)$, gdzie współczynnik $k > 0$ jest parametrem modulacji. Reprezentacja analityczna sygnału AM i jego postać rzeczywista są zatem określone wzorami:

$$z_{\mathrm{AM}}(t) = Y_0[1 + kx(t)]\,\mathrm{e}^{\mathrm{j}\Omega t}, \tag{10.50}$$

$$y_{\mathrm{AM}}(t) = Y_0[1 + kx(t)]\cos\Omega t = Y_0\cos\Omega t + kY_0 x(t)\cos\Omega t. \tag{10.51}$$

Wynika stąd, że w przypadku modulacji AM obwiednia sygnału $Y(t) = Y_0|1 + kx(t)|$, a jego pulsacja chwilowa $\omega(t) = \Omega$.

Tak więc, w systemie AM obok składowej iloczynowej AM-SC $kY_0x(t)\cos\Omega t$ jest przesyłany dodatkowo sygnał fali nośnej $Y_0\cos\Omega t$. Jak pokażemy dalej, obecność fali nośnej w sygnale zmodulowanym umożliwia zastosowanie do demodulacji sygnału AM prostego układu demodulatora, nie wymagającego generowania sygnału fali nośnej w odbiorniku. Unika się tym samym trudności związanych z zapewnieniem jej odpowiedniej synchronizacji.

Widmo sygnału AM ma postać:

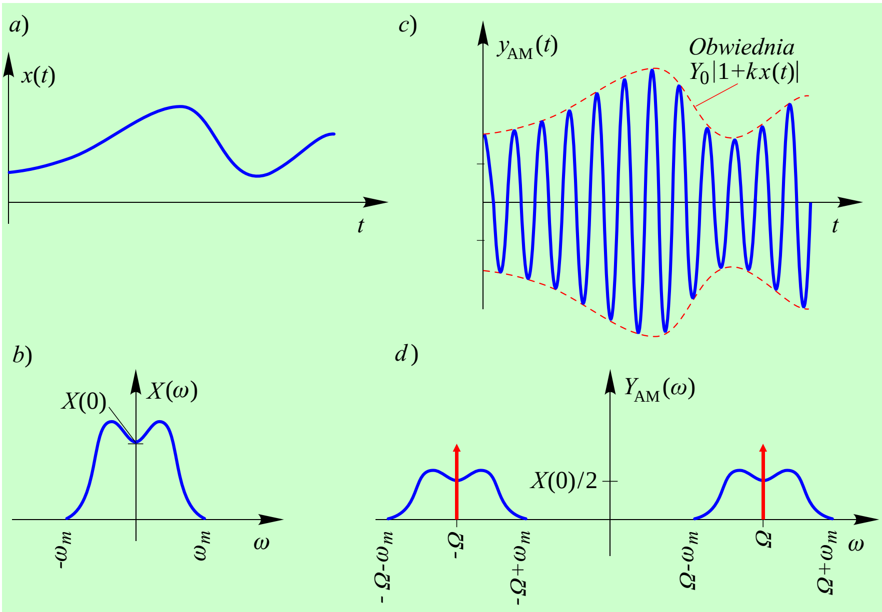
$$Y_{\mathrm{AM}}(\omega) = \pi Y_0[\delta(\omega-\Omega) + \delta(\omega+\Omega)] + \frac{kY_0}{2}[X(\omega-\Omega) + X(\omega+\Omega)]. \tag{10.52}$$

W porównaniu z widmem sygnału AM-SC zawiera ono dodatkowo parę dystrybucji Diraca w punktach $\pm\Omega$. Wykresy sygnałów modulującego i zmodulowanego AM oraz ich widm są przedstawione na rys. 10.8.

Moc sygnału AM jest określona wzorem

$$P_y = \frac{1}{2}Y_0^2 + \frac{1}{2}k^2 Y_0^2 P_x, \tag{10.53}$$

gdzie P_x jest mocą sygnału modulującego. Pierwszy składnik jest mocą składowej nośnej, drugi zaś – mocą zawartą we wstęgach bocznych.



Rys. 10.8. Sygnał modulujący (a), jego widmo (b) oraz sygnał zmodulowany AM (c) i jego widmo (d)

10.3.2. Szerokość pasma sygnału AM

Szerokość pasma sygnału AM jest identyczna jak sygnału AM-SC i określona wzorem:

$$B_{\mathrm{AM}} = 2f_m, \tag{10.54}$$

gdzie f_m jest maksymalną częstotliwością widma sygnału modulującego $x(t)$.

10.3.3. Współczynnik głębokości modulacji

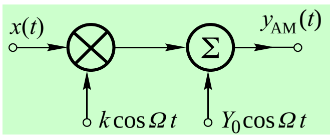
Ważnym parametrem charakteryzującym właściwości sygnału AM jest *współczynnik głębokości modulacji* określony wzorem:

$$m = \frac{\max m(t) - \min m(t)}{\max m(t) + \min m(t)}. \tag{10.55}$$

Wartość tego współczynnika dobiera się tak, aby zapewnić odpowiedni stosunek mocy składowej nośnej do mocy składowej AM-SC sygnału AM oraz aby uniknąć zjawiska przemodulowania sygnału (por. p. 10.3.5). Do interpretacji współczynnika głębokości modulacji powrócimy jeszcze w p. 10.3.6.

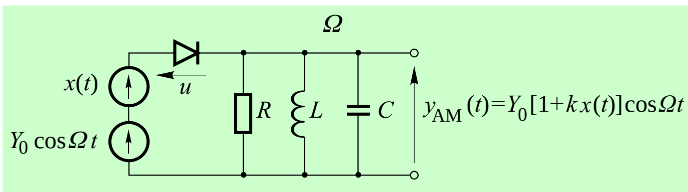
10.3.4. Generacja sygnału AM

Schemat blokowy modulatora AM jest przedstawiony na rys. 10.9. Składowa AM-SC sygnału AM może być wytworzona za pomocą modulatora iloczynowego, np. modulatora zrównoważonego z rys. 10.6. Składowa ta jest następnie dodawana w sumatorze do sygnału fali nośnej.



Rys. 10.9. Schemat blokowy modulatora AM

Innym rozwiązaniem układowym jest *modulator prostownikowy* pokazany na rys. 10.10. W układzie tym suma napięciowych sygnałów modulującego i fali nośnej jest doprowadzona do elementu nieliniowego o charakterystyce $i = au^2 + bu$, utworzonego przez połączenie szeregowe diody D i oporu R. Na oporze R pojawia się wyprostowany sygnał napięcia zawierający m.in. składową AM. Obwód rezonansowy LC dostrojony do pulsacji nośnej Ω odfiltrowuje pozostałe składowe i w rezultacie na wyjściu układu otrzymujemy sygnał AM.

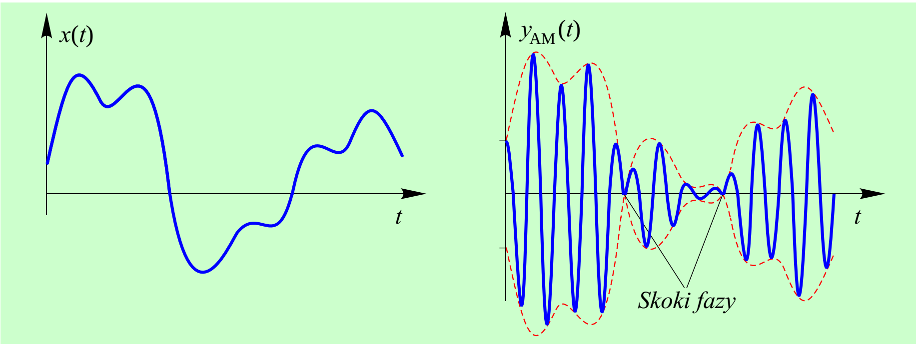


Rys. 10.10. Modulator prostownikowy AM

10.3.5. Demodulacja sygnału AM. Detektor obwiedni

Podobnie jak w przypadku sygnału AM-SC, demodulację sygnału AM można zrealizować za pomocą detektora koherentnego (rys. 10.7). Rozwiązanie to ma jednak wspomnianą wcześniej wadę związaną z koniecznością stosowania lokalnego generatora fali nośnej w odbiorniku i jego dokładnej synchronizacji.

Obecność fali nośnej w sygnale AM umożliwia jego demodulację za pomocą znacznie prostszego układu nazywanego *detektorem obwiedni*. Aby układ ten odtwarzał niezniekształcony sygnał informacyjny, wartość współczynnika k musi być tak dobrana, aby $1 + kx(t) \geqslant 0$ dla każdego t. Przy spełnieniu tego warunku kształt obwiedni sygnału AM jest identyczny jak sygnału modulującego i nie występuje *przemodulowanie* sygnału. Zjawisko przemodulowania jest zilustrowane na rys. 10.11 i objawia się skokowymi zmianami fazy sygnału AM w punktach przejścia jego obwiedni przez zera. Przemodulowanie nie powoduje zniekształceń sygnału zdemodulowanego, gdy stosowany jest detektor koherentny. Natomiast w przypadku stosowania detektora obwiedni jest ono przyczyną silnych zniekształceń.

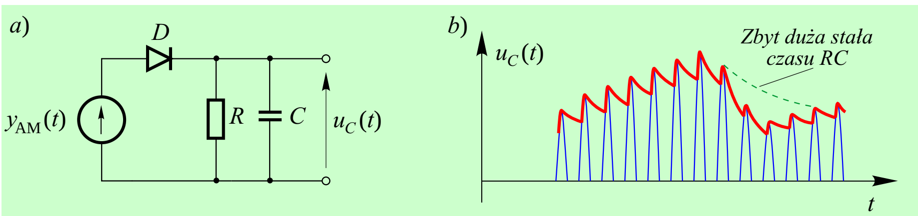


Rys. 10.11. Sygnał modulujący (a) i przemodulowany sygnał AM (b)

Detektor obwiedni (rys. 10.12a) jest prostym układem nieliniowym zawierającym diodę i dwójnik równoległy RC. Jego zasadę działania można wyjaśnić następująco. Załóżmy, że przed podaniem na wejście sygnału napięciowego AM układ był w stanie spoczynku. Pierwszy dodatni impuls fali zmodulowanej ładuje kondensator C do wartości szczytowej napięcia (rys. 10.12b). Proces ładowania trwa, dopóki wartość chwilowa szybko malejącego tylnego zbocza impulsu opadnie poniżej wartości napięcia na kondensatorze, które maleje znacznie wolniej ze względu na dostatecznie dużą wartość stałej czasowej RC obwodu. Następuje wówczas odcinek czasu, w którym dioda jest spolaryzowana zaporowo, a kondensator rozładowuje się w odizolowanym obwodzie RC. Napięcie na kondensatorze maleje wykładniczo z szybkością określoną przez stałą czasową RC do chwili, gdy wartość chwilowa następnego dodatniego impulsu fali zmodulowanej przekroczy wartość tego napięcia. Dioda zaczyna wówczas ponownie przewodzić i następuje krótkotrwałe doładowanie kondensatora do kolejnej wartości szczytowej, po czym proces powtarza się. Tym samym układ detektora śledzi obwiednię sygnału AM i odtwarza ją na swoim wyjściu. Sygnał wyjściowy różni się jeszcze od sygnału modulującego składową stałą. Wyeliminowanie tej składowej nie przedstawia już trudności.

Z przebiegu napięcia $u_C(t)$ na wyjściu detektora obwiedni przedstawionego na rys. 10.12b wynika, że odtworzona obwiednia sygnału wykazuje niepożądane tętnienia o częstotliwościach zbliżonych do częstotliwości fali nośnej. Tętnienia te można wyeliminować, stosując dodatkowy filtr dolnoprzepustowy. Należy podkreślić, że w praktyce są one dużo mniejsze, niż wynikałoby to z przedstawionego wykresu. Na wykresie tym częstotliwość oscylacji fali zmodulowanej została bowiem celowo zmniejszona w stosunku do faktycznej częstotliwości fali nośnej, tak aby można było w sposób graficznie przejrzysty zilustrować zasadę działania detektora obwiedni.

Wartość stałej czasowej RC powinna być z jednej strony dostatecznie duża, tak aby napięcie na kondensatorze było podtrzymywane i nie występowały duże tętnienia, z drugiej jednak strony dobór zbyt dużej stałej czasowej może spowodować, że proces rozładowywania się kondensatora nie będzie nadążał za zmianami obwiedni. Sytuacja, w której dobrana została zbyt duża wartość stałej czasowej została zilustrowana na rys. 10.12b.

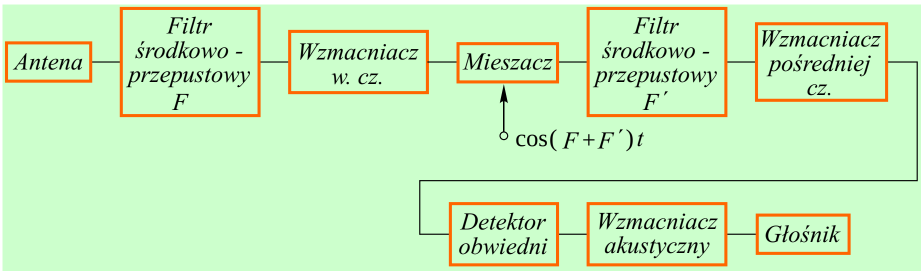


Rys. 10.12. Detektor obwiedni (a) i sygnał na jego wyjściu (b)

10.3.6. Odbiór superheterodynowy

W praktycznych systemach AM demodulację sygnału realizuje się z wykorzystaniem pośredniego stopnia *przemiany częstotliwości* zwanego *mieszaczem*. Odbiorniki pracujące w systemach z pośrednią przemianą częstotliwości noszą nazwę *odbiorników superheterodynowych*. Schemat blokowy odbiornika superheterodynowego jest przedstawiony na rys. 10.13.

Zadaniem mieszacza jest przesunięcie widma odebranego sygnału AM, skupionego wokół częstotliwości nośnej F, w okolice częstotliwości pośredniej F', stałej dla danego odbiornika (najczęściej równej 455 kHz). W tym celu w odbiorniku jest generowany sygnał harmoniczny o częstotliwości sumacyjnej $F'' = F + F'$ (lub rzadziej różnicowej $F'' = F - F'$), który jest mnożony w układzie mieszacza przez odebrany sygnał AM przed jego demodulacją. W efekcie widmo sygnału AM (wystarczy rozważyć tylko jego część prawostronną) zostaje rozszczepione bez zmiany kształtu na dwie części, przesunięte do częstotliwości $F'' - F = (F + F') - F = F'$ oraz $F'' + F = (F + F') + F = 2F + F'$. Druga z tych części, skupiona wokół częstotliwości $2F + F'$, jest odfiltrowywana za pomocą filtru środkowoprzepustowego dostrojonego do częstotliwości pośredniej F'. W efekcie na wyjściu filtru środkowoprzepustowego otrzymujemy sygnał AM, którego częstotliwością nośną jest częstotliwość pośrednia F'. Sygnał ten, po odpowiednim wzmocnieniu, jest następnie poddany właściwej demodulacji w układzie detektora obwiedni.



Rys. 10.13. Schemat blokowy odbiornika superheterodynowego

Zaleta odbioru superheterodynowego polega na tym, że wszelkie operacje na sygnałach występujących po układzie mieszacza są dokonywane na stałej częstotliwości pośredniej F', niezależnie od częstotliwości nośnej F sygnału odbieranego. Nie wymaga to kłopotliwego przestrajania wzmacniaczy występujących po układzie mieszacza, a jedynie dostrojenia wejściowego filtru środkowoprzepustowego do częstotliwości nośnej F odbieranego sygnału i dostrojenia częstotliwości nośnej lokalnego generatora do częstotliwości sumacyjnej $F + F'$. Dostrojenie obu tych częstotliwości jest w praktyce realizowane jednocześnie.

Należy podkreślić, że odbiór superheterodynowy można stosować także w innych systemach modulacji, zarówno amplitudowej, jak i kątowej. W systemach radiowych FM przesyłających sygnały w paśmie częstotliwościach 88–108 MHz jako częstotliwość pośrednią przyjmuje się 10,7 MHz.

10.3.7. Przypadek modulacji AM jednym tonem

Często dostateczny pogląd o istotnych cechach systemu modulacji analogowej daje przeanalizowanie najprostszego przypadku modulacji, jakim jest modulacja pojedynczym sygnałem harmonicznym. Mówimy wówczas, że fala nośna jest modulowana *jednym tonem*.

Sygnał AM zmodulowany jednym tonem $x(t) = X_0 \cos \omega_0 t$ o pulsacji $\omega_0 \ll \Omega$ ma postać:

$$y_{\mathrm{AM}}(t) = Y_0(1 + kX_0 \cos \omega_0 t) \cos \Omega t. \tag{10.56}$$

Ponieważ, jak można łatwo pokazać na podstawie wzoru ogólnego (10.55), współczynnik kX_0 (z założenia dodatni) jest równy współczynnikowi głębokości modulacji m, wzór (10.56) możemy zapisać w postaci:

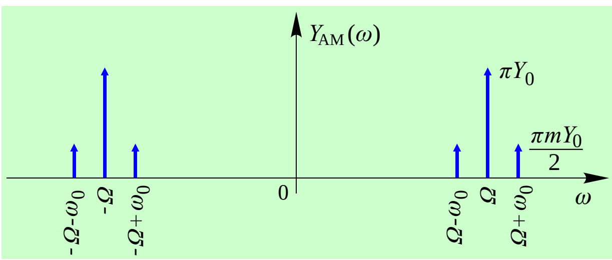
$$\begin{aligned} y_{\mathrm{AM}}(t) &= Y_0(1 + m \cos \omega_0 t) \cos \Omega t = \\ &= Y_0 \cos \Omega t + \frac{mY_0}{2} \left[\cos(\Omega + \omega_0)t + \cos(\Omega - \omega_0)t \right]. \end{aligned} \tag{10.57}$$

Sygnał AM zmodulowany jednym tonem jest zatem sumą trzech składowych harmonicznych, tj. składowej nośnej o pulsacji Ω oraz składowych bocznych o pulsacjach sumacyjnej $\Omega + \omega_0$ i różnicowej $\Omega - \omega_0$. Amplitudy obu składowych bocznych są jednakowe i $(m/2)$-raza mniejsze od amplitudy składowej nośnej.

Widmo sygnału AM zmodulowanego jednym tonem ma postać:

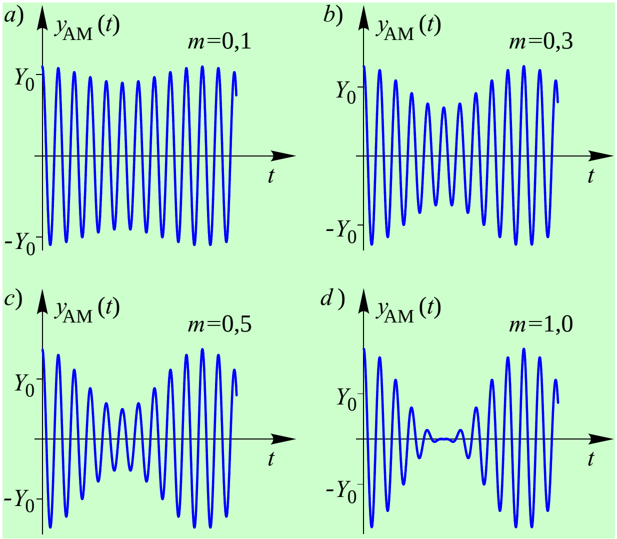
$$\begin{aligned} Y_{\mathrm{AM}}(\omega) = \; & \pi Y_0[\delta(\omega - \Omega) + \delta(\omega + \Omega)] + \\ & + \frac{\pi m Y_0}{2} \big[\delta(\omega - \Omega - \omega_0) + \delta(\omega + \Omega + \omega_0) + \delta(\omega - \Omega + \omega_0) + \\ & + \delta(\omega + \Omega - \omega_0) \big]. \end{aligned} \tag{10.58}$$

Widmo to jest pokazane na rys. 10.14. Składa się ono z trzech par dystrybucji występujących w punktach: $\pm\Omega$ (prążki nośne), $\pm(\Omega + \omega_0)$ (prawe prążki boczne) oraz $\pm(\Omega - \omega_0)$ (lewe prążki boczne). Część prawostronną widma stanowi zespół trzech charakterystycznych prążków: głównego w punkcie Ω oraz bocznych o jednakowych wysokościach w punktach $\Omega \pm \omega_0$. Szerokość pasma sygnału zmodulowanego jednym tonem wyrażona w hercach jest równa $2f_0$, gdzie f_0 jest częstotliwością sygnału modulującego.



Rys. 10.14. Widmo sygnału AM zmodulowanego jednym tonem

Przebiegi sygnału AM zmodulowanego jednym tonem dla różnych współczynników głębokości modulacji m są przedstawione na rys. 10.15. Na rys. 10.15d pokazano przypadek pełnej modulacji ($m = 1$).



Rys. 10.15. Sygnały AM zmodulowane jednym tonem dla różnych wartości współczynnika głębokości modulacji m

10.3.8. Sprawność energetyczna systemu AM

W celu przesłania informacji zawartej w sygnale zmodulowanym wystarczy transmitować jedynie wstęgi boczne. Transmisja fali nośnej w sygnale AM jest zabiegiem wyłącznie technicznym, podyktowanym prostotą demodulacji tego sygnału za pomocą detektora obwiedni. Można zatem powiedzieć, że moc potrzebna do wysłania fali nośnej jest z punktu widzenia transmisji informacji mocą bezużyteczną. Z praktycznego punktu widzenia jest więc uzasadnione, aby była ona jak najmniejsza.

W celu rozpatrzenia tego zagadnienia zdefiniujemy *współczynnik sprawności energetycznej* systemu AM określony jako stosunek

$$\varkappa = \frac{P_b}{P_y} 100\% \tag{10.59}$$

mocy P_b wstęg bocznych do całkowitej mocy P_y sygnału AM. Ponieważ moc P_y jest sumą mocy wstęg bocznych i mocy fali nośnej, współczynnik sprawności energetycznej systemu AM możemy wyrazić w postaci (por. wzór (10.53)):

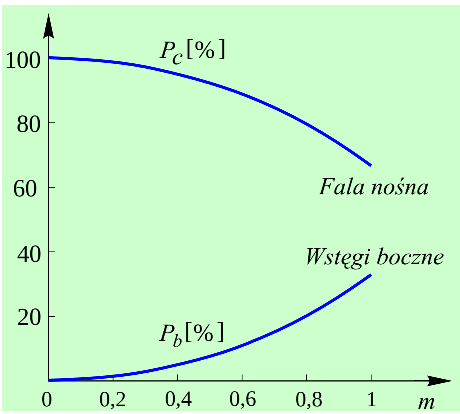
$$\varkappa = \frac{\frac{1}{2}k^2 Y_0^2 P_x}{\frac{1}{2}Y_0^2 + \frac{1}{2}k^2 Y_0^2 P_x} = \frac{k^2 P_x}{1 + k^2 P_x} 100\%, \tag{10.60}$$

gdzie P_x jest mocą sygnału modulującego. W przypadku modulacji jednym tonem $P_x = (X_0^2)/2$, a więc uwzględniając, że $kX_0 = m$, współczynnik sprawności energetycznej systemu AM można wyrazić przez współczynnik głębokości modulacji m:

$$\varkappa = \frac{\frac{1}{2}m^2}{1 + \frac{1}{2}m^2} = \frac{m^2}{2 + m^2}100\%. \tag{10.61}$$

Ze wzoru (10.61) wynika, że sprawność energetyczna systemu AM rośnie monotonicznie ze wzrostem współczynnika m. Ponieważ warunek nieprzemodulowania ogranicza wartości m do jedności, zatem współczynnik sprawności energetycznej osiąga największą wartość $\varkappa_{\max} = 33\%$ dla $m = 1$, tj. dla pełnej modulacji. Oznacza to, że co najmniej $2/3$ emitowanej mocy nadajnika sygnału AM przypada na falę nośną. W praktyce wartości współczynnika m wybiera się z przedziału $0{,}25 \div 0{,}4$. Sprawność energetyczna systemów AM wynosi więc zaledwie kilka procent.

Wykresy procentowego udziału mocy fali nośnej i mocy wstęg bocznych w całkowitej mocy sygnału AM zmodulowanego jednym tonem w funkcji współczynnika m przedstawiono na rys. 10.16. Mimo że wykresy te zostały sporządzone dla szczególnego przypadku modulacji jednym tonem, dają one dostateczny pogląd na problem sprawności energetycznej systemów AM. Niska sprawność energetyczna tych systemów jest ich istotną wadą, a zarazem ceną, jaką należy zapłacić za prostotę operacji demodulacji. Zauważmy, że sprawność energetyczna systemu AM-SC, rozumiana w sensie definicji (10.59), wynosi 100%. Mimo niskiej sprawności energetycznej systemy AM są – dzięki prostej realizacji układu odbiorczego – stosowane powszechnie w radiofonii w zakresie częstotliwości nośnych do 30 MHz.



Rys. 10.16. Procentowy udział mocy fali nośnej i mocy wstęg bocznych w całkowitej mocy sygnału AM zmodulowanego jednym tonem w funkcji współczynnika głębokości modulacji m

10.4. Modulacje jednowstęgowe SSB-SC i SSB

10.4.1. Sygnał SSB-SC

Gdyby sygnał informacyjny $x(t)$ był przesyłany w swoim paśmie podstawowym, bez dokonywania na nim operacji modulacji, zajmowałby w kanale transmisyjnym pasmo o szerokości ω_m, gdzie ω_m jest największą pulsacją widma sygnału. Jeśli jest on przesyłany w systemie modulacji AM-SC lub AM, zajmuje w kanale pasmo o szerokości $2\omega_m$, a więc dwukrotnie większej. Zauważmy jednak, że informacja zawarta w sygnale AM-SC lub AM jest przesyłana w sposób nadmiarowy. Ze względu na symetrię wstęg bocznych do przesłania pełnej informacji o sygnale $x(t)$ wystarczy bowiem transmisja tylko jednej z nich – górnej lub dolnej. Systemy modulacji, w których przesyłana jest tylko jedna wstęga boczna noszą nazwę *jednowstęgowych*.

W przypadku sygnału jednowstęgowego SSB-SC funkcja modulująca ma postać:

$$m(t) = x(t) \pm \mathrm{j}\,\hat{x}(t), \tag{10.62}$$

gdzie $x(t)$ jest sygnałem modulującym. Jeśli w sygnale SSB-SC jest transmitowana tylko wstęga górna, we wzorze (10.62) występuje znak „$+$”, natomiast jeśli transmitowana jest tylko wstęga dolna – znak „$-$”. Wynika stąd, że jeśli transmitowana jest wstęga górna, funkcja modulująca jest sygnałem analitycznym $z_x(t)$ sygnału modulującego $x(t)$, jeśli zaś transmitowana jest wstęga dolna, funkcja modulująca jest sygnałem analitycznym sprzężonym $z_x^*(t)$.

Zgodnie ze wzorem ogólnym (10.37), reprezentacje analityczne sygnałów SSB-SC zawierających wstęgę górną i odpowiednio wstęgę dolną są określone wzorami:

$$z^g_{\text{SSB-SC}}(t) = z_x(t) z_c(t), \tag{10.63}$$

$$z^d_{\text{SSB-SC}}(t) = z_x^*(t) z_c(t), \tag{10.64}$$

lub wzorem

$$z_{\text{SSB-SC}}(t) = Y_0\sqrt{x^2(t) + \hat{x}^2(t)}\,\mathrm{e}^{\mathrm{j}[\Omega t \pm \operatorname{arctg}[\hat{x}(t)/x(t)]]}, \tag{10.65}$$

gdzie

$$Y_{\text{SSB-SC}}(t) = Y_0\sqrt{x^2(t) + \hat{x}^2(t)} \tag{10.66}$$

jest obwiednią sygnału SSB-SC oraz

$$\psi_{\text{SSB-SC}}(t) = \Omega t \pm \operatorname{arctg}[\hat{x}(t)/x(t)] \tag{10.67}$$

jego kątem. Ze wzorów tych wynika, że sygnał SSB-SC jest sygnałem zmodulowanym zarówno amplitudowo, jak i kątowo.

Wydzielając we wzorach (10.63) i (10.64) część rzeczywistą, otrzymamy postać rzeczywistą sygnału SSB-SC:

$$y_{\text{SSB-SC}}(t) = Y_0[x(t)\cos\Omega t \mp \hat{x}(t)\sin\Omega t] = x_I(t)\cos\Omega t \mp x_Q(t)\sin\Omega t, \quad (10.68)$$

gdzie znak „–" odpowiada wstędze górnej, a znak „+" wstędze dolnej. Odnotujmy, że wzór (10.68) stanowi zapis Rice'a (10.14) sygnału SSB-SC, przy czym $x_I(t) = Y_0 x(t)$ jest jego składową synfazową, a $x_Q(t) = Y_0\hat{x}(t)$ – składową kwadraturową.

10.4.2. Widmo sygnału SSB-SC

Pokażemy obecnie, że widmo sygnału SSB-SC zawiera istotnie tylko jedną ze wstęg bocznych. W tym celu wyznaczymy najpierw widmo $Z^g_{\text{SSB-SC}}(\omega)$ sygnału analitycznego sygnału SSB-SC reprezentującego wstęgę górną (wzór (10.63)).

Na podstawie twierdzenia 3.9 o splocie w dziedzinie częstotliwości z zależności (10.63) wynika, że

$$Z^g_{\text{SSB-SC}}(\omega) = \frac{1}{2\pi}\left[Z_x(\omega) * Z_c(\omega)\right], \quad (10.69)$$

gdzie $Z_x(\omega)$ jest widmem sygnału analitycznego sygnału modulującego, a $Z_c(\omega)$ widmem sygnału analitycznego fali nośnej. Ponieważ, zgodnie ze wzorami (10.34) i (10.36), $Z_x(\omega) = 2X(\omega)\boldsymbol{1}(\omega)$ oraz $Z_c(\omega) = 2\pi Y_0\delta(\omega-\Omega)$, zatem korzystając z właściwości splotu dystrybucji Diraca (1.19) wnioskujemy, iż

$$\begin{aligned} Z^g_{\text{SSB-SC}}(\omega) &= \frac{1}{2\pi}\left[2\pi Y_0\delta(\omega-\Omega) * 2X(\omega)\boldsymbol{1}(\omega)\right] = \\ &= 2Y_0 X(\omega-\Omega)\boldsymbol{1}(\omega-\Omega). \end{aligned} \quad (10.70)$$

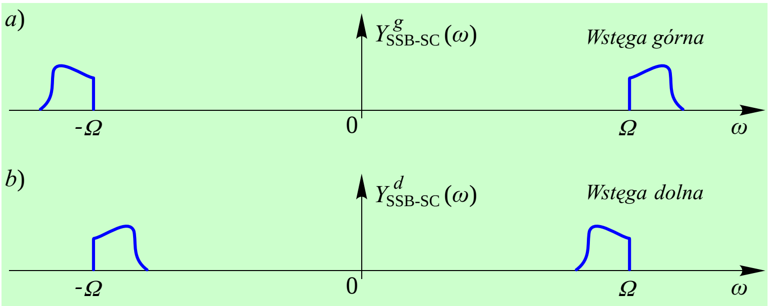
Widmo $Z_{\text{SSB-SC}}(\omega)$ jest oczywiście prawostronne i – jak wynika ze wzoru (10.70) – jest różne od zera jedynie dla pulsacji leżących powyżej pulsacji nośnej Ω, tj. pulsacji odpowiadających wstędze górnej. Korzystając teraz z zależności (10.35) i uwzględniając, że sygnał modulujący $x(t)$ jest z założenia rzeczywisty, a więc że jego widmo jest hermitowskie (tzn. spełnia równość $X(\omega) = X^*(-\omega)$ – por. p. 3.2.4), otrzymujemy ostatecznie:

$$\begin{aligned} Y^g_{\text{SSB-SC}}(\omega) &= \frac{1}{2}\left[Z^g_{\text{SSB-SC}}(\omega) + Z^{g*}_{\text{SSB-SC}}(-\omega)\right] = \\ &= Y_0[X(\omega-\Omega)\boldsymbol{1}(\omega-\Omega) + X^*(-\omega-\Omega)\boldsymbol{1}(-\omega-\Omega)] = \\ &= Y_0[X(\omega-\Omega)\boldsymbol{1}(\omega-\Omega) + X(\omega+\Omega)\boldsymbol{1}(-\omega-\Omega)]. \end{aligned} \quad (10.71)$$

Identyczny rezultat otrzymalibyśmy, obliczając transformaty Fouriera obu stron równości (10.68) (przy założeniu znaku „−") i stosując przy tym twierdzenia o modulacji (3.26) i (3.27) oraz zależność (10.26).

Widmo (10.71) jest przedstawione na rys. 10.17a. Postępując analogicznie można pokazać, że widmo sygnału SSB reprezentującego wstęgę dolną (znak „+" we wzorze (10.68)) ma postać (rys. 10.17b):

$$Y^{d}_{\text{SSB-SC}}(\omega) = Y_0[X(\omega - \Omega)\boldsymbol{1}(-\omega + \Omega) + X(\omega + \Omega)\boldsymbol{1}(\omega + \Omega)]. \tag{10.72}$$



Rys. 10.17. Widmo sygnału SSB-SC: wstęga górna (a) oraz wstęga dolna (b)

10.4.3. Szerokość pasma sygnału SSB-SC

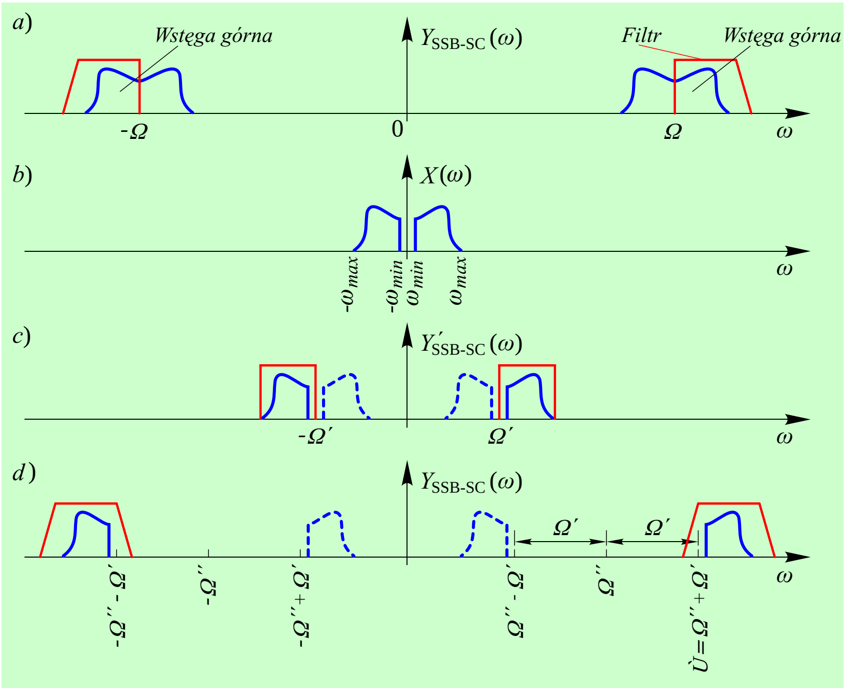
Ponieważ w widmie sygnału SSB-SC występuje tylko jedna wstęga boczna, jego szerokość pasma wyrażona w hercach jest określona wzorem:

$$B_{\text{SSB-SC}} = f_m, \tag{10.73}$$

gdzie f_m jest maksymalną częstotliwością widma sygnału modulującego $x(t)$. Do transmisji sygnału SSB-SC wymagane jest zatem dwukrotnie mniejsze pasmo, niż w przypadku systemów dwuwstęgowych AM i AM-SC.

10.4.4. Generacja sygnału SSB-SC

Najprostszą koncepcyjnie metodą generacji sygnału SSB-SC jest *metoda filtracji*, polegająca na wytworzeniu sygnału AM-SC i odfiltrowaniu za pomocą filtru środkowoprzepustowego jednej z jego wstęg bocznych. W przypadku generacji sygnału SSB-SC reprezentującego wstęgę górną (rys. 10.18a) filtr ten powinien przepuścić bez zmian składowe o pulsacjach większych od pulsacji nośnej Ω i wyeliminować składowe o pulsacjach mniejszych od Ω. Powinien on mieć zatem bardzo stromą charakterystykę filtracji w otoczeniu pulsacji Ω.

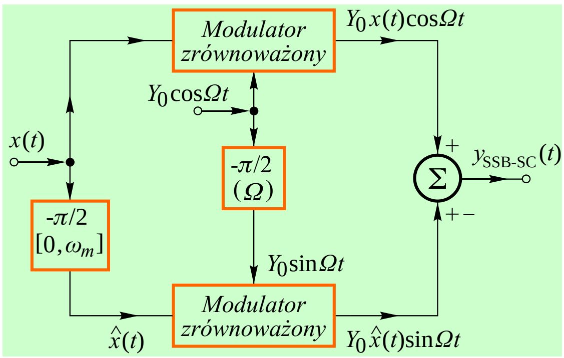


Rys. 10.18. Generacja sygnału SSB-SC metodą filtracji z pośrednią przemianą częstotliwości

Wymagania co do stromości zbocza charakterystyki filtru są mniej krytyczne, jeśli w widmie sygnału modulującego nie występują składowe o małych pulsacjach lub ich gęstość widmowa jest pomijalna (rys. 10.18b). Sytuacja taka ma miejsce np. w przypadku sygnału mowy, którego dolna częstotliwość pasma wynosi 200 ÷ 300 Hz, a więc w otoczeniu pulsacji nośnej Ω występuje pewne pasmo puste. Jednak nawet w takim przypadku realizacja filtru o dostatecznie stromej charakterystyce jest trudna, jeśli pulsacja nośna Ω jest duża. Generacja sygnału SSB-SC jest wówczas realizowana dwustopniowo. Najpierw wytwarzany jest sygnał AM-SC o pośredniej pulsacji nośnej Ω' (tzw. *podnośnej*), znacznie mniejszej od pulsacji Ω (rys. 10.18c). Z sygnału tego jest odfiltrowywana jedna ze wstęg bocznych za pomocą łatwiejszego do realizacji filtru środkowoprzepustowego. Otrzymany w ten sposób sygnał jednowstęgowy jest z kolei wykorzystywany do ponownej modulacji AM-SC fali nośnej o pulsacji Ω'', takiej że $\Omega'' + \Omega' = \Omega$, gdzie Ω jest właściwą pulsacją nośną. W efekcie widmo z rys. 10.18c zostaje rozszczepione na cztery części przesunięte do punktów $\pm\Omega'' \pm \Omega'$ (rys. 10.18d). Zbędne składowe widma są odfiltrowywane za pomocą łatwiejszego do realizacji filtru środkowoprzepustowego o łagodnie opadających zboczach. W efekcie otrzymujemy sygnał SSB-SC o pożądanym usytuowaniu widma na skali pulsacji.

Metoda filtracji jest powszechnie stosowana do generacji sygnałów w systemach wielokrotnej telefonii jednowstęgowej. W przypadku wysokich wymagań co do dokładności tłumienia zbędnej wstęgi bocznej stosowane są filtry kwarcowe.

Inną metodą generacji sygnału SSB-SC jest *metoda przesuwania fazy*. Koncepcja tej metody jest oparta na zapisie Rice'a (10.68). Zgodnie ze wzorem (10.68), w celu wytworzenia sygnału SSB-SC należy wygenerować dwa sygnały AM-SC $Y_0 x(t) \cos \Omega t$ oraz $Y_0 \hat{x}(t) \sin \Omega t$ i utworzyć ich różnicę (wstęga górna) lub sumę (wstęga dolna). Schemat blokowy generatora sygnału SSB-SC działającego w oparciu o metodę przesuwania fazy jest przedstawiony na rys. 10.19. Generację sygnałów AM-SC w torze synfazowym i kwadraturowym można zrealizować za pomocą modulatorów iloczynowych, natomiast w celu wytworzenia sygnału $\hat{x}(t)$ należy zastosować filtr Hilberta. Jak podkreślaliśmy, filtr Hilberta o stałym opóźnieniu fazy $\pi/2$ w całym paśmie pulsacji jest fizycznie nierealizowalny. Można jednak zrealizować pasmowy *przesuwnik fazy* zapewniający z dobrym przybliżeniem opóźnienie fazy o $\pi/2$ w skończonym paśmie sygnału modulującego $x(t)$. Układ generatora sygnału SSB-SC działający według metody przesuwania fazy jest nazywany *modulatorem Hartleya*.



Rys. 10.19. Generacja sygnału SSB-SC za pomocą modulatora Hartleya

Niekiedy, w celu uniknięcia trudności w zapewnieniu przez przesuwnik fazy stałego opóźnienia $\pi/2$ w całym paśmie sygnału modulującego, fazę sygnału $x(t)$ przesuwa się przed podaniem na wejścia modulatorów iloczynowych o wartość α w górnym torze i o wartość β w dolnym torze, tak aby $\alpha - \beta = \pi/2$. Utrzymanie stałej różnicy $\alpha - \beta = \pi/2$ w paśmie sygnału modulującego jest już łatwiejsze, niż utrzymanie stałego opóźnienia $\pi/2$. Taka metoda generacji sygnału SSB-SC jest nazywana *metodą kompensacji fazy*.

Jeszcze inną metodą generacji sygnału SSB-SC jest *metoda Weavera* (por. [2], p. 8.2.2.3). Metoda ta nie wymaga stosowania pasmowych przesuwników fazy, można ją jednak stosować tylko wtedy, kiedy pasmo sygnału modulującego jest ograniczone od dołu.

10.4.5. Demodulacja sygnału SSB-SC

Podobnie jak w przypadku sygnału AM-SC, demodulację sygnału SSB-SC można zrealizować za pomocą detektora koherentnego (rys. 10.7). Układ ten wykonuje operację mnożenia sygnału SSB-SC przez sygnał fali nośnej generowany przez lokalny generator odbiornika. W wyniku tej operacji powstaje sygnał

$$\begin{aligned} y_{\text{SSB-SC}}(t)\cos\Omega t &= Y_0[x(t)\cos\Omega t \mp \hat{x}(t)\sin\Omega t]\cos\Omega t = \\ &= Y_0[x(t)\cos^2\Omega t \mp \hat{x}(t)\sin\Omega t\cos\Omega t] = \\ &= \frac{1}{2}Y_0 x(t) + \frac{1}{2}Y_0[x(t)\cos 2\Omega t \mp \hat{x}(t)\sin 2\Omega t]. \end{aligned} \tag{10.74}$$

Sygnał ten jest sumą składowej proporcjonalnej do sygnału informacyjnego i składowej SSB-SC o dwukrotnie większej pulsacji nośnej 2Ω. Składowa SSB-SC jest odfiltrowywana za pomocą filtru dolnoprzepustowego RC i na wyjściu detektora otrzymujemy składową informacyjną.

Demodulacja sygnału SSB-SC za pomocą detektora koherentnego ma omówione wcześniej wady związane z koniecznością dokładnej synchronizacji częstotliwości i fazy fal nośnych generowanych w nadajniku i odbiorniku. Jak pamiętamy (por. p. 10.2.5), w przypadku koherentnej demodulacji sygnału AM-SC stały błąd fazy nie powoduje zniekształceń sygnału zdemodulowanego, a jego efektem jest jedynie tłumienie sygnału. Inaczej przedstawia się sytuacja w przypadku koherentnej demodulacji sygnału SSB-SC. Jeśli występuje błąd fazy $\Delta\varphi$ lokalnego generatora fali nośnej, to (przy założeniu zerowego błędu częstotliwości) w wyniku mnożenia sygnału odebranego przez tę falę otrzymujemy sygnał:

$$\begin{aligned} y_{\text{SSB-SC}}(t)\cos(\Omega t + \Delta\varphi) &= Y_0[x(t)\cos\Omega t\cos(\Omega t + \Delta\varphi) \mp \\ &\quad \mp \hat{x}(t)\sin\Omega t\cos(\Omega t + \Delta\varphi)] = \\ &= \frac{Y_0}{2}\left[x(t)\cos\Delta\varphi \pm \hat{x}(t)\sin\Delta\varphi\right] + \\ &\quad + \frac{Y_0}{2}\left[x(t)\cos(2\Omega t + \Delta\varphi) \mp \hat{x}(t)\sin(2\Omega t + \Delta\varphi)\right]. \end{aligned} \tag{10.75}$$

Po odfiltrowaniu drugiej składowej na wyjściu demodulatora otrzymujemy sygnał

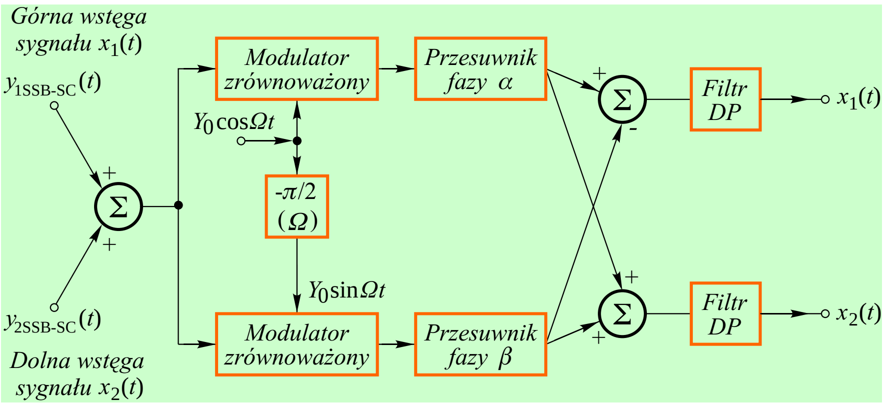
$$y_d(t) = \frac{Y_0}{2}\left[x(t)\cos\Delta\varphi \pm \hat{x}(t)\sin\Delta\varphi\right]. \tag{10.76}$$

Widzimy więc, że oprócz składowej proporcjonalnej do sygnału informacyjnego sygnał zdemodulowany zawiera niepożądaną składową proporcjonalną do jego transformaty Hilberta. Aby zbadać jaki efekt wywołuje ta składowa w dziedzinie widmowej, wyznaczymy widmo sygnału zdemodulowanego:

$$\begin{aligned} Y_d(\omega) &= \frac{Y_0}{2}\left[X(\omega)\cos\Delta\varphi \pm (-\mathrm{j}\operatorname{sgn}\omega)X(\omega)\sin\Delta\varphi\right] = \\ &= \frac{Y_0}{2}X(\omega)\,\mathrm{e}^{\mp\mathrm{j}\Delta\varphi(\operatorname{sgn}\omega)}\,. \end{aligned} \tag{10.77}$$

Wynika stąd, że każda składowa widma sygnału informacyjnego jest przesuwana w fazie o stałą wartość $\Delta\varphi$. Gdy błąd fazy $\Delta\varphi$ jest mały, wówczas w przypadku transmisji sygnału mowy zniekształcenia fazy nie są zbyt dokuczliwie. Ucho ludzkie jest bowiem mało wrażliwe na zniekształcenia fazowe. Dla większych wartości błędu $\Delta\varphi$ wywołują one efekt głosu Kaczora Donalda. Jeśli natomiast transmitowany jest sygnał muzyki lub wizyjny, zniekształcenia te są nieakceptowalne.

Inną metodą demodulacji, wolną od powyższych wad, ale wymagającą zastosowania bardziej złożonego układu demodulacyjnego, jest *metoda kompensacyjna*. Układ dwukanałowego demodulatora kompensacyjnego sygnału SSB-SC jest przedstawiony na rys. 10.20.



Rys. 10.20. Schemat blokowy dwukanałowego kompensacyjnego demodulatora sygnału SSB-SC

Układ ten jest przeznaczony do jednoczesnego odbioru dwóch sygnałów SSB-SC $y_{1\text{SSB-SC}}(t)$ i $y_{2\text{SSB-SC}}(t)$ o tej samej pulsacji nośnej Ω, przy czym pierwszy z nich reprezentuje wstęgę górną, drugi zaś – wstęgę dolną. Sygnały modulujące $x_1(t)$ i $x_2(t)$ tych sygnałów pochodzą z niezależnych źródeł informacji. Oba sygnały są podane na wejście sumatora. Suma sygnałów jest mnożona w układach modulatorów iloczynowych przez lokalną falę nośną nieprzesuniętą i opóźnioną o $\pi/2$. Przesuwniki fazy przesuwają fazy sygnałów iloczynowych o wartości α oraz β, tak aby ich różnica była równa $\pi/2$. Utrzymanie stałej różnicy $\alpha - \beta = \pi/2$ zapewnia eliminację ewentualnych błędów fazy lokalnego generatora. Sygnały z wyjść przesuwników fazowych są następnie po zsumowaniu i odjęciu rozdzielone na dwa kanały i filtrowane za pomocą filtrów dolnoprzepustowych. Na wyjściach tych filtrów otrzymujemy zdemodulowane sygnały informacyjne $x_1(t)$ i $x_2(t)$. Zasadę działania dwukanałowego demodulatora kompensacyjnego można bez trudu wyjaśnić, przyjmując harmoniczne modele sygnałów modulujących i wykonując kolejne operacje na sygnałach wynikające ze struktury demodulatora.

10.4.6. Sygnał SSB

W systemie SSB wraz z sygnałem SSB-SC jest przesyłany dodatkowy sygnał fali nośnej o dostatecznie dużym poziomie. Podobnie jak w systemie dwuwstęgowym AM umożliwia to uniknięcie trudności z synchronizacją częstotliwości i fazy lokalnego generatora fali nośnej w odbiorniku.

W przypadku sygnału SSB funkcja modulująca ma postać:

$$m(t) = 1 + x(t) \pm \mathrm{j}\,\hat{x}(t), \tag{10.78}$$

przy czym znak „+” odpowiada wstędze górnej, a znak „−” wstędze dolnej. Uwzględniając tę funkcję we wzorze ogólnym (10.37) i wydzielając w nim część rzeczywistą, otrzymujemy postać rzeczywistą sygnału SSB:

$$\begin{aligned} y_{\mathrm{SSB}}(t) &= Y_0[1 + x(t)\cos\Omega t \mp \hat{x}(t)\sin\Omega t], \\ &= Y_0\sqrt{[1+x^2(t)] + \hat{x}^2(t)}\cos\left(\Omega t \pm \operatorname{arctg}\frac{\hat{x}(t)}{1+x(t)}\right). \end{aligned} \tag{10.79}$$

Obwiednia i kąt sygnału SSB są więc określone wzorami:

$$Y_{\mathrm{SSB}}(t) = Y_0\sqrt{[1+x^2(t)] + \hat{x}^2(t)}, \tag{10.80}$$

$$\psi_{\mathrm{SSB}}(t) = \Omega t \pm \operatorname{arctg}\frac{\hat{x}(t)}{1+x(t)}, \tag{10.81}$$

skąd wynika, że sygnał SSB jest zmodulowany zarówno amplitudowo, jak i kątowo.

Widmo sygnału SSB różni się od widma sygnału SSB-SC (10.71) lub (10.72) jedynie dwoma dodatkowymi dystrybucjami Diraca w punktach $\pm\Omega$. Szerokość pasma sygnału SSB jest więc identyczna jak sygnału SSB-SC i określona wzorem $B_{\mathrm{SSB}} = f_m$, gdzie f_m jest największą częstotliwością w widmie sygnału modulującego $x(t)$.

Obecność fali nośnej w sygnale SSB umożliwia zastosowanie do jego demodulacji dużo prostszego układu detektora obwiedni. System SSB, podobnie jak system AM, charakteryzuje się jednak niską sprawnością energetyczną. Większość mocy sygnału SSB jest przenoszona w składowej nośnej.

10.5. Porównanie systemów modulacji amplitudy. Modulacja VSB

10.5.1. Zestawienie zalet i wad systemów modulacji amplitudy

Zaletą systemu AM-SC jest 100% sprawność energetyczna. Wadą jest konieczność stosowania do demodulacji sygnału AM-SC detektora koherentnego, wymagającego generowania w odbiorniku wysokostabilnej fali nośnej o zsynchronizowanej częstotliwości i fazie z częstotliwością i fazą fali nośnej nadajnika. Wadą jest także, choć w mniejszym stopniu, dwukrotnie szersze pasmo w porównaniu z pasmem sygnałów jednowstęgowych oraz dość złożony układ modulatora zrównoważonego generującego sygnał AM-SC.

Zaletą systemu AM są proste układy generacji (modulator prostownikowy) i demodulacji (detektor obwiedni). Do wad należy niska, zaledwie kilkuprocentowa sprawność energetyczna i dwukrotnie szersze pasmo niż w systemach jednowstęgowych.

Do zalet systemu SSB-SC należy zaliczyć 100% sprawność energetyczną oraz wąskie pasmo. Wadę stanowią natomiast wysokie wymagania co do dokładności układów generujących sygnał SSB-SC bez względu na zastosowaną metodę generacji (filtracyjną, przesuwania fazy, kompensacji fazy lub Weavera).

Z kolei wśród zalet systemu SSB należy wymienić wąskie pasmo i możliwość stosowania detektora obwiedni. Wadą jest natomiast niska sprawność energetyczna.

Oprócz dwukrotnie węższego pasma systemy jednowstęgowe mają jeszcze jedną przewagę nad dwuwstęgowymi. Są one bardziej odporne na zjawisko *zaniku selektywnego* (ang. *fading*). Przyczyną tego zjawiska są odbicia sygnału od jonosfery i ziemi podczas jego propagacji od nadajnika do odbiornika, wskutek czego dociera on do odbiornika różnymi drogami. W efekcie tego poszczególne składowe częstotliwościowe sygnału są odbierane z różnymi opóźnieniami. Ponieważ warunki jonosferyczne zmieniają się przypadkowo, opóźnienia te są także przypadkowe i mogą powodować silne zniekształcenia fazowe sygnału. Zanik selektywny jest szczególnie odczuwalny w zakresie dużych częstotliwości w przypadku transmisji sygnałów o szerokim paśmie, a więc sygnałów fonicznych muzyki i sygnałów wizyjnych. Sygnał mowy jest stosunkowo mało czuły na zniekształcenia wywołane zanikiem selektywnym.

10.5.2. Systemy z falą nośną wprowadzoną w odbiorniku

Próbę połączenia zalet systemów modulacji amplitudy bez fali nośnej i z falą nośną stanowią systemy z falą nośną wprowadzoną po stronie odbiorczej przed demodulacją sygnału. W systemach tych transmitowane są sygnały bez fali nośnej AM-SC lub SSB-SC, do których w odbiorniku jest dodawana fala nośna o dostatecznie dużym poziomie. Zabieg ten umożliwia demodulację sygnału odebranego za pomocą detektora obwiedni. Pozostają jednak problemy z zapewnieniem synchronizacji częstotliwości i fazy lokalnej fali nośnej. Można wykazać, że zniekształcenia spowodowane błędami częstotliwości i fazy lokalnej fali nośnej mają podobny charakter, jak w przypadku koherentnego odbioru sygnałów AM-SC i SSB-SC.

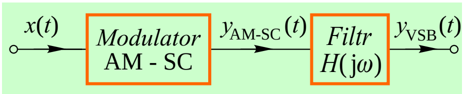
Systemy SSB-SC z falą nośną wprowadzoną w odbiorniku zachowują zalety systemów SSB-SC, tj. wąskie pasmo i 100% sprawność energetyczną, a ponadto umożliwiają demodulację sygnału za pomocą detektora obwiedni, a więc mają zaletę właściwą dla systemów z falą nośną. Połączenie tych zalet decyduje o częstym zastosowaniu modulacji SSB-SC z falą nośną wprowadzoną w odbiorniku do transmisji sygnału mowy w systemach telefonicznych i radiofonicznych. W przypadku sygnałów o szerokim paśmie, np. wizyjnych, występują kłopoty z generacją sygnału SSB-SC. W metodzie filtracyjnej są one związane z trudnościami w realizacji filtru o dostatecznie stromym zboczu charakterystyki, w metodzie przesuwania fazy i metodzie kompensacji – z zapewnieniem w szerokim paśmie stałego przesunięcia fazy, natomiast w metodzie Weavera – ze znacznym stopniem skomplikowania modulatora. Ta ostatnia metoda może być zresztą, jak wspomnieliśmy, stosowana jedynie w przypadku sygnałów modulujących o ograniczonym od dołu widmie, co wyklucza jej zastosowanie do sygnałów wizyjnych.

10.5.3. System VSB

Z dokonanego porównania amplitudowych systemów modulacji wynika, że żaden z nich nie łączy w sobie zalet systemów jednowstęgowych (wąskie pasmo), systemów bez fali nośnej (100% sprawność energetyczna) i systemów z falą nośną (prostota układu demodulacji). Rozwiązaniem kompromisowym jest koncepcja systemu modulacji VSB *z częściowo stłumioną wstęgą boczną* (lub *ze szczątkową wstęgą boczną*, od ang. *Vestigial Sideband*).

W systemie VSB (rys. 10.21) generowany jest najpierw sygnał dwuwstęgowy bez fali nośnej AM-SC. Sygnał ten jest następnie filtrowany za pomocą filtru środkowoprzepustowego o specjalnie dobranej charakterystyce amplitudowo-fazowej $H(\mathrm{j}\,\omega)$, którego zadaniem jest silne, lecz niepełne stłumienie dolnej wstęgi bocznej sygnału AM-SC. Jeśli przez $h(t)$ oznaczymy odpowiedź impulsową tego filtru, to sygnał VSB możemy zapisać w postaci:

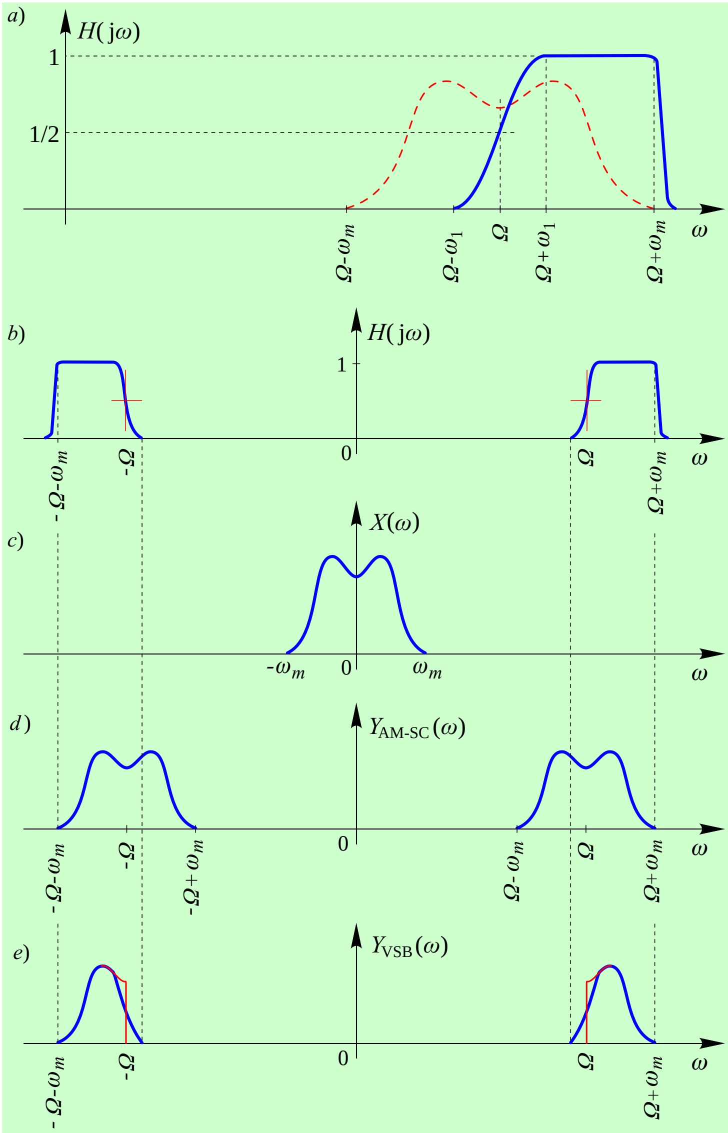
$$y_{\text{VSB}}(t) = y_{\text{AM-SC}}(t) * h(t). \tag{10.82}$$



Rys. 10.21. Schemat blokowy generatora sygnału VSB

Prawostronna część charakterystyki amplitudowej $|H(\mathrm{j}\,\omega)|$ filtru jest pokazana na rys. 10.22a. Linią przerywaną wykreślono na nim dodatkowo widmo prawostronne sygnału AM-SC. Charakterystyka $|H(\mathrm{j}\,\omega)|$ jest ukształtowana w taki sposób, aby w przedziale pulsacji $\Omega + \omega_1 \leqslant \omega \leqslant \Omega + \omega_m$ składowe widma sygnału AM-SC były przenoszone bez zmian, tzn. jej wartość w tym przedziale jest równa 1. Pulsacja ω_m jest największą pulsacją widma sygnału modulującego, a pulsacja ω_1 określa szerokość szczątkowej wstęgi bocznej sygnału VSB. W zakresie pulsacji $\Omega - \omega_m \leqslant \omega \leqslant \Omega - \omega_1$ charakterystyka filtru jest równa zeru, a więc składowe widma sygnału AM-SC z tego zakresu są całkowicie stłumione. Natomiast w zakresie pulsacji $\Omega - \omega_1 \leqslant \omega \leqslant \Omega + \omega_1$ charakterystyka filtru jest rosnąca, przy czym – co jest warunkiem istotnym – wykazuje symetrię nieparzystą względem osi rzędnych w punkcie $\omega = \Omega$ odpowiadającym pulsacji nośnej, w którym przybiera wartość 1/2. Powyżej pulsacji $\Omega + \omega_m$ kształt charakterystyki może być dowolny. Wartość pulsacji ω_1, decydującej o punkcie odcięcia $\Omega - \omega_1$ wstęgi dolnej, jest niewielkim ułamkiem (w praktyce kilkanaście procent) szerokości ω_m pasma podstawowego sygnału modulującego. W efekcie filtr przenosi bez zmian większość wstęgi górnej sygnału AM-SC (powyżej pulsacji $\Omega + \omega_1$), a ponadto małą jej część w niewielkim stopniu stłumioną (w zakresie $\Omega \leqslant \omega \leqslant \Omega + \omega_1$) oraz silnie stłumioną szczątkową wstęgę dolną (w zakresie $\Omega - \omega_1 \leqslant \omega \leqslant \Omega$).

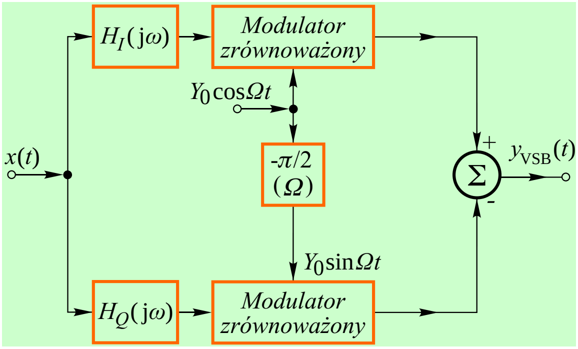
Na rys. 10.22b–e przedstawiono widmową interpretację koncepcji generacji sygnału VSB drogą filtracji sygnału AM-SC. Dla wygody ilustracji przyjęto rzeczywisty charakter widm. Podkreślmy, że warunek nieparzystej symetrii przedniego zbocza charakterystyki filtru w przedziale $\Omega - \omega_1 \leqslant \omega \leqslant \Omega + \omega_1$ jest kluczowy dla prawidłowej generacji sygnału VSB. Tylko wówczas bowiem w wyniku koherentnej demodulacji tego sygnału (przeniesienia jego widma $Y_{\text{VSB}}(\omega)$ do punktu $\omega = 0$) nastąpi pełna kompensacja szczątkowej części wstęgi dolnej sygnału VSB leżącej w przedziale $\Omega - \omega_1 \leqslant \omega \leqslant \Omega + \omega_1$ ze szczątkową częścią jego wstęgi górnej leżącą w przedziale $-\Omega - \omega_1 \leqslant \omega \leqslant -\Omega + \omega_1$. W konsekwencji na wyjściu detektora koherentnego zostanie odzyskane niezniekształcone widmo sygnału modulującego.



Rys. 10.22. Interpretacja koncepcji generacji sygnału VSB

10.5.4. Generacja sygnału VSB

Z uwagi na trudności w realizacji filtru środkowoprzepustowego o opisanych wyżej właściwościach omówiona metoda generacji sygnału VSB ma znaczenie jedynie teoretyczne. Koncepcja tej metody jest jednak wykorzystywana w praktyce do generacji sygnału VSB za pomocą układu pokazanego na rys. 10.23. Układ ten jest modyfikacją modulatora Hartleya sygnału SSB-SC z rys. 10.19. Funkcję filtru środkowoprzepustowego $H(\mathrm{j}\,\omega)$ przejmują dwa prostsze do realizacji filtry dolnoprzepustowe $H_I(\mathrm{j}\,\omega)$ i $H_Q(\mathrm{j}\,\omega)$, umieszczone odpowiednio w torze synfazowym i kwadraturowym modulatora.



Rys. 10.23. Modulator sygnału VSB z modyfikacją składowej kwadraturowej

Przeanalizujemy warunki, jakie muszą spełniać filtry dolnoprzepustowe $H_I(\mathrm{j}\,\omega)$ i $H_Q(\mathrm{j}\,\omega)$, aby układ modulatora z rys. 10.23 był równoważny układowi z rys. 10.21. Z dokładnej analizy sygnału VSB wynika (por. np. [2], p. 8.2.2.7), że operacje wykonywane przez filtry dolnoprzepustowe $H_I(\mathrm{j}\,\omega)$ i $H_Q(\mathrm{j}\,\omega)$ na sygnale modulującym $x(t)$ będą dawały ten sam efekt, co operacja wykonywana przez filtr środkowoprzepustowy $H(\mathrm{j}\,\omega)$ na sygnale AM-SC, jeśli ich odpowiedzi impulsowe będą związane z odpowiedzią impulsową $h(t)$ filtru $H(\mathrm{j}\,\omega)$ zależnościami:

$$h_I(t) = h(t)\cos\Omega t, \tag{10.83}$$

$$h_Q(t) = -h(t)\sin\Omega t. \tag{10.84}$$

Z twierdzeń o modulacji (3.26) i (3.27) wynika zatem, że charakterystyki amplitudowo-fazowe filtrów dolnoprzepustowych powinny mieć postacie:

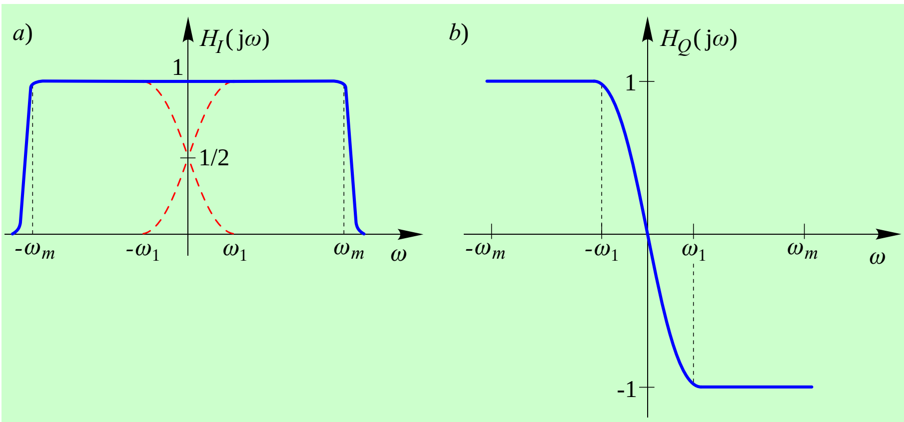
$$H_I(\mathrm{j}\,\omega) = \frac{1}{2}\left[H[\mathrm{j}\,(\omega-\Omega)] + H[\mathrm{j}\,(\omega+\Omega)]\right], \tag{10.85}$$

$$H_Q(\mathrm{j}\,\omega) = -\frac{1}{\mathrm{j}\,2}\left[H[\mathrm{j}\,(\omega-\Omega)] - H[\mathrm{j}\,(\omega+\Omega)]\right]. \tag{10.86}$$

Rozpatrzmy dokładniej równość (10.85). Wynika z niej, że charakterystyka $H_I(\mathrm{j}\,\omega)$ filtru dolnoprzepustowego w torze synfazowym modulatora powstaje w wyniku przesunięcia charakterystyki $H(\mathrm{j}\,\omega)$ z rys. 10.22b w lewo o odcinek Ω oraz w prawo o ten sam odcinek i zsumowania obu przesuniętych części ze współczynnikiem 1/2. W wyniku tych przesunięć powstają cztery części, z których dwie przesunięte do punktu $\omega = 0$ nałożą się na siebie, a dwie pozostałe zostaną przesunięte do punktów $\pm 2\Omega$. Ponieważ na charakterystykę $H(\mathrm{j}\,\omega)$ jest narzucony warunek nieparzystej symetrii względem rzędnych $\omega = \pm\Omega$ wokół punktu $1/2$, zatem w przedziale pulsacji $|\omega| \leqslant \omega_1$ nastąpi pełna kompensacja zboczy obu przesuniętych części. Oznacza to, że w przedziale $|\omega| \leqslant \omega_m$, obejmującym pasmo sygnału modulującego, charakterystyka $H_I(\mathrm{j}\,\omega)$ powinna być stała i równa 1 (rys. 10.24a):

$$H_I(\mathrm{j}\,\omega) = 1, \qquad |\omega| \leqslant \omega_m. \tag{10.87}$$

Poza pasmem $|\omega| \leqslant \omega_m$ charakterystyka $H_I(\mathrm{j}\,\omega)$ może być dowolna. Podobna analiza prowadzi do wniosku, że charakterystyka $H_Q(\mathrm{j}\,\omega)$ filtru w torze kwadraturowym powinna mieć w paśmie $|\omega| \leqslant \omega_m$ kształt pokazany na rys. 10.24b. Poza tym pasmem jej kształt jest dowolny. Przy spełnieniu tych warunków na wyjściu modulatora z rys. 10.23 występuje sygnał VSB o pożądanym kształcie widma.



Rys. 10.24. Charakterystyki filtrów dolnoprzepustowych $H_I(\mathrm{j}\,\omega)$ i $H_Q(\mathrm{j}\,\omega)$ w modulatorze sygnału VSB

Zauważmy, że jeśli omówione warunki są spełnione, filtr $H_I(\mathrm{j}\,\omega)$ w torze synfazowym nie zmienia sygnału modulującego $x(t)$, a tym samym nie wpływa na zmianę składowej synfazowej w porównaniu z modulatorem Hartleya sygnału SSB-SC z rys. 10.19. Sygnał $x(t)$ jest natomiast filtrowany filtrem $H_Q(\mathrm{j}\,\omega)$ w torze kwadraturowym, a więc modyfikacji ulega tylko składowa kwadraturowa sygnału SSB-SC (por. Tabl. 7.1). Zwróćmy także uwagę, że jeśli wartość pulsacji ω_1, regulującej szerokość odciętego pasma wstęgi dolnej, będzie dążyć do zera, to dolna pulsacja odcięcia $\Omega - \omega_1$ będzie zbliżać się do pulsacji Ω, filtr $H_I(\mathrm{j}\,\omega)$ będzie dążyć do filtru wszechprzepustowego, a filtr $H_Q(\mathrm{j}\,\omega)$ – do filtru Hilberta o charakterystyce $-\mathrm{j}\,\mathrm{sgn}\,\omega$. W konsekwencji na wyjściu modulatora otrzymamy sygnał SSB-SC. Wynika stąd, że sygnał jednowstęgowy SSB-SC można traktować jako graniczny przypadek sygnału VSB ze szczątkową wstęgą boczną.

10.5.5. Sygnał telewizyjny

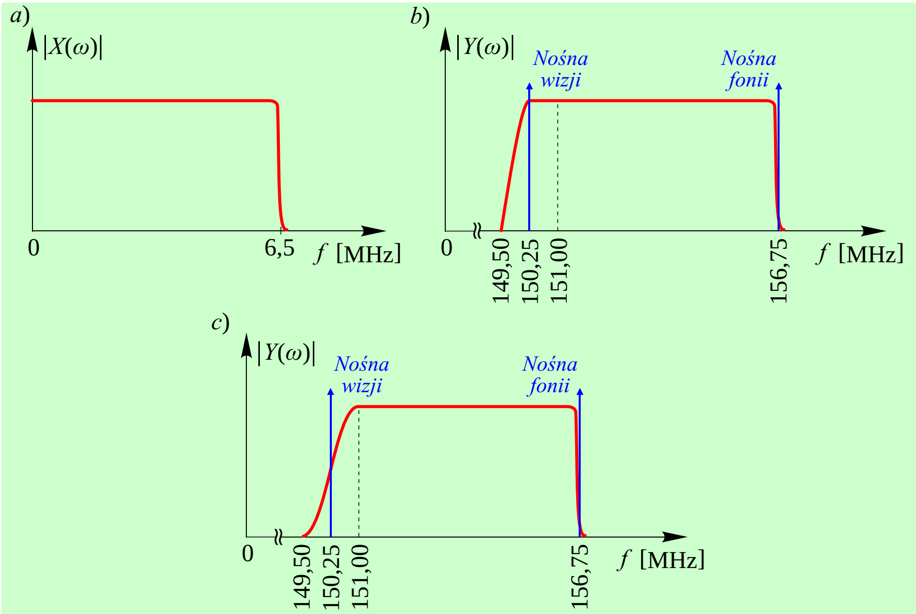
Praktyczne znaczenie modulacji VSB wynika przede wszystkim z faktu, że jest ona stosowana do przesyłania sygnału wizji w systemach telewizyjnych powszechnego użytku. Sygnał wizji jest sygnałem szerokopasmowym, przy czym udział składowych niskoczęstotliwościowych w jego widmie jest znaczący. W warunkach polskich pasmo podstawowe sygnału wizji zajmuje przedział częstotliwości $0 \div 6{,}5$ MHz. W przybliżeniu można przyjąć, że jego gęstość widmowa w tym przedziale jest stała. Wykres idealizowanego widma amplitudowego sygnału wizji przedstawiono na rys. 10.25a.

W systemach telewizyjnych powszechnego użytku dąży się do jak najprostszych i tanich rozwiązań układowych odbiornika. Narzuca to stosowanie do demodulacji sygnału odebranego detektora obwiedni. To z kolei wymaga przesyłania w sygnale zmodulowanym dodatkowej fali nośnej, a więc zastosowania modulacji AM lub SSB. Z drugiej strony pożądane jest, aby sygnał transmitowany zajmował w kanale jak najwęższe pasmo. Jeśli sygnał wizji byłby transmitowany w systemie AM, należałoby zarezerwować w kanale pasmo około 13 MHz, a więc bardzo szerokie. Z kolei, w każdym z omówionych sposobów generacji sygnału SSB występują trudności w niezniekształconym przekazie składowych niskoczęstotliwościowych sygnału wizyjnego.

Kompromisowym rozwiązaniem, spełniającym w zadowalającym stopniu wymienione wymagania, jest zastosowanie systemu modulacji VSB z dodatkową falą nośną. System ten umożliwia znaczne zmniejszenie pasma sygnału w porównaniu z systemem AM, a ponadto – dzięki obecności fali nośnej – zastosowanie do demodulacji detektora obwiedni.

Ze względów technicznych, związanych z trudnościami realizacji odpowiednich filtrów kształtujących sygnał VSB w warunkach dużych wymaganych mocy nadajnika, w rzeczywistych systemach telewizyjnych jest transmitowany nie w pełni ukształtowany sygnał VSB. Przykładowe widmo sygnału transmitowanego przy założeniu, że fala nośna sygnału wizji ma częstotliwość 150,25 MHz, jest pokazane na rys. 10.25b. Zawiera ono pełną wstęgę górną sygnału leżącą w paśmie $150{,}25 \div 156{,}75$ MHz, nieukształtowaną jeszcze część wstęgi dolnej w paśmie $149{,}5 \div 150{,}25$ MHz oraz prążek fali nośnej wizji o częstotliwości 150,25 MHz. Szerokość pasma sygnału zmodulowanego redukuje się zatem z 13 MHz (w przypadku, gdyby sygnał wizyjny był zmodulowany w systemie AM), do około 7,25 MHz. Należy dodać, że w sygnale telewizyjnym obok sygnału wizji jest transmitowany ponadto niezależnie sygnał fonii (zmodulowany w systemie FM). Częstotliwość nośna sygnału fonii jest równa z reguły częstotliwości krańca pasma transmitowanego sygnału wizji, a więc w omawianym przypadku równa 156,75 MHz.

Właściwe kształtowanie przedniego zbocza widma sygnału VSB za pomocą odpowiednich filtrów następuje w odbiorniku przy małym poziomie mocy



Rys. 10.25. Widmo naturalne sygnału wizyjnego (a), widmo sygnału transmitowanego (b) oraz widmo sygnału VSB ukształtowane w odbiorniku (c)

sygnału. Filtry te zapewniają omówioną wyżej nieparzystą symetrię tego zbocza. Dopiero tak ukształtowany sygnał VSB podlega demodulacji za pomocą detektora obwiedni. Amplituda fali nośnej jest przy tym tłumiona w przybliżeniu dwukrotnie. Widmo sygnału VSB ukształtowanego w odbiorniku jest pokazane na rys. 10.25c. Wprawdzie detektor obwiedni wprowadza pewne nieliniowe zniekształcenia odzyskanej obwiedni, jednak ze względu na cechy statystyczne sygnału wizji zniekształcenia te są niewielkie.

Słownik

amplituda chwilowa (obwiednia rzeczywista)
: moduł sygnału analitycznego reprezentującego dany sygnał

częstotliwość chwilowa
: pochodna argumentu sygnału analitycznego reprezentującego dany sygnał podzielona przez 2π

drganie uogólnione
: reprezentacja sygnału (określona względem ustalonej pulsacji ω_0) w postaci

oscylacyjnego sygnału nieharmonicznego o zmiennych w czasie amplitudzie i fazie, którego pulsacja chwilowa oscyluje wokół pulsacji ω_0

faza chwilowa
(określona względem ustalonej pulsacji ω_0) różnica między argumentem sygnału analitycznego reprezentującego dany sygnał i liniową funkcją czasu $\omega_0 t$

filtr Hilberta (kwadraturowy)
filtr, którego odpowiedzią na dany sygnał wejściowy jest transformata Hilberta tego sygnału

funkcja modulująca
funkcja określająca rodzaj modulacji analogowej; iloczyn funkcji modulującej i sygnału analitycznego fali nośnej jest równy sygnałowi analitycznemu sygnału zmodulowanego w danym systemie modulacji analogowej

kanał transmisyjny
środowisko, w którym sygnał jest przekazywany od nadajnika do odbiornika (np. kabel, wolna przestrzeń lub światłowód)

kąt chwilowy
argument sygnału analitycznego reprezentującego dany sygnał

modulacja
operacja na sygnale mająca na celu dopasowanie właściwości widmowych sygnału do charakterystyki kanału transmisyjnego, z reguły przenosząca widmo sygnału w zakres wielkich częstotliwości z zachowaniem informacji zawartej w sygnale

modulacja AM
modulacja analogowa dwuwstęgowa z falą nośną

modulacja AM-SC
modulacja analogowa dwuwstęgowa bez fali nośnej

modulacja amplitudy
modulacja analogowa, w której w zależności od bieżących wartości sygnału informacyjnego uzmienniana jest amplituda harmonicznej fali nośnej

modulacja SSB
modulacja analogowa jednowstęgowa z falą nośną

modulacja SSB-SC
modulacja analogowa jednowstęgowa bez fali nośnej

modulacja VSB
modulacja analogowa z częściowo stłumioną wstęgą boczną

obwiednia zespolona
funkcja zespolona, której częścią rzeczywistą jest składowa synfazowa, częścią urojoną – składowa kwadraturowa, a modułem – obwiednia rzeczywista sygnału

odbiór superheterodynowy
odbiór sygnałów zmodulowanych z pośrednią przemianą częstotliwości; w wyniku tej przemiany widmo sygnału odebranego jest przesuwane w odbiorniku w okolice pewnej ustalonej, zawsze jednakowej częstotliwości

przekształcenie Hilberta
przekształcenie całkowe przyporządkowujące sygnałowi inny sygnał (transformatę Hilberta)

przemiana częstotliwości
operacja wykonywana w odbiorniku polegająca na przesunięcia częstotliwości nośnej sygnału zmodulowanego do innej częstotliwości

przemodulowanie
efekt występujący przy zbyt dużych wartościach chwilowych sygnału modulującego w systemie modulacji AM w stosunku do amplitudy fali nośnej; objawia się skokowymi zmianami fazy sygnału AM w punktach przejścia jego obwiedni przez zera

pulsacja chwilowa
pochodna argumentu sygnału analitycznego reprezentującego dany sygnał

składowe synfazowa i kwadraturowa
składowe wolnozmienne będące czynnikami w rozkładzie sygnału reprezentowanego drganiem uogólnionym na dwie części związane z szybkozmiennymi składowymi kosinusoidalną i sinusoidalną; część rzeczywista i odpowiednio część urojona obwiedni zespolonej sygnału

sygnał analityczny
reprezentacja sygnału w postaci sygnału zespolonego, którego częścią rzeczywistą jest dany sygnał, a częścią urojoną jego transformata Hilberta

sygnał pilotujący
sygnał nośny transmitowany dodatkowo wraz z sygnałem zmodulowanym w niektórych systemach modulacji

transformata Hilberta
sygnał otrzymany w wyniku przekształcenia Hilberta danego sygnału

współczynnik głębokości modulacji
parametr określający właściwości sygnałów zmodulowanych amplitudowo

współczynnik sprawności energetycznej systemu modulacji
współczynnik określający procentowy udział mocy sygnału informacyjnego w całkowitej mocy sygnału zmodulowanego

zanik selektywny
zjawisko zniekształceń fazowych transmitowanego sygnału spowodowane jego odbiciami od jonosfery i ziemi podczas transmisji od nadajnika do odbiornika i docieraniu wskutek tego do odbiornika po różnych drogach propagacji

Literatura

[1] Haykin S.: Systemy telekomunikacyjne. WKiŁ, Warszawa, 1999.

[2] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

Lekcja 11

Modulacje analogowe kąta

W lekcji 11 omówimy analogowe modulacje kąta PM i FM. Podkreślimy, że dzięki rozszerzeniu pasma sygnału w systemach modulacji kąta uzyskuje się większą odporność sygnału na zakłócenia w porównaniu z systemami modulacji amplitudowej. Wprowadzimy podstawowe parametry charakteryzujące sygnały zmodulowane kątowo: dewiację fazy (wskaźnik modulacji) oraz dewiację częstotliwości. Przedyskutujemy krótko przypadek modulacji wąskopasmowych kąta. Dokładniej omówimy modulacje szerokopasmowe PM i FM. W szczególności przedyskutujemy zależność szerokości pasma sygnałów zmodulowanych w obu systemach od pasma sygnału modulującego. Omówimy także zasady generacji i demodulacji sygnałów zmodulowanych kątowo. Na zakończenie dokonamy porównania systemów modulacji amplitudy z systemami modulacji kąta.

11.1. Ogólna charakterystyka systemów modulacji kąta.

11.1.1. Wymiana szerokości pasma na stosunek sygnał–szum w systemach modulacji kąta

Modulacje amplitudowe były pierwszymi znanymi i wdrożonymi w praktyce systemami modulacji sygnałów. Stosowane były do przesyłania sygnałów radiowych i telefonicznych. W systemach telefonicznych są one wykorzystywane po dzień dzisiejszy, natomiast w systemach radiofonicznych powszechnego użytku stosuje się obecnie przede wszystkim modulacje kąta.

Sygnały zmodulowane kątowo zajmują w kanale transmisyjnym znacznie szersze pasmo w porównaniu z sygnałami zmodulowanymi amplitudowo. Przy niskich w pierwszych latach rozwoju radiofonii częstotliwościach nośnych uniemożliwiało to przesyłanie zbyt wielu sygnałów zmodulowanych kątowo przez jeden kanał. Nie istniały wówczas także odpowiednie elementy elektroniczne, za pomocą których można byłoby efektywnie generować i demodulować sygnały zmodulowane kątowo. Jednak w miarę postępów technologicznych i przesuwania się częstotliwości nośnych w coraz to wyższy zakres, do transmisji sygnałów radiowych zaczęto stosować w początkach lat trzydziestych ubiegłego stulecia modulacje kąta.

Dzięki szerszemu pasmu systemy modulacji kąta mają istotną przewagę nad systemami modulacji amplitudy. Przy założeniu takiej samej transmitowanej mocy sygnału użytecznego w obu typach systemów są one bardziej odporne na szumy i zakłócenia. Właściwość ta jest konsekwencją ogólnego podstawowego prawa telekomunikacji, zgodnie z którym, przy określonych zdolnościach transmisyjnych kanału, zwiększanie pasma zajętego w kanale przez transmitowany sygnał umożliwia przesyłanie go przy mniejszym stosunku sygnał-szum. Oznacza to, że przy zadanym poziomie szumów, ta sama informacja zawarta w sygnale może być przesłana przy mniejszej emitowanej mocy sygnału użytecznego. Dobierając odpowiednie wartości parametrów sygnału zmodulowanego kątowo, mamy możliwość regulowania szerokości jego pasma i tym samym wpływania na odporność systemu na zakłócenia. Możliwość wymiany między szerokością pasma a stosunkiem sygnał-szum w systemach modulacji kąta przesądziła o ich powszechnym obecnie zastosowaniu do transmisji fonii w systemach radiowych i telewizyjnych. Podkreślmy, że w przypadku modulacji amplitudowych kompromis taki nie jest możliwy.

11.1.2. Sygnały PM i FM

W systemach modulacji kąta amplituda sygnału zmodulowanego jest stała w czasie, a zmianom zależnym od sygnału modulującego ulega tylko kąt. Sygnał zmodulowany kątowo możemy zatem zapisać w ogólnej postaci:

$$y(t) = Y_0 \cos \psi(t), \tag{11.1}$$

gdzie $Y_0 = \mathrm{const}$ jest amplitudą, a $\psi(t)$ – zmiennym w czasie kątem. W zależności od sposobu uzmienniania kąta $\psi(t)$ w takt zmian sygnału modulującego rozróżniamy dwa podstawowe rodzaje modulacji kątowej: modulację fazy PM i modulację częstotliwości FM.

W przypadku modulacji fazy PM funkcja modulująca $m(t)$ ma postać:

$$m(t) = \mathrm{e}^{\mathrm{j}k_p x(t)}, \tag{11.2}$$

gdzie $k_p > 0$ jest parametrem modulacji, a $x(t)$ – sygnałem modulującym. Zgodnie ze wzorem ogólnym (10.38), sygnał analityczny sygnału PM jest zatem opisany wzorem:

$$z_{\mathrm{PM}}(t) = Y_0\,\mathrm{e}^{\mathrm{j}k_p x(t)}\,\mathrm{e}^{\mathrm{j}\Omega t} = Y_0\,\mathrm{e}^{\mathrm{j}[\Omega t + k_p x(t)]}, \tag{11.3}$$

skąd wynika postać rzeczywista sygnału PM:

$$y_{\mathrm{PM}}(t) = Y_0 \cos[\Omega t + k_p x(t)]\,. \tag{11.4}$$

Amplituda chwilowa sygnału PM jest więc stała:

$$Y_{\mathrm{PM}}(t) = Y_0 = \mathrm{const}, \tag{11.5}$$

a kąt chwilowy zmienia się w czasie według zależności:

$$\psi_{\mathrm{PM}}(t) = \Omega t + k_p x(t)\,. \tag{11.6}$$

Oznacza to, że faza chwilowa sygnału PM (określona względem pulsacji nośnej Ω) jest opisana wzorem $\varphi_{\mathrm{PM}}(t) = k_p x(t)$, a więc zmienia się w czasie proporcjonalnie do sygnału modulującego $x(t)$.

W przypadku modulacji FM funkcja modulująca $m(t)$ ma postać:

$$m(t) = \mathrm{e}^{\mathrm{j}k_f \int x(t)\,\mathrm{d}t}, \tag{11.7}$$

gdzie $k_f > 0$ jest parametrem modulacji. Postacie analityczna i rzeczywista sygnału FM są zatem określone wzorami:

$$z_{\mathrm{FM}}(t) = Y_0\,\mathrm{e}^{\mathrm{j}k_f \int x(t)\,\mathrm{d}t}\,\mathrm{e}^{\mathrm{j}\Omega t} = Y_0\,\mathrm{e}^{\mathrm{j}\left[\Omega t + k_f \int x(t)\,\mathrm{d}t\right]}, \tag{11.8}$$

$$y_{\mathrm{FM}}(t) = Y_0 \cos\left[\Omega t + k_f \int x(t)\,\mathrm{d}t\right]\,. \tag{11.9}$$

Wynika stąd, że amplituda chwilowa sygnału FM jest stałą funkcją czasu:

$$Y_{\mathrm{FM}}(t) = Y_0 = \mathrm{const}, \tag{11.10}$$

a jego kąt chwilowy zmienia się według zależności:

$$\psi_{\mathrm{FM}}(t) = \Omega t + k_f \int x(t)\,\mathrm{d}t\,. \tag{11.11}$$

Faza chwilowa $\varphi_{\mathrm{FM}}(t) = k_f \int x(t)\,\mathrm{d}t$ sygnału FM (określona względem pulsacji nośnej Ω) jest zatem proporcjonalna do całki sygnału modulującego $x(t)$.

Z podanych zależności wynika, że częstotliwość chwilowa sygnału PM

$$f_{\mathrm{PM}}(t) = \frac{1}{2\pi}\frac{\mathrm{d}\psi_{\mathrm{PM}}(t)}{\mathrm{d}t} = F + \frac{k_p}{2\pi}\frac{\mathrm{d}x(t)}{\mathrm{d}t} \tag{11.12}$$

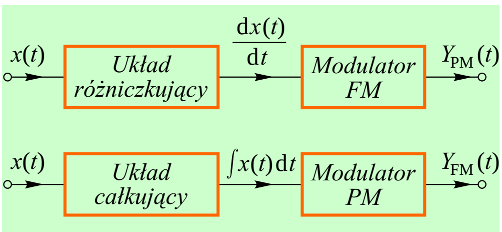
zmienia się wokół częstotliwości nośnej $F = \Omega/2\pi$ proporcjonalnie do pochodnej sygnału modulującego $x(t)$, a częstotliwość chwilowa sygnału FM

$$f_{\mathrm{FM}}(t) = \frac{1}{2\pi}\frac{\mathrm{d}\psi_{\mathrm{FM}}(t)}{\mathrm{d}t} = F + \frac{k_f}{2\pi}x(t) \tag{11.13}$$

zmienia się wokół częstotliwości nośnej F proporcjonalnie do sygnału modulującego $x(t)$.

Należy podkreślić, że w obu systemach modulacji kąta modulowana jest zarówno faza chwilowa, jak i częstotliwość chwilowa. Modulacja jednej z tych wielkości pociąga za sobą modulację drugiej. Różnica polega na sposobie oddziaływania sygnałem modulującym na te wielkości. W przypadku modulacji PM sygnał modulujący oddziaływuje bezpośrednio na fazę chwilową, która zmienia się proporcjonalnie do tego sygnału. Natomiast w przypadku modulacji FM sygnał modulujący oddziaływuje bezpośrednio na częstotliwość chwilową, wywołując jej proporcjonalne zmiany. Różnica ta sprawia, że mimo wielu formalnych podobieństw systemów modulacji PM i FM, z punktu widzenia zastosowań technicznych ich właściwości są odmienne.

Z porównania wzorów (11.4) i (11.9) wynika, że sygnał PM zmodulowany sygnałem $x(t)$ jest równoważny sygnałowi FM zmodulowanemu sygnałem pochodnej $\mathrm{d}x(t)/\,\mathrm{d}t$, natomiast sygnał FM zmodulowany sygnałem $x(t)$ jest równoważny sygnałowi PM zmodulowanemu sygnałem całki $\int x(t)\,\mathrm{d}t$. Modulację PM można zatem zrealizować z wykorzystaniem modulatora FM, podając na jego wejście zróżniczkowany sygnał modulujący, a modulację FM można zrealizować z wykorzystaniem modulatora PM, podając na jego wejście scałkowany sygnał modulujący (rys. 11.1).



Rys. 11.1. Realizacja modulatora PM za pomocą modulatora FM i układu różniczkującego (a) oraz realizacja modulatora FM za pomocą modulatora PM i układu całkującego (b)

Podobieństwa między systemami PM i FM sprawiają, że analiza formalna właściwości obu rodzajów sygnałów zmodulowanych kątowo, a więc ich widm, szerokości pasma, sprawności energetycznej, a także metod modulacji i demodulacji, przebiega podobnie. Wystarczy omówić dokładnie jeden z tych systemów i wskazać na występujące różnice. Z tego względu omówimy dalej dokładnie jedynie system PM (punkty E-L), a w ostatnim punkcie (punkt M) dokonamy obszernego porównania obu systemów. Mimo że w praktyce częściej stosowana jest modulacja FM, wybór systemu PM został tu podyktowany nieco prostszą postacią odpowiednich zależności.

11.1.3. Dewiacja fazy i dewiacja częstotliwości

Podstawowymi parametrami sygnałów zmodulowanych kątowo określającymi ich właściwości widmowe są dewiacja fazy $\Delta\varphi$, oraz dewiacja częstotliwości ΔF.

Definicja 11.1. *Dewiacją fazy* sygnału zmodulowanego kątowo nazywamy maksymalną bezwzględną wartość jego fazy chwilowej lub równoważnie, maksymalną bezwzględną odchyłkę jego kąta chwilowego od liniowej zmiany tego kąta:

$$\Delta\varphi = |\varphi(t)|_{\max} = |\psi(t) - \Omega t|_{\max} . \tag{11.14}$$

Dewiacja fazy jest nazywana także *wskaźnikiem* lub *indeksem modulacji* i oznaczana β.

Definicja 11.2. *Dewiacją częstotliwości* sygnału zmodulowanego kątowo nazywamy maksymalną bezwzględną odchyłkę jego częstotliwości chwilowej od częstotliwości nośnej:

$$\Delta f = |f(t) - F|_{\max} . \tag{11.15}$$

Z definicji 11.1 oraz wzorów (11.6) i (11.11) wynika, że wskaźniki modulacji sygnałów PM i FM są określone wzorami:

$$\Delta\varphi_{\mathrm{PM}} \triangleq \beta_{\mathrm{PM}} = k_p |x(t)|_{\max}, \tag{11.16}$$

$$\Delta\varphi_{\mathrm{FM}} \triangleq \beta_{\mathrm{FM}} = k_f \left| \int x(t)\,\mathrm{d}t \right|_{\max} . \tag{11.17}$$

Tak więc wskaźnik modulacji sygnału PM jest proporcjonalny do maksymalnej bezwzględnej wartości sygnału modulującego, a wskaźnik modulacji sygnału FM jest proporcjonalny do maksymalnej bezwzględnej wartości całki sygnału modulującego.

Z kolei z definicji 11.2 i wzorów (11.12) i (11.13) wynika, że dewiacje częstotliwości sygnałów PM i FM są określone wzorami:

$$\Delta f_{\mathrm{PM}} = \frac{k_p}{2\pi} \left| \frac{\mathrm{d}x(t)}{\mathrm{d}t} \right|_{\max} , \tag{11.18}$$

$$\Delta f_{\mathrm{FM}} = \frac{k_f}{2\pi} |x(t)|_{\max} . \tag{11.19}$$

Wynika stąd, że dewiacja częstotliwości sygnału PM jest proporcjonalna do maksymalnej bezwzględnej wartości pochodnej sygnału modulującego, a dewiacja częstotliwości sygnału FM jest proporcjonalna do maksymalnej bezwzględnej wartości sygnału modulującego.

W praktycznych systemach PM i FM jest spełniony warunek $F \gg \Delta f$. Częstotliwość chwilowa odchyla się zatem niewiele od częstotliwości nośnej. Ale nawet w ramach tych niewielkich odchyleń ich zakres ma decydujący wpływ na charakter sygnału zmodulowanego. Na wartość dewiacji częstotliwości można wpływać przez dobór parametru k_p (w przypadku modulacji PM) lub parametru k_f (w przypadku modulacji FM). Jeśli wartości tych współczynników są tak dobrane, że wskaźnik modulacji β danego systemu, określony wzorem (11.16) lub (11.17), spełnia warunek $\beta \ll 1$, modulację nazywamy *wąskopasmową* . . . Jeśli warunek ten nie jest spełniony, modulację nazywamy *szerokopasmową* .

11.1.4. Sygnał wąskopasmowy PM

W przypadku wąskopasmowej modulacji PM wskaźnik modulacji $\beta_{\mathrm{PM}} = k_p|x(t)|_{\max} \ll 1$. Funkcję modulującą (11.2) można wówczas przybliżyć dwoma pierwszymi wyrazami jej rozwinięcia w szereg potęgowy $m(t) = \mathrm{e}^{\mathrm{j}k_p x(t)} \approx 1 + \mathrm{j}\, k_p x(t)$. W konsekwencji wzory (11.3) i (11.4) przybierają następujące przybliżone postacie:

$$z_{\mathrm{PM}}(t) \approx Y_0 \left[1 + \mathrm{j}\, k_p x(t)\right] \mathrm{e}^{\mathrm{j}\Omega t}, \tag{11.20}$$

$$y_{\mathrm{PM}}(t) \approx Y_0 \cos \Omega t - Y_0 k_p x(t) \sin \Omega t\,. \tag{11.21}$$

Zauważmy, że struktura wzoru (11.21), opisującego sygnał PM w przypadku małych wartości wskaźnika modulacji, jest identyczna, jak wzoru (10.51) opisującego sygnał AM. Oba sygnały różnią się jedynie fazą wstęg bocznych. Wynika stąd, że dla małych wartości wskaźnika β_{PM} szerokość pasma sygnału PM jest równa w przybliżeniu szerokości pasma sygnału AM i określona wzorem:

$$B_{\mathrm{PM}} \approx 2 f_m, \tag{11.22}$$

gdzie f_m jest maksymalną częstotliwością widma sygnału modulującego. Jest to pasmo wąskie, co uzasadnia nazwę tego typu modulacji.

Mimo wyraźnych podobieństw modulacja wąskopasmowa PM i modulacja AM różnią się jednak w istotny sposób. W przeciwieństwie do modulacji AM modulacja wąskopasmowa PM należy do tzw. *modulacji kwadraturowych* (por. [3], p. 4.3.4-B)

11.2. Modulacja szerokopasmowa PM

11.2.1. Sygnał szerokopasmowy PM. Przypadek modulacji jednym tonem

Modulacja wąskopasmowa PM nie znajduje szerszego zastosowania w praktyce. Dla małych wartości wskaźnika modulacji β_{PM} nie uzyskujemy bowiem efektu rozszerzenia pasma w stosunku do sygnału AM i tym samym zwiększenia odporności sygnału PM na zakłócenia. Efekt ten występuje dopiero dla dostatecznie dużych wartości wskaźnika modulacji.

W przypadku ogólnym, tj. przy założeniu dowolnego sygnału modulującego $x(t)$, pełna analiza sygnału szerokopasmowego PM jest bardzo złożona. Wystarczający pogląd na strukturę czasową i częstotliwościową sygnału szerokopasmowego PM można jednak uzyskać rozpatrując przypadki modulacji jednym i dwoma tonami.

Załóżmy, że w systemie PM falę nośną modulujemy sygnałem harmonicznym $x(t) = X_0 \sin \omega_0 t$ o pulsacji $\omega_0 \ll \Omega$. Sygnał zmodulowany (11.4) przybiera wówczas postać:

$$y_{\text{PM}}(t) = Y_0 \cos\left(\Omega t + k_p X_0 \sin \omega_0 t\right) . \tag{11.23}$$

Kąt chwilowy i częstotliwość chwilowa sygnału PM zmodulowanego jednym tonem zmieniają się zatem w czasie zgodnie z zależnościami:

$$\psi_{\text{PM}}(t) = \Omega t + k_p X_0 \sin \omega_0 t, \tag{11.24}$$

$$\begin{aligned} f_{\text{PM}}(t) = \frac{1}{2\pi}\frac{\mathrm{d}\psi_{\text{PM}}(t)}{\mathrm{d}t} &= F + \frac{k_p X_0 \omega_0}{2\pi}\cos \omega_0 t = \\ &= F + k_p X_0 f_0 \cos \omega_0 t\,. \end{aligned} \tag{11.25}$$

Wynika stąd, że wskaźnik modulacji i dewiacja częstotliwości są określone wzorami:

$$\beta_{\text{PM}} = k_p X_0, \tag{11.26}$$

$$\Delta f_{\text{PM}} = k_p X_0 f_0 = \beta_{\text{PM}} f_0\,. \tag{11.27}$$

Przy ustalonych wartościach parametru k_p i amplitudy sygnału modulującego X_0 wskaźnik modulacji β_{PM} sygnału PM zmodulowanego jednym tonem jest wielkością stałą, niezależną od częstotliwości f_0 sygnału modulującego, natomiast jego dewiacja częstotliwości Δf_{PM} wzrasta proporcjonalnie ze wzrostem częstotliwości f_0.

11.2.2. Widmo sygnału PM zmodulowanego jednym tonem

Sygnał analityczny (11.3) sygnału PM w przypadku modulacji jednym tonem przybiera postać (dla wygody pomijamy indeks PM w oznaczeniu wskaźnika modulacji):

$$z_{\text{PM}}(t) = Y_0\, \mathrm{e}^{\mathrm{j}[\Omega t + k_p X_0 \sin \omega_0 t]} = Y_0\, \mathrm{e}^{\mathrm{j}[\Omega t + \beta \sin \omega_0 t]} = Y_0\, \mathrm{e}^{\mathrm{j}\beta \sin \omega_0 t}\, \mathrm{e}^{\mathrm{j}\Omega t}\,. \tag{11.28}$$

Ponieważ funkcja $\mathrm{e}^{\mathrm{j}\beta \sin \omega_0 t}$ występująca we wzorze (11.28) jest funkcją okresową o okresie $T_0 = 2\pi/\omega_0$, można ją rozwinąć w zespolony szereg Fouriera (2.57) o współczynnikach (2.58). W analizie matematycznej dowodzi się, że k-ty współczynnik tego szeregu jest wartością funkcji Bessela $J_k(x)$ pierwszego rodzaju i k-tego rzędu w punkcie x równym wartości wskaźnika modulacji β. Zespolony szereg Fouriera funkcji $\mathrm{e}^{\mathrm{j}\beta \sin \omega_0 t}$ ma zatem postać:

$$\mathrm{e}^{\mathrm{j}\beta \sin \omega_0 t} = \sum_{k=-\infty}^{\infty} J_k(\beta)\, \mathrm{e}^{\mathrm{j}k\omega_0 t}, \tag{11.29}$$

gdzie funkcje Bessela są określone wzorem:

$$J_k(x) = \sum_{n=0}^{\infty} (-1)^n \frac{\left(\frac{x}{2}\right)^{2n+k}}{n!(n+k)!}\,. \tag{11.30}$$

Uwzględniając szereg (11.29) we wzorze (11.28), otrzymujemy:

$$z_{\mathrm{PM}}(t) = Y_0\, \mathrm{e}^{\mathrm{j}\Omega t} \sum_{k=-\infty}^{\infty} J_k(\beta)\, \mathrm{e}^{\mathrm{j}k\omega_0 t} = Y_0 \sum_{k=-\infty}^{\infty} J_k(\beta)\, \mathrm{e}^{\mathrm{j}(\Omega + k\omega_0)t}, \tag{11.31}$$

a po przejściu do postaci rzeczywistej:

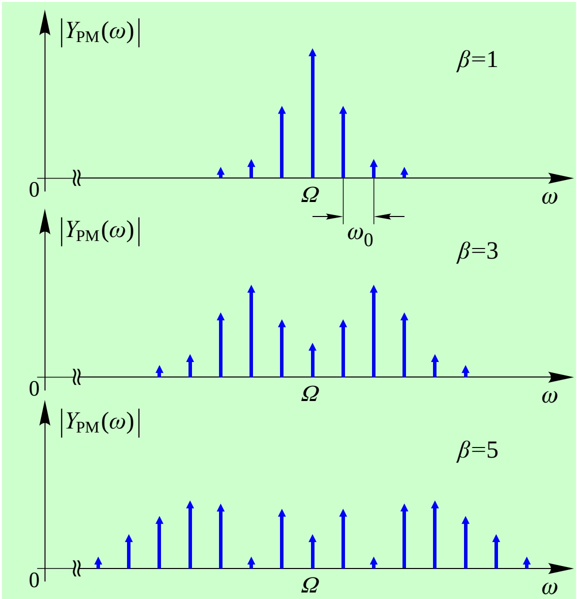
$$y_{\mathrm{PM}}(t) = Y_0 \sum_{k=-\infty}^{\infty} J_k(\beta) \cos(\Omega + k\omega_0)t\,. \tag{11.32}$$

Ponieważ dla funkcji Bessela (11.30) zachodzą równości: $J_k(\beta) = J_{-k}(\beta)$ dla k parzystych oraz $J_k(\beta) = -J_{-k}(\beta)$ dla k nieparzystych, wzór (11.32) można zapisać w równoważnej rozwiniętej postaci:

$$\begin{aligned} y_{\mathrm{PM}}(t) \; = \; Y_0 \Big\{ & J_0(\beta) \cos \Omega t + J_1(\beta) \left[\cos(\Omega + \omega_0)t - \cos(\Omega - \omega_0)t\right] + \\ & + J_2(\beta) \left[\cos(\Omega + 2\omega_0)t + \cos(\Omega - 2\omega_0)t\right] + \\ & + J_3(\beta) \left[\cos(\Omega + 3\omega_0)t - \cos(\Omega - 3\omega_0)t\right] \ldots \Big\}\,. \end{aligned} \tag{11.33}$$

Ze wzoru (11.33) wynika wprost struktura czasowa i częstotliwościowa sygnału PM zmodulowanego jednym tonem. Sygnał ten jest nieskończoną sumą sygnałów harmonicznych o dyskretnych pulsacjach $\Omega \pm k\omega_0$, $k = 0{,}1, \ldots$, zawierającą falę nośną o pulsacji Ω oraz nieskończoną liczbę par fal bocznych o pulsacjach sumacyjnych i różnicowych $\Omega \pm k\omega_0$, $k = 1{,}2, \ldots$ Amplituda fali nośnej jest określona przez wartość $J_0(\beta)$ funkcji Bessela zerowego rzędu w punkcie β, zaś amplitudy k-tej pary fal bocznych przez wartości $J_k(\beta)$ funkcji Bessela rzędu k w punkcie β, przy czym dla parzystych k fazy fal bocznych są zgodne, a dla nieparzystych k – przeciwne.

Ze struktury czasowej sygnału PM zmodulowanego jednym tonem wynika prążkowy charakter jego widma. Widmo amplitudowe sygnału PM o składa się z prążków nośnych w punktach $\pm\Omega$, wokół których występuje nieskończona liczba par prążków bocznych w punktach $\pm(\Omega \pm k\omega_0)$, przy czym dla ustalonego k prążki boczne mają jednakową wysokość. Układ tych prążków, a zarazem udział poszczególnych składowych harmonicznych w strukturze sygnału PM, zależy silnie od wartości wskaźnika modulacji β. Na rys. 11.2 pokazano przykładowe prawostronne widma amplitudowe sygnału PM zmodulowanego jednym tonem dla kilku wartości wskaźnika modulacji β i ustalonych wartości pulsacji nośnej Ω i pulsacji modulującej ω_0.

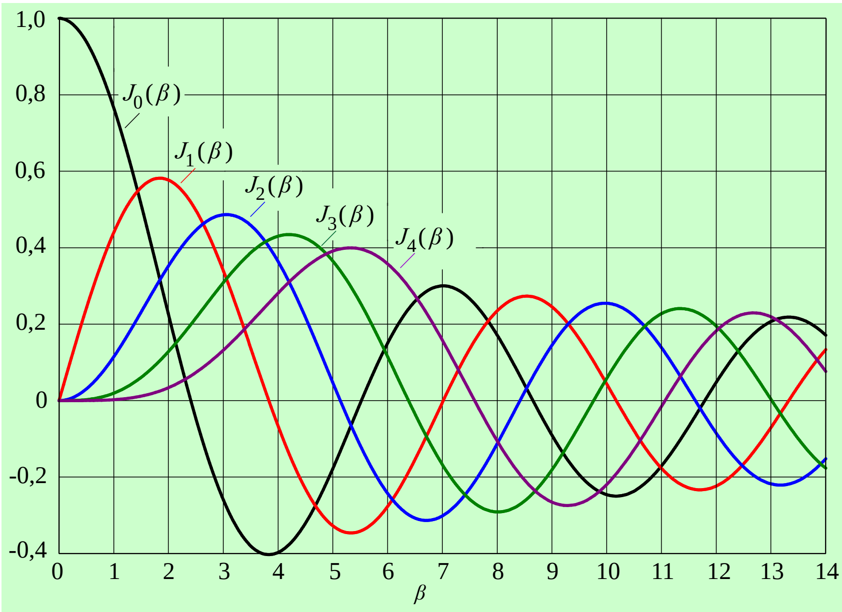


Rys. 11.2. Unormowane prawostronne widma amplitudowe sygnału PM zmodulowanego jednym tonem dla różnych wartości wskaźnika modulacji β

11.2.3. Szerokość pasma sygnału PM zmodulowanego jednym tonem

Ze wzoru (11.33) wynika, że sygnał PM ma teoretycznie pasmo nieograniczone. Z dokładnej analizy, uwzględniającej właściwości funkcji Bessela, wynika jednak, że prawie cała moc sygnału PM jest zawarta w paśmie ograniczonym. Dla zastosowań praktycznych ważne znaczenie ma oszacowanie szerokości tego pasma w zależności od wartości wskaźnika modulacji β.

Wykresy funkcji Bessela kilku pierwszych rzędów dla dodatnich wartości zmiennej β są pokazane na rys. 11.3. Wszystkie funkcje Bessela, poza funkcją zerowego rzędu, przybierają w punkcie $\beta = 0$ wartość równą zeru, przy czym im rząd funkcji jest wyższy, tym dłuższy jest odcinek w otoczeniu punktu $\beta = 0$, w którym wartości funkcji są bardzo małe (po którym funkcja „odrywa się” od osi odciętych). Dla $\beta < 0{,}1$ znaczące wartości mają tylko funkcje rzędu zerowego i pierwszego. Wynika stąd, że sygnał PM zawiera w tym przypadku tylko składową nośną i jedną parę składowych bocznych o pulsacjach $\Omega \pm \omega_0$. Pozostałe składowe boczne są pomijalne. Przypadek ten odpowiada modulacji wąskopasmowej PM.



Rys. 11.3. Funkcje Bessela pierwszego rodzaju

Dla większych wartości β znaczące wartości przybierają także funkcje Bessela wyższych rzędów, ale dla każdego β można znaleźć takie K, że dla $k > K$ wartości te można uznać za nieznaczące. Na przykład, dla $\beta = 3$ mamy (por. rys. 11.2b): $J_0(3) = -0{,}2601$, $J_1(3) = 0{,}3391$, $J_2(3) = 0{,}4861$, $J_3(3) = 0{,}3091$, $J_4(3) = 0{,}1320$, $J_5(3) = 0{,}0430$, $J_6(3) = 0{,}0114$, $J_7(3) = 0{,}0025$, $J_8(3) = 0{,}0005, \ldots$ Dalsze wartości są jeszcze mniejsze i bardzo szybko maleją ze wzrostem rzędu k. Oznacza to, że dla rzędów k większych od pewnej liczby K, ustalonej na podstawie określonego kryterium, prążki boczne widma sygnału PM zmodulowanego jednym tonem są na tyle małe, że można je pominąć. Jeżeli za nieznaczące przyjmiemy wartości funkcji Bessela mniejsze od 0,1 (co oznacza, jak dalej pokażemy, że moc wznoszona przez daną falę boczną do mocy całego sygnału PM jest mniejsza niż 1%), to dla $\beta = 3$ za pomijalne uznajemy wszystkie fale boczne o pulsacjach $\Omega \pm k\omega_0$, takich że $k > 4$. Jeśli przyjmiemy bardziej ostre kryterium i będziemy uznawać za nieznaczące wartości funkcji Bessela mniejsze od 0,01 (co odpowiada 0,1% mocy wnoszonej), to za pomijalne uważamy fale boczne o pulsacjach $\Omega \pm k\omega_0$, gdzie $k > 7$.

W praktyce, w celu oszacowania efektywnej szerokości pasma sygnału PM zmodulowanego jednym tonem, przyjmuje się jeszcze ostrzejsze kryterium oparte na stosunku sumarycznej mocy wnoszonej przez falę nośną oraz przez pierwszych K znaczących par fal bocznych do mocy całego sygnału. Za znaczące uważa się te fale, których sumaryczna moc przekroczy 99% mocy całkowitej.

Ze wzoru (11.33) wynika, że moce wnoszone przez każdą parę fal bocznych wynoszą $Y_0^2 J_k^2(\beta)$. Całkowita moc sygnału PM zmodulowanego jednym tonem jest więc określona wzorem:

$$P_y = \frac{1}{2} Y_0^2 \left[J_0^2(\beta) + 2 \sum_{k=1}^{\infty} J_k^2(\beta) \right]. \tag{11.34}$$

Ponieważ z teorii funkcji Bessela wiadomo, że dla każdego β

$$J_0^2(\beta) + 2 \sum_{k=1}^{\infty} J_k^2(\beta) = 1,$$

zatem całkowita moc sygnału PM wynosi:

$$P_y = \frac{1}{2} Y_0^2 \tag{11.35}$$

i jest równa mocy niezmodulowanej fali nośnej. Z uwagi na stałość w czasie amplitudy chwilowej sygnału PM wyniku tego można było się intuicyjnie spodziewać. Tak więc, jeśli przyjmiemy kryterium 99% mocy sumarycznej, to liczbę znaczących par fal bocznych przy ustalonej wartości wskaźnika modulacji β wyznaczamy jako najmniejszą liczbę K spełniającą nierówność:

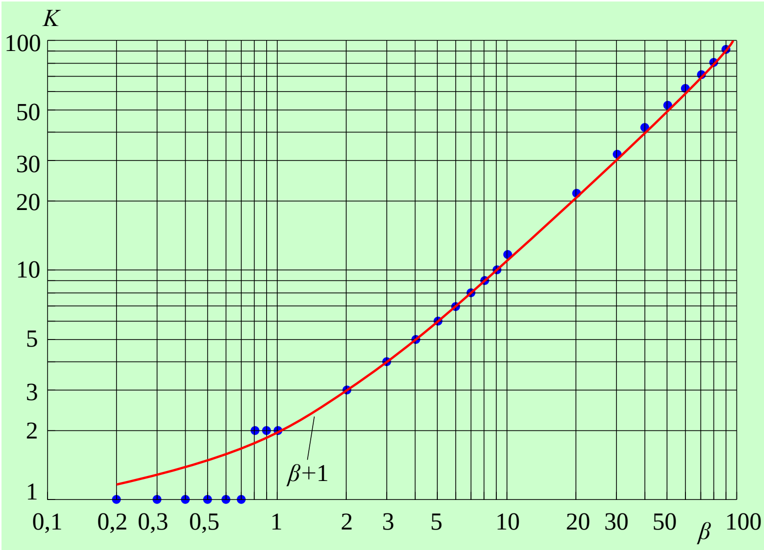
$$J_0^2(\beta) + 2 \sum_{k=1}^{K} J_k^2(\beta) \geqslant 0{,}99\,. \tag{11.36}$$

Na podstawie analizy numerycznej nierówności (11.36) na rys. 11.4 wykreślono w skali logarytmicznej zależność $K(\beta)$. Dodatkowo na rysunku tym wykreślono prostą $\beta + 1$. Jak widzimy, dla $\beta > 0{,}7$ wyznaczone punkty $K(\beta)$ leżą prawie dokładnie na prostej $\beta + 1$. Wnioskujemy stąd, że jeśli tylko $\beta > 0{,}7$, to przy przyjętym kryterium za znaczące prążki możemy uznać prążki zawarte w paśmie $\Omega - (\beta + 1)\omega_0 \leqslant \omega \leqslant \Omega + (\beta + 1)\omega_0$, gdzie ω_0 jest pulsacją sygnału modulującego. Szerokość tego pasma wynosi $2(\beta+1)\omega_0$ i rośnie ze wzrostem β (por. rys. 11.2). Efektywną szerokość pasma sygnału PM zmodulowanego jednym tonem (wyrażoną w hercach) możemy zatem obliczyć w przybliżeniu ze wzoru:

$$B_{\mathrm{PM}} \approx 2(\beta_{\mathrm{PM}} + 1) f_0\,. \tag{11.37}$$

W przypadku, gdy wartość wskaźnika modulacji β_{PM} jest dostatecznie duża (umownie przyjmuje się $\beta_{\mathrm{PM}} > 10$), szerokość pasma sygnału PM można obliczać ze wzoru:

$$B_{\mathrm{PM}} \approx 2\beta_{\mathrm{PM}} f_0\,. \tag{11.38}$$



Rys. 11.4. Logarytmiczny wykres zależności liczby K znaczących par fal bocznych sygnału PM zmodulowanego jednym tonem w funkcji wskaźnika modulacji β

Ponieważ wskaźnik modulacji β_{PM} jest dla modulacji PM jednym tonem stały, niezależny od częstotliwości f_0 sygnału modulującego (por. wzór (11.26)), zatem szerokość pasma sygnału PM rośnie proporcjonalnie ze wzrostem tej częstotliwości. Zauważmy, że dla dostatecznie dużych wartości wskaźnika modulacji szerokość pasma sygnału PM jest w przybliżeniu β_{PM}-krotnie większa od szerokości pasma sygnału AM zmodulowanego tym samym sygnałem. Dokładniejszą analizę zależności szerokości pasma sygnału PM od częstotliwości sygnału modulującego można znaleźć w [2], p. 8.2.3.1.

11.2.4. Sprawność energetyczna systemu PM

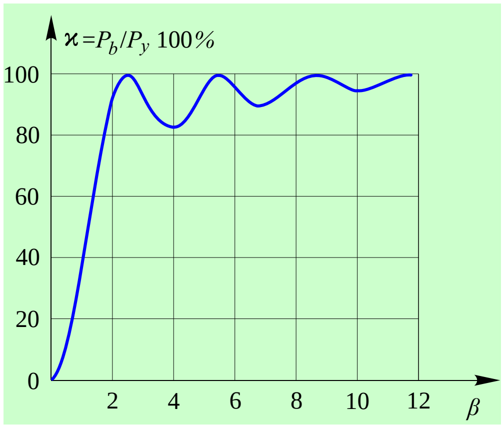
Z wykresu funkcji Bessela rzędu zerowego (rys. 11.3) wynika, że dla pewnych wartości wskaźnika modulacji β funkcja ta przybiera wartości równe zeru. Kilka pierwszych takich wartości β_i podano w tablicy 11.1. Ponieważ funkcja Bessela rzędu zerowego określa amplitudę fali nośnej, wynika stąd, że dla podanych w tablicy wartości β_i w sygnale PM zmodulowanym jednym tonem nie wystąpi fala nośna. Sprawność energetyczna modulacji PM (rozumiana w sensie definicji (10.59)) jest zatem dla tych wartości wskaźnika modulacji równa 100%.

Dla wartości β różnych od podanych w tablicy sprawność energetyczna modulacji PM jest oczywiście mniejsza od 100%. Z dokładnej analizy wynika jednak, że jeśli tylko $\beta > 2$, sprawność energetyczna systemów PM (a także FM) jest wysoka i dużo większa niż systemu AM. Pod tym względem systemy modulacji

Tablica 11.1. Pierwiastki równania $J_0(\beta) = 0$

i	1	2	3	4	5	6
β_i	2,405	5,520	8,654	11,792	14,931	18,071

kąta mają wyraźną przewagę. Wykres sprawności energetycznej systemu modulacji PM przy założeniu modulacji jednym tonem w zależności od wskaźnika modulacji β przedstawiono na rys. 11.5.



Rys. 11.5. Sprawność energetyczna systemu PM w przypadku modulacji jednym tonem w funkcji wskaźnika modulacji β

11.2.5. Przypadek modulacji dwoma tonami

Załóżmy teraz, że fala nośna jest zmodulowana w systemie PM sygnałem $x(t) = X_1 \sin \omega_1 t + X_2 \sin \omega_2 t$, $\omega_1 \neq \omega_2$. Sygnał zmodulowany przybiera w tym przypadku postać:

$$y_{\mathrm{PM}}(t) = Y_0 \cos(\Omega t + \beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t), \tag{11.39}$$

gdzie przez $\beta_1 = k_p X_1$ i $\beta_2 = k_p X_2$ oznaczono wskaźniki modulacji określone osobno dla obu składowych modulujących. Sygnał analityczny sygnału (11.39) jest określony wzorem:

$$z_{\mathrm{PM}}(t) = Y_0\, \mathrm{e}^{\mathrm{j}[\Omega t + \beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t]} = Y_0\, \mathrm{e}^{\mathrm{j}\beta_1 \sin \omega_1 t}\, \mathrm{e}^{\mathrm{j}\beta_2 \sin \omega_2 t}\, \mathrm{e}^{\mathrm{j}\Omega t}\,. \tag{11.40}$$

Korzystając z rozwinięć funkcji $e^{j\beta_1 \sin\omega_1 t}$ oraz $e^{j\beta_2 \sin\omega_2 t}$ w zespolony szereg Fouriera (11.29), przekształcimy ten wzór do postaci:

$$\begin{aligned} z_{PM}(t) &= Y_0 \left[\sum_{k=-\infty}^{\infty} J_k(\beta_1)\, e^{jk\omega_1 t}\right] \left[\sum_{l=-\infty}^{\infty} J_l(\beta_2)\, e^{jl\omega_2 t}\right] e^{j\Omega t} = \\ &= Y_0 \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} J_k(\beta_1) J_l(\beta_2)\, e^{j(\Omega + k\omega_1 + l\omega_2)t}\,. \end{aligned} \tag{11.41}$$

Powracając do zapisu rzeczywistego, otrzymujemy:

$$y_{PM}(t) = Y_0 \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} J_k(\beta_1) J_l(\beta_2) \cos(\Omega + k\omega_1 + l\omega_2)t\,. \tag{11.42}$$

Ze wzoru (11.42) wynika, że sygnał PM zmodulowany dwoma tonami zawiera nie tylko falę nośną o pulsacji Ω ($k, l = 0$) i fale boczne o pulsacjach $\Omega \pm k\omega_1$ i $\Omega \pm l\omega_2$, $k, l = 1, 2, \ldots$, związane z modulacją własną tonami ω_1 i ω_2, ale również fale boczne o *pulsacjach skrośnych* $\Omega \pm k\omega_1 \pm l\omega_2$, $k, l = 1, 2, \ldots$ Amplitudy poszczególnych fal bocznych są przy tym określone iloczynami wartości funkcji Bessela w punktach β_1 i β_2. W przypadku modulacji wieloma tonami efekt *modulacji skrośnej* potęguje się, a sygnał zmodulowany będzie zawierać fale boczne o pulsacjach będących wszystkimi kombinacjami liniowymi typu $\Omega \pm k\omega_1 \pm l\omega_2 \pm m\omega_3 \pm \ldots$. Efekt modulacji skrośnej występuje oczywiście także w przypadku modulacji sygnałem o widmie ciągłym, z tym że ma on wówczas charakter ciągły.

Należy podkreślić, że w przypadku modulacji AM wieloma tonami efekt modulacji skrośnej nie występuje. Sygnał ten zawiera bowiem tylko falę nośną i pary fal bocznych o pulsacjach $\Omega \pm \omega_k$, gdzie ω_k są pulsacjami poszczególnych tonów. Wynika stąd, że widma sygnałów PM (a także sygnałów FM) mają w porównaniu z widmami sygnałów AM strukturę znacznie bogatszą i urozmaiconą.

Ze względu na efekt modulacji skrośnej modulacje kątowe, w odróżnieniu od modulacji amplitudowych, są nazywane modulacjami *typu nieliniowego*. Pojęcia tego nie należy oczywiście mylić z charakterem samej operacji modulacji, która zarówno w przypadku modulacji kątowych, jak i amplitudowych, jest w swojej istocie operacją nieliniową.

11.2.6. Efektywna szerokość pasma sygnału PM zmodulowanego dowolnym sygnałem

Oszacowanie szerokości pasma zajmowanego przez sygnał PM w przypadku modulacji dowolnym sygnałem jest skomplikowane. Wzorować się tu można na analizie widma sygnału zmodulowanego jednym i dwoma tonami. Z uogólnienia tej analizy, popartej wynikami pomiarów doświadczalnych, wynika, iż dla wartości wskaźnika modulacji $\beta > 0{,}7$ dobre oszacowanie szerokości pasma sygnału

PM stanowi najczęściej stosowany w tym celu *wzór Carsona*:

$$B_{\text{PM}} \approx 2(\beta_{\text{PM}} + 1) f_m, \tag{11.43}$$

w którym f_m oznacza największą częstotliwość widma sygnału modulującego. W przypadku dostatecznie dużych wartości wskaźnika modulacji ($\beta_{\text{PM}} > 10$) można przyjąć oszacowanie:

$$B_{\text{PM}} \approx 2\beta_{\text{PM}} f_m \,. \tag{11.44}$$

Oszacowania (11.43) i (11.44) są uogólnieniami oszacowań (11.37) (11.38) dla przypadku modulacji jednym tonem. Podobnie jak w tamtym przypadku, ze względu na stałą wartość wskaźnika β_{PM}, szerokość pasma sygnału PM rośnie proporcjonalnie do maksymalnej częstotliwości f_m widma sygnału modulującego. Z porównania wzorów (10.54) i (11.44) wynika, że do przesłania sygnału PM wymagane jest pasmo w przybliżeniu β_{PM}-krotnie większe niż w przypadku sygnału AM zmodulowanego tym samym sygnałem.

Zależność szerokości pasma sygnału PM od szerokości pasma sygnału modulującego stanowi istotną wadę modulacji PM. Powoduje ona, że w praktycznych zastosowaniach preferowana jest modulacja FM. Do kwestii tej powrócimy jeszcze w p. 11.3.3.

11.3. Generacja i demodulacja sygnałów zmodulowanych kątowo

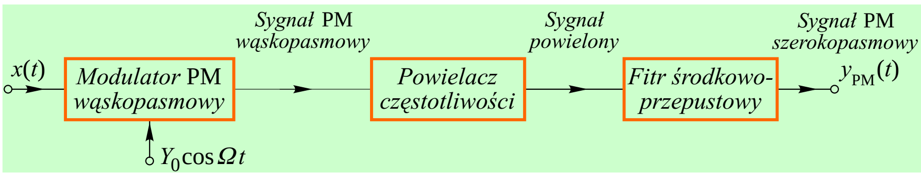
11.3.1. Generacja sygnału PM

Pierwszą historycznie metodą generacji sygnału PM była metoda zaproponowana na początku lat trzydziestych ubiegłego stulecia przez Armstronga. Koncepcja Armstronga polega na wytworzeniu szerokopasmowego sygnału PM p. 11.1.4 z wąskopasmowego sygnału PM przez odpowiednie przesunięcie widma sygnału wąskopasmowego w zakres dostatecznie dużych częstotliwości, tak aby wraz z przesunięciem widma następowało rozszerzenie jego pasma.

Wytworzenie wąskopasmowego sygnału PM nie nastręcza trudności. Należy w tym celu wykonać operacje wynikające ze wzoru (11.21) stosując np. generator Hartleya (rys. 10.19). Natomiast przesunięcie widma można zrealizować za pomocą układu nieliniowego o charakterystyce potęgowej $y_{\text{wy}}(t) = a[y_{\text{we}}(t)]^n$, gdzie n jest dostatecznie duże. Rozważmy przypadek $n = 2$, tj. układ podnoszący do kwadratu. Jeżeli na wejście tego układu podamy wąskopasmowy sygnał PM

$y_{\mathrm{we}}(t) = Y_0 \cos[\Omega t + k_p x(t)]$ o małej wartości parametru k_p, to na wyjściu wystąpi sygnał $y_{\mathrm{wy}}(t) = (a/2)Y_0^2\{1+\cos[2\Omega t+2k_p x(t)]\}$. Po usunięciu składowej stałej otrzymujemy sygnał PM o dwukrotnie większej pulsacji nośnej i dwukrotnie większym wskaźniku modulacji (por. wzór (11.16)). W ogólnym przypadku, gdy układ nieliniowy jest opisany charakterystyką potęgową rzędu n, po odfiltrowaniu zbędnych składowych otrzymamy sygnał o n krotnie większej pulsacji nośnej i n krotnie większym wskaźniku modulacji. Dobierając odpowiednio wartości parametrów Ω i k_p sygnału wąskopasmowego PM i rząd n charakterystyki elementu nieliniowego, można uzyskać sygnał szerokopasmowy PM o żądanym wskaźniku modulacji, a więc założonej szerokości pasma. Schemat blokowy modulatora Armstronga przedstawiono na rys. 11.6.

Modulator sygnału PM pracujący według metody Armstronga jest układem dość złożonym i wymaga zastosowania bardzo dokładnych elementów składowych, zwłaszcza doboru odpowiedniej charakterystyki elementu nieliniowego. Ze względu na to, że w metodzie Armstronga jest wytwarzany najpierw wąskopasmowy sygnał PM, jest ona nazywana metodą pośrednią generacji sygnału PM.



Rys. 11.6. Schemat blokowy szerokopasmowego modulatora PM Armstronga

Alternatywną metodą generacji szerokopasmowego sygnału PM jest zastosowanie *diody pojemnościowej* oraz oscylatora sterowanego napięciem zwanego *oscylatorem VCO* (ang. *Voltage Controlled Oscillator*). Oscylator VCO jest generatorem drgań harmonicznych, których częstotliwość jest zmieniana w odpowiedni sposób w zależności od bieżących wartości sygnału modulującego. Zasadę jego działania można wyjaśnić, odwołując się do zwykłego generatora sygnału harmonicznego o stałej pulsacji drgań. W przypadku, gdy generatorem takim jest np. układ trójnikowy Hartleya, zawierający dwie indukcyjności L_1 i L_2 oraz pojemność C (por. np. [4], p. 10.3.2), pulsacja generowanego sygnału harmonicznego jest określona wzorem $\Omega = 1/\sqrt{(L_1+L_2)C} = 1/\sqrt{LC}$, $L \triangleq L_1 + L_2$. Przy ustalonych wartościach elementów L_1, L_2, C pulsacja ta jest stała. Jeśli teraz jeden z elementów, np. pojemność C, będzie zmieniać się w czasie według pewnej funkcji $C = C(t)$, to także pulsacja generowanych drgań będzie ulegać zmianom w czasie. Zmiany pulsacji chwilowej drgań oscylatora są wówczas opisane wzorem:

$$\Omega(t) = \frac{1}{\sqrt{LC(t)}}\,. \tag{11.45}$$

Z drugiej strony, jak wynika ze wzoru (11.12), pulsacja chwilowa sygnału PM powinna zmieniać się wokół pulsacji nośnej Ω proporcjonalnie do pochodnej sygnału modulującego:

$$\Omega_{\text{PM}}(t) = \Omega + k_p \frac{\mathrm{d}x(t)}{\mathrm{d}t}\,. \tag{11.46}$$

Z porównania wzorów (11.45) i (11.46) wynika zatem, że musi zachodzić równość:

$$\Omega(t) = \frac{1}{\sqrt{LC(t)}} = \Omega + k_p \frac{\mathrm{d}x(t)}{\mathrm{d}t}\,. \tag{11.47}$$

Zbadamy warunki, przy których równość ta będzie spełniona. Załóżmy, że pojemność $C(t)$ zmienia się w czasie wokół pewnej stałej pojemności C_0 według funkcji $C(t) = C_0 + C_1(t)$. Mamy wówczas:

$$\Omega(t) = \frac{1}{\sqrt{LC(t)}} = \frac{1}{\sqrt{L[C_0 + C_1(t)]}} = \frac{1}{\sqrt{LC_0}} \frac{1}{\sqrt{1 + \dfrac{C_1(t)}{C_0}}}. \tag{11.48}$$

Załóżmy dalej, że zmiany pojemności wokół wartości C_0 są małe, tzn. $C_0 \gg \max_t |C_1(t)|$. Drugi czynnik we wzorze (11.48) możemy wtedy rozwinąć w szereg potęgowy i uwzględnić w nim tylko dwa pierwsze wyrazy. Otrzymamy następujące przybliżenie:

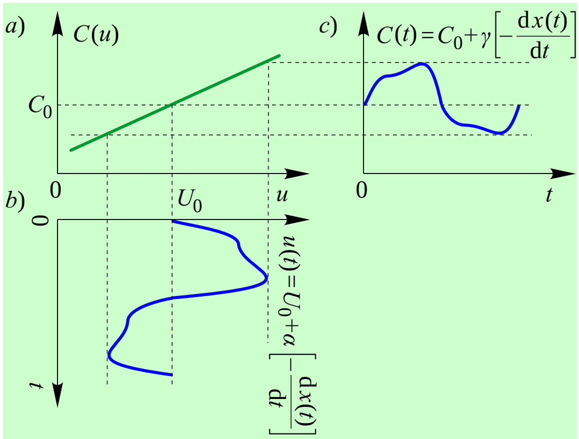
$$\Omega(t) \approx \frac{1}{\sqrt{LC_0}} \left[1 - \frac{C_1(t)}{2C_0}\right]. \tag{11.49}$$

Porównując wzory (11.47) i (11.49) widzimy, że równość (11.47) będzie spełniona z dobrym przybliżeniem, jeśli wartość C_0 będzie dobrana tak, aby $\Omega = 1/\sqrt{LC_0}$ oraz jeśli zmiany funkcji $C_1(t)$ wokół C_0 będą małe i proporcjonalne do pochodnej sygnału modulującego wziętej ze znakiem ujemnym:

$$C_1(t) \approx -\frac{2k_p C_0}{\Omega} \left[-\frac{\mathrm{d}x(t)}{\mathrm{d}t}\right]. \tag{11.50}$$

Wynika stąd, że w celu realizacji generatora sygnału PM należy dysponować elementem pojemnościowym sterowanym napięciem, którego pojemność w pewnym zakresie zmian napięcia sterującego zmienia się liniowo (rys. 11.7a). Jeżeli na zaciski takiego elementu podamy sygnał napięcia $u(t) = U_0 + \alpha[-\mathrm{d}x(t)/\mathrm{d}t]$ (rys. 11.7b), gdzie U_0 jest napięciem stałym, takim że $C(U_0) = C_0$, $\alpha > 0$ – współczynnikiem proporcjonalności, a $x(t)$ – sygnałem modulującym, to pojemność będzie zmieniać się wokół stałej wartości C_0 zgodnie z zależnością $C(t) = C_0 + \gamma[\,\mathrm{d}x(t)/\,\mathrm{d}t]$ (rys. 11.7c). Dla dostatecznie małych wartości współczynnika $\gamma = 2k_p C_0/\Omega$ spełnione będą zatem warunki, przy których zachodzi z dobrym przybliżeniem równość (11.47).

W celu wytworzenia pożądanego przebiegu napięcia sterującego $u(t)$ należy podać sygnał modulujący $x(t)$ o zmienionej polaryzacji na układ różniczkujący, dobrać odpowiednią wartość współczynnika α skalującego sygnał wyjściowy



Rys. 11.7. Element pojemnościowy o liniowej charakterystyce $C(u)$

i dodać do niego sygnał stały U_0. Natomiast jako element pojemnościowy o liniowej charakterystyce napięciowo-pojemnościowej $C(u)$ można wykorzystać diodę spolaryzowaną zaporowo. Pojemność złącza diodowego spolaryzowanego zaporowo zmienia się bowiem w pewnym zakresie tak, jak pokazano na rys. 11.7a. Wartości U_0 i C_0 określają stałonapięciowy punkt pracy diody. Dioda o takiej charakterystyce jest nazywana *diodą pojemnościową* (*waraktorem* lub *warikapem*).

11.3.2. Demodulacja sygnału PM

Do demodulacji wąskopasmowych sygnałów PM można stosować detektor koherentny (rys. 11.8). W przeciwieństwie do detektora koherentnego sygnału AM-SC (por. p. 10.2.4, rys. 10.7) faza lokalnej fali nośnej generowanej w odbiorniku musi różnić się w tym przypadku od fazy fali nośnej nadajnika o $\pi/2$. Detektor realizuje operację mnożenia sygnału PM (11.4) przez lokalną falę nośną $\sin\Omega t$. Sygnał iloczynowy jest przepuszczony przez filtr dolnoprzepustowy, który odfiltrowuje zbędne składowe skupione wokół pulsacji 2Ω. Na wyjściu filtru pozostaje składowa niskoczęstotliwościowa:

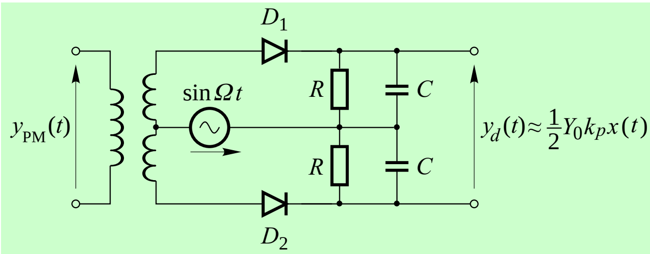
$$y_d(t) = \frac{1}{2}Y_0 \sin\left[k_p x(t)\right] . \tag{11.51}$$

Jeśli wskaźnik modulacji $\beta_{\text{PM}} = k_p|x(t)|_{\max}$ sygnału PM (por. wzór (11.16)) jest dostatecznie mały ($< 0{,}1$ radiana) to $\sin[k_p x(t)] \approx k_p x(t)$ i na wyjściu demodulatora otrzymujemy w przybliżeniu sygnał

$$y_d(t) \approx \frac{1}{2}Y_0 k_p x(t) \tag{11.52}$$

proporcjonalny do sygnału modulującego.

Detektor koherentny wymaga bardzo precyzyjnej kontroli fazy lokalnej fali nośnej odbiornika, tak aby była ona przesunięta dokładnie o $\pi/2$ względem fazy fali nośnej nadajnika. Detekcja koherentna sygnału PM ma więc tę samą wadę co w przypadku demodulacji sygnału AM-SC. W przypadku zbyt dużych wartości wskaźnika modulacji mogą wystąpić zniekształcenia nieliniowe sygnału zdemodulowanego. Liniowość demodulacji dla większych wartości wskaźnika modulacji (do około 1,5 radiana) można uzyskać kształtując odpowiednio przebiegi sygnałów na wejściach modulatora iloczynowego.

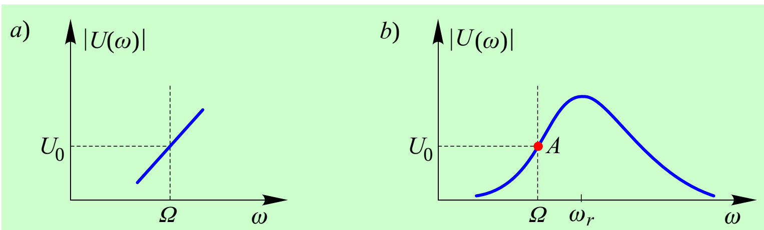


Rys. 11.8. Schemat blokowy koherentnego detektora wąskopasmowego sygnału PM

W przypadku szerokopasmowych sygnałów PM koncepcja demodulacji polega na przetworzeniu sygnału PM w odpowiedni sygnał AM, a następnie zdemodulowaniu sygnału AM w układzie detektora obwiedni. Zgodnie z tą koncepcją w demodulatorze wytwarzany jest sygnał napięciowy AM $u(t)$, którego amplituda chwilowa (obwiednia) $U(t)$ jest proporcjonalna do pulsacji chwilowej sygnału PM.

Komentarz. W rzeczywistości sygnał PM jest przetwarzany na złożony sygnał, będący jedynie w przybliżeniu sygnałem AM. Jednak informacja o sygnale modulującym jest zachowana w obwiedni tego sygnału.

W celu dokonania konwersji chwilowych zmian pulsacji sygnału PM na chwilowe zmiany amplitudy sygnału AM należy dysponować przetwornikiem „pulsacja-amplituda napięcia” o liniowej charakterystyce w pewnym zakresie zmian pulsacji. Ponieważ zmiany pulsacji chwilowej sygnału PM następują wokół pulsacji nośnej Ω, charakterystyka tego przetwornika powinna być liniowa w otoczeniu punktu Ω (rys. 11.9a).

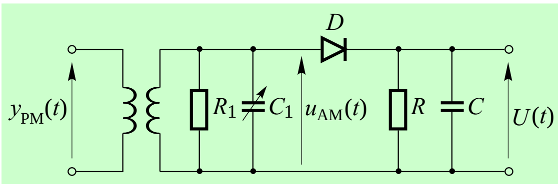


Rys. 11.9. Charakterystyka liniowego przetwornika „pulsacja-amplituda napięcia” (a) oraz charakterystyka amplitudowa równoległego obwodu rezonansowego (b)

Zauważmy, że funkcję takiego przetwornika może pełnić równoległy obwód rezonansowy, o charakterystyce „pulsacja-amplituda napięcia” (charakterystyce amplitudowej) pokazanej na rys. 11.9b (por. np. [5], p. 4.6.2). Pulsacja rezonansowa ω_r tego obwodu powinna być większa od pulsacji nośnej, a punkt pracy A przy braku modulacji powinien leżeć na najbardziej liniowym odcinku lewego zbocza tej charakterystyki. Jeżeli taki obwód rezonansowy będzie pobudzany sygnałem PM, to na jego zaciskach wystąpi sygnał napięcia, którego obwiednia jest w przybliżeniu opisana zależnością:

$$U(t) \approx \alpha \left[\Omega + k_p \frac{\mathrm{d}x(t)}{\mathrm{d}t} \right]. \tag{11.53}$$

Aby odzyskać sygnał modulujący $x(t)$, należy zdemodulować sygnał napięcia występujący na obwodzie rezonansowym za pomocą detektora obwiedni, a następnie, po usunięciu składowej stałej, otrzymany sygnał podać jeszcze na układ całkujący. Układ realizujący wszystkie te operacje nosi nazwę *dyskryminatora częstotliwości*. Jego schemat jest pokazany na rys. 11.10.

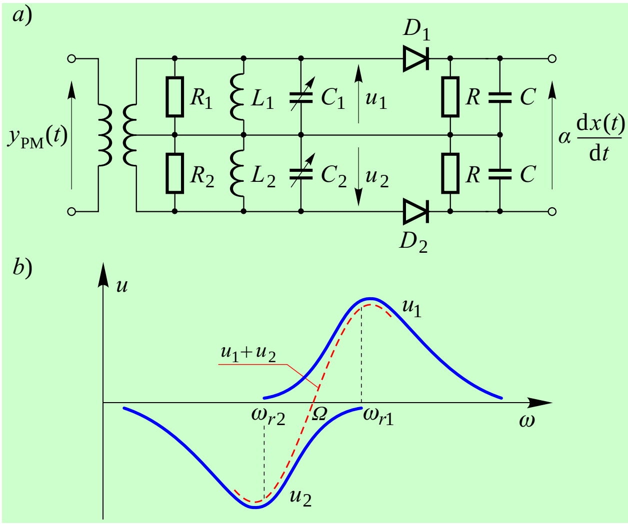


Rys. 11.10. Jednoobwodowy dyskryminator częstotliwości

Diodowy dyskryminator częstotliwości z pojedynczym obwodem rezonansowym charakteryzuje się małą czułością, tj. małym nachyleniem zbocza charakterystyki, oraz niewielkim zakresem liniowości tej charakterystyki w otoczeniu punktu Ω. Pod tym względem znacznie lepsze właściwości wykazuje dwuobwodowy *zrównoważony dyskryminator częstotliwości* (rys. 11.11a). Zawiera on dwa obwody rezonansowe pracujące w przeciwfazie o pulsacjach rezonansowych $\omega_{r1} > \Omega$ oraz $\omega_{r2} < \Omega$ (rys. 11.11b). Wypadkowa charakterystyka dyskryminatora jest zaznaczona linią przerywaną. Jak widać zakres liniowości charakterystyki w otoczeniu punktu Ω jest dużo większy, a ponadto zwiększa się jej nachylenie.

Na zaciskach obwodów występują sygnały napięciowe AM $u_1(t)$ i $u_2(t)$, które są demodulowane za pomocą detektorów obwiedni. Na wyjściu dyskryminatora wystąpi wypadkowy sygnał obwiedni proporcjonalny do pochodnej sygnału modulującego. Sygnał na wyjściu dyskryminatora należy jeszcze podać na układ całkujący. Dwuobwodowy diodowy dyskryminator częstotliwości ma nie tylko większą czułość i szerszy zakres liniowości, ale także – w przypadku prawidłowego zrównoważenia obwodów – eliminuje ewentualne zniekształcenia sygnału. Wymaga natomiast bardzo dokładnego zestrojenia obu obwodów.

Wadą diodowych dyskryminatorów częstotliwości jest konieczność stosowania elementów indukcyjnych, kłopotliwych w produkcji i wymagających dokładnego strojenia. Mimo to dyskryminatory częstotliwości były do niedawna stosowane powszechnie do demodulacji sygnałów PM i FM. Obecnie są one coraz częściej



Rys. 11.11. Dwuobwodowy dyskryminator częstotliwości (a) i jego charakterystyka (b)

zastępowane przez *demodulatory kwadraturowe* i *demodulatory z pętlą fazową PLL* (ang. *Phase-Locked Loop* – por. np. [1], p. 3.12) realizowane w technologii scalonej.

11.3.3. Porównanie systemów PM i FM

Jak stwierdziliśmy w punkcie 11.1.2, sygnał FM zmodulowany sygnałem $x(t)$ jest równoważny sygnałowi PM zmodulowanemu sygnałem całkowym $v(t) = \int x(t)\,\mathrm{d}t$. Oznacza to, że w celu dokonania analizy sygnału FM zmodulowanego sygnałem $x(t)$ można wykorzystać wyniki przeprowadzonej wyżej analizy sygnału PM przy założeniu, że jest on modulowany sygnałem całkowym $v(t)$. Z porównania tych dwóch sygnałów wynikać będą podobieństwa i zasadnicze różnice między systemami FM i PM.

Załóżmy najpierw, że parametr k_f jest tak dobrany, że wskaźnik modulacji systemu FM spełnia nierówność $\beta_{\mathrm{FM}} = k_f|\int x(t)\,\mathrm{d}t|_{\max} < 0{,}1$. Odpowiada to przypadkowi wąskopasmowej modulacji FM. Powtarzając rozumowanie z p. 11.1.4, wąskopasmowy sygnał FM możemy zapisać w postaci:

$$y_{\mathrm{FM}}(t) \approx Y_0 \cos \Omega t - Y_0 \left[k_f \int x(t)\,\mathrm{d}t\right] \sin \Omega t\,. \tag{11.54}$$

Zgodnie z twierdzeniem 3.7 o całkowaniu (wzór (3.29)) widma sygnałów $x(t)$ i $v(t)$ są związane zależnością $V(\omega) = X(\omega)/(\mathrm{j}\,\omega)$. Sygnał $x(t)$ i sygnał całkowy $v(t)$ mają więc taką samą pulsację graniczną. Wynika stąd, że szerokość pasma wąskopasmowego sygnału FM zmodulowanego sygnałem $x(t)$ jest taka sama, jak szerokość pasma wąskopasmowego sygnału PM zmodulowanego sygnałem całkowym $v(t)$ i jest określona wzorem (por. wzór (11.22)):

$$B_{\mathrm{FM}} \approx 2f_m, \tag{11.55}$$

gdzie f_m jest największą częstotliwością widma sygnału $x(t)$. W przypadku wąskopasmowych modulacji PM i FM zależność szerokości pasma sygnału od maksymalnej częstotliwości sygnału modulującego jest zatem identyczna.

Inaczej przedstawia się sytuacja w przypadku szerokopasmowych sygnałów PM i FM. Aby zbadać występujące różnice, wystarczy rozpatrzyć przypadek modulacji FM jednym tonem i uogólnić otrzymane rezultaty na przypadek modulacji dowolnym sygnałem. Załóżmy, że fala nośna jest modulowana w systemie FM sygnałem $x(t) = X_0 \cos \omega_0 t$. Zgodnie ze wzorem ogólnym (11.9) sygnał zmodulowany FM przybiera wówczas postać:

$$y_{\mathrm{FM}}(t) = Y_0 \cos\left(\Omega t + k_f \frac{X_0}{\omega_0} \sin \omega_0 t\right). \tag{11.56}$$

Jego kąt chwilowy i częstotliwość chwilowa zmieniają się według zależności:

$$\psi_{\mathrm{FM}}(t) = \Omega t + k_f \frac{X_0}{\omega_0} \sin \omega_0 t, \tag{11.57}$$

$$\begin{aligned} f_{\mathrm{FM}}(t) = \frac{1}{2\pi} \frac{\mathrm{d}\psi_{\mathrm{FM}}(t)}{\mathrm{d}t} &= \frac{1}{2\pi} (\Omega + k_f X_0 \cos \omega_0 t) = \\ &= F + \frac{k_f X_0}{2\pi} \cos \omega_0 t\,. \end{aligned} \tag{11.58}$$

Wskaźnik modulacji i dewiacja częstotliwości są zatem określone wzorami:

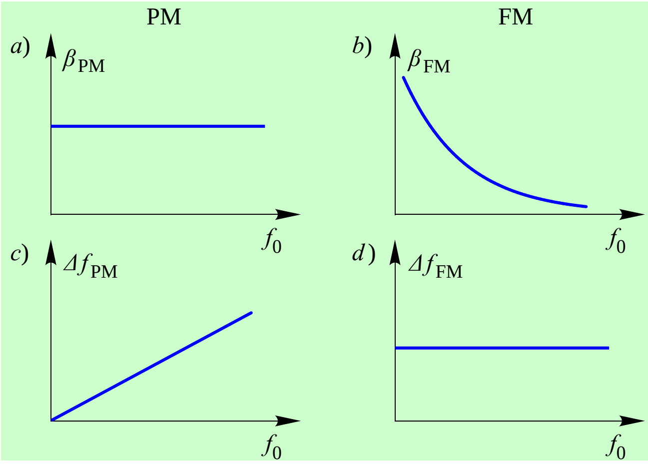
$$\beta_{\mathrm{FM}} = k_f \frac{X_0}{\omega_0}, \tag{11.59}$$

$$\Delta f_{\mathrm{FM}} = \frac{k_f X_0}{2\pi}\,. \tag{11.60}$$

Wynika stąd, że przy ustalonych wartościach parametru k_f i amplitudy sygnału modulującego X_0 wskaźnik modulacji β_{FM} sygnału FM zmodulowanego jednym tonem jest odwrotnie proporcjonalny do częstotliwości f_0 sygnału modulującego, natomiast jego dewiacja częstotliwości Δf_{FM} jest stała, niezależna od częstotliwości f_0. Przypomnijmy, że w przypadku sygnału PM zmodulowanego jednym tonem wskaźnik modulacji β_{PM} jest stały w funkcji częstotliwości f_0 sygnału modulującego, a dewiacja częstotliwości Δf_{PM} wzrasta proporcjonalnie do częstotliwości f_0. Wykresy zależności wskaźnika modulacji i dewiacji częstotliwości od częstotliwości f_0 sygnału modulującego dla sygnałów PM i FM zmodulowanych jednym tonem przedstawiono na rys. 11.12.

Ponieważ przyjęty sygnał modulujący i jego sygnał całkowy mają tę samą częstotliwość, szerokość pasma sygnału FM zmodulowanego jednym tonem można obliczać ze wzoru (11.37):

$$B_{\mathrm{FM}} \approx 2\,(\beta_{\mathrm{FM}} + 1)\, f_0, \tag{11.61}$$



Rys. 11.12. Wykresy wskaźnika modulacji (a, b) i dewiacji częstotliwości (c, d) sygnałów PM i FM zmodulowanych jednym tonem w funkcji częstotliwości f_0 sygnału modulującego

a dla dostatecznie dużych wartości wskaźnika modulacji ($\beta_{FM} > 10$) – ze wzoru (11.38):

$$B_{FM} \approx 2\beta_{FM} f_0 \,. \tag{11.62}$$

Korzystając z zależności (11.59) i (11.60), wzory (11.61) i (11.62) określające szerokość pasma sygnału FM możemy wyrazić przez dewiację częstotliwości:

$$B_{FM} \approx 2(\Delta f_{FM} + f_0), \tag{11.63}$$

$$B_{FM} \approx 2\Delta f_{FM} \,. \tag{11.64}$$

Wynika stąd, że dla dostatecznie dużych wartości wskaźnika modulacji ($\beta_{FM} > 10$) szerokość pasma sygnału FM zmodulowanego jednym tonem nie zależy od częstotliwości sygnału modulującego i jest równa w przybliżeniu podwojonej dewiacji częstotliwości.

Podobny wniosek jest słuszny w przypadku sygnału FM zmodulowanego dowolnym sygnałem o maksymalnej częstotliwości widma f_m. Jeśli wskaźnik modulacji jest dostatecznie duży, szerokość pasma sygnału FM jest niezależna od częstotliwości f_m i określona wzorem (11.64), a więc jest równa podwojonej wartości dewiacji częstotliwości. Ponieważ dewiacja częstotliwości w systemie FM jest w pełni określona jedynie przez parametr k_f modulatora oraz maksymalną bezwzględną wartość sygnału modulującego, zatem przez odpowiedni dobór wartości tych parametrów można z góry ustalić pożądaną szerokość pasma, niezależnie od cech widmowych sygnału modulującego. Właściwość ta w zasadniczy sposób odróżnia sygnał FM od sygnału PM, którego szerokość pasma rośnie proporcjonalnie do częstotliwości f_m.

Niezależność szerokości pasma sygnału FM od szerokości pasma sygnału modulującego ma decydujące znaczenie dla zastosowań praktycznych i z tego względu do transmisji sygnału fonii w systemach radiofonicznych w zakresie fal UKF oraz w systemach telewizyjnych powszechnego użytku jest stosowana wyłącznie szerokopasmowa modulacja FM. Stosuje się ją także w radiotelefonii oraz mikrofalowych liniach radiowych naziemnych i satelitarnych. Modulację PM wykorzystuje się często jako etap pośredni generacji sygnału FM. W warunkach europejskich sygnały radiowe FM są przesyłane w zakresie częstotliwości 87,5–108 MHz, przy czym na każdy z nich jest przeznaczone pasmo o szerokości 0,2? MHz.

11.3.4. Porównanie sygnałów zmodulowanych amplitudowo i kątowo

Sygnały zmodulowane amplitudowo mają węższe pasmo. Z jednej strony jest to ich zaletą (jeśli w danym paśmie chcemy przesyłać jak najwięcej sygnałów), z drugiej zaś – wadą (jeśli chcemy uzyskać wysoką jakość transmisji). Ze względu na prostotę układów odbiorczych modulacja AM jest powszechnie stosowana do transmisji sygnałów radiowych w systemach radiofonicznych powszechnego użytku w zakresie fal średnich i krótkich. Sygnały AM są jednak mało odporne na zakłócenia, w szczególności na zjawisko zaniku selektywnego, które jest bardziej odczuwalne w zakresie fal krótkich. Dlatego w tym zakresie jest stosowana niekiedy modulacja AM-SC bez fali nośnej.

Ze względu na dwukrotnie mniejsze pasmo i mniejszą moc wymaganą do transmisji, w systemach radiowych specjalnego przeznaczenia (np. łączności krótkofalarskiej, systemach radiotelefonicznych służb miejskich, takich jak policja, straż pożarna itp.) jest często stosowana modulacja SSB-SC. Umożliwia ona około 16-krotny zysk mocy w porównaniu z modulacją AM przy założeniu takiego samego zasięgu. Jednakże realizacja modulatorów i demodulatorów w tych systemach wymaga znacznie większych nakładów sprzętowych.

Wyraźną poprawę odporności na zakłócenia, a tym samym poprawę jakości transmisji, uzyskuje się w systemach modulacji kąta. Sygnały zmodulowane kątowo wymagają jednak do transmisji znacznie szerszego pasma i bardziej złożonych układów modulatorów i demodulatorów. W przeciwieństwie do sygnałów PM, szerokość pasma sygnałów FM jest praktycznie stała, niezależna od pasma sygnału modulującego. Z tego względu w praktyce jest zwykle stosowana modulacja FM. Sygnał FM jest jednak bardziej wrażliwy na zakłócenia wynikające z wielodrogowości propagacji (np. w wyniku odbić od budynków, kompleksów leśnych itp.), szczególnie dokuczliwe podczas jazdy samochodem.

Słownik

dewiacja częstotliwości
maksymalna bezwzględna odchyłka częstotliwości chwilowej sygnału od częstotliwości nośnej

dewiacja fazy (wskaźnik, indeks modulacji)
maksymalna bezwzględna wartość fazy chwilowej sygnału (maksymalna bezwzględna odchyłka kąta chwilowego sygnału od liniowej zmiany tego kąta)

dyskryminator częstotliwości
układ demodulatora sygnałów zmodulowanych kątowo

mieszacz
układ realizujący przemianę częstotliwości

modulacja fazy PM
modulacja analogowa, w której w zależności od bieżących wartości sygnału informacyjnego uzmienniana jest bezpośrednio faza harmonicznej fali nośnej

modulacja kąta
modulacja analogowa, w której w zależności od bieżących wartości sygnału informacyjnego uzmienniany jest kąt harmonicznej fali nośnej; obejmuje modulację fazy i modulację częstotliwości

modulacja skrośna
efekt powstawania składowych harmonicznych o częstotliwościach sumacyjnych i różnicowych (skrośnych) występujący w modulacjach kąta wieloma tonami

Literatura

[1] Haykin S.: Systemy telekomunikacyjne. WKiŁ, Warszawa, 1999.

[2] Szabatin J.: Podstawy teorii sygnałów. WKiŁ, Warszawa, wyd. 3, 2000.

[3] Szabatin J., Radecki K. (red.): Teoria sygnałów i modulacji, ćwiczenia laboratoryjne. OWPW, Warszawa, 2001.

[4] Filipkowski A.: Uklady elektroniczne analogowe i cyfrowe. WNT, Warszawa, 1993.

[5] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom II. WNT, Warszawa, wyd. 3, 1998.

Lekcja 12

Modulacje impulsowe

W przeciwieństwie do analogowych systemów modulacji, w których parametry harmonicznej fali nośnej są zmieniane w czasie w sposób ciągły, w impulsowych systemach modulacji zmiany jednego z parametrów fali nośnej następują w czasie skokowo. Parametr ten jest uzmienniany w zależności od wartości kolejnych próbek sygnału informacyjnego.

Podstawowym ciągiem impulsów, pełniącym w impulsowych systemach modulacji rolę fali nośnej, jest unipolarna fala prostokątna (rys. 9.2b). W zależności od rodzaju modulacji zmianom, zgodnym z wartościami bieżących próbek sygnału informacyjnego, podlega albo amplituda impulsów albo ich czas trwania albo położenie. Zmiany te zachodzą w dyskretnych chwilach czasu, ale uzmienniany parametr może przybierać wartości w zbiorze ciągłym lub w zbiorze dyskretnym. Jeśli modulowany parametr fali nośnej przybiera w kolejnych dyskretnych chwilach wartości ze zbioru ciągłego, modulację nazywamy *modulacją impulsową analogową*. Jeśli zbiór tych wartości jest skończony, co umożliwia ich zakodowanie i transmisję impulsów w postaci cyfrowej jako ciągu *impulsów kodowych*, modulację nazywamy *modulacją impulsową cyfrową* lub *modulacją impulsowo-kodową*. Ten drugi rodzaj modulacji impulsowej można traktować jako etap pośredni między impulsowymi a cyfrowymi systemami modulacji.

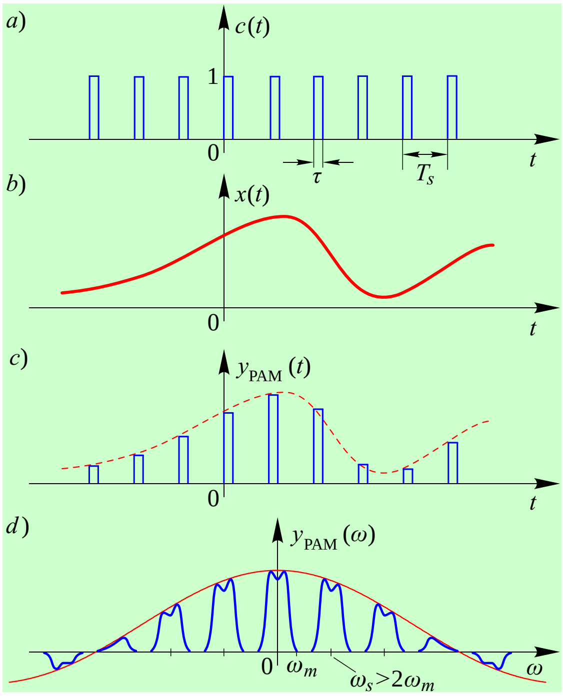
W lekcji omówimy trzy podstawowe analogowe modulacje impulsowe: PAM, PDM i PPM oraz cyfrowe modulacje impulsowe: PCM i modulację przyrostową delta (DM).

Przedyskutujemy zasady generacji i demodulacji sygnałów w każdym z tych systemów modulacji. Rozpatrzymy także sposoby fizycznej reprezentacji znaków binarnych w modulacjach impulsowo-kodowych.

12.1. Modulacja PAM

12.1.1. Sygnał PAM

Modulacja amplitudy impulsów PAM (ang. *Pulse Amplitude Modulation*) jest najprostszą z analogowych modulacji impulsowych. Polega ona na uzmiennieniu amplitudy impulsów unipolarnej prostokątnej fali nośnej o okresie T_s, czasie trwania τ i jednostkowej amplitudzie (rys. 12.1a), zgodnie z wartościami bieżących próbek sygnału modulującego $x(t)$ pobieranych w chwilach nT_s (rys. 12.1b). W efekcie powstaje ciąg równoodległych impulsów prostokątnych o jednakowym czasie trwania τ i amplitudach równych próbkom $x(nT_s)$ sygnału modulującego (rys. 12.1c). Należy jednak podkreślić, że do wytworzenia sygnału PAM można stosować także ciągi wąskich impulsów o kształcie innym niż prostokątny.



Rys. 12.1. Fala nośna sygnału PAM (a), sygnał modulujący (b), sygnał zmodulowany PAM (c) i jego widmo (d)

Z określenia sygnału PAM wynika, że w celu jego wytworzenia należy spróbkować sygnał informacyjny $x(t)$ w systemie próbkowania chwilowego omówionym w p. 5.3.2-B. Przy założeniu prostokątnego kształtu impulsów w wyniku tej operacji otrzymujemy sygnał PAM

$$y_{\mathrm{PAM}}(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \sqcap \left(\frac{t - \tau/2 - nT_s}{\tau} \right). \tag{12.1}$$

Jego widmo jest określone wzorem (por. wzór (6.30)):

$$Y_{\mathrm{PAM}}(\omega) = \frac{\tau}{T_s} \operatorname{Sa}(\omega\tau/2) \, \mathrm{e}^{-\mathrm{j}(\omega\tau/2)} \sum_{n=-\infty}^{\infty} X(\omega - n\omega_s), \tag{12.2}$$

gdzie $X(\omega)$ jest widmem sygnału modulującego $x(t)$, a $\omega_s = 2\pi/T_s$ – pulsacją próbkowania. Jeśli okres T_s fali nośnej jest dobrany tak, że $\omega_s > 2\omega_m$, gdzie ω_m jest największą pulsacją widma sygnału modulującego, to widmo sygnału PAM jest ciągiem kopii widma $X(\omega)$ odseparowanych od siebie pewnymi pasmami wolnymi i zniekształconych multiplikatywnie obwiednią o kształcie funkcji Sa (rys. 12.1d).

12.1.2. Odtwarzanie sygnału informacyjnego z sygnału PAM. Efekt aperturowy

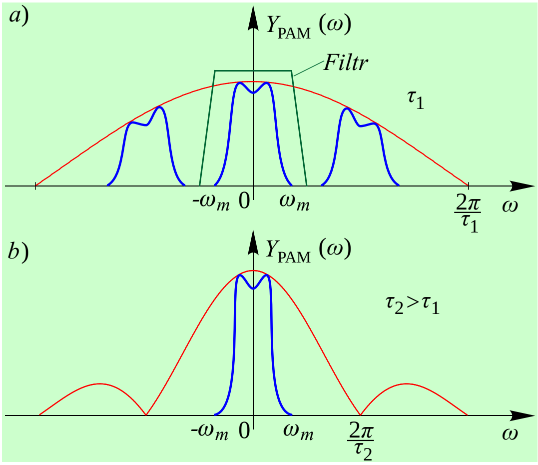
Jeśli sygnał PAM jest transmitowany w swoim paśmie podstawowym, to w celu odtworzenia po stronie odbiorczej sygnału informacyjnego $x(t)$ należy zastosować filtr dolnoprzepustowy o płaskiej charakterystyce w przedziale $|\omega| \leqslant \omega_m$, który wydziela z widma sygnału odebranego jego środkowy segment (rys. 12.2a). Jednak nawet wówczas, gdy charakterystyka filtru będzie idealnie płaska w tym przedziale, odzyskane widmo będzie zniekształcone amplitudowo obwiednią

$$\tau |\operatorname{Sa}(\omega\tau/2)| \tag{12.3}$$

wynikającą ze skończonego czasu trwania τ impulsów fali nośnej (por. 6.4.2). Najbardziej zniekształcane są składowe widmowe leżące na krańcach przedziału $|\omega| \leqslant \omega_m$. Tego typu zniekształcenia amplitudowe sygnału noszą nazwę *efektu aperturowego*. Efekt ten występuje również w przypadku innych fal nośnych niż fala prostokątna. Odmienny będzie tylko kształt zniekształcającej obwiedni.

Zniekształcenia sygnału odtworzonego są tym większe, im dłuższe są impulsy fali nośnej. Zwiększanie długości impulsów pociąga bowiem za sobą zawężenie listka głównego obwiedni (12.3), w wyniku czego ma ona w przedziale $|\omega| \leqslant \omega_m$ bardziej stromy kształt (rys. 12.2b). Z drugiej jednak strony wydłużanie impulsów jest korzystne z uwagi na węższe pasmo zajęte przez sygnał PAM w kanale. Aby skompensować efekt aperturowy, sygnał z wyjścia filtru dolnoprzepustowego jest podawany na filtr korekcyjny o charakterystyce amplitudowej będącej odwrotnością obwiedni (12.3):

$$|H_{kor}(\mathrm{j}\,\omega)| = \frac{1}{\tau |\operatorname{Sa}(\omega\tau/2)|}. \tag{12.4}$$



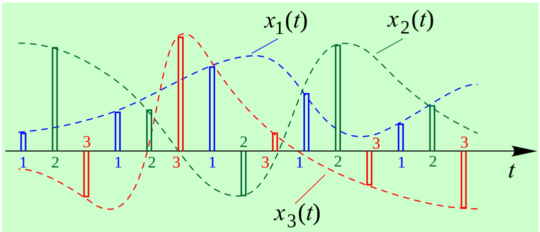
Rys. 12.2. Zniekształcenia aperturowe sygnału PAM w przypadku krótkich (a) i długich (b) impulsów fali nośnej

W praktyce, ze względu na sposób organizacji transmisji sygnałów PAM (por. p. 12.1.3), czas trwania τ impulsów fali nośnej dobiera się tak, aby jej współczynnik wypełnienia $\tau/T_s \ll 1$. Wówczas $2\pi/\tau \gg \omega_m$, a więc obwiednia (12.3) jest w przedziale $|\omega| \leqslant \omega_m$ płaska. Zniekształcenia aperturowe są w takim przypadku pomijalne i filtr korekcyjny nie jest stosowany.

12.1.3. Transmisja sygnału PAM w systemie zwielokrotnienia czasowego

W systemach modulacji analogowej informacja zawarta w sygnale modulującym jest transmitowana w czasie w sposób ciągły. W systemie PAM przekazywanie informacji odbywa się za pomocą krótkich impulsów o czasie trwania τ. System ten wymaga zatem znacznie mniejszych średnich mocy nadajnika. Sygnał PAM zajmuje jednak w kanale szerokie pasmo – tym szersze, im krótsze są impulsy fali nośnej (por. rys. 12.2). Sygnały PAM nie mogą być zatem przekazywane w systemach zwielokrotnienia częstotliwościowego, których zasadą jest rozłączność widmowa sygnałów transmitowanych przez jeden kanał nadawczy.

Wykorzystując fakt, że między kolejnymi impulsami sygnałów PAM występują długie przerwy w transmisji, sygnały te można przesyłać w systemach zwielokrotnienia czasowego (por. p. 9.2.1). Zasada zwielokrotnienia czasowego w odniesieniu do sygnałów PAM polega na tym, że w przerwach między impulsami jednego sygnału PAM są transmitowane przez ten sam kanał impulsy innych sygnałów PAM. Wprawdzie sygnały te mają pokrywające się pasma, ale ich impulsy są przekazywane w ściśle określonych rozłącznych oknach czasowych i mają uporządkowaną strukturę czasową pozostającą pod kontrolą odpowiedniego układu synchronizującego. Umożliwia to rozdzielenie w odbiorniku impulsów pochodzących od poszczególnych sygnałów na różne tory odbiorcze i odtworzenie sygnałów informacyjnych w tych torach za pomocą filtrów dolnoprzepustowych. Zasadę zwielokrotnienia czasowego sygnałów w przypadku stopnia zwielokrotnienia równego 3 zilustrowano na rys. 12.3.



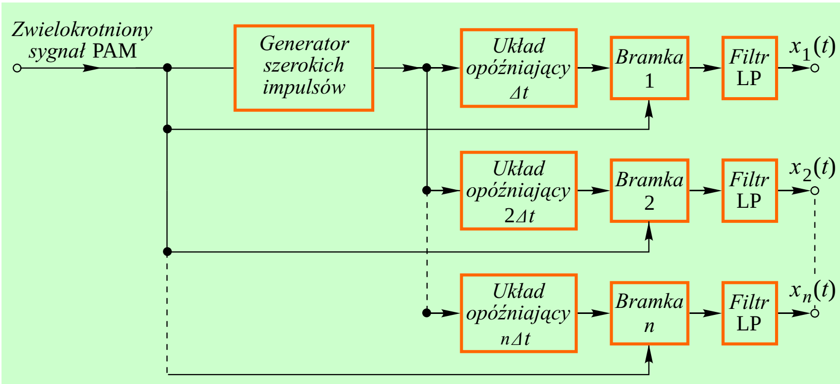
Rys. 12.3. Zwielokrotniony czasowo sygnał PAM

12.1.4. Demodulacja zwielokrotnionego czasowo sygnału PAM

Na rys. 12.4a pokazano schemat blokowy demodulatora zwielokrotnionego sygnału PAM. Przemieszane w czasie impulsy zwielokrotnionego sygnału PAM, o tym samym dla każdego z sygnałów okresie próbkowania T_s, są podawane na wejścia bramek. W każdej chwili nT_s w generatorze szerokich impulsów jest wytwarzany impuls o nieco dłuższym czasie trwania w porównaniu z czasem trwania impulsów informacyjnych. Impuls ten, po odpowiednim proporcjonalnie wzrastającym opóźnieniu $k\Delta t$, jest podawany na drugie wejścia bramek w poszczególnych torach odbiorczych. Na obu wejściach k-tej bramki następuje koincydencja czasowa szerszego impulsu opóźnionego o $k\Delta t$ oraz k-tego impulsu informacyjnego i impuls ten przechodzi na wyjście k-tej bramki. Jednocześnie bramka ta nie przepuszcza impulsów pochodzących od pozostałych transmitowanych sygnałów. Tym samym następuje rozdzielenie impulsów na poszczególne tory odbiorcze. Po filtracji dolnoprzepustowej na poszczególnych wyjściach otrzymuje się odtworzone sygnały informacyjne.

12.1.5. Modulacja PAM-AM

Sygnał PAM ma teoretycznie pasmo nieskończone. Większość jego mocy jest jednak skupiona w paśmie $|\omega| \leqslant 2\pi/\tau$ obejmującym listek główny widma (por. rys. 12.2). W widmie sygnału PAM dominują więc niekorzystne z punktu widzenia transmisji składowe niskoczęstotliwościowe. Aby uniknąć kłopotów z transmisją składowej stałej i składowych o niskich częstotliwościach, sygnał PAM jest



Rys. 12.4. Schemat blokowy demodulatora zwielokrotnionego sygnału PAM

często poddawany dodatkowej modulacji AM, w wyniku której jego pasmo zostaje przeniesione w zakres dużych częstotliwości. Tego rodzaju systemy transmisji są nazywane systemami PAM-AM. W odbiorniku sygnałów PAM-AM następuje najpierw demodulacja sygnału AM za pomocą detektora obwiedni, a następnie właściwa demodulacja sygnału PAM.

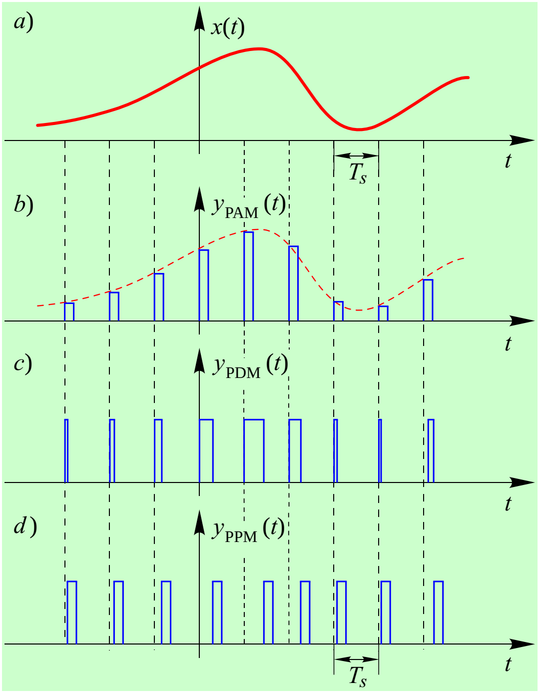
12.2. Modulacje PDM i PPM

12.2.1. Sygnały PDM i PPM

W analogowych systemach modulacji jako fala nośna jest wykorzystywany sygnał harmoniczny. W wyniku modulacji uzmienniany jest jeden z trzech jego parametrów, tj. amplituda, faza lub częstotliwość.

Fala unipolarna, wykorzystywana jako fala nośna w impulsowych systemach modulacji, jest scharakteryzowana także trzema parametrami: amplitudą impulsów, ich szerokością i częstotliwością powtarzania. W systemie PAM, który jest odpowiednikiem analogowych systemów modulacji amplitudy, uzmienniana jest amplituda impulsów fali nośnej. Uzmiennieniu mogą jednak podlegać również pozostałe dwa parametry. W systemie PDM (ang. *Pulse Duration Modulation*) uzmienniana jest w zależności od bieżących próbek sygnału informacyjnego szerokość impulsów fali nośnej, natomiast w systemie PPM (ang. *Pulse Position Modulation*) – ich położenie względem nominalnych chwil nT_s występowania impulsów. Można powiedzieć, że odpowiednikiem modulacji PDM wśród modulacji analogowych jest modulacja PM, zaś modulacji PPM – modulacja FM.

Sygnały zmodulowane w systemach PDM i PPM, wraz z odpowiadającym im sygnałem PAM, są przedstawione na rys. 12.5.



Rys. 12.5. Sygnały: PAM (b), PDM (c) oraz PPM (d) zmodulowane tym samym sygnałem informacyjnym (a)

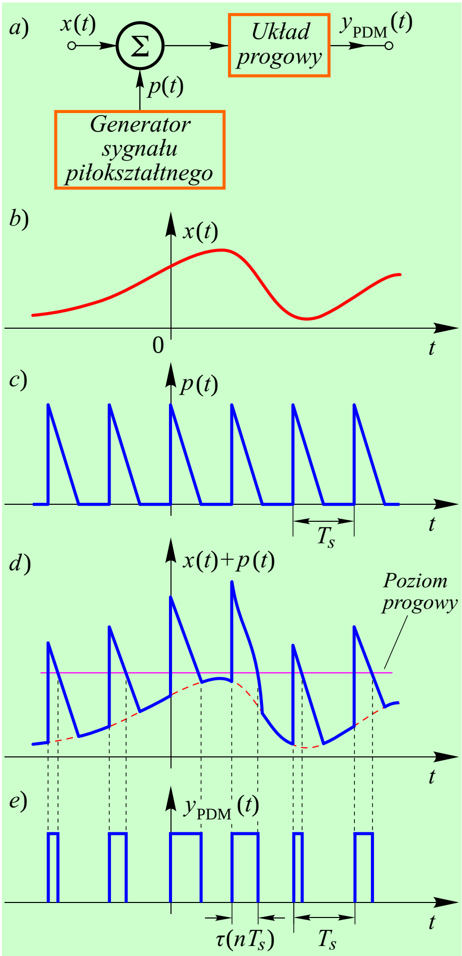
12.2.2. Generacja i demodulacja sygnałów PDM i PPM

W systemie PDM szerokości $\tau(nT_s)$ kolejnych impulsów są związane z wartościami próbek $x(nT_s)$ sygnału informacyjnego zależnością:

$$\tau(nT_s) = a_0 + a_1 x(nT_s)\,. \tag{12.5}$$

Stałe a_0 i a_1 dobiera się tak, aby $\tau(nT_s) > 0$ dla najmniejszej chwilowej wartości sygnału informacyjnego oraz $\tau(nT_s) < T_s$.

Sygnał PDM można wytworzyć w układzie, którego schemat blokowy jest przedstawiony na rys. 12.6a. Na rys. 12.6 są także pokazane sygnały występujące w poszczególnych punktach układu. Układ składa się z generatora przebiegu piłokształtnego, sumatora i układu progowego. Generator wytwarza okresowy sygnał piłokształtny $p(t)$ (rys. 12.6c), o okresie T_s równym okresowi próbkowania sygnału informacyjnego. Sygnał ten jest dodawany w sumatorze do sygnału informacyjnego (rys. 12.6b). Sygnał z wyjścia sumatora (rys. 12.6d) jest podawany z kolei na układ progowy. Na wyjściu układu progowego pojawia się sygnał o stałej wartości dodatniej jedynie w tych przedziałach czasu, w których wartości sygnału na jego wejściu przekraczają ustaloną wartość progową. W pozostałych przedziałach czasu wartości sygnału na wyjściu układu progowego są równe zeru. W ten sposób układ formuje ciąg impulsów prostokątnych, których czasy trwania $\tau(nT_s)$ są proporcjonalne do wartości kolejnych próbek sygnału informacyjnego. Otrzymany na wyjściu układu sygnał PDM jest pokazany na rys. 12.6e. Parametry sygnału piłokształtnego i wartość progu dobiera się tak, aby szerokości generowanych impulsów spełniały nierówność $0 < \tau(nT_s) < T_s$.



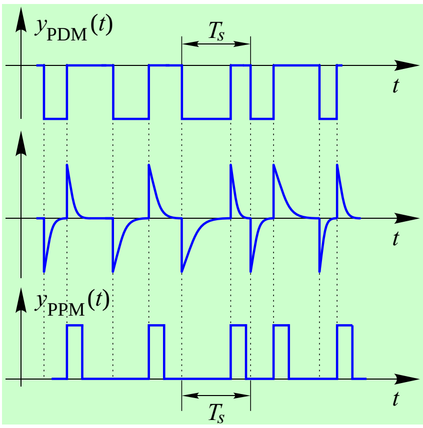
Rys. 12.6. Schemat blokowy generatora sygnału PDM (a) oraz sygnały występujące w jego poszczególnych punktach (b-e)

Przesyłanie sygnałów w systemie PDM wymaga większej średniej mocy nadajnika w porównaniu z systemami PAM i PPM. Związane jest to ze znacznym średnim czasem trwania impulsów sygnału PDM. Z tego względu modulacja PDM jest rzadko stosowana w praktyce. Jest ona natomiast wykorzystywana zwykle jako etap pośredni generacji sygnału PPM, który wymaga mniejszych transmitowanych mocy ze względu na stały i krótki czas trwania impulsów. System PPM wykazuje ponadto większą odporność na szumy i zakłócenia w porównaniu z systemami PAM i PDM.

Metoda wytwarzania sygnału PPM z sygnału PDM jest zilustrowana na rys. 12.7. Impulsy sygnału PDM, po zmianie ich polaryzacji (rys. 12.7a), są podawane na układ różniczkujący o małej stałej czasowej, który wytwarza na wyjściu wąskie impulsy szpilkowe (rys. 12.7b). Dodatnie impulsy szpilkowe, występujące w punktach tylnych zboczy impulsów PDM, wyzwalają generator wąskich impulsów prostokątnych o jednakowym czasie trwania i jednakowej amplitudzie (rys. 12.7c). Informacja o sygnale modulującym zapamiętana w szerokości impulsów zostaje tym samym w bardzo prosty sposób przekodowana na przesunięcie impulsów względem chwil próbkowania nT_s.

Analiza widmowa sygnałów PDM i PPM jest bardzo złożona, a wyznaczenie dokładnych postaci widm tych sygnałów i tym samym dobranie odpowiednich filtrów rekonstruujących widmo sygnału informacyjnego jest praktycznie niemożliwe. Demodulację tych sygnałów przeprowadza się zwykle drogą pośrednią, zamieniając je w odbiorniku na sygnał PAM i stosując filtrację dolnoprzepustową.

Podobnie jak sygnał PAM, sygnały PDM i PPM zawierają składową stałą i charakteryzują się dużą gęstością widmową w zakresie niskich częstotliwości. Z tego względu są one transmitowane często po dodatkowej modulacji AM, przesuwającej ich widma w zakres dużych częstotliwości.



Rys. 12.7. Przetwarzanie sygnału PDM w sygnał PPM

12.2.3. Porównanie impulsowych systemów modulacji

Spośród omówionych impulsowych systemów modulacji najbardziej korzystne właściwości ma system PPM. Jego największą zaletą jest większa odporność na zakłócenia w porównaniu z pozostałymi dwoma systemami modulacji. Z dokładnej analizy wynika, że w systemie PPM uzyskuje się poprawę stosunku sygnał-szum proporcjonalną do kwadratu szerokości pasma. Ponadto wymaga on znacznie mniejszej średniej transmitowanej mocy niż system PDM. Z tego względu w praktyce preferowana jest modulacja PPM.

12.3. Modulacje impulsowo-kodowe

12.3.1. Modulacja PCM

W analogowych systemach modulacji impulsowej PAM, PDM i PPM informacja jest przesyłana w dyskretnych chwilach czasu w wielkości analogowej przybierającej wartości w zbiorze ciągłym. W systemach impulsowo-kodowych przesyłanie informacji następuje w sposób dyskretny zarówno w czasie, jak i w amplitudzie. Dyskretyzację w amplitudzie uzyskuje się w wyniku operacji kwantowania, która umożliwia zakodowanie przesyłanej informacji słowami binarnymi o skończonej długości słowa. W systemach impulsowo-kodowych na sygnale informacyjnym są zatem wykonywane trzy podstawowe operacje, charakterystyczne dla przetwarzania analogowo-cyfrowego sygnałów: próbkowanie, kwantowanie i kodowanie (por. rys. 6.2).

Podstawową modulacją impulsowo-kodową jest modulacja PCM (ang. *Pulse Code Modulation*). Zasadę działania systemu PCM wyjaśnimy na przykładzie kodowania sygnału PAM. Kodowaniu mogą jednak równie dobrze podlegać sygnały PDM lub PPM. W wyniku kwantowania zakres zmienności amplitudy impulsów sygnału PAM jest dzielony na skończoną liczbę przedziałów kwantyzacji, z reguły równą 2^n, gdzie n jest liczbą całkowitą dodatnią. Poszczególne poziomy kwantyzacji można wówczas zakodować słowami binarnymi o stałej długości słowa równej n. Na rys. 12.8 podano przykład kodowania impulsów sygnału PAM przy założeniu, że zakres zmienności ich amplitudy jest podzielony na $16 = 2^4$ poziomów ponumerowanych od 0 do 15. Poziomy te zakodowano naturalnym kodem o stałej długości słowa równej 4 w taki sposób, że poszczególnym poziomom odpowiadają słowa kodowe stanowiące zapis binarny ich numerów.

Po zakodowaniu sygnału PAM informacja źródłowa ma już postać cyfrową. W celu jej przesłania znakom binarnym „1” i „0” należy przypisać konkretną postać fizyczną, tzn. ustalić ich reprezentację za pomocą odpowiednich sygnałów elektrycznych. W praktyce stosowane są różne rodzaje fizycznych reprezentacji znaków binarnych tworzące tzw. *kod sygnałowy*. Na rys. 12.8 przyjęto kod sygnałowy, w którym znak binarny „1” jest reprezentowany krótkim dodatnim

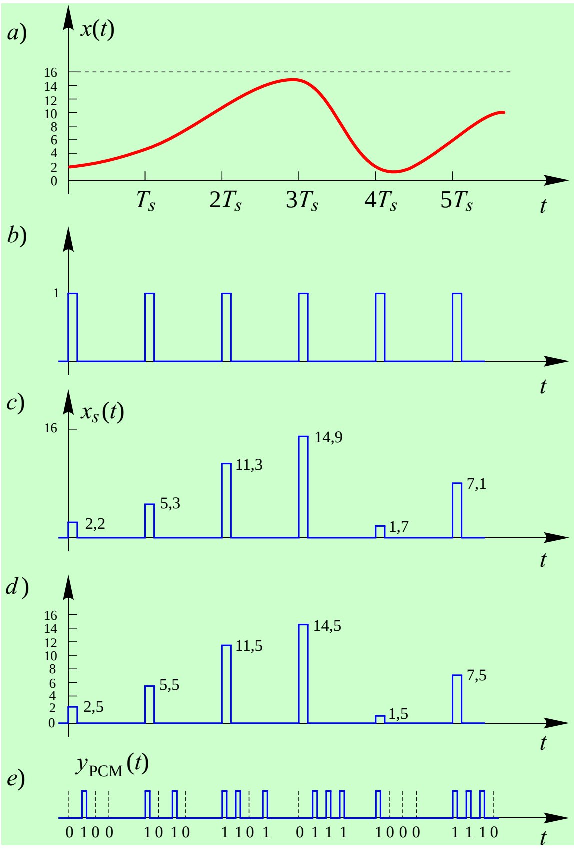
Numer poziomu kwantyzacji	*Słowo kodowe*	*Ciąg impulsów kodowych*
	$2^3\,2^2\,2^1\,2^0$	$2^3\,2^2\,2^1\,2^0$
0	0 0 0 0	
1	0 0 0 1	
2	0 0 1 0	
3	0 0 1 1	
4	0 1 0 0	
5	0 1 0 1	
6	0 1 1 0	
7	0 1 1 1	
8	1 0 0 0	
9	1 0 0 1	
10	1 0 1 0	
11	1 0 1 1	
12	1 1 0 0	
13	1 1 0 1	
14	1 1 1 0	
15	1 1 1 1	

Rys. 12.8. Kodowanie sygnału PAM

impulsem o ustalonej amplitudzie i czasie trwania znacznie mniejszym od czasu transmisji znaku (w rozpatrywanym przypadku równego około jednej czwartej okresu próbkowania T_s), natomiast znak binarny „0” jest reprezentowany poziomem zerowym (impulsem zerowym). Transmitowany sygnał PCM tworzy ciąg dodatnich impulsów kodowych (gdy transmitowany jest znak binarny „1”) przedzielonych przerwami (gdy transmitowany jest znak binarny „0”). Inne sposoby reprezentacji fizycznej znaków binarnych zostaną omówione w p. 12.3.4.

Na rys. 12.9 zilustrowano kolejne etapy przetwarzania napięciowego sygnału informacyjnego $x(t)$ w sygnał PCM. Dla wygody przyjęto, że wartości tego sygnału zmieniają się w zakresie od 0 do 16 umownych znormalizowanych jednostek napięcia (rys. 12.9a). Zakres ten podzielono na 16 przedziałów kwantyzacji o jednakowej szerokości równej 1.

W pierwszym kroku przetwarzania sygnał informacyjny jest spróbkowany chwilowo, tj. zmodulowany w systemie PAM przy założeniu prostokątnej fali nośnej (rys. 12.9b). W wyniku modulacji powstaje sygnał PAM pokazany na rys. 12.9c. Na rysunku tym podano znormalizowane (analogowe) wartości amplitud kolejnych impulsów tego sygnału. W wyniku kwantowania dokładne wartości amplitud są przybliżane najbliższymi liczbami równymi połowie wartości między kolejnymi poziomami kwantyzacji, a więc amplituda 11,3 jest przybliżona wartością 11,5, amplituda 14,9 – wartością 14,5 itd. Na rys. 12.9d przedstawiono sygnał PCM przy założeniu omówionego wyżej kodu sygnałowego, przy czym ze względu na prostszą realizację operacji demodulacji tego sygnału (por. p. 12.3.3)



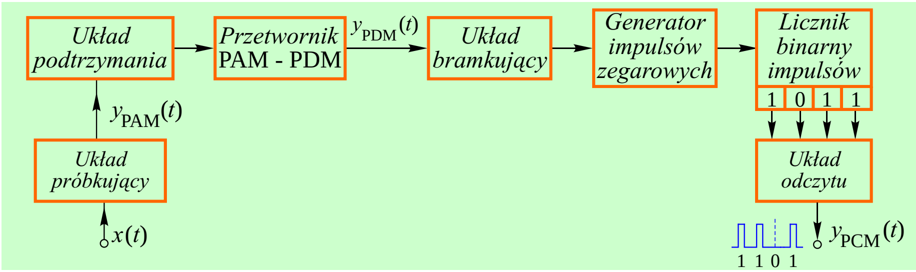
Rys. 12.9. Sygnały występujące w kolejnych etapach generacji sygnału PCM

kolejność impulsów kodowych w każdym okresie próbkowania została odwrócona.

12.3.2. Generacja sygnału PCM

Schemat blokowy modulatora generującego sygnał PCM jest przedstawiony na rys. 12.10. W pierwszym kroku sygnał informacyjny $x(t)$ jest próbkowany chwilowo w układzie próbkującym. Na wyjściu tego układu otrzymujemy sygnał PAM.

Przetwarzanie sygnału PAM w sygnał PCM odbywa się w cyklach, których czas trwania jest równy okresowi próbkowania T_s. W kolejnych cyklach przetwarzane są kolejne impulsy sygnału PAM. Przetwarzanie impulsów prześledzimy na przykładzie impulsu o amplitudzie 11,3 (trzeci kolejny impuls na rys. 12.9c). W chwili rozpoczęcia cyklu impuls ten jest podany na układ podtrzymania i jednocześnie uruchamia generator impulsów zegarowych, który w jednym okresie



Rys. 12.10. Schemat blokowy generatora sygnału PCM

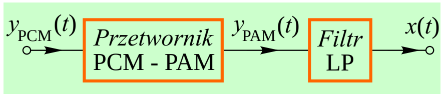
próbkowania generuje maksymalnie 16 równoodległych impulsów. Impulsy zegarowe są zliczane przez licznik wyposażony w czteropozycyjny rejestr. Układ podtrzymania zapamiętuje wartość amplitudy impulsu i steruje przetwornikiem sygnału PAM w sygnał PDM. Przetwornik ten generuje impuls prostokątny, którego szerokość jest proporcjonalna do zapamiętanej wartości amplitudy. Tylne zbocze tego impulsu zamyka bramkę, która z kolei stopuje generator impulsów zegarowych. Liczba impulsów zliczonych do tej pory w liczniku (w rozpatrywanym przypadku równa 11) jest proporcjonalna do czasu trwania impulsu PDM, a zarazem do amplitudy impulsu PAM. W chwili zastopowania generatora impulsów zegarowych liczba ta zostaje zapamiętana w rejestrze. Informacja zawarta w rejestrze jest wyprowadzana przez układ odczytu w kolejności od najmniej do najbardziej znaczącej pozycji i na tej podstawie jest formowany ciąg impulsów sygnałowych sygnału PCM odpowiadający przetwarzanemu impulsowi PAM. Zauważmy, że funkcję układu kwantowania spełnia tu niejako automatycznie licznik. Przed rozpoczęciem kolejnego cyklu licznik zostaje wyzerowany, po czym następuje przetwarzanie kolejnego impulsu PAM. Na rys. 12.10 pokazano ciąg impulsów PCM odpowiadający impulsowi PAM o amplitudzie 11,3.

Trzy pierwsze układy modulatora sygnału PCM stanowią w istocie rzeczy modulator PDM. Modulator ten można oczywiście zrealizować w inny sposób, np. wykorzystując generator impulsów piłokształtnych. Pozostałe układy są prostymi typowymi układami cyfrowymi.

12.3.3. Demodulacja sygnału PCM

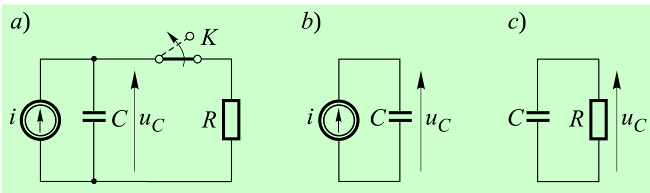
Operację demodulacji sygnału PCM można zrealizować za pomocą równie prostego układu, którego schemat blokowy jest pokazany na rys. 12.11. Zawiera on przetwornik sygnału PCM w sygnał PAM i filtr dolnoprzepustowy formujący po stronie odbiorczej analogowy sygnał informacyjny.

Przetworzenie sygnału PCM w sygnał PAM można zrealizować za pomocą prostego układu RC (rys. 12.12a). Na wejście tego układu jest podawany ciąg impulsów kodowych sygnału PCM odpowiadający słowu kodowemu kodującemu bieżącą próbkę sygnału informacyjnego. Impulsy tego ciągu są podawane na



Rys. 12.11. Schemat blokowy demodulatora sygnału PCM

układ w odwrotnym porządku, tj. od najmniej do najbardziej znaczącej pozycji. Założymy, że czas trwania τ dodatnich impulsów kodowych jest znacznie krótszy od czasu $\Delta t \approx T_s/4$ transmisji jednego znaku binarnego. Źródło impulsów powinno być źródłem prądowym zbliżonym do źródła idealnego o nieskończenie dużym oporze wewnętrznym. Przednie zbocza dodatnich impulsów kodowych przełączają klucz K w pozycję rozwartą, natomiast tylne zbocza powodują zamknięcie klucza.



Rys. 12.12. Realizacja układowa przetwornika PCM-PAM

Rozpatrzmy proces ładowania kondensatora C pojedynczym kodowym impulsem prądowym o amplitudzie I_0 i czasie trwania τ przy założeniu idealnego źródła prądowego (rys. 12.12b). Równanie operatorowe tego obwodu ma postać (por. [1], p. 6.2.4-B):

$$u_C(s) = \frac{1}{sC} i(s) + \frac{u_0}{s}, \tag{12.6}$$

gdzie u_0 jest napięciem początkowym na kondensatorze. Ponieważ w czasie trwania impulsu można go opisać skokiem jednostkowym $i(t) = I_0 \mathbf{1}(t)$, zatem $i(s) = I_0/s$. Uwzględniając to we wzorze (12.6), otrzymujemy:

$$u_C(s) = \frac{I_0}{s^2 C} + \frac{u_0}{s},$$

a stąd, po obliczeniu odwrotnej transformaty Laplace'a:

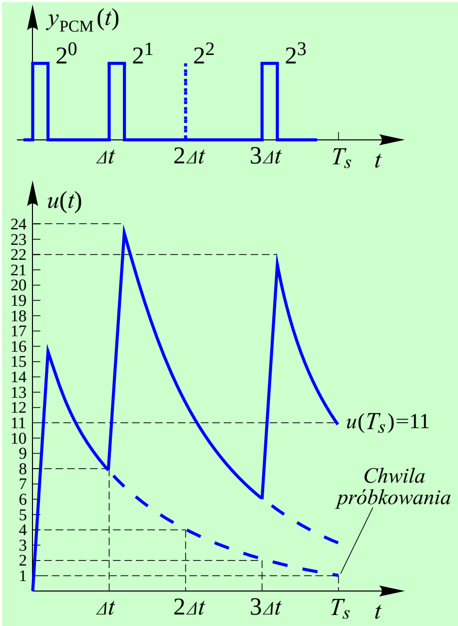
$$u_C(t) = \frac{I_0}{C} t + u_0, \qquad 0 < t \leqslant \tau\,. \tag{12.7}$$

Widzimy więc, że ładowanie kondensatora ma charakter ładunkowy, tzn. w czasie τ trwania impulsu napięcie na kondensatorze przyrasta liniowo od wartości początkowej u_0 zawsze o stałą wartość proporcjonalną do ładunku $I_0\tau$ zawartego w impulsie. Jeżeli napięcie na kondensatorze będziemy mierzyć w skali znormalizowanej, to amplitudę I_0 impulsów dobiera się tak, aby każdy impuls, niezależnie od wartości napięcia początkowego, powodował przyrost napięcia na kondensatorze o 2^n znormalizowanych jednostek, gdzie n jest długością słowa kodowego kodującego próbki sygnału informacyjnego (w rozpatrywanym przykładzie przyrost o 16 jednostek). Tylne zbocze impulsu przełącza po czasie τ klucz K do pozycji zamkniętej i w pozostałej części przedziału Δt następuje rozładowanie kondensatora przez opór R w obwodzie RC pokazanym na rys. 12.12c. Stała czasowa RC tego obwodu jest dobrana tak, aby spełniona była równość $RC = 1{,}44\Delta t$. Można łatwo sprawdzić, że jeśli spełniony jest warunek $\tau \ll \Delta t$ (czas τ można pominąć), to przy takim doborze wartości stałej czasowej napięcie na kondensatorze zmaleje wykładniczo po czasie Δt dokładnie do połowy swojej wartości początkowej, tj. wartości w chwili wystąpienia tylnego zbocza impulsu kodowego. Jeśli po czasie Δt w ciągu impulsów kodowych wystąpi kolejny impuls dodatni (transmitowany jest znak „1”), kondensator podładuje się o 2^n jednostek, po czym napięcie na nim maleje z taką samą prędkością. Jeśli zaś po czasie Δt w ciągu impulsów kodowych wystąpi impuls zerowy (transmitowany jest znak „0”), napięcie na kondensatorze maleje nadal i po czasie $2\Delta t$ zmaleje do $1/4$ wartości początkowej. Po kolejnym zerowym impulsie kodowym napięcie na kondensatorze zmaleje do $1/8$ wartości początkowej itd.

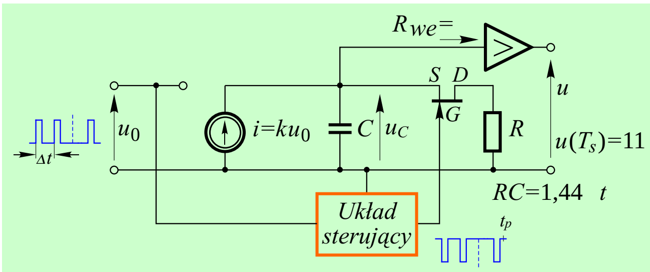
Po czasie T_s, tj. na końcu cyklu przetwarzania, napięcie na kondensatorze jest próbkowane i formowany jest impuls PAM o amplitudzie równej zarejestrowanej wartości próbki napięcia. Przed rozpoczęciem następnego cyklu przetwarzania napięcie na kondensatorze zostaje wyzerowane, po czym następuje przetwarzanie kolejnego ciągu impulsów kodowych. Można bez trudu sprawdzić, że przy założonej stałej czasowej RC obwodu znormalizowane wartości próbek napięcia w chwilach nT_s będą zawsze równe numerowi poziomu kwantyzacji sygnału PCM reprezentowanemu przez przetwarzany ciąg impulsów kodowych.

Procedura przetwarzania sygnału PCM w sygnał PAM została zilustrowana na rys. 12.13 na przykładzie ciągu impulsów 1101 odpowiadającego jedenastemu poziomowi kwantyzacji. Analizując wykres napięcia na kondensatorze należy uwzględnić, że impulsy kodowe są podawane na układ demodulatora PCM w odwrotnym porządku. W celu odtworzenia sygnału informacyjnego, sygnał PAM należy jeszcze podać na wejście filtru dolnoprzepustowego.

Jedna z możliwych realizacji układowych przetwornika PCM-PAM jest przedstawiona na rys. 12.14. Układ spełnia jednocześnie funkcję przetwarzania sygnału i sterowania kluczem K. Kluczowanie jest zrealizowanie za pomocą tranzystora JFET sterowanego impulsami kodowymi o odwróconej polaryzacji. Ujemne impulsy zatykają tranzystor w obwodzie bramka-źródło (G-S) i w czasie ich trwania kanał źródło-dren (S-D) stanowi bardzo duży opór. W przerwach między impulsami tranzystor jest włączony, a opór kanału S-D jest bardzo mały w porównaniu z oporem R i nie wpływa praktycznie na stałą czasową RC.



Rys. 12.13. Wykres napięcia na kondensatorze przetwornika PCM-PAM



Rys. 12.14. Realizacja układowa przetwornika PCM-PAM na tranzystorze JFET

12.3.4. Modulacja delta (DM)

Jeśli próbki sygnału informacyjnego są pobierane z częstotliwością Nyquista, to w przedziale czasu między dwoma sąsiadującymi próbkami może nastąpić wyraźna zmiana sygnału i różnica między wartościami tych próbek może być znaczna. W konsekwencji między dwoma słowami kodowymi kodującymi te słowa mogą wystąpić znaczne różnice. Pociąga to za sobą konieczność dokonywania w procesie kodowania zmian na wielu pozycjach kolejnych słów kodowych. Koncepcja modulacji delta polega na próbkowaniu sygnału informacyjnego z częstotliwością dużo większą od częstotliwości Nyquista i tym samym zmniejszeniu dynamiki zmian wartości kolejnych próbek, a zarazem zwiększeniu korelacji między nimi. Zmiany wartości sygnału od próbki do próbki są wówczas na tyle nie-znaczne, że informacja o nich (a dokładniej o tym, czy sygnał wzrósł, czy zmalał) może być zakodowana za pomocą tylko jednego znaku binarnego.

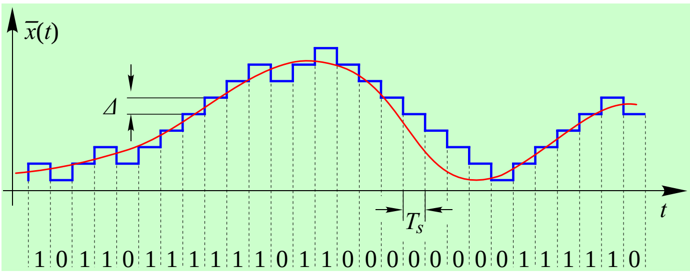
Zasada działania modulatora sygnału DM polega na śledzeniu przyrostów sygnału informacyjnego $x(t)$ w chwilach próbkowania i wytworzeniu na tej podstawie jego schodkowej aproksymacji $\bar{x}(t)$ (rys. 12.15a). W każdej chwili próbkowania nT_s wyznaczana jest próbka $x(nT_s)$ sygnału informacyjnego i porównywana z aproksymowaną wartością sygnału $\bar{x}(nT_s - T_s)$ w poprzedniej chwili próbkowania. Następnie badany jest znak różnicy

$$e(nT_s) = x(nT_s) - \bar{x}(nT_s - T_s) \tag{12.8}$$

i na tej podstawie tworzony sygnał binarny

$$\bar{e}(nT_s) = \Delta \operatorname{sgn}[e(nT_s)] \tag{12.9}$$

przybierający wartość $+\Delta$, jeśli różnica (12.8) jest dodatnia (aktualna próbka sygnału jest większa od wartości aproksymowanej), oraz wartość $-\Delta$, jeśli różnica ta jest ujemna (aktualna próbka sygnału jest mniejsza od wartości aproksymowanej).



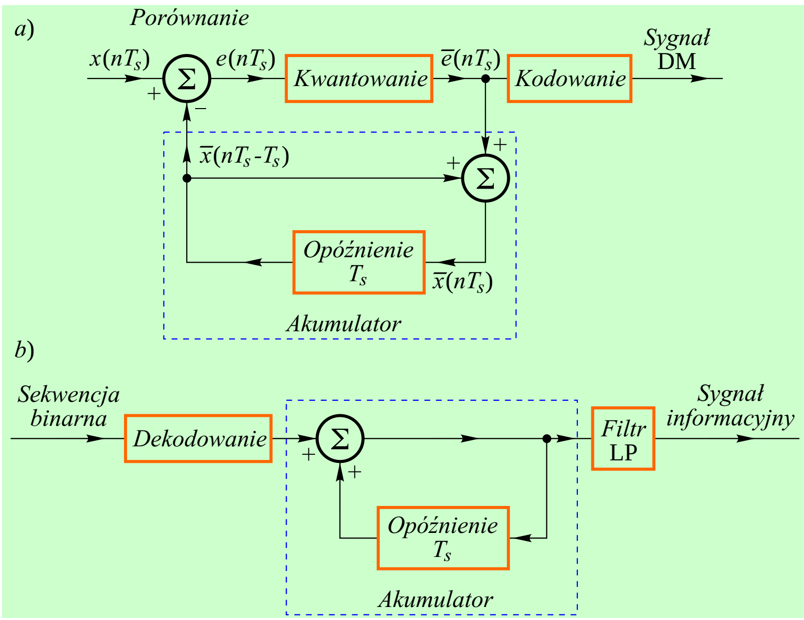
Rys. 12.15. Zasada generacji sygnału DM (a) i sekwencja kodowa (b)

Sygnał binarny $\bar{e}(nT_s)$ powstaje więc w wyniku skwantowania sygnału $e(nT_s)$ na dwa poziomy $\pm\Delta$. Na podstawie sygnału $\bar{e}(nT_s)$ jest wyznaczana następna wartość schodkowego sygnału aproksymującego

$$\bar{x}(nT_s) = \bar{x}(nT_s - T_s) + \bar{e}(nT_s), \tag{12.10}$$

która jest utrzymywana przez odcinek czas T_s. W ten sposób jest tworzony schodkowy sygnał aproksymujący $\bar{x}(t)$. Jeśli sygnał $x(t)$ nie zmienia się między próbkami zbyt szybko, to błąd aproksymacji $x(t)-\bar{x}(t)$ jest utrzymywany w przedziale $[-\Delta, +\Delta]$. W końcowym etapie modulacji delta wyznaczana jest sekwencja kodowa kodująca binarnie sygnał $\bar{e}(nT_s)$. Wartości $\bar{e}(nT_s) = \Delta$ przyporządkowywany jest znak binarny „1", a wartości $\bar{e}(nT_s) = -\Delta$ – znak binarny „0" (rys. 12.15b). Sekwencja ta jest przesyłana w postaci odpowiedniego kodu impulsowego do odbiornika sygnału DM. Do odbiornika jest zatem przesyłana zakodowana informacja jedynie o kolejnych przyrostach sygnału aproksymującego.

Zaletą modulacji DM są bardzo proste układy modulatora i demodulatora. Schemat blokowy modulatora sygnału DM działającego zgodnie z algorytmem (12.8)–(12.10) jest pokazany ma rys. 12.16a. Demodulator na podstawie przychodzącej do odbiornika sekwencji binarnej wytwarza ciąg dodatnich i ujemnych impulsów prostokątnych, które są akumulowane w pętli sprzężenia zwrotnego (rys. 12.16b) i na tej podstawie jest odtwarzana w odbiorniku schodkowa aproksymacja sygnału informacyjnego. Sygnał schodkowy jest jeszcze podawany na wejście wygładzającego filtru dolnoprzepustowego o częstotliwości odcięcia równej maksymalnej częstotliwości pasma sygnału informacyjnego, którego zadaniem jest usunięcie z widma sygnału schodkowego składowych o dużych częstotliwościach związanych ze skokowymi zmianami tego sygnału.

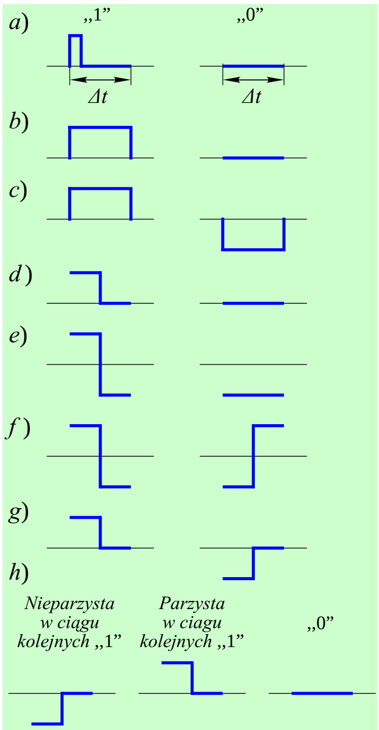


Rys. 12.16. Schemat blokowy modulatora (a) i demodulatora (b) sygnału DM

Wadą modulacji delta jest stały krok przyrostów sygnału aproksymującego o $+\Delta$ lub $-\Delta$, niezależny od szybkości zmian sygnału. Ustalona wartość przyrostu Δ może być za mała, gdy sygnał zmienia się zbyt szybko od próbki do próbki, lub za duża, gdy zmiany sygnału są bardzo wolne. Może to być źródłem nieakceptowalnych błędów aproksymacji. W celu uniknięcia tych błędów w praktyce stosowane są często bardziej rozwinięte systemy modulacji impulsowej takie jak *modulacja delta-sigma*, *różnicowa modulacja impulsowo-kodowa*, a ostatnio coraz częściej ich wersje adaptacyjne, w których krok aproksymacji $\Delta(nT_s)$ jest zmienny i dobierany w każdej chwili próbkowania adaptacyjnie w zależności od obserwowanej szybkości zmian sygnału informacyjnego.

12.3.5. Sposoby fizycznej reprezentacji znaków binarnych w modulacjach impulsowo-kodowych

W systemach modulacji impulsowo-kodowych wykorzystywane są różne kody sygnałowe, w których stosowane są różne sposoby reprezentacji znaków binarnych „1” i „0” za pomocą impulsów elektrycznych. Najczęściej spotykane kody sygnałowe są zestawione na rys. 12.17.



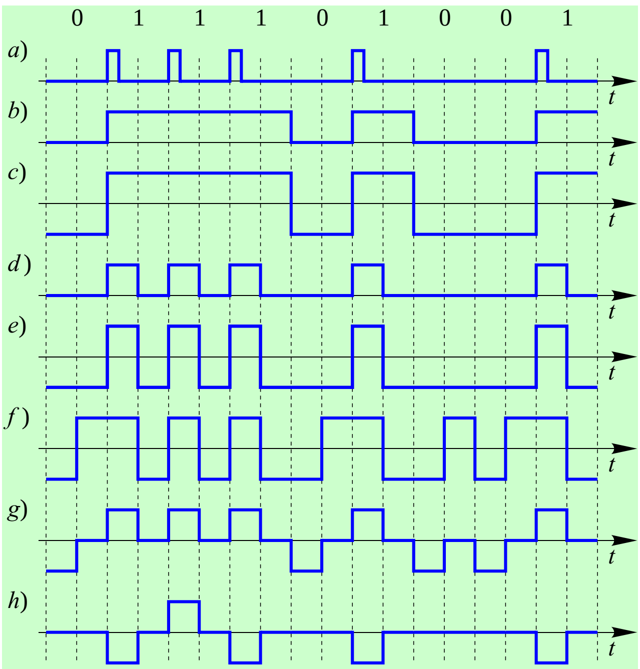
Rys. 12.17. Różne sposoby fizycznej reprezentacji znaków binarnych „1” i „0”

Na rys. 12.17a przedstawiono omówiony w p. 12.3.1 kod sygnałowy, w którym znak binarny „1” jest reprezentowany krótkim impulsem dodatnim, a znak binarny „0” – poziomem zerowym. Zbliżonym sposobem kodowania jest reprezentacja znaku „1” impulsem dodatnim o stałej amplitudzie w całym czasie Δt transmisji pojedynczego znaku oraz znaku „0” – poziomem zerowym (rys. 12.17b). Na rys. 12.17c przedstawiono kod sygnałowy *NRZ bez powrotu do zera* (ang. *nonreturn-to-zero*), w którym znaki „1” i „0” są reprezentowane impulsami o czasie trwania Δt, jednakowych amplitudach i przeciwnych polaryzacjach. Z kolei w kodzie *RZ z powrotem do zera* (ang. *return-to-zero*), często stosowanym w praktyce, znak „1” jest reprezentowany impulsem dodatnim o czasie trwania równym połowie czasu Δt, zaś znak „0” – poziomem zerowym (rys. 12.17d).

Osobną grupę kodów sygnałowych stanowią kody, w których znaki binarne są reprezentowane przejściami między poziomami sygnału. Na rys. 12.17e przedstawiono kod *bifazowy jednoelementowy*, w którym znak „1” jest reprezentowany przejściem od poziomu dodatniego do ujemnego, a znak „0” – poziomem ujemnym. Odmianą tego kodu jest kod *bifazowy dwuelementowy* nazywany także *kodem Manchester*. W kodzie tym znak binarny „1” jest reprezentowany przejściem od poziomu dodatniego do ujemnego, a znak binarny „0” – przejściem odwrotnym od poziomu ujemnego do dodatniego (rys. 12.17f).

Do kodów kodujących znaki binarne za pomocą przejść należą także *kody ternarne*, w których wykorzystywane są trzy poziomy sygnału: dodatni, ujemny i zerowy. W kodzie *ternarnym RZ* z powrotem do zera znaki „1” i „0” są reprezentowane impulsami dodatnimi i odpowiednio ujemnymi, trwającymi przez połowę czasu Δt transmisji znaku i powracającymi w drugiej połowie przedziału Δt do poziomu zerowego (rys. 12.17g). Wariantem kodu ternarnego RZ jest kod *ternarny bipolarny RZ*. W kodzie tym znak „0” jest zakodowany poziomem zerowym, natomiast znak „1” jest zakodowany albo impulsem ujemnym albo dodatnim, w zależności od tego, czy znak „1” jest nieparzysty czy parzysty w ciągu kolejnych znaków „1”. Impulsy te trwają przez połowę przedziału Δt i w drugiej połowie tego przedziału powracają do poziomu zerowego (rys. 12.17h).

Na rys. 12.18 przedstawiono sygnały otrzymane w wyniku zakodowania wymienionymi wyżej kodami sygnałowymi sekwencji binarnej 011101001.



Rys. 12.18. Sekwencja binarna 011101001 zakodowana kodami sygnałowymi z rys. 12.17

Sygnały elektryczne transmitowane w systemach modulacji impulsowo-kodowych są oczywiście sygnałami losowymi (pokazane na rys. 12.18 konkretne przebiegi tych sygnałów są nazywane ich *realizacjami*). Właściwości widmowe tych sygnałów zależą silnie od rodzaju użytego kodu sygnałowego. Stosowanie w praktyce bardziej złożonych kodów sygnałowych komplikuje układy generacji, jest jednak podyktowane chęcią dopasowania widma sygnału do charakterystyki częstotliwościowej kanału transmisyjnego. W praktyce dąży się przede wszystkim do takiego ukształtowania widma, aby nie zawierało ono niekorzystnych do transmisji składowej stałej i składowych niskoczęstotliwościowych. Pod tym względem kody z kodowaniem znaków binarnych poziomami (rys. 12.17a-d) nie są dogodne, charakteryzują się bowiem dużą gęstością widmową składowych o niskich częstotliwościach, które są zniekształcane podczas transmisji. Wady tej nie ma natomiast np. kod bifazowy dwuelementowy Manchester (rys. 12.17f), a także kod bipolarny ternarny RZ (rys. 12.17h). Widma sygnałów zakodowane tymi kodami nie zawierają składowej stałej i mają małą gęstość widmową w otoczeniu częstotliwości zerowej. Z tego względu są one preferowane w praktyce.

Słownik

efekt aperturowy

zniekształcenia widma sygnału PAM wynikające ze skończonego czasu trwania transmitowanych impulsów

kod sygnałowy impulsowy

sposób fizycznej reprezentacji znaków binarnych „1” oraz „0” za pomocą impulsów elektrycznych

modulacja amplitudy impulsów PAM

modulacja impulsowa, w której zgodnie z bieżącymi wartościami próbek sygnału jest uzmienniana amplituda impulsów unipolarnej prostokątnej fali nośnej

modulacja częstotliwości FM

modulacja analogowa, w której w zależności od bieżących wartości sygnału informacyjnego uzmienniana jest bezpośrednio częstotliwość harmonicznej fali nośnej

modulacja delta DM
modulacja impulsowo-kodowa, w której informacja o transmitowanym sygnale jest przesyłana za pomocą znaków binarnych kodujących przyrosty wartości sygnału od próbki do próbki

modulacja impulsowa
modulacja, w której parametry unipolarnej fali nośnej są uzmienniane w dyskretnych chwilach zgodnie z wartościami bieżących próbek sygnału modulującego

modulacja impulsowo-kodowa
modulacja impulsowa, w której wartości próbek sygnału informacyjnego są kwantowane, a następnie kodowane słowami binarnymi; słowom binarnym są przyporządkowywane za pomocą odpowiedniego kodu sygnałowego ciągi impulsów elektrycznych, które są następnie transmitowane do odbiornika

modulacja położenia impulsów PDM
modulacja impulsowa, w której zgodnie z bieżącymi wartościami próbek sygnału informacyjnego uzmienniane jest przesunięcie impulsów unipolarnej prostokątnej fali nośnej względem chwil próbkowania

modulacja szerokości impulsów PDM
modulacja impulsowa, w której zgodnie z bieżącymi wartościami próbek sygnału informacyjnego uzmienniana jest szerokość impulsów unipolarnej prostokątnej fali nośnej

próbkowanie chwilowe
sposób formowania sygnału spróbkowanego przez wymnożenie impulsów unipolarnej fali prostokątnej przez wartości sygnału analogowego w chwilach próbkowania; w wyniku otrzymuje się sygnał PAM

próbkowanie idealne
teoretyczny sposób formowania sygnału spróbkowanego przez wymnożenie sygnału analogowego przez okresowy ciąg impulsów Diraca; w wyniku mnożenia otrzymuje się impulsowy sygnał spróbkowany

próbkowanie naturalne
sposób formowania sygnału spróbkowanego przez wymnożenie sygnału analogowego przez unipolarną falę prostokątną o małym współczynniku wypełnienia

Literatura

[1] Osiowski J., Szabatin J.: Podstawy teorii obwodów, tom II. WNT, Warszawa, wyd. 3, 1998.

Lekcja 13

Modulacje cyfrowe

Lekcja 13, ostatnia, jest poświęcona modulacjom cyfrowym. Rozpoczniemy od ogólnej charakterystyki cyfrowych systemów modulacji i ich klasyfikacji. Omówimy zasady reprezentacji geometrycznej sygnałów zmodulowanych cyfrowo. Wprowadzimy pojęcia przestrzeni sygnałów, ich konstelacji oraz odległości w tej przestrzeni. Rozpatrzymy zagadnienie odbioru sygnałów zmodulowanych cyfrowo i zasady ich detekcji. Dokładniej rozpatrzymy modulacje dwuwartościowe 2PSK i 2FSK. Przeanalizujemy także krótko modulacje wielowartościowe PSK i FSK oraz modulację QAM. Na zakończenie dokonamy porównania cyfrowych systemów modulacji sygnałów.

13.1. Ogólna charakterystyka modulacji cyfrowych

Modulacje cyfrowe są stosowane do transmisji danych binarnych w określonych wąskich rozłącznych pasmach kanału transmisyjnego przeznaczonych na transmisję poszczególnych sygnałów. Wykorzystuje się je np. w łączności modemowej, internetowej, radiowej w paśmie mikrofalowym i satelitarnej. Wprawdzie sygnały zmodulowane w systemach modulacji impulsowo-kodowej mogą być również transmitowane w wąskich pasmach, możliwe jest to jednak po zastosowaniu dodatkowej modulacji analogowej, np. modulacji AM. W przypadku modulacji cyfrowych wąskopasmowy charakter sygnałów wynika natomiast bezpośrednio z samej istoty zastosowanego sposobu modulacji.

W cyfrowych systemach modulacji informacja o sygnale zakodowana w sekwencji znaków binarnych „1” i „0” lub sekwencji grup tych znaków o określonej długości jest przesyłana do odbiornika w postaci ciągu krótkich impulsów harmonicznych, których amplituda, faza początkowa lub częstotliwość jest uzmienniana w zależności od transmitowanego strumienia danych. W bardziej złożonych systemach uzmiennieniu mogą podlegać jednocześnie dwa z tych parametrów.

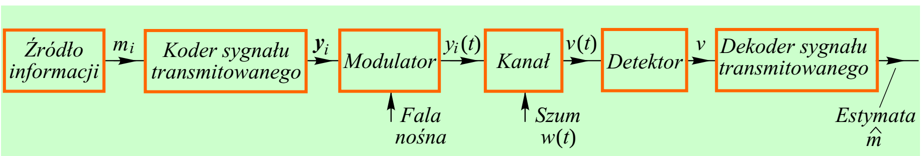
13.1.1. Schemat cyfrowego systemu modulacji

Ogólny schemat blokowy cyfrowych systemów modulacji jest przedstawiony na rys. 13.1. W systemach tych źródło informacji wysyła co T sekund jeden z M *symboli* $m_1, \ldots, m_M$. W przypadku, gdy $M = 2$, system modulacji nazywamy *binarnym* (2-*wartościowym*). W ogólnym przypadku system nazywamy M-*wartościowym*.

Liczba M jest z reguły dodatnią całkowitą potęgą dwójki, a transmitowane symbole są reprezentowane słowami binarnymi o długości $\log_2 M$. Na przykład, w systemach binarnych, najprostszych i najczęściej stosowanych, oba symbole m_1 i m_2 są utożsamiane z symbolami binarnymi „1” oraz „0” (*bitami*). W systemach 4-wartościowych poszczególne symbole są reprezentowane ciągami „11” „10” „01” „00” (*dwubitami*), itd. Z reguły zakłada się, że transmitowane symbole są równoprawdopodobne.

Odcinek czasu o długości T, w którym jest transmitowany pojedynczy symbol, będziemy nazywać *przedziałem symbolowym* (ang. *symbol interval*). Odcinek T obejmuje $\log_2 M$ odcinków czasu o długości $T_b = T / \log_2 M$ równej czasowi trwania bitu. Przedział T_b będziemy nazywać *przedziałem bitowym* (ang. *bit interval*). W przypadku binarnych systemów modulacji $T_b = T$.

Komentarz. Termin *bit* jest używany w dwóch znaczeniach. Oznacza on przede wszystkim jednostkę ilości informacji. Stosowany jest także umownie, aczkolwiek niezbyt precyzyjnie, na określenie jednej pozycji binarnej, na której można zapisać znak binarny „1” lub „0”. W podanych określeniach termin *bit* jest użyty oczywiście w tym drugim znaczeniu.



Rys. 13.1. Schemat blokowy cyfrowego systemu modulacji

Symbole wysyłane ze źródła są kodowane w koderze sygnału transmitowanego, którego zadaniem jest przyporządkowanie symbolowi m_i odpowiedniego N-elementowego wektora liczbowego $\boldsymbol{y}_i = [y_{i\,1}, y_{i\,2}, \ldots, y_{i\,N}]^T$, $i = 1, \ldots, M$, (por. p. 7.4.2-A, B), gdzie indeks ^{T} oznacza transpozycję. Informacja zakodowana w wektorach $\boldsymbol{y}_i$ steruje modulatorem. Zadaniem modulatora jest wytworzenie w kolejnych przedziałach symbolowych sygnałów $y_i(t)$ będących odcinkami harmonicznej fali nośnej o długości T. Sygnały $y_i(t)$ są następnie transmitowane przez kanał do odbiornika, przy czym parametry tych sygnałów: amplituda, faza lub częstotliwość są zmieniane w zależności od aktualnego wektora $\boldsymbol{y}_i$, a więc w zależności od symbolu m_i przesyłanego w danym przedziale czasu T. Ponieważ zmiany tych parametrów następują skokowo, mówimy o *kluczowaniu amplitudy* (*systemy ASK – Amplitude-Shift Keying*), *kluczowaniu fazy* (*systemy PSK – Phase-Shift Keying*) lub *kluczowaniu częstotliwości* (*systemy FSK – Frequency-Shift Keying*).

W czasie transmisji sygnały $y_i(t)$ są zniekształcane na skutek występowania w kanale sygnału zakłócającego $w(t)$. Najczęściej zakłada się, że $w(t)$ jest addytywnym *białym szumem gaussowskim*. Kanał nazywamy wówczas *kanałem AWGN* (ang. *Additive White Gaussian Noise*). W każdym przedziale czasu T na wejściu odbiornika pojawia się zatem sygnał zniekształcony $v(t) = y_i(t) + w(t)$, który przy dużym poziomie szumu może znacznie odbiegać od sygnału transmitowanego. Zadaniem odbiornika jest podjęcie w każdym przedziale czasu T decyzji, który z sygnałów $y_i(t)$ został faktycznie w tym przedziale wysłany. Decyzja ta powinna być oparta na określonej *regule decyzyjnej* i podjęta w sposób optymalny w sensie ustalonego kryterium optymalności. W kategoriach teorii optymalnego odbioru oznacza to, że zadaniem odbiornika jest wyznaczenie w każdym przedziale symbolowym T estymaty $\hat{m}$ transmitowanego symbolu m_i, optymalnej w sensie tego kryterium.

13.1.2. Właściwości sygnałów zmodulowanych cyfrowo

Transmisja sygnałów ASK i FSK może być koherentna lub niekoherentna, podczas gdy sygnały PSK są zawsze transmitowane w systemach koherentnych. Jak pamiętamy, system modulacji nazywamy koherentnym, jeśli faza sygnału nadawanego jest znana po stronie odbiorczej.

Sygnały transmitowane w systemach PSK i FSK charakteryzują się stałą w czasie amplitudą chwilową. Właściwości tej nie mają systemy ASK, ich amplituda chwilowa zmienia się bowiem w zależności od transmitowanych symboli. Podczas transmisji przez mikrofalowe linie radiowe oraz łącza satelitarne sygnały ASK są zatem bardziej narażone na nieliniowe zniekształcenia spowodowane nieliniowościami układów wchodzących w skład łącza transmisyjnego. Z tego względu w zastosowaniach praktycznych są preferowane systemy PSK i FSK i na tych systemach skoncentrujemy przede wszystkim naszą uwagę.

Należy jednak podkreślić, że w praktyce stosuje się również różne warianty wymienionych wyżej podstawowych rodzajów modulacji cyfrowych, takie jak *kluczowanie częstotliwości z ciągłą fazą* (*systemy CPFSK* – ang. *Continuous-Phase Frequency-Shift Keying*), szczególny przypadek modulacji CPFSK z *minimalnym* *rozstawem częstotliwości* (*systemy MSK* – ang. *Minimum Shift Keying*), *różniczkowe kluczowanie fazy* (*systemy DPSK* – ang. *Differential Phase-Shift Keying*), czy *kwadraturowa modulacja amplitudy* (*systemy QAM* – ang. *Quadrature Amplitude Modulation*), która jest kombinacją modulacji PSK i ASK.

We współczesnych systemach transmisji danych coraz częściej stosowane są złożone wielowartościowe rodzaje modulacji cyfrowych. Podstawowe znaczenie mają jednak nadal najprostsze z nich binarne modulacje PSK i FSK, nazywane modulacjami *2PSK* i *2FSK*. Modulacje te omówimy w p. 13.3. Przedtem jednak zapoznamy się z ogólnymi zasadami reprezentacji sygnałów zmodulowanych cyfrowo w odpowiednich przestrzeniach funkcyjnych oraz sposobami ich optymalnego odbioru.

13.2. Geometryczna reprezentacja sygnałów zmodulowanych cyfrowo

13.2.1. Przestrzeń i konstelacja sygnałów

Analizę sygnałów zmodulowanych cyfrowo, a zwłaszcza rozwiązanie zagadnienia ich optymalnego odbioru, ułatwia wprowadzenie geometrycznej reprezentacji sygnałów. W ujęciu geometrycznym sygnały $y_1(t), \dots, y_M(t)$, transmitowane w kolejnych przedziałach symbolowych T, są reprezentowane wektorami w zwykłej skończenie wymiarowej przestrzeni wektorowej. Zasada tej reprezentacji polega na przedstawieniu sygnałów $s_i(t)$ w postaci kombinacji liniowych

$$\begin{aligned}
y_1(t) &= y_{11}\varphi_1(t) + y_{12}\varphi_2(t) + \ldots + y_{1N}\varphi_N(t) \\
y_2(t) &= y_{21}\varphi_1(t) + y_{22}\varphi_2(t) + \ldots + y_{2N}\varphi_N(t) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
y_M(t) &= y_{M1}\varphi_1(t) + y_{M2}\varphi_2(t) + \ldots + y_{MN}\varphi_N(t)
\end{aligned} \tag{13.1}$$

pewnych standardowych rzeczywistych sygnałów $\varphi_1(t), \dots, \varphi_N(t)$ określonych w przedziale czasu $0 \leqslant t < T$ i ortonormalnych w tym przedziale, tj. spełniających dla $k, l = 1, ldots, N$ warunek (por. definicje 2.3–2.5 i wzór (2.42)):

$$\begin{aligned}
(\varphi_k(t), \varphi_l(t))_{L^2(0,T)} &= \textstyle\int_0^T \varphi_k(t)\varphi_l(t)\,\mathrm{d}t = \\
&= \begin{cases} 1, & k = l, \\ 0, & k \neq l. \end{cases}
\end{aligned} \tag{13.2}$$

Zgodnie ze wzorem ogólnym (2.33), współczynniki y_{ij}, $i = 1, \ldots, M$, $j = 1, \ldots, N$, kombinacji liniowych 13.1 są określone wyrażeniem:

$$y_{ij} = (y_i, \varphi_j)_{L^2(0,T)} = \int_0^T y_i(t)\varphi_j(t)\,\mathrm{d}t\,. \tag{13.3}$$

Sygnały $y_i(t)$ można zatem traktować jako zbiór M punktów $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iN})$ w N-wymiarowej *przestrzeni funkcyjnej* $\mathscr{P}$ rozpiętej na bazie ortonormalnej $\{\varphi_1(t), \ldots, \varphi_N(t)\}$ (por. p. 2.3.4). Przestrzeń ta jest nazywana *przestrzenią sygnałów*, zaś zbiór punktów $\boldsymbol{y}_i$, reprezentujących transmitowane sygnały $y_i(t)$, nosi nazwę *konstelacji sygnałów*.

Z określenia przestrzeni $\mathscr{P}$ wynika, że każdy sygnał $x(t)$ należący do tej przestrzeni można przedstawić w postaci skończonego szeregu:

$$x(t) = \sum_{j=1}^{N} x_j\varphi_j(t), \tag{13.4}$$

którego współczynniki $x_j = (x, \varphi_j)_{L^2(0,T)}$ są rzutem sygnału $x(t)$ na sygnały bazowe $\varphi_1(t), \ldots, \varphi_N(t)$. Istnieje przy tym wzajemnie jednoznaczne odwzorowanie przestrzeni $\mathscr{P}$ w N-wymiarową przestrzeń wektorową $\mathscr{R}^N$, które sygnałowi $x(t) \in \mathscr{P}$ przyporządkowuje wektor $\boldsymbol{x} = [x_1, \ldots, x_N]^T \in \mathscr{R}^N$ o współrzędnych będących współczynnikami rozwinięcia (13.4). Przypomnijmy, że odwzorowanie to zachowuje normę i iloczyn skalarny (por. komentarz w p. 2.3.4).

13.2.2. Odległość między sygnałami

Rozpatrywanie sygnałów jako elementów przestrzeni funkcyjnej $\mathscr{P}$ rozpiętej na ortonormalnej bazie $\{\varphi_1(t), \ldots, \varphi_N(t)\}$ umożliwia jednoznaczne określenie odległości między dwoma sygnałami należącymi do tej przestrzeni. Odległość $\rho(x, y)$ między dwoma sygnałami $x(t)$, $y(t) \in \mathscr{P}$ jest określona wzorem (por. wzór (2.7)):

$$\rho(x, y) = \left[\int_0^T [x(t) - y(t)]^2\,\mathrm{d}t\right]^{1/2}. \tag{13.5}$$

Jeżeli sygnały $x(t)$, $y(t) \in \mathscr{P}$ są reprezentowane w przestrzeni $\mathscr{R}^N$ wektorami $\boldsymbol{x} = [x_1, \ldots, x_N]^T$ i odpowiednio $\boldsymbol{y} = [y_1, \ldots, y_N]^T$, to z właściwości zachowania norm w przestrzeniach $\mathscr{P}$ i $\mathscr{R}^N$ wynika, że:

$$\rho(x, y) = \rho(\boldsymbol{x}, \boldsymbol{y}) = \left[\sum_{j=1}^{N} (x_j - y_j)^2\right]^{1/2}. \tag{13.6}$$

Oznacza to, że odległość między dwoma sygnałami $x(t)$ i $y(t)$ przestrzeni $\mathscr{P}$ może być mierzona zwykłą euklidesowską odległością między odpowiadającymi im wektorami $\boldsymbol{x}$ i $\boldsymbol{y}$ w przestrzeni $\mathscr{R}^N$.

13.2.3. Detekcja sygnałów zmodulowanych cyfrowo

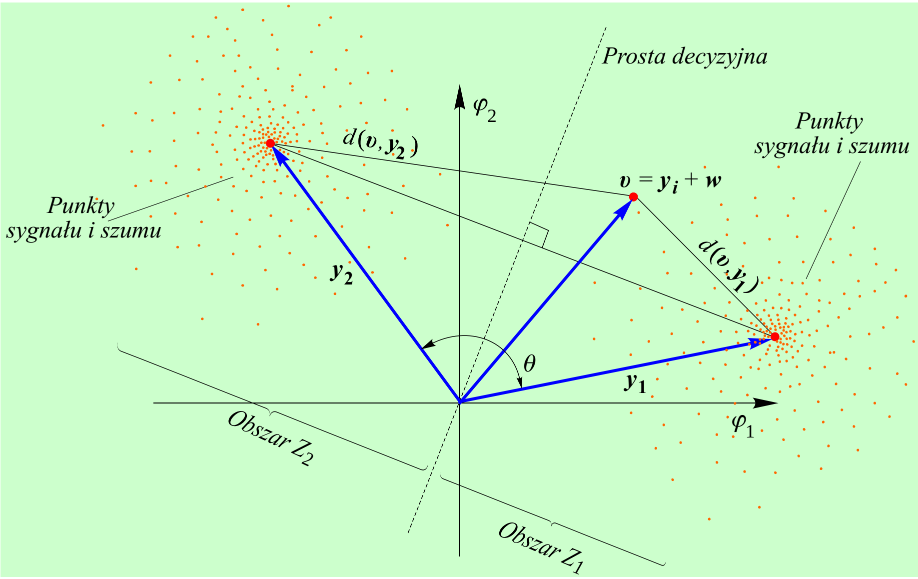
Sformalizowanie pojęcia odległości między sygnałami ma fundamentalne znaczenie dla prawidłowego przeprowadzenia operacji demodulacji. Demodulacja sygnałów zmodulowanych cyfrowo polega na detekcji sygnałów nadawanych w kolejnych przedziałach symbolowych T na podstawie obserwacji sygnałów odbieranych w tych przedziałach. Inaczej mówiąc, demodulacja sygnału sprowadza się do rozstrzygnięcia w każdym przedziale T, który z sygnałów $y_i(t)$ został nadany, jeśli odebrany został sygnał zakłócony $v(t) = y_i(t) + w(t)$.

Sygnałowi $v(t)$ odpowiada pewien wektor $\boldsymbol{v} = \boldsymbol{y}_i + \boldsymbol{w}$, gdzie $\boldsymbol{y}_i$ jest wektorem przyporządkowanym sygnałowi niezakłóconemu $y_i(t)$, natomiast $\boldsymbol{w}$ jest wektorem w pełni określonym przez przebieg (realizację) $w(t)$ szumu na odcinku czasu T. Ponieważ szum jest sygnałem losowym, długość i kierunek wektora $\boldsymbol{w}$ są także losowe. Przyjmiemy upraszczające założenie, że realizacja szumu $w(t)$ w przedziale czasu T należy do przestrzeni $\mathscr{P}$ rozpiętej na ortonormalnej bazie $\{\varphi_1(t), \ldots, \varphi_N(t)\}$, a więc do tej samej przestrzeni, do której należą sygnały transmitowane $y_k(t)$, $k = 1, \ldots, M$. Przy tym założeniu elementem przestrzeni $\mathscr{P}$ będzie także sygnał odebrany $v(t)$. Sygnał ten może różnić się znacznie od wszystkich sygnałów $y_k(t)$, w tym także od faktycznie nadanego sygnału $y_i(t)$. Zadaniem układu detekcji po stronie odbiorczej jest podjęcie decyzji, który z sygnałów $y_k(t)$ został nadany, tak aby zminimalizować prawdopodobieństwo popełnienia błędu. Reguła decyzyjna (optymalna w sensie kryterium największej wiarygodności – por. np. [2], p. ?), sprowadza się do wyznaczenia najmniejszej spośród odległości $\rho(\boldsymbol{v}, \boldsymbol{y}_k)$, $k = 1, \ldots, M$, między wektorem $\boldsymbol{v}$, reprezentującym sygnał odebrany $v(t)$, i wektorami $\boldsymbol{y}_k$, reprezentującymi sygnały $y_k(t)$. Za nadany uznajemy ten z sygnałów $y_k(t)$, którego wektor $\boldsymbol{y}_k$ jest najbliższy wektorowi $\boldsymbol{v}$.

13.2.4. Podział przestrzeni sygnałów na obszary decyzyjne

Optymalna reguła decyzyjna dzieli przestrzeń sygnałów na *obszary decyzyjne*. Interpretacja tego podziału jest przedstawiona na rys. 13.2 dla przypadku $M = 2$ i $N = 2$ przy założeniu, że transmitowane są dwa sygnały $y_1(t)$ i $y_2(t)$ o jednakowych amplitudach. Równość amplitud oznacza równość długości wektorów $\boldsymbol{y}_1$ i $\boldsymbol{y}_2$ reprezentujących te sygnały. Sygnał odebrany jest reprezentowany wektorem $\boldsymbol{v} = \boldsymbol{y}_i + \boldsymbol{w}$. W ilustrowanym przypadku dwuwymiarowa przestrzeń (płaszczyzna) sygnałów jest podzielona na dwa obszary decyzyjne Z_1 i Z_2 dwusieczną kąta θ między wektorami $\boldsymbol{y}_1$ i $\boldsymbol{y}_2$. Jeśli punkt odpowiadający sygnałowi odebranemu znajduje się w obszarze Z_1, podejmowana jest decyzja, że został nadany sygnał $y_1(t)$. Jeśli natomiast należy on do obszaru Z_2, podejmowana jest decyzja, że został nadany sygnał $y_2(t)$.

Tak określona reguła decyzyjna minimalizuje średnie prawdopodobieństwo błędnej transmisji znaku binarnego (tzw. *bitową stopę błędów BER*; ang. *Bit-Error-Rate*). W podobny sposób kształtowane są reguły decyzyjne dla M-wartościowych systemów modulacji dla $M > 2$. W każdym przypadku minimalizowana jest bitowa stopa błędów BER.



Rys. 13.2. Dwuwymiarowa przestrzeń sygnałów i jej podział na obszary decyzyjne

Z omówionej wyżej optymalnej reguły decyzyjnej wynika praktyczna realizacja detektorów rozstrzygających, do którego z obszarów decyzyjnych należy odebrany sygnał. Układ detektora może być zrealizowany w postaci banku korelatorów wyznaczających iloczyny skalarne $v_j = (v, \varphi_j)_{L^2(0,T)}$ sygnału odebranego $v(t)$ z funkcjami bazowymi $\varphi_j(t)$. Iloczyny te są następnie porównywane z wartościami odpowiednio ustalonych progów. Innym rozwiązaniem układowym jest bank filtrów dopasowanych (por. rys. 13.7). W następnym punkcie omówimy konkretne przykłady realizacji układowej detektorów dla poszczególnych rodzajów modulacji.

13.3. Modulacje PSK i FSK

13.3.1. Modulacja 2PSK

W systemie modulacji cyfrowej 2PSK z binarnym kluczowaniem fazy nadawane są dwa sygnały impulsowe o postaci:

$$y_i(t) = Y_0 \cos(\Omega t + \phi_i) = \sqrt{\frac{2E_b}{T_b}} \cos(\Omega t + \phi_i), \quad 0 \leqslant t \leqslant T_b, \quad i = 1, 2, \tag{13.7}$$

gdzie Y_0 jest amplitudą, Ω – pulsacją nośną, a ϕ_i – fazą chwilową:

$$\phi_i = \begin{cases} 0, & \text{gdy} \quad i = 1, \\ \pi, & \text{gdy} \quad i = 2\,. \end{cases} \tag{13.8}$$

Oba sygnały mają jednakowy czas trwania T_b, równy czasowi transmisji jednego znaku binarnego (i równy w tym przypadku czasowi T transmisji symbolu) oraz jednakową energię $E_b = Y_0^2 T_b/2$. W zapisie analitycznym tych sygnałów wygodnie jest wyrażać amplitudę Y_0 przez parametry E_b i T_b.

Czas trwania T_b sygnałów $y_i(t)$ powinien obejmować całkowitą liczbę okresów fali nośnej. Aby warunek ten był spełniony, wartość częstotliwości $F = \Omega/2\pi$ fali nośnej musi być całkowitą wielokrotnością częstotliwości kluczowania $1/T_b$, tzn. musi zachodzić równość $F = k\,/T_b$, gdzie k jest dużą liczbą całkowitą.

Sygnały $y_1(t)$ i $y_2(t)$ reprezentują symbole binarne „1” oraz „0”:

$$\begin{aligned} y_1(t) &= \sqrt{\frac{2E_b}{T_b}} \cos \Omega t \quad &&– \quad \text{symbol „1”}, \\ y_2(t) &= -\sqrt{\frac{2E_b}{T_b}} \cos \Omega t \quad &&– \quad \text{symbol „0”}. \end{aligned} \tag{13.9}$$

Z postaci tych sygnałów wynika, że przestrzeń sygnałów w przypadku modulacji 2PSK jest przestrzenią jednowymiarową ($N = 1$), której baza zawiera tylko jeden element:

$$\varphi_1(t) = \sqrt{\frac{2}{T_b}} \cos \Omega t, \quad 0 \leqslant t \leqslant T_b\,. \tag{13.10}$$

Sygnały $y_1(t)$ i $y_2(t)$ wyrażają się przez ten element następująco:

$$\begin{aligned} y_1(t) &= \sqrt{E_b}\varphi_1(t), \quad &&0 \leqslant t \leqslant T_b, \\ y_2(t) &= -\sqrt{E_b}\varphi_1(t), \quad &&0 \leqslant t \leqslant T_b\,. \end{aligned} \tag{13.11}$$

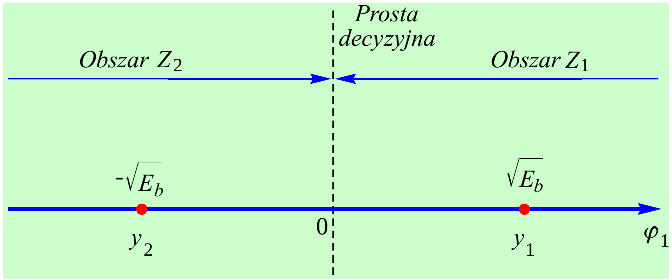
Przestrzeń i konstelacja sygnałów 2PSK są przedstawione na rys. 13.3. Tworzą ją dwa punkty y_1 i y_2 należące do przestrzeni sygnałów (w tym przypadku jest nią linia prosta), o współrzędnych:

$$y_{11} = \int_0^{T_b} y_1(t)\varphi_1(t)\,\mathrm{d}t = \sqrt{E_b} \tag{13.12}$$

oraz

$$y_{21} = \int_0^{T_b} y_2(t)\varphi_1(t)\,\mathrm{d}t = -\sqrt{E_b}\,. \tag{13.13}$$

Odległość między sygnałami $y_1(t)$ i $y_2(t)$ w przypadku modulacji 2PSK wynosi więc $2\sqrt{E_b}$.



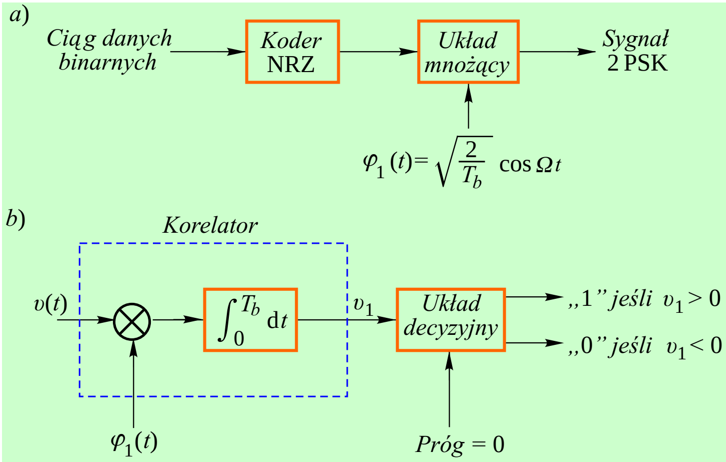
Rys. 13.3. Przestrzeń i konstelacja sygnałów 2PSK

Prosta sygnałów jest podzielona punktem $v = 0$ na dwa obszary decyzyjne Z_1 i Z_2, które stanowią dwie półproste: $0 < v < \infty$ i odpowiednio $-\infty < v < 0$. Jeżeli współrzędna odebranego sygnału należy do półprostej $0 < v < \infty$, układ decyzyjny podejmuje decyzję, że został nadany sygnał $y_1(t)$ (znak binarny „1”). Jeżeli natomiast współrzędna odebranego sygnału należy do półprostej $-\infty < v < 0$, układ decyzyjny podejmuje decyzję, że został nadany sygnał $y_2(t)$ (znak binarny „0”).

13.3.2. Generacja i demodulacja sygnałów 2PSK

Schemat blokowy modulatora sygnałów 2PSK jest przedstawiony na rys. 13.4a. Wynika on bezpośrednio ze wzorów (13.9)–(13.11). Dane binarne w postaci ciągu symboli „1” i „0” są doprowadzone do układu, który zgodnie z kodem sygnałowym NRZ (por. rys. 12.17c) generuje bipolarny sygnał prostokątny. W poszczególnych przedziałach bitowych T_b sygnał ten przybiera wartość $\sqrt{E_b}$, gdy transmitowane są znaki binarne „1” oraz wartość $-\sqrt{E_b}$, gdy transmitowane są znaki binarne „0” (por. rys. 12.18c). Ze wzorów (13.9) wynika, że w celu wytworzenia sygnału 2PSK wystarczy tak uformowaną falę prostokątną podać na jedno z wejść układu mnożącego, na którego drugie wejście jest podany sygnał bazowy $\varphi_1(t) = \sqrt{2/T_b}\cos\Omega t$ pełniący zarazem rolę harmonicznej fali nośnej. Na wyjściu układu mnożącego otrzymujemy zmodulowany sygnał 2PSK. Sygnały na obu wejściach układu mnożącego są sterowane wspólnym zegarem.

Układ demodulatora sygnału 2PSK jest przedstawiony na rys. 13.4b. Sygnał odebrany $v(t)$, będący sumą sygnału 2PSK i szumu $w(t)$, jest mnożony w każdym przedziale bitowym przez koherentny sygnał nośny $\varphi_1(t)$ wytwarzany przez



Rys. 13.4. Schematy blokowe modulatora (a) i demodulatora (b) sygnałów 2PSK

lokalny generator fali harmonicznej o tej samej częstotliwości. Sygnał iloczynowy jest z kolei podawany na wejście integratora, na którego wyjściu w chwili kończącej kolejny przedział bitowy otrzymujemy pewną liczbę $v_1 = (v, \varphi_1)_{L^2(0,T_b)}$. Generator lokalny, układ mnożący i integrator tworzą tzw. *detektor korelacyjny*. Wyjście v_1 detektora korelacyjnego jest porównywane z progiem równym zeru i układ decyzyjny podejmuje decyzję. Jeżeli $v_1 > 0$, zostaje podjęta decyzja o nadaniu znaku binarnego „1”, jeżeli zaś $v_1 < 0$, zostaje podjęta decyzja o nadaniu znaku binarnego „0”. W przypadku, gdy $v_1 = 0$, decyzja jest losowa. Operacje te są powtarzane w kolejnych przedziałach bitowych.

13.3.3. Modulacja 2FSK

W systemie modulacji 2FSK z binarnym kluczowaniem częstotliwości w każdym przedziale bitowym nadawany jest jeden z dwóch sygnałów:

$$y_i(t) = Y_0 \cos 2\pi F_i t = \sqrt{\frac{2E_b}{T_b}} \cos 2\pi F_i t, \qquad i = 1, 2, \qquad 0 \leqslant t \leqslant T_b. \quad (13.14)$$

Oba sygnały mają jednakowy czas trwania T_b, jednakowe amplitudy Y_0 i energie E_b oraz różne częstotliwości F_1 i $F_2 > F_1$. Sygnał $y_1(t)$ o częstotliwości nośnej F_1 reprezentuje znak binarny „1”, natomiast sygnał $y_2(t)$ o częstotliwości nośnej F_2 reprezentuje znak binarny „0”. Różnica $F_2 - F_1$ nosi nazwę *rozstawu częstotliwości*.

Przyjmiemy założenie, że częstotliwości nośne F_i obu sygnałów są dobrane tak, aby spełniona była równość:

$$F_i = \frac{n_0 + i}{T_b}, \qquad i = 1, 2, \quad (13.15)$$

gdzie n_0 jest odpowiednio dużą liczbą całkowitą. Modulacja 2FSK przy takim doborze częstotliwości nośnych jest nazywany *modulacją Sunde'a*. Jak można łatwo sprawdzić, w systemie modulacji Sunde'a zapewniona jest ciągłość fazy sygnału 2FSK w każdej chwili czasu, a więc także w chwilach kluczowania częstotliwości, a ponadto oba sygnały $y_1(t)$ i $y_2(t)$ są ortogonalne. Rozstaw częstotliwości w przypadku modulacji 2FSK Sunde'a wynosi $1/T_b$, a więc jest równy odwrotności czasu transmisji jednego znaku binarnego.

Przestrzeń sygnałów 2FSK jest dwuwymiarowa ($N = 2$). Jej bazę tworzą dwa sygnały:

$$\varphi_i(t) = \sqrt{\frac{2}{T_b}} \cos 2\pi F_i t, \qquad i = 1, 2, \qquad 0 \leqslant t \leqslant T_b, \tag{13.16}$$

gdzie częstotliwości nośne F_i są określone wzorem (13.15). Sygnały $y_1(t)$ i $y_2(t)$ wyrażają się przez sygnały bazowe następująco:

$$y_1(t) = \sqrt{E_b}\varphi_1(t), \qquad y_2(t) = \sqrt{E_b}\varphi_2(t), \qquad 0 \leqslant t \leqslant T_b\,. \tag{13.17}$$

Są one zatem reprezentowane wektorami

$$\boldsymbol{y}_1 = \left[\sqrt{E_b}, 0\right]^T, \qquad \boldsymbol{y}_2 = \left[0, \sqrt{E_b}\right]^T. \tag{13.18}$$

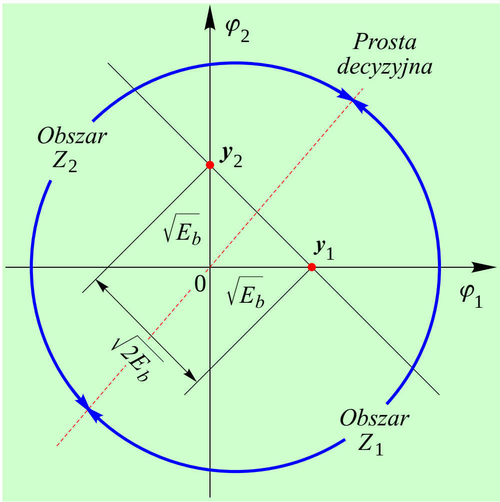
Konstelacja sygnałów 2FSK jest przedstawiona na rys. 13.5. Odległość między sygnałami 2FSK $y_1(t)$ i $y_2(t)$ przy założeniu tej samej energii E_b jest mniejsza niż w przypadku sygnałów 2PSK i wynosi $\sqrt{2E_b}$. Prosta decyzyjna jest symetralną odcinka łączącego punkty $\boldsymbol{y}_1$ i $\boldsymbol{y}_2$ i dzieli płaszczyznę sygnałów na dwa obszary decyzyjne Z_1 i Z_2. Jeżeli odebranemu zakłóconemu szumem sygnałowi $v(t)$ odpowiada w przestrzeni sygnałów wektor $\boldsymbol{v} = [v_1, v_2]^T$ o współrzędnych

$$v_j = \int_0^{T_b} v(t)\varphi_j(t)\,\mathrm{d}t, \qquad j = 1, 2, \tag{13.19}$$

należących do obszaru Z_1, podejmowana jest decyzja o nadaniu znaku „1". Gdy współrzędne te należą do obszaru Z_2, podejmowana jest decyzja o nadaniu znaku „0".

13.3.4. Generacja i demodulacja sygnałów 2FSK

Schemat blokowy modulatora sygnału 2FSK jest pokazany na rys. 13.6a. Ciąg znaków binarnych „1" i „0" jest podawany na układ kodera poziomu, który wytwarza unipolarny sygnał prostokątny przyjmujący stały poziom $\sqrt{E_b}$ w tych przedziałach bitowych, w których nadawany jest znak binarny „1" i poziom zero, gdy nadawany jest znak binarny „0". Koder poziomu koduje zatem ciąg danych binarnych kodem sygnałowym pokazanym na rys. 12.17b. W dolnym torze modulatora jest włączony inwerter, na którego wyjściu występuje odwrócony sygnał

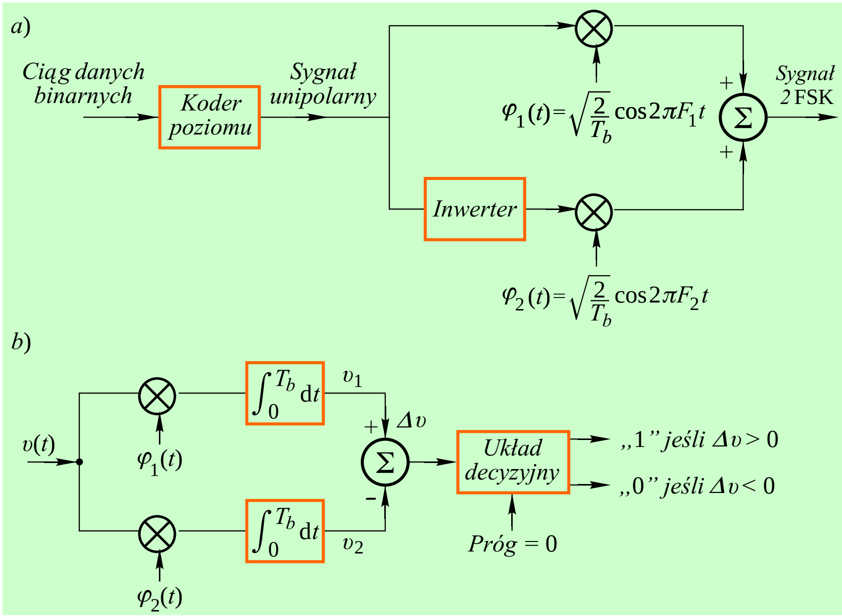


Rys. 13.5. Przestrzeń i konstelacja sygnałów 2FSK

unipolarny, tj. sygnał przyjmujący poziom zero, gdy na wyjściu kodera występuje poziom $\sqrt{E_b}$, oraz poziom $\sqrt{E_b}$, gdy na wyjściu kodera występuje poziom zero. Tak więc, jeśli nadawany jest znak binarny „1", włączony jest lokalny generator fali nośnej $\varphi_1(t) = \sqrt{2/T_b}\cos 2\pi F_1 t$ o częstotliwości F_1 w torze górnym i wyłączony lokalny generator fali nośnej $\varphi_2(t) = \sqrt{2/T_b}\cos 2\pi F_2 t$ o częstotliwości F_2 w torze dolnym. W przypadku, gdy nadawany jest znak binarny „0", sytuacja jest odwrotna. Na wyjściu sumatora otrzymujemy zmodulowany sygnał 2FSK. Oba generatory lokalne są w odpowiedni sposób synchronizowane, tak aby zapewnić ciągłość fazy sygnału zmodulowanego.

Innym sposobem generacji sygnału 2FSK jest zastosowanie oscylatora VCO (por. p. 11.3.1) kluczowanego sygnałem binarnym unipolarnym z wyjścia kodera. Także i w tym przypadku jest zapewniona ciągłość fazy sygnału 2FSK.

Odbiór sygnałów 2FSK może być koherentny lub niekoherentny. Schemat blokowy odbiornika koherentnego jest pokazany na rys. 13.6b. Składa się on z dwóch detektorów korelacyjnych zawierających lokalne generatory bazowych przebiegów nośnych $\varphi_1(t)$ i $\varphi_2(t)$. Sygnał odebrany $v(t)$ jest mnożony przez sygnały bazowe w układach mnożących. Sygnały iloczynowe z wyjść układów mnożących są podawane na wejścia integratorów, które zgodnie ze wzorem (13.19) obliczają w chwilach kończących kolejne przedziały bitowe współrzędne v_1 i v_2 odebranego sygnału. Decyzja jest podejmowana na podstawie znaku różnicy $\Delta v = v_1 - v_2$ tych współrzędnych, która jest porównywana w układzie decyzyjnym z progiem równym zeru. Gdy $\Delta v > 0$, tzn. współrzędna v_1 jest większa od współrzędnej v_2, wówczas punkt $\boldsymbol{v}$, odpowiadający na płaszczyźnie sygnałów z rys. 13.5 sygnałowi odebranemu, leży w obszarze decyzyjnym Z_1 i podejmowana jest decyzja o nadaniu znaku binarnego „1". Gdy $\Delta v < 0$, punkt $\boldsymbol{v}$ leży w obszarze decyzyjnym Z_2 i podejmowana jest decyzja o nadaniu znaku binarnego „0".

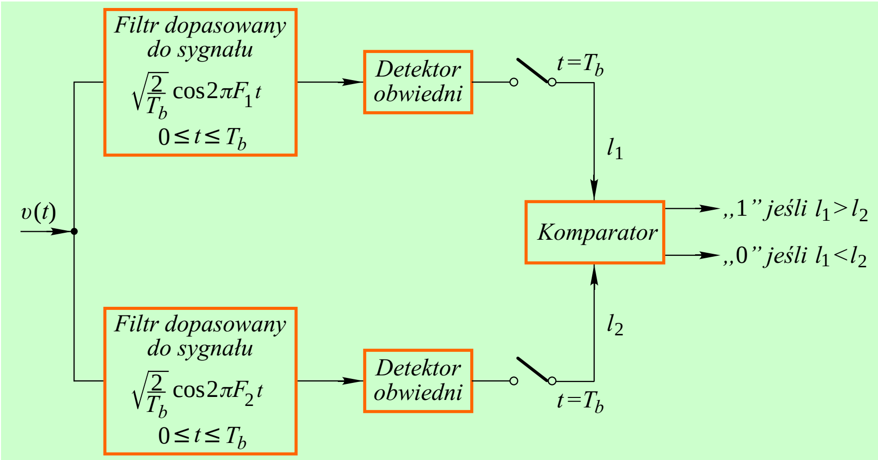


Rys. 13.6. Schematy blokowe modulatora (a) i demodulatora (b) sygnałów 2FSK

W przypadku odbioru niekoherentnego sygnałów 2FSK demodulator (rys. 13.7) zawiera dwa filtry o odpowiedziach impulsowych $h_i(t) = \sqrt{2/T_b}\varphi_i(T_b - t)$, $i = 1, 2$. Filtry o takich odpowiedziach impulsowych nazywamy *dopasowanymi* do sygnałów bazowych $\varphi_i(t)$. Sygnały z wyjść filtrów dopasowanych są podawane na detektory obwiedni (por. p. 10.3.5). Z kolei sygnały wyjściowe detektorów obwiedni są próbkowane w chwili $t = T_b$ (próbkowanie w tej chwili zapewnia największy stosunek sygnał-szum dla pobranej próbki). Jeżeli wartość l_1 próbki obwiedni w górnym kanale odbiornika jest większa od wartości próbki l_2 w kanale dolnym, podejmowana jest decyzja o nadaniu znaku binarnego „1”. W przeciwnym przypadku podejmowana jest decyzja o nadaniu znaku binarnego „0”.

13.3.5. M-wartościowe modulacje PSK i FSK

W podobny sposób jak w binarnych systemach 2PSK i 2FSK są generowane sygnały w M-wartościowych systemach kluczowania fazy i częstotliwości, gdzie $M = 2^k$, $k = 2, 3, \ldots$ W systemach tych informacja źródłowa jest przesyłana w postaci M symboli $m_1, \ldots, m_M$, reprezentowanych ciągami binarnymi o długości k. Każdy symbol jest przesyłany w odcinku czasu o długości T równej długości przedziału symbolowego i obejmującym $\log_2 M$ przedziałów bitowych T_b. W modulatorze symbolom tym są przyporządkowane sygnały $y_1(t), \ldots, y_M(t)$,



Rys. 13.7. Schemat blokowy demodulatora sygnału 2FSK w przypadku odbioru niekoherentnego

będące odcinkami harmonicznej fali nośnej o długości T, których faza albo częstotliwość przybiera jedną z M, zwykle równoodległych, wartości. Sygnały te są następnie transmitowane do odbiornika.

Sygnały generowane w systemie 4PSK, nazywanym także systemem QPSK, zostały omówione już wcześniej (przy innych oznaczeniach) w przykładzie 2.1 (p. 2). Posłużył on nam wówczas do interpretacji pojęcia przestrzeni funkcyjnej sygnałów oraz jej bazy i wymiaru. Obecnie przykład 2.1 możemy prześledzić ponownie pod kątem zastosowania tych pojęć do opisu, analizy, a przede wszystkim optymalnej detekcji sygnałów zmodulowanych cyfrowo. Wynika z niego m.in., że przestrzeń sygnałów QPSK jest dwuwymiarowa. Przebiegi sygnałów generowanych w systemie QPSK zostały przedstawione na rys. 2.1, a ich konstelacja – na rys. 2.2.

Generacja i demodulacja wielowartościowych sygnałów PSK i FSK jest realizowana według podobnych zasad, co w przypadku modulacji binarnych. Wzrasta jedynie stopień złożoności modulatorów i demodulatorów. Nie byłoby zatem celowe stosowanie wielowartościowych systemów modulacji cyfrowej, gdyby nie wynikające stąd korzyści. Zastępowanie przez projektantów systemów transmisji danych prostych modulacji binarnych znacznie bardziej złożonymi modulacjami wielowartościowymi jest podyktowane przede wszystkim chęcią coraz bardziej efektywnego wykorzystania zdolności transmisyjnych kanału, a więc dążeniem do przesyłania jak największej informacji w jak najwęższym paśmie i przy najmniejszej możliwej bitowej stopie błędów BER.

Ogólnie można powiedzieć, że przy ustalonej wymaganej wartości BER, ustalonej mocy nadajnika i stałym poziomie (stałej gęstości widmowej) szumu, zwiększanie liczby M umożliwia przesłanie tej samej informacji w węższym paśmie. Na przykład, dokładna analiza wykazuje, że przy tej samej wartości BER, tej samej mocy nadajnika i tym samym poziomie szumu, w systemie QPSK można w ciągu jednej sekundy przesłać taką samą ilość informacji co w systemie 2PSK, zajmując przy tym w kanale dwa razy mniejsze pasmo. Oznacza to, że w systemie QPSK w tym samym paśmie można przesłać w ciągu jednej sekundy dwukrotnie (w systemach M-PSK $\log_2 M$-krotnie) więcej informacji w porównaniu z systemem 2PSK (por. p. 13.4.7). Oznacza to także, że w systemie QPSK w tym samym paśmie można przesyłać informację z tą samą szybkością co w systemie 2PSK, ale przy mniejszej mocy nadajnika.

Nieco gorsze właściwości mają wielowartościowe systemy FSK. Przy założonej wartości BER umożliwiają one także redukcję mocy nadajnika, ale uzyskuje się to kosztem poszerzenia ich pasma.

13.3.6. Modulacja QAM

Z porównania konstelacji sygnałów 2PSK (rys. 13.3) i QPSK (rys. 2.2) wynika, że – przy założonej wartość energii E_b przypadającej na jeden transmitowany bit – odległość między sygnałami w obu systemach jest jednakowa i równa $2\sqrt{E_b}$. Wynika to z faktu, że w systemie QPSK $Y_0 = \sqrt{2E/T}$, $E = 2E_b$ oraz $T = 2T_b$. Jednak począwszy od $M = 8$ w systemach M-PSK odległość między najbliżej położonymi punktami konstelacji maleje. Oznacza to trudniejsze warunki poprawnego rozróżniania transmitowanych sygnałów $y_i(t)$ po stronie odbiorczej przez układ detektora, a w konsekwencji – zmniejszenie prawdopodobieństwa poprawnej detekcji sygnału M-PSK. Tak więc, mimo że systemy M-PSK są potencjalnie bardziej atrakcyjne dla zastosowań praktycznych z uwagi na węższe pasmo zajęte w kanale niezbędne do przesłania tej samej informacji w ciągu sekundy, ich odporność na zakłócenia maleje ze wzrostem M.

Ze względu na mniejszą odporność na zakłócenia systemów M-PSK są one w praktyce wypierane przez wielowartościowe kwadraturowe modulacje amplitudy QAM z jednoczesnym kluczowaniem amplitudy i fazy. Na przykład, w systemie 32-wartościowym QAM, stosowanym jako jeden ze standardów w łączności modemowej, transmitowane symbole są kodowane za pomocą czterech wartości amplitudy impulsów i ośmiu wartości fazy początkowej. Przy założonej energii E_b wymaganej do przesłania jednego bitu, M-wartościowy system QAM umożliwia podobną redukcję pasma jak system M-PSK i zapewnia przy tym większe prawdopodobieństwo poprawnej detekcji. Jest to konsekwencją korzystniejszej konstelacji sygnału M-QAM. Odległości między poszczególnymi jej punktami są bowiem większe niż w przypadku sygnału M-PSK. Systemy QAM wymagają jednak bardzo dokładnego zapewnienia liniowości kanału w całym zakresie zmian amplitudy sygnałów.

Dokładniejsze omówienie wielowartościowych systemów modulacji cyfrowych można znaleźć np. w [1], p. 8.13, 8.19.

13.4. Analiza widmowa sygnałów zmodulowanych cyfrowo

13.4.1. Związek między widmem sygnału a widmem jego obwiedni zespolonej

W badaniu właściwości i porównaniu różnych systemów modulacji cyfrowej pod kątem ich zastosowań w praktyce ważną rolę odgrywa analiza widmowa sygnałów zmodulowanych. Ponieważ sygnały te są wąskopasmowe, w celu zbadania ich widm wygodnie jest posłużyć się ich reprezentacją za pomocą drgania uogólnionego (por. p. 10.1.4). Reprezentację tę wykorzystamy do ustalenia związku między widmem sygnału a widmem jego obwiedni zespolonej. Na tej podstawie będzie można w prosty sposób wyznaczać widma sygnałów zmodulowanych cyfrowo.

Sygnały zmodulowane cyfrowo są ciągami sygnałów $y_i(t)$ transmitowanych w kolejnych odcinkach czasu T zgodnie z bieżącymi symbolami m_i. Ponieważ informacja źródłowa zakodowana w tych symbolach jest losowa, sygnały zmodulowane cyfrowo mają charakter losowy. Jednak podobnie jak w przypadku analizy widmowej sygnałów zmodulowanych analogowo, w celu zbadania podstawowych cech widmowych sygnałów zmodulowanych cyfrowo wystarczy rozpatrzyć deterministyczne modele sygnału modulującego i sygnału zmodulowanego.

Założymy, że sygnał $y(t)$ zmodulowany cyfrowo jest reprezentowany drganiem uogólnionym określonym względem pulsacji Ω (por. definicję 10.5):

$$y(t) = Y(t)\cos[\Omega t + \varphi(t)], \tag{13.20}$$

gdzie $Y(t)$ jest amplitudą chwilową (obwiednią rzeczywistą), a $\varphi(t)$ – fazą chwilową. W przypadku sygnałów z kluczowaniem amplitudy i fazy jako pulsację Ω przyjmuje się pulsację nośną, natomiast w przypadku sygnałów z kluczowaniem częstotliwości – średnią arytmetyczną pulsacji nośnych (por. p. 13.4.3).

Drganie uogólnione (13.20) można przedstawić w postaci:

$$y(t) = \mathrm{Re}\left\{Y(t)\,\mathrm{e}^{\mathrm{j}[\Omega t+\varphi(t)]}\right\} = \mathrm{Re}\left\{\boldsymbol{Y}(t)\,\mathrm{e}^{\mathrm{j}\Omega t}\right\}, \tag{13.21}$$

gdzie

$$\boldsymbol{Y}(t) = Y(t)\,\mathrm{e}^{\mathrm{j}\varphi(t)} \tag{13.22}$$

jest obwiednią zespoloną sygnału $y(t)$. Przypomnijmy, że obwiednia zespolona jest sygnałem wolnozmiennym w porównaniu z falą nośną.

Ponieważ część rzeczywistą liczby zespolonej z można zapisać jako $\mathrm{Re}\,z = (z + z^*)/2$, wzór (13.21) możemy przepisać w postaci:

$$y(t) = \frac{\boldsymbol{Y}(t)\,\mathrm{e}^{\mathrm{j}\Omega t} + \boldsymbol{Y}^*(t)\,\mathrm{e}^{-\mathrm{j}\Omega t}}{2}. \tag{13.23}$$

Wnioskujemy stąd, że widmo $Y(\omega)$ sygnału zmodulowanego $y(t)$ można wyznaczyć na podstawie widma $\boldsymbol{Y}(\omega)$ jego obwiedni zespolonej $\boldsymbol{Y}(t)$. Związek między tymi widmami wynika wprost ze struktury równości (13.23). Obliczając transformaty Fouriera (w sensie granicznym) obu stron tej równości i korzystając przy tym z twierdzenia o przesunięciu w dziedzinie częstotliwości (o mnożeniu przez sygnał $\exp(\mathrm{j}\,\Omega t)$ w dziedzinie czasu) oraz z właściwości (3.17) przekształcenia Fouriera, zgodnie z którą, jeśli $\mathscr{F}[\boldsymbol{Y}(t)] = \boldsymbol{Y}(\omega)$, to $\mathscr{F}[\boldsymbol{Y}^*(t)] = \boldsymbol{Y}^*(-\omega)$, otrzymujemy:

$$\begin{aligned} Y(\omega) &= \int_{-\infty}^{\infty} y(t)\,\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \\ &= \frac{1}{2}\left[\int_{-\infty}^{\infty} \boldsymbol{Y}(t)\,\mathrm{e}^{-\mathrm{j}(\omega-\Omega)t}\,\mathrm{d}t + \int_{-\infty}^{\infty} \boldsymbol{Y}^*(t)\,\mathrm{e}^{-\mathrm{j}(\omega+\Omega)t}\,\mathrm{d}t\right] = \\ &= \frac{1}{2}\left[\boldsymbol{Y}(\omega-\Omega) + \boldsymbol{Y}^*(-\omega-\Omega)\right] . \end{aligned} \tag{13.24}$$

Związek (13.24) stanowi punkt wyjścia do analizy widmowej sygnałów zmodulowanych cyfrowo.

13.4.2. Widmo sygnału 2PSK zmodulowanego okresową falą prostokątną

Założymy, że w systemie 2PSK ciąg danych binarnych jest sekwencją naprzemiennie powtarzanych znaków binarnych „0” i „1” (rys. 13.8a). W modulatorze sygnału 2PSK (por. rys. 13.4a) koder NRZ wytwarza wówczas okresową bipolarną falę prostokątną $x(t)$ o okresie $2T_b$ i pulsacji podstawowej $\omega_0 = \pi/T_b$ (rys. 13.8b), która jest następnie mnożona przez falę nośną w układzie mnożącym. W wyniku otrzymujemy szczególny przypadek sygnału 2PSK zmodulowanego deterministyczną falą prostokątną (rys. 13.8c). Przez analogię można powiedzieć, że przypadek ten odpowiada modulacji jednym tonem w analogowych systemach modulacji. Modulująca fala prostokątna pełni tu taką samą rolę prostego testowego deterministycznego sygnału modulującego, jak sygnał harmoniczny w przypadku modulacji analogowych. Analiza widma sygnału 2PSK zmodulowanego okresową falą prostokątną posłuży nam do wyciągnięcia ogólnych wniosków o strukturze widmowej sygnałów 2PSK.

Ze sposobu generacji sygnału 2PSK zmodulowanego okresową falą prostokątną $x(t)$ wynika, że sygnał ten jest iloczynem sygnału $x(t)$ i fali nośnej $Y_0 \cos\Omega t$, a więc jest równoważny sygnałowi AM-SC, którego funkcja modulująca $m(t)$ jest równa fali prostokątnej $x(t)$. Obwiednia zespolona $\boldsymbol{Y}(t) = Y_0 x(t)$ tego sygnału jest w tym przypadku rzeczywista, wobec czego wzór (13.24) upraszcza do postaci:

$$Y(\omega) = \frac{1}{2}[\boldsymbol{Y}(\omega-\Omega) + \boldsymbol{Y}(\omega+\Omega)], \tag{13.25}$$

gdzie widmo $\boldsymbol{Y}(\omega) = Y_0 X(\omega)$ obwiedni zespolonej sygnału zmodulowanego jest określone wzorem (por. wzór (3.53)):

$$\boldsymbol{Y}(\omega) = 2\pi Y_0 \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \mathrm{Sa}\left(\frac{k\pi}{2}\right) \delta\left(\omega - k\frac{\pi}{T_b}\right) . \tag{13.26}$$

Widmo amplitudowe $|\boldsymbol{Y}(\omega)|$ obwiedni zespolonej sygnału 2PSK zmodulowanego okresową falą prostokątną jest pokazane na rys. 13.8e. Z jego struktury wynika bezpośrednio struktura widma amplitudowego $|Y(\omega)|$ sygnału zmodulowanego $y(t)$. Widmo to jest rozszczepionym na dwie części widmem amplitudowym jego obwiedni zespolonej. Części te są przesunięte do punktów $\pm\Omega$ i mają dwukrotnie mniejszą gęstość widmową. Prawostronna część widma składa się z zespołu prążków skupionych wokół pulsacji nośnej Ω i oddalonych od tej pulsacji o $k\omega_0 = k\pi/T_b$, gdzie k jest liczbą nieparzystą (prążki dla k parzystych są równe zeru). Prążki te leżą na obwiedni $\pi Y_0 |\,\mathrm{Sa}[(\omega-\Omega)T_b/2]|$ i maleją w miarę oddalania się od pulsacji Ω, przy czym ich wysokości są odwrotnie proporcjonalne do k. Widmo $|Y(\omega)|$ nie zawiera prążka w punkcie Ω (sygnał zmodulowany nie zawiera składowej nośnej) i teoretycznie jest nieskończone. Jednak zasadnicza część mocy sygnału zmodulowanego jest zawarta w pierwszej parze prążków bocznych o pulsacjach $\Omega \pm \pi/T_b$. Moc zawarta w dalszych parach prążków bocznych maleje proporcjonalnie do k^2.

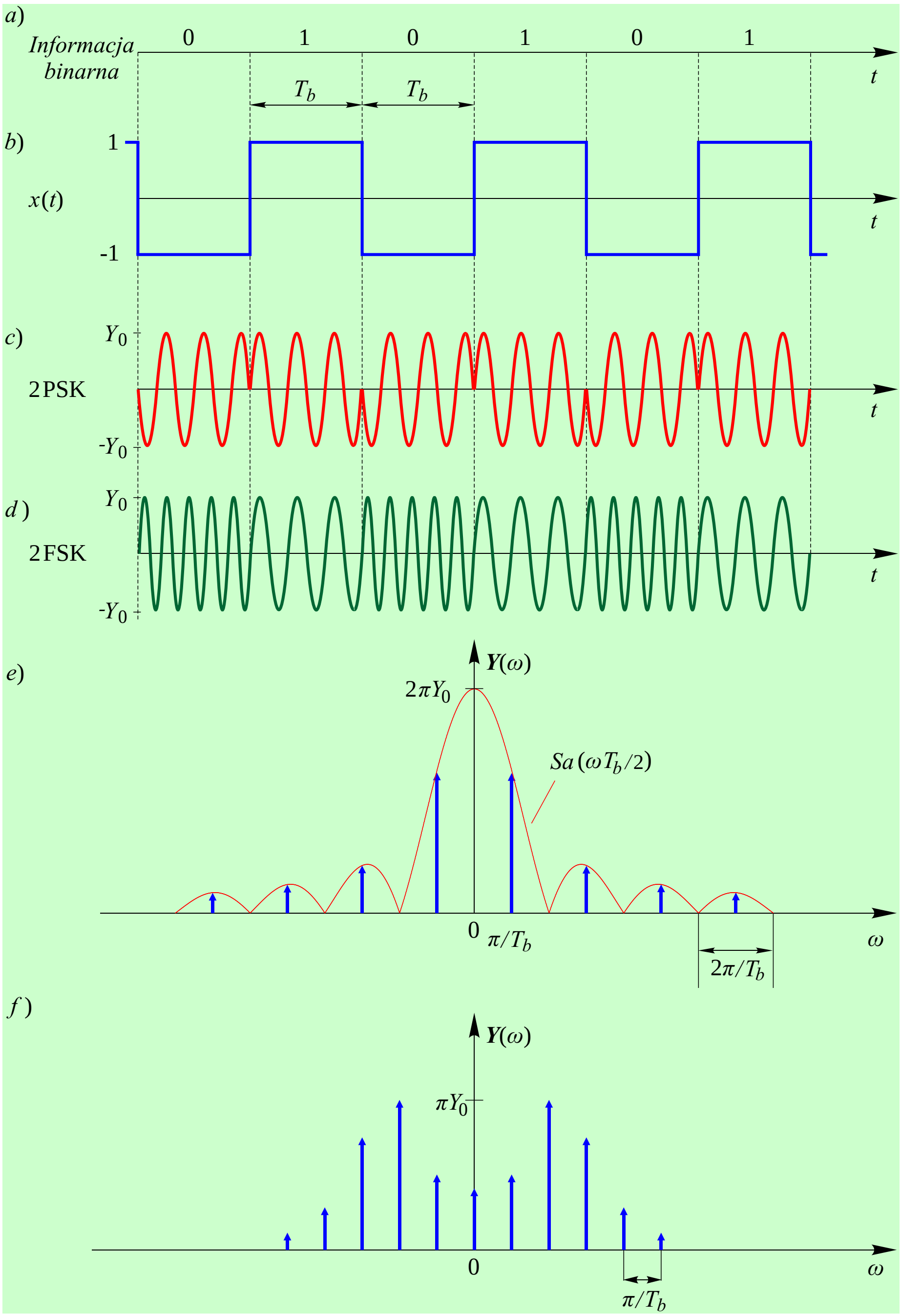
13.4.3. Widmo sygnału 2FSK zmodulowanego okresową falą prostokątną

Z kolei ze sposobu generacji sygnału 2FSK (por. p. 13.3.4) wynika, że sygnał 2FSK zmodulowany okresową falą prostokątną (rys. 13.8d) jest równoważny zmodulowanemu tą samą falą prostokątną sygnałowi FM o częstotliwości nośnej $F = (F_1 + F_2)/2$ równej średniej arytmetycznej częstotliwości F_1 i F_2 oraz dewiacji częstotliwości $\Delta F = (F_2 - F_1)/2$ równej połowie rozstawu częstotliwości $F_2 - F_1$. Analiza widma obwiedni zespolonej tego sygnału jest znacznie bardziej złożona. Można wykazać ([3], rozdz. 9), że widmo to jest określone wzorem:

$$\boldsymbol{Y}(\omega) = \pi Y_0 \sum_{k=-\infty}^{\infty} \left\{\mathrm{Sa}\left[\frac{\pi}{2}(k-l)\right] + (-1)^k \,\mathrm{Sa}\left[\frac{\pi}{2}(k+l)\right]\right\} \delta\left(\omega - k\frac{\pi}{T_b}\right), \tag{13.27}$$

gdzie $l = (F_2 - F_1)T_b$ jest rozstawem częstotliwości unormowanym względem czasu T_b trwania jednego znaku binarnego. Zauważmy, że w przypadku modulacji 2FSK Sunde'a, określonej warunkiem (13.15), unormowany rozstaw częstotliwości wynosi $l = 1$.

Widmo amplitudowe obwiedni zespolonej sygnału 2FSK zmodulowanego okresową falą prostokątną dla $l = 2{,}5$ jest pokazane na rys. 13.8f. Wynika stąd, że



Rys. 13.8. Sygnały 2PSK i 2FSK zmodulowane okresową falą prostokątną: ciąg znaków binarnych (a), sygnał modulujący (b), sygnał zmodulowany 2PSK (c), sygnał zmodulowany 2FSK (d), widmo obwiedni zespolonej sygnału 2PSK (e) oraz widmo obwiedni zespolonej sygnału 2FSK dla parametru $l = 2{,}5$ (f)

widmo amplitudowe sygnału 2FSK zawiera prążki rozłożone symetrycznie wokół pulsacji $\Omega = (\Omega_1 + \Omega_2)/2$ i oddalone od siebie o π/T_b. Podobnie jak w przypadku modulacji 2PSK, widmo to jest teoretycznie nieskończone. Ponieważ jednak sygnał 2FSK zmodulowany okresową falą prostokątną jest równoważny sygnałowi FM, więc – tak jak dla każdego sygnału FM – prążki leżące dostatecznie daleko od pulsacji Ω są na tyle małe, że można je pominąć. Tak więc również w przypadku sygnału 2FSK zmodulowanego okresową falą prostokątną zasadnicza część mocy jest zawarta w wąskim paśmie wokół pulsacji nośnej.

13.4.4. Widma mocy sygnałów 2PSK i 2FSK zmodulowanych dowolnym sygnałem

Przeprowadzona w poprzednich punktach analiza widmowa sygnałów 2PSK i 2FSK dla szczególnego przypadku okresowej prostokątnej fali modulującej daje jedynie orientacyjny pogląd na ogólny charakter widm tych sygnałów. W rzeczywistości sygnały modulujące są pewnymi nieokresowymi falami prostokątnymi, a więc widma sygnałów 2PSK i 2FSK są ciągłe. W celu dokładnego zbadania ich właściwości konieczne jest uwzględnienie losowego charakteru sygnałów modulującego i zmodulowanego oraz wprowadzenie ich odpowiednich modeli losowych. Wymagałoby to szerszego omówienia elementów teorii procesów stochastycznych, a w szczególności wprowadzenia opisu sygnałów losowych w dziedzinie częstotliwości. Z konieczności ograniczymy się do podania podstawowych wiadomości niezbędnych do sformułowania ostatecznych wniosków.

Podstawową charakterystyką sygnałów losowych w dziedzinie częstotliwości jest *widmo mocy* opisujące rozkład mocy sygnału wzdłuż osi pulsacji. Widmo mocy analogowego sygnału losowego $x(t)$ jest oznaczane zwykle $S_x(\omega)$. Odpowiednikiem związku (13.24) w przypadku sygnałów losowych jest związek:

$$S_y(\omega) = \frac{1}{4}\left[S_{\boldsymbol{Y}}(\omega - \Omega) + S_{\boldsymbol{Y}}(\omega + \Omega)\right], \tag{13.28}$$

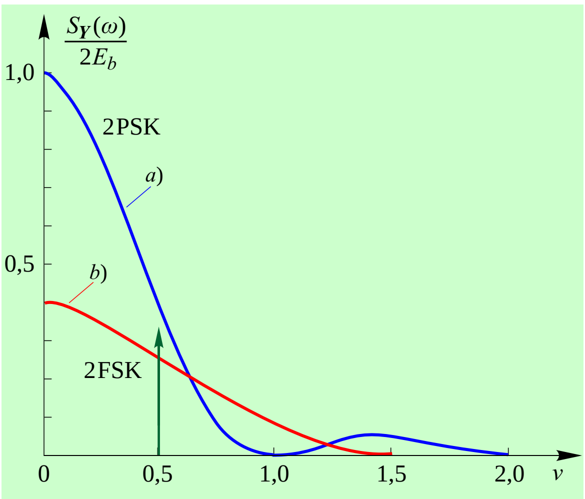
gdzie Ω jest pulsacją nośną, $S_y(\omega)$ widmem mocy sygnału zmodulowanego oraz $S_{\boldsymbol{Y}}(\omega)$ widmem mocy jego obwiedni zespolonej. Związek ten jest słuszny, jeśli oba składniki widma (13.28) są symetryczne względem pulsacji $\pm\Omega$. Warunek ten dla większości sygnałów zmodulowanych cyfrowo jest spełniony. Tak więc również w przypadku sygnałów losowych do wyznaczenia widma mocy sygnału zmodulowanego cyfrowo wystarcza znajomość widma mocy jego wolnozmiennej obwiedni zespolonej.

W przypadku modulacji 2PSK pulsacja Ω we wzorze (13.28) jest pulsacją nośną sygnału 2PSK, natomiast w przypadku modulacji 2FSK – średnią arytmetyczną $(\Omega_1 + \Omega_2)/2$ pulsacji nośnych. Pełna analiza widm mocy obwiedni zespolonej sygnałów 2PSK i 2FSK jest złożona. Można wykazać ([1],p. 8.20), że jeśli znaki binarne „1" i „0" przesyłane w poszczególnych przedziałach bitowych

T_b są statystycznie niezależne i równoprawdopodobne, to widmo mocy obwiedni zespolonej sygnału 2PSK jest opisane wzorem:

$$S_{\boldsymbol{Y}}(\omega) = 2E_b \, \mathrm{Sa}^2\left(\frac{\omega T_b}{2}\right), \tag{13.29}$$

gdzie E_b oznacza, tak jak poprzednio, energią pojedynczego impulsu $y_i(t)$ przypadającą na jeden transmitowany bit. Wykres prawostronnego widma (13.29) unormowanego względem wartości $2E_b$ w funkcji zmiennej unormowanej $\nu = \omega T_b/2\pi$ jest pokazany na rys. 13.9a. Zmienna ν jest częstotliwością unormowaną względem czasu T_b trwania znaku binarnego. Z wykresu tego i z zależności (13.28) wynika, że widmo mocy $S_y(\omega)$ sygnału 2PSK jest skupione wokół pulsacji $\pm\Omega$ i ma charakterystyczną strukturę listkową typu Sa^2. Maksima listków bocznych maleją zatem w miarę oddalania się od punktów $\pm\Omega$ z kwadratem pulsacji. Zasadnicza część mocy sygnału 2PSK jest skupiona w pasmach $\Omega - 2\pi/T_b \leqslant \omega \leqslant \Omega + 2\pi/T_b$ oraz $-\Omega - 2\pi/T_b \leqslant \omega \leqslant -\Omega + 2\pi/T_b$ o szerokości $4\pi/T_b$ obejmujących listki główne jego widma mocy.



Rys. 13.9. Unormowane widma mocy obwiedni zespolonych sygnałów: 2PSK (a) oraz sygnału 2FSK w przypadku modulacji Sunde'a (b)

Przy tych samych założeniach widmo mocy obwiedni zespolonej sygnału 2FSK w przypadku modulacji Sunde'a (tj. dla unormowanego rozstawu częstotliwości $l = 1$) jest opisane wzorem:

$$S_{\boldsymbol{Y}}(\omega) = \pi \frac{E_b}{T_b}\left[\delta\left(\omega - \frac{\pi}{T_b}\right) + \delta\left(\omega + \frac{\pi}{T_b}\right)\right] + \frac{8\pi^2 E_b \cos^2\left(\frac{\omega T_b}{2}\right)}{\left(\omega^2 T_b^2 - \pi^2\right)^2}. \tag{13.30}$$

Widmo to składa się z części dystrybucyjnej i części ciągłej o strukturze listkowej. Prawostronna część widma (13.30) unormowanego względem wartości $2E_b$

została wykreślona w funkcji zmiennej unormowanej $\nu = \omega T_b/2\pi$ na rys. 13.9b. Zauważmy, że ponieważ granica tego widma przy $\nu \to 0{,}5$ jest różna od zera, pierwsze jego zero występuje dopiero dla wartości $\nu = 1{,}5$.

Z wykresu 13.8b i wzoru (13.28) wynika, że widmo mocy sygnału 2FSK w przypadku modulacji Sunde'a składa się z części dystrybucyjnych i części ciągłych skupionych wokół punktów $\pm\Omega$, gdzie $\Omega = (\Omega_1 + \Omega_2)/2$. Składniki dystrybucyjne występują w punktach odległych o π/T_b od punktów $\pm\Omega$ (o $0{,}5$ w skali zmiennej ν). Natomiast maksima listków bocznych części ciągłych maleją w miarę oddalania się od punktów $\pm\Omega$ z czwartą potęgą pulsacji, a więc bardzo szybko. Zasadnicza część mocy sygnału 2FSK Sunde'a jest skupiona w pasmach $\Omega - 3\pi/T_b \leqslant \omega \leqslant \Omega + 3\pi/T_b$ oraz $-\Omega - 3\pi/T_b \leqslant \omega \leqslant -\Omega + 3\pi/T_b$, a więc szerszych niż w przypadku sygnału 2PSK.

Należy podkreślić, że obecność składowych harmonicznych w sygnale 2FSK Sunde'a, związanych ze składnikami dystrybucyjnymi jego widma mocy, jest wykorzystywana w praktyce do synchronizacji pracy odbiornika i nadajnika w przypadku koherentnego odbioru sygnałów 2FSK. Dodajmy także, że jeżeli częstotliwości nośne F_1 i F_2 sygnału 2FSK nie spełniają warunku Sunde'a (13.15), to jego widmo mocy maleje wolniej przy oddalaniu się od pulsacji $\pm\Omega$ (z kwadratem pulsacji).

13.4.5. Efektywność widmowa systemów 2PSK i 2FSK

Jak już podkreślaliśmy, w praktyce w systemach modulacji cyfrowych dąży się do przesłania jak największej informacji w jak najwęższym paśmie. Ważnym parametrem, określającym miarę wykorzystania pasma w danym systemie modulacji, jest jego *efektywność widmowa* (ang. *spectral efficiency*) zdefiniowana jako:

$$\rho = \frac{R_b}{B}\left[\frac{\mathrm{bit\,/s}}{\mathrm{Hz}}\right], \tag{13.31}$$

gdzie $R_b = 1/T_b$ jest *bitową szybkością transmisji* (ang. *bit rate*), wyrażoną w bitach na sekundę, zaś B – miarą szerokości pasma zajętego przez sygnał wyrażoną w hercach. Efektywność widmowa ρ systemu modulacji określa zatem liczbę bitów przesłanych w tym systemie w ciągu jednej sekundy przypadającą na 1 Hz pasma sygnału.

Bitowa szybkość transmisji R_b jest określona jednoznacznie. Trudności występują natomiast z jednoznacznym określeniem szerokości pasma B sygnału. Z teoretycznego punktu widzenia pasma sygnałów zmodulowanych cyfrowo są bowiem nieskończone (por. wzory (13.29) i (13.30) oraz wykresy na rys. 13.9). Jednak zasadnicza część mocy tych sygnałów jest zawsze zawarta w pewnym skończonym paśmie, którego szerokość może być określana w różny sposób, zgodnie z przyjętą arbitralnie definicją. Najczęściej jako szerokość pasma sygnału zmodulowanego cyfrowo przyjmuje się szerokość listka głównego jego widma mocy.

Z wykresów przedstawionych na rys. 13.9 wynika, że – przy tak przyjętej mierze – szerokości pasma sygnałów modulowanych w systemach 2PSK i 2FSK Sunde'a (wyrażone w hercach) wynoszą odpowiednio $B_{\text{2PSK}} = 2/T_b$ oraz $B_{\text{2FSK}} = 3/T_b$. Efektywności widmowe tych systemów są zatem równe $\rho_{\text{2PSK}} = 1/2$ oraz $\rho_{\text{2FSK}} = 1/3$. Widzimy więc, że pod tym względem system 2PSK ma przewagę nad systemem 2FSK Sunde'a.

13.4.6. Efektywność widmowa systemów M – PSK i M – FSK

Aby przybliżyć problemy, które muszą rozwiązywać w praktyce projektanci systemów transmisji danych, celowe jest porównanie efektywności widmowych systemów 2PSK i 2FSK z systemami wielowartościowymi PSK i FSK. Porównanie to przeprowadzimy przy założeniu jednakowych przedziałów bitowych T_b i jednakowych energii E_b przypadających na jeden transmitowany bit, tj. jednakowych amplitud sygnałów $y_i(t)$ w różnych systemach.

Z uwagi na zależność (13.28) wystarczy dokonać porównania widm mocy obwiedni zespolonych sygnałów. Stosunkowo prosto można wyprowadzić ogólny wzór określający widmo mocy obwiedni zespolonej sygnałów M-PSK. Można pokazać, że dla tych sygnałów:

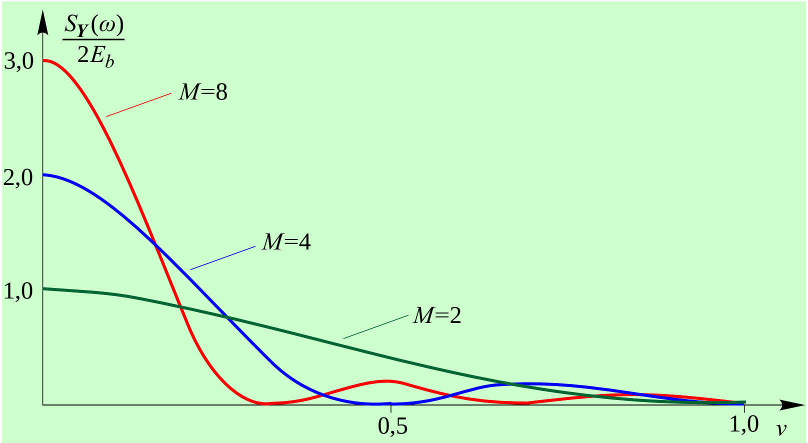
$$S_{\mathbf{Y}}(\omega) = (2E_b \log_2 M)\ \mathrm{Sa}^2\left[\frac{\omega T_b \log_2 M}{2}\right]. \tag{13.32}$$

Widzimy więc, że widma mocy obwiedni zespolonej sygnałów M-PSK mają identyczny kształt jak w przypadku sygnału 2PSK (por. wzór (13.29)), przy czym ulegają one zwężeniu w stosunku $\log_2 M$, a jednocześnie ich gęstości widmowe wzrastają ($\log_2 M$)-krotnie. Na rys. 7.68 pokazano widma mocy obwiedni zespolonej sygnałów M-PSK dla przypadków $M = 2$, 4 i 8. Widma te wykreślono w unormowanej względem wartości $2E_b$ skali rzędnych w funkcji unormowanej częstotliwości $(\omega/2\pi)T_b$. Szerokość pasma sygnałów M-PSK (wyrażona w hercach), zdefiniowana jako szerokość głównego listka widm (13.32), jest określona wzorem:

$$B_{\text{M-PSK}} = \frac{2}{T} = \frac{2}{T_b \log_2 M}. \tag{13.33}$$

gdzie $T = T_b \log_2 M$ jest przedziałem symbolowym. W porównaniu z systemem 2PSK szerokość pasma maleje zatem w stosunku $\log_2 M$.

Analiza widma mocy obwiedni zespolonej sygnałów M-FSK jest w ogólnym przypadku bardzo złożona. Względnie prosto szerokość pasma tych sygnałów można oszacować w przypadku, gdy częstotliwości nośne F_i sygnałów $y_i(t)$



Rys. 13.10. Unormowane widma mocy sygnałów M-PSK

przesyłanych w poszczególnych przedziałach symbolowych o długości T są określone wzorem $F_i = F_0 + i/2T$, $i = 1, \ldots, M$, gdzie F_0 jest pewną częstotliwością. Odstępy między kolejnymi częstotliwościami F_i są wówczas jednakowe i równe $1/2T$, a rozstaw częstotliwości unormowany względem czasu T wynosi $(F_{i+1} - F_i)T = 1/2$. W takim przypadku zasadnicza część mocy sygnału jest zawarta w paśmie o szerokości M-krotnie większej od odstępu między kolejnymi częstotliwościami:

$$B_{\text{M-FSK}} = \frac{M}{2T} = \frac{M}{2T_b \log_2 M}\,. \tag{13.34}$$

Zauważmy, że ze wzrostem M szerokość pasma wielowartościowych sygnałów PSK maleje, podczas gdy w przypadku wielowartościowych sygnałów FSK – rośnie.

Uwzględniając wzory (13.33) i (13.34) we wzorze definicyjnym (13.31), otrzymujemy wyrażenia określające efektywności widmowe systemów M-PSK i M-FSK:

$$\rho_{\text{M-PSK}} = \frac{R_b}{B_{\text{M-PSK}}} = \frac{\log_2 M}{2}, \tag{13.35}$$

$$\rho_{\text{M-FSK}} = \frac{R_b}{B_{\text{M-FSK}}} = \frac{2\log_2 M}{M}\,. \tag{13.36}$$

Wartości tych efektywności zestawiono w tablicy 13.1. Jak widzimy, efektywność widmowa systemów M-PSK rośnie ze wzrostem M, natomiast systemów M-FSK – maleje ze wzrostem M. Dla $M > 4$ systemy M-PSK są bardziej korzystne pod względem wykorzystania pasma transmisji od systemów M-FSK.

Tablica 13.1. Porównanie efektywności widmowej systemów M-PSK i M-FSK

M	2	4	8	16	32
ρ_{PSK} [bitów/s/Hz]	0,5	1	1,5	2	2,5
ρ_{PSK} [bitów/s/Hz]	1	1	0,75	0,5	0,3125

13.4.7. Porównanie 2-wartościowych i M-wartościowych cyfrowych systemów modulacji

Szerokość pasma i efektywność widmowa nie są jedynymi parametrami określającymi jakość danego systemu modulacji cyfrowej. Ważnym z praktycznego punktu widzenia parametrem jest średnia transmitowana moc zapewniająca uzyskanie wymaganej bitowej stopy błędów BER. Przy założonej wartości BER dąży się z reguły do osiągnięcia jak największej efektywności widmowej przy minimalnym poziomie średniej mocy, a dokładniej minimalnym stosunku sygnał-szum w odbiorniku. Jak wynika z tablicy 13.1, efektywność widmowa modulacji 2PSK jest dwukrotnie mniejsza niż modulacji 2FSK. W przypadku modulacji wielowartościowych, począwszy od $M = 8$, modulacje M-PSK są efektywniejsze widmowo w porównaniu z modulacjami M-FSK. Jednakże wymagany stosunek sygnał-szum zapewniający założoną wartość BER w systemach M-PSK rośnie, a w systemach M-FSK maleje ze wzrostem M. Pod tym względem modulacje M-FSK są korzystniejsze od modulacji M-PSK.

Jeżeli w systemie modulacji nie ma ograniczenia na szerokość B pasma transmisji i może być ono poszerzane, to korzystnie jest stosować modulacje M-FSK. Szerokość pasma sygnałów M-FSK rośnie bowiem ze wzrostem M, co umożliwia redukcję średniej transmitowanej mocy stałej wartości BER. Umożliwia to także zmniejszenie wartości BER przy zachowaniu tej samej mocy. Jeżeli natomiast szerokość pasma transmisji jest ograniczona, to wybór danego systemu modulacji jest dokonywany na drodze kompromisu między szerokością pasma, a wymaganym stosunkiem sygnał-szum (założoną wartością BER).

Słownik

bitowa stopa błędu BER
: stosunek bitów odebranych błędnie w cyfrowym systemie modulacji do wszystkich wysłanych bitów

bitowa szybkość transmisji
liczba bitów transmitowanych w cyfrowym systemie modulacji w ciągu 1 sekundy

dwubity
dwuelementowe ciągi znaków binarnych

efektywność widmowa
liczba bitów transmitowanych w ciągu 1 sekundy w cyfrowym systemie modulacji przypadająca na 1 Hz pasma zajętego w kanale przez sygnał

kanał AWGN
kanał transmisyjny, w którym na sygnał użyteczny oddziaływuje addytywnie szum biały gaussowski

konstelacja sygnałów
zbiór punktów przyporządkowanych sygnałom transmitowanym w systemie modulacji cyfrowej w wyniku odwzorowania geometrycznego przestrzeni sygnałów w przestrzeń wektorową

korelator
układ wyznaczający iloczyn skalarny dwóch sygnałów

metody geometryczne
metody reprezentacji i analizy sygnałów, w których sygnały są odwzorowywane w punkty zwykłej przestrzeni wektorowej

modulacja 2FSK
dwuwartościowa (binarna) modulacja z kluczowaniem częstotliwości

modulacja 2PSK
dwuwartościowa (binarna) modulacja z kluczowaniem fazy

modulacja QAM
modulacja cyfrowa, w której w zależności od transmitowanego symbolu w każdym przedziale symbolowym uzmienniane są skokowo amplituda i faza harmonicznej fali nośnej

modulacja QPSK
modulacja cyfrowa z czterowartościowym kluczowaniem fazy

modulacja z kluczowaniem amplitudy ASK
modulacja cyfrowa, w której w zależności od transmitowanego symbolu w każdym przedziale symbolowym uzmienniana jest skokowo amplituda harmonicznej fali nośnej

modulacja z kluczowaniem częstotliwości FSK
modulacja cyfrowa, w której w zależności od transmitowanego symbolu w

każdym przedziale symbolowym uzmienniana jest skokowo częstotliwość harmonicznej fali nośnej

modulacja z kluczowaniem fazy PSK
modulacja cyfrowa, w której w zależności od transmitowanego symbolu w każdym przedziale symbolowym uzmienniana jest skokowo faza harmonicznej fali nośnej

obszary decyzyjne
obszary, na które jest podzielona przestrzeń sygnałów w cyfrowym systemie modulacji

przedział bitowy
odcinek czasu, w którym w systemie modulacji cyfrowej transmitowany jest pojedynczy znak binarny (bit)

przedział symbolowy
odcinek czasu, w którym w systemie modulacji cyfrowej transmitowany jest pojedynczy symbol (ciąg znaków binarnych)

reguła decyzyjna
reguła, na podstawie której w odbiorniku jest podejmowana decyzja, który z sygnałów transmitowanych w systemie modulacji cyfrowej został faktycznie nadany w danym przedziale symbolowym

rozstaw częstotliwości
odległość między częstotliwościami nośnymi w cyfrowym systemie modulacji z kluczowaniem częstotliwości

szum biały gaussowski
sygnał losowy, którego gęstość widmowa jest stała w całym paśmie częstotliwości i którego wartości są zmiennymi losowymi o rozkładzie gaussowskim

Literatura

[1] Haykin S.: Systemy telekomunikacyjne. WKiŁ, Warszawa, 1999.
[2] Kryterium ML?
[3] Lahti B.P.: Systemy telekomunikacyjne. WNT, Warszawa, 1972.

Dodatki

Skorowidz

N

O

Z

Ź