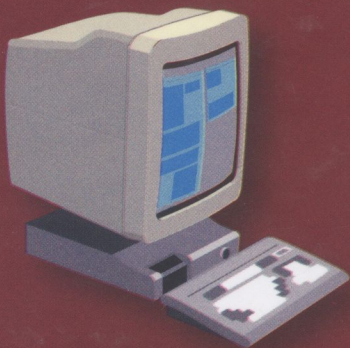


621.391

0 75



**Основи теорії
ІНФОРМАЦІЇ
та
КОДУВАННЯ**

Підручник

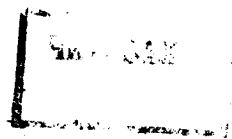
БДІ. 391(075)
0-75

ОСНОВИ ТЕОРІЇ ІНФОРМАЦІЇ ТА КОДУВАННЯ

За редакцією І.В. Кузьміна

Видання 3-тє перероблене та доповнене

Затверджено Міністерством освіти і науки України
як підручник для студентів ВНЗ, які навчаються
за напрямом підготовки "Радіоелектронні апарати",
"Телекомунікації", "Комп'ютерна інженерія"



Хмельницький 2009

УДК 621.39+004.415(075.8)

ББК 32.968

О 75

*Затверджено Міністерством освіти і науки України
як підручник для студентів вищих навчальних закладів,
лист № 14/18.2-680 від 30.03.2005*

Авторський колектив:

І. В. Кузьмін, І. В. Троцишин, А. І. Кузьмін,
В. О. Кедрус, В. Р. Любчик

Рецензенти:

- В. Д. Циделко* – д-р техн. наук, проф., зав. кафедрою
Національного технічного університету України “КПІ”;
В. М. Локазюк – д-р техн. наук, проф., зав. кафедрою
Хмельницького національного університету;
О. Д. Азаров – д-р техн. наук, проф., дир. ін-ту інформаційних
технологій та комп’ютерної інженерії
Вінницького національного технічного університету

454 129

О 75

Основи теорії інформації та кодування : підручник /
[І. В. Кузьмін, І. В. Троцишин, А. І. Кузьмін, В. О. Кедрус,
В. Р. Любчик] ; за ред. І. В. Кузьміна. – 3-тє вид., перероб.
та допов. – Хмельницький : ХНУ, 2009. – 373 с.
ISBN 978-966-330-063-4

Розглядаються загальні поняття і визначення інформації, інформаційних систем, детермінованих і випадкових сигналів, їх інформаційні характеристики, модуляція та перетворення, ефективність і завадостійкість кодування та передачі інформації, теорія статистичних рішень і інформаційна оцінка автоматизованих систем контролю і управління.

Третє видання доповнено розділами з модуляції і перетворення сигналів, стиснення та архівації інформації.

Для студентів ВНЗ напрямів підготовки “Радіоелектронні апарати”, “Телекомунікації” і “Комп’ютерна інженерія”. Може бути корисним спеціалістам, що досліджують інформаційні системи.

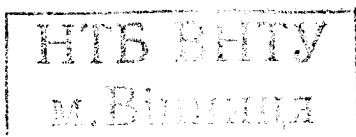
УДК 621.39+004.415(075.8)

ББК 32.968

© Авторський колектив, 2009

© ХНУ, оригінал-макет, 2009

ISBN 978-966-330-063-4



Широке впровадження автоматизованого управління неможливе без використання засобів зв'язку, телемеханіки та обчислювальної техніки.

У процесі автоматизованого управління технікою, виробничими процесами і галузями народного господарства, а також в процесах творчої діяльності людини і суспільних явищ відбувається інтенсивний обмін інформацією між окремими ланками систем управління, людиною і технікою, людиною і природою та окремими людьми.

Комплексна автоматизація та удосконалення електронних цифрових обчислювальних машин супроводжується різким зростанням обсягу і швидкості передачі та обробки інформації. Одночасно підвищуються вимоги до достовірності передачі та обробки інформації.

З усієї різноманітності сучасних технічних систем можна виділити в окрему групу так звані інформаційні системи, призначені для передачі, перетворення і збереження інформації. До цієї групи систем можна віднести системи зв'язку, телемеханічні, локаційні, навігаційні і телевізійні, обчислювальні та інформаційно-вимірювальні, автоматизовані системи управління і контролю тощо.

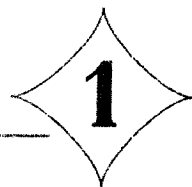
Інформаційні системи істотно відрізняються від енергетичних приладів і систем (двигунів, електричних генераторів, ліній електропередач, транспортних засобів, доменних печей тощо), основою функціонування яких є процеси передачі і перетворення енергії. Основою функціонування інформаційних систем є процеси передачі, перетворення і накопичення інформації. Тому критерієм якості роботи інформаційних пристроїв слугує їх здатність передавати, накопичувати або перетворювати необхідну кількість інформації в одиницю часу при допустимих викраденнях і затратах, а не коефіцієнт корисної дії, як в енергетичних установках. В інформаційних системах енергетичні співвідношення відіграють другорядну роль, а сама енергія є характеристикою сигналу, який використовується лише як транспортний засіб.

Однак при різкому зростанні потоку інформації в найближчому майбутньому виникне потреба у значних витратах енергії на передачу інформації. Це обумовить необхідність, окрім кількісної оцінки інформації, здійснювати енергетичну оцінку інформаційних процесів. Вже тепер одним з критеріїв оцінки систем космічного зв'язку є інформаційно-енергетичний критерій, що характеризує витрату енергії сигналу на передачу одиниці кількості інформації при даній потужності шумів, що припадає на 1 Гц смуги пропускання каналу зв'язку.

У посібнику основна увага приділяється теорії отримання, накопичення, передачі і перетворення інформації, аналізу і синтезу інформаційних процесів та інформаційних систем. Передбачається, що теорія логічних і обчислювальних машин, а також теорія автоматичного регулювання вивчаються в окремих самостійних курсах.

Посібник складається з тринадцяти розділів. У першому дається поняття про інформаційні системи, основою функціонування яких є процеси передачі і перетворення інформації; розкриваються роль і місце інформаційних систем в народному господарстві; викладаються основні визначення та поняття, що стосуються основ теорії проходження сигналів у цілому і теорії інформації частково. У другому та третьому розділах розглянуто сигнал як носій інформації. Наводяться характеристики основних типів детермінованих і випадкових сигналів, способи їх математичного опису і перетворення. У четвертому розділі викладені основні положення теорії квантування сигналів за часом і рівнем. У п'ятому розділі розглянуто основні положення модуляції та демодуляції, у шостому – основи теорії ефективного і завадостійкого кодування. У сьомому і восьмому розділах викладені інформаційні моделі сигналів і процесів передачі інформації, дані поняття ентропії як міри невизначеності та кількості інформації, наведені способи кількісної оцінки ентропії та інформації, способи оцінки швидкості передачі інформації та пропускну здатності інформаційного каналу, способи узгодження сигналів з каналом. У дев'ятому розділі викладені основи оптимального прийому і обробки інформації, різні методи фільтрації і накопичення інформації. Десятий розділ присвячений основам оцінки ефективності інформаційних систем і методам її підвищення. В одинадцятому розділі розглядаються завадостійкість інформаційних систем і способи її підвищення. У дванадцятому розділі викладаються методи інформаційної оцінки ефективності і якості роботи автоматизованих систем контролю і управління, практичного використання апарату теорії інформації для розв'язання окремих задач контролю і управління. У тринадцятому розділі описані алгоритми архізації зображень. Основним математичним апаратом, що використовується в книзі, є апарат теорії інформації, теорії ймовірностей, а також теорії випадкових процесів.

ІНФОРМАЦІЯ ТА ІНФОРМАЦІЙНІ СИСТЕМИ



1.1 Основні поняття і визначення

Виробничі процеси, а також процеси живої природи пов'язані з отриманням, передачею, перетворенням, накопиченням, збереженням і відображенням інформації.

На сьогодні існують різні визначення інформації. Зазвичай під інформацією в широкому сенсі розуміють нові відомості про навколишній світ, які ми отримуємо в результаті взаємодії з ним, пристосування до нього і зміни його в процесі пристосування.

Рекомендується наступне визначення інформації: *інформація* – це відомості, що є об'єктом зберігання, передачі і перетворення.

Інформація – це насамперед відомості, які повинні бути використані (відомості про стан природи, стан і положення в часі та просторі окремих об'єктів, величину контрольованих параметрів тощо).

У філософії розгорнулася боротьба між матеріалістами та ідеалістами навколо розуміння природи інформації. Ідеалісти намагаються відірвати інформацію від матерії, перетворити її в деяку духовну субстанцію або ж представити у вигляді комплексу відчуттів суб'єкта. Інформація є властивістю матерії. Інформація не є матерією, і не є енергією, але нерозривно пов'язана з першою.

Деякі ідеалісти намагаються стверджувати, що саме поняття інформації дає основу для примирення ідеалізму і матеріалізму. Інформація розглядається ними як "нейтральний елемент", що не є ні матерією, ні свідомістю, на основі чого робиться висновок, що кібернетика надає наукові дані для подолання "обмеженості" як матеріалістичних, так і ідеалістичних концепцій.

Поняття інформації є науковим підтвердженням гіпотези, що вся матерія має властивість відображення. При цьому треба мати на увазі, що поняття "відображення" як властивість всієї матерії більш загальне, ніж поняття "інформація", що є поняттям кібернетики і зустрічається тільки в організованих системах – машинах, живих організаціях і суспільстві.

Потрібно розрізняти поняття "інформація" і "повідомлення". Повідомлення – це форма представлення інформації. Наприклад, при телеграфній передачі повідомленням є текст телеграми, що являє собою

послідовність різних символів. При розмові повідомлення являє собою механічні коливання із різною частотою та інтенсивністю голосових зв'язок людини. При телевізійних чорно-білих передачах повідомлення – це зміна у часі яскравості елементів зображення, що передається. При кольорових телевізійних передачах повідомлення є також зміною і кольору елементів зображення тощо.

Структурна схема інформаційної системи (рис. 1.1) найбільш повно відображає систему зв'язку, але може бути використана для характеристики й інших інформаційних систем (автоматизованого управління, контролю, біологічних, суспільних тощо). Система складається із відправника інформації, лінії зв'язку і приймача інформації.

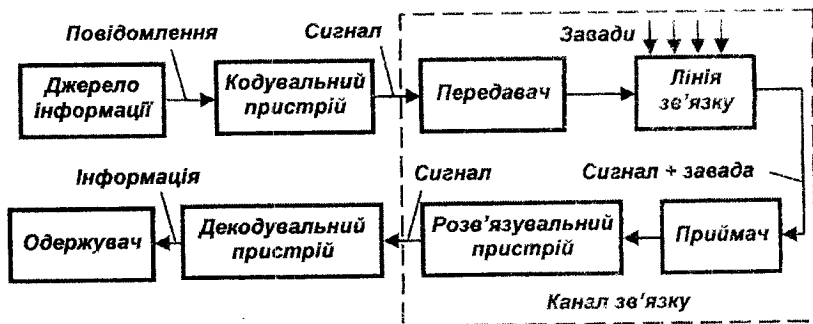


Рис. 1.1

Повідомлення для передачі його за відповідною адресою повинно бути попередньо перетвореним у сигнал. Під сигналом розуміється змінна у часі фізична величина, що відображає повідомлення. *Сигнал* – це матеріальний носій повідомлення. Фізичне середовище, по якому відбувається передача сигналів від передавача до приймача, називається *лінією зв'язку*.

У сучасній техніці знайшли застосування електричні, електромагнітні, світлові, механічні, звукові, ультразвукові сигнали. Для передачі повідомлень необхідно застосувати той носій, який здатен ефективно розповсюджуватися лінією зв'язку системи. Наприклад, по провідній лінії зв'язку найбільш легко проходять постійний струм і змінні струми порівняно невисоких частот (практично не більше декількох десятків кілогерц). По радіолінії ефективно розповсюджуються лише електромагнітні коливання високих частот (від сотень кілогерц до десятків тисяч мегагерц).

Всі повідомлення за характером зміни у часі можна поділити на неперервні та дискретні. Неперервні у часі повідомлення відобра-

жуються неперервною функцією часу. Дискретні у часі повідомлення характеризуються тим, що надходять у певні моменти часу і описуються дискретною функцією часу. Оскільки повідомлення носять випадковий характер, то неперервні повідомлення описуються випадковою функцією часу, а дискретні – як послідовність випадкових подій.

Повідомлення можна також поділити на неперервні і дискретні за множиною. Неперервні за множиною повідомлення характеризуються тим, що функція, яка їх описує, може приймати неперервну множину значень (континуум значень) у певному інтервалі.

Дискретні за множиною повідомлення – це повідомлення, які можуть бути описані за допомогою скінченного набору чисел або дискретних значень деякої функції.

Дискретності за множиною і часом не пов'язані між собою. Тому можливі наступні типи повідомлень:

- а) неперервні за множиною і часом або просто неперервні (рис. 1.2);
- б) неперервні за множиною і дискретні за часом (рис. 1.3);
- в) дискретні за множиною і неперервні за часом (рис. 1.4);
- г) дискретні за множиною і часом або просто дискретні (рис. 1.5).

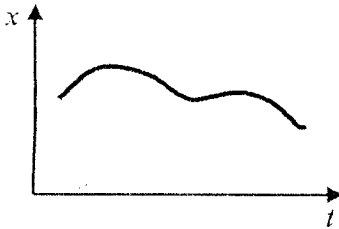


Рис. 1.2



Рис. 1.3

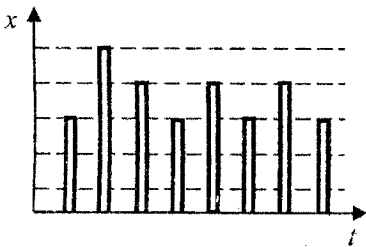


Рис. 1.4

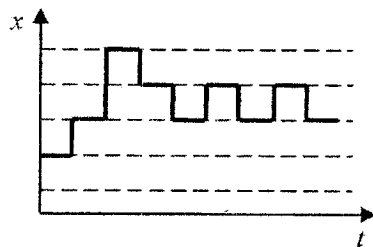


Рис. 1.5

У широкому сенсі перетворення повідомлень у сигнал називають кодуванням повідомлень. У вузькому – **кодування** – це відображення дискретних повідомлень сигналами у вигляді певних послідовних символів. Пристрій, який здійснює кодування, називається кодером (кодувальним пристроєм).

За допомогою кодувального пристрою повідомлення, яке може мати будь-яку фізичну природу (наприклад, звукове коливання, зображення), перетворюється у первинний сигнал, для технічних пристроїв – зазвичай у електричний сигнал. При телеграфній передачі таким сигналом слугує послідовність електричних імпульсів постійного струму і різної тривалості. У телефонії – це електричні коливання звукової частоти.

Для перетворення первинного сигналу у вигляд, придатний для використання у лінії зв'язку, існує передавач.

У передавальному пристрої здійснюється вплив на один чи декілька параметрів носія за законом, прийнятим при кодуванні повідомлень. Цей процес називається модуляцією, а параметри, що модулюються, називаються інформативними. Іноколи у значенні “модуляція” вживається термін “маніпуляція”.

Неперервні і дискретні повідомлення не завжди передаються з допомогою неперервних і дискретних сигналів. У деяких випадках неперервні повідомлення перетворюються у дискретний сигнал, а дискретні повідомлення – у неперервний сигнал.

При передачі сигнали піддаються впливу завад. Під завадами розуміються будь-які зовнішні збурення чи впливи (атмосферні завади, впливи сторонніх джерел сигналів), а також викривлення сигналів у самій апаратурі (апаратурні завади), що викликають випадкове відхилення прийнятого повідомлення сигналу від переданого.

На приймальній стороні здійснюється відновлення переданого повідомлення за прийнятим сигналом. При цьому спочатку здійснюється демодуляція сигналу, в результаті якої відновлюється первинний сигнал.

Внаслідок різних викривлень і впливів завад прийнятий сигнал може суттєво відрізнитися від переданого. Тому проводиться обробка прийнятого сигналу з метою найбільш повного вилучення з нього інформації. Ці функції виконуються розв'язувальним пристроєм. Процедура обробки сигналу є досить різноманітними. Вони можуть зводитись до фільтрації сигналів, обмеження, інтегрування (або диференціювання), перемноження сигналів, складання тощо.

Після обробки здійснюється декодування сигналу, тобто перетворення його у повідомлення, що потрапляє до одержувача. При цьому сигнал повинен бути приведений до вигляду, зручного для сприйняття одержувачем. Наприклад, при передачі мови або музики прийнятий сигнал перетворюється у звукові механічні коливання, якщо одержу-

вачем є людина. Якщо ж прийняті сигнали призначено для магнітофонного запису, то вони перетворюються в електричні сигнали, зручні для фіксації на магнітній плівці.

Операції демодуляції, обробки і декодування або дві з них, можуть виконуватись одним пристроєм. Наприклад, при прийомі телеграфних сигналів на слух оператор здійснює як обробку, так і декодування прийнятих сигналів.

Аналогічно, на передавальній стороні в одному пристрої можуть суміщатися функції кодування і модуляції.

Часто на практиці необхідно забезпечити в інформаційних системах незалежну передачу повідомлень від декількох джерел. Використання для кожного джерела повідомлень окремої лінії зв'язку є економічно не вигідним. Наприклад, для сучасних технічних систем передачі інформації лінії зв'язку є найбільш дорогими ланками. Тому виникає задача побудови систем, що використовують одну лінію зв'язку для передачі повідомлень від декількох джерел. Такі системи називаються багатоканальними.

Очевидно, у багатоканальних системах сигнали, що надходять від різних джерел повідомлень, повинні володіти певними ознаками належності до власного джерела – канальними ознаками. Це дозволить розрізнити їх на приймальній стороні. Для цього у таких системах, окрім вказаних раніше ланок, на передавальній стороні необхідно мати формувач канальних ознак – апаратуру ущільнення, а на приймальній стороні – роздільник сигналів – апаратуру розділення.

Сукупність пристроїв, що забезпечують незалежну передачу сигналів різних джерел повідомлень по одній лінії зв'язку, називається **каналом зв'язку**.

1.2 Критерії оцінки ефективності якості

При синтезі та аналізі інформаційних систем необхідно оперувати певними якісними і кількісними показниками, що дозволяють здійснювати оцінку систем, а також порівнювати їх з іншими, подібними їм системами.

Всі задачі, які стоять перед інформаційними системами, зводяться до двох основних проблем: забезпечення високої ефективності та надійності системи.

Під **ефективністю інформаційної системи** розуміється її здатність забезпечити передачу заданої кількості інформації найбільш економічним способом. Очевидно, що ефективність системи перш за все визначається максимально можливою швидкістю передачі інформації.

Швидкість передачі інформації визначається кількістю інформації, яка передається за одиницю часу. Максимально можливу швидкість передачі інформації прийнято називати *пропускнуою здатністю* системи. Пропускна здатність характеризує потенційні можливості системи.

Оскільки ефективність характеризує економічність системи, то найбільш повна її кількісна оцінка буде характеризуватися співвідношенням переданої інформації до витрат, пов'язаних із передачею даної кількості інформації. Для передачі інформації за допомогою сигналу необхідні певні витрати потужності. Крім цього, для передачі та обробки сигналів потрібен певний час. Зрештою, кожний сигнал займає в каналі певний діапазон частот. Тому ефективність характеризує здатність системи забезпечити передачу даної кількості інформації з найменшими витратами потужності сигналів, часу і смуги частот.

Слід зазначити, що відомі й інші критерії оцінки ефективності систем. Наведене раніше визначення ефективності відноситься до чисто інформаційних систем.

Під надійністю технічної системи розуміється її здатність до безвідмовної роботи протягом заданого проміжку часу і в заданих умовах експлуатації. Тому під відмовою розуміється подія, що призводить до неможливості використання хоча б однієї із робочих властивостей системи. Кількісно надійність оцінюється різними показниками. Одним із таких показників є ймовірність безвідмовної роботи системи протягом заданого проміжку часу.

Однак для інформаційних систем дуже важливою характеристикою є *інформаційна надійність*, під якою будемо розуміти здатність системи передавати інформацію з досить високою достовірністю. Достовірність кількісно характеризується ступенем відповідності прийнятого повідомлення переданому. Цю величину прийнято називати ймовірністю.

Зниження достовірності передачі повідомлень називається їх викривленням, що здійснюється під впливом зовнішніх і внутрішніх (апаратурних) завод. Здатність інформаційної системи протистояти шкідливій дії завод називається *завадостійкістю*.

У подальшому будемо оцінювати системи не лише в інформаційному аспекті, але й за ефективністю (в інформаційному значенні) і завадостійкістю.

Теорія інформації і кодування встановлює критерії оцінки завадостійкості та ефективності інформаційних систем, а також вказує загальні шляхи підвищення завадостійкості та ефективності. Умови, за яких може бути досягнуто підвищення завадостійкості та ефективності, є суперечливими. Підвищення завадостійкості практично завжди супроводжується погіршенням ефективності і, навпаки, покращення ефективності негативно зазначається на завадостійкості систем.

1.3 Предмет і метод теорії інформації

Теорія інформації – розділ кібернетики, що займається математичним описом і оцінкою методів передачі, зберігання, вилучення і класифікації інформації. Курс теорії інформації являє собою єдину наукову дисципліну, основу якої складають математичні моделі сигналів, кількісна оцінка інформації, теорія оцінки ефективності і завадостійкості інформаційних систем.

Без застосування теорії інформації стає практично неможливим створення надскладних систем управління космічними кораблями і ракетами, систем зв'язку і телебачення з використанням штучних супутників Землі, складних обчислювальних і керуючих систем тощо.

Метод теорії інформації є сукупністю прийомів дослідження інформаційних систем (наприклад, методи оцінки інформаційної спроможності джерела інформації, пропускну здатності систем передачі інформації, інформаційної ємності пристроїв, ємності запам'ятовуючих пристроїв). При цьому в основі методу лежить матеріалістична діалектика, об'єктивні закони природи і суспільства, об'єктивні закони дійсності.

Зовнішній світ, що оточує людину, впливає на неї через органи чуття з підсилювачами або без підсилювачів відчуття. Органи чуття дають людині інформацію про оточуючу дійсність. Інформація перетворюється нервовою системою і мозком, на допомогу якому може підключатися ЕОМ або найпростіші вимірювальні пристрої, а потім, після відповідного відбору, переробки і накопичення використовується людиною для відповідних впливів у відповідь (рис. 1.6).

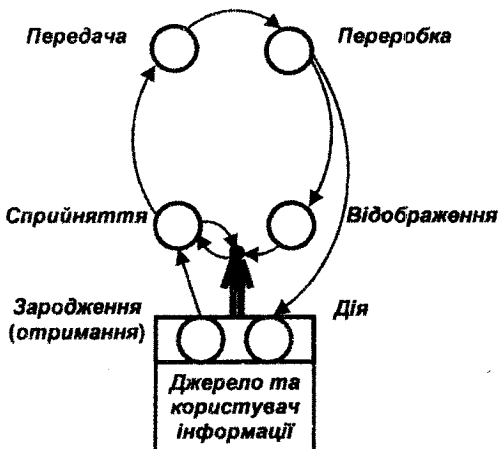


Рис. 1.6

У ролі підсилювачів відчуття можуть виступати прості і складні вимірювальні системи (електронний мікроскоп, радіотелескоп, космічні та астрономічні системи тощо). Можливості рухової системи людини (виконавчих органів) можуть підсилити знаряддя праці (знаряддя виробництва).

Отримання інформації відбувається в процесі часткового або повного знімання невизначеностей стану чи положення об'єкта у просторі і часі, в процесі пізнання стану і положення об'єкта.

Контрольні питання

1. Що розуміється під інформаційними системами?
2. Що таке інформація, повідомлення, сигнал?
3. Який зв'язок існує між поняттями "інформація", "повідомлення", "зв'язок"?
4. У чому полягає суть перетворень повідомлень у сигнал?
5. Що таке лінія зв'язку, канал зв'язку?
6. Що розуміється під багатоканальними системами?
7. У чому полягає предмет і метод теорії інформації?
8. Що розуміють під ефективністю та завадостійкістю інформаційних систем?
9. Назвіть та охарактеризуйте основні фази перетворення інформації, а також системи обміну інформації?

МАТЕМАТИЧНІ МОДЕЛІ ДЕТЕРМІНОВАНИХ СИГНАЛІВ

2

2.1 Загальна характеристика сигналів

Сигнал – змінна фізична величина, що забезпечує передачу інформації по лінії зв'язку. У технічних інформаційних системах використовуються в більшості випадків електричні сигнали.

Вся різноманітність сигналів, що використовуються в інформаційних системах, за своїми особливостями може бути поділена на дві основні групи: детерміновані і випадкові. Детерміновані сигнали характеризуються тим, що в будь-які моменти часу значення є відомими величинами. Сигнал, значення якого в будь-які моменти часу будуть випадковими величинами, називається випадковим.

Розподіл сигналів на детерміновані і випадкові є умовним, тому що детермінованих сигналів у точному їхньому розумінні в природі немає. На практиці не може бути заздалегідь точно передбачене значення сигналу в будь-які моменти часу, в іншому випадку сигнал не ніс би корисної інформації. Крім того, будь-який реальний сигнал випадковий у силу впливу на нього численних випадкових чинників.

Незважаючи на це, дослідження детермінованих сигналів є дуже важливим з двох причин:

1) математичний апарат, який використовується для аналізу детермінованих сигналів, набагато простіше апарату аналізу випадкових сигналів;

2) висновки, отримані у результаті досліджень детермінованих сигналів, можуть бути в багатьох випадках використані для аналізу випадкових сигналів.

Оскільки в інформаційних системах застосовуються переважно електричні засоби передачі повідомлень (провідні і радіозасоби), то обмежимося розглядом тільки електричних сигналів.

Електричні сигнали є носіями інформації лише при певному впливі (модуляції) на один або декілька їхніх параметрів. Такі параметри прийнято називати інформативними або кодовими.

Сигнали, як і повідомлення, за своєю структурою, поділяються на:

- безперервні за рівнем (множиною) і часом;
- дискретні за рівнем і часом;
- дискретні за рівнем і безперервні за часом;
- безперервні за рівнем і дискретні за часом.

Сигнали, безперервні за рівнем і часом, називають безперервними. Сигнали, дискретні за рівнем і часом, називаються дискретними. Сигнал, що є дискретним тільки за часом або тільки за рівнем, прийнято називати дискретно-безперервним. На практиці з цієї групи сигналів переважне застосування знайшли сигнали, дискретні тільки за часом. У зв'язку з цим надалі під дискретно-безперервними сигналами будемо розуміти сигнали, дискретні за часом і безперервні за рівнем.

Прикладами безперервних сигналів є постійний струм і напруга або гармонічні коливання струму (напруги). У першому випадку інформативним параметром сигналу може бути тільки величина струму або напруги. У другому випадку – частота або фаза коливань.

Прикладом дискретно-безперервних сигналів є послідовність прямокутних імпульсів струму або напруги. Тут як інформативний параметр може бути використана амплітуда, тривалість, частота проходження і фаза імпульсів (тимчасове зрушення імпульсів у кожному періоді посилок відносно опорного моменту часу).

Дискретні сигнали можуть представлятися у вигляді окремих найпростіших елементів (посилок) або сукупності таких елементів.

Для аналізу і синтезу інформаційних систем необхідно знати не тільки характеристики цих систем у вигляді операторів, але і математичні моделі сигналів.

Залежно від методів аналізу інформаційних систем відповідно до форми представлення операторів застосовуються ті або інші способи представлення сигналів. До основних варто віднести представлення сигналу:

- у вигляді деякої функції часу $x(t)$;
- в операторній формі $x(p)$;
- у вигляді деякої функції частоти $x(\omega)$.

При аналізі процесу проходження довільного детермінованого сигналу через лінійну систему або систему, близьку до лінійної, зручно представляти його у вигляді належним чином обраної сукупності елементарних сигналів. Знаючи реакцію системи на елементарний сигнал, можна, користуючись методом суперпозиції, визначити реакцію системи на сигнал довільної форми.

До елементарних детермінованих сигналів відносять одиничну функцію, ідеальний одиничний імпульс і синусоїдальний вплив.

Одиничний імпульс та одинична функція використовуються при тимчасовому представленні сигналів.

Властивості одиничної функції визначаються співвідношеннями:

$$1(t - \tau) = \begin{cases} 0, & \text{при } t < \tau; \\ 1, & \text{при } t \geq \tau, \end{cases} \quad (2.1)$$

де $1(t - \tau)$ – одинична функція; t – час, τ – момент початку дії одиничної функції.

Таким чином, одинична функція – це тимчасова функція, що при будь-якому $t < \tau$ тотожно дорівнює нулю, а при будь-якому $t \geq \tau$ дорівнює одиниці. Одинична функція відбиває стрибкоподібну зміну в момент часу $t = \tau$ величини від нуля до одиниці (рис. 2.1).

Властивості одиничного імпульсу (δ -функції) визначаються співвідношеннями:

$$\delta(t - \tau) = \begin{cases} 0, & \text{при } t \neq \tau; \\ \infty, & \text{при } t = \tau; \end{cases} \quad (2.2)$$

$$\int_0^t \delta(t - \tau) dt = 1(t - \tau), \quad 0 < \tau < t,$$

де $\delta(t - \tau)$ – дельта-функція; t – час, τ – момент дії імпульсу.

Отже, одиничний імпульс – це ідеалізований сигнал, що характеризується нескінченно малою тривалістю, нескінченно великим рівнем і площею, рівною одиниці (рис. 2.2).

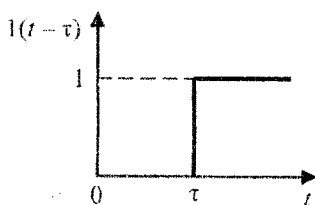


Рис. 2.1

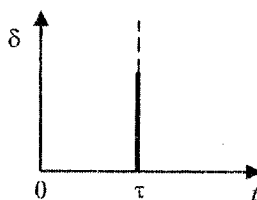


Рис. 2.2

При аналізі багатьох типів технічних інформаційних систем, особливо при дослідженні пропускнув спроможності і селективності систем, широко використовуються синусоїдальні елементарні сигнали. У зв'язку з цим великий інтерес становить частотне представлення сигналу, тобто представлення моделі сигналу у вигляді сукупності елементарних синусоїдальних сигналів.

2.2 Частотне представлення детермінованих сигналів

У частотному вигляді можуть представлятися як періодичні, так і неперіодичні детерміновані сигнали. Необхідно зазначити, що в реальних умовах періодичні сигнали не існують, тому що ідеальний

періодичний сигнал нескінченний у часі, у той час як усякий реальний сигнал має початок і кінець. Однак у багатьох випадках скінченністю часу дії сигналу можна знехтувати і для аналізу припустимо використовувати апарат, придатний для ідеальних періодичних сигналів.

2.2.1 Періодичні сигнали

Розглянемо сигнал (рис. 2.3), що виражається довільною періодичною функцією $x(t)$.

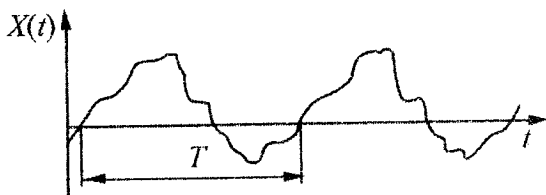


Рис. 2.3

Відомо, що будь-яка періодична функція, що задовольняє умовам Діріхле¹, може бути подана у вигляді нескінченної в загальному випадку суми гармонічних складових – рядом Фур'є.

Відомі дві форми розкладання в ряд Фур'є: тригонометрична і комплексна. Тригонометрична форма розкладання має вигляд:

$$x(t) = \frac{1}{2} A_0 + \sum_{k=-\infty}^{\infty} A_k \cos(K\omega_0 t - \varphi_k), \quad (2.3)$$

де $\frac{1}{2} A_0$ – постійна складова функції $x(t)$; $A_k \cos(K\omega_0 t - \varphi_k)$ – k -та гармонічна складова; A_k , $K\omega_0$, φ_k – амплітуда, частота і початкова фаза k -ї гармонічної складової $\omega_0 = \frac{2\pi}{T}$ – частота основної (першої) гармоніки; T – період зміни функції $x(t)$.

У математичному відношенні зручніше оперувати комплексною формою ряду Фур'є, що представляється у вигляді:

$$x(t) = \frac{1}{2} \sum_{k=-\infty}^{\infty} A_k \exp\{jK\omega_0 t\}, \quad (2.4)$$

¹Умова Діріхле полягає у наступному: функція повинна бути обмеженою, кусково-неперервною і мати протягом періоду скінченну кількість екстремальних значень.

де $A_k = A_k \exp \{-j \varphi_k\}$ – комплексна амплітуда гармонічної складової частоти $\omega_k = K \omega_0$.

Суми, обумовлені виразами (2.3) і (2.4), будуть тотожними при виконанні умов:

$$A_{-k} = A_k; \quad \varphi_{-k} = \varphi_k; \quad \varphi_0 = 0. \quad (2.5)$$

При цьому модуль комплексної амплітуди дорівнює амплітуді відповідної гармонічної складової, а аргумент дорівнює початковій фазі складової. Комплексна амплітуда визначається через тимчасову функцію $x(t)$ за формулою:

$$A_k = \frac{2}{T} \int_{t_1}^{t_1+T} x(t) \exp\{-jK\omega_0 t\} dt. \quad (2.6)$$

Сукупність амплітуд і відповідних частот гармонік прийнято називати спектром амплітуд. Сукупність початкових фаз і відповідних частот гармонік називається спектром фаз.

Спектр амплітуд і спектр фаз однозначно визначають сигнал. Однак для багатьох практичних задач досить обмежитися розглядом тільки спектра амплітуд.

На рис. 2.4 наведені зображення спектри амплітуд і фаз періодичного сигналу. Окремі спектральні складові в графічному зображенні спектра амплітуд називаються спектральними лініями.

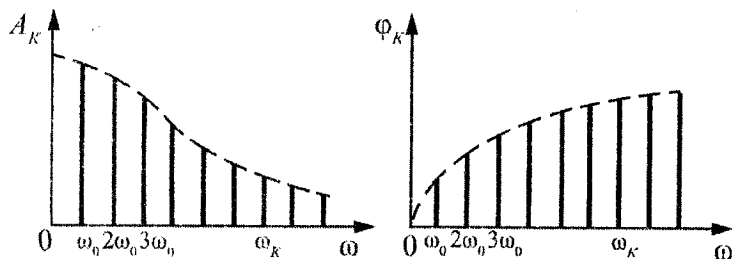


Рис. 2.4

Характерною рисою спектра періодичного сигналу є його переривистість (дискретність). Відстань між сусідніми спектральними лініями однакова і дорівнює частоті основної гармоніки.

Розглянемо як приклад послідовність прямокутних імпульсів тривалістю τ , амплітудою h , із періодом проходження T (рис. 2.5).

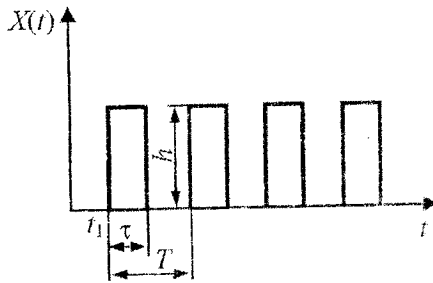


Рис. 2.5

Функція $x(t)$, що описує такий сигнал, може бути представлена наступним чином:

$$x(t) = \begin{cases} h, & \text{при } t_1 + i \leq t \leq t_1 + \tau + iT; \\ 0, & \text{при } t_1 + \tau + iT < t < t_1 + (i+1)T; \quad i=0, 1, 2, \dots, \infty. \end{cases} \quad (2.7)$$

Функція $x(t)$ може бути полана рядом Фур'є:

$$\begin{aligned} x(t) &= \frac{1}{2} \sum_{k=-\infty}^{\infty} A_k \exp\{jK\omega_0 t\} = \\ &= \frac{1}{2} A_0 + \sum_{k=1}^{\infty} [A_k \exp\{jK\omega_0 t\} + A_{-k} \exp\{-jK\omega_0 t\}], \end{aligned} \quad (2.8)$$

де $\omega_0 = \frac{2\pi}{T}$; A_k, A_{-k} – комплексні амплітуди k -ї гармоніки;

$\frac{1}{2} A_0$ – постійна складова сигналу:

$$A_k = A_k \exp\{j\varphi_k\}; \quad A_{-k} = A_k \exp\{-j\varphi_k\}.$$

Відповідно до (2.6) комплексна амплітуда:

$$\begin{aligned} A_k &= \frac{2}{T} \int_{t_1}^{t_1+T} x(t) \exp\{-jK\omega_0 t\} dt = \frac{2}{T} \int_{-\tau/2}^{\tau/2} h \exp\{-jK\omega_0 t\} dt = \\ &= \frac{2h}{T} \frac{\exp\{-jK\omega_0 t\}}{-jK\omega_0} \Bigg|_{-\tau/2}^{\tau/2} = 2 \frac{\tau h}{T} \frac{\sin \frac{K\omega_0 t}{2}}{\frac{K\omega_0 t}{2}}. \end{aligned} \quad (2.9)$$

Постійна складова сигналу може бути отримана з виразу (2.9) при $k = 0$:

$$\frac{1}{2} A_0 = \frac{1}{T} \frac{\tau}{2} \int_{-\tau/2}^{\tau/2} h dt = \frac{\tau h}{T}.$$

Таким чином, розклад у ряд Фур'є періодичної послідовності прямокутних імпульсів представляється у вигляді:

$$x(t) = \frac{\tau h}{T} \left[1 + 2 \sum_{k=1}^{\infty} \frac{\sin \frac{K \omega_0 t}{2}}{\frac{K \omega_0 t}{2}} \cos K \omega_0 t \right]. \quad (2.10)$$

Як видно з виразу (2.10), при черговому прирості частоти на величину $2\pi/\tau$ – фаза гармонік змінюється на величину π .

Спектр амплітуд показаний на рис. 2.6, причому обвідна його визначається рівнянням:

$$A(\omega) = 2 \frac{\tau h}{T} \left[\frac{\sin \frac{\omega t}{2}}{\frac{\omega t}{2}} \right], \quad (2.11)$$

де $\omega = K \omega_0$ – частота для K -ї гармоніки.

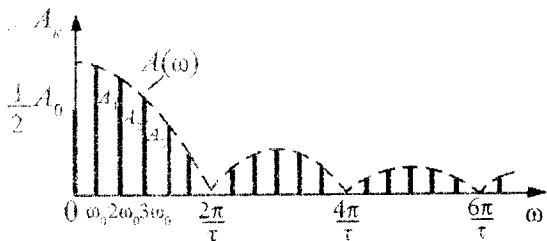


Рис. 2.6

Форма обвідної спектра амплітуд визначається видом функції

$$\left[\frac{\sin \frac{\omega t}{2}}{\frac{\omega t}{2}} \right], \quad \text{причому: } \lim_{\omega \rightarrow 0} \left[\frac{\sin \frac{\omega t}{2}}{\frac{\omega t}{2}} \right] = 1;$$

$$\left| \frac{\sin \frac{\omega \tau}{2}}{\frac{\omega \tau}{2}} \right| = 0 \text{ при } \omega = \frac{n\pi}{\tau},$$

де n – парне число.

2.2.2 Неперіодичні сигнали

Будь-який неперіодичний сигнал (рис. 2.7) можна розглядати як періодичний, період зміни якого дорівнює нескінченності. У зв'язку з цим розглянутий раніше спектральний аналіз періодичних процесів може бути узагальнений і на неперіодичний сигнал.

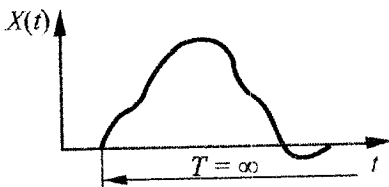


Рис. 2.7

Розглянемо, як змінюється спектр періодичного сигналу при необмеженому збільшенні періоду зміни сигналу. При збільшенні періоду T інтервали між суміжними частотами в спектрі сигналу і амплітуди спектральних складових зменшуються і в межах при $T \rightarrow \infty$ стають нескінченно малими величинами.

При цьому ряд Фур'є, що відображає спектральне розкладання періодичного сигналу, перетворюється в інтеграл Фур'є, що відображає спектральне розкладання неперіодичного сигналу.

Комплексна форма інтеграла Фур'є має вигляд:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega) \exp(j\omega t) d\omega, \quad (2.12)$$

де $S(j\omega) = S(\omega) \exp j\varphi(\omega)$ – спектральна щільність сигналу;
 $S(\omega) = |S(j\omega)|$ – амплітудно-частотна характеристика сигналу; $\varphi(\omega)$ – фазочастотна характеристика сигналу.

Вираз (2.12) називають формулою зворотного перетворення Фур'є.

Представлення неперіодичної функції інтегралом Фур'є можливо при дотриманні умов, що функція $x(t)$:

- задовольняє умовам Діріхле;
- абсолютно інтегровальна¹, тобто:

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty.$$

¹Цій умові задовольняє будь-який сигнал.

Таким чином, спектр неперіодичного сигналу на відміну від спектра періодичного сигналу є неперервним і являє собою суму нескінченного числа гармонічних складових, з нескінченно малими амплітудами.

Амплітуди гармонічних складових, виходячи з (2.12), можна представити у вигляді:

$$dA = \frac{1}{2\pi} S(j\omega) d\omega,$$

звідки спектральна щільність визначається виразом:

$$S(j\omega) = 2\pi \frac{dA}{d\omega}.$$

Спектральна щільність пов'язана з часовою функцією сигналу через пряме перетворення Фур'є:

$$S(j\omega) = \int_{-\infty}^{\infty} x(t) \exp\{-j\omega t\} dt. \quad (2.13)$$

Спектральна щільність однозначно відображає неперіодичний сигнал і задовольняє умовам (рис. 2.8):

$$-- \lim_{\omega \rightarrow \infty} S(\omega) = 0;$$

-- модуль спектральної щільності є парною, а аргумент – непарною функцією частоти, тобто:

$$S(\omega) = S(-\omega); \quad \varphi(\omega) = -\varphi(-\omega).$$

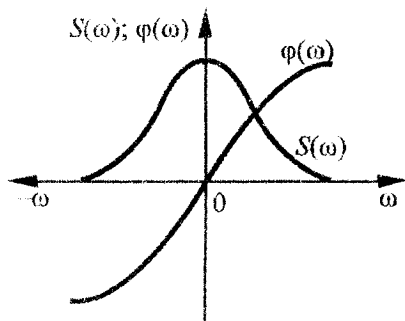


Рис. 2.8

2.2.3 Енергетичне тлумачення спектра сигналу

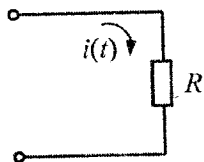


Рис. 2.9

Розглянемо розподіл потужності в спектрі періодичного сигналу. Для цього припустимо, що сигнал являє собою струм $i(t)$, що протікає по резистору R (рис. 2.9) і описується складною періодичною функцією часу з періодом зміни T .

Середня потужність, що виділяється на резисторі R :

$$P_{\text{сеп}} = \frac{R}{T} \int_0^T i^2(t) dt = RI^2,$$

де $I^2 = \frac{1}{T} \int_0^T i^2(t) dt$ – квадрат діючого значення струму.

Якщо представити струм $i(t)$ рядом Фур'є, отримаємо наступний вираз для квадрата діючого значення струму:

$$\begin{aligned} I^2 &= \frac{1}{T} \int_0^T \left[\frac{1}{2} A_0 + \sum_{k=1}^{\infty} A_k \cos(K\omega_0 t - \varphi_k) \right]^2 dt = \frac{1}{4T} \int_0^T A_0^2 dt + \\ &+ \frac{A_0}{T} \sum_{k=1}^{\infty} \int_0^T A_k \cos(K\omega_0 t - \varphi_k) dt + \frac{1}{T} \sum_{i=1}^{\infty} \sum_{l=1, l \neq i}^{\infty} \int_0^T A_i A_l \cos(i\omega_0 t - \varphi_i) \times \\ &\times \cos(l\omega_0 t - \varphi_l) dt + \frac{1}{T} \sum_{k=1}^{\infty} A_k^2 \cos^2(K\omega_0 t - \varphi_k) dt = \frac{A_0^2}{4} + \frac{1}{2} \sum_{k=1}^{\infty} A_k^2. \end{aligned}$$

Таким чином, середня потужність:

$$P_{\text{сеп}} = R \left[\left(\frac{A_0}{2} \right)^2 + \frac{1}{2} \sum_{k=1}^{\infty} A_k^2 \right] \quad (2.14)$$

Отже, середня потужність, що виділяється складним періодичним струмом на резисторі, дорівнює сумі середніх потужностей, що виділяються на цьому резисторі окремими гармоніками струму і його постійною складовою.

Розглянемо тепер розподіл енергії в спектрі неперіодичного сигналу. Енергія, що виділяється сигналом (струмом) в резисторі з опором в 1 Ом , визначається виразом:

$$W = \int_{-\infty}^{\infty} [x(t)]^2 dt. \quad (2.15)$$

Для визначення розподілу енергії по спектру неперіодичного сигналу визначимо енергію W через модуль спектральної щільності сигналу $S(\omega)$. Квадрат модуля спектральної щільності представимо у вигляді:

$$[S(\omega)]^2 = S(j\omega)S(-j\omega),$$

де $S(-j\omega)$ – комплексно-спряжена функція для спектральної щільності $S(j\omega)$

Згідно з виразом (2.13):

$$S(-j\omega) = \int_{-\infty}^{\infty} x(t) \exp\{j\omega t\} dt.$$

Інтеграл від квадрата модуля спектральної щільності:

$$\int_{-\infty}^{\infty} [S(\omega)]^2 d\omega = \int_{-\infty}^{\infty} S(j\omega) S(-j\omega) d\omega = \int_{-\infty}^{\infty} S(j\omega) \left[\int_{-\infty}^{\infty} x(t) \exp\{j\omega t\} dt \right] d\omega. \quad (2.16)$$

Змінивши у виразі (2.16) порядок інтегрування, отримаємо:

$$\int_{-\infty}^{\infty} [S(\omega)]^2 d\omega = \int_{-\infty}^{\infty} x(t) \left[\int_{-\infty}^{\infty} S(j\omega) \exp\{j\omega t\} d\omega \right] dt = 2\pi \int_{-\infty}^{\infty} [x(t)]^2 dt.$$

Таким чином, енергія сигналу:

$$W = \int_{-\infty}^{\infty} [x(t)]^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} [S(\omega)]^2 d\omega = \frac{1}{\pi} \int_{-\infty}^{\infty} [S(\omega)]^2 d\omega. \quad (2.17)$$

Вираз (2.17), який отримав назву рівняння Парсеваля, показує, що енергія сигналу може бути представлена у вигляді суми нескінченно малих доданків $\frac{1}{\pi} [S(\omega)]^2 d\omega$, що відповідають нескінченно малим ді-

лянкам частотного спектра (рис. 2.10). Вираз $\frac{1}{\pi} [S(\omega)]^2 d\omega$ являє собою енергію, що знаходиться в спектральних складових сигналу, розташованих в смузі частот $d\omega$ в околі частоти ω . Таким чином, квадрат модуля спектральної щільності характеризує розподіл по спектру енергії сигналу.

Якщо задана енергія сигналу ΔW в певній смузі частот $\Delta\omega$ в межах частот ω_k (рис. 2.11), то модуль спектральної щільності у точці ω_k може бути знайденим з наближеного рівняння:

$$S(\omega_k) \approx \sqrt{\frac{\pi \Delta W}{\Delta\omega}}.$$

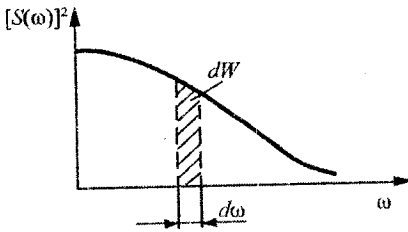


Рис. 2.10

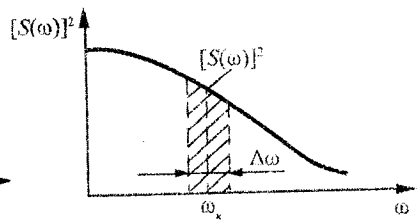


Рис. 2.11

Практична ширина спектра сигналу. Кожний реальний сигнал має скінченну тривалість, відповідно, володіє нескінченним частотним спектром. Практично всі канали зв'язку мають обмежену смугу пропускання. Відповідно, при передачі сигналу через реальний канал зв'язку може бути передана лише частина його частотного спектра. Тому необхідно потурбуватися про те, щоб забезпечити пропускання через канал зв'язку найбільш суттєвої частини спектра. У зв'язку з цим введено поняття практичної ширини спектра сигналу.

За практичну ширину спектра сигналу приймають діапазон частот, в межах якого знаходиться найбільш суттєва частина спектра сигналу. Вибір практичної ширини спектра сигналу визначається двома критеріями: енергетичним і допустимих спотворень форми сигналу.

З енергетичної точки зору практична ширина спектра періодичного сигналу визначається як область частот, в межах якої зосереджена переважна частина всієї потужності сигналу. Якщо розглянути приклад 2.2, то для періодичної послідовності імпульсів прямокутної форми тривалістю $\tau = T/2$ достатньо практичну ширину спектра вибрати рівною

$$3\omega_0 = 6 \frac{\pi}{T} = 3 \frac{\pi}{t}. \text{ У цій області частот зосереджено 95 \% всієї потужності сигналу.}$$

Розглянемо одиночний прямокутний імпульс тривалістю τ і величиною h (рис. 2.12).

Спектральна щільність такого сигналу визначається виразом:

$$S(j\omega) = \int_{-\infty}^{\infty} x(t) \exp\{-j\omega t\} dt = \int_{-\tau/2}^{\tau/2} h \exp\{-j\omega t\} dt = \tau h \frac{\sin \frac{\omega \tau}{2}}{\frac{\omega \tau}{2}} \quad (2.18)$$

і представлена на рис. 2.13.

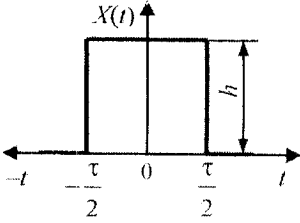


Рис. 2.12

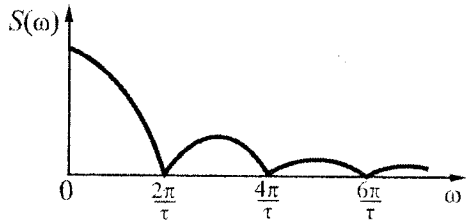


Рис. 2.13

Відповідно до (2.17) енергія сигналу, що зосереджена в смузі частот від 0 до ω_1 :

$$W_1 = \frac{1}{\pi} \int_0^{\omega_1} [S(\omega)]^2 d\omega = \frac{\tau^2 h^2}{\pi} \int_0^{\omega_1} \left[\frac{\sin \frac{\omega\tau}{2}}{\frac{\omega\tau}{2}} \right]^2 d\omega.$$

Відносна величина енергії одиначного імпульсу зосереджена в смузі частот від 0 до ω_1 , виражається функцією:

$$\lambda(\omega_1) = \frac{W_1}{W_0} = \frac{\tau}{\pi} \int_0^{\omega_1} \left[\frac{\sin \frac{\omega\tau}{2}}{\frac{\omega\tau}{2}} \right]^2 d\omega, \quad (2.19)$$

де $W_0 = \tau h^2$ – повна енергія одиначного прямокутного імпульсу.

Графік функції $\lambda(\omega_1)$, наведений на рис. 2.14, називається інтегральною кривою розподілу енергії сигналу в спектрі.

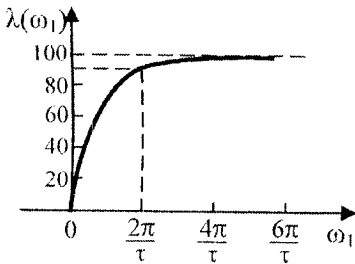


Рис. 2.14

З рисунка видно, що в смузі частот від 0 до $\omega_1 = 2\pi/\tau$ зосереджено більше 90 % всієї енергії сигналу і цей діапазон частот може бути використаний як практична ширина спектра одиначного прямокутного імпульсу. При цьому подальше збільшення практичної ширини спектра веде до незначного збільшення енергії в даному діапазоні частот, тому що при $\omega_1 > 2\pi/\tau$ крива $\lambda(\omega_1)$ досить полого.

Також визначається практична ширина спектра неперіодичних сигналів будь-якої форми.

У тих випадках, коли важливо зберегти потрібну форму сигналу, використовують другий критерій вибору практичної ширини спектра сигналу.

Відомо, що неспотворена передача форми сигналу буде забезпечена при умові, що комплексний коефіцієнт передачі каналу зв'язку має вигляд:

$$K(j\omega) = K \exp\{-j\omega T_0\}, \quad (2.20)$$

тобто амплітудна і фазова характеристики каналу повинні визначатися законом:

$$K(\omega) = K = \text{const}; \quad \varphi(\omega) = \omega T_0.$$

Передача сигналів по реальному каналу, що має обмежену смугу пропускання, супроводжується спотворенням форми сигналів. На практиці важливе значення має забезпечення достатньо високого значення крутизни або допустимого значення тривалості фронту імпульсів.

Для оцінки впливу ширини смуги пропускання каналу зв'язку на спотворення форми сигналів розглянемо як приклад проходження прямокутного імпульсу тривалістю τ і величиною U через канал зв'язку, що являє собою ідеальний фільтр низьких частот. Коефіцієнт передачі цього фільтра виражається залежністю (2.20). При цьому в діапазоні частот $0 \leq \omega \leq \omega_b$ коефіцієнт передачі $K(\omega) = K = \text{const}$ і аргумент $\varphi(\omega) = \omega T_0$; за межами цього діапазону $K(\omega) = 0$ (рис. 2.15).

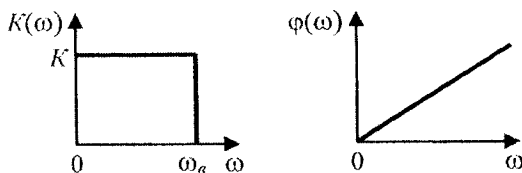


Рис. 2.15

У теорії кіл показано, що вихідний сигнал в цьому випадку може бути представлений аналітично у вигляді:

$$a_{\text{вих}}(t) = UK \left\{ \int_0^{\omega_b} \frac{\sin \omega(t - T_0)}{\omega} d\omega - \int_0^{\omega_b} \frac{\sin \omega [t - (T_0 + \tau)]}{\omega} d\omega \right\}.$$

Форми переднього та заднього фронтів імпульсу спотворюються однаково. На рис. 2.16 показана форма переднього фронту вихідного сигналу.

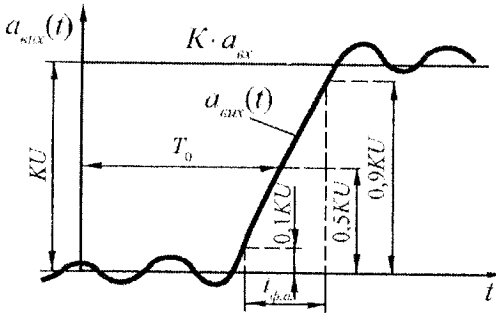


Рис. 2.16

Щоб вихідний сигнал зміг досягнути найбільшого значення, активна довжина переднього фронту $t_{ф.а}$ повинна бути не більше довжини вхідного імпульсу τ . При цьому, як показав аналіз, повинна бути справедливою умова:

$$\tau \cong \frac{0,4}{f_d} \quad \text{або} \quad f_d \cong \frac{0,4}{\tau} \quad (2.21)$$

Вираз (2.21) можна використовувати для вибору практичної ширини спектра прямокутного імпульсу при заданій довжині.

2.3 Приклади

Приклад 2.1. Знайти спектр послідовності косинусоїдальних імпульсів (рис. 2.17).

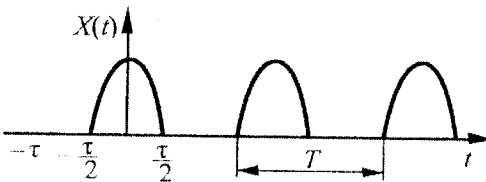


Рис. 2.17

Розв'язування. Функція $x(t)$, що описує цей сигнал, може бути подана наступним чином:

$$x(t) = \begin{cases} h \cos \omega_0 t, & \text{при } -\left(\frac{\tau}{2} - iT\right) \leq t \leq \frac{\tau}{2} + iT; \\ 0, & \text{при } -\left(\frac{3\tau}{2} + iT\right) < t < -\left(\frac{\tau}{2} + iT\right) \end{cases}$$

$$\omega_0 = \frac{2\pi}{T}; \quad T = 2\tau \quad \text{та} \quad \frac{\tau}{2} + iT < t < \frac{3\tau}{2} + iT; \quad i = 0, 1, 2, \dots$$

Комплексна амплітуда сигналу відповідно до (2.6):

$$\begin{aligned} A_k &= \frac{2}{T} \int_{-\tau/2}^{\tau/2} h \cos \omega_0 t \exp\{-jK\omega_0 t\} dt = \\ &= \frac{2h}{T} \int_{-\tau/2}^{\tau/2} \frac{\exp\{j\omega_0 t\} + \exp\{-j\omega_0 t\}}{2} \exp\{-jK\omega_0 t\} dt = \\ &= \frac{h}{2} \left[\int_{-\tau/2}^{\tau/2} \exp\{-j(K-1)\omega_0 t\} dt + \int_{-\tau/2}^{\tau/2} \exp\{-j(K+1)\omega_0 t\} dt \right] = \\ &= \frac{h}{T} \left[\frac{\exp\left\{j(K-1)\omega_0 \frac{\tau}{2}\right\} - \exp\left\{-j(K-1)\omega_0 \frac{\tau}{2}\right\}}{j(K-1)\omega_0} + \right. \\ &\quad \left. + \frac{\exp\left\{j(K+1)\omega_0 \frac{\tau}{2}\right\} - \exp\left\{-j(K+1)\omega_0 \frac{\tau}{2}\right\}}{j(K+1)\omega_0} \right] = \\ &= \frac{h}{\pi(K-1)} \sin(K-1)\omega_0 \frac{\tau}{2} + \frac{h}{\pi(K+1)} \sin(K+1)\omega_0 \frac{\tau}{2} = \\ &= -\frac{h}{\pi(K-1)} \cos \frac{K\omega_0 \tau}{2} + \frac{h}{\pi(K+1)} \cos \frac{K\omega_0 \tau}{2}. \end{aligned}$$

Оскільки

$$\left| \cos \frac{K\omega_0\tau}{2} \right| = \left| \cos \frac{K\pi}{2} \right| = \begin{cases} 1, & \text{при парному } K; \\ 0, & \text{при непарному } K, \end{cases}$$

то спектр сигналу містить тільки парні гармоніки, при цьому комплексна амплітуда:

$$\dot{A}_k = \frac{-2h}{\pi(K^2 - 1)}.$$

Модуль комплексної амплітуди:

$$A_k = \frac{2h}{\pi(K^2 - 1)}.$$

Графік спектра амплітуд зображений на рис. 2.18.

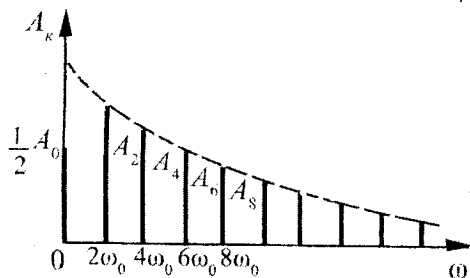


Рис. 2.18

Приклад 2.2. Визначити практичну ширину спектра періодичної послідовності прямокутних імпульсів при ширині імпульсів τ , рівній половині періоду проходження імпульсів T , якщо потрібно врахувати всі гармонічні складового сигналу, що містять не менше 95 % загальної потужності сигналу.

Розв'язування. Як раніше було встановлено, у складі спектра такого сигналу містяться тільки непарні гармонічні складові. Із виразу (2.14) видно, що потужність сигналу може бути подана у вигляді:

$$\begin{aligned} P_0 &= \frac{h^2}{2} + \frac{1}{2} \left(\frac{2h}{\pi} \right)^2 + \frac{1}{2} \left(\frac{2h}{3\pi} \right)^2 + \frac{1}{2} \left(\frac{2h}{5\pi} \right)^2 + \frac{1}{2} \left(\frac{2h}{7\pi} \right)^2 + \dots = \\ &= h^2 \left[\frac{1}{4} + \frac{2}{\pi^2} \left(1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \dots \right) \right] = \frac{1}{2} h^2. \end{aligned}$$

Якщо спектр сигналу обмежити першою гармонікою, то в ньому

буде міститися $\frac{P_1}{P_0} 100 = \frac{h^2 \left(\frac{1}{4} + \frac{2}{\pi^2} \right) 200}{h^2} = 90\%$ усієї потужності сигналу. При урахуванні і третьої гармоніки відсотковий зміст потужності

в спектрі буде $\frac{P_3}{P_0} 100 = \frac{h^2 \left(\frac{1}{4} + \frac{2}{\pi^2} \left(1 + \frac{1}{9} \right) \right) 200}{h^2} = 95\%$.

Таким чином, практичну ширину спектра сигналу можна вважати рівною $3\omega_0 \approx 3 \frac{\pi}{\tau}$.

Приклад 2.3. Знайти спектр одиничного імпульсу високочастотних коливань (рис. 2.19).

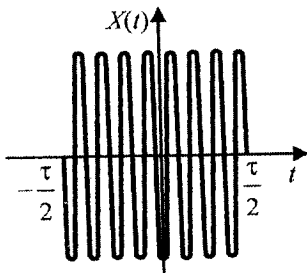


Рис. 2.19

Розв'язування. Функція $x(t)$, що описує сигнал, може бути представлена у вигляді:

$$x(t) = \begin{cases} h \cos \omega_0 t, & \text{при } -\frac{\tau}{2} \leq t \leq \frac{\tau}{2}; \\ 0, & \text{при } \frac{\tau}{2} < t < -\frac{\tau}{2}. \end{cases}$$

Спектральна щільність такого сигналу дорівнює:

$$\begin{aligned} S(j\omega) &= \int_{-\infty}^{\infty} x(t) \exp\{-j\omega t\} dt = \int_{-\tau/2}^{\tau/2} h \cos \omega_0 t \exp\{-j\omega t\} dt = \\ &= \frac{h}{2} \int_{-\tau/2}^{\tau/2} \exp\{j(\omega_0 - \omega)t\} dt + \frac{h}{2} \int_{-\tau/2}^{\tau/2} \exp\{-j(\omega_0 + \omega)t\} dt = \\ &= \frac{h}{2} \left[\frac{\exp\left\{j(\omega_0 - \omega)\frac{\tau}{2}\right\} - \exp\left\{-j(\omega_0 - \omega)\frac{\tau}{2}\right\}}{j(\omega_0 - \omega)} + \right. \end{aligned}$$

$$+ \frac{\exp\left\{-j(\omega_0 + \omega)\frac{\tau}{2}\right\} - \exp\left\{j(\omega_0 + \omega)\frac{\tau}{2}\right\}}{-j(\omega_0 + \omega)} \Bigg] = \frac{h}{(\omega_0 - \omega)} \sin(\omega_0 - \omega)\frac{\tau}{2} +$$

$$+ \frac{h}{(\omega_0 + \omega)} \sin(\omega_0 + \omega)\frac{\tau}{2} = \frac{h\tau}{2} \frac{\sin(\omega_0 - \omega)\frac{\tau}{2}}{(\omega_0 - \omega)\frac{\tau}{2}} + \frac{h\tau}{2} \frac{\sin(\omega_0 + \omega)\frac{\tau}{2}}{(\omega_0 + \omega)\frac{\tau}{2}}.$$

З порівняння отриманого виразу з виразом (2.18) для спектра одиничного імпульсу такої ж тривалості і розміру h , але без високочастотного заповнення, видно, що стосовно спектра прямокутного імпульсу спектр імпульсу високочастотних коливань зміщений на величину несучої ω_0 і розширений у два рази за рахунок появи дзеркального відображення спектра.

Графік модуля спектральної щільності одиничного імпульсу високочастотних коливань показаний на рис. 2.20.

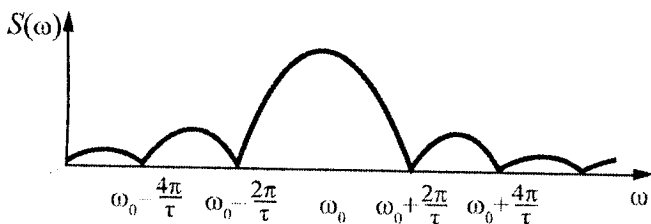


Рис. 2.20

Приклад 2.4. Визначити спектр експоненціального імпульсу (рис. 2.21).

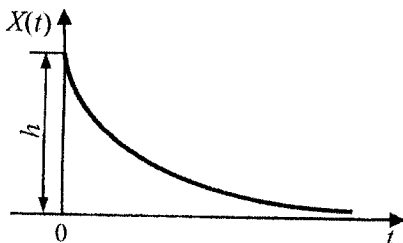


Рис. 2.21

Розв'язування. Експоненціальний імпульс визначається функцією:

$$x(t) = \begin{cases} h \exp\{-\beta t\}, & \text{при } t \geq 0; \\ 0, & \text{при } t < 0. \end{cases}$$

Спектральна щільність такого сигналу:

$$S(j\omega) = \int_0^{\infty} h \exp\{-\beta t\} \exp\{-j\omega t\} dt = \frac{h}{\beta + j\omega}.$$

Модуль і фаза спектральної щільності визначаються виразами:

$$S(\omega) = \frac{h}{\sqrt{\beta^2 + \omega^2}}; \quad \varphi(\omega) = \operatorname{arctg} \frac{\omega}{\beta}.$$

На рис. 2.22 наведені графіки модуля та фази спектральної щільності експоненціального імпульсу.

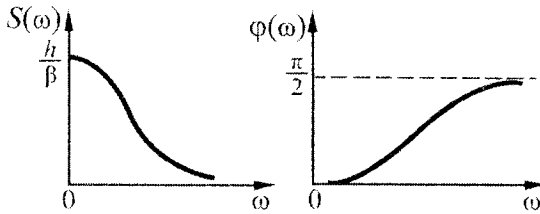


Рис. 2.22

Приклад 2.5. Визначити спектр сигналу включення величиною U (рис. 2.23).

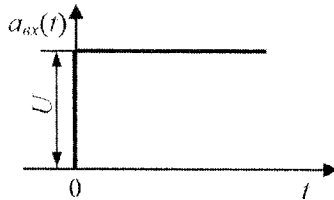


Рис. 2.23

Розв'язування. Сигнал включення визначається функцією:

$$x(t) = U1(t) = \begin{cases} U, & \text{при } t \geq 0; \\ 0, & \text{при } t < 0. \end{cases}$$

Одинична функція $1(t)$ не задовольняє умові абсолютної інтегрованості, отже, до неї не можна застосувати перетворення Фур'є. Однак

її можна розглядати як таку, що утворена з імпульсу експоненціальної форми (рис. 2.21) при необмеженому зменшенні його коефіцієнта затухання $\beta \rightarrow 0$.

Відповідно до цього спектральна щільність сигналу включення буде дорівнювати:

$$S(j\omega) = \lim_{\beta \rightarrow 0} \frac{U}{\beta + j\omega} = \frac{U}{j\omega} = \frac{U}{\omega} \exp\left\{-j\frac{\pi}{2}\right\},$$

звідки модуль і фаза спектральної щільності визначаються виразами:

$$S(\omega) = \frac{U}{\omega}; \quad \varphi(\omega) = \frac{\pi}{2}.$$

Модуль і фаза спектральної щільності сигналу включення наведені на рис. 2.24.

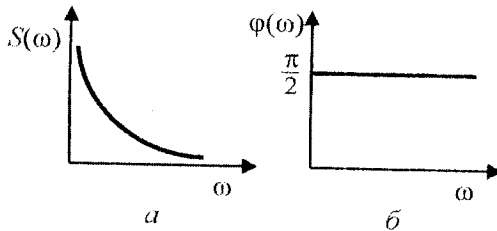


Рис. 2.24

Приклад 2.6. Знайти спектр δ -функції.

Розв'язування. Дельта-функцію можна трактувати як граничну форму прямокутного імпульсу тривалості τ і амплітуди $1/\tau$, що отримується при $\tau \rightarrow 0$ (рис. 2.25).

Тоді згідно з виразом (2.18), прийнявши амплітуду імпульсу рівною $h = 1/\tau$, отримаємо:

$$S(j\omega) = \lim_{\tau \rightarrow 0} \frac{\sin \frac{\omega\tau}{2}}{\frac{\omega\tau}{2}} = 1.$$

Модуль і фаза спектральної щільності дорівнюють $S(\omega) = 1$; $\phi(\omega) = 0$. Графік спектральної щільності δ -функції показаний на рис. 2.26.

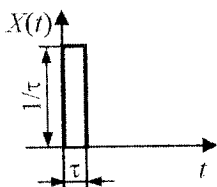


Рис. 2.25

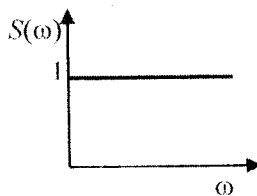


Рис. 2.26

Контрольні питання

1. Які детерміновані сигнали відносяться до елементарних сигналів?
2. Які способи представлення моделей сигналів вам відомі?
3. У чому полягають переваги частотного методу представлення сигналів?
4. При яких умовах періодична функція може бути подана рядом Фур'є?
5. Що розуміється під спектром амплітуд і спектром фаз?
6. Які характерні риси спектра періодичного сигналу?
7. При яких умовах неперіодична функція може бути подана інтегралом Фур'є?
8. Які властивості спектральної щільності сигналу?
9. Яким способом можна одержати спектр неперіодичного сигналу безпосередньо зі спектра відповідного періодичного сигналу?
10. Як можна енергетично витлумачити спектр періодичного і неперіодичного сигналів?
11. Що розуміється під практичною шириною спектра періодичного і неперіодичного сигналів? Які існують критерії вибору практичної ширини спектра сигналу?
12. Як виглядає спектр експоненціального імпульсу, сигналу включення і δ -функції?

МАТЕМАТИЧНІ МОДЕЛІ ВИПАДКОВИХ СИГНАЛІВ

3

3.1 Випадкові сигнали та їх ймовірнісні характеристики

У реальних умовах всі сигнали носять випадковий характер. Внаслідок цього, адресату заздалегідь не відомо, яким буде сигнал, що передається. Але не можна стверджувати, що приймальна сторона не має ніяких попередніх (апріорних) даних про сигнали. По-перше, попередньо відома вся множина, вся гама можливих сигналів. По-друге, відомі, як правило, дані про очікувану ймовірність тих чи інших сигналів із загального переліку сигналів. Наприклад, при передачі мовного тексту відомо, які букви використовуються; які букви, сполучення букв і слів повинні передаватись частіше, а які рідше.

Таким чином, попередні відомості, які ми маємо про сигнали, носять статистичний характер. Тому для дослідження проходження сигналів через інформаційні системи слід застосовувати статистичні методи.

Доцільність застосування статистичних методів обумовлена ще й тим, що на сигнал діють завади, які являють собою, як правило, невідому функцію часу. Основним змістом задачі прийому сигналів на фоні завад є найбільш повне отримання інформації з сигналу. Успішний розв'язок цієї задачі може бути досягнутий тільки на основі використання статистичних методів прийому. Доцільність використання статистичних методів викликана ще й тим, що характеристики систем, по яких проходять сигнали, під впливом різноманітних зовнішніх і внутрішніх факторів можуть змінюватись випадковим чином.

Процес, описом якого є випадкова функція часу, називається **випадковим**. Іноді замість терміну “випадковий процес” застосовують у тому ж розумінні терміни “стохастичний процес” або “ймовірнісний процес”. Конкретний вид, що приймає випадковий процес в результаті дослідження, називається **реалізацією процесу**. Окремі спостереження випадкового процесу, що відбуваються в однотипних системах при однакових умовах дослідження (рис. 3.1), дають різні реалізації випадкового процесу $x_1(t)$, $x_2(t)$, ..., $x_k(t)$. Вид функції $x_k(t)$ випадковим чином змінюється від одного дослідження до іншого. Сукупність реалізацій випадкового процесу, отриманих в результаті дослідів, називається **ансамблем реалізацій випадкового процесу** $x(t)$.

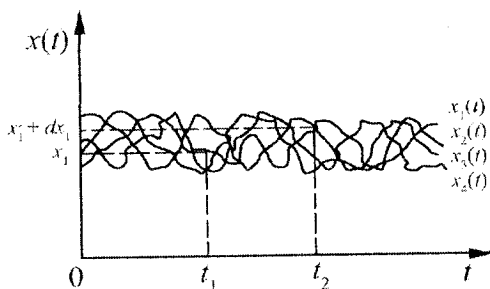


Рис. 3.1

Величина k -ї реалізації випадкового процесу в певний момент часу (наприклад, $t = t_1$) називається **вибіркою** випадкового процесу $x_k(t_1)$. Сукупність значень вибірок в певний момент часу (наприклад, $t = t_1$) утворює випадкову величину $x(t_1)$.

Ймовірність того, що в певний момент часу $t = t_1$ величина x знаходиться в інтервалі між x_1 та $x_1 + \Delta x$, становить:

$$P[x_1 \leq x(t_1) \leq x_1 + dx] = \omega_1(x, t_1) dx, \quad (3.1)$$

де $\omega_1(x, t_1)$ — одновимірна щільність ймовірності або одновимірна функція розподілу випадкового процесу $x(t)$.

Щільність ймовірності $\omega_1(x, t_1)$ є в загальному випадку функцією часу, що залежить від t_1 , і частинною похідною від інтегральної функції розподілу $F_1(x, t_1) = P(x_1 < x)$:

$$\omega_1(x, t_1) = \frac{\partial F_1(x, t_1)}{\partial x}.$$

На рис. 3.2 наведений графік нормального закону розподілу щільності ймовірності випадкової величини x , що найбільш часто зустрічається на практиці.

Математичний опис цього закону має вигляд:

$$\omega_1(x, t_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[x(t_1) - a]^2}{2\sigma^2}\right\}, \quad (3.2)$$

де a та σ — математичне сподівання та середньоквадратичне відхилення випадкової величини x .

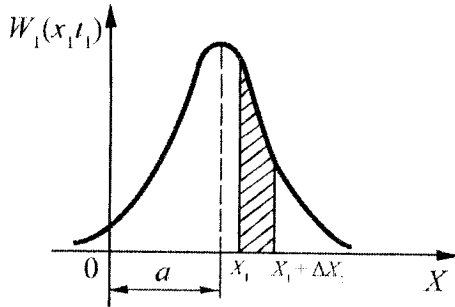


Рис. 3.2

Заштрихована на рис. 3.2 площа під кривою розподілу дорівнює ймовірності знаходження x в інтервалі від x_1 до $x_1 + \Delta x_1$:

$$P[x_1 \leq x(t_1) \leq x_1 + \Delta x_1] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x_1}^{x_1 + \Delta x_1} \exp\left\{-\frac{[x(t_1) - a]^2}{2\sigma^2}\right\} dx. \quad (3.3)$$

При будь-якому законі розподілу щільності ймовірності справедлива рівність:

$$\int_{-\infty}^{\infty} \omega_1(x, x_1) dx = 1. \quad (3.4)$$

Одновимірний закон розподілу щільності ймовірності є найпростішою статистичною характеристикою випадкового процесу. Він дає уявлення про процес лише в окремі, фіксовані моменти часу, характеризує процес статично і не дає уявлення про динаміку його розвитку.

Для більш повної характеристики випадкового процесу необхідно знати зв'язок між ймовірними значеннями випадкової функції у будь-які моменти часу t_1 та t_2 . Цей зв'язок виражається через двовимірну щільність ймовірності і формулюється наступним чином: ймовірність знаходження будь-якої з функцій $x_k(t)$, які входять у сукупність функцій $x(t)$, в інтервалі $(x_1; x_1 + \Delta x_1)$ у момент часу t_1 та в інтервалі $(x_2; x_2 + \Delta x_2)$ у момент часу t_2 дорівнює:

$$P[x_1 \leq x(t_1) \leq x_1 + \Delta x_1; x_2 \leq x(t_2) \leq x_2 + \Delta x_2] = \omega_2(x_1, t_1; x_2, t_2) dx_1 dx_2, \quad (3.5)$$

де $\omega_2(x_1, t_1; x_2, t_2)$ – двовимірна щільність ймовірності або двовимірна функція розподілу випадкового процесу $x(t)$.

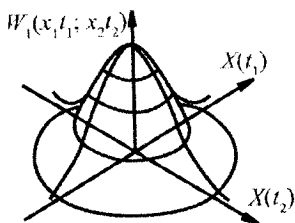


Рис. 3.3

На рис. 3.3 наведена поверхня дво-вимірної нормального закону розподілу щільності ймовірності для випадкових величин з нульовим середнім значенням в моменти часу t_1 та t_2 .

Розмірковуючи аналогічним чином, можна ввести поняття про тривимірну $\omega_2(x_1, t_1; x_2, t_2; x_3, t_3)$, а також n -вимірну $\omega_n(x_1, t_1; x_2, t_2; x_3, t_3, \dots, x_n, t_n)$ щільності ймовірності випадкового процесу $x(t)$.

Тоді ймовірність складної події, яка полягає в тому, що в момент часу t_1 функція $x(t)$ знаходиться в інтервалі $(x_1; x_1 + \Delta x)$ в момент часу t_2 функція $x(t)$ – в інтервалі $(x_2; x_2 + \Delta x)$, в момент часу t_n – в інтервалі $(x_n; x_n + \Delta x_n)$ і т. д., становить:

$$P[x_1 \leq x(t_1) \leq x_1 + \Delta x_1; x_2 \leq x(t_2) \leq x_2 + \Delta x_2; \dots, x_n \leq x(t_n) \leq x_n + \Delta x_n] = \int \omega_n(x_1 t_1; x_2 t_2; \dots; x_n t_n) dx_1, dx_2, \dots, dx_n. \quad (3.6)$$

Чим більше число n , тим точніше n -вимірна функція розподілу характеризує статистичні властивості випадкового процесу. Проте n -вимірні функції розподілу можуть бути отримані за допомогою достатньо важкої та трудомісткої обробки множини реалізацій випадкового процесу. При використанні n -вимірних функцій розподілу зустрічаються суттєві математичні труднощі. Тому на практиці частіше всього оперують скінченною кількістю числових характеристик, які дають, безумовно, менш повну характеристику випадкового процесу, але достатні для вирішення низки важливих задач і, крім того, можуть бути отримані шляхом порівняно простої обробки реалізацій випадкового процесу.

3.2 Числові характеристики випадкового процесу

Як числові характеристики випадкового процесу найчастіше застосовуються моментні функції. Найпростішими моментними функціями, які в основному використовуються для характеристики випадкових процесів, є моменти розподілу перших двох порядків: математичне сподівання, дисперсія і кореляційна функція випадкового процесу.

Математичне сподівання (або перший початковий момент одновимірної закону розподілу) виражається формулою:

$$a[x(t_1)] = m_1[x(t_1)] = \int_{-\infty}^{\infty} x \omega_1(x, t_1) dx. \quad (3.7)$$

Фізично математичне сподівання виражає середнє значення сукупності вибірок випадкового процесу (випадкової величини $x(t_1)$) у визначений момент часу t_1 .

Дисперсія (або другий центральний момент одновимірного закону розподілу) – це математичне сподівання квадрата відхилення величин $x(t_1)$ від математичного сподівання $x(t_1)$ у певний момент часу t_1 . Дисперсія визначається виразом:

$$D[x(t_1)] = m_1 \left\{ [x(t_1) - m_1 \{x(t_1)\}]^2 \right\} = \int_{-\infty}^{\infty} \{x_1 - a[x(t_1)]\}^2 \omega_1(x, t_1) dx. \quad (3.8)$$

Дисперсія виражає міру розкиду значень випадкової величини $x(t_1)$ навколо математичного сподівання, тобто – “ступінь випадковості” величини $x(t_1)$. Корінь квадратний з дисперсії прийнято називати **середнім квадратичним відхиленням** випадкової величини. Середнє значення квадрата випадкової величини $x(t_n)$ – початковий момент другого порядку виражається формулою:

$$\sigma^2 [x(t_1)] = m_1 [x^2(t_1)] = \int_{-\infty}^{\infty} x^2 \omega_1(x, t_1) dx. \quad (3.9)$$

Фізично початковий момент другого порядку виражає повну середню потужність випадкової величини, в той час як другий центральний момент (дисперсія) – середню потужність відхилення випадкової величини відносно математичного сподівання. Додатний корінь $\sigma[x(t_1)]$ з цієї величини називається середнім квадратичним значенням $x(t_1)$.

З виразу (3.8) можна встановити наступну залежність між дисперсією та середнім квадратичним значенням випадкової величини:

$$\begin{aligned} D[x(t_1)] &= \int_{-\infty}^{\infty} \{x^2 - 2ax[x(t_1)] + a^2[x(t_1)]\} \omega_1(x, t_1) dx = \\ &= \int_{-\infty}^{\infty} x^2 \omega_1(x, t_1) dx - 2a[x(t_1)] \int_{-\infty}^{\infty} x \omega_1(x, t_1) dx + a^2[x(t_1)] \int_{-\infty}^{\infty} \omega_1(x, t_1) dx. \end{aligned}$$

З урахуванням того, що $\int_{-\infty}^{\infty} \omega_1(x, t_1) dx = 1$, можна отримати:

$$D[x(t_1)] = \sigma^2[x(t_1)] - a^2[x(t_1)]. \quad (3.10)$$

Таким чином, дисперсія дорівнює різниці квадратів середньоквадратичного значення та математичного сподівання.

При $a = 0$ дисперсія $D[x(t_1)]$ збігається з квадратом середньоквадратичного значення $\sigma^2[x(t_1)]$ випадкової величини $x(t_1)$.

Математичне сподівання, дисперсія та середньоквадратичне значення випадкової величини ϵ в загальному випадку функцією часу. Вони наближено характеризують поведінку випадкового процесу в окремі моменти часу, але зовсім не характеризують зв'язок між значеннями випадкового процесу в різні моменти часу. Цей зв'язок виражає **кореляційна функція** $K_{xx}(t_1, t_2)$ – другий змішаний початковий момент – яка визначається як середнє значення добутку значень випадкової функції $x(t)$ в моменти часу t_1 та t_2 :

$$K_{xx}(t_1, t_2) = m_1[x(t_1)x(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \omega_2(x_1, t_1; x_2, t_2) dx_1 dx_2. \quad (3.11)$$

Зв'язок між значеннями двох випадкових процесів $x(t_1)$ і $y(t_2)$ в моменти часу t_1 і t_2 відповідно виражає взаємна кореляційна функція $K_{xy}(t_1, t_2)$, яка визначається як середнє значення добутку $x(t_1)y(t_2)$:

$$K_{xy}(t_1, t_2) = m_1[x(t_1)y(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 y_2 \omega_2(x_1, t_1; y_2, t_2) dx_1 dy_2. \quad (3.12)$$

Іноді розглядають нормовану автокореляційну r_{xx} і взаємну кореляційну r_{xy} функції (коефіцієнти кореляції), які визначаються виразами:

$$r_{xx}(t_1, t_2) = \frac{K_{xx}(t_1, t_2)}{\sigma[x(t_1)]\sigma[x(t_2)]}, \quad (3.13)$$

$$r_{xy}(t_1, t_2) = \frac{K_{xy}(t_1, t_2)}{\sigma[x(t_1)]\sigma[y(t_2)]}. \quad (3.14)$$

Нормовані кореляційні функції зручні тим, що вони не перевищують одиниці за абсолютною величиною при $a = 0$.

Математичне сподівання, дисперсія, квадрат середньоквадратичного значення і кореляційна функція, які визначаються виразами (3.7)–(3.9) і (3.11), отримані шляхом осереднення за великою кількістю реалізацій випадкового процесу для фіксованих моментів часу. Осереднені характеристики можуть бути також отримані шляхом обробки однієї з реалізацій випадкового процесу на достатньо великому інтервалі часу.

Середнє за часом значення випадкового процесу визначається виразом:

$$\bar{x} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_i(t) dt, \quad (3.15)$$

де $x_i(t)$ – реалізація випадкового процесу $x(t)$; T – час його спостереження.

Аналогічно використовують поняття середнього за часом значення від функції $x^2(t)$, від квадрату різниці $[x(t) - \bar{x}]^2$ та від добутку $x(t)x(t + \tau)$, які визначаються відповідними виразами:

$$\bar{x}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_i^2(t) dt; \quad (3.16)$$

$$\begin{aligned} \overline{[x(t) - \bar{x}]^2} &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [x_i(t) - \bar{x}]^2 dt = \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_i^2(t) dt - \lim_{T \rightarrow \infty} \frac{\bar{x}}{2T} \int_{-T}^T x_i(t) dt + \bar{x}^2 = \bar{x}^2 - \overline{x^2}; \end{aligned} \quad (3.17)$$

$$\overline{x(t)x(t + \tau)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_i(t)x_i(t + \tau) dt. \quad (3.18)$$

Для різних реалізацій випадкового процесу отримують різні значення середнього за часом значення від $x(t)$, $[x(t) - \bar{x}]$ та $x(t + \tau)$.

Якщо припустити, що $x(t)$ являє собою зміну напруги (чи струму), то фізично (3.15) дорівнює потужності постійної складової, яка розсіюється на опорі в 1 Ом. У зв'язку з цим вважають, що середнє за часом \bar{x} виражає потужність постійної складової реалізації випадкового процесу $x_i(t)$. Аналогічно можна вважати, що вираз (3.16) виражає повну середню потужність “випадкової” складової процесу $x(t)$.

Якщо випадковий сигнал є дискретним, то числові характеристики визначаються виразами:

$$\begin{aligned} \bar{x} &= \sum_{k=1}^n P_k x_k; & \bar{x}^2 &= \sum_{k=1}^n P_k x_k^2; \\ \overline{[x_k - \bar{x}]^2} &= \sum_{k=1}^n P_k (x_k - \bar{x})^2; & \overline{x_k x_{k+c}} &= \sum_{k=1}^n P_{k,k+c} x_k x_{k+c}, \end{aligned} \quad (3.19)$$

де P_k – апіорна ймовірність випадкової величини x_k ; $P_{k,k+c}$ – спільна апіорна ймовірність величини x_k та x_{k+c} ; n – кількість значень випадкової величини x .

3.3 Стаціонарні випадкові процеси

В інформаційних системах дуже часто зустрічаються випадкові процеси, які приблизно одночасно відбуваються у часі. Ці процеси мають вигляд неперервних випадкових коливань навколо деякого середнього значення, причому ні середнє значення, ні характер коливань не зазнають суттєвих змін у часі. Такі випадкові процеси називаються **стаціонарними**. Як приклад стаціонарних випадкових процесів можна привести шуми на виході електронних пристроїв, випадкові коливання у колах живлення тощо.

У будь-якій динамічній системі випадковий процес розпочинається з так званого “перехідного” процесу і потім переходить в установлений режим, який з деяким наближенням можна вважати стаціонарним. Строго кажучи, стаціонарні процеси нескінченні в часі, тобто не мають ні початку, ні кінця. Таких процесів практично немає. Однак більшість випадкових процесів на окремих відрізках часу з певним наближенням можна вважати стаціонарними.

Відомі два поняття стаціонарності: стаціонарність у вузькому сенсі і стаціонарність у широкому сенсі.

Під стаціонарними процесами у вузькому сенсі розуміють випадкові процеси, для яких функція розподілу щільності ймовірності $\omega_n = (x_1, t_1; x_2, t_2; \dots; x_n, t_n)$ довільного порядку n не змінюється при будь-якому зсуві всієї групи точок t_1, t_2, \dots, t_n вздовж осі часу, тобто для будь-яких n та τ :

$$\omega_n = (x_1, t_1; x_2, t_2; \dots; x_n, t_n) = \omega_n(x_1, t_1 + \tau; x_2, t_2 + \tau; \dots; x_n, t_n + \tau). \quad (3.20)$$

З наведеного випливає, що для стаціонарних процесів:

а) одновимірна функція розподілу щільності ймовірності не залежить від часу, тобто:

$$\omega_1(x, t_1) = \omega_1(x_1, t_1 + \tau) = \omega_1(x); \quad (3.21)$$

б) двовимірна функція розподілу щільності ймовірності залежить тільки від різниці часу, тобто:

$$\omega_2(x_1, t_1; x_2, t_2) = \omega_2(x_1, x_2, t_2 - t_1) = \omega_2(x_1, x_2, \tau); \quad (3.22)$$

в) тривимірна функція розподілу щільності ймовірності залежить тільки від двох різниць часу $t_2 - t_1 = \tau_1$ та $t_3 - t_1 = \tau_2$, тобто:

$$\begin{aligned} \omega_3(x_1, t_1; x_2, t_2; x_3, t_3) &= \omega_3(x_1, x_2, x_3; t_2 - t_1; t_3 - t_1) = \\ &= \omega_3(x_1, x_2, x_3; \tau_1, \tau_2) \text{ і т. д.} \end{aligned} \quad (3.23)$$

Оскільки математичне сподівання та дисперсія виражаються через одновимірну функцію розподілу щільності ймовірності, то на основі висновку а) можна стверджувати, що для стаціонарного процесу математичне сподівання та дисперсія не залежать від часу. Внаслідок залежності двовимірної функції розподілу тільки від різниці часу $\tau = t_2 - t_1$, кореляційна функція стаціонарного процесу також залежить від різниці часу τ . Таким чином, для стаціонарних процесів:

$$a[x(t)] = a(x); \quad D[x(t)] = D(x); \quad K_{xx}(t_1, t_2) = K_{xx}(\tau). \quad (3.24)$$

На практиці найбільш часто зустрічаються випадкові процеси, для яких при виконанні умов (3.24) моменти вищих порядків залежать від часу. Тому поняття стаціонарності доцільно розширити, прийнявши за основу визначення стаціонарності умови (3.24).

У зв'язку з цим введено поняття стаціонарності у широкому сенсі (або стаціонарності в сенсі А.Я. Хінчина), відповідно до якого стаціонарними процесами є такі випадкові процеси, в яких математичне сподівання і дисперсія не залежать від часу, а кореляційна функція залежить тільки від різниці $\tau = t_2 - t_1$.

Випадкові процеси, стаціонарні у вузькому сенсі, будуть завжди стаціонарними в широкому сенсі, але не навпаки.

Стаціонарні випадкові процеси за своєю природою простіші, чим нестаціонарні, і описуються більш простими характеристиками. Оскільки стаціонарні процеси зустрічаються на практиці дуже часто, одержала широке застосування спеціальна теорія стаціонарних випадкових процесів. Розділ теорії, присвячений вивченню лише тих властивостей процесів, що визначаються моментами перших двох порядків, називається кореляційною теорією випадкових процесів.

Оскільки властивості стаціонарного процесу багато в чому визначаються властивостями кореляційної функції, то для вивчення стаціонарного процесу потрібно, в першу чергу, визначити властивості кореляційної функції.

3.4 Властивості кореляційної функції стаціонарного випадкового процесу

Визначимо, як поводиться кореляційна функція при необмеженому збільшенні різницевого часового інтервалу $\tau = t_2 - t_1$.

У міру збільшення часового інтервалу τ залежність між величинами $x(t)$ та $x(t + \tau)$ слабшає. У межах $t \rightarrow \infty$ ці розміри стають не-

залежними. Тоді з урахуванням того, що математичне сподівання добутку випадкових незалежних розмірів дорівнює добутку математичних сподівань співмножників і що для стаціонарного процесу математичне сподівання не залежить від часу, отримаємо:

$$\begin{aligned} \lim_{\tau \rightarrow \infty} K_{xx}(\tau) &= \lim_{\tau \rightarrow \infty} m_1 [x(t)x(t + \tau)] = \\ &= \lim_{\tau \rightarrow \infty} (m_1 [x(t)]m_1 [x(t + \tau)]) = a^2(x). \end{aligned} \quad (3.25)$$

Таким чином, при необмеженому збільшенні аргументу τ кореляційна функція прямує до квадрата математичного сподівання випадкового процесу. Отже, при $\tau \rightarrow \infty$ кореляційна функція дорівнює потужності постійної складової реалізації випадкового стаціонарного процесу.

При зменшенні часового інтервалу τ залежність між розмірами $x(t)$ та $x(t + \tau)$ підсилюється і у межі при $\tau \rightarrow 0$ одержимо:

$$\lim_{\tau \rightarrow 0} K_{xx}(\tau) = \lim_{\tau \rightarrow 0} m_1 [x(t)x(t + \tau)] = m_1 [x^2(t)] = m_2 [x(t)]. \quad (3.26)$$

Таким чином, при $\tau \rightarrow 0$ кореляційна функція дорівнює початковому моменту другого порядку функції $x(t)$.

Отже, дисперсія стаціонарного випадкового процесу:

$$D(x) = K_{xx}(0) - K_{xx}(\infty). \quad (3.27)$$

В силу незалежності функції розподілу щільності ймовірності стаціонарного випадкового процесу від початку відліку часу кореляційна функція є парною функцією τ , тобто:

$$K_{xx}(\tau) = K_{xx}(-\tau). \quad (3.28)$$

Можна показати, що кореляційна функція за абсолютним значенням максимальна при $\tau = 0$. Дійсно, якщо розглянути математичне сподівання квадрата суми або різниці величин $x(t)$ і $x(t + \tau)$, то отримаємо:

$$\begin{aligned} m_1 \left\{ [x(t) \pm x(t + \tau)]^2 \right\} &= m_1 [x^2(t)] \pm 2m_1 [x(t)x(t + \tau)] + \\ &+ m_1 [x^2(t + \tau)] = 2K_{xx}(0) \pm 2K_{xx}(\tau). \end{aligned}$$

Оскільки середнє значення додатної величини (квадрата суми або різниці двох величин) не може бути від'ємним, то:

$$2K_{xx}(0) \pm 2K_{xx}(\tau) \geq 0.$$

Звідки:

$$|K_{xx}(\tau)| \leq |K_{xx}(0)|. \quad (3.29)$$

На рис. 3.4 показані типові криві кореляційної функції $K_{xx}(\tau)$. Як видно з рис. 3.4, асимптотичне наближення функції $K_{xx}(\tau)$ до сталого значення $a^2(x)$ може відбуватися як монотонно (рис. 3.4, а), так і немонотонно (рис. 3.4, б).

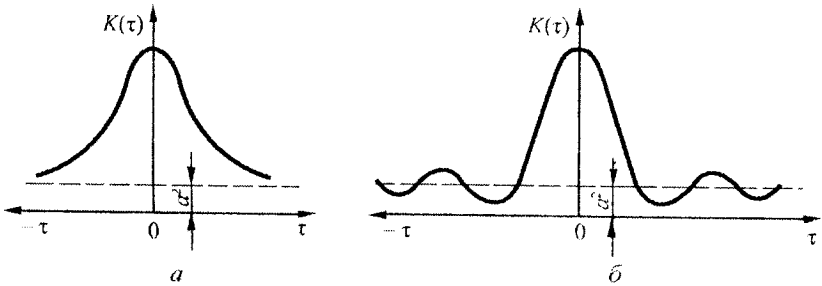


Рис. 3.4

На практиці часто замість випадкового процесу $x(t)$ розглядають його відхилення від математичного сподівання $x_0(t) = x(t) - a(t)$, яке називають пульсаціями або флуктуаціями процесу (центрованим випадковим процесом).

Кореляційна функція пульсацій стаціонарного випадкового процесу дорівнює:

$$\begin{aligned} K_{xx}^0(\tau) &= m_1[x_0(t)x_0(t+\tau)] = m_1\{[x(t) - a(x)][x(t+\tau) - a(x)]\} = \\ &= m_1[x(t)x(t+\tau)] - a(x)m_1[x(t+\tau)] - a(x)m_1[x(t)] + a^2(x) = \quad (3.30) \\ &= K_{xx}(\tau) - a^2(x). \end{aligned}$$

З виразу (3.30) випливає, що математичне сподівання пульсацій дорівнює нулю, а їхня дисперсія:

$$D^0(x) = K_{xx}^0(0). \quad (3.31)$$

Відношення

$$r_x^0(\tau) = \frac{K_{xx}(\tau)}{D^0(x)} \quad (3.32)$$

називається нормованою кореляційною функцією (коефіцієнтом кореляції) пульсації випадкового процесу (або випадкового процесу з нульовим середнім). Типові криві нормованої кореляційної функції пульсації показані на рис. 3.5.

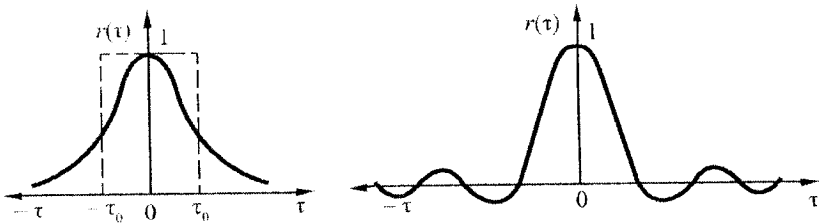


Рис. 3.5

Для чисто випадкового стаціонарного процесу завжди можна вказати таке значення інтервалу $\tau > \tau_0$, що при $\tau = \tau_0$ величини $x(t)$ та $x(t + \tau)$ можна вважати практично незалежними, причому практична незалежність розуміється в тому значенні, що при $\tau > \tau_0$ абсолютний розмір коефіцієнта кореляції залишається менше заданого розміру:

$$|r_x^0(\tau)| < \delta. \quad (3.33)$$

Величина δ зазвичай приймається рівною 0,05. Інтервал τ_0 називається часом кореляції випадкового процесу. Час кореляції іноді визначається як половина ширини основи прямокутника одиничної висоти, площа якого дорівнює площі під графіком коефіцієнта кореляції (див. рис. 3.5), тобто:

$$\tau_0 = \frac{1}{2} \int_{-\infty}^{\infty} r_x^0(\tau) d\tau = \frac{1}{K_{xx}^0(0)} \int_{-\infty}^{\infty} K_{xx}^0(\tau) d\tau. \quad (3.34)$$

3.5 Ергодичність стаціонарних процесів

Існує клас випадкових процесів, що володіють важливою для практичного застосування властивістю ергодичності.

Випадковий процес називається ергодичним, якщо осереднення за множиною з ймовірністю, як завгодно близькою до одиниці, дорівнює осередненню за часом.

Отже, для ергодичних процесів справедливі рівності:

$$\begin{aligned}
 a(x) &= m_1 [x(t)] = \bar{x}; \\
 D(x) &= m_1 \left[x(t) - m_1 \{x(t)\} \right]^2 = \overline{[x(t) - \bar{x}]^2}; \\
 \sigma^2(x) &= m_1 [x^2(t)] = \bar{x}^2; \\
 K_{xx}(\tau) &= m_1 [x(t)x(t + \tau)] = \overline{x(t)x(t + \tau)}.
 \end{aligned}
 \tag{3.35}$$

Розгорнувши вирази середніх за множиною і за часом, одержимо:

$$\begin{aligned}
 a(x) &= \int_{-\infty}^{\infty} x \omega_1(x) dx = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_i(t) dt; \\
 D(x) &= \int_{-\infty}^{\infty} [x - a(x)]^2 \omega_1(x) dx = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \int_{-T}^T [x_i(t) - x]^2 dt; \\
 \sigma^2(x) &= \int_{-\infty}^{\infty} x^2 \omega_1(x) dx = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_i^2(t) dt; \\
 K_{xx}(\tau) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \omega_2(x_1, x_2, \tau) dx_1 dx_2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_i(t) x_i(t + \tau) dt.
 \end{aligned}
 \tag{3.36}$$

Ергодична властивість випадкового процесу має велике практичне значення. При дослідженні таких процесів немає потреби вивчати велику сукупність реалізацій, а досить лише однієї реалізації, що спостерігається протягом тривалого проміжку часу. Наприклад, статистичні властивості флукуаційних шумів на виході електронних підсилювачів можна вивчати протягом досить тривалого часу на одному підсилювачі, а потім результати цього дослідження поширити на всі ідентичні пристрої.

3.6 Спектральна щільність випадкового процесу

При вивченні детермінованих сигналів дуже зручним виявився гармонічний аналіз. У зв'язку з цим бажано використовувати апарат перетворень Фур'є і до випадкових процесів. Однак безпосереднє застосування класичного гармонічного аналізу до випадкових процесів неможливо в основному з двох причин:

1) реалізації випадкового процесу $x_k(t)$ зазвичай не задовольняють умові абсолютної інтегрованості:

$$\int_{-\infty}^{\infty} |x_k(t)| dt < \infty;$$

2) для випадкового процесу $x(t)$ частотний спектр (у класичному розумінні) також є випадковою функцією.

Для узагальнення апарата перетворень Фур'є на випадкові процеси необхідно використовувати таку характеристику процесу, що має не випадковий вигляд. Такою характеристикою є середня потужність процесу.

Шляхом осереднення за ансамблем реалізації отримано вираз для енергетичної характеристики сукупності реалізації випадкового процесу:

$$G(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} K(\tau) \exp\{-j\omega\tau\} d\tau = \frac{2}{\pi} \int_0^{\infty} K(\tau) \cos\omega\tau d\tau. \quad (3.37)$$

Цей вираз являє собою пряме перетворення Фур'є для кореляційної функції. Зворотне перетворення Фур'є має вигляд:

$$K(\tau) = \frac{1}{2} \int_{-\infty}^{\infty} G(\omega) \exp\{j\omega\tau\} d\omega = \int_{-\infty}^{\infty} G(\omega) \cos\omega\tau d\omega. \quad (3.38)$$

Перетворення (3.37) і (3.38), що зв'язують функції $G(\omega)$ та $K(\tau)$, називаються перетвореннями Хінчина-Вінера.

Оскільки для стаціонарних ергодичних процесів осереднення за множиною може бути замінено осередненням за часом, то функцію $G(\omega)$ можна представити у вигляді:

$$G(\omega) = \lim_{T \rightarrow \infty} \frac{|S_k(j\omega)|^2}{\pi T}. \quad (3.39)$$

Таким чином, енергетичний спектр стаціонарного ергодичного процесу може бути вирахований двома шляхами:

а) безпосереднім спостереженням однієї реалізації $x_k(t)$ і знаходженням границі (3.39);

б) знаходженням перетворення Фур'є кореляційної функції.

Функція $G(\omega)$ відіграє велику роль при дослідженні перетворення випадкових сигналів лінійними системами. Зручність спектральної щільності $G(\omega)$ полягає в тому, що функція $G(\omega)$ може існувати навіть тоді, коли не існує перетворення Фур'є від функції $x_k(t)$.

Для з'ясування фізичного змісту функції $G(\omega)$ прийемо у виразі (3.38) $\tau = 0$:

$$K(0) = \int_0^{\infty} G(\omega) d\omega. \quad (3.40)$$

Оскільки $K(0)$ виражає потужність випадкового сигналу, то $G(\omega)$ дає осереднену енергетичну картину розподілу потужності сигналу за частотним спектром. Вираз (3.40) показує, що потужність випадкового сигналу може бути подана у вигляді нескінченної суми нескінченно малих доданків $G(\omega)d\omega$, що відповідають потужностям, які містяться у межах спектра $[\omega; \omega + d\omega]$.

3.7 Широко- та вузькосмугові процеси

З властивості перетворень Фур'є випливає, що при розтягу функції $K(\tau)$ її частотний спектр $G(\omega)$ стискується і, навпаки, при звууженні $K(\tau)$ частотний спектр $G(\omega)$ розширюється (рис. 3.6).

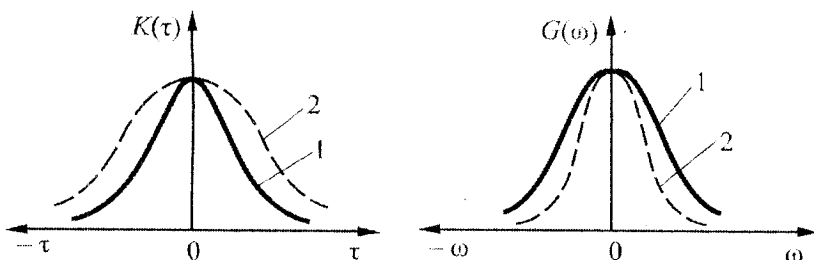


Рис. 3.6

У зв'язку з цим розглянемо граничний і найбільш цікавий випадок $G(\omega) = G_0 = \text{const}$, тобто випадок, коли спектральна щільність потужності рівномірна на всіх частотах і спектр сигналу необмежений.

Випадковий процес, що має рівномірний на всіх частотах спектр, називають **білим шумом**. Функція спектральної щільності білого шуму подана на рис. 3.7, а.

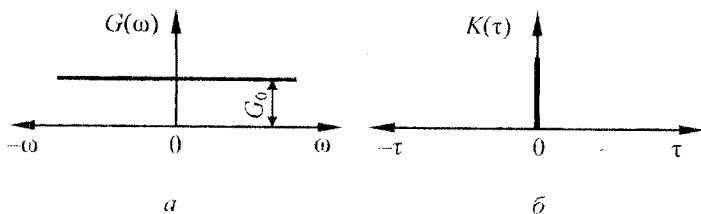


Рис. 3.7

Кореляційна функція білого шуму:

$$K(\tau) = \frac{1}{2} \int_{-\infty}^{\infty} G_0 \exp\{j\omega\tau\} d\omega = \pi G_0 \delta(\tau), \quad (3.41)$$

оскільки інтеграл

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{j\omega\tau\} d\omega = \delta(\tau).$$

Таким чином, кореляційна функція білого шуму виражається дельта-функцією (рис. 3.7, б):

$$K(\tau) = \begin{cases} \infty, & \text{при } \tau = 0; \\ 0, & \text{при } \tau \neq 0. \end{cases}$$

Очевидно, білий шум $x(t)$ характеризується тим, що значення $x(t)$ у будь-які два (навіть як завгодно близькі) моменти часу некорельовані. Такий випадковий процес іноді називають абсолютно випадковим.

Необхідно підкреслити, що поняття “білий шум” засновано на спектральній властивості випадкового процесу і цілком не пов’язано з законами розподілу щільності ймовірності. Зокрема, якщо білий шум має нормальний закон розподілу, його називають нормальним білим шумом.

Білий шум у точному розумінні є ідеалізацією, що ніколи не зустрічається в реальних умовах, хоча б тому, що досить близькі значення випадкової функції практично завжди залежні, а також і тому, що реальні процеси мають скінченну потужність, а для білого шуму повна потужність процесу нескінченна. Однак подібна ідеалізація у багатьох важливих практичних випадках значно спрощує математичний аналіз і не вносить істотних похибок.

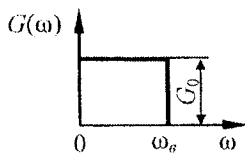


Рис. 3.8

Кореляційна функція такого сигналу:

$$K(\tau) = \int_0^{\omega_B} G_0 \cos \omega\tau d\omega = G_0 \omega_B \frac{\sin \omega_B \tau}{\omega_B \tau} = P_0 \frac{\sin \omega_B \tau}{\omega_B \tau}, \quad (3.42)$$

де $P_0 = G_0 \omega_B$ – середня потужність процесу.

Графік кореляційної функції (3.42) наведено на рис. 3.9.

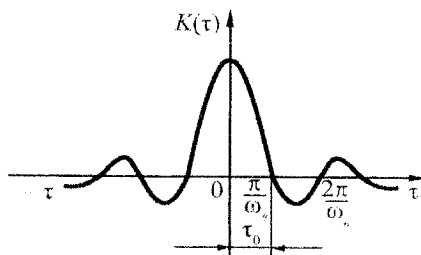


Рис. 3.9

Для такого процесу інтервал кореляції τ_0 має скінченну величину, яку можна визначити, наприклад, як інтервал між точкою $\tau = 0$ і точкою, де $K(\tau)$ перший раз перетворюється в нуль, тобто $\tau_0 = \pi/\omega_0$.

Отже, обмеження спектра тягне за собою появу кореляції, причому в міру скорочення смуги частот $\Omega = \omega_0$ інтервал кореляції збільшується.

Якщо для випадкового процесу спектр безперервний і зосереджений біля деякої фіксованої частоти ω_0 , причому виконується умова:

$$\frac{\Omega_c}{\omega_0} \ll 1, \quad (3.43)$$

то такий процес називається **вужькосмуговим**.

Якщо вужькосмуговий спектр має максимум при ω_0 і симетричний щодо цієї точки (рис. 3.10), то кореляційна функція процесу:

$$K(\tau) = \int_{-\infty}^{\infty} G(\omega) \cos \omega \tau d\omega = \int_{-\infty}^{\omega_0} G(\omega_0 - \omega) \cos(\omega_0 - \omega) \tau d\omega = \\ = \left[\int_{-\infty}^{\omega_0} G(\omega_0 - \omega) \cos \omega \tau d\omega \right] \cos \omega_0 \tau + \left[\int_{-\infty}^{\omega_0} G(\omega_0 - \omega) \sin \omega \tau d\omega \right] \sin \omega_0 \tau.$$

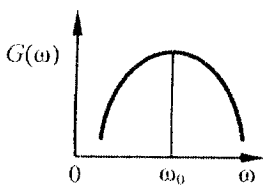


Рис. 3.10

Оскільки за умовою смуга спектра мізерно мала в порівнянні з частотою ω_0 , то верхні межі інтегрування без особливої похибки можна поширити до нескінченності. Приймаючи до уваги, що другий інтеграл, як інтеграл від синуса в нескінченних межах, дорівнює нулю, одержимо:

$$K(\tau) \cong \cos \omega_0 \tau \int_{-\infty}^{\infty} G^*(\omega) \cos \omega \tau d\omega = b(\tau) \cos \omega_0 \tau, \quad (3.44)$$

де

$$G^*(\omega) = G(\omega_0 - \omega);$$

$$b(\tau) = \int_{-\infty}^{\infty} G^*(\omega) \cos \omega \tau d\omega = 2 \int_0^{\infty} G^*(\omega) \cos \omega \tau d\omega.$$

З виразу (3.44) видно, що кореляційна функція вузькосмугового процесу, спектр якого розташований симетрично біля деякої частоти ω_0 , дорівнює помноженій на $\cos \omega_0 \tau$ кореляційній функції $b(t)$, яка відповідає спектру $G^*(\omega)$, отриманому з початкового зміщенням на величину ω_0 в область низьких частот.

3.8 Ефективна ширина спектра випадкового процесу

Під ефективною або еквівалентною шириною спектра прийнято розуміти ширину смуги частот, у межах якої зосереджена переважна частина (зазвичай 95 %) потужності процесу. Однак при аналізі випадкових процесів, що характеризуються нерівномірним спектром із яскраво вираженим максимумом (рис. 3.11), часто користуються поняттям ефективною ширини спектра, обумовленою виразом:

$$\Omega_c = \frac{\int_0^{\infty} G(\omega) d\omega}{G_{\max}} = \frac{K(0)}{G_{\max}}, \quad (3.45)$$

де G_{\max} – найбільше значення функції спектральної щільності.

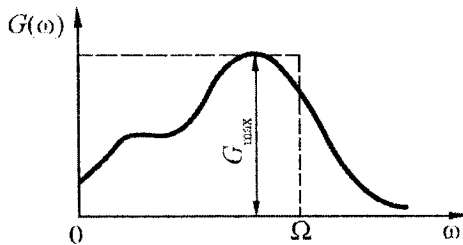


Рис. 3.11

Оскільки інтервал кореляції визначається із врахуванням виразу (3.37) співвідношенням

$$\tau_0 = \frac{\int_0^{\infty} |K(\tau)| d\tau}{K(0)} = \frac{\pi G(0)}{2K(0)},$$

то може бути встановлений наступний зв'язок між інтервалом кореляції та ефективною шириною спектра процесу:

$$\tau_0 = \frac{\pi G(0)}{2G_{\max} \Omega_e}. \quad (3.46)$$

3.9 Приклади

Приклад 3.1

У вимірювальному приладі відстань між сусідніми мітками шкали постійна і дорівнює α . При округленні відліку до найближчої цілої поділки похибка округлення за абсолютною величиною не перевищує половини відстані між сусідніми мітками шкали.

Знайти щільність розподілу ймовірності, математичне сподівання і дисперсію похибки округлення.

Розв'язування. Похибку при округленні відліку можна розглядати як випадкову величину, яка може приймати з рівною ймовірністю будь-які значення в межах від $-\alpha/2$ до $+\alpha/2$. Отже, щільність розподілу випадкової величини x постійна в межах від $-\alpha/2$ до $+\alpha/2$ і дорівнює нулю за цими межами (рис. 3.12).

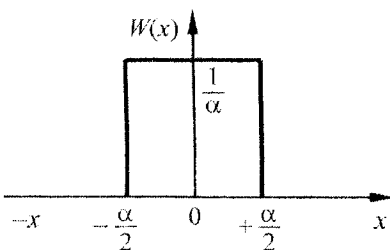


Рис. 3.12

Оскільки повинна бути справедлива рівність:

$$\int_{-\alpha/2}^{\alpha/2} \omega_1(x) dx = \omega_1(x) \int_{-\alpha/2}^{\alpha/2} dx = 1,$$

то:

$$\omega_1(x) = \frac{1}{\alpha}.$$

Аналітичний закон рівномірного розподілу можна записати так:

$$\omega_1(x) = \begin{cases} 0, & \text{при } x < -\alpha/2; \\ 1/\alpha, & \text{при } -\alpha/2 \leq x \leq \alpha/2; \\ 0, & \text{при } x > \alpha/2. \end{cases}$$

Математичне сподівання і дисперсія похибки округлення відповідно до (3.7) і (3.8):

$$a(x) = \int_{-\infty}^{\infty} x \omega_1(x) dx = \int_{-\alpha/2}^{\alpha/2} x \frac{1}{\alpha} dx = \frac{1}{\alpha} \frac{x^2}{2} \Big|_{-\alpha/2}^{\alpha/2} = \frac{1}{\alpha} \left(\frac{\alpha^2}{8} - \frac{\alpha^2}{8} \right) = 0;$$

$$\begin{aligned} D(x) &= \int_{-\infty}^{\infty} [x - a(x)]^2 \omega_1(x) dx = \int_{-\alpha/2}^{\alpha/2} x^2 \frac{1}{\alpha} dx = \\ &= \frac{1}{\alpha} \frac{x^3}{3} \Big|_{-\alpha/2}^{\alpha/2} = \frac{1}{\alpha} \left(\frac{\alpha^3}{24} + \frac{\alpha^3}{24} \right) = \frac{\alpha^2}{12}. \end{aligned}$$

Приклад 3.2

Контрольно-вимірювальний пристрій має систематичну похибку 30 мВ і випадкову похибку, розподілену за нормальним законом із середнім квадратичним відхиленням 10 мВ. Знайти ймовірність того, що загальна похибка пристрою прийме значення, що належить інтервалу 10–50 мВ.

Розв'язування. Ймовірність того, що похибка x пристрою не перевищить за абсолютною величиною заданої межі, дорівнює:

$$P(10 \leq x \leq 50) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{10}^{50} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx.$$

Введемо нову змінну:

$$Z = \frac{x-a}{\sigma}.$$

Звідси: $x = \sigma z + a$; $dx = \sigma dz$.

Тоді:

$$P(10 \leq x \leq 50) = \frac{1}{\sqrt{2\pi}} \int_{\frac{10-a}{\sigma}}^{\frac{50-a}{\sigma}} \exp\left[-\frac{z^2}{2}\right] dz =$$

$$\begin{aligned}
 &= \frac{1}{\sqrt{2\pi}} \int_{\frac{10-a}{\sigma}}^0 \left\{ -\frac{z^2}{2} \right\} dz + \frac{1}{\sqrt{2\pi}} \int_0^{\frac{50-a}{\sigma}} \exp\left\{ -\frac{z^2}{2} \right\} dz = \\
 &= \frac{1}{\sqrt{2\pi}} \int_0^{\frac{50-a}{\sigma}} \exp\left\{ -\frac{z^2}{2} \right\} dz - \frac{1}{\sqrt{2\pi}} \int_0^{\frac{10-a}{\sigma}} \exp\left\{ -\frac{z^2}{2} \right\} dz.
 \end{aligned}$$

Користуючись функцією Лапласа:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left\{ -\frac{z^2}{2} \right\} dz,$$

остаточно отримаємо:

$$P(10 \leq x \leq 50) = \Phi\left(\frac{50-a}{\sigma}\right) - \Phi\left(\frac{10-a}{\sigma}\right).$$

У нашому випадку: $a = 30$ мВ та $\sigma = 10$ мВ.

За допомогою таблиці функції Лапласа (додаток А) знайдемо:

$$P = (10 \leq x \leq 50) = \Phi(2) - \Phi(-2) = 2\Phi(2) = 2 \cdot 0,4772 = 0,9544.$$

Приклад 3.3

Спільна щільність ймовірності випадкових величин x і y має вигляд:

$$\begin{aligned}
 \omega_2(x, y) = & \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-K_{xy}^2}} \exp\left\{ -\frac{1}{2(1-K_{xy}^2)} \left[\frac{(x-a_x)^2}{\sigma_x^2} - \right. \right. \\
 & \left. \left. \frac{2K_{xy}(x-a_x)(y-a_y)}{\sigma_x\sigma_y} + \frac{(y-a_y)^2}{\sigma_y^2} \right] \right\},
 \end{aligned}$$

де $a_x, a_y, \sigma_x, \sigma_y, K_{xy}$ – параметри розподілу.

Визначити щільності ймовірностей $\omega_1(x)$ і $\omega_2(y)$ випадкових величин x і y .

Розв'язування. Щільності ймовірностей $\omega_1(x)$ і $\omega_2(y)$ можуть бути визначені, виходячи з наступних властивостей двовимірної інтегральної функції розподілу $F_2(x, y)$:

$$F_2(x, y) = \int_{-\infty}^x \int_{-\infty}^y \omega_2(x, y) dx dy,$$

$$F_2(x, \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} \omega_2(x, y) dx dy = F_1(x);$$

$$F_2(\infty, y) = \int_{-\infty}^{\infty} \int_{-\infty}^y \omega_2(x, y) dx dy = F_1(y).$$

де $F_1(x)$ і $F_2(y)$ – одновимірні інтегральні функції розподілу випадкових величин x і y .

Тоді одновимірні щільності ймовірностей $\omega_2(x)$ і $\omega_2(y)$ можуть бути визначені зі співвідношень:

$$\omega_1(x) = \frac{dF_1(x)}{dx} = \int_{-\infty}^{\infty} \omega_2(x, y) dy;$$

$$\omega_1(y) = \frac{dF_1(y)}{dy} = \int_{-\infty}^{\infty} \omega_2(x, y) dx;$$

Зробивши заміну змінних $\frac{x - \alpha_x}{\sqrt{2\sigma_x}} = U$, $\frac{y - \alpha_y}{\sqrt{2\sigma_y}} = V$, одержимо:

$$\omega_1(x) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-K_{xy}^2}} \exp\left\{-\frac{U^2}{1-K_{xy}^2}\right\} \times$$

$$\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{1-K_{xy}^2}V^2 + \frac{2K_{xy}}{1-K_{xy}^2}U\right\} dV.$$

$$\omega_1(y) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-K_{xy}^2}} \exp\left\{-\frac{V^2}{1-K_{xy}^2}\right\} \times$$

$$\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{1-K_{xy}^2}U^2 + \frac{2K_{xy}}{1-K_{xy}^2}U\right\} dU.$$

Відомо, що:

$$\int_{-\infty}^{\infty} \exp\{-P^2x^2 \pm gx\} dx = \frac{\sqrt{\pi}}{P} \exp\left\{\frac{g^2}{4P^2}\right\},$$

тоді

$$\omega_1(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\{-U^2\}$$

або

$$\omega_1(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{x^2}{2\sigma_x^2}\right\}.$$

Таким чином, величина x підпорядкована гаусівському закону з параметрами α_x та σ_x .

Зробивши аналогічні обчислення, отримасмо:

$$\omega_1(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{(y - a_y)^2}{2\sigma_y^2}\right\}.$$

Приклад 3.4

Кореляційна функція процесу $x(t)$ задана виразом:

$$K_x(\tau) = A \exp\{-\alpha|\tau|\},$$

де $\alpha > 0$ (рис. 3.13). Знайти спектральну щільність.

Розв'язування. Відповідно до (3.37) спектральна щільність:

$$G(\omega) = \frac{2}{\pi} \int_0^{\infty} K_x(\tau) \cos \omega\tau d\tau = \frac{2A}{\pi} \int_0^{\infty} \exp\{-\alpha\tau\} \cos \omega\tau d\tau = \frac{2A\alpha}{\pi(\alpha^2 + \omega^2)}.$$

Графік $G(\omega)$ показаний на рис. 3.14.

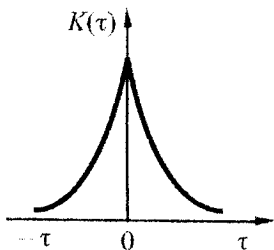


Рис. 3.13

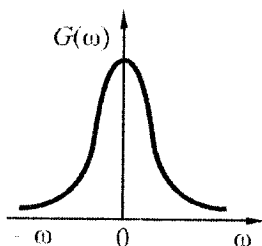


Рис. 3.14

Приклад 3.5

Знайти спектральну щільність для процесу з кореляційною функцією виду δ -функції.

Розв'язування. Виходячи з визначення δ -функції як границі прямокутного імпульсу $\delta_\tau(t)$, тривалості τ і висоти $1/\tau$ при $\tau \rightarrow 0$ інтеграл $\int_{-\infty}^{\infty} x(t)\delta(t)dt$, де $x(t)$ – довільна функція, може бути представлений у вигляді:

$$\int_{-\infty}^{\infty} x(t)\delta(t)dt = \lim_{\tau \rightarrow 0} \int_{-\infty}^{\infty} x(t)t\delta_\tau(t)dt = \lim_{\tau \rightarrow 0} \int_{-\tau/2}^{\tau/2} x(t)\frac{1}{\tau}dt.$$

Згідно з теоремою про середнє значення:

$$\lim_{\tau \rightarrow 0} \int_{-\tau/2}^{\tau/2} x(t)\frac{1}{\tau}dt = \lim_{\tau \rightarrow 0} x(\theta)\frac{1}{\tau} \int_{-\tau/2}^{\tau/2} dt,$$

де $x(\theta)$ – середнє значення $x(t)$ у межах $\left[-\frac{\tau}{2}; \frac{\tau}{2}\right]$.

$$|\theta| \leq \tau$$

Оскільки інтеграл $\int_{-\tau/2}^{\tau/2} \frac{1}{\tau}dt = 1$, то:

$$\lim_{\tau \rightarrow 0} x(\theta) \int_{-\tau/2}^{\tau/2} \frac{1}{\tau}dt = \lim_{\substack{\tau \rightarrow 0 \\ |\theta| \leq \tau}} x(\theta) = x(\theta).$$

Тоді, прийнявши у формулі (3.37) для спектральної щільності $K_x(\tau) = \delta(\tau)$ та $\exp\{j\omega\tau\} = x(t)$, з урахуванням отриманих співвідношень, одержимо:

$$\begin{aligned} G(\omega) &= \frac{1}{\pi} \int_{-\infty}^{\infty} K_x(\tau) \exp\{-j\omega\tau\} d\tau = \frac{1}{\pi} \int_{-\infty}^{\infty} \delta(\tau) \exp(-j\omega\tau) d\tau = \\ &= \frac{1}{\pi} \left. \exp\{-j\omega\tau\} \right|_{\tau=0} = \frac{1}{\pi}. \end{aligned}$$

Отже, при кореляційній функції типу δ -функції спектр рівномірний на всіх частотах (сигнал типу “білий шум” – див. рис. 3.7, а, б).

Приклад 3.6

Знайти кореляційну функцію для процесу зі спектральною щільністю (рис. 3.15) у вигляді δ -функції – $G(\omega) = \delta(\omega)$.

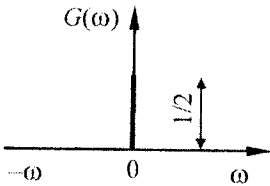


Рис. 3.15

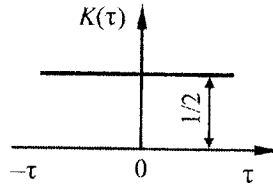


Рис. 3.16

Розв'язування. Відповідно до загальної формули (3.38) і виходячи з визначення d -функції як межі прямокутної функції шириною Ω і висотою $1/\Omega$ при $\Omega \rightarrow \infty$, отримуємо:

$$\begin{aligned} K(\tau) &= \frac{1}{2} \int_{-\infty}^{\infty} G(\omega) \exp\{j\omega\tau\} d\omega = \frac{1}{2} \int_{-\infty}^{\infty} \delta(\omega) \exp\{j\omega\tau\} d\omega = \\ &= \lim_{\Omega \rightarrow \infty} \frac{1}{\Omega} \int_{-\Omega/2}^{\Omega/2} \exp\{j\omega\tau\} \frac{1}{\Omega} d\omega = \frac{1}{2} \left| \exp\{j\omega\tau\} \right|_{\omega=0} = \frac{1}{2}. \end{aligned}$$

Графік кореляційної функції поданий на рис. 3.16.

Приклад 3.7

Білий шум із спектральною щільністю $G(\omega) = G_0 = \text{const}$ проходить через підсилювач із частотною характеристикою у формі гаусової кривої:

$$K(\omega) = K_0 \exp\left\{-\frac{(\omega - \omega_0)^2}{2\beta^2}\right\}.$$

Знайти спектральну щільність і кореляційну функцію вихідного сигналу.

Розв'язування. Спектральна щільність вихідного сигналу:

$$G_1(\omega) = G_0 |K(\omega)|^2 = G_0 K_0^2 \exp\left\{-\frac{(\omega - \omega_0)^2}{\beta^2}\right\}.$$

Таким чином, спектр шумів на виході підсилювача також має форму гаусової кривої.

Для випадку, коли спектр сигналу вузькосмуговий, симетричний щодо частоти і має в цій точці максимум, кореляційна функція відповідно до виразу (3.44):

$$\begin{aligned} K(\tau) &= \left[2 \int_0^{\infty} G_1(\omega_0 - \omega) \cos \omega \tau d\omega \right] \cos \omega_0 \tau = \\ &= 2G_0 K_0^2 \cos \omega_0 \tau \int_0^{\infty} \exp\left\{-\frac{\omega^2}{\beta^2}\right\} \cos \omega \tau d\omega = \sqrt{\pi} G_0 K_0^2 \beta \exp\left\{-\frac{\beta^2 \tau^2}{4}\right\} \cos \omega_0 \tau. \end{aligned}$$

Приклад 3.8

Знайти енергетичний спектр випадкового процесу, що має кореляційну функцію гаусового виду:

$$K(\tau) = K(0) \exp\{-\beta^2 \tau^2\}.$$

Визначити ширину спектра й інтервал кореляції процесу.

Розв'язування. Згідно з виразом (3.37) енергетичний спектр:

$$\begin{aligned} G(\omega) &= \frac{2}{\pi} \int_0^{\infty} K(\tau) \cos \omega \tau d\tau = \frac{2}{\pi} \int_0^{\infty} K(0) \exp\{-\beta^2 \tau^2\} \cos \omega \tau d\tau = \\ &= \frac{K(0)}{\beta \sqrt{\pi}} \exp\left\{-\frac{\omega^2}{4\beta^2}\right\}. \end{aligned}$$

Отже, випадковий процес гаусового виду має і такий самий енергетичний спектр.

Ефективна ширина спектра відповідно до виразу (3.49):

$$\Omega_e = \frac{\int_0^{\infty} G(\omega) d\omega}{G_{\max}} = \frac{\int_0^{\infty} \frac{K(0)}{\beta \sqrt{\pi}} \exp\left\{-\frac{\omega^2}{4\beta^2}\right\} d\omega}{\frac{K(0)}{\beta \sqrt{\pi}}} = \beta \sqrt{\pi}.$$

Інтервал кореляції процесу:

$$\tau_0 = \frac{1}{K(0)} \int_0^{\infty} K(\tau) d\tau = \int_0^{\infty} \exp\{-\beta^2 \tau^2\} d\tau = \frac{\sqrt{\pi}}{2\beta}.$$

Приклад 3.9

Знайти кореляційну функцію і спектральну щільність для стаціонарного випадкового сигналу:

$$x(t) = A \sin(\omega_0 t + \varphi),$$

де A і ω_0 – постійні амплітуда і кутова частота; φ – випадкова початкова фаза, рівномірно розподілена на інтервалі $(-\pi; \pi)$.

Розв'язування. За визначенням кореляційної функції (3.11):

$$K_x(\tau) = m_1[x(t)x(t+\tau)] = \int_{-\pi}^{\pi} A^2 \sin(\omega_0 t + \varphi) \sin(\omega_0 t + \omega_0 \tau + \varphi) \omega_1(\varphi) d\varphi.$$

У зв'язку з тим, що:

$$\omega_1(\varphi) = \frac{1}{2\pi};$$

$$\sin \alpha \sin \beta = \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)],$$

знаходимо:

$$K_x(\tau) = \frac{A^2}{2} \cos \omega_0 \tau.$$

Спектральна щільність відповідно до виразу (3.37):

$$\begin{aligned} G(\omega) &= \frac{1}{\pi} \int_{-\infty}^{\infty} K_x(\tau) \exp\{-j\omega\tau\} d\tau = \frac{A^2}{2\pi} \int_{-\infty}^{\infty} \cos \omega_0 \tau \exp\{-j\omega\tau\} d\tau = \\ &= \frac{A^2}{4\pi} \int_{-\infty}^{\infty} [\exp\{j(\omega_0 - \omega)\tau\} + \exp\{-j(\omega_0 + \omega)\tau\}] d\tau. \end{aligned}$$

У зв'язку з тим, що $\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{j\omega\tau\} d\tau = \delta(\omega)$, отримаємо:

$$G(\omega) = \frac{A^2}{2} [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)].$$

Контрольні питання

1. Які труднощі виникають при використанні n -вимірних щільностей розподілу можливостей випадкового процесу для аналізу передачі інформації?

2. Якими осередненими характеристиками описуються випадкові процеси?
3. Що розуміють під математичним сподіванням і дисперсією випадкового процесу?
4. Що розуміють під середнім квадратичним значенням і середнім квадратичним відхиленням випадкового розміру?
5. Що розуміють під кореляційною функцією випадкового процесу?
6. Що розуміють під стаціонарністю випадкового процесу в широкому і вузькому сенсах; чи завжди процеси, стаціонарні у вузькому сенсі, будуть стаціонарними в широкому сенсі, і навпаки?
7. Які основні властивості кореляційної функції стаціонарного процесу?
8. Що розуміють під часом кореляції випадкового процесу?
9. Які випадкові процеси називаються ергодичними?
10. Чому неможливе безпосереднє застосування класичного гармонічного аналізу до випадкових процесів?
11. Який зв'язок між спектральною щільною і кореляційною функцією?
12. Який випадковий процес називають білим шумом?
13. Яка кореляційна функція білого шуму?
14. Що розуміють під ефективною шириною спектра випадкового процесу?
15. Який зв'язок між інтервалом кореляції і ефективною шириною спектра випадкового процесу?

4.1 Способи квантування сигналів

У багатьох випадках первинні сигнали в інформаційних системах неперервні як за множиною, так і за часом. Передача таких повідомлень ускладнюється у зв'язку з появою апаратурних похибок, похибок від нестабільності параметрів ліній зв'язку тощо. З метою усунення цих похибок роблять перетворення неперервних сигналів у дискретні. Дискретна форма представлення сигналів дає значні переваги при збереженні й опрацюванні інформації. Нарешті, дискретизація сигналів дозволяє використовувати ті самі пристрої (канали зв'язку, пристрої опрацювання інформації) для більшого числа різноманітних сигналів.

Перетворення неперервних сигналів у дискретні називається **квантуванням сигналів**.

Необхідно розрізняти квантування за часом і квантування за рівнем. Квантування за часом полягає в заміні неперервного сигналу $x(t)$ дискретним за часом сигналом $x_k(t)$, значення якого для фіксованих моментів часу $t_0, t_1, t_2, \dots, t_n$ збігаються відповідно з миттєвими значеннями неперервного сигналу (рис. 4.1, а, б). Квантування за рівнем полягає в заміні неперервної множини значень сигналу $x(t)$ множиною дискретних значень. При цьому шкала можливих значень сигналу розбивається на певну кількість рівнів і неперервне значення сигналу замінюється найближчим із цих рівнів (рис. 4.1, в). Часто сигнал піддається квантуванню як за часом, так і за рівнем.

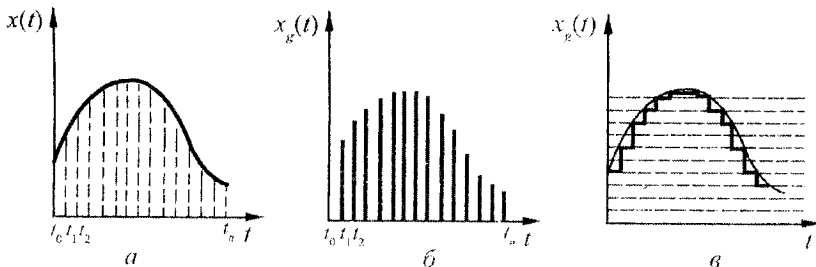


Рис. 4.1

4.2 Квантування за часом

При квантуванні за часом неперервна за аргументом функція $x(t)$, що описує сигнал, перетвориться у функцію $x_k(t)$ дискретного аргументу. Таке перетворення може бути виконане шляхом взяття відліків функції $x(t)$ у визначені дискретні моменти часу $t_0, t_1, t_2, \dots, t_n$. У результаті функція $x(t)$ замінюється сукупністю миттєвих значень $x(t_i)$, $i = 0, 1, 2, \dots, n$.

Тимчасовий інтервал $T_k = t_i - t_{i-1}$ між двома сусідніми фіксованими моментами часу, у яких задається дискретна функція, називається інтервалом тимчасового квантування. Розмір, обернений інтервалу тимчасового квантування називається частотою квантування:

$$f_k = \frac{1}{T_k}. \quad (4.1)$$

За миттєвими значеннями $x(t_i)$ можна відновити вихідну функцію з визначеною точністю. Функцію, отриману в результаті відновлення за відліками $x(t_i)$ називають відтворювальною. Очевидно, із зменшенням інтервалу тимчасового квантування відтворювальна функція буде з більшою точністю відображати вихідну функцію $x(t)$.

Проте, зменшення T_k призводить до збільшення часу зайнятості каналу зв'язку, потребує використання каналів із більшою пропускну здатністю. Тому інтервали T_k потрібно вибирати так, щоб за відліковими значеннями $x(t_i)$ можна було б із заданою точністю одержати вихідну функцію.

Квантування сигналів за часом може бути рівномірним і нерівномірним.

При рівномірному квантуванні функції $x(t)$ інтервал тимчасового квантування T_k постійний. Його розмір вибирається на основі апріорних відомостей про характеристику сигналу. При нерівномірному квантуванні інтервал між відліками змінюється за випадковим законом або з урахуванням зміни характеристик сигналу (адаптивне квантування). Адаптивне квантування реалізувати складніше, ніж рівномірне, проте воно дозволяє значно скоротити кількість надлишкових відліків a , отже, і надлишкову для споживача інформацію.

Відомо декілька критеріїв вибору інтервалів тимчасового квантування. До таких критеріїв належать: частотний критерій В.А. Котельникова, кореляційний критерій Н.А. Железнова, критерій допустимого відхилення відтворювальної функції від вихідної та інші.

4.3 Рівномірне квантування за часом

При рівномірному квантуванні точки відліків t_i функції, що відображає вихідний сигнал, рівномірно розмішені по осі t .

На рис. 4.2 подана спрощена функціональна схема пристрою, що забезпечує рівномірне квантування за часом сигналів.

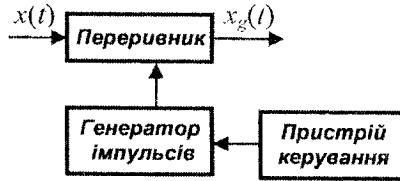


Рис. 4.2

Генератор імпульсів виробляє керуючі сигнали, що надходять із частотою f_k , з якою необхідно квантувати вхідний сигнал $x(t)$. Переривник забезпечує відліки миттєвих значень сигналу $x(t)$ у моменти надходження керуючих сигналів генератора імпульсів. Пристрій керування встановлює частоту генератора імпульсів і визначає час роботи пристрою.

4.3.1 Частотний критерій Котельникова

Даний критерій вибору частоти квантування базується на теоремі Котельникова, що формулюється у такий спосіб: якщо неперервна функція $x(t)$ задовольняє умовам Діріхле (обмежена, кусково-неперервна і має скінченне число екстремумів) і її спектр обмежений деякою частотою f_c , то вона цілком визначається послідовністю своїх

значень у точках, що відстоять на відстані $T_k = \frac{1}{2f_c}$ одна від одной.

Аналітично теорема Котельникова виражається інтерполяційним рядом:

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t) \frac{\sin[\omega_c(t - k\Delta t)]}{\omega_c(t - k\Delta t)}, \quad (4.2)$$

$$\text{де } \Delta t = \frac{\pi}{\omega_c} = \frac{1}{2f_c}.$$

Безпосередньо з виразу (4.2) випливає, що неперервна функція з обмеженим спектром може бути представлена у вигляді суми нескінченно великої кількості членів, кожний із яких є добутком функції виду $\frac{\sin y}{y}$ (функції відліку) і коефіцієнта $x(k\Delta t)$, що визначає значення функції $x(t)$ у моменти відліку.

Функція відліків подана графічно на рис. 4.3, де введене позначення $\tau = t - r\Delta t$. Ця функція в момент часу $t = r\Delta t$ досягає максимального значення і дорівнює одиниці. У момент часу $t = (r + i)\Delta t$, де $i = 1, 2, 3, \dots, \infty$, функція відліків перетворюється в нуль.

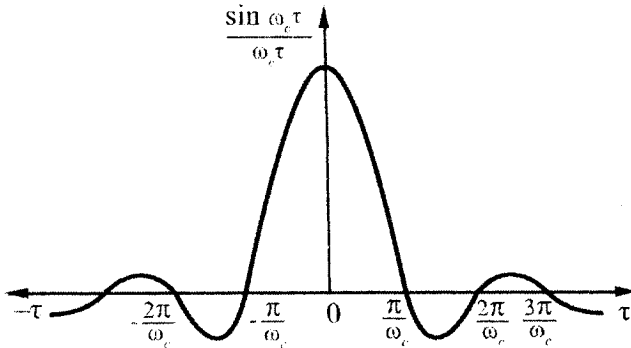


Рис. 4.3

Як відомо, функція виду $\frac{\sin \omega_c \tau}{\omega_c \tau}$ являє собою реакцію ідеального

фільтра нижніх частот із граничною частотою ω_c на δ -функцію. Тоді, якщо через такий фільтр пропустити квантований за часом сигнал із частотою квантування $f_k = 2f_c = \omega_c / \pi$, то підсумовуючи вихідні сигнали фільтра, отримується вихідний неперервний сигнал $x(t)$.

Усе це свідчить про фундаментальне значення теореми Котельникова, тому що вона вказує на можливість відновлення вихідного неперервного сигналу за його відліковими значеннями. Проте при практичному застосуванні теореми Котельникова виникає низка принципових ускладнень.

Як видно з виразу (4.2), для точного відновлення вихідної функції необхідно одержати і підсумувати реакції фільтра на вхідні імпульси на всій осі часу від $+\infty$ до $-\infty$ або хоча б достатньо великої кількості імпульсів до і після апроксимуючої ділянки функції. Практично реалі-

зувати це важко. Далі, функції відліків, що генеруються фільтром низьких частот, повинні мати нескінченну протяжність у часі як для додатних, так і для від'ємних значень t . Такі фільтри фізично нездійсненні. Нарешті, на практиці доводиться мати справу з сигналами, обмеженими в часі, які мають нескінченно широкий спектр, що суперечить основній умові теореми Котельникова.

Проте на практиці ніколи не потрібно ідеально точного відтворення передавального сигналу.

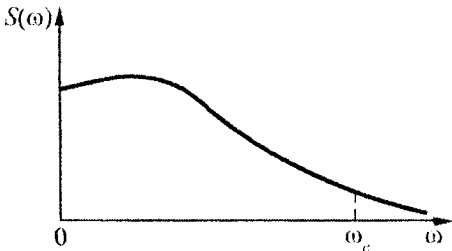


Рис. 4.4

Тому з метою використання теореми Котельникова для квантування сигналів реальний спектр сигналу, що сягає від нуля до нескінченності, умовно обмежують деяким діапазоном частот від нуля до ω_c , в якому зосереджена основна частина енергії спектра (рис. 4.4).

Енергія частини спектра сигналу, що відсікається, характеризує похибку, яка виникає за рахунок обмеження спектра частотою. Цю похибку можна оцінити відношенням енергії, що міститься в частині спектра, яка відсікається, до загальної енергії сигналу:

$$\gamma_{\omega_c} = \frac{\int_0^{\omega_c} |S(j\omega)|^2 d\omega}{\int_0^{\infty} |S(j\omega)|^2 d\omega}. \quad (4.3)$$

Похибку зручніше оцінювати в приведених значеннях. Тоді дисперсія приведеної похибки, що виникає в результаті обмеження частотного спектра сигналу:

$$D_{\omega_c} = \frac{\frac{1}{\pi T_c} \int_0^{\omega_c} |S(j\omega)|^2 d\omega}{(X_{\max} - X_{\min})^2}, \quad (4.4)$$

де $\frac{1}{\pi T_c} \int_0^{\omega_c} |S(j\omega)|^2 d\omega$ – середня потужність частини, що відсі-

кається, частотного спектра; T_c – тривалість сигналу; X_{\max} , X_{\min} – граничні значення функції $x(t)$, яка описує сигнал.

Дисперсію поданої похибки зручніше представити у вигляді:

$$D_{\omega_c} = \frac{\gamma \omega_c P_T}{(X_{\max} - X_{\min})^2}, \quad (4.5)$$

де $P_T = \frac{1}{\pi T_c \omega_c} \int_{-\infty}^{\infty} |S(j\omega)|^2 d\omega = \int_{X_{\min}}^{X_{\max}} x^2 \omega(X) dx$ – середня потуж-

ність сигналу.

Вираз (4.5) за заданим розміром D_{ω_c} і при відомих P_T , X_{\max} , X_{\min} та спектрі функції $x(t)$ дозволяє визначити частоту ω_c , що обмежує спектр.

4.3.2 Кореляційний критерій Железнова

В основі моделі сигналів, запропонованої Железновим, лежать такі припущення:

- сигнал $x(t)$ являє собою недетермінований стаціонарний або квазістаціонарний процес;
- тривалість сигналу T_c скінченна;
- спектр сигналу суцільний і відмінний від нуля на всій осі частот;
- інтервал кореляції сигналу $\tau_0 \ll T_c$;
- функція кореляції сигналу дорівнює нулю поза інтервалом τ_0 ;
- миттєва потужність сигналу обмежена і не може змінюватися стрибком.

Оскільки сигнал передбачається випадковим, обмеженим за часом і має необмежений спектр, то модель сигналу Железнова значно ближче до реальних сигналів, ніж модель Котельникова. Єдиним обмеженням у цій моделі є обмеження, що накладаються на функцію кореляції, тому що передбачається, що $K(\tau) = 0$ при $\tau > \tau_0$.

Железновим було доведено, що стаціонарний або квазістаціонарний сигнал із необмеженим спектром визначається з як завгодно малою помилкою послідовністю його миттєвих значень, що проходять один за одним через інтервали Δt , якщо $\Delta t \leq \tau_0$, а тривалість сигналу $T_c \ll \tau_0$.

Для оцінки похибки, що виникає внаслідок припущення, що поза інтервалом τ_0 функція кореляції дорівнює нулю, можна скористатися виразом (3.46) для визначення ефективної ширини спектра сигналу:

$$\Omega_e = \frac{\pi G(0)}{2G_{\max} \tau_0}.$$

Тоді шукана похибка може бути оцінена відношенням:

$$\gamma_{\Omega_c} = \frac{\int_{-\infty}^{\infty} G(\omega) d\omega}{\int_0^{\Omega_c} G(\omega) d\omega}. \quad (4.6)$$

4.3.3 Критерій допустимого відхилення

Труднощі практичної реалізації методів Котельникова і Желєзна обумовлюють необхідність використання у деяких випадках інших методів вибору необхідної частоти квантування сигналів за часом. У тих випадках, коли закон зміни неперервної функції з визначеною достовірністю відомий, доцільним може виявитися метод, який ґрунтується на заміні вихідної функції апроксимуючою. У загальному випадку вихідна функція апроксимується поліномом, крива якого збігається з кривою функції для заданих дискретних моментів часу. При цьому інтервали тимчасового квантування повинні вибиратися таким чином, щоб у їхніх межах максимальний розмір або середнє квадратичне значення відхилення апроксимуючої функції від вихідної не перевищували допустимих значень.

На практиці найбільш часто застосовується кусково-лінійна або ступінчаста апроксимація.

При кусково-лінійній апроксимації поліномом першого порядку всі точки кривої вихідної функції, що відповідають відліковим моментам часу, з'єднуються відрізками прямих (рис. 4.5).

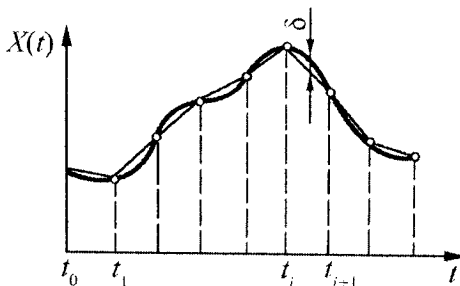


Рис. 4.5

Якщо дотримуватися критерію найбільшого відхилення, то вибір частоти квантування потрібно робити з умови, щоб відхилення апроксимуючої ламаної прямої від вихідної функції на кожному з ін-

тервалів дискретизації не перевищувало б заданого значення. Задача може бути вирішена за допомогою інтерполяційної формули Ньютона, відповідно до якої значення функції для будь-якого моменту часу в середині інтервалу $\Delta t = t_{i+1} - t_i$ визначається виразом:

$$x^*(t) = x(t_i) + a_i [x(t_{i+1}) - x(t_i)],$$

де

$$a_i = \frac{t - t_i}{t_{i+1} - t_i}.$$

Похибка апроксимації δ визначається залишковим членом $L(t)$ інтерполяційної формули:

$$|\delta| = |L(t)| = x^*(t) - x(t).$$

В аналізованому випадку залишковий член виражається наступним чином:

$$L(t) = \frac{1}{2} \frac{d^2 x(t)}{dt^2} (t - t_i)(t - t_{i+1}).$$

Очевидно, максимальне значення похибки апроксимації:

$$|\delta|_{\max} = |L(t)|_{\max} = \frac{1}{2} \left(\frac{d^2 x(t)}{dt^2} \right)_{\max} \left(\frac{t_{i+1} - t_i}{2} \right)^2 = \frac{1}{8} \left(\frac{d^2 x(t)}{dt^2} \right)_{\max} \Delta t^2. \quad (4.7)$$

Отже, задаючись допустимою похибкою апроксимації δ_{\max} , можна визначити інтервал і частоту квантування:

$$\Delta t = \sqrt{\frac{8\delta_{\max}}{\left(\frac{d^2 x(t)}{dt^2} \right)_{\max}}}; \quad f_k = \sqrt{\frac{\left(\frac{d^2 x(t)}{dt^2} \right)_{\max}}{8\delta_{\max}}}. \quad (4.8)$$

При ступінчастій апроксимації поліномом нульового порядку про значення функції в будь-який момент часу t' в інтервалі $t_i \leq t' \leq t_{i+1}$ судять за значенням функції $x(t)$, зафіксованої в точці t_i . Залишковий член інтерполяційної формули визначається в цьому випадку виразом:

$$L(t) = \frac{dx(t)}{dt} (t - t_i).$$

Максимальне значення похибки апроксимації:

$$|\delta|_{\max} = |L(t)|_{\max} = \left(\frac{dx(t)}{dt} \right)_{\max} \Delta t, \quad (4.9)$$

звідки необхідні значення інтервалу і частоти квантування визначаються виразами:

$$\Delta t = \frac{|\delta|_{\max}}{\left(\frac{dx(t)}{dt} \right)_{\max}}; \quad f_k = \frac{\left(\frac{dx(t)}{dt} \right)_{\max}}{|\delta|_{\max}}. \quad (4.10)$$

Критерій найбільшого відхилення відноситься до детермінованого випадку, коли передбачається, що закон зміни вихідної функції $x(t)$ відомий. Оскільки реальні сигнали практично завжди мають випадковий характер, то цей критерій є дещо неточним. При випадковому характері вихідної функції доцільно користуватися критерієм середнього квадратичного відхилення.

Розглянемо кусково-лінійну інтерполяцію для випадку квантування випадкового стаціонарного ергодичного процесу $x(t)$.

Апроксимуюча функція для i -го інтервалу квантування має вигляд:

$$x^*(t) \approx x(t_i) + a_i [x(t_{i+1}) - x(t_i)],$$

де

$$a_i = \frac{t - t_i}{\Delta t}.$$

Поточне значення похибки квантування в момент часу t :

$$\delta(t) = x^*(t) - x(t_i + t) = a_i x(t_{i+1}) + (1 - a_i)x(t_i) - x(t_i + t). \quad (4.11)$$

На сьогодні існує значна кількість способів адаптивного квантування. Найбільш ефективними і простими серед них є способи, засновані на порівнянні сигналу $x(t)$ і апроксимуючої функції, сформованої в процесі опрацювання сигналу $x(t)$.

Адаптивне квантування може бути реалізоване шляхом зміни типу і порядку апроксимуючої функції при постійному інтервалі тимчасового квантування або шляхом зміни інтервалу квантування при незмінній апроксимуючій функції. На практиці найбільше застосування

одержали способи з адаптацією за розміром інтервалу квантування, що використовують поліном нульового і першого порядку.

При адаптивному квантуванні з використанням поліномів нульового порядку проводиться порівняння поточного значення сигналу $x(t)$ із значенням попередньої вибірки $x(t_i)$ сигналу. Модуль різниці $\Delta x(t) = x(t) - x(t_i)$ порівнюється з допустимим значенням похибки δ_{\max} . Момент такого відліку t_{i+1} сигналу $x(t)$ визначається виконанням умови:

$$|\Delta x(t)| = \delta_{\max}.$$

Один із способів адаптивного квантування з використанням поліномів першого порядку полягає в тому, що після чергового відліку функції $x(t)$ у момент часу t_i формується функція:

$$y(t) = x(t_i) + \left. \frac{dx(t)}{dt} \right|_{t=t_i} t. \quad (4.12)$$

Потім обчислюється різниця $\Delta x(t) = y(t) - x(t_i)$ і момент такого відліку t_{i+1} визначається при виконанні умови:

$$|\Delta x(t)| = \delta_{\max}.$$

4.4 Квантування за рівнем

При квантуванні за рівнем неперервні множини значень функції $x(t)$ замінюються множинами дискретних значень. Для цього в діапазоні неперервних значень функції $x(t)$ вибирається скінченне число дискретних значень цієї функції (дискретних рівнів) і в процесі квантування значення функції $x(t)$ у кожний момент часу замінюється найближчим дискретним значенням. У результаті квантування утворюється ступінчаста функція $x_q(t)$.

Квантування за рівнем практично може здійснюватися двома способами. При першому способі миттєве значення функції $x(t)$ замінюється меншим дискретним значенням (рис. 4.6, а). При другому способі квантування миттєве значення функції замінюється найближчим меншим або більшим дискретним значенням залежно від того, яке з цих значень ближче до миттєвого значення функції. У цьому випадку перехід ступінчастої функції з одної сходинки на іншу відбувається в ті моменти, коли початкова неперервна функція $x(t)$ перетинає середину між відповідними сусідніми дискретними рівнями (рис. 4.6, б).

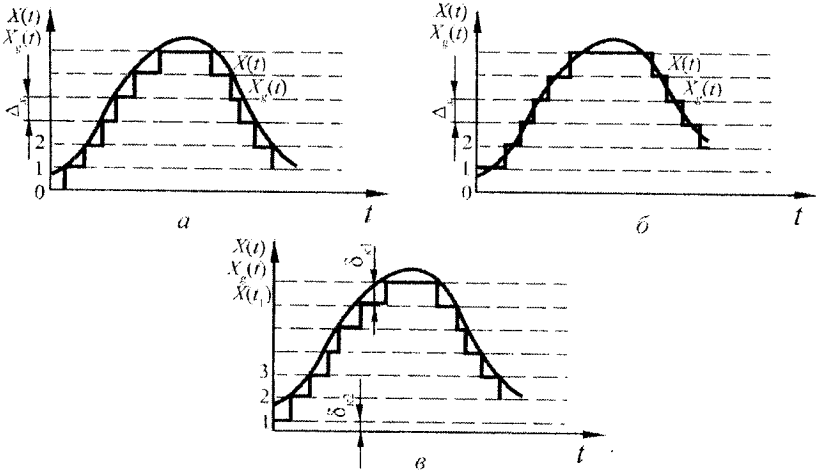


Рис. 4.6

Практично в процесі квантування можуть мати місце випадки, коли дискретні рівні не фіксовані щодо нульового рівня $x(t)$. У таких випадках квантування здійснюється шляхом заміни значення $x(t)$ найближчим меншим дискретним значенням (рис. 4.6, в).

Відстань між сусідніми дискретними рівнями називається **інтервалом** або **кроком квантування** Δx .

Розрізняють рівномірне квантування за рівнем, при якому крок квантування постійний, і нерівномірне, коли крок квантування непостійний. На практиці переважно застосування одержало рівномірне квантування у зв'язку з простотою його технічної реалізації.

Квантування за рівнем широко використовується в системах передачі інформації, інформаційно-вимірювальних системах, при автоматичному керуванні, обробці даних за допомогою ЦЕОМ тощо.

Оскільки в процесі квантування за рівнем значення сигналу відображається рівнем квантування, то при передачі, зберіганні або обробці значення сигналу може бути зображене номером-кодом відповідного дискретного рівня.

4.5 Рівномірне квантування за рівнем

Квантування сигналу за рівнем здійснюється за допомогою нелінійного елемента – квантизатора (рис. 4.7, а), амплітудна характеристика якого визначається способом квантування. На рис. 4.7, б показана

характеристика квантизатора при квантуванні способом заміни випадкової величини $x(t)$ найближчим меншим дискретним значенням. Характеристика на рис. 4.7, *в* відповідає способу квантування шляхом заміни $x(t)$ найближчим меншим або більшим дискретним рівнем. У тому випадку, коли дискретні рівні не фіксовані відносно нульового рівня $x(t)$, характеристика квантизатора може зміщуватися випадково відносно характеристики, яка представлена на рис. 4.7, *б*, у вертикальному напрямку в межах від 0 до Δ_x .

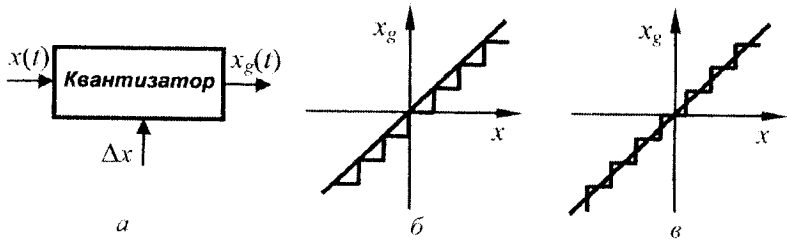


Рис. 4.7

Внаслідок квантування функції за рівнем з'являються методичні похибки, тому що дійсне миттєве значення функції замінюється дискретним значенням. Ця похибка, яка отримала назву похибки квантування за рівнем або шумом квантування, має випадковий характер. Абсолютне її значення в кожен момент часу визначається різницею між квантованим значенням $x_g(t)$ і дійсним миттєвим значенням $x(t)$ функції.

$$\delta_x(t) = x_g(t) - x(t).$$

На рис. 4.8, *а*, *б* показаний характер зміни абсолютного значення похибки квантування для тих способів квантування, які наведені на рис. 4.7, *б*, *в* відповідно. При квантуванні методом заміни дійсного миттєвого значення функції $x(t)$ найближчим меншим дискретним значенням абсолютна похибка квантування буде завжди від'ємною та знаходитиметься в діапазоні від 0 до Δ_x .

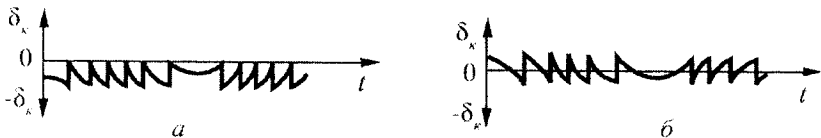


Рис. 4.8

Закон розподілу цієї похибки залежить від закону розподілу $x(t)$.

Нехай функція $x(t)$ підпорядковується визначеному закону розподілу $\omega(x)$. Розіб'ємо діапазон зміни функції $x(t)$ на інтервали величиною Δ_x (рис. 4.9).

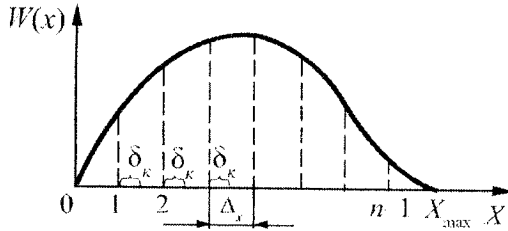


Рис. 4.9

Нехай δ_k – деякі відхилення дійсних значень функції $x(t)$ від найближчого меншого дійсного значення похибки квантування. Ймовірність появи помилки δ_k може бути визначена як ймовірність $P(\delta_k)$ попадання значення функції $x(t)$ в проміжок δ_k будь-якого з інтервалів квантування:

$$P(\delta_k) = \sum_{i=0}^{n-1} \int_{i\Delta_x}^{i\Delta_x + \delta_k} \omega(x) dx.$$

Диференціюючи обидві частини даного рівняння за δ_k , знайдемо диференціальний закон розподілу похибки квантування:

$$\frac{dP(\delta_k)}{d\delta_k} = \omega(\delta_k) = \sum_{i=0}^{n-1} \omega(i\Delta_x + \delta_k).$$

Помноживши обидві частини рівності на Δ_x , отримаємо:

$$\omega(\delta_k) \Delta_x = \sum_{i=0}^{n-1} \omega(i\Delta_x + \delta_k) \Delta_x. \quad (4.13)$$

Праву частину останнього рівняння можна розглядати як приблизне зображення площі фігури, яка утворена між віссю і кривою. Це наближення буде тим точнішим, чим більшою буде кількість проміжків квантування. Звідси випливає, що в проміжку при $n \rightarrow \infty$ або $\Delta_x \rightarrow 0$ вираз приводиться до вигляду:

$$\omega(\delta_k) \Delta_x = \int_0^{x_{\max}} \omega(x) dx. \quad (4.14)$$

Оскільки права частина виражає ймовірність знаходження функції в проміжку від 0 до x_{\max} і, відповідно, дорівнює одиниці, то щільність ймовірності розподілення похибки квантування буде з достатнім наближенням визначатися виразом:

$$\omega(\delta_k) = \frac{1}{\Delta_x}. \quad (4.15)$$

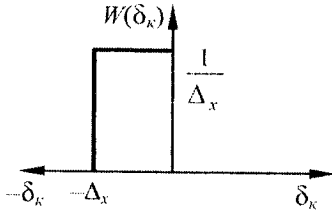


Рис. 4.10

$$\omega(\delta_k) = \begin{cases} \frac{1}{\Delta_x}, & \text{при } -\Delta_x \leq \delta_k \leq 0; \\ 0, & \text{при } \delta_k > 0 \text{ і } \delta_k < -\Delta_x. \end{cases}$$

Математичне сподівання похибки квантування:

$$a(\delta_k) = \int_{\Delta_x}^0 \delta_k \omega(\delta_k) d\delta_k = \frac{1}{\Delta_x} \int_{-\Delta_x}^0 \delta_k d\delta_k = -\frac{\Delta_x}{2}. \quad (4.16)$$

Дисперсія похибки квантування:

$$D(\delta_k) = \int_{-\Delta_x}^0 [\delta_k - a(\delta_k)]^2 \omega(\delta_k) d\delta_k = \int_{-\Delta_x}^0 \left[\delta_k + \frac{\Delta_x}{2} \right]^2 \frac{1}{\Delta_x} d\delta_k = \frac{\Delta_x^2}{12}. \quad (4.17)$$

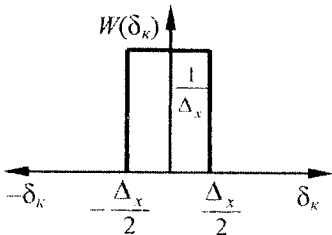


Рис. 4.11

Таким чином, можна вважати, що при достатньо великому числі рівнів квантування її похибка підпорядковується рівномірному закону розподілу ймовірності (див. рис. 4.10).

Згідно з ним можна записати, що:

При квантуванні методом заміни дійсного миттєвого значення функції найближчим меншим або більшим дискретним значенням похибка квантування також відповідає закону розподілу рівної ймовірності, але змінюється в межах від $-\Delta_x/2$ до $+\Delta_x/2$ (рис. 4.11).

У даному випадку справедливо, що:

$$\omega(\delta_\kappa) = \begin{cases} \frac{1}{\Delta_x}, & \text{при } -\frac{\Delta_x}{2} \leq \delta_\kappa \leq \frac{\Delta_x}{2}; \\ 0, & \text{при } \delta_\kappa > \frac{\Delta_x}{2} \text{ або } \delta_\kappa < -\frac{\Delta_x}{2}. \end{cases}$$

Математичне сподівання похибки квантування: $a(\delta_\kappa) = 0$.

Дисперсія похибки:

$$D(\delta_\kappa) = \int_{-\frac{\Delta_x}{2}}^{\frac{\Delta_x}{2}} \delta_\kappa^2 \omega(\delta_\kappa) d\delta_\kappa = \frac{\Delta_x}{12}.$$

При відсутності фіксації дискретних рівнів, квантування виконується практично шляхом заміни миттєвого значення функції найближчим меншим дискретним рівнем і нижній дискретний рівень приймається за перший. Внаслідок цього будуть мати місце дві похибки квантування: $\delta_{\kappa 1}$, яка змінюється в межах від 0 до $-\Delta_x$, і $\delta_{\kappa 2}$, яка змінюється в межах від 0 до $+\Delta_x$. Ці похибки незалежні і відповідають законам рівномірного розподілу ймовірності (рис. 4.12, а, б).

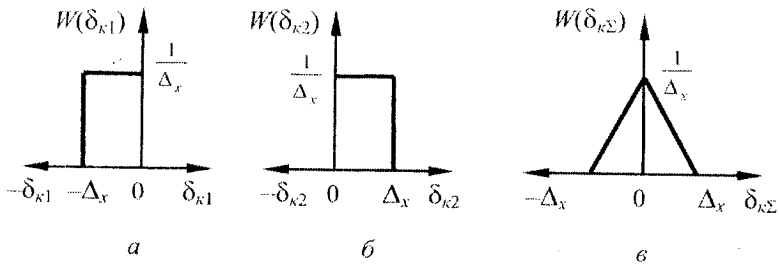


Рис. 4.12

Сумарна похибка квантування

$$\delta_{\kappa\Sigma} = \delta_{\kappa 1} + \delta_{\kappa 2}$$

змінюється в межах від $-\Delta_x$ до $+\Delta_x$. Закон розподілу сумарної похибки є композицією законів розподілу похибок $\delta_{\kappa 1}$ і $\delta_{\kappa 2}$ і виражається розподілом Сімпсона (рис. 4.12, в).

Аналітично цей розподіл виражається наступним чином:

$$\omega(\delta_{k\Sigma}) = \begin{cases} 1 + \frac{\delta_{k\Sigma}}{\Delta_x}, & \text{при } -\Delta_x \leq \delta_{k\Sigma} \leq 0; \\ 1 - \frac{\delta_{k\Sigma}}{\Delta_x}, & \text{при } 0 \leq \delta_{k\Sigma} \leq \Delta_x; \\ 0, & \text{при } \delta_{k\Sigma} > \Delta_x \text{ і } \delta_{k\Sigma} < -\Delta_x. \end{cases}$$

Математичне сподівання сумарної похибки квантування:

$$a(\delta_{k\Sigma}) = 0.$$

Дисперсія похибки:

$$D(\delta_{k\Sigma}) = \int_{-\Delta_x}^0 \delta_{k\Sigma}^2 \frac{\left(1 + \frac{\delta_{k\Sigma}}{\Delta_x}\right)}{\Delta_x} d\delta_{k\Sigma} + \int_0^{\Delta_x} \frac{1 - \frac{\delta_{k\Sigma}}{\Delta_x}}{\Delta_x} d\delta_{k\Sigma} = \frac{\Delta_x^2}{6}. \quad (4.18)$$

Таким чином, у випадку відсутності фіксації дискретних рівнів відповідно початкового рівня функції, яка квантується, середньоквадратична похибка квантування збільшується в $\sqrt{2}$ разів.

4.6 Нерівномірне квантування за рівнем

Нерівномірне квантування може бути використане в тих випадках, коли ймовірність розподілу значення функції $x(t)$ за шкалою рівнів неоднакова. Основна ціль такого квантування – зменшення осередненої за дисперсією похибки квантування.

Ідея нерівномірного квантування полягає в тому, що значення функції $x(t)$, ймовірність виникнення яких велика, передаються з меншою похибкою квантування, а малоймовірні значення – з більшою похибкою квантування.

Для реалізації такого квантування первинна функція піддається нелінійному перетворенню $y(t)$. Суть нелінійного перетворення показана на рис. 4.13, де $\omega(x)$ – крива розподілу щільності ймовірності значення функції $x(t)$.

Величина y квантується рівномірно. При цьому приріст функцій Δ_y і Δ_x зв'язаний співвідношенням $\Delta_y = \Delta_x y'(x)$.

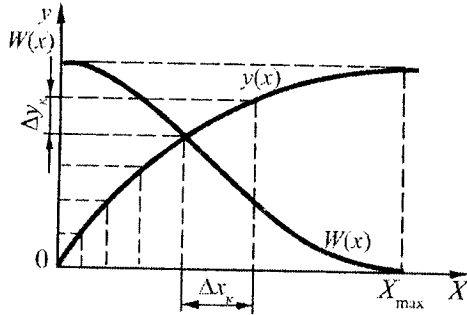


Рис. 4.13

При великій кількості рівнів квантування таке співвідношення справедливе і для похибок квантування, тобто:

$$\delta_{y_k} = \delta_{x_k} y'(x).$$

Для дисперсій похибок одержимо:

$$D_{y_k} = D_{x_k} [y'(x)]^2.$$

Тоді осереднена за параметром дисперсія похибки квантування функції $x(t)$:

$$D_{x_k} = \int_{x_{\min}}^{x_{\max}} \omega(x) \frac{D_{y_k}}{[y'(x)]^2} dx = D_{y_k} \int_{x_{\min}}^{x_{\max}} \frac{\omega(x)}{[y'(x)]^2} dx.$$

У даному випадку множник D_{y_k} винесений за знак інтеграла, тому що квантування функції $y(t)$ рівномірне і дисперсія похибки квантування постійна в межах всього діапазону зміни величин x та y .

Виграш в похибці квантування оцінюється співвідношенням:

$$B = \frac{D_{y_k}}{D_{x_k}} = \frac{1}{\int_{x_{\min}}^{x_{\max}} \frac{\omega(x)}{[y'(x)]^2} dx}. \quad (4.20)$$

Для отримання максимального виграшу в похибці квантування необхідно знайти таку функцію $y'(x)$, яка мінімізувала б інтеграл:

$$\int_{x_{\min}}^{x_{\max}} \frac{\omega(x)}{[y'(x)]^2} dx.$$

Задача відноситься до типу варіаційних. Розв'язування цієї задачі зводиться до розв'язування диференціального рівняння:

$$\frac{\partial F}{\partial y} - \frac{\partial}{\partial y'} \left(\frac{\partial F}{\partial y'} \right) = 0, \quad (4.21)$$

$$\text{де } F = \frac{\omega(x)}{[y'(x)]^2}.$$

$$\text{У нашому випадку } \frac{\partial F}{\partial y} = 0; \quad \frac{\partial F}{\partial y'} = -2\omega(x)[y'(x)]^{-3}.$$

Тоді рівняння (4.21) набуде вигляду:

$$\frac{\partial \omega(x)}{\partial x} [y'(x)]^{-3} - 3\omega(x)[y'(x)]^{-4} y''(x) = 0.$$

Інтегруючи це рівняння розділенням змінних, одержимо:

$$y'_{opt}(x) = C_0 \sqrt{\omega(x)}. \quad (4.22)$$

Тоді оптимальне нелінійне перетворення за критерієм мінімуму дисперсії похибки дорівнює:

$$y_{opt}(x) = C_0 \int_{x_{\min}}^{x_{\max}} \sqrt[3]{\omega(x)} dx + C_1. \quad (4.23)$$

Коефіцієнти C_0 і C_1 визначаються з умови $y = y_{\max}$ або $y = y_{\min}$, при $x = x_{\min}$ або $x = x_{\max}$ відповідно. Отже $C_1 = y_{\min}$ та

$$C_0 = \frac{y_{\max} - y_{\min}}{\int_{x_{\min}}^{x_{\max}} \sqrt[3]{\omega(x)} dx}.$$

4.7 Приклади

Приклад 4.1

Визначити похибку квантування в часі сигналу скінченної довжини. Відомо наступне:

1) сигнал $x(t)$ з рівною ймовірністю може приймати будь-які значення в межах від 0 до x_{\max} ;

2) частота квантування ω_k вибрана таким чином, що відповідна величина площі енергетичного спектра сигналу, який знаходиться в межах частот від $\omega_0 = \frac{1}{2}\omega_k$ до $\omega = \infty$, дорівнює 5 %.

Розв'язування. Для рішення задачі скористаємося виразом (4.5). За умовою відносна величина площі ділянки спектра сигналу, що відсікається, дорівнює $\gamma_{\omega_k} = 0,05$. Середня потужність сигналу:

$$P_T = \int_0^{x_{\max}} x^2 \omega(X) dx.$$

Оскільки сигнал за шкалою рівнів розподілений з рівною ймовірністю, то $\omega(X) = \frac{1}{x_{\max}}$. Тоді:

$$P_T = \int_0^{x_{\max}} x^2 \frac{1}{x_{\max}} dx = \frac{x_{\max}^2}{3}.$$

У такому випадку шукана похибка квантування:

$$D_{\omega_c} = \frac{0,05 x_{\max}^2}{3 x_{\max}^2} 100\% = 1,7\%.$$

Приклад 4.2

Знайти частоту квантування в часі експоненціального сигналу $x(t) = A_0 \exp\{-\alpha t\}$, $t \geq 0$, якщо відносна величина ділянки площі енергетичного спектра, яка відсікається, дорівнює 5 %; $\alpha = 10 \text{ c}^{-1}$.

Розв'язування. З виразу (4.3) маємо:

$$\gamma_{\omega_c} = \frac{\int_0^{\omega_c} |S(j\omega)|^2 d\omega}{\int_0^{\infty} |S(j\omega)|^2 d\omega} = 0,05. \quad (4.24)$$

Спектральна щільність сигналу:

$$S(j\omega) = \int_0^{\infty} x(t) \exp\{-j\omega t\} dt = \int_0^{\infty} A_0 \exp\{-(\alpha + j\omega t)\} dt = \frac{A_0}{\alpha + j\omega}.$$

Модуль спектральної щільності:

$$|S(j\omega)| = \frac{A_0}{\sqrt{\alpha^2 + \omega^2}}.$$

Згідно з (2.17) енергія сигналу:

$$W_0 = \frac{1}{\pi} \int_0^{\infty} |S(j\omega)|^2 d\omega = \frac{1}{\pi} \int_0^{\infty} \frac{A_0^2}{\alpha^2 + \omega^2} d\omega = \frac{A_0^2}{2\alpha}.$$

Енергія сигналу, зосереджена в діапазоні частот від $\omega = \omega_c$ до $\omega = \infty$, дорівнює:

$$W_{\omega_c} = \frac{1}{\pi} \int_{\omega_c}^{\infty} |S(j\omega)|^2 d\omega = \frac{1}{\pi} \int_{\omega_c}^{\infty} \frac{A_0^2}{\alpha^2 + \omega^2} d\omega = \frac{A_0^2}{\pi\alpha} \left(\frac{\pi}{2} - \arctg \frac{\omega_c}{\alpha} \right).$$

Підставляючи знайдені значення інтервалів у (4.24) і виконуючи перетворення, одержимо:

$$\arctg \frac{\omega_c}{\alpha} = \frac{\pi}{2} - \frac{\pi}{2} \gamma_{\omega_c},$$

звідки

$$\operatorname{tg} \left(\frac{\pi}{2} - \frac{\pi}{2} \gamma_{\omega_c} \right) = \frac{\omega_c}{\alpha}$$

або

$$\operatorname{ctg} \left(\frac{\pi}{2} \gamma_{\omega_c} \right) = \frac{\omega_c}{\alpha}.$$

Використовуючи формулу розкладу:

$$\operatorname{ctg} x = \frac{1}{x} - \left[\frac{x}{3} + \frac{x^3}{45} + \dots \right],$$

і для малих значень x обмежуючись першим членом розкладу, одержимо:

$$\frac{2}{\pi \gamma_{\omega_c}} = \frac{\omega_c}{\alpha}$$

або остаточно:

$$\omega_c = \frac{2\alpha}{\pi \gamma_{\omega_c}} = \frac{2 \cdot 10}{\pi \cdot 0,05} = 127,5 \text{ с}^{-1}.$$

Необхідна частота квантування сигналу відповідно до теореми Котельникова:

$$\omega_x = 2\omega_c = 225 \text{ с}^{-1}.$$

Приклад 4.3

Визначити частоту, з якою слід квантувати у часі стаціонарний випадковий процес із кореляційною функцією $K_x(\tau) = K_x(0)\exp\{-0,1|\tau|\}$. Розрахунки зробити, виходячи з критерію середньоквадратичного відхилення зі ступінчастою апроксимацією. Допустиме значення дисперсії похибки квантування 1 %.

Розв'язування. З виразу (4.11) визначаємо відносно значення дисперсії похибки квантування:

$$\gamma_D = \frac{2[K_x(0) - K_x(\Delta t)]}{K_x(0)} = 2[1 - \exp\{-0,1|\Delta t|\}].$$

Тоді інтервал тимчасового квантування:

$$\Delta_t = 10 \ln \frac{1}{1 - 0,5\gamma_D} = 10 \ln \frac{1}{1 - 0,5 \cdot 0,01} = 0,05 \text{ с}.$$

Частота квантування:

$$f_k = \frac{1}{\Delta t} = 20 \text{ Гц}.$$

Приклад 4.4

Переданий по каналу зв'язку сигнал квантується за рівнем способом заміни його миттєвих значень найближчим меншим квантованим рівнем. Визначити необхідну кількість рівнів квантування сигналу за умови, що наведена середньоквадратична похибка квантування не перевищує 0,3 %.

Розв'язування. При заданому способі квантування похибка квантування від'ємна і може приймати значення від 0 до $-\Delta_x$, де Δ_x – крок квантування.

Середнє квадратичне значення похибки квантування відповідно до виразу (4.17):

$$\sigma_k = \frac{\Delta_x}{2\sqrt{3}}.$$

Приведена середня квадратична похибка:

$$\gamma_{\kappa} \% = \frac{\sigma_{\kappa}}{x_{\max}} \cdot 100 = \frac{\Delta_x}{2\sqrt{3}N\Delta_x} 100 = \frac{1}{2\sqrt{3}N} 100,$$

де N – кількість інтервалів, на які розбивається динамічний діапазон сигналу при квантуванні.

Оскільки кількість рівнів квантування M на одиницю перевищує кількість інтервалів квантування, то:

$$M = N + 1 = \frac{50}{\sqrt{3}\gamma_{\kappa} \%} + 1 = \frac{50}{\sqrt{3} \cdot 0,1} + 1 = 97.$$

Приклад 4.5

У процесі передачі сигналу, що представляє напругу постійного струму, величина якого змінюється в межах від 0 до 100 В, відбувається послідовне перетворення його в тимчасовий t_x і в число-імпульсний код. Перетворення напруги U_x в часовий інтервал t_x здійснюється за допомогою час-імпульсного перетворювача ЧІП, наступне перетворення в число-імпульсний код здійснюється шляхом заповнення тимчасового інтервалу їх імпульсами генератора еталонної частоти (рис. 4.14).

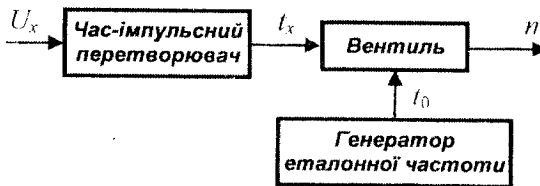


Рис. 4.14

Визначити сумарну відносну середню квадратичну похибку перетворення напруги в код, якщо:

– коефіцієнт перетворення час-імпульсного перетворювача $K_{\text{ЧІП}} = 10^4$ С/В;

– номінальне значення частоти еталонного генератора $f_0^{\text{н}} = 100$ кГц;

– відносне значення середньоквадратичної похибки час-імпульсного перетворювача $\gamma_{\text{ЧІП}} = 0,1$ %;

– відносне значення середньої квадратичної похибки еталонного генератора $\gamma_{f_0} = 0,05$ %.

Розв'язування. Кількість імпульсів n еталонного генератора, що укладаються в тимчасовому t_x і надходять на вихід вентиля (рис. 4.14):

$$n \cong t_x f_0 = K_{\text{ШИ}} f_0 U_x.$$

Загальна середня квадратична похибка перетворювача, приведена до виходу пристрою, визначається залежністю:

$$\sigma_{\Sigma} = \sqrt{\left(\frac{\partial n}{\partial t_x} \sigma_{t_x}\right)^2 + \left(\frac{\partial n}{\partial f_0} \sigma_{f_0}\right)^2} + \sigma_k^2 = \sqrt{(f_0 \sigma_{\text{ШИ}})^2 + (t_x \sigma_{f_0})^2} + \sigma_k^2.$$

де $\sigma_{\text{ШИ}}$, σ_{f_0} , σ_k – середні квадратичні похибки час-імпульсного перетворювача, еталонного генератора і перетворювача тимчасового інтервалу в число імпульсів відповідно.

Перетворення тимчасового інтервалу t_x у число імпульсів зводиться по суті до квантування за рівнем сигналу, поданого тимчасовим інтервалом t_x . При цьому крок квантування дорівнює $\Delta_x = t_0 = \frac{1}{f_0}$.

Оскільки немає синхронізації між посланнями імпульсів еталонного генератора і початком тимчасового інтервалу t_x , то маємо по суті випадок квантування, коли дискретні рівні не фіксовані щодо нульового рівня вхідного сигналу (див. рис. 4.6, в). При цьому похибка квантування розподілена за законом Сімпсона, а середнє квадратичне значення цієї похибки: $\sigma_k = \frac{\Delta_x}{\sqrt{6}} = \frac{1}{\sqrt{6}}$, тому що крок квантування Δ_x ,

приведений до виходу пристрою, дорівнює одиниці (одному імпульсу). Тоді загальна середньоквадратична похибка перетворювача:

$$\sigma_{\Sigma} = \sqrt{(f_0 \sigma_{\text{ШИ}})^2 + (t_x \sigma_{f_0})^2} + \frac{1}{6}.$$

Переходячи до відносних величин, одержимо:

$$\begin{aligned} \gamma_{\Sigma} \% = \frac{\sigma_{\Sigma}}{n} \cdot 100 &= \sqrt{\left(\frac{\sigma_{\text{ШИ}}}{t_x}\right)^2 + \left(\frac{\sigma_{f_0}}{f_0}\right)^2} + \frac{1}{6} \left(\frac{100}{K_{\text{ШИ}} f_0 U_x}\right)^2 = \\ &= \sqrt{(0,1)^2 + (0,05)^2} + \frac{1}{6} \left(\frac{100}{10^{-4} 10^5 10^2}\right)^2 = 0,12 \%. \end{aligned}$$

Контрольні питання

1. У чому сутність квантування безперервного сигналу за часом і рівнем?
2. Яка мета квантування сигналів?
3. Сформулюйте критерій Котельникова. У чому полягає його фундаментальне значення для теорії та техніки передачі повідомлень?
4. У чому полягають труднощі практичної реалізації критерію Котельникова?
5. Що таке функція відліків і який її вигляд?
6. Сформулюйте критерій Железнова і зіставте його з критерієм Котельникова.
7. У чому сутність критерію найбільшого відхилення з кусково-лінійною апроксимацією вихідної функції?
8. У чому сутність критерію найбільшого відхилення зі ступінчастою апроксимацією вихідної функції?
9. У чому полягає обмеженість практичного використання критеріїв найбільшого відхилення?
10. У чому сутність критерію середнього квадратичного відхилення?
11. У чому сутність адаптивного квантування за часом і з якою метою воно використовується?
12. Які відомі способи квантування сигналів за рівнем?
13. Який характер похибок квантування сигналів за рівнем?
14. У чому сутність нерівномірного квантування за рівнем і в яких випадках використовують цей вид квантування?

МОДУЛЯЦІЯ І ДЕМОДУЛЯЦІЯ

5

При створенні систем передачі інформації в більшості випадків виявляється, що спектр вихідного сигналу, що підлягає передачі, зосереджений аж ніяк не на тих частотах, які ефективно пропускає наявний канал зв'язку. Крім того, дуже часто необхідно в тому самому каналі зв'язку передавати кілька сигналів одночасно. Одним зі способів розв'язання цієї задачі є використання частотного поділу каналів, при якому різні сигнали займають смуги частот, що не перекриваються.

У багатьох випадках потрібно, щоб переданий сигнал був вузько-смуговим. Це означає, що ефективна ширина спектра набагато менше його центральної частоти:

$$\Delta f \ll f_0.$$

Перераховані причини приводять до необхідності такої трансформації вихідного сигналу, щоб вимоги, які висунуті до смуги частот, яку займає сигнал, були виконані, а сам вихідний сигнал можна було потім відновити.

Розв'язання зазначеної проблеми досягається при використанні модуляції, сутність якої полягає в наступному. Формується деяке коливання (найчастіше гармонійне), яке називають *несучим коливанням*, або просто *несучою*, і який-небудь із параметрів цього коливання змінюється в часі пропорційно вихідному сигналу. Вихідний сигнал називають *модулюючим* сигналом, а результуюче коливання із параметрами, що змінюються в часі – *модульованим* сигналом. Зворотний процес – виділення модулюючого сигналу із модульованого коливання – називається *демодуляцією*.

Запишемо гармонічний сигнал у загальному вигляді:

$$S(t) = A \cos(\omega_0 t + \varphi_0).$$

Такий сигнал має три параметри: амплітуду A , частоту ω_0 і початкову фазу φ_0 . Кожний з них можна зв'язати із модулюючим сигналом, одержавши у такий спосіб три основних види модуляції: *амплітудну*, *частотну* і *фазову*. Частотна і фазова модуляції тісно взаємопов'язані, оскільки обидві вони впливають на аргумент функції косинус. Тому ці два види модуляції мають загальну назву – *кутова модуляція*.

У сучасних системах передачі цифрової інформації також одержала поширення **квадратурна модуляція**, при якій одночасно змінюються амплітуда і фаза сигналу.

5.1 Амплітудна модуляція

При амплітудній модуляції (АМ) відповідно до модулюючого сигналу змінюється амплітуда несучого коливання:

$$S_{AM}(t) = A(t) \cos(\omega_0 t + \varphi_0).$$

Однак, якщо амплітуду просто зробити прямо пропорційною модулюючому сигналу, можливе виникнення наступної проблеми. Як правило модулюючий сигнал є двополярним (знакозмінним). Розглянемо, наприклад, такий сигнал (рис. 5.1, а):

$$S_M(t) = 3 \cos(2\pi t) - \sin\left(6\pi t + \frac{\pi}{4}\right).$$

Якщо ми безпосередньо використовуємо його як амплітудну функцію $A(t)$, вийде наступне (рис. 5.1, б). Множення двополярного модулюючого сигналу на несуче коливання дає неправильну амплітудну обвідну.

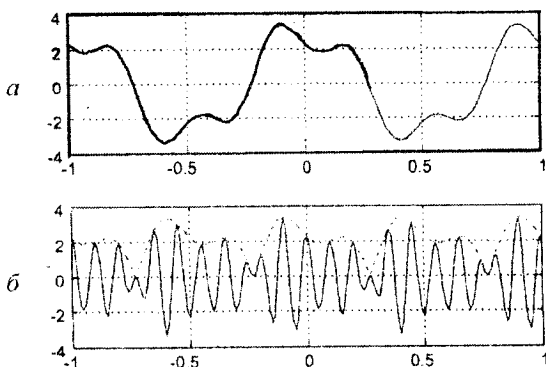


Рис. 5.1

З рис. 5.1, б видно, що амплітудна обвідна, котра буде виділена в процесі демодуляції, у даному випадку виявляється неправильно – вона відповідає **модулю вихідного сигналу**.

Тому при реалізації АМ до модулюючого сигналу попередньо додають постійну складову, щоб зробити його однополярним:

$$A(t) = A_0 + ks_M(t).$$

Для нашого прикладу досить буде постійної складової, рівної чотирьом (рис. 5.2). Додавання постійної складової робить модулюючий сигнал однополярним.

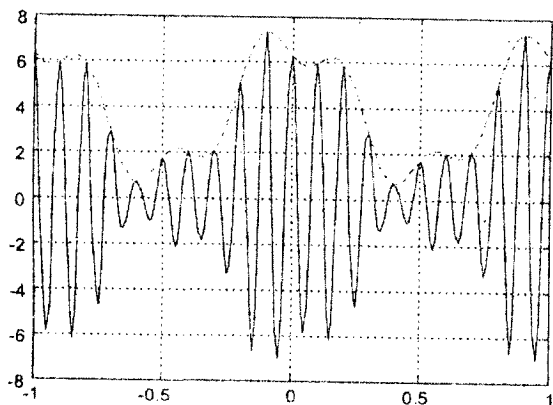


Рис. 5.2

Тепер форма амплітудної обвідної відповідає модулюючому сигналу з точністю до постійної складової, котра легко може бути вилучена після демодуляції.

Отже, остаточно можна записати АМ-сигнал у такому вигляді:

$$S_{AM}(t) = (A_0 + ks_M(t)) \cos(\omega_0 t + \varphi_0). \quad (5.1)$$

5.1.1 Однотональна амплітудна модуляція

Для розуміння суті амплітудної модуляції та спектральної структури АМ-сигналу корисно докладніше розглянути окремий випадок, коли модулюючий сигнал є гармонічним:

$$S_{AM}(t) = A_M \cos(\Omega t + \Phi_0).$$

$$S_{AM}(t) = (A_0 + A_M \cos(\Omega t + \Phi_0)) \cos(\omega_0 t + \varphi_0). \quad (5.2)$$

Відношення між амплітудами модулюючого сигналу A_M і несучого коливання A_0 називається **коефіцієнтом модуляції** або **глибиною модуляції**:

$$m = \frac{A_M}{A_0}.$$

З урахуванням цього можна записати:

$$S_{AM}(t) = (1 + m \cos(\Omega t + \Phi_0)) \cos(\omega_0 t + \varphi_0).$$

На рис. 5.3 показаний вигляд однотонового АМ-сигналу при різних значеннях коефіцієнта модуляції: а) $m = 0$ (немодульована несуча); б) $m = 0,5$; в) $m = 1$.

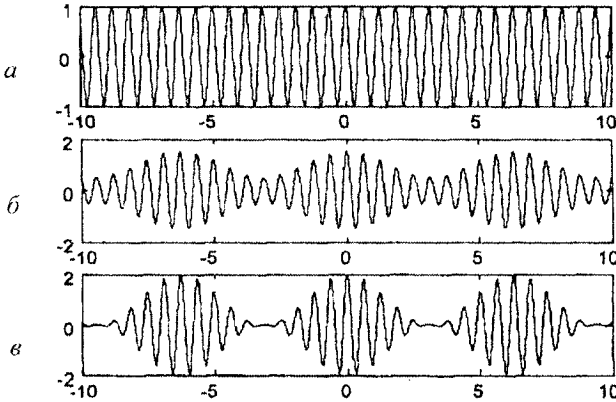


Рис. 5.3

Очевидно, що максимальне значення обвідної однотонового АМ-сигналу досягається тоді, коли обидва косинуси рівні 1:

$$A_{\max} = A(1 + m).$$

Мінімальне значення обвідної відповідає тим моментам, коли косинус модулюючого сигналу дорівнює -1 : $A_{\min} = A(1 - m)$.

Звідси випливає формула, що дозволяє обчислити коефіцієнт модуляції m за результатами вимірювання (наприклад, за допомогою осцилографа) максимальної і мінімальної амплітуд сигналу:

$$m = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}}.$$

Звичайно коефіцієнт модуляції повинен знаходитись у діапазоні $0 \dots 1$. При $m > 1$ має місце **перемодуляція**. Підстановка таких значень у формулу (5.4) дає результат, показаний на рис. 5.4.

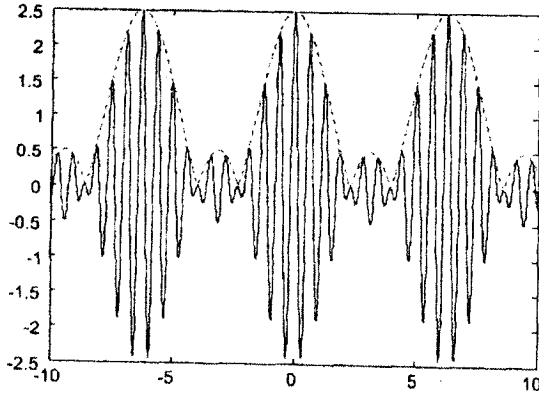


Рис. 5.4 – Однотональний АМ-сигнал у випадку перемодуляції ($m = 1,5$)

Як уже зазначалося, амплітудна обвідна при перемодуляції спотворюється. Однак, і цей режим може бути корисний на практиці.

Проаналізуємо спектральний склад такого коливання. Для цього спочатку розкриємо дужки у виразі для однотонального АМ-сигналу, а потім виконаємо тригонометричні перетворення:

$$\begin{aligned}
 S_{AM}(t) &= (A_0 \cos(\omega_0 t + \varphi_0) + A_0 m \cos(\Omega t + \Phi_0)) \cos(\omega_0 t + \varphi_0) = \\
 &= A_0 \cos(\omega_0 t + \varphi_0) + \frac{A_0 m}{2} \cos((\omega_0 + \Omega)t + \varphi_0 + \Phi_0) + \\
 &\quad + \frac{A_0 m}{2} \cos((\omega_0 - \Omega)t + \varphi_0 - \Phi_0).
 \end{aligned} \quad (5.3)$$

Результат перетворення показує, що однотональний АМ-сигнал складається із трьох гармонічних складових, одна з яких являє собою несуче коливання з частотою ω_0 , а інші дві (їх називають бічними частотами) віддалені від нього вгору і вниз по частоті на величину Ω . Амплітуда несучого коливання дорівнює A_0 і не залежить від рівня модулюючого сигналу. Амплітуди бічних частот, рівні $A_0/2$, навпаки, пропорційні коефіцієнту модуляції m .

Для верхньої бічної частоти початкові фази несучої і модулюючого сигналу додаються, а для нижньої – віднімаються.

Амплітудний (а) і фазовий (б) спектри однотонального АМ-сигналу показані на рис. 5.5.

Слід підкреслити, що в загальному випадку отримане представлення однотонального АМ-сигналу не є його розкладом у ряд Фур'є.

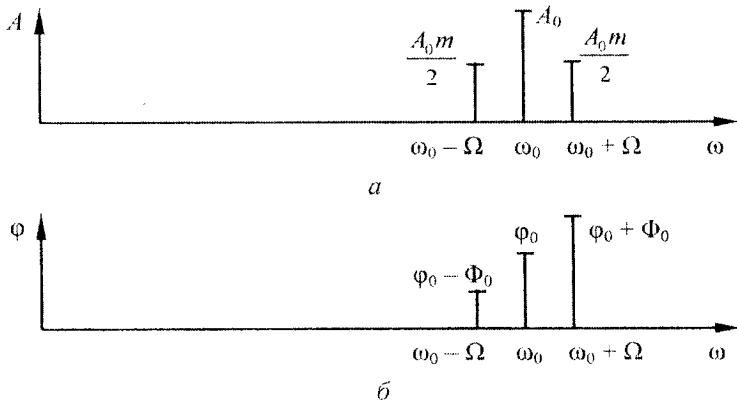


Рис. 5.5

Таке представлення у вигляді суми гармонічних функцій може розглядатися як ряд Фур'є тільки в тому випадку, коли для всіх трьох частот спектральних складових існує загальний дільник. Це можливо, якщо відношення ω_0/Ω є раціональним дробом.

З графіків видно, що ширина спектра однотонового АМ-сигналу в два рази перевищує частоту модулюючого сигналу: $\Delta\omega = 2\Omega$.

Щоб зрозуміти, як із трьох гармонічних складових із постійною амплітудою складається сигнал із мінливою амплітудою, побудуємо векторну діаграму. Для цього представимо кожне із трьох гармонічних коливань як дійсну частину комплексної експоненти:

$$S_{AM}(t) = \operatorname{Re} \left[e^{j\omega_0 t} \left(A_0 e^{j\varphi_0} + \frac{A_0 m}{2} e^{j(\Omega t + \omega_0 + \varphi_0)} + \frac{A_0 m}{2} e^{j(-\Omega t + \omega_0 - \varphi_0)} \right) \right].$$

Побудуємо векторну діаграму, що демонструє підсумовування цих складових (рис. 5.6).

При додаванні коливань з різними частотами постійно змінюються їх взаємні фазові співвідношення. Вектори бічних частот обертаються в різні сторони з кутовою швидкістю Ω щодо вектора несучого коливання.

У результаті коливання бічних частот виявляються то синфазними із несучим коливанням, збільшуючи його амплітуду, то протифазними з ним (тоді амплітуда сигналу зменшується). Нарешті, вектор несучої частоти обертається з кутовою швидкістю ω_0 . Значення АМ-сигналу визначається проекцією результуючого вектора на дійсну (Re) горизонтальну вісь.

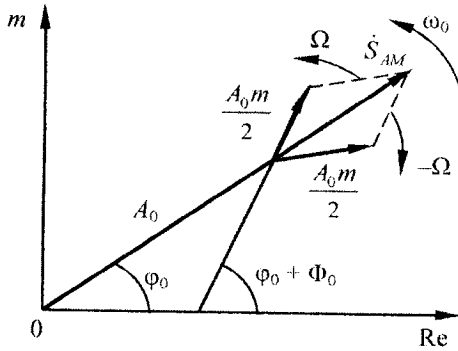


Рис. 5.6

5.1.2 Амплітудно-модульований сигнал у загальному випадку

АМ-сигнал – результат множення модулюючого сигналу (з доданою постійною складовою) на гармонічне несуче коливання. Формула, що зв’яже спектр АМ-сигналу зі спектром модулюючого сигналу:

$$\dot{S}_{AM}(\omega) = \frac{1}{2}e^{j\varphi_0}\dot{S}_A(\omega + \omega_0) + \frac{1}{2}e^{-j\varphi_0}\dot{S}_A(\omega - \omega_0). \quad (5.4)$$

Спектр обвідної $A(t)$ при амплітудній модуляції зсувається в область несучої частоти $\pm\omega_0$, “роздвоюючись” і зменшуючись у два рази за рівнем. Покажемо це на графіку, задавши яку-небудь функцію для спектра обвідної $\dot{S}_A(\omega)$. Спектри обвідної (пунктирна лінія) і АМ-сигналу (суцільна лінія) наведені на рис. 5.7.

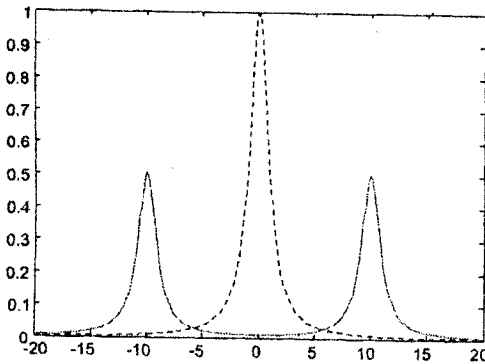


Рис. 5.7

Отже, спектр АМ-сигналу в загальному випадку містить несучу частоту (рівень якої визначається постійною складовою обвідної), а також верхню і нижню бічні смуги.

З графіків видно, що ширина спектра АМ-сигналу вдвічі більше максимальної (граничної) частоти модулюючого сигналу: $\Delta\omega = \Omega_{\max}$.

Обчислимо значення спектральної функції АМ-сигналу на несучій частоті:

$$\dot{S}_{AM}(\omega_0) = \frac{1}{2} e^{-j\varphi_0} \dot{S}_A(0) + \frac{1}{2} e^{j\varphi_0} \dot{S}_A(2\omega_0).$$

Перший доданок результату – як і повинно бути, є ділена навпіл постійна складова модулюючого сигналу. А от другий доданок являє собою “хвіст” від другої “половинки” спектра, сконцентрованої в області від’ємних частот, в околі частоти ω_0 .

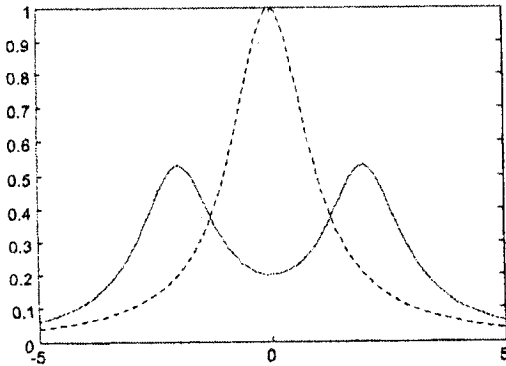


Рис. 5.8

Оскільки всі наявні реальні сигнали мають скінченну тривалість (нескінченно протяжний спектр), дане явище накладання “хвостів” завжди буде мати місце. У більшості практичних ситуацій, однак, несуча частота значно перевищує ефективну граничну частоту спектра обвідної, і тоді впливом такого ефекту можна знехтувати.

Графічно проілюструємо накладання “хвостів” зсунутих копій спектра, зменшивши в розглянутому прикладі несучу частоту (рис. 5.8). При недостатньо високій несучій частоті спектр АМ-сигналу може бути істотно несиметричним щодо несучої частоти через накладання “хвостів”.

5.1.3 Енергетичні співвідношення в АМ-сигналі

Повернемося до розгляду однотонального АМ-сигналу, щоб з’ясувати, як розподіляється потужність у його спектрі.

Для початку визначимо пікову потужність однотонального АМ-сигналу. Його максимальна амплітуда дорівнює $A_0(1+m)$, отже, пікова потужність складає: $P_{\max} = A_0^2(1+m)^2$.

Тепер визначимо середню потужність. Як уже зазначалося, АМ-сигнал в загальному випадку не є періодичним, тому для розрахунку середньої потужності необхідно застосувати граничний перехід:

$$P_{CP} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s^2(t) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \left(A_0 (1 + m \cos(\Omega t + \Phi_0)) \right) \times \\ \times \cos(\omega_0 t + \phi_0) \Big)^2 dt = \frac{A_0^2}{2} + \frac{A_0^2 m^2}{4}.$$

Той же результат можна одержати і без обчислення інтеграла. Досить згадати, що потужність гармонійного коливання з амплітудою A дорівнює $A^2/2$, і що гармонійні коливання різних частот є некорельованими, а тому їхні потужності можна складати. Далі залишається тільки застосувати все сказане до представлення однотонового АМ-сигналу у вигляді суми гармонійних складових.

Перший доданок не залежить від коефіцієнта модуляції і являє собою потужність немодульованої несучої. Корисна потужність, яка закладена в бічних частотах, представлена другим доданком.

Введемо в розгляд **коефіцієнт корисної дії** (ККД) амплітудної модуляції, визначивши його як відношення потужності бічних частот до загальної середньої потужності сигналу:

$$\eta_{AM} = \frac{A_0^2 \frac{m^2}{4}}{A_0^2 \left(\frac{1}{2} + \frac{m^2}{4} \right)} = \frac{m^2}{m^2 + 2}.$$

Побудуємо графік залежності ККД від коефіцієнта модуляції (рис. 5.9). Результати неутішні – навіть при максимально допустимому значенні коефіцієнта модуляції ($m = 1$) ККД складає лише 33 %, тобто дві третини потужності витрачається на передачу марної в інформаційному відношенні несучої.

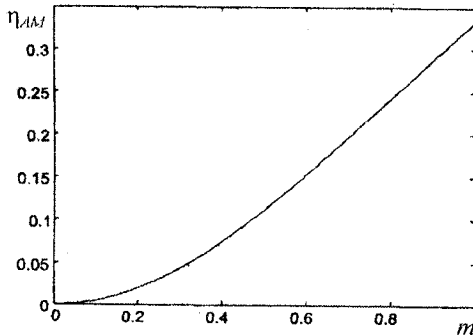


Рис. 5.9

Історично АМ була першим практично освоєним видом модуляції. Однак низький ККД і ширина спектра, яка удвічі перевищує ширину спектра модулюючого сигналу привели до того, що сфера застосування АМ стала досить вузькою. На сьогодні АМ застосовується для радіомовлення на порівняно низьких частотах (у діапазонах довгих, середніх і коротких хвиль) і для передачі зображення у телепередачах.

5.1.4 Демодуляція АМ-сигналів

Демодуляція АМ-сигналів може бути виконана декількома способами. Найпростіший шлях – імітувати роботу аналогового двонапівперіодного детектора. Ми обчислюємо модуль вхідного АМ-сигналу, а потім згладжуємо отримані однополярні косинусоїдальні імпульси, пропускаючи їх через ФНЧ. Двонапівперіодне детектування АМ-сигналу – однополярні імпульси (пунктирна лінія) і результат їх згладжування (суцільна лінія) – показано на рис. 5.10.

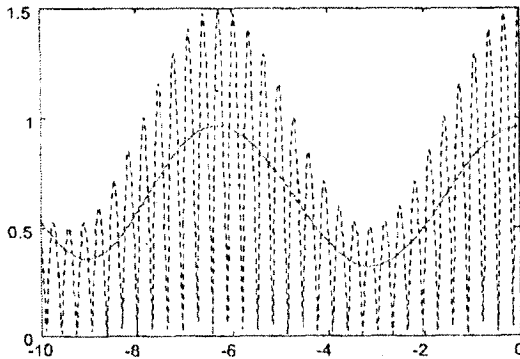


Рис. 5.10

Цей спосіб, мабуть, не буде працювати правильно у випадку перемодуляції.

Наступний метод – так зване **синхронне детектування**, суть якого полягає в множенні частоти сигналу на опорне коливання з несучою частотою:

$$\begin{aligned} y(t) &= S_{AM}(t) \cos(\omega_0 t + \varphi_0) = A(t) \cos^2(\omega_0 t + \varphi_0) = \\ &= \frac{1}{2} A(t) + \frac{1}{2} A(t) \cos(2\omega_0 t + 2\varphi_0). \end{aligned} \quad (5.5)$$

Результат множення містить два доданки. Перший – це шукана амплітудна функція, другий – АМ-сигнал із несучою частотою $2\omega_0$.

Цей високочастотний сигнал легко усувається шляхом пропущення сигналу через ФНЧ. Синхронне детектування АМ-сигналу – результат множення на опорне коливання (пунктирна лінія) і виділений низько-частотний сигнал (суцільна лінія) – показано на рис. 5.11.

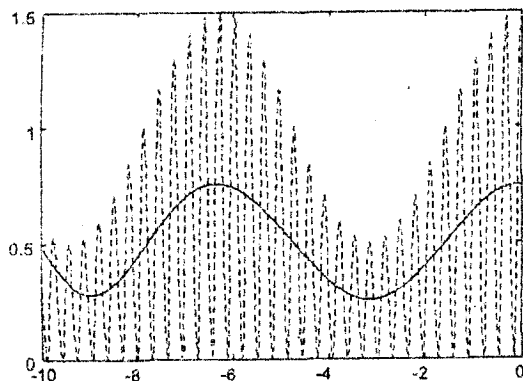


Рис. 5.11

Однак у даному випадку необхідно мати точний збіг початкових фаз і частот опорного коливання демодулятора і несучого коливання АМ-сигналу. При збігу частот, але розбіжності початкових фаз вихідний низькочастотний сигнал виявляється помноженим на косинус фазової похибки:

$$\begin{aligned} y(t) &= S_{AM}(t) \cos(\omega_0 t + \varphi_0) = A(t) \cos(\omega_0 t + \varphi_0) \cos(\omega_0 t + \varphi_0) = \\ &= \frac{1}{2} A(t) \cos(\omega_0 t + \varphi_0) + \frac{1}{2} A(t) \cos(2\omega_0 t + \varphi_0 + \varphi). \end{aligned}$$

Таким чином, при наявності фазової похибки рівень корисного сигналу на виході демодулятора спадає, а при похибці, яка рівна 90° , стає рівним нулю.

При наявності частотного зсуву між несучим і опорним коливаннями ситуація стає ще гіршою – вихідний низькочастотний сигнал виявляється помноженим на гармонійне коливання з різницевою частотою:

$$\begin{aligned} y(t) &= S_{AM}(t) \cos((\omega_0 + \Delta\omega)t) = A(t) \cos(\omega_0 t + \varphi_0) \cos((\omega_0 + \Delta\omega)t) = \\ &= \frac{1}{2} A(t) \cos(\Delta\omega t + \varphi_0) + \frac{1}{2} A(t) \cos((2\omega_0 + \Delta\omega)t + \varphi_0). \end{aligned}$$

У результаті вихідний сигнал буде пульсувати із частотою $\Delta\omega$. Це явище називається **биттям**, а різниця частот $\Delta\omega$ – **частотою биття**.

Для підтримки частотної і фазової синхронізації між несучим і опорним коливаннями використовуються системи фазової автопідналадки частоти (ФАПЧ), розгляд яких виходить за рамки змісту цієї книги.

Достоїнством синхронного детектування є те, що воно дозволяє правильно демодувати сигнал навіть у випадку перемодуляції, адже формула (5.5) не перестає бути правильною у випадку знакозмінної функції $A(t)$.

5.2 Різновиди амплітудної модуляції

Спроби поліпшити характеристики АМ привели до розробки декількох її модифікацій, які будуть розглянуті у підрозділі.

5.2.1 Амплітудна модуляція із заглушеною несучою

Перше, що спадає на думку при міркуванні щодо підвищення ККД амплітудної модуляції – це ідея видалити марне несуче коливання, усе-таки відмовившись від додавання постійної складової до модулюючого сигналу. Такий спосіб називається АМ із **заглушеною несучою** (АМ-ПН): $s(t) = S_M(t) \cos(\omega_0 t + \varphi_0)$.

Побудуємо графік такого сигналу при одностональній модуляції (рис. 5.12).

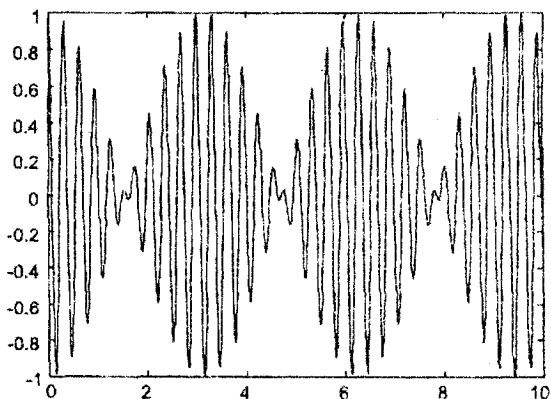


Рис. 5.12

Енергетичний вигравш при цьому, звичайно, великий (відповідно до введеного нами визначення, ККД стає рівним 100 %).

Ширина спектра АМ-сигналу із заглушеною несучою така сама, як у випадку звичайної АМ (адже заглушена лише середня (несуча) частота, а бічні частоти залишилися на місці).

Таким чином, АМ із заглушеною несучою має визначені переваги у порівнянні зі звичайною АМ. Однак цей спосіб модуляції не одержав широкого поширення, і пов'язано це із проблемами, що виникають при демодуляції сигналу.

5.2.2 Демодуляція АМ із заглушеною несучою

Як було зазначено у п. 5.1.4, формула (5.5) справедлива як для однополярної, так і для знакозмінної амплітудної функції $A(t)$. Тому демодуляція АМ із заглушеною несучою може виконуватися шляхом синхронного детектування. При цьому зберігає силу все сказане про необхідність точної відповідності частот і початкових фаз несучого й опорного коливаль.

Для полегшення правильного відновлення несучої іноді застосовують наступний прийом. На передавальній стороні несуче коливання заглушується не цілком. Його “залишок” із невеликою амплітудою (він називається *пілот-сигналом*) використовують для синхронізації частоти і фази несучого коливання на приймальній стороні.

5.2.3 Односмугова модуляція

Розглянута в попередньому пункті двосмугова АМ із заглушеною несучою має переваги перед звичайною АМ тільки в енергетичному сенсі – за рахунок усунення несучого коливання. Ширина ж спектра при цьому залишається рівною подвоєній частоті модулюючого сигналу.

Однак можна легко помітити, що спектри двох бічних смуг АМ-сигналу є дзеркальним відображенням один одного, тобто вони несуть ту саму інформацію. Тому одну із бічних смуг можна видалити. Модуляція, що виходить у результаті видалення однієї із бічних смуг, називається *односмуговою модуляцією*.

Залежно від того, яка бічна смуга зберігається, говорять про односмугову модуляцію із використанням *верхньої* чи *нижньої бічної смуги*. Формування односмугового сигналу найпростіше пояснити, привівши кілька спектральних графіків (рис. 5.13): a – спектр модулюючого сигналу; b – спектр односмугового сигналу із верхньою бічною смугою; c – спектр односмугового сигналу із нижньою бічною смугою.

По суті, при односмуговій модуляції відбувається просто зсув спектра сигналу в околі частоти несучого коливання. На відміну від АМ кожна “половинка” спектра зміщується у своєму напрямі: область додатних частот – до ω_0 , а область від’ємних частот – до $-\omega_0$.

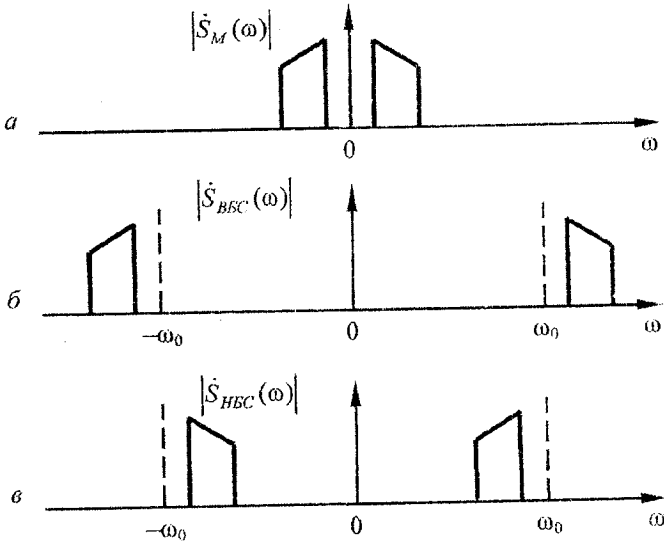


Рис. 5.13

Очевидно, що ширина спектра односмугового сигналу дорівнює ширині спектра модулюючого сигналу. Таким чином, спектр односмугового сигналу виявляється в два рази вужче, ніж при звичайній АМ.

На відміну від попередніх випадків тут нам не вдасться простими засобами виразити зв'язок між модульованим і модулюючим сигналом. Щоб зробити це, доведеться скористатися перетворенням Гілберта і поняттям аналітичного сигналу.

Отже, насамперед, ми сформуємо із модулюючого сигналу **аналітичний сигнал**, що має **однобічний спектр**. Множення цього сигналу на $\exp(j\omega_0 t)$ зсуває його однобічний спектр на ω_0 вправо (вгору по частоті), формуючи однобічний спектр односмугового сигналу із верхньою бічною смугою.

Нарешті, щоб перейти від аналітичного сигналу назад до дійсного, потрібно взяти дійсну частину. Формування сигналу із нижньою бічною смугою описується аналогічно, тільки множити аналітичний сигнал потрібно на $\exp(-j\omega_0 t)$ – тоді його спектр зсуветься вліво, в область від'ємних частот, і займе положення нижньої бічної смуги.

Тепер залишилося записати все сказане математично:

$$\begin{aligned}
 S_{SSB}(t) &= \operatorname{Re}((x(t) + jx_{\perp}(t))\exp(\pm j\omega_0 t)) = \\
 &= x(t) \cos \omega_0 t \mp x_{\perp}(t) \sin \omega_0 t.
 \end{aligned}
 \tag{5.6}$$

Знак “мінус” в остаточній формулі відповідає виділенню верхньої бічної смуги, “плюс” – нижньої.

Отже, односмуговий сигнал можна представити як суму двох АМ-сигналів, несучі коливання яких мають ту саму частоту, але зсунуті по фазі одна щодо іншої на 90° . Амплітудними функціями цих АМ-сигналів є модулюючий сигнал і його квадратурне доповнення. Залежно від того, складаються ці два АМ-сигнали чи віднімаються (а точніше, яка із двох несучих випереджає іншу по фазі), формується односмуговий сигнал із верхньою чи нижньою бічною смугою.

Амплітудна обвідна односмугового сигналу зовсім не схожа на низькочастотний модулюючий сигнал. Перевіримо це на конкретному прикладі. Випадок гармонічної модуляції цього разу розглядати не будемо – він занадто простий (синусоїда із частотою Ω при односмуговій модуляції перетвориться в синусоїду із частотою $\omega_0 \pm \Omega$; вибір знака визначається тим, яка бічна смуга використовується). Візьмемо модулюючий сигнал, який складається з двох гармонік (рис. 5.14).

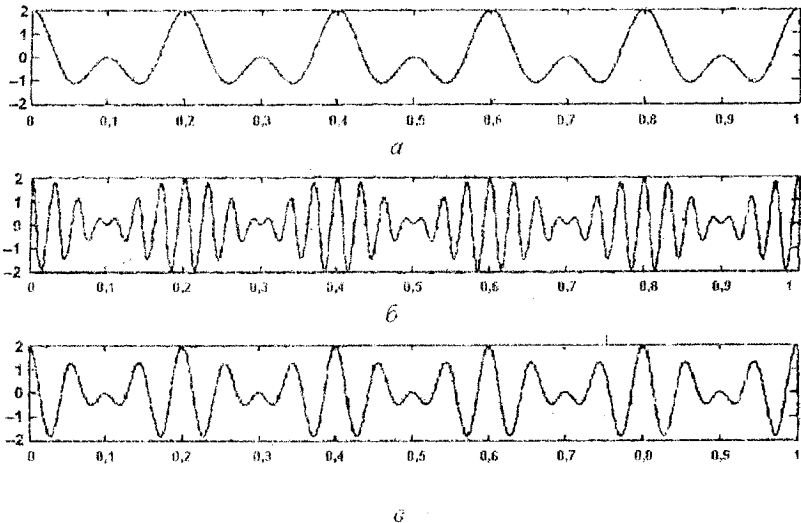


Рис. 5.14

Односмугова модуляція з різним рівнем несучого коливання – від повного заглушення до повного збереження – знайшла широке застосування в техніці професійного радіозв'язку; радіоматори використовують її під назвою SSB.

5.2.4 Демодуляція односмугового сигналу

На рис. 5.14 показана односмугова модуляція: a – модулюючий сигнал; b – сигнал із верхньою бічною смугою; c – сигнал із нижньою бічною смугою. Незважаючи на те що візуально односмуговий сигнал значно відрізняється від звичайного АМ-сигналу, його демодуляція можлива тим же методом синхронного детектування – шляхом множення на опорне коливання:

$$\begin{aligned} y(t) &= S_{SSB}(t) \cos \omega_0 t = (x(t) \cos \omega_0 t \pm x_{\perp}(t) \sin \omega_0 t) \cos \omega_0 t = \\ &= \frac{1}{2} x(t) + \frac{1}{2} x(t) \cos 2\omega_0 t \pm \frac{1}{2} x_{\perp}(t) \sin 2\omega_0 t. \end{aligned}$$

Результат множення (рис. 5.15, a) містить два доданки. Перший – це модулюючий сигнал, другий – односмуговий сигнал на подвоєній несучій $2\omega_0$. Залишається лише виділити модулюючий сигнал за допомогою ФНЧ. Відфільтрований демодульований сигнал показано на рис. 5.15, b .

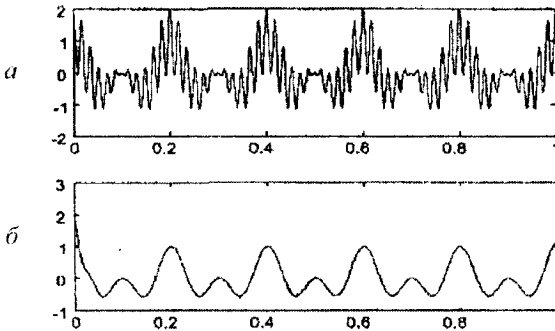


Рис. 5.15

Ефекти, що виникають при наявності частотного або фазового зсуву опорного коливання, у випадку односмугового сигналу виявляються не так, як при демодуляції АМ. Спочатку розглянемо ситуацію, коли є тільки фазовий зсув, а похибка по частоті відсутня:

$$\begin{aligned} y(t) &= s_{SSB}(t) \cos(\omega_0 t + \varphi) = (x(t) \cos \omega_0 t \pm x_{\perp}(t) \sin \omega_0 t) \times \\ &\quad \times \cos(\omega_0 t + \varphi) = \frac{1}{2} (x(t) \cos \varphi \pm x_{\perp}(t) \sin \varphi + \\ &\quad + \frac{1}{2} (x(t) \cos(2\omega_0 t + \varphi) \mp x_{\perp}(t) \sin(2\omega_0 t + \varphi)). \end{aligned}$$

Низькочастотна складова у цьому випадку являє собою лінійну комбінацію модулюючого сигналу $x(t)$ і його квадратурного доповнення $x_{\perp}(t)$. Зі спектральної точки зору це означає фазовий зсув всіх частотних складових сигналу на φ . Форма сигналу при цьому спотворюється. Чи прийнятні такі спотворення, залежить від характеру переданого сигналу. Якщо односмугова модуляція використовується для передачі звукового сигналу, фазові перекручування несуттєві, оскільки вухо людини нечутливе до фазових співвідношень, які виникають в акустичному сигналі.

Нехай тепер опорне колювання має частоту $\omega_0 + \Delta\omega$. Щоб одержати результат множення, досить підставити в попередню формулу значення $\varphi = \Delta\omega t$:

$$y(t) = s_{SSB}(t) \cos(\omega_0 + \Delta\omega)t = \frac{1}{2} \left(x(t) \cos(\Delta\omega t) \pm x_{\perp}(t) \sin(\Delta\omega_0 t) + \right. \\ \left. + \frac{1}{2} \left(x(t) \cos((2\omega_0 t + \Delta\omega)t) \pm x_{\perp}(t) \sin(2\omega_0 t + \Delta\omega)t \right) \right).$$

Низькочастотна складова тут фактично являє собою односмуговий сигнал із несучою частотою $\Delta\omega$. З частотної точки зору це означає зсув спектра сигналу, що модулює, на $\Delta\omega$. Це, зрозуміло, значно більш серйозні спотворення, ніж у попередньому випадку. Скажімо, при передачі музики спотворення такого роду зовсім неприйнятні. При передачі ж мовного сигналу (а професійний радіозв'язок – це основна сфера застосування односмугової модуляції) зсув спектра призводить до спотворення тембру голосу, але розбірливість мови (артикуляція) зберігається при величині зсуву до декількох десятків, а для досвідченого оператора – до декількох сотень герц.

5.2.5 Полярна модуляція

Полярна модуляція не є окремим видом модуляції. Це скоріше демонстрація застосування амплітудної модуляції для розв'язання конкретної технічної задачі. Мова йтиме про те, як реалізується **стереофонічне радіомовлення** в УКХ-діапазоні.

Для здійснення стереомовлення необхідно передавати два сигнали $S_L(t)$ і $S_R(t)$ (лівого і правого каналів) одночасно. У той же час при розробці системи такого мовлення накладається ще і вимога сумісності із уже наявними монофонічними приймачами. Тому для стереомовлення модифікується низькочастотний модулюючий сигнал, який і надходить на вхід модулятора передавача (у передавачі використовується кутова модуляція, яка буде розглянута далі). Низькочастотна складова модулюючого сигналу, яка лежить у звуковому діапазоні, для

сумісності з монофонічними приймачами, повинна являти собою монофонічний сигнал, тобто суму сигналів лівого і правого каналів:

$$S_{MONO}(t) = S_L(t) + S_R(t).$$

В області більш високих (ультразвукових) частот модулюючого сигналу за допомогою амплітудної модуляції передається додатковий сигнал, що дозволяє згодом виділити із отриманої суміші сигнали $S_L(t)$ і $S_R(t)$ кожен окремо. Оскільки монофонічний сигнал – це сума двох каналів, цим додатковим компонентом, що дозволяє відновити вихідні сигнали двох каналів, природно вибрати їхню різницю:

$$S_{DIFF}(t) = S_L(t) - S_R(t).$$

Розглянемо конкретний приклад, задавши для правого і лівого каналів гармонічні сигнали різних частот: $S_L(t) = A_L \cos(\Omega_L t)$, $S_R(t) = A_R \cos(\Omega_R t)$.

Графіки сигналів $S_L(t)$, $S_R(t)$, $S_{MONO}(t)$ і $S_{DIFF}(t)$ для даного випадку показані на рис. 5.16: *a* – сигнали лівого і правого каналів; *б* – сумарний і різницевий сигнали.

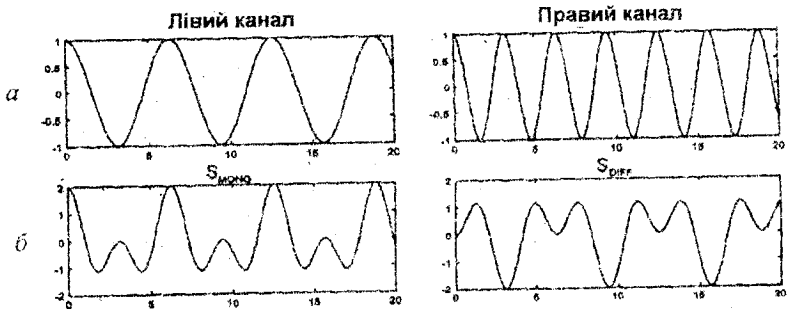


Рис. 5.16

Складаючи і віднімаючи сигнали $S_{MONO}(t)$ і $S_{DIFF}(t)$, можна відновити сигнали лівого і правого каналів:

$$S_L(t) = \frac{S_{MONO}(t) + S_{DIFF}(t)}{2}, \quad S_R(t) = \frac{S_{MONO}(t) - S_{DIFF}(t)}{2}.$$

Монофонічний сигнал, як уже було сказано, для сумісності повинен передаватися як ϵ , а різницевий сигнал модулює несучу частоту, розташовану трохи вище звукового діапазону частот (її називають що **піднесучою**, оскільки весь сигнал, про формування якого йде

мова, потім використовується для кутової модуляції несучого коливання радіопередавача): $S(t) = S_{MONO}(t) + (A_0 + S_{DIFF}(t))\cos(\omega_0 t)$.

Побудуємо графік сигналу, що виходить під час формування такого сигналу (рис. 5.17).

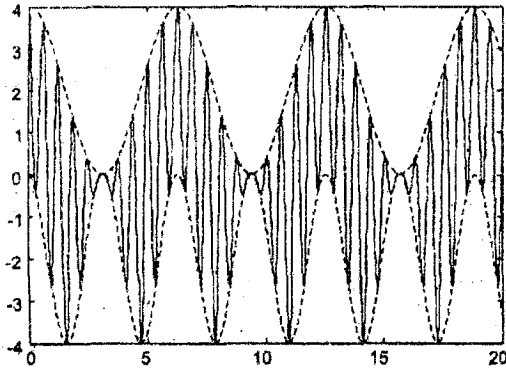


Рис. 5.17

Сигнал, що виходить, (його називають **компаративним стереосигналом**), як видно з рисунка, завдяки наявності низькочастотної складової має цікаву властивість: його верхня і нижня (додатна і від'ємна) обвідні не збігаються. Більш того, кожна з них із точністю до постійної складової відповідає сигналу одного із стереоканалів: верхня (додатна) обвідна — це лівий канал, а нижня (від'ємна) — правий. Така структура сигналу дозволяє легко виділити з нього сигнали стереоканалів.

Однак у показаному на рис. 5.17 вигляді сигнал для передачі не використовується. Справа в тому, що через наявність несучого коливання **розмах сигналу** (різниця між його максимальним і мінімальним значеннями) сильно збільшується у порівнянні із монофонічним сигналом, що небажано. Тому несуче коливання частково або цілком заглушують, так, що різницевий сигнал передається в режимі **перемодуляції**.

Вітчизняний стандарт, відповідно до якого ведеться радіомовлення в діапазоні 65...74 МГц, передбачає часткове заглушення частоти, що є піднесучою, яка рівна 31,25 кГц. Закордонний стандарт, який використовується у діапазоні 88...108 МГц, передбачає повне заглушення частоти піднесучої, рівної 38 кГц. При цьому для забезпечення можливості правильного відновлення коливання, піднесучої у стереодекодері приймача до складу сигналу вводиться **пілот-сигнал** на половинній частоті піднесучої (19 кГц).

Зобразимо композитний стереосигнал (рис. 5.18) для випадку повного заглушення піднесучої (пілот-сигнал додавати не будемо):

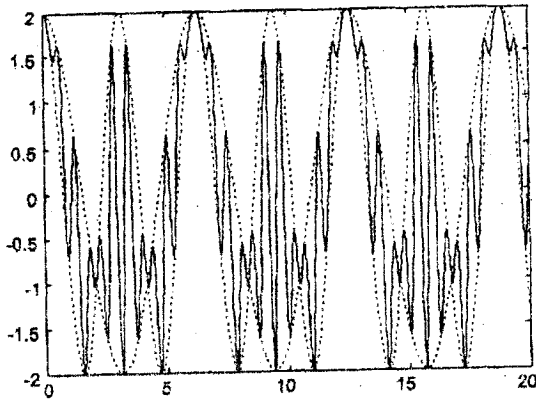


Рис. 5.18

Із графіка видно, що завдяки заглушенню піднесучої розмах сигналу істотно (у два рази) зменшився.

5.3 Кутова модуляція

Фазова і частотна модуляція тісно пов'язані одна із одною, завдяки чому й одержали загальну назву **кутова модуляція** (КМ). Тому насамперед у цьому підрозділі потрібно розглянути подібність і розбіжність цих двох видів модуляції.

5.3.1 Фазова модуляція

Назви двох розглянутих видів модуляції, як і у випадку із АМ, вказують на параметр несучого коливання, що лінійно зв'язаний із сигналом що модулює.

Нехай модулюючий сигнал визначає початкову фазу несучого коливання: $\varphi(t) = ks_M(t)$.

Звідси ми одержуємо сигнал із **фазовою модуляцією** (ФМ):

$$S_{\varphi_M}(t) = A \cos(\omega_0 t + ks_M(t)). \quad (5.7)$$

Весь аргумент функції косинуса, взятий цілком, називається **повною фазою** коливання: $\Psi(t) = \omega_0 t + ks_M(t)$.

Кругова частота коливання згідно з визначенням являє собою швидкість зміни початкової фази. Подібно тому як у випадку нерівномірного руху вводиться поняття миттєвої швидкості (рівної похідній від координати за часом), для коливань з кутовою модуляцією вводиться поняття **миттєвої частоти**, рівної похідній від повної фази за часом:

$$\omega(t) = \frac{d\Psi}{dt} = \omega_0 + k \frac{ds_M}{dt}.$$

Отже, у випадку фазової модуляції змінюється не тільки початкова фаза, але і миттєва частота коливання. Відповідно, повна фаза може бути знайдена шляхом інтегрування миттєвої частоти: $\Psi(t) = \int \omega(t) dt$.

5.3.2 Частотна модуляція

Тепер можна ввести поняття частотної модуляції (ЧМ), при якій модулюючий сигнал лінійно пов'язаний із миттєвою частотою коливання:

$$\omega(t) = \omega_0 + k s_M(t).$$

Добавка у вигляді константи ω_0 необхідна для того, щоб зробити коливання високочастотним.

Повна фаза знаходиться шляхом інтегрування:

$$\Psi(t) = \omega_0 t + k \int s_M(t) dt + \varphi_0, \quad (5.8)$$

де φ_0 – довільна постійна інтегрування.

Нарешті, сам ЧМ-сигнал має наступний вигляд:

$$S_{c_{\text{М}}}(t) = A \cos(\omega_0 t + k \int s_M(t) dt + \varphi_0).$$

Як бачимо, початкова фаза коливання при частотній модуляції зазнає змін, пропорційних інтегралу від сигналу, що модулює:

$$\varphi(t) = k \int s_M(t) dt + \varphi_0.$$

Таким чином, частотна і фазова модуляції виявляються взаємозалежними: якщо змінюється початкова фаза коливання, змінюється і його миттєва частота, і навпаки. З цієї причини два цих види модуляції поєднують під загальною назвою “кутова модуляція”.

Зі сказаного можна зробити кілька висновків.

1. За формою коливання з кутовою модуляцією не можна визначити, ФМ це чи ЧМ. Для цього необхідно знати ще і модулюючий сигнал.

2. Якщо пропустити модулюючий сигнал через ідеальний пристрій, що диференціює, а потім подати його на частотний модулятор, вийде фазова модуляція (верхня гілка на рис. 5.19).

3. Якщо пропустити сигнал, що модулює, через ідеальний пристрій, що інтегрує, а потім подати його на фазовий модулятор, вийде частотна модуляція (нижня гілка на рис. 5.19).

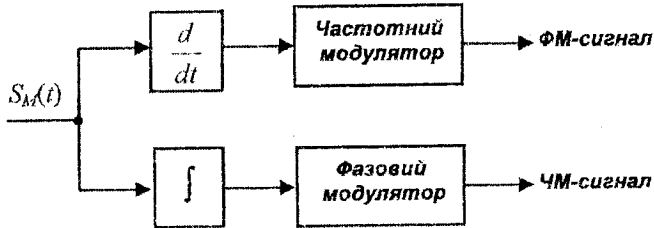


Рис. 5.19

У зведеній таблиці 5.1 показано, як зв'язані із модулюючим сигналом різні характеристики модульованого коливання при фазовій і частотній модуляції.

Таблиця 5.1

| Параметр | ФМ | ЧМ |
|-----------------|--|---|
| Початкова фаза | $\varphi(t) = ks_M(t)$ | $\varphi(t) = k \int s_M(t) dt + \varphi_0$ |
| Повна фаза | $\Psi(t) = \omega_0 t + ks_M(t)$ | $\Psi(t) = \omega_0 t + k \int s_M(t) dt + \varphi_0$ |
| Миттєва частота | $\omega(t) = \omega_0 + k \frac{ds_M}{dt}$ | $\omega(t) = \omega_0 + ks_M(t)$ |

5.3.3 Гармонійна кутова модуляція

Аналогічно тому, як ми це робили для амплітудної модуляції, розглянемо випадок гармонійного модулюючого сигналу. Початкова фаза коливання змінюється при цьому за гармонійним законом:

$$\varphi(t) = \beta \sin(\Omega t).$$

Коефіцієнт β називається **індексом кутової модуляції**. Він визначає інтенсивність коливань початкової фази. Повна фаза вийде шляхом додавання лінійного доданка $\omega_0 t$: $\Psi(t) = \omega_0 t + \beta \sin(\Omega t)$.

Нарешті, сам сигнал із гармонійною ЧМ:

$$S(t) = A \cos(\omega_0 t + \beta \sin(\Omega t)).$$

При зміні початкової фази змінюється і миттєва частота:

$$\omega(t) = \frac{d\Psi}{dt} = \omega_0 + \beta\Omega \cos(\Omega t).$$

У даному випадку миттєва частота змінюється також за гармонічним законом. Як видно із отриманої формули її максимальне відхилення від середнього значення ω_0 складає $\beta\Omega$. Ця величина називається **девіацією частоти** і позначається ω_d . Таким чином, ми отримали важливу формулу, яка показує, що індекс кутової модуляції дорівнює відношенню девіації частоти до частоти модулюючого сигналу: $\beta = \omega_d / \Omega$.

Отже, при гармонічній КМ і початкова фаза, і миттєва частота змінюються за гармонічним законом. Розходження між частотною і фазовою модуляціями починають виявлятися при зміні частоти модулюючого сигналу Ω .

При ФМ індекс β є характеристичним параметром модуляції і від частоти сигналу, що модулює, не залежить. Девіація частоти виявляється прямо пропорційною Ω : $\beta = \text{const}$, $\omega_d = \beta\Omega$.

При ЧМ характеристичним параметром, який не залежить від частоти модулюючого сигналу, є девіація частоти ω_d . Індекс модуляції в цьому випадку виявляється обернено пропорційним Ω : $\omega_d = \text{const}$, $\beta = \omega_d / \Omega$.

Залежності індексу модуляції та девіації частоти від частоти модулюючого сигналу у випадку ФМ і ЧМ представлені на рис. 5.20.

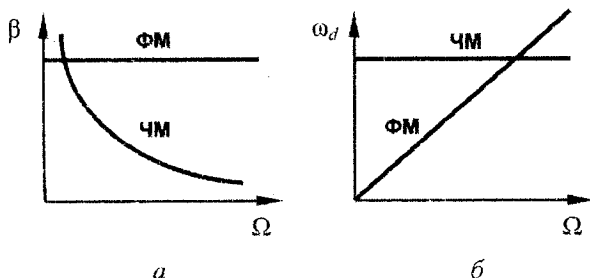


Рис. 5.20

5.3.4 Спектр сигналу із гармонічною кутовою модуляцією

У цьому підрозділі ми спочатку одержимо точний вираз для спектра радіосигналу із гармонічною КМ, а потім розглянемо його наближений варіант, справедливий для малих індексів модуляції ($\beta \ll 1$).

Для розрахунку спектрального представлення сигналу з гармонічною КМ, насамперед, представимо сигнал у вигляді дійсної частини комплексної експоненти:

$$S_{KM}(t) = \operatorname{Re} \left(A \exp(j\omega_0 t + j\varphi_0 + j\beta \sin(\Omega t + \Phi_0)) \right) = \\ = \operatorname{Re} \left(A \exp(j\omega_0 t + j\varphi_0) \exp(j\beta \sin(\Omega t + \Phi_0)) \right).$$

Тепер скористаємося представленням виразу $\exp(j\beta \sin x)$ у вигляді ряду Фур'є:

$$\exp(j\beta \sin x) = \sum_{k=-\infty}^{\infty} J_k(\beta) \exp(jkx). \quad (5.9)$$

Тут $J_k(\beta)$ – функція Бесселя першого роду k -го порядку від аргументу β . Використовуючи представлення (5.9), можна записати:

$$S'_{KM}(t) = \operatorname{Re} \left(A \exp(j\omega_0 t + j\varphi_0) \sum_{k=-\infty}^{\infty} J_k(\beta) \exp(jk(\Omega t + \Phi_0)) \right) = \\ = A \sum_{k=-\infty}^{\infty} J_k(\beta) \cos((\omega_0 + k\Omega)t + \varphi_0 + k\Phi_0).$$

Як бачимо, спектр сигналу містить нескінченну кількість складових із частотами $\omega_0 + k\Omega$, $k = 0, \pm 1, \pm 2, \dots$. Амплітуда k -ї складової дорівнює $AJ_k(\beta)$, тобто пропорційна функції Бесселя k -го порядку, аргументом якої є індекс модуляції β .

Функції Бесселя мають коливальний характер (графіки декількох з них наведені на рис. 5.21), тому спектр при видаленні від несучої частоти ω_0 спадає немонотонно.

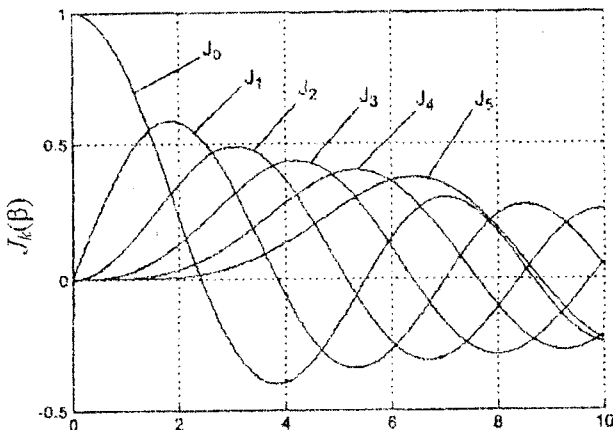


Рис. 5.21

Для прикладу на рис. 5.22 наведені амплітудні спектри сигналу з гармонічною КМ при індексі модуляції, рівному: а) $\beta = 1$, б) $\beta = 10$, в) $\beta = 100$. На рисунку передбачається, що $\omega_0 \gg \Omega$, тоді наявністю "хвостів", які заповзають із області від'ємних частот, можна знехтувати.

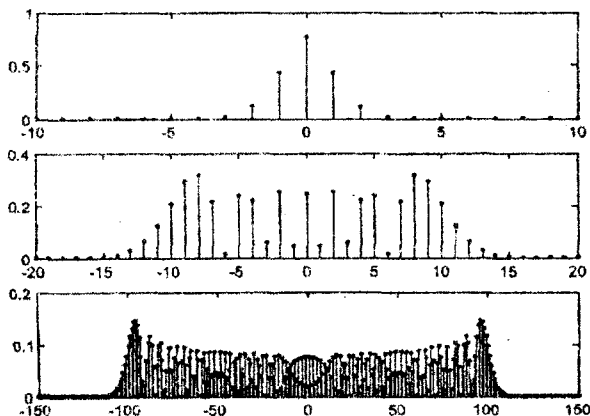


Рис. 5.22

Тепер наближено розглянемо окремий випадок малого індексу модуляції ($\beta \ll 1$). Почнемо з того, що застосуємо до сигналу з гармонічною КМ тригонометричні перетворення, щоб розкрити косинус суми:

$$S_{KM}(t) = A \cos(\omega_0 t + \varphi_0) \cos(\beta \sin(\Omega t + \Phi_0)) - \\ - A \sin(\omega_0 t + \varphi_0) \sin(\beta \sin(\Omega t + \Phi_0)).$$

Оскільки ми вважаємо, що $\beta \ll 1$, можна приблизно прийняти:

$$\cos(\beta \sin(\Omega t + \Phi_0)) \approx 1, \\ \sin(\beta \sin(\Omega t + \Phi_0)) \approx \beta \sin(\Omega t + \Phi_0).$$

З урахуванням цього:

$$S_{KM}(t) = A \cos(\omega_0 t + \varphi_0) - A \sin(\omega_0 t + \varphi_0) \beta \sin(\Omega t + \Phi_0).$$

Залишається представити останній доданок у вигляді напіврізниць косинусів:

$$S_{KM}(t) \approx A \cos(\omega_0 t + \varphi_0) - \frac{\beta A}{2} \cos((\omega_0 + \Omega)t + \varphi_0 + \Phi_0) + \\ + \frac{\beta A}{2} \cos((\omega_0 - \Omega)t + \varphi_0 - \Phi_0).$$

Отриманий результат нагадує отриманий вираз (5.3) для АМ-сигналу із гармонічною модуляцією – теж три складових із тими ж частотами, та й амплітуди їх розраховуються аналогічно (тільки замість коефіцієнта амплітудної модуляції m у формулі фігурує індекс кутової модуляції β). Однак є принципова відмінність, що перетворює амплітудну модуляцію в кутову – знак “мінус” перед одним із доданків, що відповідають бічним частотам. Чому цей знак виявляється настільки важливим, ми побачимо трохи нижче, при створенні векторної діаграми, а поки відзначимо головне: щоб перетворити сигнал із гармонічною АМ у сигнал із гармонічною КМ, досить змінити на 180° початкову фазу однієї із бічних частот.

Ще один спосіб перетворити АМ у КМ – змінити фазу складової із частотою несучої на 90° .

Амплітудний і фазовий спектри сигналу із КМ при малому індексі модуляції (її ще називають **вузькосмуговою КМ**) показані на рис. 5.23 (пунктирна лінія демонструє фазу верхньої бічної частоти для випадку АМ).

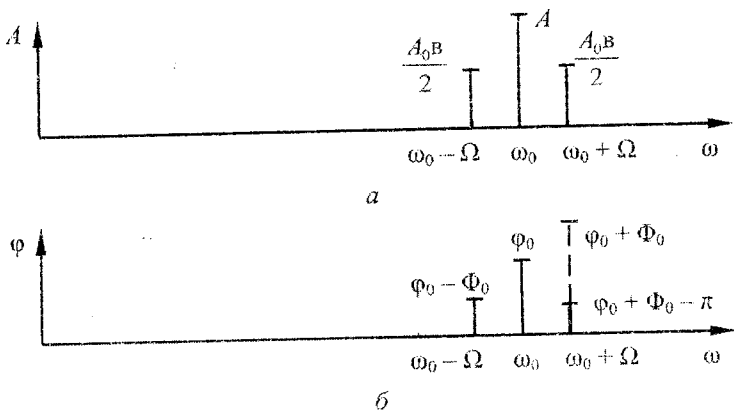


Рис. 5.23 – Амплітудний (а) і фазовий (б) спектри сигналу із гармонічною КМ при $\beta \ll 1$

Тепер побудуємо векторну діаграму аналогічно тому, як ми робили це для гармонічної АМ. Оскільки спектральні представлення сигналів із гармонічною АМ і КМ розрізняються лише знаком перед одним із доданків, докладно коментувати побудову векторної діаграми сигналу із гармонічною КМ, показаної на рис. 5.24, немає необхідності. Скажемо тільки, що для більшої наочності при побудові обране досить велике значення індексу модуляції ($\beta = 0,5$).

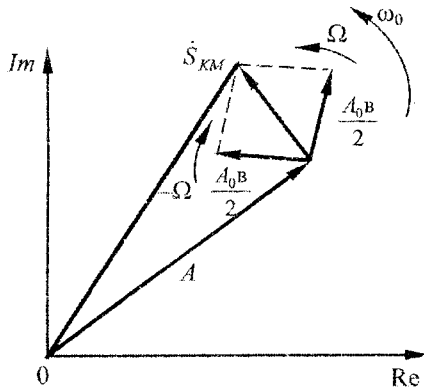


Рис. 5.24

Із векторної діаграми видно, що при показаному додаванні векторів міняється не тільки початкова фаза, але й амплітуда результуючого вектора. Це пов'язано із тим, що формула, згідно з якою здійснюється побудова діаграми, є наближеною. Точний розрахунок амплітудної обвідної і фазової функції даного сигналу можна виконати, скориставшись перетворенням Гілберта.

5.3.5 Ширина спектра сигналу з гармонічною КМ

При малому індексі модуляції, коли можна використовувати розглянуте наближене спектральне представлення, ширину спектра можна прийняти рівною 2Ω (те саме мало місце у випадку АМ-сигналу). Складніше оцінити ефективну ширину спектра у випадку великого індексу модуляції β . (У випадку КМ мова йде саме про **ефективну** ширину спектра, оскільки спектр, містить нескінченне число складових.)

При фіксованому аргументі β функції Бесселя згасають із зростанням (по модулю) їхнього порядку k . Якщо оцінити, як співвідносяться індекс модуляції β і номер функції Бесселя k , починаючи із якого абсолютні величини функцій Бесселя стають мізерними, аналіз покаже, що із зростанням порядку k при фіксованому аргументі β функції Бесселя згасають. Звичайно вважають, що при оцінюванні ефективної ширини спектра у випадку $\beta \gg 1$ можна знехтувати складовими із номерами, які за абсолютною величиною переважають індекс кутової модуляції β . Таким чином, залишаються складові, для яких $|k| \leq \beta$, і **ефективна ширина спектра** виявляється рівною: $\Delta\omega_{\text{ЕФ}} = 2\beta\Omega$.

З врахуванням наведеного співвідношення, що визначає індекс кутової модуляції як відношення між дев'яцією частоти і частотою

модуючого сигналу, ефективна ширина спектра виявляється рівною подвоєній девіації частоти: $\Delta\omega_{\text{ЕЦ}} = 2\omega_d$.

Отже, залежно від значення індексу модуляції β можна навести дві формули для розрахунку ефективної ширини спектра КМ-сигналу:

– при $\beta \ll 1$ ширина спектра дорівнює подвоєній частоті модулюючого сигналу: $\Delta\omega_{\text{ЕЦ}} = 2\Omega$;

– при $\beta \gg 1$ ширина спектра дорівнює подвоєній девіації частоти: $\Delta\omega_{\text{ЕЦ}} = 2\omega_d$.

На відміну від АМ, у випадку кутової модуляції одержати аналітичний вираз для спектра при довільному модулюючому сигналі не вдається. Навіть спроба розглянути двотональний модулюючий сигнал робить аналітичні викладення набагато складнішими, ніж при гармонійній модуляції.

5.3.6 Демодуляція сигналів із кутовою модуляцією

Як і у випадку АМ, демодуляція КМ-сигналу може виконуватися різними способами. Найбільш радикальний підхід – обчислити аналітичний сигнал і виділити його фазову функцію. Подальші дії залежать від виду кутової модуляції. Результат фазової (суцільна лінія) і частотної (пунктирна лінія) демодуляції сигналу із гармонічною КМ показано на рис. 5.25.

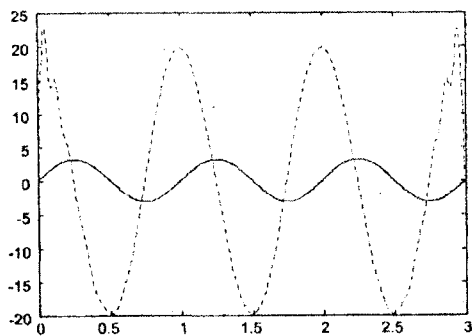


Рис. 5.25

Для реалізації демодуляції в реальному масштабі часу необхідна послідовна обробка відліків вхідного сигналу, що надходять. Це можна реалізувати, використовуючи для одержання аналітичного сигналу наближену гільбертівську фільтрацію у часовій області.

Ще однією альтернативою, придатною для реалізації в реальному масштабі часу, є **квадратурна обробка**. При цьому вхідний сигнал множить на два опорних коливання, зсув по фазі між якими складає 90° :

$$\begin{aligned}
 y_I(t) &= S_{KM}(t) \cos \omega_0 t = A \cos(\omega_0 t + \varphi(t)) \cos \omega_0 t = \\
 &= \frac{A}{2} \cos \varphi(t) + \frac{A}{2} \cos(2\omega_0 t + \varphi(t)), \\
 y_Q(t) &= S_{KM}(t) \sin \omega_0 t = A \cos(\omega_0 t + \varphi(t)) \sin \omega_0 t = \\
 &= \frac{A}{2} \sin \varphi(t) + \frac{A}{2} \sin(2\omega_0 t + \varphi(t)).
 \end{aligned}$$

Кожен із результатів множення містить два доданки. Один із них – низькочастотний (косинус чи синус початкової фази), інший – високочастотний (КМ-сигнал із несучою частотою $2\omega_0$). Низькочастотні складові виділяються за допомогою ФНЧ:

$$y'_I(t) = \frac{A}{2} \cos \varphi(t), \quad y'_Q(t) = \frac{A}{2} \sin \varphi(t).$$

Подальші дії, так само як і раніш, залежать від виду кутової модуляції. Для демодуляції ФМ нам необхідно обчислити фазу отриманої пари квадратурних складових:

$$\begin{aligned}
 x_{\text{ФМ}}(t) &= -\arg(y'_I(t) + jy'_Q(t)) = -\arg\left(\frac{A}{2} \cos \varphi(t) - j \frac{A}{2} \sin \varphi(t)\right) = \\
 &= -\arg\left(\frac{A}{2} \exp(-j\varphi(t))\right) = \varphi(t).
 \end{aligned}$$

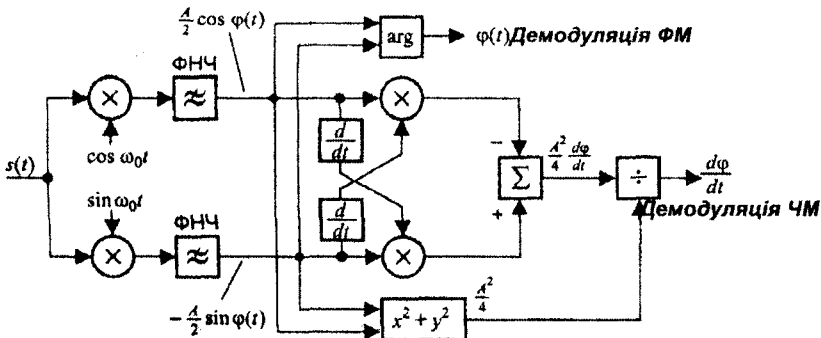


Рис. 5.26

Для демодуляції ЧМ отриману фазову функцію необхідно про- диференціювати:

$$\begin{aligned}
 x_{\text{ДМ}} &= \frac{dx_{\text{ДМ}}(t)}{dt} = \frac{d}{dt} \left(-\arg \left(y'_1(t) + jy'_Q(t) \right) \right) = \\
 &= -\frac{d}{dt} \arctg \left(\frac{y'_Q(t)}{y'_1(t)} \right) = \frac{\frac{dy'_1}{dt} y'_Q(t) - \frac{dy'_Q}{dt} y'_1(t)}{y_1'^2(t) + y_Q'^2(t)}.
 \end{aligned}$$

Структурна схема демодулятора, що вийшов із рівняння, показана на рис. 5.26. Її перевагою є те, що при високій несучій частоті вхідні блоки (генератор опорного колювання й помножувачі) можуть бути виконані в аналоговому вигляді. Ще одним способом демодуляції сигналів із КМ є використання систем ФАПЧ. Розгляд таких систем виходить за рамки цього видання.

5.4 Методи здійснення зовнішньої частотної модуляції радіосигналів

5.4.1 Принципи здійснення ЗЧМ

Проблема здійснення керованої зміни частоти радіосигналів є не лише завданням радіопередавальної техніки (здійснення ЧМ), а також безпосередньо стосується методів та пристроїв формування радіовимірювальних сигналів (характеріографи, вимірювачі АЧХ, ФЧХ, радіолокатори, хромографи тощо). Тому розробка нових методів є досить актуальною у обох вказаних галузях.

Насамперед визначимо, в чому полягають методичні та практичні обмеження існуючих методів, а також намітимо шляхи їх подолання, які можуть бути реалізовані в нових методах та пристроях. Відомі класичні методи керування частотою генератора радіосигналів поділяються на три напрямки: прямий, посередній та синтетичний (рис. 5.27).

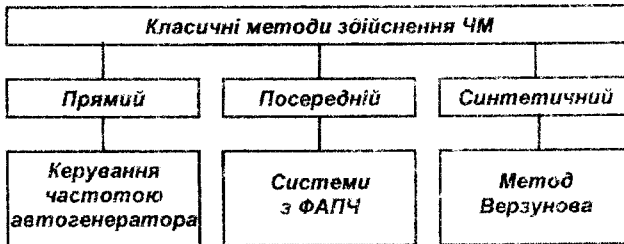


Рис. 5.27

На сучасному етапі розвитку техніки синтезу частот найбільшого поширення для здійснення ЧМ набули прямі методи, які широко представлені у виробках провідних фірм Hewlett-Packard, Qwelcom та інших. У той же час системи активного синтезу із ФАПЧ відтіснені у діапазон частот вище 0,5...1,0 ГГц і широко використовуються лідером радіоелектронного обладнання фірмою Motorola. Синтетичний метод Верзунова широкого розповсюдження не знайшов.

Необхідно відзначити, що перераховані методи мають істотний недолік, який полягає у неможливості здійснення ЧМ після того як сигнал від високостабільного генератора вийшов назовні (прямий синтез), і потребує системи ФАПЧ, аналогічної тій, що застосовується при посередньому або синтетичному синтезі. Без вказаних методів вони не допускають здійснення зовнішньої ЧМ (ЗЧМ), яка, наприклад, використовується в оптичних системах або при фазовому методі отримання ЧМ через ФМ.

Ознакою ЗЧМ є структурна схема котра функціонально має вигляд рис. 5.28, а, на відміну від рис. 5.28, б, де зображена внутрішня (пряма) ЧМ.

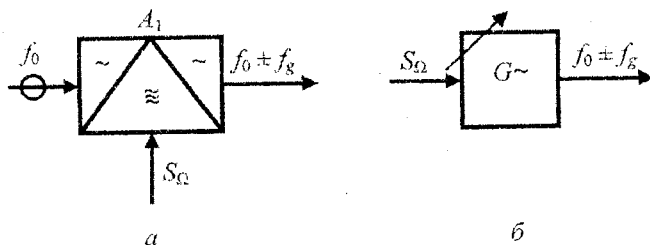


Рис. 5.28

Функціональний перетворювач A_1 є пристроєм для здійснення ЗЧМ і може бути використаний у вигляді аналогового або цифрового перетворювача, тому і методи їх реалізації можуть бути різними, головним залишається лише рівняння перетворення, котре можна записати у загальному виразі як систему рівнянь, враховуючи полярність S_Ω -сигналу модуляції:

$$\begin{cases} f_{\text{вих}} = f_0 + (\text{sign} f_D) \cdot f_D; \\ f_D = \kappa S_\Omega; \\ \text{sign} f_D = \text{sign} S_\Omega, \end{cases}$$

де $f_{\text{вих}}$ – частота ЗЧМ сигналу; f_D – девіація частоти ЗЧМ; κ – коефіцієнт перетворення, Гц/В; S_Ω – сигнал модуляції.

Знакова обробка $\text{sign}f_D = \text{sign}S_\Omega$ потрібна для однозначного перетворення різнополярних сигналів S_Ω .

Класичними представниками аналогового здійснення ЗЧМ можна вважати оптичні системи із використанням акустооптичних модуляторів (АОМ), котрі працюють в режимі дифракції Рамана–Ната або Вульфа–Брегга, а також їх фазовий аналог – механічний круговий фазообертач у режимі неперервного зсуву фаз (фазообертання), що використовується для змінення частоти (в даному випадку на фіксоване значення f_D).

Вказані схеми застосовуються для здійснення односмугової модуляції та активного змінення частоти оптичного (рис. 5.29, а) та радіосигналів (рис. 5.29, б).

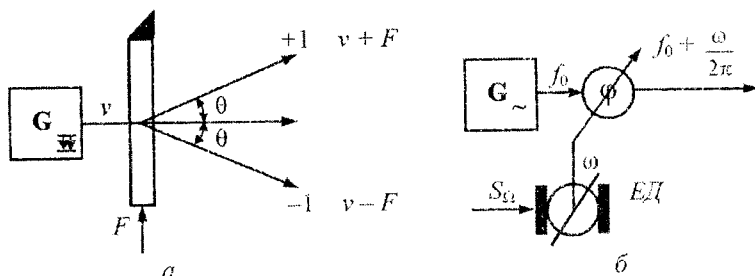


Рис. 5.29

У режимі дифракції Рамана–Ната (рис. 5.29, а) маємо одночасно +1 та -1 порядки дифракції та відповідні зсуви частот $\nu + F$ та $\nu - F$.

Для механічного фазообертача (рис. 5.29, б), вал якого обертається із кутовою швидкістю від електродвигуна (ЕД) сигнал на виході отримає зсув до значення: $f_{\text{вих}} = f_0 \pm \frac{\omega}{2\pi}$, де знак \pm залежить від напрямку фазообертання.

Тому при аналізі, дослідженні та розробці нових методів здійснення ЗЧМ вказані методи були взяті за прототип, для якого було встановлено загальні особливості, і на основі методів теорії фазочастотних вимірювань та перетворень параметрів радіосигналів (ФЧВ і ПР) було запропоновано систему їх класифікації.

Оскільки, згідно з визначеннями теорії ФЧВ і ПР основною інформаційною величиною що характеризує фазочастотні параметри радіосигналу є $\Psi(t)$ – повний фазовий зсув, його похідна $d\Psi(t)/dt$ і характер поведінки у часі, а також враховуючи, що для радіосигналів замість поняття частота f_0 необхідно використовувати **секвенційність** f_s

(частотність), завдання здійснення ЗЧМ зводиться до керованого синтезу секвентності.

Фазочастотний синтез має два методи реалізації: аналоговий та цифровий. Він базується на штучному створенні обертання вектора миттєвої фази сигналу в часі, із урахуванням відповідних обмежень, які є очевидними. Наприклад, ривок фази не повинен перевищувати $\Delta\varphi \leq 180^\circ$ або при реверсивній зміні $\Delta\varphi^+ \leq 90^\circ$, $\Delta\varphi^- \leq 90^\circ$.

На рис. 5.30 зображені найбільш характерні положення вектора $\varphi(t)$ на фазовому колі.

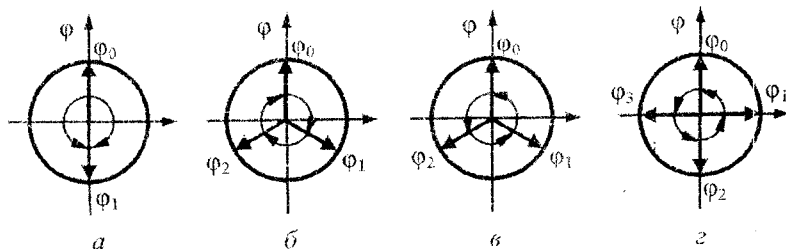


Рис. 5.30

Для $\Delta\varphi = \varphi_1 - \varphi_2 = 180^\circ$ (рис. 5.30, а) видно, що напрям зсуву частоти є невизначеним. І дійсно в спектрі такого сигналу відсутня несуча (класична 100-відсоткова балансна модуляція, рис. 5.31, а).

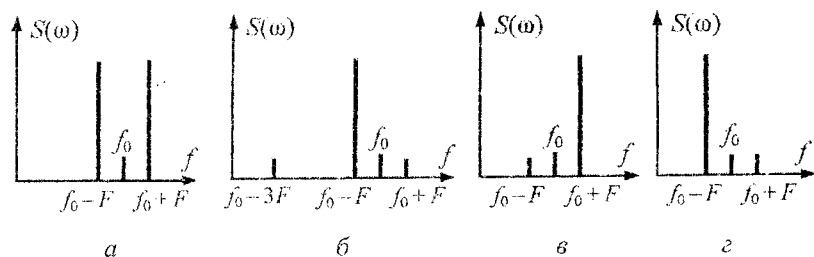


Рис. 5.31

Графіки спектральних складових на рис. 5.31 є якісними і наведені для пояснення принципів перетворення при дискретній зміні $\Psi(t)$, що і відображено побічними складовими.

Якщо сигнали є синусними і дискретизація відсутня, то вихідні сигнали будуть зсунутими на F і побічних складових не буде. У той же час відомо, що створити та підтримувати відповідне фазування $\varphi = 0^\circ$,

$\varphi = 120^\circ$, $\varphi = 240^\circ$ (трифазна система), або забезпечити більш ефективну у спектральному відношенні чотирифазну або квадратурну ($\varphi = 0^\circ$, $\varphi = 90^\circ$, $\varphi = 180^\circ$, $\varphi = 270^\circ$) систему є досить проблематичним, а їх розходження по фазі та амплітуді буде викликати появу побічних продуктів у спектрі вихідного сигналу, рівень котрих визначається аналогічно ступеню заглушення неробочих частот при фазовому способі формування односмугової модуляції. Тому дискретне і особливо цифрове перетворення відліків $\varphi_i(t)$ та $A_i(t)$ стає досить вискооефективним, враховуючи що використання квадратурних складових є однією із умов здійснення цифрової обробки радіосигналів.

Звідси можна сформулювати **обов'язкову умову для здійснення ЗЧМ**: сигнал, для якого здійснюється ЗЧМ, повинен бути представлений у вигляді векторів відліку фази і максимальний ривок $\Delta\varphi \leq 180^\circ$, (інтерпретація теореми відліків на фазовій осі сигналу).

Перехід до розгортки:

$$\Psi(t) = \sum_{i=1}^{\infty} \varphi_i(t), \quad (5.15)$$

Із врахуванням, що $\text{mod}_{2\pi} \Psi(t) \equiv f_s$, можна отримати об'ємну проекцію процесу здійснення ЗЧМ. На рис. 5.32 показано розташування векторів при квадратурній зміні поточного значення фазового зсуву.

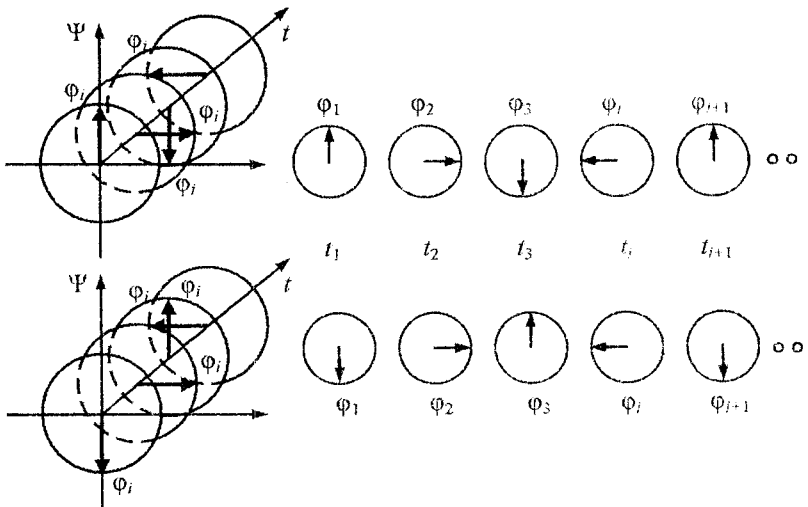


Рис. 5.32

Зрозуміло, що зображені положення векторів $\Psi(t)$ показані для випадку модуляції функцією $\text{sign}S_\Omega$; для модуляції S_Ω за законом

$\Psi(t) = \Omega t$ необхідно брати прирости $\Delta\varphi$ як площу під кривою модуляції, тобто при рівномірній дискретизації для функції $\text{sign}S_{\Omega}$:

$$\Delta\varphi_1 = \Delta\varphi_2 = \Delta\varphi_3 = \dots \Delta\varphi = 90^\circ.$$

Для закону модуляції S_{Ω} значення будуть обчислюватися як інтеграли:

$$\begin{aligned} \Delta\varphi_1 &= \int_0^{T/8} \sin 2\pi Ft dt = \dots; & \Delta\varphi_2 &= \int_{T/8}^{T/4} \sin 2\pi Ft dt = \dots; \\ \Delta\varphi_3 &= \int_{T/4}^{3T/8} \sin 2\pi Ft dt = \dots; & \Delta\varphi_4 &= \int_{3T/8}^{T/2} \sin 2\pi Ft dt = \dots \end{aligned}$$

При виконанні умови і підтриманні величини поточного значення приросту фази згідно $\Delta\varphi_i = \int_{t_i}^{t_{i+1}} \sin 2\pi Ft dt$ буде спостерігатися мінімальна кількість та рівень побічних спектральних складових секвени f_s .

Враховуючи що радіосигнали можуть бути представлені у вигляді аналітичного сигналу $A(t) = \text{Re}S(t) + j\text{Im}S(t)$ стає зрозумілим, що даній моделі представлення буде відповідати квадратурна система (рис. 5.30, а), а трифазна (рис. 5.30, б, в) може бути приведена до неї.

Таким чином **другою необхідною умовою для здійснення ЗЧМ є розкладання відомого сигналу на квадратурні складові**.

Дійсно, квадратурні складові застосовуються у механічних фазообертачах БІФ-114, 116 (індуктивних), а також і у ємнісних. У той же час безпосереднє використання механічних фазообертачів для ЗЧМ неможливе через значну інерційність механічної системи і цілковиту непридатність для роботи навіть у звуковому діапазоні частот зсуву. Вказаних недоліків позбавлені електронні фазообертачі дискретного типу або елементи секвентного синтезу.

Ще одним важливим аспектом ЗЧМ є необхідність перетворення $A(t) \rightarrow f(t)$, тобто амплітуди сигналу модуляції $A(t)$ в частоту девіації $f(t)$, що можливо здійснити різними способами, забезпечуючи лише однозначність перетворення.

5.4.2 Класифікація методів здійснення ЗЧМ

Так спираючись на загальні принципи теорії ФЧВ і ПР, а також на встановлені необхідні умови для здійснення ЗЧМ, можна запропонувати наступну класифікаційну структуру методів ЗЧМ (рис. 5.33).

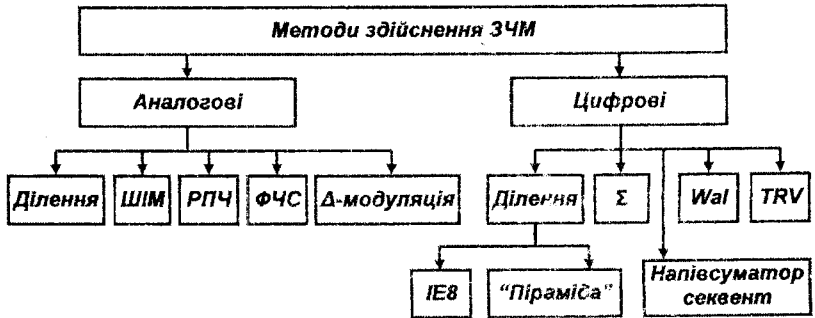


Рис. 5.33

Для пояснення суті запропонованої класифікації наведемо декотрі пояснення перетворень, які необхідно здійснити над сигналами для отримання ЗЧМ.

Оскільки аналогові методи майже не використовуються, то їх опис дамо в загальному вигляді. Так всі виділені способи здійснення ЗЧМ аналоговими методами полягають у використанні операції зміни частоти, до котрих належать “ділення” та її відомий прототип регенеративний подільник частоти (РПЧ). У прямому вигляді ділення частоти аналоговими методами досить проблематичне, тому прості алгоритми невідомі.

Найбільш широко при аналогових методах ЗЧМ застосовується фазочастотний синтез (ФЧС), до якого відносяться відомі методи ЗЧМ оптичних сигналів, та використання аналогових фазообертачів у динамічному режимі. Однак використовувати вказані розробки безпосередньо для ЗЧМ не завжди є можливим, враховуючи інерційність механічних частин, хоча перехід до електронних фазообертачів частково вирішує дану проблему.

5.4.3 Аналогові методи здійснення ЗЧМ

Аналогові методи на основі ФЧС можуть використовувати як синусні сигнали так і гармонічні із формою прямокутних імпульсів, при цьому перетворення вже не відповідає чіткому визначенню “частота” і найбільш відповідним буде поняття “секвентність”. Найбільш повно сутність аналогового здійснення ЗЧМ пояснює рис 5.29, б, при умові реверсивної роботи ЕД. Основною перевагою аналогового методу при виконанні умов ідеального фазування та рівності і стабільності амплітуд фазових векторів є можливість повного усунення побічних складових перетворення, котрі в цифрових системах присутні завжди, виходячи із наявності шумів квантування.

Системи секвентного синтезу ЗЧМ дозволяють її здійснення із використанням ще двох аналогових методів регулювання імпульсної послідовності: широтно-імпульсною модуляції (ШІМ) і Δ -модуляції.

Основним недоліком методу ШІМ є досить значна нерівномірність регулювання кількості імпульсів у пачці, що робить його зовсім непридатним для систем радіозв'язку, але допустимим для різних інтегруючих перетворень, де важливою є "секвентність", а не "частота".

Системи із Δ -модуляцією навпаки досить добре стикуються із аналоговими системами передачі, і можуть одночасно поєднувати функції інформаційного вимірювального перетворення та передачі інформації.

Оскільки для вимірювальних задач аналогові методи потребують додаткового перетворення, то їх використання не знайшло широкого впровадження, на відміну від цифрових методів.

5.4.4 Цифрові методи здійснення ЗЧМ

Цифрові методи мають один принциповий недолік, що полягає у наявності методичної похибки цифрового перетворення, але вказана похибка може бути зведена у реальній системі до необхідного рівня -- 40...60 дБ при використанні 10–16-розрядних АЦП-ЦАП.

Тому основним принципом роботи цифрових модуляторів для ЗЧМ є використання цифрового впливу на секвентність імпульсних послідовностей та виконання над ними лінійних перетворень фазочастотних параметрів.

Найбільш простим є метод ділення цифрової послідовності, який реалізується за відомим алгоритмом прорідження імпульсної послідовності (лічильник типу "ІЕ8"), або застосування структур із архітектурою подільника типу "піраміда", що дозволяє значно покращити спектральну однорідність вихідного сигналу. На відміну від "ІЕ8" "Піраміда" може керуватися непозиційним кількісним кодом, що в свою чергу дозволяє перетворювати у ЗЧМ відносну кількість поданих та відсутніх сигналів керування.

Відомий класичний метод арифметичного суматора у режимі накопичення суми особливих коментарів не потребує, і найбільш масово використовується при ЗЧМ. Однак принциповий недолік цього методу полягає в тому, що він обмежений половиною тактової частоти і не дозволяє в повній мірі розкрити всі можливості цифрового методу здійснення ЗЧМ. Вказаного недоліку позбавлений "напівсуматор секвент", який є раціональною комбінацією "піраміди" та суми.

У багатьох випадках більш ефективним може бути застосування цифрових методів здійснення ЗЧМ, які використовують алгоритм формування функцій Уолша ($Wal(I, \theta)$), або його більш широкої модифікації – використання функцій трансверсії ($Trv(I, \theta)$).

Наведена класифікація опирається на уже встановлені особливості теорії синтезу секвент, і використовує описані методи та принципи, незалежно від того, стосується це цифрових чи аналогових методів. Використання системного підходу та спирання на запропоновані методи дозволяє уже на стадії попереднього оцінювання та розрахунків вибрати метод, що найкращим чином задовольняє всім умовам функціонування модулятора ЧМС, який на відміну від класичних виконує ЗЧМ сигналу.

У цілому запропоновані методи не лише вирішують питання здійснення ЧМ, а й дозволяють розв'язати багато питань, які виходять за рамки існуючих методів.

5.5 Квадратурна модуляція

Ми розглянули випадки, коли амплітуда і початкова фаза гармонічного коливання піддавалися модуляції кожна окремо. Однак можна змінювати ці два параметри одночасно, одержавши за рахунок цього можливість передавати два сигнали відразу:

$$S(t) = A(t) \cos(\omega_0 t + \varphi(t)).$$

Таку модуляцію можна було б назвати **амплітудно-фазовою**. Однак два модулюючі сигнали виявляються у цьому випадку “нерівноправними”, тому що вони модулюють істотно різні параметри несучого коливання. Можна зробити ситуацію більш “симетричною”, перетворивши форму представлення сигналу. Для початку розкриємо косинус суми:

$$S(t) = A(t) \cos(\omega_0 t) \cos \varphi(t) - A(t) \sin(\omega_0 t) \sin \varphi(t).$$

Тепер сигнал виявився представленим у вигляді суми двох АМ-коливань. Їх несучі – $\cos(\omega_0 t)$ і $\sin(\omega_0 t)$ – зсунуті по фазі на 90° одна відносно іншої, а амплітудні функції дорівнюють $A(t)\cos\varphi(t)$ та $A(t)\sin\varphi(t)$. Позначимо ці амплітудні функції як $a(t)$ і $b(t)$, та використаємо їх як нову пару модулюючих сигналів (замість амплітуди і початкової фази):

$$S(t) = a(t)\cos(\omega_0 t) + b(t)\sin(\omega_0 t). \quad (5.10)$$

Таке представлення розглянутого сигналу називається **квадратурним**, а даний спосіб модуляції – **квадратурною модуляцією** (КАМ). Модулюючі сигнали у цьому випадку виявляються цілком рівноправними.

Варто підкреслити, що **квадратурна і амплітудно-фазова модуляції** – це різні представлення того самого сигналу. Розходження

між ними полягає тільки в тому, як пари модулюючих сигналів керують параметрами результуючого коливання.

5.5.1 Спектр сигналу із квадратурною модуляцією

У випадку амплітудно-фазового представлення сигналу записати аналітично вираз для спектра неможливо. А от для квадратурного представлення одержати спектральну функцію не складно. Оскільки КАМ-сигнал являє собою суму двох АМ-сигналів, ми можемо скористатися формулою (5.4) і записати:

$$\dot{S}(t) = \frac{1}{2} \dot{A}(\omega + \omega_0) + \frac{1}{2} \dot{A}(\omega - \omega_0) - \frac{1}{2} j \dot{B}(\omega + \omega_0) + j \frac{1}{2} \dot{B}(\omega - \omega_0).$$

Отже, аналогічно тому, що відбувається при амплітудній модуляції, спектри модулюючих сигналів “роздвоюються” і “переїжджають” в окіл частоти несучої $\pm\omega_0$. Якщо спектри модулюючих сигналів, $a(t)$ і $b(t)$ займають одну і ту ж смугу частот (як звичайно і буває), тоді вони будуть перекриватися і після зсуву в область несучої частоти. Однак при цьому спектр, що відповідає синусній несучій, додатково множиться на $\pm j$. Саме це дає можливість розділити квадратурні складові при прийомі сигналу.

5.5.2 Демодуляція сигналу з квадратурною модуляцією

Як і інші різновиди АМ, квадратурно-модульований сигнал може бути демодульований шляхом множення на опорне коливання. Однак оскільки КАМ-сигнал являє собою суму двох АМ-сигналів, тоді й опорних коливань повинно бути два – зі зсувом фаз на 90° :

$$\begin{aligned} y_I(t) &= s_{КАМ}(t) \cos \omega_0 t = (a(t) \cos \omega_0 t + b(t) \sin \omega_0 t) \cos \omega_0 t = \\ &= \frac{1}{2} a(t) + \frac{1}{2} a(t) \cos 2\omega_0 t + \frac{1}{2} b(t) \sin 2\omega_0 t, \\ y_Q(t) &= s_{КАМ}(t) \sin \omega_0 t = (a(t) \cos \omega_0 t + b(t) \sin \omega_0 t) \sin \omega_0 t = \\ &= \frac{1}{2} b(t) + \frac{1}{2} a(t) \sin 2\omega_0 t + \frac{1}{2} b(t) \cos 2\omega_0 t. \end{aligned} \quad (5.11)$$

Результат кожного множення містить три доданки. Один з них є низькочастотним і являє собою модулюючу функцію $a(t)$ або $b(t)$ зі зменшенням вдвоє рівнем. Інші два доданки утворюють КАМ-сигнал із несучою частотою $2\omega_0$. Тому корисні складові легко виділяються шляхом пропускання результатів множення через ФНЧ. Демодуляція КАМ-сигналу в дійсному (a) і комплексному (b) представленні показана на рис. 5.34.

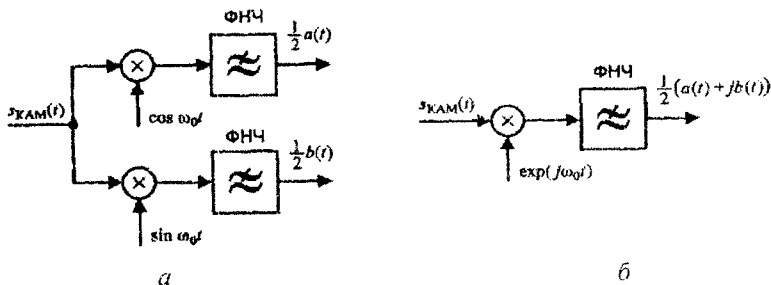


Рис. 5.34

Якщо не обмежуватися дійсними сигналами, можна представити демодуляцію КАМ трохи простіше. При цьому потрібен всього один канал, а опорним коливанням слугує комплексна експонента $\exp(j\omega_0 t)$:

$$\begin{aligned}
 y(t) &= s_{КАМ}(t) \exp(j\omega_0 t) = (a(t) \cos \omega_0 t + b(t) \sin \omega_0 t) \exp(j\omega_0 t) = \\
 &= \frac{1}{2}(a(t) + jb(t)) + \frac{1}{2}(a(t) - jb(t)) \exp(j2\omega_0 t).
 \end{aligned}$$

Результат комплексного множення містить низькочастотний доданок, що представляє обидві модулюючі функції і високочастотний комплексний доданок, що містить множник $\exp(j2\omega_0 t)$. Низькочастотний доданок зазвичай виділяється за допомогою ФНЧ. Структурна схема комплексного варіанта демодулятора показана на рис. 5.34, б. Дійсна і уявна частини комплексного вихідного сигналу відповідають двом вихідним сигналам дійсного демодулятора.

Під час демодуляції дуже важливе точне дотримання частоти і початкової фази опорного коливання. При наявності фазової похибки $\Delta\varphi$ на виходах демодулятора будуть отримані лінійні комбінації модулюючих функцій:

$$\operatorname{Re}((a(t) + jb(t)) \exp(j\Delta\varphi)) = a(t) \cos \Delta\varphi - b(t) \sin \Delta\varphi,$$

$$\operatorname{Im}((a(t) + jb(t)) \exp(j\Delta\varphi)) = a(t) \sin \Delta\varphi - b(t) \cos \Delta\varphi.$$

При наявності похибки за частотою фазовий зсув лінійно змінюється в часі, так що досить підставити в попередню формулу $\Delta\varphi = \Delta\omega t$:

$$\operatorname{Re}((a(t) + jb(t)) \exp(j\Delta\omega t)) = a(t) \cos \Delta\omega t - b(t) \sin \Delta\omega t,$$

$$\operatorname{Im}((a(t) + jb(t)) \exp(j\Delta\omega t)) = a(t) \sin \Delta\omega t - b(t) \cos \Delta\omega t.$$

Таким чином, вихідні сигнали демодулятора виявляються промодульованими, але амплітудою із частотою биттів.

5.6 Способи модуляції, які використовуються при передачі цифрової інформації

На сьогодні усе більша частина інформації, переданої різноманітними каналами зв'язку, існує у цифровому вигляді. Це означає, що передачі підлягає не безупинний (аналоговий) модулюючий сигнал, а послідовність цілих чисел n_0, n_1, n_2, \dots , що можуть приймати значення із деякої фіксованої скінченної множини. Ці числа, названі **символами**, надходять від джерела інформації з періодом T , а частота, що відповідає цьому періоду, називається **символьною швидкістю**: $f_T = 1/T$.

Часто використовуваним на практиці варіантом є двійкова послідовність символів, коли кожне з чисел n , може приймати одне із двох значень – 0 чи 1.

Послідовність переданих символів є також дискретним сигналом. Оскільки символи приймають значення із скінченної множини, цей сигнал фактично є **квантованим**, тобто його можна назвати **цифровим сигналом**. Далі у цьому пункті будуть розглядатися питання, пов'язані із перетворенням цього цифрового сигналу в аналоговий модульований сигнал.

Типовий підхід при здійсненні передачі дискретної послідовності символів полягає в наступному. Кожному із можливих значень символів зіставляється деякий набір параметрів несучого коливання. Ці параметри підтримуються постійними протягом інтервалу T , тобто до приходу наступного символу. Фактично це означає перетворення послідовності чисел $\{n_k\}$ у східчастий сигнал $s_k(t)$ із використанням кусково-постійної інтерполяції; $s_k(t) = f(n_k)$, $kT \leq t < (k+1)T$, де f – деяка функція перетворення.

Отриманий сигнал далі використовується як модулюючий сигнал звичайним способом.

Такий спосіб модуляції, коли параметри несучого коливання міняються стрибкоподібно, називається маніпуляцією. Залежно від того, які саме параметри змінюються, розрізняють амплітудну (АМп), фазову (ФМп), частотну (ЧМп) і квадратурну (КАМп) маніпуляції. Крім того, при передачі цифрової інформації може використовуватися несуче коливання, відмінне за формою від гармонічного. Так, при використанні як несучого коливання послідовності прямокутних імпульс-

сів можливі амплітудно-імпульсна (АІМ), широко-імпульсна (ШІМ) і часо-імпульсна (ЧІМ) модуляції.

5.6.1 Частотна маніпуляція

При частотній маніпуляції (ЧМп) кожному можливому значенню переданого символу зіставляється своя частота. Протягом кожного символного інтервалу передається гармонійне коливання із частотою, що відповідає поточному символу.

Як приклад сформуємо двопозиційний (бінарний) ЧМп-сигнал, у якому можливим значенням символів 0 і 1 відповідають частоти 1000 та 1400 Гц. Символьна швидкість буде дорівнює 400 символам в секунду, а частота дискретизації – 8 кГц (рис. 5.35).

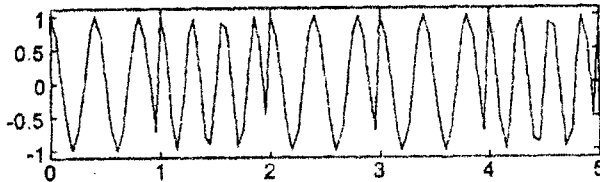


Рис. 5.35

Оскільки ми формували кожен посилку незалежно, сформований сигнал містить стрибки – це добре видно на графіку (рис. 5.35). Чим більш гладким є сигнал, тим швидше убуває його спектр. Тому можна спробувати зробити спектр ЧМп-сигналу компактніше, змінивши спосіб формування так, щоб усунути стрибки.

Такий спосіб формування сигналу називається **частотною маніпуляцією із безупинною фазовою функцією**. При цьому формується лінійно змінна повна фаза коливання, а передані символи керують швидкістю її зміни. Можна сказати і так: передані символи переключують значення **миттєвої частоти**; ця частота інтегрується, даючи безупинну фазову функцію; косинус такої повної фази теж буде безупинною функцією. Сформуємо зазначеним способом ЧМп-сигнал із тими ж параметрами, що й у попередньому прикладі (рис. 5.36).

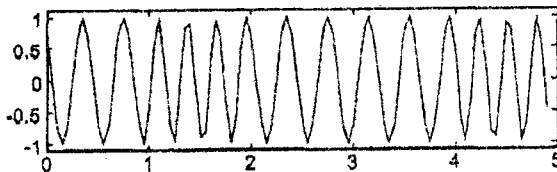


Рис. 5.36

Як бачимо, стрибки зникли. Тепер порівнюємо спектри потужності при двох варіантах формування ЧМп-сигналу (рис. 5.37).

Порівняння графіків (рис. 5.37) показує, що при наявності стрибків фази (рис. 5.37, *а*) в спектрі присутні яскраво виражені піки на частотах, кратних символній швидкості 400 Гц. У сигналі із безупинною фазовою функцією (рис. 5.37, *б*) такі піки відсутні, і спектр у цілому вищий більш компактним.

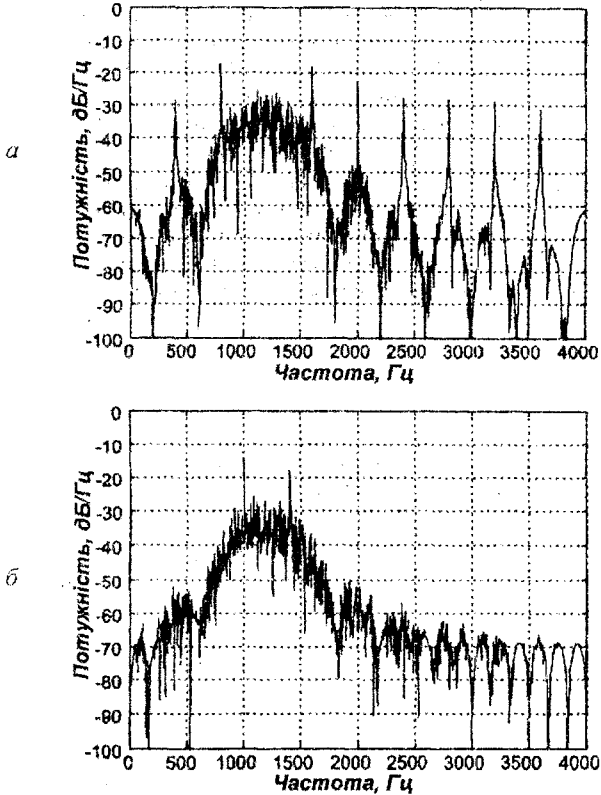


Рис. 5.37

Прийм ЧМп-сигналу, як правило, здійснюється кореляційним методом. Сутність його полягає в обчисленні взаємної кореляції між прийнятим сигналом і коливаннями-зразками (опорними сигналами), що являють собою гармонічні коливання із частотами, які використовуються для маніпуляції. Як вихідний символ вибирається той, частота якого виявляється максимально корельованою із вхідним сигналом.

Кореляційний прийом може бути **когерентним** чи **некогерентним**. Когерентний алгоритм може використовуватися, якщо відома початкова фаза коливання. Опорними сигналами при цьому слугують дійсні синусоїди із потрібними частотами і початковими фазами. Для ухвалення рішення про прийнятий символ порівнюються речовинні результати обчислення кореляційних сум.

На практиці, однак, початкова фаза, як правило, невідома. У таких випадках застосовують **некогерентний кореляційний прийом**, при якому опорні сигнали являють собою комплексні експоненти із потрібними частотами. Для ухвалення рішення про прийнятий символ порівнюються **модулі комплексних результатів обчислення кореляційних сум**. Модуль комплексної ВКФ не залежить від початкових фаз сигналів, однак завадостійкість такого алгоритму трохи гірша, ніж у випадку когерентного прийому.

5.6.2 Мінімальна частотна маніпуляція

Для підвищення завадостійкості ЧМп бажано, щоб посліжки, що відповідають різним символам, були некорельовані. Вважаючи початкові фази посліжок нульовими, ЧМп-сигнали для символів 0 та 1 можна записати так: $s_0(t) = A \cos \omega_0 t$, $0 \leq t \leq T$, $s_1(t) = A \cos \omega_1 t$, $0 \leq t \leq T$.

Їх ВКФ при нульовому тимчасовому зсуві дорівнює:

$$\begin{aligned} B_{01}(0) &= \int_0^T s_0(t) s_1(t) dt = A^2 \int_0^T \cos \omega_0 t \cos \omega_1 t dt \\ &= \frac{A^2 \sin(\omega_1 + \omega_0)T}{2(\omega_1 + \omega_0)} + \frac{A^2 \sin(\omega_1 - \omega_0)T}{2(\omega_1 - \omega_0)}. \end{aligned}$$

Якщо $(\omega_1 + \omega_0)T \gg 1$, тоді перший доданок значно менше другого і ним можна знехтувати:

$$B_{01}(0) \approx \frac{A^2 \sin(\omega_1 - \omega_0)T}{2(\omega_1 - \omega_0)}.$$

Це значення дорівнює нулю при $(\omega_1 + \omega_0)T = k\pi$, де k – ціле число, не рівне нулю. Таким чином, мінімальне значення відстані між частотами маніпуляції, при якому посліжки, що відповідають різним символам, виявляються некорельованими, складає:

$$\Delta\omega_{\min} = \frac{\pi}{T}, \quad \Delta f_{\min} = \frac{1}{2T} = \frac{f_T}{2}, \quad (5.12)$$

де f_T – символічна швидкість.

Двопозиційна (бінарна) ЧМп, частоти якої обрані згідно (5.12), одержала назву *мінімальної частотної маніпуляції* (МЧМн).

5.6.3 Амплітудна маніпуляція

Амплітудна маніпуляція (АМп), при якій стрибкоподібно міняється амплітуда несучого коливання, є частковим випадком квадратурної маніпуляції (див. п. 5.6.5). Тому тут ми тільки побудуємо як приклад графік АМп-сигналу (рис. 5.38).

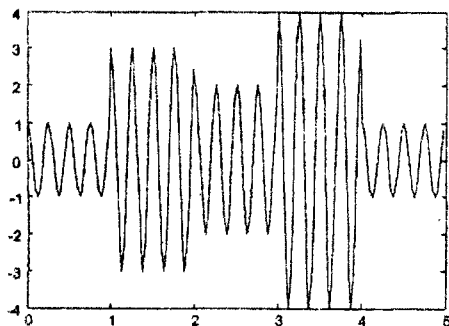


Рис. 5.38

5.6.4 Фазова маніпуляція

Фазова маніпуляція (ФМп), при якій стрибкоподібно міняється фаза несучого коливання, теж є частковим випадком квадратурної маніпуляції. На практиці фазова маніпуляція використовується при невеликому числі можливих значень початкової фази – як правило, 2, 4 чи 8. Крім того, при прийомі сигналу складно виміряти *абсолютне значення початкової фази*; значно простіше визначити *відносний фазовий зсув* між двома сусідніми символами. Тому звичайно використовується *фазорізницева маніпуляція* (синоніми – диференціальна фазова маніпуляція, відносна фазова маніпуляція).

Побудуємо графік сигналу із чотирипозиційною фазовою маніпуляцією (рис. 5.39).

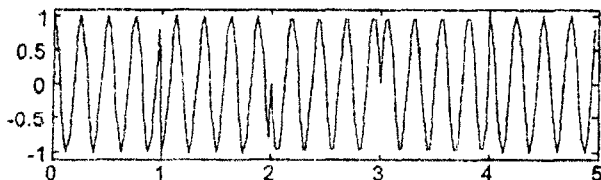


Рис. 5.39

5.6.5 Квадратурна маніпуляція

При квадратурній маніпуляції (КАМн) кожному із можливих значень дискретного символу C_k , ставиться у відповідність пара величин – амплітуди синфазної та квадратурної складових або, що еквівалентно, амплітуда і початкова фаза несучого коливання:

$$C_k \rightarrow (a_k, b_k),$$

$$s(t) = a_k \cos \omega_0 t + b_k \sin \omega_0 t, \quad kT \leq t \leq (k+1)T$$

або

$$C_k \rightarrow (A_k, \varphi_k),$$

$$s(t) = A_k \cos(\omega_0 t + \varphi_k), \quad kT \leq t \leq (k+1)T.$$

Параметри аналогового коливання, зіставлені дискретному символу C_k , зручно представляти у вигляді комплексного числа в алгебраїчній $(a_k + jb_k)$ чи експонентній $(A_k \exp(j\varphi_k))$ формі. Сукупність цих комплексних чисел для всіх можливих значень дискретного символу називається **сигнальним сузір'ям**.

У вітчизняній літературі також використовується громіздкий і не надто наочний термін “**сигнально-кодова конструкція**” (СКК). У даній книзі зроблений вибір на користь більш виразного слова, що запам'ятовується, – “сузір'я”.

При представленні дискретного символу комплексним числом C_k сигнал із квадратурною маніпуляцією можна записати у такий спосіб:

$$s(t) = \operatorname{Re}(C_k \exp(-j\omega_0 t)), \quad kT \leq t \leq (k+1)T.$$

На практиці використовуються сузір'я, що містять від чотирьох точок до декількох тисяч. На рис. 5.40 показані деякі сузір'я, які використовуються **модемами**, що призначені для передачі даних по телефонним лініям: a – 16; b – 128; v – 640 точок.

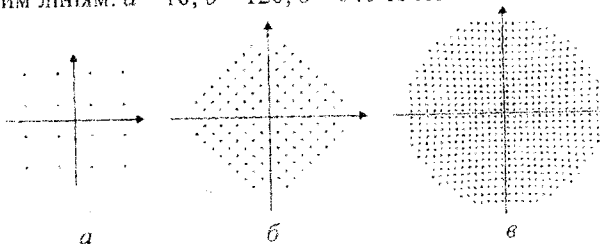


Рис. 5.40

Графік сигналу із квадратурною маніпуляцією виявляється не дуже наочним через змішаний (амплітудно-фазовий) характер модуляції.

Зміни амплітуди і фази при переході від символу до символу можуть бути невеликими і погано помітними на графіку.

Побудуємо графік сигналу, сформованого з використанням 16-точкового "квадратного" сузір'я, показаного на рис. 5.40, *a*. Оскільки нас не цікавить конкретний спосіб зв'язку дискретних символів і точок сузір'я, ми просто утворимо вектори амплітуд синфазної і квадратурної складових, значення яких випадково обрані з набору $\{-3, -1, 1, 3\}$ (рис. 5.41):

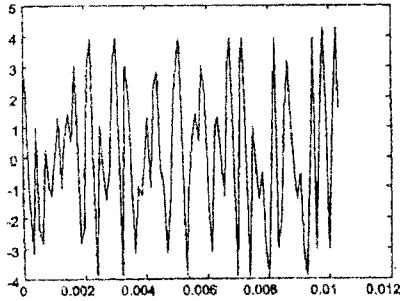


Рис. 5.41

Параметри сформованого сигналу (структура сузір'я, значення символів швидкості і несучої частоти) відповідають модему, що передає дані зі швидкістю 9600 біт/с відповідно до Рекомендації ITU-T V.32.

При квадратурній маніпуляції можуть мінятися і амплітуда, і початкова фаза несучого коливання, тому амплітудна і фазова маніпуляції є окремими випадками квадратурної — потрібно лише використовувати відповідні сузір'я. Наведемо графіки сузір'їв, що відповідають восьмипозиційній амплітудній (*a*) і фазовій (*б*) маніпуляції (рис. 5.42).

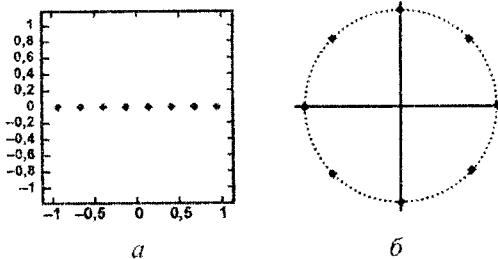


Рис. 5.42

5.6.6 Заводистійкість різних видів маніпуляції

За рахунок використання двовимірного характеру гармонічного несучого коливання (під двовимірністю тут розуміється наявність двох параметрів, які можна незалежно змінювати) квадратурна маніпуляція

забезпечує більшу завадостійкість (тобто меншу ймовірність помилки) ніж АМп і ФМп. Завадостійкість тим вище, чим більше відстань d між найближчими точками сузір'я на комплексній площині. При цьому для коректності порівняння різних сузір'їв у них повинні бути однаковими, крім числа точок, середньоквадратичні амплітуди. Порівняємо для прикладу завадостійкість 16-позиційних амплітудної (a), фазової (b) і квадратурної ($в$) маніпуляцій (рис. 5.43).

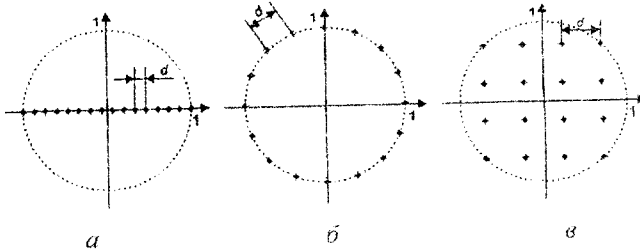


Рис. 5.43

Результати розрахунків мінімальної відстані між точками, середньоквадратичного рівня сигналів, а також міжточкової відстані, нормованої до цього рівня, зведені у таблицю 5.1.

Таблиця 5.2

| Вид маніпуляції | Мінімальна міжточкова відстань d_{min} | Середньоквадратичний рівень сигналу σ | Нормована відстань d_{min}/σ |
|-----------------|--|--|--|
| АМп | $\frac{2}{5} = 0,1333$ | $\frac{\sqrt{85}}{15} \approx 0,6146$ | $\frac{2}{\sqrt{85}} \approx 0,2169$ |
| ФМп | $2 \sin\left(\frac{\pi}{16}\right) \approx 0,39$ | 1 | $2 \sin\left(\frac{\pi}{16}\right) \approx 0,39$ |
| КАМп | $\frac{\sqrt{2}}{3} \approx 0,47$ | $\frac{\sqrt{5}}{3} \approx 0,745$ | $\frac{\sqrt{2}}{\sqrt{5}} \approx 0,63$ |

Аналіз даних з таблиці показує, що квадратурна маніпуляція забезпечує максимальну міжточкову відстань, нормовану до середньоквадратичного рівня сигналу.

5.6.7 Демодуляція сигналів із квадратурною маніпуляцією

Демодулюється сигнал із квадратурною маніпуляцією так само, як і у випадку аналогової квадратурної модуляції – сигнал множиться на два несучих коливання, зсунутих по фазі один щодо другого на 90° ,

а результати множення пропускаються через ФНЧ. На виході цих ФНЧ будуть отримані аналогові сигнали синфазної і квадратурної складових. Далі ці сигнали дискретизуються з частотою, рівною символній швидкості. Пари відліків синфазної і квадратурної складових утворюють комплексне число, і найближча до цього числа точка сузір'я, яке використовується, (а точніше – відповідний цій точці інформаційний символ) видається як вихідний результат.

5.6.8 Формування спектра

Якщо параметри модуляції аналогового сигналу підтримуються постійними протягом символного такту і на початку нового такту змінюються стрибкоподібно, це призводить до появи стрибків і у сформованому сигналі. Спектр сигналу, що містить стрибки, згасає із зростанням частоти повільно – пропорційно $1/\omega$. Щоб зробити спектр більш компактним, необхідно забезпечити гладкість сигналу, а це, у свою чергу, означає гладкість модулюючої функції. Отже, замість стрибкоподібною зміни параметрів модуляції нам необхідно виконати **інтерполяцію** між точками сузір'я, що відповідають послідовним символам.

Відповідно до теореми Котельникова, ми можемо з'єднати відліки, що надходять із символною швидкістю F_d , плавною функцією, що займає смугу частот від нуля до $F_d/2$. У цьому випадку квадратурно-маніпульований сигнал буде займати смугу частот шириною F_d . Однак повільне загасання функцій $\sin c$, що складають базис Котельникова, робить незручною інтерполяцію на їхній основі. Найбільше поширення при інтерполяції відліків для цифрової модуляції одержав SQRT-варіант **фільтра із косинусоїдальним згладжуванням АЧХ**.

Фільтр, який використовується для інтерполяцій, визначає форму спектра КАМп – сигналу, тому його називають **формуючим фільтром**, а сам процес інтерполяції – **формуванням спектра**.

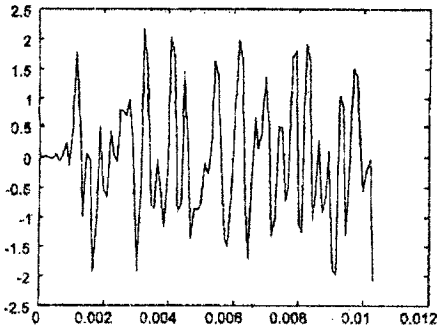


Рис. 5.44

Стрибокподібну зміну параметрів модуляції можна розглядати як використання формуючого фільтра з прямокутною імпульсною характеристикою, тривалість якої дорівнює символічному інтервалу. Повторимо формування 16-позиційного квадратурно-маніпульованого сигналу (див. рис. 5.41), використовуючи цього разу формуючий фільтр із косинусоїдальним згладжуванням АЧХ (рис. 5.44).

Порівнюємо спектри потужності сигналів, наведених на рис. 5.41 і 5.44, щоб наочно показати вплив формуючого фільтра. На рис. 5.45 наведені спектри щільності потужності сигналів із квадратурною маніпуляцією при відсутності (верхній графік) і наявності (нижній графік) формуючого фільтра.

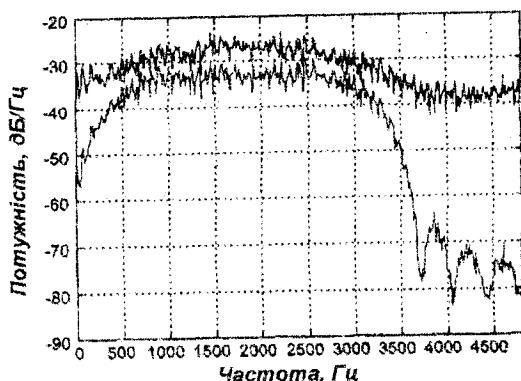


Рис. 5.45

З графіків видно, що при використанні формуючого фільтра спектр сигналу виявляється значно компактніше. При прийомі такого сигналу як ФНЧ необхідно використовувати такий же фільтр, як для формування спектра. Послідовне використання двох SQRT-фільтрів із косинусоїдальним згладжуванням дає результуючу імпульсну характеристику, для якої значення рівні нулю в точках, зсунутих на ціле число символів щодо піка. Це дозволяє при правильному виборі моментів узяття відліків усунути перешкоди від сусідніх символів (так звану **міжсимвольну інтерференцію**).

На рис. 5.46, а показана "глазкова" діаграма для даного сигналу. "Глазкова" діаграма являє собою "осцилограму" сигналу, побудовану при тривалості "прямого ходу розгорнення", рівному одному символічному такту, і нескінченному "часі післясвітіння екрана". У точках оптимальної дискретизації лінії на такій діаграмі утворюють вузькі пучки, вільний простір між якими за формою нагадує розкрите око. У даному випадку видно, що вибирати елементи з вектора потрібно, по-

чинаючи із першого. Оскільки сигнал є комплексним, наведені окремі графіки для його дійсної та уявної частин. На рис. 5.46, б наведена діаграма розсіювання, отримана при прийомі даного сигналу. Завдяки використанню погоджених один із одним фільтрів на передавальній і приймальній сторонах розкид точок виявляється значно менше, ніж на наведеному раніше рис. 5.42.

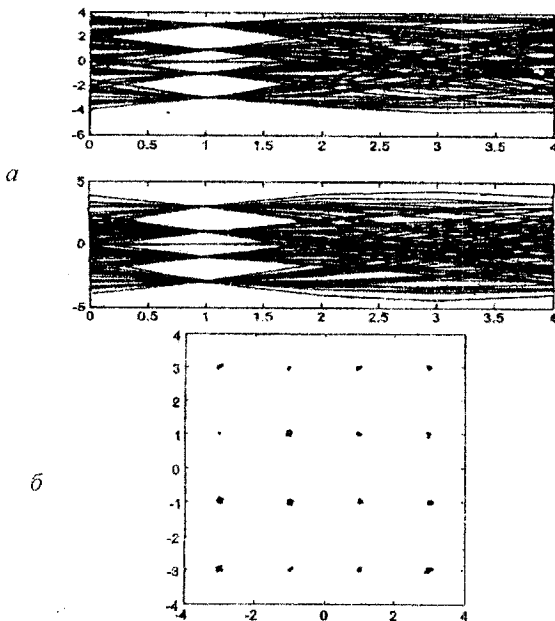


Рис. 5.46

При проходженні сигналу через канал зв'язку, що має *частотну дисперсію*, яка вносить різну групову затримку на різних частотах, символи виявляються “розмазаними” у часі і “наповзають” один на одного. У цьому випадку усунути міжсимвольну інтерференцію цілком не вдається. Щоб мінімізувати її, використовують *адаптивні фільтри*, параметри яких автоматично підбудовуються під характеристики оброблюваного сигналу.

5.6.9 Широтно-імпульсна модуляція

При широтно-імпульсній модуляції (ШІМ) як несуче коливання використовується періодична послідовність прямокутних імпульсів, а інформаційним параметром, пов'язаним із дискретним модулюючим сигналом, є тривалість цих імпульсів (рис. 5.47).

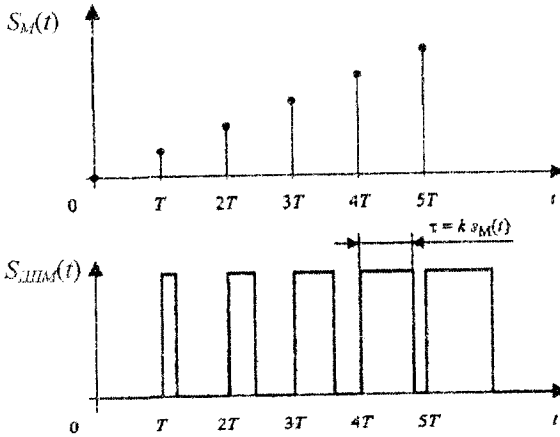


Рис. 5.47

Періодична послідовність прямокутних імпульсів однакової тривалості має постійну складову, обернено пропорційну **шпаруватості імпульсів**, тобто прямо пропорційну їхній тривалості. Пропустивши імпульси через ФНЧ із частотою зрізу, значно меншою, чим частота проходження імпульсів, цю постійну складову можна легко виділити, одержавши постійну напругу. Якщо тривалість імпульсів буде різною, ФНЧ виділить повільно мінливу напругу, що відслідковує закон зміни тривалості імпульсів. Таким чином, за допомогою ШІМ можна створити нескладний ЦАП: значення відліків сигналу кодуються тривалістю імпульсів, а ФНЧ перетворить імпульсну послідовність у плавно мінливий сигнал. Саме такі ЦАП використовуються, наприклад, для реалізації налаштування на канали, регулювань гучності, яскравості тощо у більшості сучасних телевізорів: команди від кнопок керувань змінюють шпаруватість імпульсів, які виробляє керуючий процесор, а кола, що згладжують, формують із цих імпульсів постійні напруги для керування аналоговими пристроями.

Число рівнів напруги, сформованих таким ЦАП, визначається числом можливої тривалості імпульсів. Якщо імпульси формуються у цифровому вигляді, крок зміни тривалості визначається частотою дискретизації імпульсної послідовності. Відношення цієї частоти до частоти дискретизації сформованого сигналу дасть можливе число рівнів.

Наприклад, якщо ми захочемо проекспериментувати із ШІМ, використовуючи звукову карту комп'ютера, нам будуть доступні всього шість рівнів (максимальна частота дискретизації, яку забезпечує звукова карта, — 48 кГц; мінімально прийнятна частота дискретизації зву-

кового сигналу – 8 кГц; звідси отримаємо $48 / 8 = 6$ рівнів). Для реальної роботи, звичайно, обмаль, а для експериментів підійде.

Спробуємо сформувати таким шляхом синусоїду. Подивимося, як виглядають вихідні відліки синусоїди (а) і відповідний ШИМ-сигнал (б) (рис. 5.48).

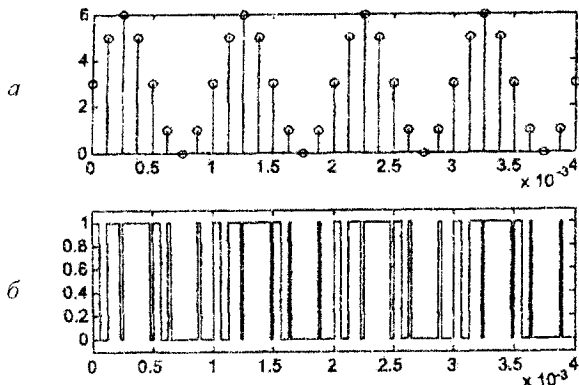


Рис. 5.48

Тепер глянемо на спектр ШИМ-сигналу (рис. 5.49). Як бачимо, найбільш інтенсивна спектральна складова (не враховуючи складової із нульовою частотою) дійсно має частоту вихідної синусоїди – 1 кГц. Виділимо її, пропустивши ШИМ-сигнал через ФНЧ із частотою зрізу, рівною частоті Найквіста, тобто 4 кГц (рис. 5.50). Результат, як бачимо, дійсно нагадує синусоїду, хоча спотворення, що залишилися, добре помітні на око – шести рівнів квантування явно обмаль.

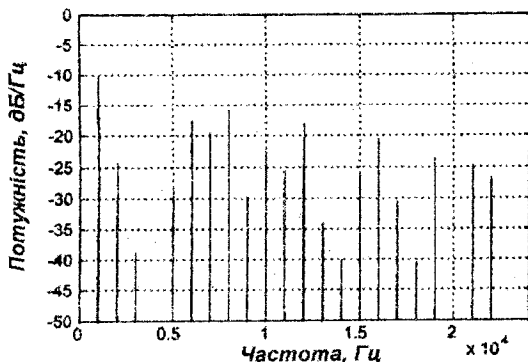


Рис. 5.49

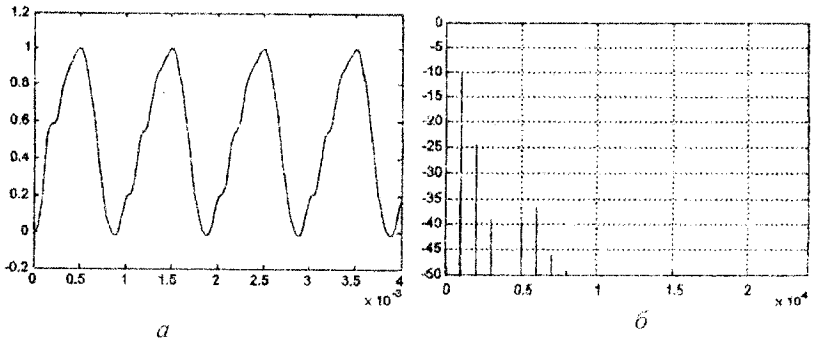


Рис. 5.50

Контрольні питання

1. Поясніть фізичний зміст модуляції та демодуляції сигналів з точки зору перетворення інформації.
2. Охарактеризуйте особливості видів амплітудної модуляції.
3. Позитивні та негативні сторони амплітудних видів модуляції.
4. Дайте визначення обвідної та спектра АМ-сигналів.
5. Опишіть методи та засоби демодуляції АМ-сигналів.
6. Охарактеризуйте методи модуляції із заглушеною несучою та їх застосування.
7. Вкажіть особливості полярної модуляції та її використання.
8. Дайте визначення кутової модуляції, порівняйте частотну та фазову модуляції.
9. Наведіть відмінні характеристики спектра сигналів із кутовою модуляцією у порівнянні з АМ.
10. Поясніть особливості методів демодуляції сигналів із кутовою модуляцією.
11. Опишіть методи та засоби здійснення зовнішньої частотної модуляції та її застосування.
12. Вкажіть на особливості квадратурної модуляції та її застосування у сучасних засобах телекомунікацій.
13. Порівняйте завадостійкість сучасних методів цифрової модуляції сигналів.
14. Опишіть імпульсні методи модуляції і їх застосування у цифрових системах телекомунікацій.

6.1 Мета кодування. Основні поняття і визначення

Як уже зазначалося, під кодуванням у широкому сенсі розуміють процес перетворення повідомлень у сигнал.

Як при передачі, так і при збереженні й опрацюванні інформації значні переваги дає дискретна форма представлення сигналів. Тому у тих випадках, коли первинні сигнали інформаційних систем є безперервними, відбувається, як правило, попереднє перетворення їх у дискретні сигнали. У зв'язку з цим термін “кодування” відносять до дискретних сигналів і під кодуванням у вузькому сенсі розуміють відображення дискретних повідомлень сигналами у вигляді певних сполучень символів.

Сукупність правил, відповідно до яких проводяться ці операції, називають *кодом*.

Отже, будемо вважати, що джерело видає деяке дискретне повідомлення a , яке можна розглядати як послідовність елементарних повідомлень a_i ($i = 1, 2, \dots, l$). Ці елементарні повідомлення будемо називати символами повідомлень, а їх сукупність $\{a_i\}$ – алфавітом джерела. Кодування полягає в тому, що послідовність символів джерела a замінюється послідовністю кодових символів – кодовою комбінацією (кововим словом).

Кодування переслідує декілька цілей. Перша з них полягає в тому, щоб представити повідомлення в такій системі символів, що забезпечувала б простоту і надійність апаратної реалізації інформаційних пристроїв та їх необхідну ефективність. Друга мета кодування полягає у тому, щоб забезпечити найкраще узгодження властивостей джерела повідомлень із властивостями каналу зв'язку. Шляхом такого узгодження домагаються виграшу в часі передачі, тобто підвищення ефективності системи. Нарешті, при наявності завад кодування може забезпечити досить високу достовірність передачі або обробки інформації. Але необхідно мати на увазі, що кодування, яке забезпечує зміну структури сигналів, жодною мірою не повинно змінювати кількість інформації, що міститься в початковому повідомленні.

Елементарними сигналами, які представляють символи кодової комбінації, у технічних інформаційних системах слугують поодинокі імпульси постійного струму (відеоімпульси) або змінного струму (радіо-

імпульси). Елементарним сигналом може слугувати пауза між імпульсами або комбінації паузи й імпульсу і т.д. Ці сигнали повинні розрізнятися за одним або декількома параметрами, які називаються кодовими ознаками. Як кодові ознаки застосовуються такі параметри, як величина, полярність, час (тривалість або фаза імпульсів), частота заповнення імпульсу тощо.

Загальна кількість символів, що складають кодову комбінацію, називається **значністю** або **довжиною коду** n . Кількість значень кодових ознак, використовуваних у кодових комбінаціях, називається **основою коду** m .

На рис. 6.1, *a* наведений приклад кодової комбінації з довжиною коду $n = 5$. Як імпульсна ознака тут використана величина імпульсів. Основа коду (кількість значень імпульсної ознаки) $m = 3$. На рис. 6.1, *б* наведена кодова комбінація з $n = 5$ і $m = 3$, у якій як імпульсна ознака використовується тривалість імпульсів.

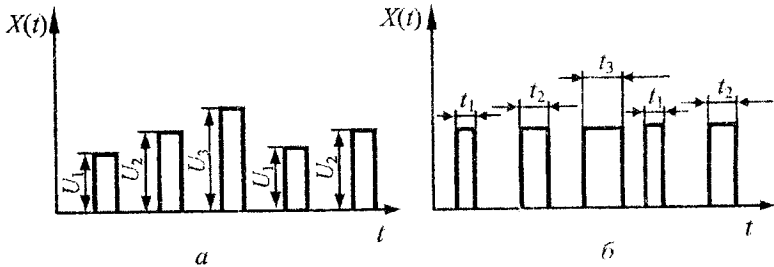


Рис. 6.1

За умовами побудови кодових комбінацій коди діляться на рівномірні і нерівномірні. У рівномірних кодах усі повідомлення передаються кодовими групами з однаковою кількістю елементів $n = \text{const}$. Так, наприклад, телеграфний код Бодо є рівномірним кодом з $n = 5$.

При використанні нерівномірних кодів різні повідомлення можуть передаватися кодовими групами, що містять неоднакову кількість елементів ($n = \text{var}$). Типовим представником цієї групи є код Морзе. Рівномірний код має більші можливості з погляду забезпечення заводозахищеності передачі, тому що втрата елементів або виникнення нових елементів у кодових комбінаціях із $n = \text{const}$ можуть бути легко виявлені. Нерівномірні коди можуть забезпечити найбільшу економічність побудови кодів і найбільшу швидкість передачі повідомлень. Такі коди використовуються при так званому статистичному кодуванні.

Водночас нерівномірні коди менш заводозахищені, чим рівномірні. Втрата або виникнення нових елементів у комбінації в результаті

дії завад можуть привести до створення нової помилкової комбінації, яка сприйматиметься на приймальній стороні як істинна. Нерівномірні коди вимагають при передачі або спеціальних роздільних символів, що вказують кінець однієї і початок іншої кодової комбінації (код Морзе вимагає наявності роздільного символу), або ж повинні будуватися так, щоб ніяка кодова комбінація не стала початком іншої.

За кількістю різних символів m у кодових комбінаціях розрізняють коди:

- одиничні: $m = 1$;
- двійкові: $m = 2$;
- багатопозиційні: $m > 2$.

В одиничному коді використовуються однакові символи. Кодові комбінації відрізняються одна від іншої лише кількістю символів (імпульсів). Такі коди називають ще число-імпульсними. Ілюстрація число-імпульсного коду наведена на рис. 6.2.

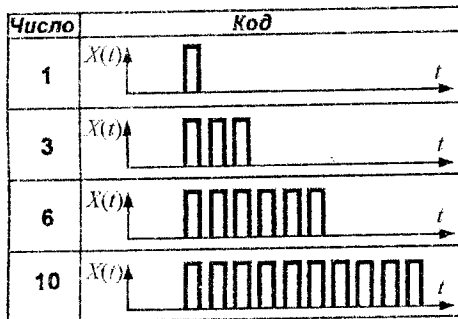


Рис. 6.2

Одиничний код відрізняється своєю простотою. Однак внаслідок того, що він нерівномірний, заводозахисненість його низька. Крім того, при передачі великої кількості повідомлень відбувається зміна в широких межах довжини коду, що викликає певні незручності. У зв'язку з цим одиничний код практично не використовується для передачі інформації з каналу зв'язку, а використовується лише при проміжних перетвореннях сигналів на передавальній і приймальній сторонах.

Найбільше поширення одержали двійкові коди. Це обумовлено наступним. Формування кодових сигналів та їх дешифрування проводяться за допомогою релейних пристроїв, здатних займати низку стійких станів. Кількість таких станів визначається основою коду. Очевидно, що найпростішими релейними пристроями є пристрої з двома станами.

До такого типу пристроїв належить більшість електромагнітних, електронних, магнітних та інших типів безконтактних реле. Крім того, потрібно також враховувати простоту зберігання інформації та виконання арифметичних і логічних операцій при двійковому кодуванні.

Багатопозиційні коди поки не знайшли широкого застосування в інформаційних системах.

За формою представлення в каналі передачі розрізняють послідовні і паралельні коди. При послідовній формі елементарні сигнали, що складають кодову комбінацію, посиляються в каналі передачі послідовно в часі. Як правило, вони розділені між собою певним часовим інтервалом.

При паралельній формі елементарні сигнали посиляються одночасно по декількох електричних колах, кількість яких відповідає кількості елементів коду.

Паралельна форма представлення коду, хоч і пов'язана з меншими втратами часу для передачі повідомлення, використовується для передачі інформації по каналу зв'язку рідко, тому що потребує значних матеріальних витрат на багатопровідні лінії зв'язку. Практично паралельна форма коду при передачі інформації по однопровідній лінії зв'язку використовується лише в тих випадках, коли як імпульсна ознака застосовується частота і на стороні прийому елементи кодової комбінації можна легко розділити за допомогою частотних фільтрів.

Паралельна форма представлення коду часто використовується при перетворенні аналогових величин в код і зворотних перетвореннях, в пристроях пам'яті, реєстрації, при логічній і математичній обробці інформації, коли важливу роль грає швидкодія.

За можливістю знаходження і виправлення помилок розрізняють прості (примітивні) і коригувальні коди. У простих кодах помилка в прийомі хоча б одного елемента кодової комбінації приводить до неправильної реєстрації повідомлення, що передається. Коригувальні коди дозволяють знаходити і виправляти помилки в кодових комбінаціях.

За основними законами кодового утворення коди можна розділити на два класи: комбінаторні (нечислові) і арифметичні (числові).

Комбінаторні коди будуються на законах теорії з'єднань. Прикладом таких кодів є код за законом сполучень. У цьому коді із m різноманітних символів утворюються кодові комбінації із $n < m$ символів. Довжина коду постійна і дорівнює n , можлива кількість кодових комбінацій $N = C_m^n$. Наприклад, код за законом сполучень може утворювати наступні комбінації із трьох елементів a, b, c по два елементи: ab, ac, bc .

Арифметичні коди засновані на відомих системах числення.

У технічних інформаційних системах використовуються здебільшого арифметичні (цифрові) коди.

6.2 Рівномірні прості цифрові коди

Системи числення, на основі яких можуть будуватись цифрові коди, діляться на непозиційні і позиційні. Із непозиційних систем можна навести як приклад римську систему числення. Значення символу в цих системах не залежить від його положення в ряду символів, що складають повідомлення, яке передається. Найбільш розповсюдженим при кодуванні на сьогодні є позиційний принцип утворення системи числення. При такому принципі значення кожного символу залежить від його положення — позиції в ряду символів, які складають число. Значення одиниці кожного наступного розряду більше значення одиниці попереднього розряду в m раз, де m — основа системи числення. При цьому будь-яке n -розрядне число N з основою m може бути представлено у вигляді суми:

$$N = \sum_{i=0}^{n-1} e_i m^i, \quad (6.1)$$

де e_i — значення розрядного коефіцієнта i -го розряду.

Кількість можливих значень розрядного коефіцієнта кожного розряду дорівнює основі вибраної системи числення.

Як приклад розглянемо чотирирозрядне десяткове число 4752. Це число може бути представлено у вигляді суми:

$$4752 = 2 \cdot 10^0 + 5 \cdot 10^1 + 7 \cdot 10^2 + 4 \cdot 10^3.$$

Кодові комбінації, що відповідають простому цифровому коду, містять кількість елементів, яка відповідає кількості розрядів числа N , що виражається даним кодом. Кожний елемент кодової комбінації може приймати m різних значень імпульсної ознаки.

Максимально можлива кількість кодових комбінацій визначається виразом:

$$N_{\max} = m^n. \quad (6.2)$$

На практиці найбільш широко використовуються двійкові коди, тобто коди, які базуються на двійковій системі числення. Математичний запис двійкового коду наступний:

$$N = \sum_{i=0}^{n-1} l_i 2^i. \quad (6.3)$$

Тут розрядні коефіцієнти l_i можуть приймати значення 0 та 1.

Максимально можлива кількість кодових комбінацій простого двійкового коду визначається виразом:

$$N_{\max} = 2^n. \quad (6.4)$$

На рис. 6.3 показана двійкова кодова комбінація довжиною $n = 7$. Як імпульсна ознака тут використовується полярність. Імпульсам додатної полярності відповідає 1, а імпульсам від'ємної полярності – 0.

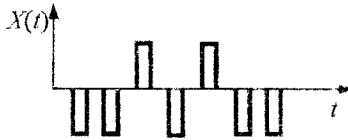


Рис. 6.3

Кодова комбінація містить дві одиниці (два додатних імпульси), але числовий вираз (вага) кожної одиниці визначається її номером (порядком). Оскільки нумерація розрядів йде справа, то кодова комбінація, показана на рис. 6.3, виражає число (в десятковому записі):

$$N = 0 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 + 0 \cdot 2^3 + 1 \cdot 2^4 + 0 \cdot 2^5 + 0 \cdot 2^6 = 20.$$

Двійковий код, який широко використовується при передачі, зберіганні та обробці інформації, незручний при вводі і виводі інформації, оскільки оператору важко оперувати з незвичними двійковими числами. Крім того запис чисел у двійковій системі числення громіздкий. Тому крім двійкової отримали розповсюдження системи з основою, рівною цілому степеню двійки (вісімкова, шістнадцяткова), які з одного боку легко зводяться як до двійкової так і до десяткової, а з другого – дають більш компактний запис.

Для позначення будь-якого числа у вісімковій системі числення використовується вісім цифр: 0, 1, 2, 3, 4, 5, 6, 7. Переведення чисел із вісімкової системи у двійкову зводиться до заміни кожної вісімкової цифри рівним їй трирозрядним двійковим числом. Наприклад, для вісімкового числа 3275 отримаємо:

$$\begin{array}{cccc} 3 & 2 & 7 & 5 \\ 011 & 010 & 011 & 101 \end{array}$$

Для представлення шістнадцяткових цифр використовуються цифри від 0 до 9 і літери латинського алфавіту від A до F. Переведення

чисел із шістнадцятковою системою в двійкову здійснюється шляхом заміни кожної шістнадцяткової цифри чотирирозрядним двійковим числом. Позначення шістнадцяткових цифр і представлення їх двійковими числами наведені у таблиці 6.1.

Таблиця 6.1

| Число | | | | | |
|------------|----------------|----------|------------|----------------|----------|
| десятькове | шістнадцяткове | двійкове | десятькове | шістнадцяткове | двійкове |
| 0 | 0 | 0000 | 8 | 8 | 1000 |
| 1 | 1 | 0001 | 9 | 9 | 1001 |
| 2 | 2 | 0010 | 10 | A | 1010 |
| 3 | 3 | 0011 | 11 | B | 1011 |
| 4 | 4 | 0100 | 12 | C | 1100 |
| 5 | 5 | 0101 | 13 | D | 1101 |
| 6 | 6 | 0110 | 14 | E | 1110 |
| 7 | 7 | 0111 | 15 | F | 1111 |

Вісімкова і шістнадцяткова системи числення використовуються при ввіді-виводі інформації, а також при програмуванні.

6.3 Складені коди

Складені коди базуються на складених системах числення, які мають дві і більше основ. При такому кодуванні числа, задані в системі з деякою основою q зображуються за допомогою цифр іншої системи числення з основою $p < q$.

Серед складених кодів найбільше застосування знайшли двійково-десятькові коди. Такі коди зазвичай використовуються як проміжні при переведенні десяткових кодів у двійкові і навпаки, наприклад, при ввіді в обчислювальну машину даних, представлених в десятковій системі, або при виводі із машини інформації для реєстрації в десятковому коді.

У двійково-десятьковому коді основною системою числення є десяткова. Однак кожна цифра десяткового числа записується у вигляді чотирирозрядного двійкового числа (тетради). Для фіксації цифр десяткового числа найбільше практичне застосування знайшли чотириелементні двійкові вагові коди 8-4-2-1, 7-4-2-1, 5-1-2-1 і 2-4-2-1. Цифри в назві коду виражають вагу одиниць у відповідних двійкових розрядах.

У таблиці 6.2 наводиться запис десяткових цифр різними типами чотириелементних двійкових кодів.

Таблиця 6.2

| Десяткове число | Код 8-4-2-1 | Код 7-4-2-1 | Код 5-1-2-1 | Код 2-4-2-1 |
|-----------------|-------------|-------------|-------------|-------------|
| 0 | 0000 | 0000 | 0000 | 0000 |
| 1 | 0001 | 0001 | 0001 | 0001 |
| 2 | 0010 | 0010 | 0010 | 0010 |
| 3 | 0011 | 0011 | 0011 | 0011 |
| 4 | 0100 | 0100 | 0111 | 0100 |
| 5 | 0101 | 0101 | 1000 | 1011 |
| 6 | 0110 | 0110 | 1001 | 1100 |
| 7 | 0111 | 1000 | 1010 | 1101 |
| 8 | 1000 | 1001 | 1011 | 1110 |
| 9 | 1001 | 1010 | 1111 | 1111 |

Оскільки за допомогою чотирьох двійкових розрядів можна утворити 16 різноманітних комбінацій, а використовується тільки 10, то двійково-десятковий код володіє деякою надликовістю.

6.4 Рефлексні (відображені) коди

У простому двійковому коді при переході від зображення одного числа до зображення сусіднього старшого або сусіднього молодшого числа може відбуватись одночасна зміна цифр у декількох розрядах. Як випливає із таблиці 6.3, при переході від зображення числа 3 до зображення числа 4, відбувається одночасна зміна в трьох розрядах, а при переході від зображення числа 7 до зображення числа 8 відбувається одночасна зміна цифр в чотирьох розрядах.

Таблиця 6.3

| Десяткове число | Двійковий код | Код Грея | Десяткове число | Двійковий код | Код Грея |
|-----------------|---------------|----------|-----------------|---------------|----------|
| 0 | 0000 | 0000 | 8 | 1000 | 1100 |
| 1 | 0001 | 0001 | 9 | 1001 | 1101 |
| 2 | 0010 | 0011 | 10 | 1010 | 1111 |
| 3 | 0011 | 0010 | 11 | 1011 | 1110 |
| 4 | 0100 | 0110 | 12 | 1100 | 1010 |
| 5 | 0101 | 0111 | 13 | 1101 | 1011 |
| 6 | 0110 | 0101 | 14 | 1110 | 1001 |
| 7 | 0111 | 0100 | 15 | 1111 | 1000 |

Це може стати джерелом значних помилок при деяких методах кодування неперервних повідомлень, наприклад, при перетворенні кута повороту вала у двійковий код. Особливістю таких перетворювачів є те,

що при переході від зображення одного числа до зображення наступного числа (більшого або меншого) може мати місце неоднозначність відліку в тих розрядах, де відбувається заміна цифр. Наприклад, при переході від зображення числа 7 до зображення числа 8 відбувається заміна цифр у всіх розрядах, тобто може мати місце неоднозначність відліку в будь-якому розряді і навіть зразу у всіх розрядах. Внаслідок цього, зокрема, може бути такий випадок, коли замість числа 1000 буде зафіксовано число 1111, тобто, буде мати місце неприпустимо велика похибка.

Для усунення цього явища використовуються спеціальні двійкові коди, в яких при переході від зображення одного числа до зображення наступного сусіднього числа змінюється значення цифри тільки одного розряду. Внаслідок цього похибка за рахунок неоднозначності зчитування не перевищує одиниці зображуваного числа. До таких кодів відноситься код Грея, який ще називають рефлексним (відображенням) кодом (див. табл. 6.3).

Код Грея є непоозиційним кодом, вага одиниці не визначається номером розряду. У цьому коді можна виділити осі симетрії (осі “відображення”), відносно яких спостерігається ідентичність елементів у деяких розрядах. Так, наприклад, має місце симетрія відносно осі, проведеної між числами 7 і 8. У комбінаціях, симетричних відносно цієї осі, ідентичні три символи молодших розрядів. Зазначена особливість стала основою для введення терміну “рефлексний код” (англ. *to reflect* – відображати).

Вага одиниці в коді Грея за абсолютною величиною є j -му розряді визначається виразом $\sum_{i=0}^j 2^i$, причому знак членів, що підсумовуються, додатний для всіх непарних одиниць в числі, записаному в коді Грея (рахувати зліва направо), і від’ємний для всіх парних одиниць. Наприклад, значення числа 1101101, записаного в коді Грея, буде:

$$\begin{aligned} [1101101]_Г &= \sum_{i=0}^6 2^i + \sum_{i=0}^5 2^i + \sum_{i=0}^4 2^i + \sum_{i=0}^3 2^i + \sum_{i=0}^2 2^i + \sum_{i=0}^1 2^i + \sum_{i=0}^0 2^i = \\ &= 2^6 - 3^3 + 2^0 = 57. \end{aligned}$$

6.5 Завадостійке кодування. Класифікація завадостійких кодів

Завадостійкі коди – один із найбільш ефективних засобів забезпечення високої правдивості передачі дискретної інформації. Створена спеціальна теорія завадостійкості кодування, яка швидко розвивається останнім часом.

Бурхливий розвиток теорії завадостійкого кодування пов'язаний з вторгненням автоматизованих систем, у яких обробка інформації, що приймається, здійснюється без участі людини. Використання для обробки інформації електронних цифрових обчислювальних машин висуває дуже високі вимоги до достовірності передачі повідомлень.

К. Шеннон сформулював теорему для випадку передачі дискретної інформації по каналу із завадами (див. п. 8.3), яка підтверджує, що ймовірність помилкового декодування приймальних сигналів може бути забезпечена скільки завгодно малою шляхом вибору відповідного методу кодування сигналів. У теоремі Шеннона не говориться про те, як потрібно будувати завадостійкі коди. Але в ній вказується на принципову можливість кодування, при якому може бути забезпечена скільки завгодно висока достовірність передачі. Це виявилось стимулом до розробки завадостійких кодів.

Під завадостійкими кодами розуміють коди, що дозволяють знаходити або знаходити і виправляти помилки, які виникають в результаті впливу завад.

Завадостійкість кодування забезпечується за рахунок введення надлишковості у кодові комбінації, тобто за рахунок того, що не всі символи в кодових комбінаціях використовуються для передачі інформації.

Всі завадостійкі коди можна поділити на два основні класи (рис. 6.4): блочні та неперервні (рекурентні або ланцюгові).

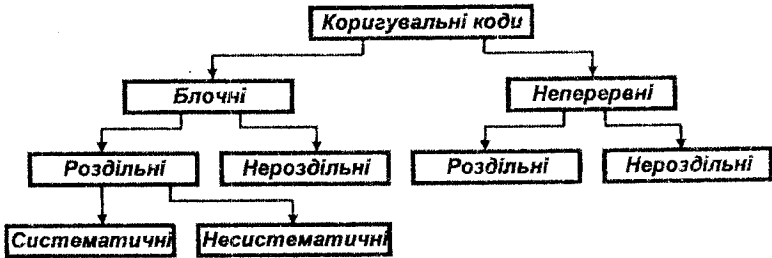


Рис. 6.4

У блочних кодах кожному повідомленню (або елементу повідомлення) з'являється кодова комбінація (блок) із визначеної кількості сигналів. Блоки кодуються і декодуються окремо один від одного.

Блочні коди можуть бути рівномірними, коли довжина кодових комбінацій n постійна, або нерівномірними, коли n непостійна.

Нерівномірні завадостійкі коди не отримали практичного застосування через складність їх технічної реалізації.

У неперервних кодах введення перебільшення в послідовність вхідних символів здійснюється без розбиття його на окремі блоки. Процеси кодування і декодування у безперервних кодах мають також безперервний характер. Цей клас кодів з'явився недавно і не одержав поки що широкого розвитку.

Як блочні, так і безперервні коди, залежно від методів внесення надлишковості поділяються на роздільні і нероздільні. У роздільних кодах чітко розмежована роль окремих символів. Одні – інформаційними, інші – перевірними і слугують для виявлення і виправлення помилок. Роздільні блочні коди називаються звичайно (n, k) -кодами, де n – довжина кодових комбінацій, k – кількість інформаційних символів у комбінаціях. Нероздільні коди не мають чіткого розподілу кодової комбінації на інформаційні і перевірні символи. Цей клас кодів поки що нечислений.

Роздільні блочні коди діляться, у свою чергу, на несистематичні і систематичні. Несистематичні роздільні коди будуються таким чином, що перевірні символи визначаються як сума підблоків довжиною l , на які поділяється блок інформаційних символів.

Більшість відомих роздільних кодів складають систематичні коди. У цих кодів перевірні символи визначаються при проведенні лінійних операцій над певними інформаційними символами. У випадку двійкових кодів кожний перевірний символ вибирається таким, щоб його сума за модулем два з певними інформаційними символами дорівнювала нулю. Декодування зводиться до перевірки на парність певних груп символів. У результаті таких перевірок дається інформація про наявність помилок, а в разі потреби – про позицію символів, де є помилки.

6.6 Основні принципи завадостійкого кодування

При подальшому розгляді завадостійких кодів будемо вважати їх блочними і рівномірними.

Для з'ясування ідеї завадостійкого кодування розглянемо двійковий код, що знайшов на практиці найбільш широке застосування.

Нагадаємо, що двійковий код – це код із основою $m = 2$.

Кількість розрядів n у кодовій комбінації прийнято називати довжиною або значимістю коду. Символи кожного розряду можуть приймати значення 0 і 1. Кількість одиниць у кодовій комбінації називають вагою кодової комбінації і позначають w .

Наприклад, кодова комбінація 100101100 характеризується значністю $n = 9$ і вагою $w = 4$.

Ступінь відмінності будь-яких двох кодових комбінацій даного коду характеризується відстанню між кодами d . Вона виражається числом позицій або символів, у яких комбінації відрізняються одна від іншої, і визначається як вага суми за модулем два цих кодових комбінацій. Наприклад, для визначення відстані між комбінаціями 100101100 і 110110101 необхідно підсумувати їх за модулем два:

$$\begin{array}{r} 100101100 \\ + 110110101 \\ \hline 010011001 \end{array}$$

Отримана в результаті підсумовування нова кодова комбінація характеризується вагою $w = 4$. Отже, відстань між вихідними кодовими комбінаціями $d = 4$.

Помилки, внаслідок впливу завад, виявляються в тому, що в одному або декількох розрядах кодової комбінації нулі переходять в одиниці і, навпаки, одиниці переходять у нулі. У результаті створюється нова помилкова кодова комбінація. Якщо помилки відбуваються тільки в одному розряді кодової комбінації, то вони називаються однократними. При наявності помилок у двох, трьох і т.д. розрядах помилки називають двократними, трікратними і т.д.

Експериментальні дослідження каналів зв'язку показали, що помилки символів при передачі по каналу зв'язку, як правило, групуються в пачки різної тривалості. Під пачкою помилок розуміють ділянку послідовності, що починається і кінчається помилково прийнятими символами. Всередині пачки можуть бути і правильно прийняті елементи.

Для зазначення місць у кодовій комбінації, де є спотворення символів, використовується вектор помилки \vec{e} . Вектор помилки n -розрядного коду — це n -розрядна комбінація, одиниці в якій указують положення спотворених символів кодової комбінації. Наприклад, якщо для n -розрядного коду вектор помилки має вигляд $\vec{e} = 01100$, то це означає, що мають місце помилки в третьому і четвертому розрядах кодової комбінації.

Вага вектора помилки w_e характеризує кратність помилки. Сума за модулем два для спотвореної кодової комбінації та вектора помилки дає вихідну неспотворену комбінацію.

Як уже зазначалося, завадостійкість кодування забезпечується за рахунок введення надлишковості в кодові комбінації. Це означає, що з n символів кодової комбінації для передачі інформації використовується $k < n$ символів. Отже, із загальної кількості $N_0 = 2^n$ можливих кодових комбінацій для передачі інформації використовується тільки $N_0 = 2^k$ комбінацій. Відповідно до цього вся множина $N_0 = 2^n$ можливих

кодівих комбінацій ділиться на дві групи. У першу групу входить множина $N_0 = 2^k$ дозволених комбінацій, друга група містить у собі множину $N_0 - N = 2^n - 2^k$ заборонених комбінацій.

Якщо на приймальній стороні встановлено, що прийнята комбінація відноситься до групи дозволених, то вважається, що сигнал прийшов без спотворень. В іншому випадку робиться висновок, що прийнята комбінація спотворена. Однак це справедливо лише для таких завад, коли виключена можливість переходу одних дозволених комбінацій в інші.

У загальному випадку кожна з N дозволених комбінацій може трансформуватися у будь-яку з N_0 можливих комбінацій, тобто усього є $N \cdot N_0$ можливих варіантів передачі (рис. 6.5), із них N варіантів безпомилкової передачі (на рис. 6.5 позначені суцільними лініями), $N \cdot (N - 1)$ варіантів переходу в інші дозволени комбінації (на рис. 6.5 позначені пунктирними лініями) і $N \cdot (N_0 - N)$ варіантів переходу в заборонені комбінації (на рис. 6.5 позначені штрих-пунктирними лініями).

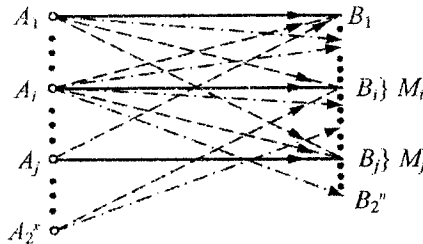


Рис. 6.5

Таким чином, не всі спотворення можуть бути виявлені. Частка виявлених помилкових комбінацій складає:

$$\frac{N(N_0 - N)}{N \cdot N_0} = 1 - \frac{N}{N_0}. \quad (6.5)$$

Для використання даного коду як виправного множина заборонених кодівих комбінацій розбивається на N підмножин, які не пересікаються. Кожна з підмножин M_k ставиться у відповідність одній з дозволених комбінацій.

Якщо прийнята заборонена комбінація належить підмножині M_i , то вважається, що передана комбінація A_i (див. рис. 6.5).

Помилка буде виправлена у тих випадках, коли отримана комбінація дійсно утворилася з комбінації A_i . Таким чином, помилка виправляється у $(N_0 - N)$ випадках, рівних кількості заборонених комбінацій. Частка виправлених помилкових комбінацій від загального числа виявлених помилкових комбінацій складася:

$$\frac{N_0 - N}{N(N_0 - N)} = \frac{1}{N}. \quad (6.6)$$

Спосіб розбиття на підмножини залежить від того, які помилки повинні виправлятися даним кодом.

6.7 Зв'язок коригувальної здатності коду з кодовою відстанню

Для оцінювання ступеня відмінності між двома довільними комбінаціями даного коду використовується характеристика, що одержала назву відстані між комбінаціями.

Найменша відстань між дозволеними кодовими комбінаціями d_{\min} — дуже важлива характеристика коду, тому що саме вона характеризує його коригувальні здатності.

Розглянемо це на конкретних прикладах.

Нехай необхідно побудувати код, що виявляє всі помилки кратності t і нижче. Побудувати такий код — означає із множини N_0 можливих вибрати N дозволених комбінацій так, щоб будь-яка з них в сумі за модулем два з будь-яким вектором помилок із вагою $w_i \leq t$ не дала б у результаті ніякої іншої дозволеної комбінації. Для цього необхідно, щоб найменша кодова відстань задовольняла умові:

$$d_{\min} \geq t + 1. \quad (6.7)$$

Як приклад розглянемо код із значністю $n = 3$. Всі можливі комбінації такого коду подані у таблиці 6.4.

Таблиця 6.4

| A_1 | A_2 | A_3 | A_4 | A_5 | A_6 | A_7 | A_8 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |

Матриця відстаней між кодовими комбінаціями має вигляд (табл. 6.5).

Таблиця 6.5

| | A_1 | A_2 | A_3 | A_4 | A_5 | A_6 | A_7 | A_8 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| A_1 | 0 | 1 | 1 | 2 | 1 | 2 | 2 | 3 |
| A_2 | 1 | 0 | 2 | 1 | 2 | 1 | 3 | 2 |
| A_3 | 1 | 2 | 0 | 1 | 2 | 3 | 1 | 2 |
| A_4 | 2 | 1 | 1 | 0 | 3 | 2 | 2 | 1 |
| A_5 | 1 | 2 | 2 | 3 | 0 | 1 | 1 | 2 |
| A_6 | 2 | 1 | 3 | 2 | 1 | 0 | 2 | 1 |
| A_7 | 2 | 3 | 1 | 2 | 1 | 2 | 0 | 1 |
| A_8 | 3 | 2 | 2 | 1 | 2 | 1 | 1 | 0 |

Для того щоб код забезпечував виявлення однократних помилок, необхідно з усієї множини $N_0 = 8$ можливих комбінацій вибрати як дозволені такі, відстань між якими була б не менше $d = 2$. Як видно з матриці відстаней, як приклад дозволених комбінацій у цьому випадку можна вибрати наступні:

$$A_1 = 000; \quad A_4 = 011; \quad A_6 = 101; \quad A_7 = 110.$$

Для виявлення двократних помилок найменша відстань повинна дорівнювати $d_{\min} = 3$. При цьому як приклад дозволених комбінацій можна вибрати:

$$A_1 = 000; \quad A_2 = 111.$$

Очевидна справедливості умови:

$$d < n. \quad (6.8)$$

Нехай тепер необхідно побудувати код, що забезпечує усунення однократних помилок. Вибираємо першу дозволена комбінація $A_1 = 000$. При наявності однократних помилок комбінація A_1 може перейти в одну з наступних заборонених комбінацій: $A_2 = 0010$, $A_3 = 010$ та $A_5 = 100$. Комбінації A_2 , A_3 і A_5 можна прийняти як підмножини заборонених комбінацій A_1 . Останнє означає, що у випадку прийому однієї з комбінацій цієї підмножини приймається рішення, що передана комбінація A_1 .

Нехай за другу дозволена комбінація вибирається комбінація, віддалена від першої на $d = 2$, наприклад, $A_4 = 011$. Їй повинна відповідати підмножина заборонених комбінацій $A_3 = 010$, $A_2 = 001$, $A_8 = 111$. Однак утворився перетин підмножин. При прийомі заборонених сигналів A_2 або A_3 не можна однозначно установити, який сигнал був переданий – A_1 або A_4 .

Якщо ж за другу дозволена комбінація вибрати комбінацію, що відстоїть від A_1 на $d = 3$, тобто комбінацію $A_8 = 111$, якій відповідає підмножина заборонених комбінацій $A_4 = 011$, $A_6 = 101$ та $A_2 = 110$, то у цьому випадку підмножини заборонених комбінацій не перетинаються. Отже, при $d = 3$ забезпечується усунення всіх однократних помилок.

У загальному випадку для усунення помилок кратності σ кодова відстань повинна задовольняти умові:

$$d_{\min} \geq 2\sigma + 1. \quad (6.9)$$

Аналогічно міркуючи, можна установити, що для виправлення всіх помилок кратності не більш σ і одночасного виявлення всіх помилок кратності не більш t (при $t \geq \sigma$) і кодова відстань повинна задовольняти умові:

$$d_{\min} \geq t + \sigma + 1. \quad (6.10)$$

6.8 Побудова кодів з заданою виправною здатністю

Дотепер при розгляді коригувальних кодів ми припускали, що їх значність n задана. Підвищення коригувальної здатності коду досягалося при зберіганні n за рахунок зменшення множини N дозволених комбінацій (або зменшення кількості k інформаційних символів). Найчастіше на практиці коди будуються в зворотному порядку: спочатку вибирається кількість інформаційних символів k , виходячи з об'єму алфавіту джерела, а потім забезпечується необхідна коригувальна здатність коду за рахунок додавання надлишкових символів.

Нехай відомий об'єм алфавіту джерела N . Необхідна кількість інформаційних символів:

$$k = \log_2 N.$$

Нехай також відомо повне число помилок E , які необхідно виправити. Завдання полягає в тому, щоб при заданих N та E визначити значність коду n , що володіє необхідними коригувальними можливостями.

Повне число помилкових комбінацій, які підлягають виправленню, дорівнює $E \cdot 2k = E \cdot N$. Оскільки кількість помилкових комбінацій дорівнює $N_0 - N$, то код забезпечує виправлення не більше $N_0 - N$ комбінацій. Отже, необхідну умову для можливості виправлення помилок можна записати у вигляді:

$$NE \leq N_0 - N,$$

звідки одержимо:

$$N_0 \geq (1 + E)N \quad (6.11)$$

або

$$N \leq \frac{2^n}{1 + E}. \quad (6.12)$$

Формула (6.12) виражає умову для вибору значності коду n .

Розглянемо окремі випадки. Якщо є помилки різної кратності, то насамперед необхідно забезпечити усунення однократних помилок, ймовірність появи яких найбільша. Можлива кількість векторів однократних помилок:

$$E = C_n^1 = n.$$

Тому залежність (6.12) прийме вигляд:

$$N = 2 \leq \frac{2^n}{1 + n}.$$

При побудові коду доцільно користуватися таблицею 6.6.

Таблиця 6.6

| n | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|------------------|------|---|-----|------|-----|----|------|------|
| $\frac{2n}{1+n}$ | 1,33 | 2 | 3,2 | 5,33 | 9,2 | 16 | 28,4 | 51,2 |

Потрібно при цьому мати на увазі, що код повинний також задовольняти вимозі:

$$d_{\min} \geq 3.$$

Якщо необхідно забезпечити усунення всіх помилок кратності від 1 до l , то потрібно врахувати, що кількість можливих помилок:

– однократних: $E_1 = C_n^1$;

– двократних: $E_2 = C_n^2$;

– l -кратних: $E_l = C_n^l$.

Загальне число помилок: $E = \sum_{i=1}^l C_n^i$.

При цьому залежність (6.12) буде мати вигляд:

$$N \leq \frac{2^n}{1 + \sum_{i=1}^l C_n^i}. \quad (6.15)$$

Умова (6.15) є нижньою оцінкою для довжини коригувального коду, тобто вона визначає необхідну мінімальну довжину коду n , що забезпечує виправлення помилок заданої кратності при відомому числі дозволених комбінацій N або числі інформаційних символів $k = \log_2 N$.

Ця ж умова є верхньою оцінкою для N або k , тобто визначає максимально можливе число дозволених комбінацій або інформаційних символів для коду довжини n , який забезпечує виправлення помилок заданої кратності.

6.9 Показники якості коригувального коду

Будь-який коригувальний код характеризується низкою показників: довжиною n , основою m , кількістю інформаційних символів k або надлишкових символів $\rho = n - k$, повним числом усіх можливих кодових комбінацій $N_0 = m^n$ (для двійкових кодів $N_0 = 2^n$), числом дозволених кодових комбінацій (потужністю коду) N , вагою кодової комбінації w , вагою вектора помилки w_i , кодовою відстанню d та ін.

Однак основним показником якості коригувального коду є його здатність забезпечити правильний прийом кодових комбінацій при наявності спотворень під впливом завад, тобто *завадостійкість коду*.

Кількісна оцінка завадостійкості коду може бути здійснена порізному. Завадостійкість коду характеризує його здатність забезпечити правильну передачу повідомлень в умовах впливу завад. Тому для кількісної оцінки завадостійкості коду доцільно використовувати ймовірність правильного прийому кодових комбінацій:

$$P_{пр} = 1 - P_{пом}, \quad (6.16)$$

де $P_{пом}$ – ймовірність помилкового прийому кодових комбінацій.

Якщо код не володіє коригувальними властивостями, то ймовірність помилкового прийому $P_{пом}$ буде дорівнює ймовірності спотворення кодових комбінацій P_k . Для коригувального коду $P_{пом} < P_k$. У реальних умовах $P_{пом} \leq 1$, тому більш зручним критерієм оцінювання завадостійкості коду є логарифмічна величина:

$$S_k = \lg \frac{1}{P_{\text{пом}}} = \lg \frac{1}{1 - P_{\text{пр}}}. \quad (6.17)$$

Іноді для оцінки якості коригувальних кодів користуються поняттям коефіцієнта виявлення:

$$K_{\text{виявл}} = \frac{P_{\text{в.пом}}}{P_k}, \quad (6.18)$$

де $P_{\text{в.пом}}$ – ймовірність виявлених помилок.

Це недостатньо повна характеристика якості коригувального коду. Нею доцільно користуватися в основному для оцінки кодів, призначених тільки для виявлення помилок.

Коригувальна можливість коду забезпечується за рахунок надлишковості, тобто подовження кодових комбінацій. При подовженні кодових комбінацій ускладнюється апаратура, збільшується час передачі й обробки інформації. Тому надлишковість також є важливою характеристикою коду.

Для оцінки надлишковості коду користуються поняттям коефіцієнта надлишковості:

$$K_{\text{надл}} = \frac{\rho}{n} = \frac{n - k}{n}, \quad (6.19)$$

де ρ – кількість надлишкових символів у кодовій комбінації.

6.10 Систематичні коди

На сьогодні найбільш широкий клас коригувальних кодів становлять систематичні коди. Ці коди відносяться до групи роздільних блочних кодів. Для систематичного коду сума за модулем два двох дозволених комбінацій також дає дозволена комбінацію. У теорії кодування широко використовується матричне представлення кодів.

Всі дозвалені кодові комбінації систематичного (n, k) -коду можна одержати, маючи k вихідних дозволених кодових комбінацій. Вихідні кодові комбінації повинні задовольняти наступним умовам.

1. У число вихідних комбінацій не повинна входити нульова.
2. Кодова відстань між будь-якими парами вихідних комбінацій не повинна бути менша $d_{\text{мін}}$.
3. Кожна вихідна комбінація, як і будь-яка ненульова дозволена комбінація, повинна містити кількість одиниць не менше $d_{\text{мін}}$.

4. Всі вихідні комбінації повинні бути лінійно незалежні, тобто жодна з них не може бути отримана шляхом підсумовування інших.

Вихідні комбінації можуть бути отримані з матриці, яка складається з k рядків та n стовпців:

$$P_{n,k} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} & b_{11} & b_{12} & \dots & b_{1p} \\ a_{21} & a_{22} & \dots & a_{2k} & b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} & b_{k1} & b_{k2} & \dots & b_{kp} \end{pmatrix}. \quad (6.20)$$

Тут символи перших k стовпців є інформаційними, а останніх p стовпців – перевірними.

Матрицю $P_{n,k}$ називають твірною.

Твірна матриця $P_{n,k}$ може бути представлена двома підматрицями – інформаційною U_k і перевірною H_p :

$$P_{n,k} = \left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1k} & b_{11} & b_{12} & \dots & b_{1p} \\ a_{21} & a_{22} & \dots & a_{2k} & b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} & b_{k1} & b_{k2} & \dots & b_{kp} \end{array} \right), \quad (6.21)$$

де

$$U_k = \left(\begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{array} \right); \quad H_p = \left(\begin{array}{cccc} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ b_{k1} & b_{k2} & \dots & b_{kp} \end{array} \right). \quad (6.22)$$

Для побудови твірної матриці зручно інформаційну матрицю брати у вигляді квадратної одиничної матриці:

$$U_k = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}. \quad (6.23)$$

При цьому перевірна підматриця H_p повинна будуватися з дотриманням наступних умов:

- кількість одиниць у рядку повинна бути не менше $d_{\min} - 1$.
- сума за модулем два двох будь-яких рядків повинна містити не менше $d_{\min} - 2$ одиниць.

Перевірні символи утворюються за рахунок лінійних операцій над інформаційними символами. Для кожної кодової комбінації повинно бути складено p незалежних сум за модулем два. Вибір інформаційних символів, що приймають участь у формуванні того або іншого перевірного символу, залежить від способу декодування коду.

Дуже зручно перевірні суми складати за допомогою перевірної матриці H , яка будується наступним чином. Спочатку будується підматриця H^1 , що є транспонованою стосовно підматриці H_p :

$$H^1 = \begin{pmatrix} b_{11} & b_{21} & \dots & b_{k1} \\ b_{12} & b_{22} & \dots & b_{k2} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ b_{1k} & b_{k2} & \dots & b_{kp} \end{pmatrix}.$$

Потім до неї справа приписується одинична матриця:

$$H = \begin{vmatrix} b_{11} & b_{21} & \dots & b_{k1} & 1 & 0 & 0 & \dots & 0 & 0 \\ b_{12} & b_{22} & \dots & b_{k2} & 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ b_{1p} & b_{2p} & \dots & b_{kp} & 0 & 0 & 0 & \dots & 0 & 1 \end{vmatrix}. \quad (6.24)$$

Алгоритм визначення перевірних символів за інформаційними за допомогою матриці (6.24) наступний. Позиції, що займаються одиницями в першому рядку підматриці H^1 , визначають інформаційні розряди, які повинні брати участь у формуванні першого перевірного розряду кодової комбінації. Позиції одиниць у другому рядку підматриці H^1 визначають інформаційні розряди, що беруть участь у формуванні другого перевірного розряду і т.д.

Нехай, наприклад, твірна матриця коду (7,4) має вигляд:

$$F_{7,4} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{vmatrix}.$$

Перевірна підматриця:

$$H_p = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{vmatrix}.$$

Транспонована підматриця стосовно H_p :

$$H^1 = \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{vmatrix}.$$

Приписуючи справа одиничну матрицю, отримаємо перевірну матрицю:

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Кодові комбінації повинні містити $p = n - k = 7 - 4 = 3$ перевірних символів. Підматриця H^1 вказує на те, що перевірні символи повинні визначатися рівностями:

$$b_1 = a_2 + a_3 + a_4;$$

$$b_2 = a_1 + a_3 + a_4;$$

$$b_3 = a_1 + a_2 + a_4.$$

Тоді для повідомлення, що представляється, наприклад, простою кодовою комбінацією 0011, перевірні символи будуть:

$$b_1 = 0 + 1 + 1 = 0;$$

$$b_2 = 0 + 1 + 1 = 0;$$

$$b_3 = 0 + 0 + 1 = 1.$$

Отже, повна кодова комбінація буде мати вигляд 0011001.

Перевірна матриця H дуже зручна для визначення місця помилки у кодовій комбінації, а, отже, виправлення помилок. Перевірка кодових комбінацій при цьому виконується шляхом підсумовування за модулем два перевірних символів кодових комбінацій і перевірних символів, обчислених за прийнятими інформаційними. У результаті буде отримана сукупність контрольних рівностей, кожна з яких представляє суму за модулем два одного з контрольних символів і певної кількості інформаційних.

Склад контрольних рівностей легко визначається з перевірної матриці H . До складу першої контрольної рівності повинні входити символи, позиції яких зайняті одиницями в першому рядку матриці H . До складу другої контрольної рівності повинні входити символи, позиції яких зайняті одиницями в другому рядку матриці H і т.д.

Так, для розглянутого раніше прикладу з кодом (7,4) ці рівності будуть мати вигляд:

$$S_1 = b_1 + a_2 + a_3 + a_4;$$

$$S_2 = b_2 + a_1 + a_3 + a_4;$$

$$S_3 = b_3 + a_1 + a_2 + a_4.$$

У результаті p таких перевірок буде отримане p -розрядне двійкове число (синдром), яке дорівнюватиме нулю при відсутності помилок і буде відмінним від нуля у випадку наявності помилок.

Якщо код призначений для виправлення помилок, то повинна бути заздалегідь визначена відповідність між видом синдрому і видом помилки, яку потрібно виправити.

Нехай у розглянутому нами прикладі з кодом (7, 4) відбулася помилка в першому розряді кодової комбінації (спотворений символ a_1). Тоді перевірні операції дадуть наступний результат:

$$S_1 = 0; \quad S_2 = 1; \quad S_3 = 0.$$

Таким чином, при помилці в першому розряді кодової комбінації буде отриманий синдром 010. У випадку помилки в другому розряді буде отриманий синдром 101 і т.д.

Якщо для виправлення однократних помилок у кодових комбінаціях одержати синдроми досить просто, то для виправлення дво-, трикратних і т.д. помилок, а також для виправлення пачок помилок побудова синдромів досить складна і у цих випадках зазвичай використовують ЕОМ.

Розглянемо деякі типи систематичних кодів.

6.11 Коди з виявленням помилок

6.11.1 Код із парним числом одиниць

Код містить лише один надлишковий символ. Вибирається надлишковий символ таким чином, щоб загальна кількість одиниць у кодовій комбінації була парною. Перевірка кодової комбінації виконується шляхом підсумовування за модулем два всіх його символів.

Надлишковість коду дорівнює:

$$K_{\text{надл}} = \frac{p}{n} = \frac{1}{n},$$

де n – довжина кодової комбінації;

p – кількість перевірних символів, $p = 1$.

Код дозволяє виявляти однократні помилки і всі помилки непарної кратності, тому що тільки в цих випадках кількість одиниць у комбінації стане непарною. Не виявляються помилки парної кратності.

6.11.2 Код із подвоєнням елементів

Код із подвоєнням елементів характеризується введенням додаткових символів для кожного символу інформаційної частини ком-

бінації, причому одиниця доповнюється нулем і перетворюється в 10, а нуль доповнюється одиницею і перетворюється в 01. Тоді вихідна комбінація 10101 буде подана у вигляді 1001100110. Показником спотворення коду буде поява в “парних” елементах сполучень виду 00 або 11.

Надлишковість коду не залежить від числа елементів кодової комбінації і дорівнює:

$$K_{\text{надл}} = 0,5.$$

Код дозволяє виявляти всі помилки, за винятком випадків, коли мають місце дві помилки в “парних” елементах. Найбільш ймовірним видом невиявлених помилок є помилка в одному із “парних” елементів.

Завадостійкість коду з подвоєнням елементів вище завадостійкості коду з парним числом одиниць. Це досягнуто за рахунок збільшення надлишковості коду й ускладнення процедури перевірки коду.

6.11.3 Інверсний код

В основу побудови інверсного коду покладений метод повторення вихідної кодової комбінації. Причому в тих випадках, коли вихідна комбінація містить парне число одиниць, друга комбінація точно відтворює вихідну. Якщо ж вихідна комбінація містить непарне число одиниць, то повторення робиться в інвертованому вигляді. Наприклад, комбінації 01010 та 01110 інверсним кодом представляються як 0101001010 та 0111010001 відповідно.

Перевірка кодової комбінації виконується в такій послідовності. Спочатку підсумовуються одиниці, що містяться в основній комбінації. Якщо їхня кількість виявиться парною, то елементи додаткової комбінації приймаються в незмінному вигляді. Після цього обидві комбінації порівнюються поелементно (перший елемент із першим, другий із другим і т.д.) і при виявленні хоча б однієї розбіжності прийнята комбінація бракується.

Якщо ж кількість одиниць основної комбінації непарна, елементи другої комбінації приймаються в інвертованому вигляді. Потім, як і в попередньому випадку, основна і додаткова комбінації порівнюються поелементно.

Твірна і перевірна матриці коду мають вигляд:

$$P_{n,k} = \left[U_k, \overline{U_k} \right]; \quad H = \left[\overline{U_k}, U_k \right],$$

де U_k – одинична матриця і $\overline{U_k}$ – матриця, отримана з одиничної шляхом заміни одиниць нулями, а нулів – одиницями.

Надлишковість коду не залежить від числа елементів кодової комбінації та дорівнює:

$$K_{надл} = 0,5.$$

Код дозволяє виявити практично всі помилки в комбінації. Помилки не будуть виявлені лише тоді, коли одночасно спотворяться два, чотири і т.д. елементи у вихідній комбінації і відповідні два, чотири і т.д. елементи додаткової комбінації.

З розглянутих кодів інверсний код має найбільшу завадостійкість.

6.12 Коди з виявленням і виправленням помилок

6.12.1 Коди Хеммінга

Відомо декілька різновидів коду Хеммінга, що характеризуються різною коригувальною здатністю. До цих кодів відносять коди з виправленням однократних помилок і коди з виправленням однократних і виявленням двократних помилок.

Код Хеммінга, що забезпечує виправлення всіх однократних помилок, повинен мати мінімальну кодову відстань $d_{min} = 3$. Довжина коду n вибирається з умови:

$$2^k \leq \frac{2^n}{1+n}, \quad (6.25)$$

де n – кількість інформаційних сигналів.

Код будується таким чином, щоб у результаті $p = n - k$ перевірок одержати p -розрядне двійкове число, що вказує номер спотвореної позиції кодової комбінації. Для цього перевірни символи повинні знаходитися на номерах позицій, що виражаються степенем двійки ($2^0, 2^1, 2^2, \dots, 2^{r-1}$), тому що кожний із них входить тільки в одне з перевірних рівнянь. Таким чином, якщо нумерувати позиції зліва направо, то контрольні символи повинні знаходитися на першій, другій, четвертій і т.д. позиціях.

Результат першої перевірки дає цифру молодшого розряду синдрому у двійковому записі. Якщо результат цієї перевірки дасть 1, то один із символів перевіреної групи спотворений. Таким чином, першою перевіркою повинні бути охоплені символи з номерами, що містяться у двійковому записі одиниці в першому розряді: 1, 3, 5, 7, 9 і т.д. Результат другої перевірки дає цифру другого розряду синдрому. Отже, другою перевіркою повинні бути охоплені символи з номерами, що містяться у двійковому записі одиниці в другому розряді: 2, 3, 6, 7, 10 і т.д.

Аналогічно при третій перевірці повинні перевірятися символи, номери яких у двійковому записі містять одиниці в третьому розряді: 4, 5, 6, 7, 12 і т.д. Таким чином, перевірні групи повинні мати вигляд:

$$\begin{aligned} S_1 &= a_1 + a_3 + a_5 + a_7 + a_9 + \dots; \\ S_2 &= a_2 + a_3 + a_6 + a_7 + a_{10} + \dots; \\ S_3 &= a_4 + a_5 + a_6 + a_7 + a_{12} + \dots; \\ S_4 &= a_8 + a_9 + a_{10} + a_{11} + a_{12} + \dots; \\ &\dots \end{aligned} \quad (6.26)$$

Перевірні матриця коду повинна мати n стовпців і ρ рядків. Кожний стовпець повинен складати двійкову комбінацію, що вказує номер відповідної позиції коду.

Наприклад, для коду довжиною $n = 9$, що забезпечує виправлення однократних помилок, кількість надлишкових символів $\rho = 9$. При цьому як перевірна може бути обрана наступна матриця:

$$H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Представимо як приклад просту двійкову комбінацію 10011 кодом Хеммінга. Оскільки інформаційними повинні бути третій, п'ятий, шостий, сьомий і дев'ятий символи, то для коду, що розглядається, $a_3 = 1$, $a_5 = 0$; $a_6 = 0$; $a_7 = 1$, $a_9 = 1$. З умови забезпечення парності сум (6.26) одержимо наступні значення перевірних символів $a_1 = 1$, $a_2 = 0$; $a_4 = 1$; $a_8 = 1$. Отже, простому п'ятиелементному коду 11011 відповідає дев'ятиелементний код Хеммінга 101100111.

Нехай тепер при передачі відбулося спотворення п'ятого символу, тобто код прийняв вигляд 101110111. Тоді в результаті першої перевірки одержимо 1, другої — 0, третьої — 1 і четвертої — 0. Таким чином, у результаті перевірок отриманий синдром 0101, що вказує на спотворення п'ятого символу. Виправлення помилки зводиться до інвертування символу на п'ятій позиції.

Код Хеммінга з кодовою відстанню $d_{\min} = 4$ утворюється шляхом додавання до коду Хеммінга з $d_{\min} = 3$ перевірного символу, що являє собою результат підсумовування за модулем два всіх символів кодової комбінації.

Операція декодування складається з двох етапів. На першому етапі визначається синдром, що відповідає коду з $d_{\min} = 3$, на другому – перевіряється останнє перевірне співвідношення.

Для розглянутого раніше коду з $d_{\min} = 4$ перевірна матриця може мати вигляд:

$$H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Додаткове перевірне співвідношення, яке вводиться для збільшення мінімальної відстані коду Хеммінга до $d_{\min} = 4$, має вигляд:

$$S_5 = a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10}.$$

Надлишковість коду Хеммінга $K_{надл} = \frac{p}{n}$ залежить від кількості інформаційних символів і при зміні k від 4 до 1013 змінюється від 0,429 до 0,098 при $d_{\min} = 3$ та від 0,5 до 0,0107 при $d_{\min} = 4$.

6.12.2 Циклічні коди

Основні властивості циклічного коду та способи побудови.

Циклічні коди одержали досить широке застосування завдяки їхній ефективності при виявленні і виправленні помилок. Схеми кодувальних і декодувальних пристроїв для цих кодів надзвичайно прості і будуються на основі звичайних регістрів зсуву.

Назва кодів виникла від їх властивості, яка полягає в тому, що кожна кодова комбінація може бути отримана шляхом циклічної перестановки символів комбінації, що належить до цього ж коду. Це означає, що якщо, наприклад, комбінація $a_0 a_1 a_2 \dots a_{n-1}$ є дозволеною комбінацією циклічного коду, то комбінація $a_{n-1} a_0 a_1 a_2 \dots a_{n-2}$ також належить цьому коду.

Циклічні коди зручно розглядати, представляючи комбінацію двійкового коду не у вигляді послідовностей нулів і одиниць, а у вигляді полінома від фіктивної змінної x , а саме.

$$G(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_1x + a_0, \quad (6.27)$$

де a_i – цифри даної системи числення (у двійковій системі 0 і 1).

Так, наприклад, двійкове семирозрядне число 1010101 може бути записане у вигляді полінома:

$$\sigma(x) = 1 \cdot x^6 + 0 \cdot x^5 + 1 \cdot x^4 + 0 \cdot x^3 + 1 \cdot x^2 + 0 \cdot x^1 + 1 \cdot x^0 = x^6 + x^4 + x^2 + 1. \quad (6.28)$$

Найбільший степінь у доданку з ненульовим коефіцієнтом називається степенем полінома.

Представлення кодових комбінацій у формі (6.28) дозволяє звести дії над комбінаціями до дії над багаточленами. При цьому додавання двійкових багаточленів зводиться до додавання за модулем два коефіцієнтів при рівних степенях змінної x ; множення виконується за звичайним правилом перемножування степеневих функцій, однак отримані при цьому коефіцієнти при рівних степенях змінної x додаються за модулем два; ділення здійснюється за правилами ділення степеневих функцій, при цьому операції віднімання замінюються операціями підсумовування за модулем два.

Представлення комбінацій у формах (6.27) і (6.28) зручно ще і тим, що згадана циклічна перестановка є результатом простого множення даного полінома на x . Дійсно, якщо одна з кодових комбінацій виражається поліномом $V(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}$, то нова комбінація за рахунок циклічного зсуву $x \cdot V(x) = a_0x + a_1x^2 + a_2x^3 + \dots + a_{n-2}x^{n-1}$. Однак в останньому члені необхідно замінити x^n на 1. Отже, нова комбінація буде:

$$V_1(x) = a_{n-1} + a_0x + a_1x^2 + \dots + a_{n-2}x^{n-1}.$$

Наприклад, циклічний зсув кодової комбінації 1010101 може бути отриманий шляхом множення полінома (6.28) на x :

$$G(x) \cdot x = x^7 + x^5 + x^3 + x.$$

Замінюючи x^7 на 1, отримаємо поліном:

$$G_1(x) = x^5 + x^3 + x + 1,$$

що відповідає кодовій комбінації 0101011.

Відповідно до визначення циклічного коду для побудови твірної матриці $P_{n,k}$ досить вибрати тільки одну вихідну n -розрядну комбінацію $V_1(x)$. Циклічним зсувом можна одержати $(n-1)$ різних комбінацій, із котрих будь-які k комбінацій можуть бути взяті як вихідні. Підсумовуючи рядки твірної матриці в усіх можливих комбінаціях, можна одер-

жати інші кодові комбінації. Можна показати, що кодові комбінації, одержувані з деякої комбінації $V_1(x)$ циклічним зсувом, задовольняють умовам, що висуваються до сукупності вихідних комбінацій.

Циклічний зсув комбінації з одиницею в старшому n -му розряді рівнозначний множенню відповідного багаточлена на x з одночасним вирахуванням із результату багаточлена $(x-1)$ або (x^n+1) , оскільки операції здійснюються за модулем два. Отже, якщо за вихідний взяти деякий поліном $P(x)$, то процес одержання базових поліномів можна представити в наступному вигляді:

$$\begin{aligned} U_1(x) &= P(x); \\ U_2(x) &= P(x)x - C_2(x^n + 1); \\ U_3(x) &= P(x)x^2 - C_3(x^n + 1); \\ &\dots\dots\dots \\ U_n(x) &= P(x)x^{n-1} - C_n(x^n + 1), \end{aligned} \tag{6.29}$$

де C_2, C_3, \dots, C_n — коефіцієнти, що приймають значення 1 при $P(x) \cdot x^i \geq (x^n - 1)$ і значення 0 при $P(x) \cdot x^i < (x^n - 1)$.

При такому способі побудови базових поліномів поліном $P(x)$ називають твірним.

Якщо прийняти умову, що поліном $P(x)$ є дільником двочлена $(x^n + 1)$, то базові комбінації, а разом із ними і всі дозволені комбінації коду набувають властивості подільності на $P(x)$. З цього випливає, що приналежність кодової комбінації до групи дозволених можна легко перевірити діленням її полінома на твірний поліном $P(x)$. Якщо залишок від ділення дорівнює нулю, то комбінація є дозволеною.

Ця властивість циклічного коду використовується для виявлення або виправлення помилок. Дійсно, якщо під впливом завад дозволена кодова комбінація трансформується в заборонену, то помилка може бути виявлена за наявністю залишку при діленні комбінації на твірний поліном $P(x)$.

Таким чином, твірний поліном $P(x)$ повинен задовольняти вимозі — він повинний бути дільником двочлена $(x^n + 1)$. Вибір $P(x)$ односторонньо визначає циклічний код і його коригувальні властивості.

Циклічний (n, k) -код може бути отриманий шляхом множення простого k -значного коду, вираженого у вигляді полінома степеня $(k-1)$, на деякий твірний поліном $P(x)$ степеня $(n-k)$.

Можлива й інша процедура отримання циклічного коду. Для цього кодова комбінація простого k -значного коду $G(x)$ множитья на

одночлен x^{n-k} , а потім ділиться на твірний поліном $P(x)$ степеня $(n-k)$. У результаті множення $G(x)$ на x^{n-k} степінь кожного одночлена, що входить в $G(x)$, підвищиться на $(n-k)$. При діленні добутку $x^{n-k}G(x)$ на твірний поліном $P(x)$ утвориться частка $Q(x)$ такого ж степеня, як і $G(x)$.

Результат множення і ділення можна представити в наступному вигляді:

$$\frac{x^{n-k}G(x)}{P(x)} = Q(x) + \frac{R(x)}{P(x)}, \quad (6.30)$$

де $R(x)$ – залишок від ділення $x^{n-k}G(x)$ на $P(x)$.

Оскільки частка $Q(x)$ має такий же степінь, як і кодова комбінація $G(x)$, то $Q(x)$ також є комбінацією простого k -значного коду.

Помноживши обидві частини рівності (6.30) на $P(x)$ і зробивши деякі перестановки, одержимо:

$$F(x) = Q(x)P(x) = x^{n-k}G(x) + R(x). \quad (6.31)$$

У правій частині виразу (6.31) знак “мінус” перед $R(x)$ замінений знаком “плюс”, тому що віднімання за модулем два зводиться до додавання.

Таким чином, кодова комбінація циклічного (n, k) -коду може бути отримана двома способами:

- 1) шляхом множення простої кодової комбінації степеня $(k-1)$ на одночлен x^{n-k} і додавання до цього добутку залишку, отриманого від ділення отриманого добутку на твірний поліном $P(x)$ степеня $(n-k)$;
- 2) шляхом множення простої кодової комбінації степеня $(k-1)$ на твірний поліном $P(x)$ степеня $(n-k)$.

При першому способі кодування перші k символів отриманої кодової комбінації збігаються з відповідними символами вихідної простої кодової комбінації.

При другому способі в отриманій кодовій комбінації інформаційні символи не завжди збігаються із символами вихідної простої комбінації. Такий спосіб легко реалізується, але внаслідок того, що в отриманих кодових комбінаціях не містяться інформаційні символи в явному вигляді, ускладнюється процес декодування.

У зв'язку з цим на практиці звичайно використовується перший спосіб отримання циклічного коду.

Матричне представлення циклічних кодів. Для формування рядків твірної матриці за першим способом утворення циклічного коду

беруть комбінації простого k -розрядного коду $G(x)$, які містять одиницю в одному розряді. Ці комбінації множаться на x^{n-k} , визначається залишок $R(x)$ від ділення отриманого добутку $x^{n-k}G(x)$ на твірний поліном і записується відповідний рядок матриці у вигляді суми добутку $x^{n-k}G(x)$ і залишку $R(x)$. При цьому твірна матриця $P_{n,k}$ представляється двома підматрицями – інформаційною U_k і додатковою H_p :

$$P_{n,k} = \left[U_k, H_p \right].$$

Інформаційна підматриця U_k являє собою квадратну одиничну матрицю з кількістю рядків і стовпців, рівною k . Додаткова підматриця H_p містить $p = n - k$ стовпців та k рядків і утворена залишками $R(x)$.

Твірна матриця дозволяє одержати k комбінацій коду. Інші комбінації утворюються підсумовуванням за модулем два рядків твірної матриці в усіх можливих сполученнях.

Нехай, наприклад, необхідно побудувати твірну матрицю (7,4) циклічного коду. Твірний поліном $P(x) = x^3 + x^2 + 1$.

Інформаційна підматриця має вигляд:

$$U_k = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix}.$$

Для одержання першого рядка додаткової підматриці перший рядок інформаційної підматриці множиться на x^3 і ділиться на твірний поліном. Це відповідає виконанню операції $\frac{0001 \cdot 1000}{1000}$. Залишок цих операцій 101 і складе перший рядок додаткової підматриці. Аналогічно визначаються інші рядки додаткової підматриці.

Остаточна твірна матриця має вигляд:

$$P_{1,4} = \begin{vmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \end{vmatrix}.$$

При другому способі утворення циклічного коду твірна матриця $P_{n,k}$ формується шляхом множення твірної полінома $P(x)$ степеня $\rho = n - k$ на одночлен x^{k-1} і наступних $k - 1$ зсувів отриманої комбінації.

Вибір твірної полінома. При побудові циклічного коду спочатку визначається кількість інформаційних розрядів k за заданим об'ємом коду. Потім знаходиться найменша довжина кодових комбінацій n , що забезпечує виявлення або виправлення помилок заданої кратності. Ця проблема зводиться до знаходження потрібного твірної полінома $P(x)$.

Як уже зазначалося, степінь твірної полінома повинна дорівнювати кількості перевірних розрядів ρ .

Оскільки в циклічному коді розпізнавачем помилок є залишки від ділення багаточлена прийнятої комбінації на твірний багаточлен, коригувальна здатність коду буде тим вище, чим більше залишків може бути утворено в результаті цього ділення.

Найбільше число залишків, рівне $2^\rho - 1$ (крім нульового), може забезпечити тільки незвідний багаточлен степеня ρ (тобто такий, що не ділиться ні на який інший багаточлен).

Відомо, що двочлен типу $(x^n + 1) = (x^{2^r-1} + 1)$, у розкладанні якого як співмножник повинен входити твірний багаточлен, має властивість бути спільним кратним для усіх без винятку незвідних поліномів степеня z і розкладається на множники з усіх незвідних поліномів, степені z , яких ділять без залишку число z .

Найпростішим циклічним кодом є код, що забезпечує виявлення однократних помилок. Вектору однократної помилки відповідає одвочлен x^i , степінь котрого i може приймати значення від 1 до n . Для того щоб могла бути виявлена помилка, одночлен x^i не повинен ділитися на твірний поліном $P(x)$. Серед незвідних багаточленів, що входять у розкладання двочлена $x^n + 1$, є багаточлен найменшої степені $x + 1$. Таким чином, твірним поліномом даного коду є двочлен $P(x) = x + 1$. Залишок від ділення будь-якого багаточлена на $x + 1$ може приймати тільки два значення: 0 і 1. Отже, при будь-якій кількості інформаційних розрядів необхідний тільки один перевірний розряд. Значення символу цього розряду забезпечує парність кількості одиниць у кодовій комбінації.

Цей циклічний код із перевіркою на парність забезпечує виявлення не тільки однократних помилок, але і всіх помилок непарної кратності.

Наприклад, для коду з $k = 4$ інформаційна підматриця має вигляд:

$$U_k = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix}.$$

Додаткову матрицю можна побудувати за залишками ділення останнього рядка інформаційної підматриці, доповненої p нулями, на твірний поліном:

| | | | |
|-----------------|-----|--------|---------|
| 10000 | $ $ | 11 | залишки |
| $\frac{11}{10}$ | $ $ | 1111 | |
| $\frac{11}{10}$ | $ $ | | 1 |
| $\frac{11}{10}$ | $ $ | | 1 |
| $\frac{11}{10}$ | $ $ | | 1 |
| $\frac{11}{1}$ | $ $ | | 1 |

Таким чином, додаткова матриця має вигляд:

$$H_p = \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \end{vmatrix}.$$

Отже, твірна матриця:

$$P_{3,4} = \begin{vmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{vmatrix}.$$

Для побудови циклічного коду, що виправляє однократні або виявляє двократні помилки, необхідно, щоб кожній одиничній помилці відповідав свій розпізнавач, тобто залишок від ділення багаточлена прийнятої комбінації на твірний багаточлен. Оскільки кількість можливих однократних помилок дорівнює n , а незвідний багаточлен степеня p може дати $2^p - 1$ ненульових залишків, то необхідною умовою виправлення будь-якої одиничної помилки є виконання нерівності:

$$2^n - 1 \geq n. \quad (6.32)$$

Звідси знаходяться степінь твірного полінома

$$\rho = n - k \geq \log_2(n + 1) \quad (6.33)$$

і загальна довжина n кодової комбінації.

У таблиці 6.7 наведені найбільші значення k та n для різних ρ .

Таблиця 6.7

| ρ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------|---|---|---|----|----|----|-----|-----|-----|------|
| N | 1 | 3 | 7 | 15 | 31 | 63 | 127 | 255 | 511 | 1023 |
| K | 0 | 1 | 4 | 11 | 26 | 57 | 120 | 247 | 502 | 1013 |

Оскільки твірний багаточлен $P(x)$ повинен входити як співмножник в розкладання двочлена $(x^n + 1) = (x^{2^2-1} + 1)$, то, використовуючи зазначені раніше властивості цього двочлена, а також умову (6.33), можна вибрати твірний поліном.

Однак не всякий багаточлен степеня ρ , що входить у розкладання $x^n + 1$, може бути використаний як твірний поліном. Необхідно, щоб для кожної із n однократних помилок забезпечувався свій, відмінний від інших, залишок від ділення прийнятої кодової комбінації на твірний поліном. Це буде мати місце за умови, якщо обраний незвідний багаточлен степеня ρ , будучи дільником двочлена $x^n + 1$, не входить у розкладання ніякого іншого двочлена $x^l + 1$, степінь якого $l < n$.

Розглянемо як приклад спосіб вибору твірного полінома для побудови циклічного коду, що містить $k = 4$ інформаційних символи і забезпечує усунення однократних помилок та виявлення двократних помилок.

Відповідно до (6.33) і таблиці 6.6 визначаємо загальну кількість символів $n = 4$ і кількість перевірних символів $\rho = 3$.

Твірний поліном $P(x)$ повинен бути степеня $\rho = 3$ і входити як співмножник в розкладання двочлена $x^n + 1 = x^{2^2-1} + 1$. Оскільки $n = 7$, то складові співмножники двочлена повинні бути незвідними поліномами, степені яких є дільниками числа $z = 3$. До чисел, на котрі $z = 3$ ділиться без залишку, відносяться 1 та 3. Отже, співмножниками двочлена $x^7 + 1$ повинні бути незвідними поліномами першого і третього степеня.

Користуючись таблицями незвідних поліномів [13], отримуємо:

$$x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1).$$

Жоден із співмножників степеня 3 не входить у розкладання іншого двочлена $x^n + 1$ степені $n < 7$. Тому кожний із цих співмножників може бути обраний як твірний поліном.

Твірні поліноми кодів, здатних виправляти помилки будь-якої кратності, можна визначати, користуючись наступним правилом Хеммінга:

1. За заданим числом інформаційних розрядів k , визначається число перевірних розрядів p , необхідне для виправлення однократних помилок, і знаходиться твірний поліном.

2. Розглядаючи отриманий (n, k) -код як некоригувальний n -розрядний код, визначають додаткові розряди для забезпечення виправлення однієї помилки в цьому коді і знаходять відповідний твірний поліном.

3. Ця процедура повторюється кілька раз, поки не буде отриманий код, що виправляє незалежні помилки до даної кратності включно.

Однак код, побудований таким чином, є неоптимальним з погляду кількості надлишкових символів. У цьому відношенні більш досконалим є код Боуза—Чоудхурі, що забезпечує мінімальну кількість перевірних символів при заданому k . Математична структура цього коду трохи відмінна від розглянутої раніше і характеризується більш складними пристроями для виявлення і виправлення помилок. Особливістю цього коду є те, що для будь-яких цілих додатних чисел z і t_u існує циклічний код значності $n = 2^z - 1$ з кодовою відстанню $d_{min} \geq 2t_u + 1$. При цьому кількість перевірних символів p не перевищує величини $z \cdot t_u$ тобто $p \leq z \cdot t_u$. Такий код гарантовано виправляє помилки кратності не більше t_u або виявляє помилки кратності не більше $2t_u$. Крім того, код виявляє всі пачки помилок, довжина яких дорівнює або менше p .

До циклічних кодів, призначених для виправлення пачок помилок, відносяться коди Файра, Абрамсона і Міласа—Абрамсона, що забезпечують виправлення однієї пачки помилок, і код Ріда—Соломона, розрахований на виправлення декількох пачок помилок.

Найбільш відомим із цієї групи кодів є код Файра.

Твірний поліном даного коду має вигляд:

$$P(x) = q(x)(x^c + 1), \quad (6.34)$$

де $q(x)$ — незвідний багаточлен степеня t , що належить показнику степеня:

$$m = 2^t - 1, \quad (6.35)$$

а число c — деяке постійне число, що не повинно ділитися на m .

Кількість перевірних символів дорівнює $c + t$, а загальна кількість символів у кодовій комбінації дорівнює найменшому спільному кратному чисел c та m .

За допомогою кодів Файра можна виправити одиничну пачку помилок довжиною b або менше й одночасно виявити пачку помилок довжиною $l \geq b$ або менше, якщо:

$$b + l - 1 \leq c, \quad b \leq m. \quad (6.36)$$

Якщо застосовувати ці коди тільки для виявлення помилок, то вони дозволяють виявити будь-яку комбінацію з двох пачок помилок, довжина найменшої з яких не перевершує t , а сума обох пачок не перевищує $c + 1$, а також будь-яку одиничну пачку помилок із довжиною, що не перевищує кількість перевірних символів $p = c + 1$.

6.12.3 Ітеративні коди

Ітеративні коди відрізняються тим, що операції кодування проводяться над сукупністю символів, що розташовуються на декількох координатах. Загальна кількість інформаційних символів у такому кодї дорівнює:

$$k = \prod_{l=1}^q k_l, \quad (6.37)$$

де k_l – кількість інформаційних символів по координаті l ;

q – кількість координат.

У загальному випадку кожний інформаційний символ входить одночасно в q різних кодових комбінацій. Комбінації інформаційних символів за кожною із координат кодуються яким-небудь лінійним кодом. Одержуваний ітеративний код також є лінійним.

Загальна кількість символів у кодовій комбінації:

$$n = \prod_{l=1}^q n_l, \quad (6.38)$$

де n_l – загальна кількість символів за координатою l .

Мінімальна кодова відстань для ітеративного коду:

$$d_{\min} = \prod_{l=1}^q d_{\min l}, \quad (6.39)$$

де $d_{\min l}$ – мінімальна кодова відстань для лінійного коду за координатою l .

Найпростішим з ітеративних кодів є двовимірний код із перевіркою на парність за рядками і стовпцями. Розташування інформаційних і перевірних символів такого коду показано на рис. 6.6.

| | | | | | | |
|-------------|-------------|-----|-------------|-----|-------------|----------------|
| a_{11} | a_{12} | ... | a_{1i} | ... | a_{1k} | $a_{1, k+1}$ |
| a_{21} | a_{22} | ... | a_{2i} | ... | a_{2k} | $a_{2, k+1}$ |
| \vdots | \vdots | | \vdots | | \vdots | \vdots |
| a_{j1} | a_{j2} | ... | a_{ji} | ... | a_{jk} | $a_{j, k+1}$ |
| \vdots | \vdots | | \vdots | | \vdots | \vdots |
| a_{l1} | a_{l2} | | a_{li} | | a_{lk} | $a_{l, k+1}$ |
| $a_{l+1,1}$ | $a_{l+1,2}$ | | $a_{l+1,i}$ | | $a_{l+1,k}$ | $a_{l+1, k+1}$ |

Рис. 6.6

Перевірні символи розташовуються в крайньому правому стовпці і нижньому рядку. Їхні значення вибираються такими, щоб суми одиниць у кожному рядку і кожному стовпці були парними.

Передача однієї комбінації ітеративного коду звичайно відбувається за рядками послідовно – від першого рядка до останнього.

Перевірка на парність також робиться послідовно за рядками.

Цей код є найпростішим ітеративним кодом з $d_{\min} = 4$. Він виявляє всі помилки непарної кратності і двократні помилки; не виявляє помилки, що мають парне число спотворених символів як за рядками, так і за стовпцями. Найпростіша помилка, що не виявляється – чотирикратна помилка, що розташовується у вершинах правильного чотирикутника (рис. 6.7, а). На рис. 6.7, б, в зображені деякі шести- та восьмикратні помилки, що не виявляються.

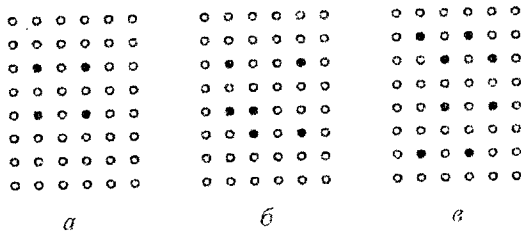


Рис. 6.7

Найпростіший ітеративний код виявляє будь-яку пачку помилок довжиною $l + 1$ і менше, де l – довжина рядка.

Процедура виправлення помилок проста. Якщо не виконується перевірка на парність для i -го рядка і j -го стовпця, то змінюється на зворотний символ, що знаходиться на перетині i -го рядка і j -го стовпця.

Багатовимірні ітеративні коди з кількістю координат $q > 2$ поки не одержали великого поширення.

Недоліком ітеративних кодів із перевітками на парність за рядками та стовпцями є їх порівняно висока надлишковість, що складає 15–20 % і значно перевищує при інших рівних умовах надлишковість циклічних кодів. Крім того, ітеративний код не здатний виправити всі помилки кратності $t_n = 2d_{\min} + 1$. Дійсно, наприклад, при ітерації двох кодів Хеммінга з $d_{\min} = 3$, що забезпечують виправлення однократних помилок, отримують код з $d_{\min} = 3 \cdot 3 = 9$, який повинен виправляти помилки кратності $t_n = 4$ і менше. Однак такий код не може виправляти чотирикратні помилки з розташуванням спотворених символів у вершинах прямокутника (див. рис. 6.7, а).

Ітеративні коди запропоновані П. Еліасом. Вони знайшли широке застосування для виявлення і виправлення помилок, що виникають при записі, збереженні і зчитуванні цифрової інформації на магнітному носії.

6.12.4 Рекурентні коди

Циклічні коди дозволяють виявляти і виправляти як одиничні і подвійні помилки, так і пачки помилок. Однак практичне застосування цих кодів для виправлення пачок помилок утруднене тим, що при не дуже великій надлишковості довжина кодових комбінацій значно перевищує довжину пачок. У цьому відношенні більш зручними є рекурентні коди. Рекурентні (безперервні) коди відрізняються тим, що в них операції кодування і декодування здійснюються безупинно над послідовністю символів без поділу їх на блоки.

Ці коди призначені в основному для виправлення пачок помилок. Формування перевірних символів здійснюється шляхом додавання двох або більше інформаційних символів, зсунутих один відносно іншого на певну відстань t , яка називається кроком додавання.

Довжина пачки помилок l , яка виправляється рекурентним кодом, залежить від кроку додавання t та визначається з умови:

$$l = 2t. \quad (6.40)$$

Мінімальна необхідна відстань між пачками помилок, при якій забезпечується виправлення всіх помилок у пачці довжиною l , дорівнює:

$$a = 6t + 1. \quad (6.41)$$

З усіх рекурентних кодів найбільше поширення одержав, так званий ланцюговий код, що відрізняється гранично простими методами кодування і декодування. У ланцюговому коді кожний перевірний символ формується шляхом додавання за модулем два двох інформаційних символів, відстань між якими дорівнює кроку додавання t .

Позначаючи послідовність інформаційних символів через $a_0a_1a_2\dots a_t a_{t+1}\dots a_{2t}a_{2t+1}\dots$, одержимо наступну послідовність перевірних символів для ланцюгового коду: $b_0 = a_0 + a_t$; $b_1 = a_1 + a_{t+1}$; ... $b_t = a_t + a_{2t}$; $b_{t+1} = a_{t+1} + a_{2t+1}$ У загальному потоці символів ланцюгового коду між кожними двома інформаційними символами міститься один перевірний $a_0b_0a_1b_1a_2b_2\dots a_t b_t a_{t+1}\dots a_{2t}b_{2t}a_{2t+1}b_{2t+1}\dots$.

Оскільки кількість перевірних символів, сформованих за деякий час, дорівнює кількості інформаційних символів, що надійшли за той же час, то надлишковість ланцюгового коду дорівнює 0,5.

На прийомі інформаційні і перевірні символи розділяються і реєструються незалежно один від одного. З прийнятої послідовності інформаційних символів формуються контрольні символи так само, як формувалися перевірні символи при передачі. Після затримки на величину $(3t + 1)$ кожний сформований контрольний символ порівнюється з відповідним прийнятим перевірним символом. У випадку спотворення одного з перевірних символів буде розбіжність відповідних контрольних і перевірних символів. Якщо ж спотворений один з інформаційних символів, то буде розбіжність двох контрольних і відповідних перевірних символів, у формуванні яких бере участь даний інформаційний символ.

З розглянутого принципу виправлення помилок у ланцюговому коді випливає, що правильне виправлення помилок можливо тільки в тому випадку, якщо два з трьох символів, які перевіряються, прийняті правильно. Ця умова виконується, якщо при кроці додавання t довжина пачки помилок не більше $2t$ і перевірні символи передаються в канал із затримкою не більшою $(3t + 1)$. Таким чином, ланцюговий код виявляє і виправляє пачки помилок порівняно просто, але ціною великої надлишковості.

В останні роки почали знаходити застосування згортувальні коди, що становлять інший різновид рекурентних кодів. Для них розроблений метод послідовного декодування, що дозволяє ефективно виправляти помилки при порівняно невеликій складності. Розгляд цих кодів виходить за рамки даного підручника.

6.13 Приклади

Приклад 6.1

Дискретне джерело видає 10 повідомлень. Яку мінімальну кількість розрядів повинні мати кодові комбінації рівномірного двійкового коду, призначеного для кодування цих повідомлень? Записати комбінації.

Розв'язування. Кількість розрядів у кодовій комбінації за виразом (6.4) дорівнює:

$$n = \log_2 10 = 3,32.$$

Округляючи отриманий результат до найближчого більшого цілого, остаточно маємо $n = 4$.

Кодові комбінації наведені у таблиці 6.8.

Таблиця 6.8

| Символ | Число | Представлення числа у вигляді суми | Кодова комбінація |
|----------|-------|---|----------------------|
| A_1 | 0 | $0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$ | 0000 |
| A_2 | 1 | $0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$ | 0001 |
| A_3 | 2 | $0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$ | 0010 |
| A_4 | 3 | $0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$ | 0011 |
| A_5 | 4 | $0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$ | 0100 |
| A_6 | 5 | $0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$ | 0101 |
| A_7 | 6 | $0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$ | 0110 |
| A_8 | 7 | $0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$ | 0111 |
| A_9 | 8 | $1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$ | 1000 |
| A_{10} | 9 | $1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$ | 1001 |

Приклад 6.2

Чому дорівнює мінімальна довжина кодів комбінацій при передачі 16 і 128 повідомлень у двійковому та вісімковому кодах?

Розв'язування. Необхідна довжина кодів комбінацій при передачі повідомлень у двійковому кодї:

$$n_{16} = \log_2 16 = 4; \quad n_{128} = \log_2 128 = 7.$$

Довжина кодів комбінацій при передачі повідомлень у вісімковому кодї:

$$n_{16} = \log_8 16 = \frac{\log_2 16}{\log_2 8} = 1,3, \text{ після округлення } - n_{16} = 2;$$

$$n_{128} = \log_8 128 = \frac{\log_2 128}{\log_2 8} = 2,33, \text{ після округлення } - n_{128} = 3.$$

Приклад 6.3

Визначити коригувальну здатність коду, що має наступні дозволені комбінації: 00000; 01110; 10101; 11011.

Розв'язування. Коригувальна здатність коду визначається мінімальною кодовою відстанню. Складемо матрицю відстаней між кодовими комбінаціями (табл. 6.9).

Таблиця 6.9

| | 0000 | 01110 | 10101 | 11011 |
|-------|------|-------|-------|-------|
| 00000 | 0 | 3 | 3 | 4 |
| 01110 | 3 | 0 | 4 | 3 |
| 10101 | 3 | 3 | 0 | 3 |
| 11011 | 4 | 3 | 4 | 0 |

Як видно з матриці, мінімальна кодова відстань $d_{\min} = 3$. Отже, цей код здатний:

- виявляти дворазові помилки;
- усувати помилки;
- усувати і виявляти одиничні помилки.

Приклад 6.4

Побудувати твірну матрицю систематичного коду, здатного виправляти одиничну помилку ($\sigma = 1$) при передачі 16 повідомлень.

Розв'язування. Оскільки кількість дозволених кодових комбінацій $N = 16$, то число інформаційних розрядів у кодових комбінаціях:

$$k = \log_2 16 = 4.$$

Отже, кількість рядків твірної матриці $P_{n,k}$ повинна дорівнювати чотирьом.

Користуючись умовою (6.13) і таблицею 6.6, знаходимо довжину коду $n = 7$. Отже, число стовпців матриці $P_{n,k}$ дорівнює семи.

Оскільки кількість перевірних розрядів $\rho = n - k = 3$, то кількість стовпців перевірної підматриці H_ρ дорівнює трьом. Кількість одиниць у кожному рядку підматриці H_ρ повинна бути не менше $d_{\min} - 1$. Для $\sigma = 1$ значення $d_{\min} = 2\sigma + 1 = 3$. Вибираємо для підматриці H_ρ наступні комбінації: 111, 110, 101, 011.

Отже, перевірна підматриця може мати вигляд:

$$H_\rho = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix}, \text{ або } H_\rho = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix}, \text{ або } H_\rho = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{vmatrix} \text{ і т.д.}$$

Інформаційну підматрицю U_k беремо у вигляді одиничної матриці з кількістю рядків і стовпців, рівним $k = 4$:

$$U_k = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}.$$

Остаточний вигляд твірної матриці:

$$P_{7,4} = \begin{vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{vmatrix}, \text{ або } P_{7,4} = \begin{vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{vmatrix},$$

$$\text{або } P_{7,4} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{vmatrix}, \text{ і т.д.}$$

Приклад 6.5

Побудувати перевірну матрицю H систематичного коду (7,4), твірна матриця якого має вигляд:

$$P_{7,4} = \begin{vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{vmatrix}.$$

Скласти рівняння для визначення перевірних символів.

Розв'язування. Знаходимо підматрицю H^1 , що є транспонованою стосовно підматриці H_p :

$$H^1 = \begin{vmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{vmatrix}.$$

Приписуючи до неї справа одиничну матрицю U_k , одержуємо перевірну матрицю:

$$H = \begin{vmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{vmatrix}.$$

Перевірні символи визначаються шляхом підсумовування за модулем два певних інформаційних символів. Алгоритм визначення перевірних символів визначається підматрицею H^1 . Позиції, що займаються одиницями в першому рядку H^1 , визначають інформаційні розряди, які повинні брати участь у формуванні першого перевірного розряду. Позиції одиниць у другому рядку визначають інформаційні розряди, що беруть участь у формуванні другого перевірного розряду. І нарешті, позиції одиниць у третьому рядку визначають інформаційні розряди, що беруть участь у формуванні третього перевірного розряду. Отже, перевірні символи визначаються рівностями:

$$b_1 = a_1 + a_2 + a_3, \quad b_2 = a_1 + a_2 + a_4, \quad b_3 = a_1 + a_3 + a_4.$$

Приклад 6.6

Побудувати комбінацію систематичного коду (7,4) для випадку, коли проста ненадлишкова комбінація має вигляд 0110.

Розв'язування. У даному випадку значення інформаційних символів наступні: $a_1 = 0$; $a_2 = 1$; $a_3 = 1$; $a_4 = 0$.

Тоді, використовуючи отримані в попередньому прикладі рівності для перевірних символів, визначаємо:

$$b_1 = 0 + 1 + 1 = 0, \quad b_2 = 0 + 1 + 0 = 1, \quad b_3 = 0 + 1 + 0 = 1.$$

Отже, повідомлення 0110, закодоване систематичним кодом (7,4), прийме вигляд: 0110011.

Приклад 6.7

Твірна матриця коду (7,4) має вигляд:

$$P_{11,7} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{vmatrix}.$$

Побудувати комбінацію систематичного коду (11,7), що відповідає ненадлишковій комбінації 1101001.

Розв'язування. Оскільки для одержання комбінації вигляду 1101001 необхідно скласти перший, другий, четвертий і сьомий рядки одиничної матриці U_k , для знаходження перевірних розрядів слід скласти ці ж рядки підматриці H_p :

$$\begin{array}{r} 1111 \\ + 1110 \\ 1011 \\ \hline 0110 \\ 1100 \end{array}$$

Таким чином, комбінація коду (11,7) має вигляд 11010011100.

Приклад 6.8

Необхідно забезпечити передачу восьми команд кодом із парною кількістю одиниць. Визначити необхідну значність коду, розрахувати надлишковість і завадостійкість коду за умови, що ймовірність спотворення одного символу комбінації дорівнює $P_e = 10^{-3}$.

Розв'язування

1. Код із парною кількістю одиниць будується шляхом додавання до інформаційних символів комбінації одного надлишкового (контрольного) символу. Причому в тих випадках, коли інформаційна частина комбінації містить парну кількість одиниць, додатковий символ представляється нулем, при непарній кількості одиниць в інформаційній групі додатковий символ представляється одиницею.

Значність коду, у такий спосіб:

$$n = k + 1,$$

де

$$k = \frac{\lg N}{\lg 2} = \frac{\lg 8}{\lg 2} = 3.$$

Отже, $n = 4$.

Можливі варіанти кодових комбінацій подані в таблиці 6.10.

Таблиця 6.10

| № з/п | Символ | | № з/п | Символ | |
|----------|---------------|------------|----------|---------------|------------|
| | інформаційний | перевірний | | інформаційний | перевірний |
| 1 | 000 | 0 | 5 | 100 | 1 |
| 2 | 001 | 1 | 6 | 101 | 0 |
| 3 | 010 | 1 | 7 | 110 | 0 |
| 4 | 011 | 0 | 8 | 111 | 1 |

Як видно з таблиці, код характеризується мінімальною відстанню $d_{\min} = 2$. Однак завдяки спеціальній логіці побудови і перевірки коду, він дозволяє виявляти не тільки одиничні помилки, але і всі помилки непарної кратності (у нашому випадку одно- і трикратні).

2. Надлишковість коду:

$$K_{\text{надл}} = \frac{P}{n} = \frac{1}{4} = 0,25.$$

3. Завадостійкість коду будемо оцінювати за формулами (6.17) та (6.18).

Для оцінювання завадостійкості коду будемо думати, що ймовірності спотворення окремих символів кодової комбінації однакові і рівні P_e , а спотворення незалежні.

Код дозволяє виявляти однократні помилки непарної кратності, тому що тільки в цих випадках кількість одиниць у комбінації стане непарною. Не виявляються помилки парної кратності.

Нехай у комбінації спотворюються два певних символи, а інші не спотворюються, тоді ймовірність такої події дорівнює $P_e^2 (1 - P_e)^{n-2}$. Оскільки таких варіантів буде C_n^2 , то ймовірність двократних помилок:

$$P_{\text{пом}}^{(2)} = C_n^2 P_e^2 (1 - P_e)^{n-2}.$$

За аналогією ймовірність чотирікратних помилок:

$$P_{\text{пом}}^{(4)} = C_n^4 P_e^4 (1 - P_e)^{n-4}.$$

Тому сумарна ймовірність появи незайдених помилок:

$$P_{n,\text{пом}} = P_{\text{пом}}^{(2)} + P_{\text{пом}}^{(4)} = C_n^2 P_e^2 (1 - P_e)^{n-2} + C_n^4 P_e^4 (1 - P_e)^{n-4}.$$

Оскільки із зростанням кратності ймовірність помилок різко спадає, то:

$$P_{n,\text{пом}} \cong C_n^2 P_e^2 (1 - P_e)^{n-2}.$$

Отже, завадостійкість коду:

$$S = \lg \frac{1}{P_{n,\text{пом}}} \cong \lg \frac{1}{C_n^2 P_e^2 (1 - P_e)^{n-2}} = \lg \frac{1}{n(n-1) P_e^2 (1 - P_e)^{n-2}} = 6,78.$$

Для знаходження коефіцієнта виявлення $K_{\text{виявл}}$ необхідно визначати ймовірність появи помилок, що виявляються, яка для коду з перевіркою на парність:

$$P_{\text{в.ком}} = C_n^1 P_e (1 - P_e)^{n-1} + C_n^3 P_e^3 (1 - P_e)^{n-3} \cong C_n^1 P_e (1 - P_e)^{n-1}.$$

Ймовірність появи всіх помилок, як тих, що виявляються, так і тих, що не виявляються, може бути визначена, виходячи з наступних міркувань. Якщо ймовірність спотворення одного символу кодової комбінації дорівнює P_e , то при незалежності помилок ймовірність того, що всі n символів комбінації не будуть спотворені, дорівнює $(1 - P_e)^n$. Тоді ймовірність спотворення кодової комбінації:

$$P_{\text{в}} = 1 - (1 - P_e)^n. \quad (5.42)$$

Отже, коефіцієнт виявлення:

$$\begin{aligned} K_{\text{виявл}} &= \frac{P_{\text{в.пом}}}{P_{\text{в}}} \cong \frac{C_n^1 P_e (1 - P_e)^{n-1}}{1 - (1 - P_e)^n} \cong \frac{n P_e (1 - P_e)^{n-1}}{n P_e} = (1 - P_e)^{n-1} = \\ &= (1 - 10^{-3})^3 = 0,991. \end{aligned}$$

Приклад 6.9

Необхідно забезпечити передачу 100 команд інверсним кодом. Визначити необхідну значність коду, розрахувати надлишковість та завадостійкість коду за умови, що ймовірність одного символу комбінації $P_e = 10^3$.

Розв'язування

1. В основу побудови інверсного коду покладений метод повторення початкової комбінації. Причому в тих випадках, коли початкова комбінація містить парне число одиниць, друга комбінація в точності відтворює початкову, якщо ж початкова комбінація містить непарне число одиниць, то повторення відбувається в інвертованому вигляді.

Кількість інформаційних символів коду повинна бути:

$$k = \frac{\lg N}{\lg 2} = \frac{\lg 100}{\lg 2} = 6,6.$$

Вибираємо $k = 7$.

Загальна кількість символів повинна в два рази перевищувати кількість інформаційних символів: $n = 2 \cdot k = 14$.

Окремі варіанти кодових комбінацій наведені в таблиці 6.11.

2. Надлишковість коду:

$$K_{\text{надл}} = \frac{p}{n} = \frac{7}{14} = 0,5.$$

Таблиця 6.11

| № з/п | Символ | | № з/п | Символ | |
|-------|---------------|------------|-------|---------------|------------|
| | інформаційний | перевірний | | інформаційний | перевірний |
| 1 | 0000000 | 0000000 | 5 | 0000100 | 1111011 |
| 2 | 0000001 | 1111110 | ... | ... | ... |
| 3 | 0000010 | 1111101 | 10 | 0001001 | 0001001 |
| 4 | 0000011 | 0000011 | 15 | 0001110 | 1110001 |

3. Завдостійкість коду оцінюється за формулами (6.17) та (6.18). Як видно з таблиці 6.11, код характеризується відстанню $d_{\min} = 4$. Однак, завдяки спеціальній логіці побудови і перевірки коду, він дозволяє практично виявляти майже всі можливі помилки.

Найбільш ймовірним видом помилок, що не виявляються, є одночасне спотворення двох символів у початковій комбінації та відповідних їм двох символів у комбінації, що повторюється.

Ймовірність одночасного спотворення якої-небудь пари символів у початковій комбінації дорівнює $C_{n/2}^2 P_e^2 (1 - P_e)^{n/2-2}$. Ймовірність одночасного спотворення двох пар відповідних символів:

$$P_{n, пом} = \left[C_{n/2}^2 P_e^2 (1 - P_e)^{n/2-2} \right]^2 \cong (C_{n/2}^2)^2 P_e^4.$$

Завдостійкість коду:

$$S = \lg \frac{1}{P_{n, пом}} \cong \lg \frac{1}{(C_{n/2}^2)^2 P_e^4} = \lg \frac{1}{(C_7^2)^2 10^{-12}} = 9,36.$$

Коефіцієнт виявлення:

$$K_{обн} = \frac{P_{e, пом}}{P_k} = 1 - \frac{P_{n, пом}}{P_k} \cong 1 - \frac{(C_{n/2}^2)^2 P_e^4}{1 - (1 - P_e)^n} \cong 1 - \frac{(C_{n/2}^2)^2 P_e^3}{n} \\ \cong 1 - \frac{(n-2)(C_{n/2}^2)^2 P_e^3}{8} = 1 - \frac{(14-2)21^2 \cdot 10^{-9}}{8} = 0,99999.$$

Приклад 6.10

Побудувати код Хеммінга, що забезпечує усунення однократних помилок і призначений для передачі 100 команд. Оцінити надлишковість і завдостійкість коду, якщо ймовірність одного символу комбінації $P_e = 10^{-3}$.

Розв'язування. Кількість інформаційних символів коду:

$$k = \frac{\lg N}{\lg 2} = \frac{\lg 100}{\lg 2} = 6,6.$$

Вибираємо $k = 7$.

Необхідну значність коду визначимо із співвідношення:

$$N \leq \frac{2^n}{1+n}$$

При $N = 100$ значність коду $n = 11$.

Кількість перевірних груп визначається кількістю надлишкових символів:

$$\rho = n - k = 4.$$

Оскільки значність коду $n = 11$, то перевірні групи повинні містити у собі наступні символи:

$$S_1 = a_1 + a_3 + a_5 + a_7 + a_9 + a_{11};$$

$$S_2 = a_2 + a_3 + a_6 + a_7 + a_{10} + a_{11};$$

$$S_3 = a_4 + a_5 + a_6 + a_7;$$

$$S_4 = a_8 + a_9 + a_{10} + a_{11}.$$

Контрольні символи: a_1, a_2, a_4, a_8 .

При складанні кодових комбінацій спочатку проставляються інформаційні символи, а потім контрольні так, щоб сума одиниць в кожній контрольній групі була парною.

Окремі варіанти кодових комбінацій наведені у таблиці 6.12.

Таблиця 6.12

| № з/п | Символ коду | | | | | | | | | | |
|-------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|
| | a_1 | a_2 | a_3 | a_4 | a_5 | a_6 | a_7 | a_8 | a_9 | a_{10} | a_{11} |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

Надлишковість коду:

$$K_{надл} = \frac{\rho}{n} = \frac{4}{11} = 0,36.$$

Оскільки код може бути використаний для виправлення однократних або виявлення двократних помилок, то ймовірність появи помилок, що не виявляються:

$$P_{\text{н.пом}} \cong C_n^3 P_e^3 (1 - P_e)^{n-3}.$$

Тому завадостійкість коду:

$$S = \lg \frac{1}{P_{\text{н.пом}}} = \lg \frac{1}{C_n^3 P_e^3 (1 - P_e)^{n-3}} \cong \lg \frac{1}{C_n^3 P_e^3} = \lg \frac{1}{C_n^3 10^{-3}} = 6,78.$$

Коефіцієнт виявлення:

$$\begin{aligned} K_{\text{виявл}} &= 1 - \frac{P_{\text{н.пом}}}{P_{\text{к}}} = 1 - \frac{C_n^3 P_e^3 (1 - P_e)^{n-3}}{1 - (1 - P_e)^n} \cong 1 - \frac{C_n^3 P_e^3 (1 - P_e)^{n-3}}{nP_e} \cong \\ &\cong 1 - \frac{(n-1)(n-2)P_e^2 (1 - P_e)^{n-3}}{6} = 0,99998. \end{aligned}$$

Приклад 6.11

Закодувати просту інформаційну групу $G(X) = 1011$ циклічним кодом, що забезпечує виявлення двократних або усунення однократних помилок.

Розв'язування. За заданою кількістю інформаційних символів $k = 4$ визначаємо значність коду. Користуючись співвідношенням (6.33) та таблицею 6.6, отримаємо $n = 7$.

Для побудови циклічного коду необхідно вибрати твірний поліном $P(x)$ степеня $\rho = n - k = 3$, який повинен входити як співмножник для розкладання двочлена $x^n + 1 = x^{2z-1} - 1$. У нашому випадку цей двочлен має вигляд $x^7 + 1$. Співмножники, що входять до нього, повинні бути незвідними поліномами, степені яких є дільниками числа $z = 3$. Отже, співмножниками двочлена $x^7 + 1$ повинні бути незвідні поліноми першої та третьої степенів.

Користуючись таблицями незвідних поліномів [13], отримаємо:

$$x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1).$$

Як твірний поліном вибираємо співмножник:

$$P(x) = x^3 + x^2 + 1.$$

Кодування здійснюємо першим способом. Для цього початкову кодову комбінацію $G(x)$ множимо на $x^{n-k} = x^3$:

$$x^{n-k}G(x) = x^3(x^3 + x + 1) = x^6 + x^4 + x^3.$$

Визначаємо залишок $R(x)$ від ділення $x^{n-k}G(x)$ на твірний багаточлен $P(x)$:

$$\begin{array}{r} x^6 + x^4 + x^3 \\ + x^6 + x^5 + x^3 \\ \hline x^5 + x^4 \\ + x^5 + x^4 + x^3 \\ \hline x^3 \end{array} \left| \begin{array}{l} x^3 + x^2 + 1 \\ x^3 + x^2 \end{array} \right.$$

Залишок $R(x) = x^2$. Отже, поліном $F(x)$ циклічного коду відповідно до (6.33) буде мати вигляд:

$$F(x) = x^{n-k}G(x) + R(x) = x^6 + x^4 + x^3 + x^2 = 1011100.$$

Приклад 6.12

Отримано повідомлення циклічним кодом:

$$F^*(x) = x^6 + x^4 + x^3 + x^2.$$

Перевірити декодуванням наявність помилок в прийнятій комбінації, якщо твірний поліном:

$$P(x) = x^3 + x^2 + 1.$$

Розв'язування. Декодування здійснюється діленням полінома отриманої комбінації на твірний поліном:

$$\begin{array}{r} x^6 + x^4 + x^3 + x^2 \\ + x^6 + x^5 + x^3 \\ \hline x^5 + x^4 + x^2 \\ + x^5 + x^4 + x^2 \\ \hline 0 \end{array} \left| \begin{array}{l} x^3 + x^2 + 1 \\ x^3 + x^2 \end{array} \right.$$

Залишок $R(x) = 0$. Отже, комбінація прийнята без спотворень.

Приклад 6.13

Отримана комбінація

$$F^*(x) = x^6 + x^4 + x^2 + 1,$$

закодована циклічним кодом. Твірний поліном:

$$P(x) = x^3 + x^2 + 1.$$

Перевірити наявність помилок у кодовій комбінації.

Розв'язування. Ділимо поліном отриманої комбінації на твірний поліном:

$$\begin{array}{r} x^6 + x^4 + x^3 + 1 \\ + x^6 + x^5 + x^3 \\ \hline x^5 + x^4 + x^3 + x^2 + 1 \\ + x^5 + x^4 + x^2 \\ \hline x^3 + 1 \\ + x^3 + x^2 + 1 \\ \hline x^2 \end{array}$$

Залишок від ділення $R(x) = x^2 \neq 0$. Отже, комбінація прийнята з помилками.

Приклад 6.14

Побудувати твірну матрицю циклічного коду при $n = 7$; $k = 4$.
Твірний поліном:

$$P(x) = x^2 + x + 1.$$

Розв'язування. Оскільки $k = 4$, інформаційна підматриця:

$$U_k = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix}.$$

Для отримання першого рядка додаткової підматриці множимо перший рядок інформаційної підматриці на x^3 і ділимо на твірний поліном, тобто виконуємо операції $\frac{0001 \cdot 1000}{1011}$. Залишок від цих операцій 011 складає перший рядок додаткової підматриці. Аналогічно визначаються інші рядки додаткової підматриці. Додаткова підматриця має вигляд:

$$H_p = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix}.$$

Отримасмо наступну твірну матрицю:

$$P_{n,k} = |U_k, H_p| = \begin{vmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{vmatrix}.$$

Для побудови твірної матриці другим способом перший рядок матриці отримусмо шляхом множення твірного полінома на x^{k-1} : $P(x) \cdot x^{k-1} = 1011 \cdot 1000 = 1011000$. Подальші рядки матриці отримусмо шляхом циклічного зсуву отриманої комбінації. Таким чином, твірна матриця:

$$P_{n,k} = \begin{vmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{vmatrix}.$$

Приклад 6.15

Визначити параметри двовимірного ітеративного коду, в якому, як код за рядками, використовується код з перевіркою на парність з параметрами $n_1 = 5$; $k_1 = 4$; $d_1 = 2$, а як код за стовпцями – код Хеммінга з параметрами $n_2 = 7$; $k_2 = 4$; $d_2 = 3$.

Розв'язування. Згідно з виразами (5.39–5.41), параметри ітеративного коду наступні:

$$n = n_1 \cdot n_2 = 35, \quad k = k_1 \cdot k_2 = 16, \quad d = d_1 \cdot d_2 = 6.$$

Контрольні питання

1. Що розуміється під кодуванням в широкому і вузькому сенсі?
2. Яка мета кодування?
3. Яка особливість рівномірних і нерівномірних кодів?
4. Що обумовило значне поширення двійкових кодів?
5. У чому суть позиційного принципу побудови кодів?
6. У чому особливість складених кодів і для яких цілей вони використовуються?
7. Для яких цілей використовуються рефлексні коди?
8. Що розуміється під завадостійкими кодами?
9. У чому полягає відмінність між блочними і безперервними кодами?

10. У чому особливість систематичних кодів?
11. У чому суть завадостійкого кодування?
12. Що розуміється під значністю та вагою кодової комбінації?
13. Як визначається відстань між кодovими комбінаціями?
14. Що таке вектор помилки?
15. Який зв'язок коригувальної здатності коду з кодовою відстанню?
16. Якими показниками можна оцінювати якість коригувальних кодів?
17. Яким вимогам повинні задовольняти початкові комбінації систематичного коду?
18. Як будується твірна матриця систематичного коду?
19. Які умови побудови перевірної підматриці?
20. Який алгоритм визначення перевірних символів за інформаційними за допомогою перевірної матриці?
21. Як визначається склад контрольних рівностей за допомогою перевірної матриці?
22. Які коригувальні властивості коду з парним числом одиниць?
23. Як будується код з подвоєнням елементів, які його коригувальні властивості?
24. Яка логіка побудови і перевірки інверсного коду, які коригувальні властивості цього коду?
25. Який принцип побудови кодів Хеммінга?
26. Яким чином складаються перевірні групи коду Хеммінга?
27. Які властивості циклічного коду?
28. Які відомі способи побудови циклічних кодів?
29. Яким чином вибирається твірний поліном циклічного коду?
30. Які основні різновиди циклічних кодів і їх властивості?
31. У чому відмінна особливість ітеративних кодів?
32. Яка процедура виявлення і виправлення помилок в ітеративних кодах?
33. Яке основне призначення рекурентних кодів і як вони будуються?
34. Який принцип побудови і дешифрування ланцюгового коду?

Будь-який інформаційний пристрій працює за єдиною схемою (рис. 7.1). На вхід пристрою подається сукупність повідомлень x_1, x_2, \dots, x_n . Задача пристрою полягає у тому, щоб передати сукупність цих повідомлень з достатньо високою достовірністю, або, іншими словами, щоб перевести вектор повідомлень на вході $X(x_1, x_2, \dots, x_n)$ у відповідний йому вектор повідомлень на виході $Y(y_1, y_2, \dots, y_n)$ без помилок або з допустимими помилками.

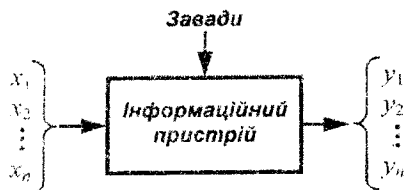


Рис. 7.1

У процесі передачі повідомлення може підлягати багаторазовим перетворенням, які істотно змінюють його фізичні характеристики. Проте інформація, яка передається, повинна залишатися інваріантною при всіх перетвореннях. Природно, що кількість інформації, яка передається одержувачу, пов'язана з невизначеністю, яка мала місце відносно повідомлення, яке передавалось. У зв'язку з цим необхідно ввести кількісну міру оцінювання повідомлення і невизначеності повідомлень, які передаються.

7.1 Кількість інформації і невизначеність. Ентропія як міра невизначеності

Фундаментальним питанням для теорії інформації є питання про кількісну міру інформації. Необхідно відмітити, що вся інформація, яку отримує споживач після того, як прийняли повідомлення, одержана в результаті експерименту. Повідомлення, отримане на приймальній

стороні, несе корисну інформацію тільки у тому випадку, якщо є невизначеність відносно стану джерела повідомлення.

Якщо експеримент має тільки один результат і не містить ніякої невизначеності, то спостерігач заздалегідь буде знати результат цього експерименту. У результаті проведення такого експерименту спостерігач не отримає ніякої інформації.

Нехай експеримент має два однакових результати. Результат контролю, наприклад, повинен вказувати, що параметр, який контролюється, знаходиться в межах норми або за її межами. Повідомлення, яке передається в цьому випадку, може мати два значення і містити визначену інформацію.

Розглянемо третій приклад – джерело напруги, вихідна напруга якого може з однаковою ймовірністю мати десять різних значень. У цьому випадку буде велика попередня невизначеність відносно джерела, а повідомлення, яке надійшло (конкретний результат експерименту), дасть більш конкретну характеристику стану входу джерела.

Розглянемо загальний випадок, коли джерело може передавати незалежні і несумісні повідомлення x_1, x_2, \dots, x_n з ймовірністю $p(x_1), p(x_2), \dots, p(x_n)$ відповідно. Природно, що чим менша апіорна ймовірність події, тим більшу кількість інформації воно несе. Так, наприклад, повідомлення про те, що літом температура повітря в Криму вище нуля, не несе істотної інформації, оскільки ймовірність такої події дуже велика. Повідомлення про те, що температура повітря на південному березі Криму в червні нижче нуля, містить значно більшу кількість інформації, оскільки така подія є дуже рідкісною.

Тому природно припустити, що кількісна міра невизначеності окремого повідомлення, а також інформація, яку воно передає, може бути величиною, зворотною її апіорній ймовірності $\frac{1}{p(x_i)}$.

Проте така міра незручна тим, що у випадку, коли експеримент буде мати тільки один результат, тобто ймовірність такої події дорівнює одиниці, і кількість інформації, згідно з прийнятою мірою, дорівнює одиниці. У дійсності ж результат такого експерименту не дає ніякої інформації. Крім того, така міра не володіє властивістю адитивності. Дійсно, якщо має місце складний випадок, який складається з двох незалежних випадків x_i і x_j , то ймовірність такого випадку буде визначатися множенням ймовірностей. Кількість інформації в складному повідомленні повинна оцінюватися величиною $\frac{1}{p(x_i)p(x_j)}$. З

цієї точки зору більш зручною є логарифмічна міра інформації:

$$I(x_i) = \log_a \frac{1}{p(x_i)}. \quad (7.1)$$

При цьому кількість інформації, яка міститься в складному повідомленні, яке показує сукупність подій x_i і x_j , буде дорівнювати:

$$I(x_i, x_j) = \log_a \frac{1}{p(x_i)p(x_j)} = \log_a \frac{1}{p(x_i)} + \log_a \frac{1}{p(x_j)} = I(x_i) + I(x_j).$$

Як бачимо, логарифмічна міра володіє властивістю адитивності. Крім того, ця міра у випадку подій з одним результатом дає нульову кількість інформації.

Вираз (7.1) характеризує кількість інформації, яка міститься в повідомленні x_i . Він характеризує також невизначеність цього повідомлення. У зв'язку з цим вираз (7.1) може бути використаний для кількісної оцінки невизначеності повідомлення:

$$H(x_i) = \log_a \frac{1}{p(x_i)}. \quad (7.2)$$

Величину (7.2), яка характеризує невизначеність окремого (i -го) повідомлення, прийнято називати **частковою ентропією**.

Кількість інформації та невизначеність для всієї сукупності випадкових повідомлень можна отримати, якщо взяти середню величину за всіма випадками:

$$I(X) = \sum_{i=1}^n p(x_i) \log_a \frac{1}{p(x_i)} = - \sum_{i=1}^n p(x_i) \log_a p(x_i), \quad (7.3)$$

$$H(X) = - \sum_{i=1}^n p(x_i) \log_a p(x_i). \quad (7.4)$$

Залежності (7.3) і (7.4) виражають середню на один випадок (повідомлення) кількість інформації та ентропії. Термін "ентропія" запозичений з термодинаміки, де аналогічні вирази характеризують середню невизначеність стану системи молекул речовини.

Незважаючи на збіг залежностей (7.3) і (7.4), ентропія $H(X)$ і кількість інформації $I(X)$ принципово відрізняються. Ентропія $H(X)$, яка виражає середню невизначеність стану джерела повідомлення, є об'єктивною характеристикою джерела повідомлення і, якщо відома статистика повідомлень, може бути підрахована апріорно, тобто до

отримання повідомлення. Кількість інформації $I(X)$ є апостеріорною характеристикою і визначає кількість інформації, отриманої з надходженням повідомлень. Ентропія $H(X)$ є мірою недостачі інформації про стан окремої системи. З надходженням інформації про стан системи ентропія останньої понижується.

Збіг виразу (7.3) з виразом (7.4) свідчить про те, що кількість отриманої інформації дорівнює чисельно ентропії, яка мала місце відносно джерела повідомлення.

На розглянутому взаємозв'язку кількості інформації з ентропією проявляється відомий діалектичний закон єднання та боротьби протилежностей, тому що інформація розглядається у зв'язку із своєю протилежністю – ентропією і, з іншого боку, розглядається як міра знищення, зняття ентропії.

Одиниці вимірювання кількості інформації та ентропії залежать від вибору основи логарифма у формулах (7.3) і (7.4). При використанні десяткових логарифмів кількість інформації та ентропія визначаються в десяткових одиницях – дитах. У випадку використання двійкових логарифмів вимірюється кількість інформації та ентропії у двійкових одиницях – бітах. При використанні натуральних логарифмів одиницею вимірювання є натуральна одиниця – нит.

При аналізі інформаційних процесів в електронних обчислювальних машинах та інших пристроях, які функціонують на основі двійкової системи числення, зручно користуватися двійковими одиницями. При аналізі процесів у пристроях, які працюють в десятковій (або двійково-десятковій) системі обчислення, доцільно користуватися дитами. У математичних викладах зручно користуватися натуральними одиницями. У подальшому будемо користуватися двійковими одиницями (дв. од.).

Міра кількості інформації у вигляді виразу (7.3) вперше була запропонована К. Шенноном у 1948 р. й потім більш точно визначена А.Я. Хінчиним.

У випадку рівної ймовірності повідомлень вираз (7.3) для кількості інформації приводиться до вигляду:

$$I(X) = -\log_2 p(x_i) = \log_2 n, \quad (7.5)$$

де $n = \frac{1}{p(x_i)}$ – кількість переданих повідомлень.

Така міра кількості інформації була запропонована у 1928 р. Р. Хартлі.

7.2 Властивості ентропії дискретних повідомлень

Формула (7.4) виражає ентропію дискретних повідомлень.

Ентропія дискретних повідомлень має наступні властивості:

- вона є величиною речовинною, обмеженою і невід'ємною. Ця властивість випливає з виразу (7.4), якщо врахувати, що ймовірності $p(x_i)$ – величини невід'ємні, замкнені на відрізку $0 \leq p(x_i) \leq 1$;
- ентропія детермінованих повідомлень дорівнює нулю;
- ентропія максимальна, якщо повідомлення рівноймовірні.

У випадку n рівноймовірних повідомлень їх ймовірності дорівнюють $p(x_1) = p(x_2) = \dots = p(x_n) = 1/n$. При цьому ентропія повідомлення буде дорівнювати:

$$H(X)_{\max} = -\sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n. \quad (7.6)$$

Як видно із виразу (7.6), у випадку рівноймовірних подій ентропія збільшується зі збільшенням кількості подій.

4. Ентропія системи двохальтернативних подій може вимірюватися у межах від нуля до одиниці. Ентропія системи двох подій:

$$\begin{aligned} H(X) &= -p(x_1) \log_2 p(x_1) - p(x_2) \log_2 p(x_2) = \\ &= -p(x_1) \log_2 p(x_1) - [1 - p(x_1)] \log_2 [1 - p(x_1)]. \end{aligned}$$

З останнього виразу видно, що ентропія дорівнює нулю, якщо $p(x_1) = 0$, $p(x_2) = 1$ або $p(x_1) = 1$, $p(x_2) = 0$. Максимум ентропії буде мати місце, коли $p(x_1) = p(x_2)$. При цьому максимальне значення ентропії: $H(X)_{\max} = -0,5 \log_2 1/2 - 0,5 \log_2 1/2 = 1$ дв. од.

Таким чином, можна стверджувати, що одна двійкова одиниця – це ентропія системи двох рівноймовірних незалежних подій.

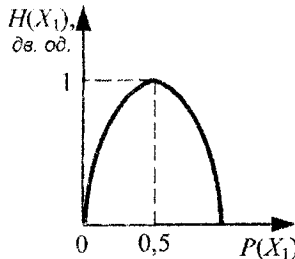


Рис. 7.2

На рис. 7.2 показана графічна залежність ентропії бінарних подій від ймовірності однієї з подій.

7.3 Ентропія неперервних подій

Неперервне повідомлення як випадкова величина характеризується диференціальним законом розподілу ймовірностей $\omega(x)$. Нехай функція розподілу щільності ймовірності неперервного повідомлення має вигляд, наведений на рис. 7.3.

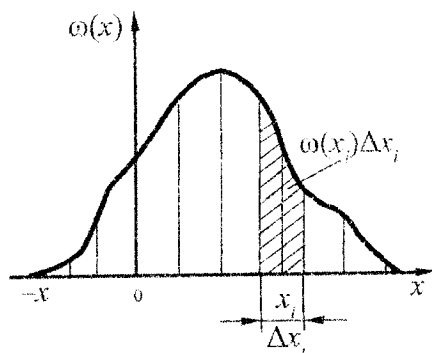


Рис. 7.3

Для визначення ентропії неперервних подій скористаємося вихідним виразом (7.4) для ентропії дискретних повідомлень. У зв'язку з цим розіб'ємо шкалу рівнів неперервних випадкових величин x на невеликі відрізки Δx_i і всередині кожної ділянки виберемо точку x_i так, щоб виконувалась умова:

$$\int_{\Delta x_i} \omega(x) dx = \omega(x_i) \Delta x_i. \quad (7.7)$$

Вираз (7.7) характеризує ймовірність $p(x_i)$ попадання випадкової величини x у відрізок Δx_i (див. рис. 7.3) при заміні неперервної випадкової величини x сукупністю дискретних повідомлень x_i , ймовірність надходження яких визначається виразом (7.4). Така заміна буде тим точніше, чим менше відрізки Δx_i .

Ентропія еквівалентного дискретного повідомлення згідно з виразом (7.4):

$$\begin{aligned}
 H(X) &= -\sum_{i=1}^n p(x_i) \log_2 p(x_i) = -\sum_{i=1}^n \omega(x_i) \Delta x_i \log_2 [\omega(x_i) \Delta x_i] = \\
 &= -\sum_{i=1}^n \omega(x_i) \Delta x_i \log_2 \omega(x_i) - \sum_{i=1}^n \omega(x_i) \Delta x_i \log_2 \Delta x_i.
 \end{aligned}$$

Щоб отримати вираз для ентропії неперервного повідомлення, зробимо граничний перехід при:

$$\begin{aligned}
 \lim_{\Delta x_i \rightarrow 0} \left\{ -\sum_{i=1}^n \omega(x_i) \Delta x_i \log_2 \omega(x_i) - \sum_{i=1}^n \omega(x_i) \Delta x_i \log_2 \Delta x_i \right\} = \\
 = -\int_{-\infty}^{\infty} \omega(x) \log_2 \omega(x) dx - \lim_{\Delta x_i \rightarrow \infty} \log_2 \Delta x_i,
 \end{aligned} \tag{7.8}$$

оскільки $\lim_{\Delta x_i \rightarrow 0} \sum_{i=1}^n \omega(x_i) \Delta x_i = \int_{-\infty}^{\infty} \omega(x) dx = 1$.

Другий член виразу (7.8) при $\Delta x_i \rightarrow 0$ прагне до нескінченності. Отже ентропія неперервного повідомлення повинна дорівнювати нескінченності. Проте в реальних умовах відлік повідомлень на приймальній стороні відбувається в дискретних точках внаслідок скінченної точності і роздільної здатності апаратури, тобто відрізки Δx_i мають скінченну величину. Тому вираз (7.8), який визначає ентропію неперервного повідомлення, має дві складові, одна з яких визначається законом розподілу повідомлень, а друга є постійною величиною і звичайно виключається з розгляду.

Перший доданок

$$h(x) = -\int_{-\infty}^{\infty} \omega(x) \log_2 \omega(x) dx \tag{7.9}$$

являє собою так звану диференціальну ентропію. Диференціальна ентропія залежить від статистики повідомлень. Якщо припустити, що величина x безрозмірна, то $h(x)$ можна вимірювати в двійкових одиницях. Потрібно мати на увазі, що на відміну від ентропії дискретних повідомлень величина $h(x)$ є відносною, тому що залежить від масштабу, тобто, від вибору одиниць вимірювання. У зв'язку з цим $h(x)$ окремо не може бути абсолютною мірою невизначеності неперервного повідомлення. Вона не володіє багатьма властивостями звичайної ентропії, частково може приймати і від'ємні значення. Інформаційне значення має не сама диференціальна ентропія, а різниця двох диференціальних ентропій.

При заданій середній потужності дисперсії максимальну ентропію має нормальний закон розподілу ймовірності. Якщо ж задана найбільша потужність, то максимальну ентропію має рівномірний закон розподілу.

7.4 Ентропія складних повідомлень

При вирішенні задач передачі інформації часто мають справу з декількома джерелами, які надають взаємопов'язані повідомлення. Сукупність повідомлень, вироблених декількома джерелами, називається *складним повідомленням*.

Нехай є два джерела повідомлень. Повідомлення першого джерела приймають значення x_1, x_2, \dots, x_n з ймовірностями $p(x_1), p(x_2), \dots, p(x_n)$, повідомлення другого джерела приймають значення y_1, y_2, \dots, y_m з ймовірностями $p(y_1), p(y_2), \dots, p(y_m)$.

Сумісну ентропію сукупності повідомлень X і Y можна представити у вигляді:

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j), \quad (7.10)$$

де $p(x_i, y_j)$ – ймовірність спільної появи повідомлень x_i і y_j .

Враховуючи, що спільна ймовірність може бути представлена у вигляді $p(x_i, y_j) = p(x_i) \cdot p(y_j/x_i)$, де $p(y_j/x_i)$ – умовна ймовірність повідомлення про надходження повідомлення, вираз (7.10.) можна привести до вигляду:

$$\begin{aligned} H(X, Y) &= - \sum_{i=1}^n \sum_{j=1}^m p(x_i) p(y_j/x_i) \log_2 [p(x_i) p(y_j/x_i)] = \\ &= - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \sum_{j=1}^m p(y_j/x_i) - \sum_{i=1}^n p(x_i) \sum_{j=1}^m p(y_j/x_i) \log_2 p(y_j/x_i). \end{aligned}$$

З урахуванням того, що $\sum_{j=1}^m p(y_j/x_i)$, тому що при наявності повідомлення x_i обов'язково буде одне з повідомлень ансамблю y_j , вираз для сумісної ентропії може бути перетворений наступним чином:

$$\begin{aligned} H(X, Y) &= - \sum_{i=1}^n p(x_i) \log_2 p(x_i) - \sum_{i=1}^n p(x_i) \sum_{j=1}^m p(y_j/x_i) \times \\ &\times \log_2 p(y_j/x_i) = H(X) + \sum_{i=1}^n p(x_i) H(Y/x_i), \end{aligned} \quad (7.11)$$

де $H(Y/x_i) = -\sum_{j=1}^m p(y_j/x_i) \log_2 p(y_j/x_i)$ – так звана часткова

умовна ентропія, яка виражає ентропію повідомлення Y за умови, що мало місце повідомлення X . Другий член виразу (6.11) являє собою усереднення $H(Y/x_i)$ за всіма повідомленнями x_1, x_2, \dots, x_n і називається середньою умовною ентропією повідомлення Y при умові надходження повідомлення X . Якщо позначити його через $H(Y/X)$, то:

$$H(X/Y) = H(X) + H(Y/X), \quad (7.12)$$

$$\begin{aligned} \text{де } H(Y/X) &= -\sum_{i=1}^n p(x_i) H(Y/x_i) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i) p(y_j/x_i) \times \\ &\times \log_2 p(y_j/x_i) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(y_j/x_i). \end{aligned}$$

Основний зміст умовної ентропії $H(Y/X)$ полягає в тому, що вона показує, яку ентропію дає повідомлення Y , коли вже відома ентропія повідомлень X .

З рівняння $H(X/Y) = H(Y/X)$ отримаємо:

$$H(X, Y) = H(Y) + H(X/Y),$$

$$\text{де } H(X/Y) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i/y_j).$$

Таким чином, сумісна ентропія двох повідомлень дорівнює сумі безумовної ентропії одного із повідомлень і умовної ентропії другого повідомлення.

При деякій множині повідомлень X, Y, Z, \dots сумісна ентропія дорівнює:

$$\begin{aligned} H(X, Y, Z, \dots) &= H(X) + H(Y, Z, \dots/X) = \\ &= H(Y) + H(X, Z, \dots/Y) = H(Z) + H(X, Y, \dots/Z) = \dots \end{aligned}$$

Можна відмітити наступні основні властивості ентропії складних повідомлень.

1. При статистично незалежних повідомленнях X і Y сумісна ентропія дорівнює сумі ентропій повідомлень:

$$H(X, Y) = H(X) + H(Y). \quad (7.13)$$

Дійсно, при статистично незалежних повідомленнях умовна ймовірність дорівнює безумовній $p(y_j/x_i) = p(y_j)$, а сумісна ймовірність дорівнює добутку ймовірностей $p(y_j/x_i) = p(x_i) + p(y_j)$. Тоді умовна ентропія дорівнює $H(Y/X) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i)p(y_j) \log_2 p(y_j)$.

Виконавши підсумовування по i , з урахуванням рівності $\sum_{i=1}^n p(x_i) = 1$, отримаємо:

$$H(Y/X) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i)p(y_j) \log_2 p(y_j) = -\sum_{j=1}^m p(y_j) \log_2 p(y_j) = H(Y).$$

Отже, що при статистично незалежних повідомленнях умовна ентропія повідомлень дорівнює безумовній.

2. При повній статистичній залежності повідомлень X та Y сумісна ентропія дорівнює безумовній ентропії одного з повідомлень.

Дійсно, повна статистична залежність відповідає випадку, коли умовні ймовірності $p(y_j/x_i)$ і $p(x_i/y_j)$ дорівнюють нулю або одиниці. У такому випадку вираз:

$$p(y_j/x_i) \log_2(y_j/x_i) = 0 \text{ і } p(x_i/y_j) \log_2(x_i/y_j) = 0$$

і, відповідно, умовні ентропії $H(Y/X) = 0$ і $H(X/Y) = 0$.

Тоді отримаємо:

$$H(X,Y) = H(X) = H(Y). \quad (7.14)$$

3. Умовна ентропія може змінюватися в межах

$$0 \leq H(Y/X) \leq H(Y).$$

Оскільки умовна ентропія додатна, дорівнює нулю при повній статистичній залежності повідомлень, максимальна при повній статистичній незалежності повідомлень і рівна безумовній ентропії, то звідси безпосередньо випливає ця властивість.

4. Для сумісної ентропії завжди справедливе співвідношення:

$$H(X,Y) \leq H(X) + H(Y). \quad (7.15)$$

Дана властивість сумісної ентропії безпосередньо випливає з попереднього співвідношення.

Якщо X і Y є неперервними випадковими величинами, то за аналогією з виразом для безумовної ентропії (7.9) вираз для ентропії об'єднання повідомлень X і Y можна представити у вигляді:

$$h(X, Y) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(X, Y) dY = \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(X) dXdY - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(Y/X) dXdY.$$

Якщо провести інтегрування у першому доданку по Y і врахувати, що $\int_{-\infty}^{\infty} \omega(X, Y) dY = \omega(X)$, отримаємо

$$h(X, Y) = - \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX - \\ - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(Y/X) dXdY = h(X) + h(Y/X), \quad (7.16)$$

де $h(X) = - \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX$ – диференціальна ентропія повідомлення X ;

$h(X/Y) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(Y/X) dXdY$ – умовна диференціальна ентропія повідомлення Y ;

$\omega(X)$ – щільність розподілу X ;

$\omega(Y/X)$ – умовна щільність розподілу Y відносно X .

7.5 Кількість інформації при неповній достовірності повідомлень

У реальних умовах передача повідомлень відбувається під дією завад. Завади спотворюють повідомлення, внаслідок чого повідомлення на приймальній стороні будуть в тому чи іншому ступені відрізнятися від переданих, тобто буде мати місце неповна достовірність передачі. Оцінимо кількість переданої інформації для випадків дискретних і неперервних повідомлень.

7.5.1 Дискретні повідомлення

Внаслідок відмінності повідомлень, що приймаються, від повідомлень, що передаються, при оцінюванні кількості інформації, що передається, доцільніше розглядати дві системи: систему повідомлень, що передаються X , і систему повідомлень, що приймаються Y .

Нехай повідомлення, що передається, може приймати значення x_1, x_2, \dots, x_n з апіорними ймовірностями $p(x_1), p(x_2), \dots, p(x_n)$ відповідно.

Повідомлення, що приймаються, характеризуються сукупністю значень y_1, y_2, \dots, y_n . Найввістий завад порушує однозначну відповідність між переданими і прийнятими повідомленнями. Оскільки завади мають випадковий характер, то при прийомі якого-небудь повідомлення y_j , неможливо точно встановити, яке повідомлення було передане. Можна говорити лише про умовну ймовірність $p(x/y_j)$, яка визначає ймовірність передачі повідомлення x_i при умові, що буде прийняте повідомлення y_j .

Оцінимо кількість інформації, яка міститься в одному з прийнятих повідомлень y_j про одне з переданих повідомлень x_i .

Умовна ймовірність $p(x/y_j)$ свідчить про те, що існує невизначеність у повідомленні y_j відносно повідомлення x_i . Ця невизначеність може бути оцінена умовною ентропією:

$$H(x_i / y_j) = -\log_2 p(x_i / y_j). \quad (7.17)$$

Таким чином, внаслідок дії завад початкова апіорна ентропія повідомлення x_i , яка визначається кількісно виразом $H(x_i) = -\log_2 p(x_i)$, знімається при отриманні повідомлення y_j не повністю, а лише зменшується до значення $H(x_i/y_j)$. Кількість отриманої інформації в цьому випадку буде дорівнювати знятій частині невизначеності:

$$I(y_j / x_i) = H(x_i) - H(x_i / y_j) = \log_2 \frac{p(x_i / y_j)}{p(x_i)}. \quad (7.18)$$

Формула (7.18) виражає кількість інформації, яка міститься у прийнятому повідомленні y_j відносно переданого x_i . Цю кількість інформації прийнято називати частковою кількістю інформації, що міститься в повідомленні y_j відносно повідомлення x_i .

Середню кількість інформації про всі x_i , що містяться в одному прийнятому повідомленні y_j , можна отримати шляхом усереднення для всіх x_i :

$$I(y_j, X) = \sum_{i=1}^n p(x_i / y_j) I(y_j, x_i) = \sum_{i=1}^n p(x_i / y_j) \log_2 \frac{p(x_i / y_j)}{p(x_i)}. \quad (7.19)$$

Формула (7.19) виражає часткову кількість інформації, яка міститься у прийнятому повідомленні y_j відносно всієї сукупності переданих повідомлень X . Нарешті, для того щоб визначити кількість інформації, що міститься у всій сукупності прийнятих повідомлень Y відносно всієї сукупності переданих повідомлень X , необхідно здійснити усереднення для всіх y_j :

$$\begin{aligned} I(Y, X) &= \sum_{j=1}^n p(y_j) I(y_j; X) = \sum_{j=1}^n p(y_j) \sum_{i=1}^n p(x_i / y_j) \log_2 \frac{p(x_i / y_j)}{p(x_i)} = \\ &= \sum_{j=1}^n \sum_{i=1}^n p(y_j) p(x_i / y_j) \log_2 \frac{p(x_i / y_j)}{p(x_i)}. \end{aligned} \quad (7.20)$$

Якщо використати рівність $p(y_j) p(x_i / y_j) = p(x_i, y_j)$, отримаємо:

$$I(Y, X) = \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i) p(y_j)}. \quad (7.21)$$

Формули (7.20) і (7.21) визначають середню кількість інформації, що міститься в X відносно Y .

Формулу (7.21) можна привести до вигляду:

$$\begin{aligned} I(Y, X) &= \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log_2 p(x_i / y_j) - \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log_2 p(x_i) = \\ &= -H(X / Y) - \sum_{i=1}^n \sum_{j=1}^n p(x_i) p(y_j / x_i) \log_2 p(x_i). \end{aligned} \quad (7.22)$$

Якщо провести перетворення в другому доданку правої частини виразу (7.22) з урахуванням $\sum_{j=1}^n p(y_j / x_i) = 1$, то отримаємо:

$$J(Y, X) = -H(X / Y) - \sum_{i=1}^n p(x_i) \log_2 p(x_i) = H(X) - H(X / Y). \quad (7.23)$$

Таким чином, середня кількість інформації, отримана при неповній достовірності повідомлень, дорівнює різниці безумовної ентропії $H(X)$, яка характеризує початкову (апріорну) невизначеність повідомлень.

Якщо використати властивість ентропії

$$H(X / Y) = H(X, Y) - H(Y),$$

вираз (7.23) можна привести до вигляду:

$$I(Y, X) = H(X) + H(Y) - H(X, Y). \quad (7.24)$$

Отже, кількість інформації, що передається, може бути виражена через суму ентропії переданого X і прийнятого Y повідомлень, за винятком сумісної ентропії $H(X, Y)$. Оскільки $H(X, Y) = H(Y, X)$, то

$$H(X) + H(Y/X) = H(Y) + H(X/Y)$$

або

$$H(X) - H(X/Y) = H(Y) - H(Y/X).$$

З останньої рівності випливає, що:

$$I(Y, X) = I(X, Y), \quad (7.25)$$

тобто кількість інформації, яка міститься в повідомленні Y відносно повідомлень X , дорівнює кількості інформації, що міститься в Y відносно X . Тому $I(Y, X)$ та $I(X, Y)$ називають також повною взаємною інформацією.

7.5.2 Неперервні повідомлення

Для визначення кількості інформації при неперервних повідомленнях використаємо вираз (7.23), який визначає кількість інформації при дискретних повідомленнях. У цьому виразі початкова і остаточна ентропії повідомлень становлять відповідно:

$$\begin{aligned} H(X) &= -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \\ H(X/Y) &= -\sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log_2 p(x_i / y_j). \end{aligned} \quad (7.26)$$

Для переходу до неперервних випадкових величин виразимо ймовірність через функції розподілу щільності ймовірності:

$$\begin{aligned} p(x_i) &= \omega(x_i) \Delta x_i; & p(y_j) &= \omega(y_j) \Delta y_j; \\ p(x_i, y_j) &= \omega(x_i, y_j) \Delta x_i \Delta y_j; & p(x_i / y_j) &= \omega(x_i / y_j) \Delta x_i, \end{aligned}$$

де Δx_i , Δy_j – елементарні дільниці, на які розбиті шкали рівнів випадкових величин x та y ; $\omega(x_i)$, $\omega(y_j)$, $\omega(x_i, y_j)$, $\omega(x_i / y_j)$ – значення функцій розподілу при аргументах $X = x_i$, $Y = y_j$.

Тоді, якщо провести в (7.26) відповідні заміни, отримаємо:

$$\begin{aligned} I(Y, X) &= -\sum_{i=1}^n \omega(x_i) \Delta x_i \log_2 [\omega(x_i) \Delta x_i] + \\ &+ \sum_{i=1}^n \sum_{j=1}^n \omega(x_i, y_j) \Delta x_i \Delta y_j \log_2 [\omega(x_i / y_j) \Delta x_i]. \end{aligned}$$

Здійснимо граничний перехід при $\Delta x_i \rightarrow 0$ і $\Delta y_j \rightarrow 0$.
 Границя виразу для початкової ентропії $H(X)$ при $\Delta x_i \rightarrow 0$:

$$\lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0}} H(X) = - \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX - \lim_{\Delta x_i \rightarrow 0} \log_2 \Delta x_i.$$

Визначимо границю для остаточної ентропії:

$$\begin{aligned} & \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0}} \left\{ \sum_{i=1}^n \sum_{j=1}^n \omega(x_i, y_j) \Delta x_i \Delta y_j \log_2 \left[\omega(x_i / y_j) \Delta x_i \right] \right\} = \\ & = \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0}} \left\{ \sum_{i=1}^n \sum_{j=1}^n \omega(x_i, y_j) \Delta x_i \Delta y_j \log_2 \omega(x_i / y_j) \right\} + \\ & + \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0}} \left\{ \sum_{i=1}^n \sum_{j=1}^n \omega(x_i, y_j) \Delta x_i \Delta y_j \log_2 \Delta x_i \right\} = \\ & = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(X/Y) dXdY + \\ & + \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0}} \left\{ \log_2 \Delta x_i \sum_{i=1}^n \sum_{j=1}^n \omega(x_i, y_j) \Delta x_i \Delta y_j \right\} = \\ & = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(X/Y) dXdY + \lim_{\Delta x_i \rightarrow 0} \log_2 \Delta x_i, \end{aligned}$$

оскільки $\sum_{i=1}^n \sum_{j=1}^n \omega(x_i, y_j) \Delta x_i \Delta y_j = 1$.

Таким чином, після здійснення граничних переходів отримаємо вираз для кількості інформації при безперервних повідомленнях:

$$\begin{aligned} I(Y, X) = h(X) - h(X/Y) = & - \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX + \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(X/Y) dXdY. \end{aligned} \tag{7.27}$$

Отже, і при безперервних повідомленнях кількість інформації, що передається, визначається різницею початкової і залишкової ентропій повідомлення.

Якщо врахувати, що:

$$\omega(X/Y) = \frac{\omega(X, Y)}{\omega(Y)} \text{ і } \omega(X) = \int_{-\infty}^{\infty} \omega(X, Y) dY,$$

то вираз (7.27) можна представити в іншому вигляді:

$$\begin{aligned} I(Y, X) &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(X) dXdY + \\ &+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \frac{\omega(X, Y)}{\omega(X)\omega(Y)} dXdY = \quad (7.28) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \frac{\omega(X, Y)}{\omega(X)\omega(Y)} dXdY. \end{aligned}$$

На відміну від диференціальної ентропії кількість інформації не залежить від масштабу безперервних повідомлень, якщо масштаби X і Y однакові.

Розглянемо два граничних випадки передачі повідомлень.

Повна статистична залежність повідомлень, що передаються X , і повідомлень, що приймаються Y . Практично це має місце при незначному рівні завад або при повній відсутності завад. У цьому випадку умовна ентропія $H(X/Y) = 0$.

Отже, кількість інформації, що міститься в Y відносно X , дорівнює ентропії повідомлень, що передаються.

Повна статистична незалежність повідомлень X та Y . Це має місце при високому рівні завад, коли завади повністю придушують корисний сигнал. У цьому випадку умовна ентропія, яка дорівнює початковій, та кількість інформації, що міститься в Y відносно X : $I(Y, X) = 0$, тобто повідомлення Y не містить ніякої інформації про повідомлення X .

Із розгляду окремих випадків виходить, що інформація, яка міститься в Y відносно X , не перевершує ентропії X . Максимальна кількість інформації у випадку абсолютно достовірної передачі повідомлень:

$$I(Y, X)_{\max} = H(X). \quad (7.29)$$

Вираз (7.29) справедливий лише для дискретних повідомлень.

При абсолютно достовірній передачі безперервних повідомлень кількість інформації не дорівнює диференціальній ентропії повідомлення. Оскільки безперервні повідомлення відтворюються з обмеженою точністю, то кількість інформації залежить не тільки від статистики по-

повідомлення $h(X)$, але і від способу його відтворення. При цьому кількість інформації, що міститься у відліках X^* відносно повідомлення X , визначиться різницею диференціальних ентропій:

$$J(X^*, X) = h(X) - h(X/X^*), \quad (7.30)$$

де (X/X^*) – умовна диференціальна ентропія, що характеризує втрату інформації за рахунок обмеженої точності відтворення повідомлення. Вираз (7.30) можна привести до вигляду:

$$I(X^*, X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, X^*) \log_2 \frac{\omega(X, X^*)}{\omega(X)\omega(X^*)} dXdY.$$

Можливі випадки, коли задаються вимоги до достовірності відтворення X . Як критерій правильності може бути використане допустиме значення середнього квадратичного відхилення X^* від X , тобто:

$$\sigma^* = \sqrt{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X - X^*)^2 \omega(X, X^*) dXdX^*}.$$

Кількість інформації $I(X^*, X)$ можна змінити, варіюючи умовною щільністю ймовірності $\omega(X/X^*) = \frac{\omega(X, X^*)}{\omega(X^*)}$. Очевидно, що

найбільш вигідним буде те значення функції $\omega(X/X^*)$, при якому $I(X^*, X)$ має найменше значення, оскільки при цьому забезпечується задана вимога до правильності відтворення при отриманні мінімальної кількості інформації. Найменше значення $I(X^*, X)$, при якому задовольняється ця вимога, називають іпсилон-ентропією. Згідно з визначенням, вона дорівнює: $H\epsilon(X) = \min I(X^*, X) = h(X) - \max h(X/X^*)$, при $\sigma^* \leq \epsilon_0$, де ϵ_0 – допустиме значення помилки відтворення. Іпсилон-ентропія визначає інформаційну місткість джерела безперервних повідомлень при заданому критерії правильності відтворення.

7.6 Ентропія і кількість інформації при статистичній залежності елементів повідомлень

При визначенні ентропії і кількості інформації, що передається, передбачалося, що елементи повідомлень статистично незалежні. Однак

в реальних умовах незалежність елементів повідомлень – явище досить рідке. Наприклад, при передачі російського тексту ймовірність появи окремих букв залежать від того, які букви їм передували. Наприклад, якщо передана буква “п”, ймовірність того, що наступною буквою може бути “а”, набагато більше, ніж ймовірність появи букви “р”. Після букви “ь” часто очікується поява сполучення “ся” і т.п. Така залежність між елементами утворилася історично в процесі тривалого формування сучасної російської мови.

Очевидно при визначенні ентропії та інформації в повідомленнях, елементи яких корельовані, не можна обмежуватися, як це ми робили досі, тільки безумовними ймовірностями елементів повідомлень. Необхідно враховувати і умовні ймовірності появи елементів.

Будемо вважати, що передається скінченне число повідомлень $x_1, x_2, x_3, \dots, x_q, x_h, x_b, x_j, \dots, x_{n-1}, x_n$. Корелятивні зв'язки між елементами повідомлень можуть розповсюджуватися на різні групи елементів. Якщо елементи повідомлень незалежні, то умовна ймовірність передачі елемента буде дорівнювати безумовній:

$$p(x_j / x_1, x_h, x_a, \dots, x_2, x_1) = p(x_j).$$

Якщо є корелятивний зв'язок тільки між двома сусідніми елементами, то ймовірність передачі будь-якого елемента повідомлення буде залежати лише від того, який був попередній символ, тобто умовна ймовірність передачі елемента x_j буде дорівнювати $p(x_j/x_i)$. У цьому випадку елементи повідомлень складають простий однозв'язковий ланцюг Маркова.

Якщо корелятивні зв'язки охоплюють три елементи повідомлень, то останні складають двозв'язковий ланцюг Маркова і умовна ймовірність передачі елемента x_j буде дорівнювати $p(x_j/x_h x_i)$.

Більшість повідомлень в реальних умовах являють собою ергодичну послідовність, у якій корелятивні зв'язки розповсюджуються на скінченне число елементів. При достатній довжині такої послідовності з достатньою точністю можуть бути визначені ймовірності та умовні ймовірності появи окремих повідомлень. Мова є типовим прикладом такої послідовності. У будь-якій книзі, виданій певною мовою (крім вузькоспеціальних), частота повторення окремих букв і їх різних сполучень буде постійною, незалежно від змісту книги.

Нехай повідомлення складають простий, тобто однозв'язковий ланцюг Маркова. У цьому випадку ентропія елемента x_j буде визначатися умовною ймовірністю $p(x_j/x_i)$. Для даного фіксованого x_j ентропія повідомлень буде визначатися частковою умовною ентропією:

$$H(X/x_i) = - \sum_{j=0}^n p(x_j/x_i) \log_2 p(x_j/x_i).$$

Зробивши усереднення по всіх x_j , отримаємо вираз для середньої ентропії повідомлення:

$$\begin{aligned} H(X) &= \sum_{i=1}^n p(x_i) H(X/x_i) = - \sum_{i=1}^n p(x_i) \sum_{j=1}^n p(x_j/x_i) \log_2 p(x_j/x_i) = \\ &= - \sum_{i=1}^n \sum_{j=1}^n p(x_i, x_j) \log_2 p(x_j/x_i). \end{aligned} \quad (7.31)$$

При наявності корелятивних зв'язків між трьома елементами ентропія повідомлень буде дорівнювати:

$$\begin{aligned} H(X) &= - \sum_{n=1}^n \sum_{i=1}^n p(x_i, x_n) \sum_{j=1}^n p(x_j/x_i, x_n) \log_2 p(x_j/x_i, x_n) = \\ &= \sum_{h=1}^n \sum_{i=1}^n \sum_{j=1}^n p(x_j, x_i, x_h) \log_2 p(x_j, x_i, x_h). \end{aligned} \quad (7.32)$$

Якщо корелятивними зв'язками охоплена велика кількість елементів, то ентропія визначається аналогічно.

При наявності корелятивних зв'язків між елементами ентропія повідомлень, а отже, і кількість інформації, що передається, зменшується, причому це зменшення буде тим інтенсивніше, чим сильніші корелятивні зв'язки і чим більша кількість елементів буде охоплена цими зв'язками.

7.7 Надлишковість повідомлення

Середня ентропія повідомлень при однаковій кількості елементів може бути різною залежно від статистичних характеристик повідомлень. Ентропія максимальна і визначається виразом (7.5), якщо елементи повідомлень рівноймовірні і взаємно незалежні. Якщо надходження елементів повідомлень не рівноймовірне, то ентропія зменшується і визначається виразом (7.4).

Ще меншою буде ентропія при наявності корелятивних зв'язків між елементами повідомлень. Для випадків, коли корелятивні зв'язки охоплюють два і три елементи, ентропія визначається виразами (7.31) і (7.32). Повідомлення, ентропія яких максимальна, є оптимальними з точки зору найбільшої кількості інформації, що передається.

Мірою кількісної оцінки того, наскільки дане реальне повідомлення за своєю ентропією відрізняється від відповідного йому оптимального повідомлення, є **коефіцієнт стиснення**:

$$\mu = \frac{H(X)}{H(X)_{\max}},$$

де $H(X)$ – ентропія реального повідомлення; $H(X)_{\max}$ – ентропія відповідного йому оптимального повідомлення.

Якщо неоптимальне й оптимальне повідомлення характеризуються однаковою загальною ентропією, то справедлива рівність:

$$nH(X) = n'H(X)_{\max},$$

де n – кількість елементів неоптимального повідомлення; n' – кількість елементів відповідного оптимального повідомлення.

Оскільки середня на елемент ентропія оптимального повідомлення максимальна, то кількість елементів неоптимального повідомлення завжди буде більшою кількості елементів відповідного йому оптимального повідомлення.

Коефіцієнт стиснення можна виразити через кількість елементів повідомлень:

$$\mu = \frac{H(X)}{H(X)_{\max}} = \frac{n'}{n}.$$

Таким чином, реальні повідомлення при однаковій інформативності мають визначену надлишковість в елементах у порівнянні з оптимальними повідомленнями. Мірою кількісної оцінки надлишковості є **коефіцієнт надлишковості**:

$$K_n = \frac{n - n'}{n} = \frac{H(X)_{\max} - H(X)}{H(X)_{\max}} = 1 - \mu.$$

Надлишковість призводить до збільшення часу передачі повідомлень, зайвого завантаження каналу зв'язку. Проте не завжди потрібно прагнути до того, щоб надлишковість $K_n = 0$. Деяка надлишковість буває корисною для забезпечення необхідної надійності систем, підвищення завадостійкості передачі повідомлень. Наприклад, іноді практикується повторна передача того самого повідомлення з метою підвищення достовірності передачі в умовах впливу завад.

Є певна надлишковість і в українській мові. Середня ймовірність повторення окремих букв в українській мові ілюструється таблицею 7.1.

Таблиця 7.1

| Буква | о | е | а | и | т | н | с | р |
|------------------------|------|-------|-------|-------|-------|-------|-------|-------|
| Ймовірність повторення | 0,11 | 0,087 | 0,075 | 0,075 | 0,065 | 0,065 | 0,065 | 0,048 |

Продовження таблиці 7.1

| Буква | в | л | к | ц | щ | э | ф |
|------------------------|-------|-------|-------|-------|-------|-------|-------|
| Ймовірність повторення | 0,048 | 0,042 | 0,034 | 0,005 | 0,004 | 0,003 | 0,002 |

Російський алфавіт містить 31 букву (за умови, якщо не розрізняти букви “е” і “ё”, а також м’який і твердий знаки). З урахуванням пробілу між буквами українська мова, таким чином, володіє 32 символами.

За умови рівноймовірності і незалежності символів середня ентропія на символ буде максимальна і дорівнює:

$$H(X)_{\max} = \log_2 32 = 5 \frac{\text{дв. од.}}{\text{символ}}$$

Якщо врахувати різну ймовірність символів, то:

$$H_1(X) = 4,39 \frac{\text{дв. од.}}{\text{символ}}$$

З урахуванням кореляції між двома символами ентропія зменшується до значення:

$$H_2(X) = 3,52 \frac{\text{дв. од.}}{\text{символ}};$$

– між трьома символами – до значення:

$$H_3(X) = 3,05 \frac{\text{дв. од.}}{\text{символ}};$$

– між вісьмома символами – до значення:

$$H_8(X) = 2 \frac{\text{дв. од.}}{\text{символ}};$$

– і далі залишається незмінною.

Отже, надлишковість російської мови складає:

$$K_{\text{надл}} = \frac{H(X)_{\max} - H(X)}{H(X)_{\max}} = \frac{5 - 2}{5} = 0,6.$$

Варто зауважити, що у всіх європейських мовах надлишковість приблизно однакова. Надлишковість розмовних мов сформувалася у результаті тривалої суспільної практики і дозволяє відновлювати цілі слова і фрази при їх спотворенні під впливом різноманітних чинників.

7.8 Приклади

Приклад 7.1

Визначити ентропію системи, що описується дискретною випадковою величиною x із наступним рядом розподілу:

$$p(x_1) = p(x_2) = p(x_3) = p(x_4) = 0,01, \quad p(x_5) = 0,96.$$

Розв'язування. Використовуючи вираз (7.4) для визначення ентропії дискретного розподілу, отримаємо:

$$H(X) = -\sum_{i=1}^5 p(x_i) \log_2 p(x_i) = -4 \cdot 0,01 \log_2 0,01 - \\ -0,96 \log_2 0,96 = 0,322 \frac{\text{дв. од.}}{\text{символ}}.$$

Приклад 7.2

Визначити ентропію повідомлення з п'яти букв, якщо загальна кількість букв у алфавіті дорівнює 32 і всі повідомлення рівноймовірні.

Розв'язування. Загальна кількість п'ятибуквених повідомлень $n = 32^5$. Використовуючи вираз (7.5) для визначення ентропії рівноймовірних повідомлень, отримаємо:

$$H(X) = \log_2 n = 5 \log_2 32 = 25 \frac{\text{дв. од.}}{\text{символ}}.$$

Приклад 7.3

Визначити диференціальну ентропію неперервного повідомлення, розподіленого за нормальним законом:

$$\omega(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - a_x)^2}{2\sigma^2} \right\}.$$

Розв'язування. Використовуючи вираз (7.9) для визначення шуканої диференціальної ентропії, отримуємо:

$$\begin{aligned}
 h(X) &= - \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-a_x)^2}{2\sigma^2}\right] \times \\
 &\times \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-a_x)^2}{2\sigma^2}\right] dX = \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-a_x)^2}{2\sigma^2}\right] dX \\
 &\times \left\{ \frac{(x-a_x)^2}{2\sigma^2} \right\} dX = \frac{\log_2 e}{2\sigma^2} \int_{-\infty}^{\infty} (x-a_x)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-a_x)^2}{2\sigma^2}\right] dX.
 \end{aligned}$$

Інтеграл першого доданка отриманого виразу дорівнює одиниці, тому що під інтегралом є функція $\omega(X)$, а $\int_{-\infty}^{\infty} \omega(X) dX = 1$. Другий доданок отриманого виразу:

$$\frac{\log_2 e}{2\sigma^2} \int_{-\infty}^{\infty} (x-a_x)^2 \omega(X) dX = \frac{\log_2 e}{2\sigma^2} \sigma^2 = \frac{1}{2} \log_2 e.$$

Таким чином,

$$h(X) = \log_2 \sqrt{2\pi\sigma^2} + \frac{1}{2} \log_2 e = \log_2 \sqrt{2\pi e\sigma^2} \text{ дв.од.}$$

Приклад 7.4

Вимірювана величина зміщується у межах від x_0 до $x_0 + a$ і розподілена за законом рівної ймовірності. Знайти диференціальну ентропію даної величини.

Розв'язування. Закон рівної ймовірності можна аналітично представити у вигляді:

$$\omega(X) = \begin{cases} \frac{1}{a}, & \text{при } x_0 \leq X \leq x_0 + a; \\ 0, & \text{при } X < x_0 \text{ і } X > x_0 + a. \end{cases}$$

Тоді, використовуючи (7.9), отримаємо шукану ентропію:

$$h(X) = - \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX = - \frac{1}{a} \int_{x_0}^{x_0+a} \log_2 \frac{1}{a} dX = \log_2 a \text{ дв. од.}$$

Приклад 7.5

Авсамблі подій $X(x_1, x_2, x_3)$ і $Y(y_1, y_2)$ об'єднані. Ймовірності спільних подій $p(X, Y)$ наведені в таблиці 7.2.

Таблиця 7.2

| y_j | x_i | | |
|-------|-------|-------|-------|
| | x_1 | x_2 | x_3 |
| y_1 | 0,1 | 0,2 | 0,3 |
| y_2 | 0,25 | 0 | 0,15 |

Визначити:

- ентропію ансамблів X та Y ;
- ентропію об'єднаного ансамблю X, Y ;
- умовні ентропії ансамблів.

Розв'язування. З очевидних рівностей

$$\sum_{j=1}^n p(x_i, y_j) = p(x_i) \quad \text{та} \quad \sum_{i=1}^n p(x_i, y_j) = p(y_j)$$

одержуємо:

$$p(x_1) = p(x_1, y_1) + p(x_1, y_2) = 0,1 + 0,25 = 0,35;$$

$$p(x_2) = p(x_2, y_1) + p(x_2, y_2) = 0,2;$$

$$p(x_3) = p(x_3, y_1) + p(x_3, y_2) = 0,3 + 0,15 = 0,45;$$

$$p(y_1) = p(x_1, y_1) + p(x_2, y_1) + p(x_3, y_1) = 0,1 + 0,2 + 0,3 = 0,6;$$

$$p(y_2) = p(x_1, y_2) + p(x_2, y_2) + p(x_3, y_2) = 0,25 + 0,15 = 0,4.$$

Тоді ентропії подій:

$$H(X) = -\sum_{i=1}^3 p(x_i) \log_2 p(x_i) = -0,35 \log_2 0,35 - 0,2 \log_2 0,2 -$$

$$-0,45 \log_2 0,45 = 1,512 \frac{\text{дв. од.}}{\text{повід.}}$$

$$H(Y) = -\sum_{j=1}^2 p(y_j) \log_2 p(y_j) = -0,6 \log_2 0,6 - 0,4 \log_2 0,4 = 0,971 \frac{\text{дв. од.}}{\text{повід.}}$$

Ентропію об'єднаного ансамблю знаходимо, використовуючи вираз (7.10):

$$H(X, Y) = -\sum_{i=1}^3 \sum_{j=1}^2 p(x_i, y_j) \log_2 p(x_i, y_j) = -0,1 \log_2 0,1 -$$

$$-0,25 \log_2 0,25 - 0,2 \log_2 0,2 - 0,3 \log_2 0,3 - 0,15 \log_2 0,15 = 2,228 \frac{\text{дв. од.}}{\text{повід.}}$$

Умовні ентропії зручно визначити, використовуючи властивість (7.12):

$$H(Y/X) = H(X, Y) - H(X) = 2,228 - 1,512 = 0,716 \frac{\text{дв. од.}}{\text{повід.}};$$

$$H(X/Y) = H(X, Y) - H(Y) = 2,228 - 0,971 = 1,257 \frac{\text{дв. од.}}{\text{повід.}}.$$

Приклад 7.6. Сигнал формується у вигляді двійкового коду з ймовірностями появи символів 1 і 0, рівними відповідно $p(x_1) = 0,6$ і $p(x_0) = 0,4$. Появи будь-якого із символів взаємопов'язані умовними ймовірностями:

– $p(x_0/x_0) = 0,1$ – ймовірність того, що після 0 буде 0;

– $p(x_1/x_0) = 0,9$ – ймовірність того, що після 0 буде 1;

– $p(x_1/x_1) = 0,1$ – ймовірність того, що після 1 буде 1;

– $p(x_0/x_1) = 0,9$ – ймовірність того, що після 1 буде 0.

Знайти ентропію сигналів.

Розв'язування. Використовуючи вираз (7.31), знаходимо шукану ентропію:

$$\begin{aligned} H(X) &= -\sum_{i=0}^1 p(x_i) \sum_{j=0}^1 p(x_j/x_i) \log_2 p(x_j/x_i) = \\ &= -(0,1 \log_2 0,1 + 0,9 \log_2 0,9)(0,6 + 0,4) = 0,467 \frac{\text{дв. од.}}{\text{елемент}}. \end{aligned}$$

Приклад 7.7

Ймовірності появи повідомлень x_1, x_2, x_3 і x_4 дорівнюють:

$$p(x_1) = \frac{1}{2}; \quad p(x_2) = \frac{1}{4}; \quad p(x_3) = p(x_4) = \frac{1}{8}.$$

Між повідомленнями є кореляційні зв'язки, описані у таблиці 7.3. Знайти ентропію повідомлень.

Таблиця 7.3

| (x_i, x_j) | $p(x_i, x_j)$ | $p(x_i/x_j)$ | (x_i, x_j) | $p(x_i, x_j)$ | $p(x_i/x_j)$ |
|--------------|---------------|--------------|--------------|---------------|--------------|
| $x_1 x_1$ | 13/32 | 13/16 | $x_3 x_1$ | 0 | 0 |
| $x_1 x_2$ | 3/32 | 3/16 | $x_3 x_2$ | 0 | 0 |
| $x_1 x_3$ | 0 | 0 | $x_3 x_3$ | 0 | 0 |
| $x_1 x_4$ | 0 | 0 | $x_3 x_4$ | 1/8 | 1 |
| $x_2 x_1$ | 1/32 | 1/8 | $x_4 x_1$ | 1/16 | 1/2 |
| $x_2 x_2$ | 1/8 | 1/2 | $x_4 x_2$ | 1/32 | 1/4 |
| $x_2 x_3$ | 3/32 | 3/8 | $x_4 x_3$ | 1/32 | 1/4 |
| $x_2 x_4$ | 0 | 0 | $x_4 x_4$ | 0 | 0 |

Розв'язування. Використовуючи вираз (7.31), отримаємо вужану ентропію:

$$\begin{aligned}
 H(X) &= -\sum_{i=1}^4 \sum_{j=1}^4 p(x_i, x_j) \log_2 p(x_i/x_j) = 13/32 \log_2 13/16 - \\
 &- 3/32 \log_2 3/16 - 1/32 \log_2 1/8 - 1/8 \log_2 1/2 - 3/32 \log_2 3/8 - \\
 &- 1/16 \log_2 1/2 - 2/32 \log_2 1/4 = 0,886 \frac{\text{дв. од.}}{\text{повід.}}
 \end{aligned}$$

Приклад 7.8

Знайти сумісну ентропію взаємопов'язаних повідомлень X і Y та умовну ентропію $h(X/Y)$, якщо двовимірний функція щільності розподілу підпорядкована нормальному закону:

$$\omega(X, Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp \left[-\frac{1}{(2-r^2)} \left(\frac{x^2}{\sigma_x^2} - \frac{2rxy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} \right) \right]$$

де r – коефіцієнт кореляції значень X та Y .

Розв'язування. Сумісна диференціальна ентропія X та Y визначається співвідношенням:

$$h(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) \log_2 \omega(X, Y) dX dY,$$

де

$$\log_2 \omega(X, Y) = -\log_2 (2\pi\sigma_x\sigma_y\sqrt{1-r^2}) - \frac{1}{2(1-r^2)} \left(\frac{x^2}{\sigma_x^2} - \frac{2rxy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} \right) \log_2 e.$$

Підставляючи значення $\log_2 \omega(X, Y)$ у вираз $h(X, Y)$ для, отримуємо:

$$h(X, Y) = \log_2(2\pi\sigma_x\sigma_y\sqrt{1-r^2}) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) dXdY + \\ + \frac{\log_2 e}{2\sigma_x^2(1-r^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X^2 \omega(X, Y) dXdY + \frac{\log_2 e}{2\sigma_y^2(1-r^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Y^2 \omega(X, Y) dXdY - \\ - \frac{r \log_2 e}{\sigma_x\sigma_y(1-r^2)} \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} XY \omega(X, Y) dXdY.$$

Оскільки:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(X, Y) dXdY = 1; \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} XY \omega(X, Y) dXdY = r\sigma_x\sigma_y;$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X^2 \omega(X, Y) dXdY = \sigma_x^2; \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Y^2 \omega(X, Y) dXdY = \sigma_y^2,$$

то в результаті отримуємо:

$$h(X, Y) = -\log_2\left(2\pi\sigma_x\sigma_y\sqrt{1-r^2}\right) + \frac{1}{1-r^2} \log_2 e - \frac{r^2}{1-r^2} \log_2 e = \\ = \log_2\left(2\pi e\sigma_x\sigma_y\sqrt{1-r^2}\right) \frac{\text{дв. од.}}{\text{відлік}}.$$

Застосовуючи вираз (7.12) і враховуючи $h(Y) = \log_2(\sqrt{2\pi e}\sigma_y)$ отримуємо значення умовної ентропії:

$$h(X/Y) = h(X, Y) - h(Y) = \log_2\left[\sqrt{2\pi e(1-r^2)}\sigma_x\right] \frac{\text{дв. од.}}{\text{відлік}}.$$

Приклад 7.9

Контрольований параметр X може приймати два значення x_0 та x_1 з рівними ймовірностями $p(x_0) = p(x_1)$. Внаслідок обмеженої точності системи контролю будуть мати місце помилки контролю, тобто замість x_1 може бути зафіксоване x_0 і, навпаки, замість x_0 зафіксоване x_1 . Умовні ймовірності таких подій дорівнюють 0,01.

Визначити кількість інформації, отриманої при контролі.

Розв'язування. Введемо позначення результатів контролю:

– y_0 – система контролю вказує, що параметр X має значення x_0 ;

– y_1 – система контролю вказує, що параметр X має значення x_1 .

Тоді умовні ймовірності помилкового контролю будуть відповідно дорівнювати $p(y_1/x_0) = 0,01$ і $p(y_0/x_1) = 0,01$.

На підставі виразу (7.23) кількість одержуваної інформації при неповній достовірності контролю буде дорівнювати різниці початкової і залишкової ентропії:

$$I(Y, X) = H(X) - H(X/Y) = -\sum_{i=1}^1 p(x_i) \log_2 p(x_i) + \\ + \sum_{j=0}^1 p(y_j) \sum_{i=0}^1 p(x_i/y_j) \log_2 p(x_i/y_j).$$

Початкова ентропія:

$$H(X) = -p(x_0) \log_2 p(x_0) - p(x_1) \log_2 p(x_1) = 2 \cdot 0,5 \log_2 0,5 = 1 \frac{\text{дв.од.}}{\text{повід.}}$$

Для визначення умовної ентропії $H(X/Y)$ необхідно знати ймовірності виду $p(y_j)$ і $p(x_i/y_j)$. Ймовірності подій $p(y_j)$ обчислюємо за формулою повної ймовірності: $p(y_j) = \sum_i p(x_i) p(y_{ji}/x_i)$. Тоді:

$$p(y_0) = p(x_0) p(y_0/x_0) + p(x_1) p(y_0/x_1) = p(x_0) [1 - p(y_1/x_0)] + \\ + p(x_1) p(y_0/x_1) = 0,5(1 - 0,01) + 0,5 \cdot 0,01 = 0,5.$$

$$p(y_1) = p(x_0) p(y_1/x_0) + p(x_1) p(y_1/x_1) = p(x_0) p(y_1/x_0) + \\ + p(x_1) [1 - p(y_0/x_1)] = 0,5.$$

Умовні ймовірності $p(x_i/y_j)$ обчислюємо за теоремою Байсса:

$$p(x_i/y_j) = \frac{p(x_i) p(y_j/x_i)}{p(y_j)}.$$

$$\text{Тоді } p(x_0/y_1) = \frac{p(x_0) p(y_1/x_0)}{p(y_1)} = \frac{0,5 \cdot 0,01}{0,5} = 0,01.$$

$$p(x_1/y_0) = \frac{p(x_1) p(y_0/x_1)}{p(y_0)} = \frac{0,5 \cdot 0,01}{0,5} = 0,01.$$

Таким чином, умовна ентропія:

$$\begin{aligned} H(X|Y) &= -p(y_0) [p(x_0/y_0) \log_2 p(x_0/y_0) + p(x_1/y_0) \log_2 p(x_1/y_0)] - \\ &\quad - p(y_1) [p(x_0/y_1) \log_2 p(x_0/y_1) + p(x_1/y_1) \log_2 p(x_1/y_1)] = \\ &= -p(y_0) \{ [1 - p(x_1/y_0)] \log_2 [1 - p(x_1/y_0)] + \\ &\quad + p(x_1/y_0) \log_2 p(x_1/y_0) \} - p(y_1) \{ p(x_0/y_1) \log_2 p(x_0/y_1) + \\ &\quad + [1 - p(x_0/y_1)] \log_2 [1 - p(x_0/y_1)] \} = 0,081 \frac{\text{дв. од.}}{\text{повід.}} \end{aligned}$$

У результаті отримаємо

$$I(Y, X) = 1 - 0,081 = 0,92 \frac{\text{дв. од.}}{\text{повід.}}$$

Приклад 7.10

Визначити кількість інформації, яка міститься в одному замірі випадкової величини x , рівномірно розподіленої в межах від 0 до 256, якщо похибка вимірювання розподілена за нормальним законом і середнє квадратичне значення похибки $\sigma = 5$.

Розв'язування. Диференціальна ентропія випадкової величини X :

$$h(x) = - \int_0^{256} \frac{1}{256} \log_2 \frac{1}{256} dx = 8 \text{ дв. од.}$$

Остаточна диференціальна ентропія визначається похибкою вимірювання і на основі прикладу 7.3:

$$h(\delta) = \log_2 \sqrt{2\pi e} \sigma = \log_2 \sqrt{2\pi e} 5 = 4,39 \text{ дв. од.}$$

Кількість інформації, отримана в результаті одного заміру, визначається різницею початкової і кінцевої ентропій:

$$I(x) = h(x) - h(\delta) = 8 - 4,39 = 3,61 \text{ дв. од.}$$

Приклад 7.11

Повна шкала вимірювального прибору містить 1000 поділок. Похибка прибору складає ± 1 % повної шкали і розподілена за законом рівної ймовірності. Визначити надлишковість шкали прибору.

Розв'язування. В оптимальному випадку при приведеній похибці, рівній $\pm 1\%$, достатньо мати 50 поділок шкали. Тоді на основі виразу (7.33) надлишковість становить:

$$K_{\text{н}} = \frac{1000 - 50}{1000} = 0,95.$$

Контрольні питання

1. Поясніть сутність поняття ентропії.
2. В яких одиницях вимірюється ентропія?
3. Сформулюйте основні властивості ентропії.
4. Як визначається ентропія дискретної системи з рівноймовірними та нерівноймовірними станами?
5. Чому дорівнює ентропія при нерівноймовірному і взаємозалежному розподілі елементів системи?
6. У чому полягає особливість визначення ентропії для неперервних розподілів?
7. Що таке умовна і сумісна ентропія?
8. Основні властивості ентропії складних повідомлень.
9. Поясніть зв'язок між ентропією та інформацією.
10. Які основні вимоги висуваються до міри кількості інформації?
11. Як кількісно оцінюється інформація при повній і неповній достовірності повідомлень?
12. Що розуміється під інсешн-ентропією?
13. Поясніть сутність поняття часткової та повної інформації.
14. Що розуміється під надлишковістю повідомлень?
15. Що є мірою кількісної оцінки надлишковості?

8.1 Узагальнені характеристики інформаційних сигналів

Будь-яку інформаційну систему можна поділити на джерело, канал передачі та одержувач інформації, а інформаційні процеси, які відбуваються в пристроях, представити у вигляді процесу передачі інформації по деякому каналу. Узагальнена схема каналу передачі інформації наведена на рис. 8.1.

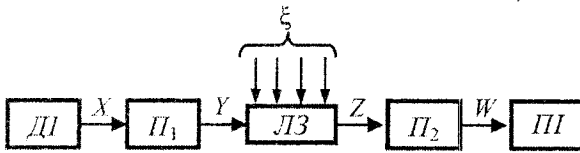


Рис. 8.1

Повідомлення X джерела інформації $ДІ$ після кодування і модуляції у перетворювачі $П_1$ перетворюються в сигнал Y , який надходить на лінію зв'язку $ЛЗ$. У результаті дії завад ξ сигнал Z на приймальній стороні відрізняється від Y . Приймальна частина містить перетворювач $П_2$, який демодулює і декодує сигнали та переробляє їх в повідомлення W , які надходять на приймач інформації $ПІ$.

Безпосереднім носієм інформації є сигнал, який являє собою фізичну величину, що випадково змінюється. На практиці при розгляді інформаційних процесів зручно оперувати узагальненими показниками, характерними для багатьох сигналів даного виду і найбільш вагомими з точки зору передачі інформації, що міститься в них. Кожен сигнал має визначену тривалість. Тривалість сигналу характеризує час передачі повідомлень, тривалість зайнятості інформаційного каналу. Таким чином, за першу узагальнену характеристику взято **час передачі сигналу** T_c .

Кожен сигнал характеризується визначеним частотним спектром. Теоретично ширина спектра скінченної тривалості необмежена. Однак вивчення спектрів реальних сигналів показує, що їх спектральна щільність зменшується із зростанням частоти. Це дозволяє при відпо-

відних умовах розглядати сигнали як процеси з обмеженим спектром. Існують різні критерії обмеження спектра сигналу. Одним з таких критеріїв є допустиме спотворення сигналу. Практика показала, що при передачі голосового сигналу розбірливість і якість мови практично повністю зберігаються при ширині спектра від 300 до 3400 Гц. Таким чином, другою узагальненою характеристикою сигналу повинна бути **ширина його частотного спектра** F_c . Третьою важливою характеристикою сигналу є його енергетична характеристика — середня потужність P_c . Однак, оскільки при передачі на сигнали завжди впливають завади, то за енергетичну характеристику сигналу доцільно брати відношення середньої потужності сигналу P_c до середньої потужності завади P_g . Зазвичай це відношення виражають в логарифмічній формі $D = \log_2(P_c / P_g)$ і називають **динамічним діапазоном**, причому, при оцінюванні інформаційної місткості зручно виражати динамічний діапазон через логарифм з основою 2.

Добуток

$$V_c = T_c F_c D_c \quad (8.1)$$

прийнято називати об'ємом сигналу. У геометричному представленні об'єм сигналу має вигляд паралелепіпеда з ребрами T_c , F_c та D_c (рис. 8.2).

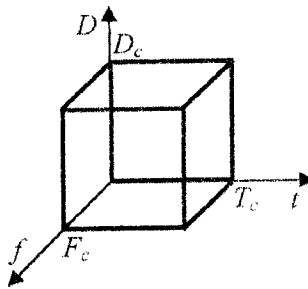


Рис. 8.2

Інформаційний канал можна характеризувати також трьома відповідними параметрами: часом використання каналу, шириною смуги частот, які пропускаються каналом T_k , і динамічним діапазоном каналу D_k , який характеризує його здатність передавати різні рівні сигналу.

Величина

$$V_k = T_k F_k D_k \quad (8.2)$$

називається **ємністю каналу**.

Неспотворена передача сигналів можлива тільки за умови, що сигнал за своїм об'ємом "вміщується" в сміність каналу (рис. 8.3).

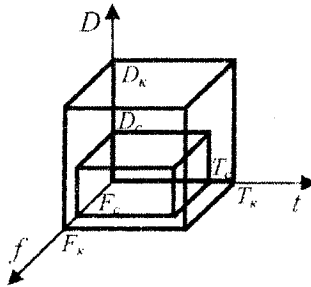


Рис. 8.3

Отже, загальна умова узгодження сигналу з каналом передачі інформації визначається співвідношенням:

$$V_c \leq V_k. \quad (8.3)$$

Однак співвідношення (8.3) виражає необхідну, але недостатню умову узгодження сигналу з каналом. Достатньою умовою є узгодження за всіма параметрами:

$$T_c \leq T_k; \quad F_c \leq F_k; \quad D_c \leq D_k. \quad (8.4)$$

Для інформаційного каналу користуються поняттями: швидкість введення інформації, швидкість передачі інформації та пропускна здатність каналу.

Під **швидкістю введення інформації** (поток інформації) $\bar{I}(X)$ розуміють середню кількість інформації, що вводиться від джерела повідомлень в інформаційний канал в одиницю часу. Це характеристика джерела повідомлень і визначається тільки статистичними властивостями повідомлень.

Швидкість передачі інформації $\bar{I}(Z, Y)$ – середня кількість інформації, що передається по каналу в одиницю часу. Вона залежить від статичних властивостей сигналу, що передається, і від властивостей каналу.

Пропускна здатність C – найбільша теоретично досяжна для даного каналу швидкість передачі інформації. Це характеристика каналу і не залежить від статистики сигналу.

З метою найбільш ефективного використання інформаційного каналу необхідно вживати заходи, щоб швидкість передачі інформації

була як можна ближче до пропускної здатності каналу. Разом з тим швидкість введення інформації не повинна перевищувати пропускну здатність каналу, інакше не вся інформація буде передана каналу.

Це основна умова динамічного узгодження джерела повідомлень та інформаційного каналу.

Одним з основних питань в теорії передачі інформації є визначення залежності швидкості передачі інформації та пропускної здатності від параметрів каналу і характеристик сигналів та завад. Ці питання були вперше глибоко досліджені К. Шенноном.

Розглянемо три види каналів: дискретний канал без завад, дискретний канал з завадами і неперервний канал з завадами.

8.2 Швидкість передачі інформації та пропускна здатність дискретного каналу без завад

Під дискретним каналом передачі інформації прийнято розуміти сукупність засобів, призначених для передачі дискретних сигналів.

На вхід такого каналу подаються дискретні повідомлення X , які утворюють первинний алфавіт x_1, x_2, \dots, x_n . Останні кодуються за допомогою перетворювача Π_1 (див. рис. 8.1) і перетворюються в кодовані сигнали Y . Для кодування використовується деякий алфавіт символів y_1, y_2, \dots, y_m , а суть кодування зводиться до представлення окремих повідомлень або послідовностей повідомлень певними комбінаціями символів алфавіту, що використовується.

Для дискретних джерел, що видають стаціонарні послідовності повідомлень досить великої тривалості T , введено поняття "типових" і "нетипових" послідовностей повідомлень. При $T \rightarrow \infty$ всі типові послідовності мають приблизно однакову ймовірність появи, в той час як сумарна ймовірність появи всіх нетипових послідовностей прямує до нуля. Зазначені властивості послідовностей дискретного джерела узагальнені теоремою асимптотичної рівномірності, яка виходить із закону великих чисел.

При $T \rightarrow \infty$ для типових послідовностей з дуже малою похибкою справедливе співвідношення:

$$N_{\text{тип}} = 2^{nH(X)}, \quad (8.5)$$

де $N_{\text{тип}}$ – кількість типових послідовностей, n – довжина послідовностей (кількість символів у послідовності); $H(X)$ – ентропія на один символ.

Для дискретного каналу швидкість передачі інформації визначається відношенням кількості інформації, яку переносять окремі сигнали, до тривалості сигналів T_c .

При деяких умовах (наприклад, ергодичному характері завад) при $T_c \rightarrow \infty$ буде типовою послідовність повідомлень і на виході каналу. Тому при $T_c \rightarrow \infty$ швидкість передачі інформації буде постійною величиною незалежно від тривалості сигналу і є осередненою характеристикою роботи інформаційного каналу.

У зв'язку з цим теоретично швидкість передачі інформації виражають таким чином:

$$\bar{I}(Z, Y) = \lim_{T_c \rightarrow \infty} \frac{I(Z, Y)}{T_c}. \quad (8.6)$$

Аналогічно швидкість введення інформації:

$$\bar{I}(X) = \lim_{T_x \rightarrow \infty} \frac{H(X)}{T_x}, \quad (8.7)$$

де $H(X)$ – ентропія послідовності повідомлень на виході джерела повідомлень; T_x – тривалість послідовності.

Для реального дискретного каналу середня швидкість передачі інформації визначається співвідношенням:

$$\bar{I}(Z, Y) = \frac{I(Z, Y)}{\bar{\tau}_c}, \quad (8.8)$$

де $I(Z, Y)$ – середня кількість інформації, яка переноситься одним символом сигналу; $\bar{\tau}_c$ – середня тривалість символу.

При відсутності статистичної залежності між сигналами:

$$\bar{\tau}_c = \sum_{i=1}^N P(y_i) \tau_{y_i}, \quad (8.9)$$

де $P(y_i)$ і τ_{y_i} – априорна ймовірність і тривалість i -го сигналу відповідно.

Швидкість введення інформації:

$$\bar{I}(X) = \frac{H(X)}{\bar{\tau}_x} = \bar{V}_x H(X), \quad (8.10)$$

де $H(X)$ – середня ентропія одного повідомлення; $\bar{\tau}_x$ – середня тривалість повідомлення; $\bar{V}_x = 1/\bar{\tau}_x$ – середня швидкість надходження повідомлень від джерела.

У каналі без завад кожному певному вхідному сигналу завжди буде відповідати один і той же сигнал на виході каналу, іншими словами, вхідні і вихідні сигнали пов'язані однозначною функціональною залежністю (рис. 8.4).

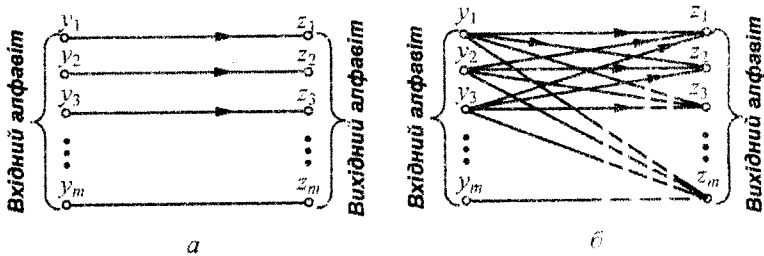


Рис. 8.4

У цьому випадку середня кількість інформації, яка переноситься одним символом, рівна ентропії символу на вході каналу:

$$I(Y) = H(Y) \frac{\text{дв. од.}}{\text{символ}}$$

Швидкість передачі інформації:

$$\bar{I}(Y) = \bar{U}_y H(Y) \frac{\text{дв. од.}}{c}, \quad (8.11)$$

де $\bar{U}_y = 1/\bar{\tau}_y$ – швидкість передачі елементарних символів сигналу; $\bar{\tau}_y$ – середня тривалість елементарних сигналів.

Пропускна здатність дискретного каналу без завад:

$$C = \max \{ \bar{U}_y H(Y) \}.$$

Вважаючи \bar{U}_y заданою, отримаємо, що максимальна швидкість передачі інформації буде забезпечена при максимальному значенні ентропії кодового сигналу:

$$C = \bar{U}_y \max \{ H(Y) \} = \bar{U}_y \log_2 n, \quad (8.12)$$

тобто при рівномірному розподілі ймовірностей і статистичної незалежності символів алфавіту сигналів.

Таким чином, швидкість передачі інформації може бути максимальною при умові, що статистичні характеристики джерела повідомлень певним чином узгоджені з властивостями інформаційного каналу. Для кожного джерела повідомлень це узгодження може бути досягнуто спеціальним вибором способу кодування сигналів.

На питання про те, в якій мірі швидкість передачі інформації може бути наближена до пропускної здатності інформаційного каналу, відповідає **теорема Шеннона** для дискретного каналу без завад. Теорема формулюється таким чином: якщо потік інформації, що виробляється джерелом, досить близький до пропускної здатності каналу, тобто якщо справедлива рівність:

$$\bar{I}(X) = C - \delta, \quad (8.13)$$

де δ – досить мала величина, то завжди можна знайти такий спосіб кодування, який забезпечить передачу всіх повідомлень, що виробляються джерелом, причому швидкість передачі інформації буде дуже близька до пропускної здатності каналу: $\bar{I}(Z, Y) = C - \delta$.

Зворотнє твердження теореми полягає в тому, що неможливо забезпечити тривалу передачу всіх повідомлень, якщо потік інформації, що виробляється джерелом, перевищує пропускну здатність каналу:

$$\bar{I}(X) > C.$$

Таким чином, теорема Шеннона стверджує, що при виконанні умови (8.13) швидкість передачі інформації може бути в принципі досить наближена до пропускної здатності каналу. Це може бути забезпечене відповідним кодуванням сигналів.

8.3 Швидкість передачі інформації та пропускна здатність дискретного каналу із завадами

При наявності завад в каналі передачі інформації порушується однозначна відповідність між вхідним і вихідним алфавітами каналу. Одному вхідному сигналу можуть відповідати різні вихідні сигнали (див. рис. 8.4).

Швидкість передачі інформації по дискретному каналу з завадами:

$$\bar{I}(Z, Y) = \bar{U}_Y [H(Y) - H(Y/Z)], \quad (8.14)$$

де $H(Y/Z)$ – залишкова ентропія сигналу, зумовлена дією завад.

Вираз для швидкості передачі інформації може бути таким:

$$\bar{I}(Z, Y) = \bar{U}_p [H(Z) - H(Z/Y)], \quad (8.15)$$

де $H(Z)$ – ентропія вихідного сигналу; $H(Z/Y)$ – умовна ентропія вихідного сигналу при відомій ентропії вхідного сигналу.

Пропускна здатність каналу:

$$C = \max \bar{I}(Z, Y) = \bar{U}_p \max [H(Z) - H(Z/Y)]. \quad (8.16)$$

Канали, у яких на кожний символ сигналу, що передається, завади впливають незалежно від того, які сигнали передавалися раніше, називаються каналами без пам'яті. У таких каналах завади не викликають додаткових корелятивних зв'язків між символами. На сьогодні основні виведення теорії інформації отримані стосовно каналів без пам'яті.

Для дискретного каналу з завадами К. Шенноном доведена наступна **теорема**: якщо потік інформації, що виробляється джерелом, досить близький до пропускної здатності каналу, тобто якщо справедлива рівність $\bar{I}(X) = C - \delta$, де δ – досить мала величина, то завжди можна знайти такий спосіб кодування, який забезпечить передачу всіх повідомлень, що виробляються джерелом, а ймовірність помилкового розпізнання будь-якого переданого повідомлення буде досить малою, тобто $P_{nr} < \eta$, де P_{nr} – ймовірність неправильного розпізнання переданого повідомлення; η – досить мала величина.

Зворотне твердження теореми полягає в тому, що якщо потік інформації джерела перевищує пропускну здатність каналу, то не існує способу кодування, що забезпечує передачу будь-якого повідомлення з малою ймовірністю помилки.

Таким чином, розглянута теорема визначає співвідношення між швидкістю створення повідомлень джерелом, пропускну здатністю каналу при наявності завад і достовірністю передачі.

Справедливість теорем Шеннона для дискретного каналу може бути показана, виходячи з теореми про асимптотичні рівноймовірності довгих послідовностей повідомлень.

Теореми Шеннона не вказують практичних шляхів знаходження оптимального коду, щоб наблизити швидкість передачі інформації до пропускної здатності каналу. Встановлено лише, що для наближення швидкості передачі до межі загальним методом для каналів з завадами і без завад є кодування довгих повідомлень. Проте значення цих теорем важко переоцінити. До К. Шеннона вважалося, що в каналі з завадами забезпечити досить малу ймовірність помилки можна тільки при наближенні швидкості передачі інформації до нуля. К. Шеннон показав, що шляхом відповідного вибору способу кодування можна

забезпечити досить малу ймовірність помилки при скінченній швидкості передачі інформації.

8.4 Швидкість передачі інформації та пропускна здатність безперервного каналу із завадами

Під безперервним каналом передачі інформації прийнято розуміти сукупність коштів, призначених для передачі безперервних сигналів. На відміну від дискретних каналів у неперервних каналах замість кодувальних і декодувальних пристроїв може використовуватися більш широкий клас різних перетворювачів. Для передачі інформації по каналу може застосовуватися модуляція одного або декількох параметрів сигналу. Незалежно від конкретного характеру перетворення сигналів вхідні і вихідні сигнали безперервного каналу задаються у вигляді ансамблів безперервних функцій з відповідною щільністю розподілу ймовірностей.

Нехай на вхід каналу надходить безперервний сигнал $Y(t)$ тривалістю T . Внаслідок впливу завад $\xi(t)$ вихідний сигнал $Z(t)$ буде відрізнятися від вхідного.

Кількість інформації про випадковий сигнал $Y(t)$, що міститься у випадковому сигналі $Z(t)$, визначається залежністю:

$$I_T(Z, Y) = h_T(Z) - h_T(Z / Y).$$

Відповідно до теореми Котельникова, безперервні сигнали $Y(t)$ і $Z(t)$ можуть бути представлені сукупностями відліків y_i і z_i у дискретні моменти часу (рис. 8.5), що являють собою випадкові величини.

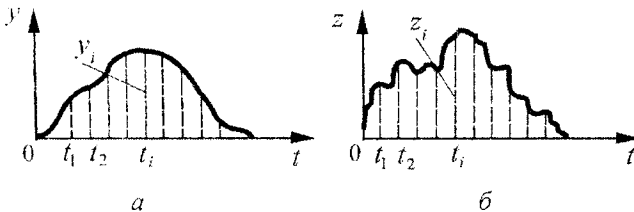


Рис. 8.5

Розподіл сукупності випадкових величин описується багатовимірною щільністю розподілу ймовірності:

$$\omega(y_1, y_2, \dots, y_m), \quad \omega(z_1, z_2, \dots, z_m).$$

Тоді диференціальна ентропія сигналу на виході каналу:

$$h_T(Z) = - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \omega(z_1, z_2, \dots, z_m) \log_2 \omega(z_1, z_2, \dots, z_m) dz_1 \dots dz_m.$$

Оскільки відповідно до критерію Н.А. Желєзнова при квантуванні випадкових сигналів за часом інтервал квантування необхідно встановлювати рівним інтервалу кореляції функції τ_0 , то випадкові величини $z_1, z_2, \dots, z_n, \dots, z_m$ можна вважати незалежними і

$$\omega(z_1, z_2, \dots, z_m) = \omega(z_1)\omega(z_2)\dots\omega(z_m).$$

Виходячи з того, що ентропія сукупності випадкових незалежних величин рівна сумі ентропій випадкових величин, отримаємо наступний вираз для диференціальної ентропії сигналу: $h_T(Z) = \sum_{i=1}^m h(z_i)$,

де $h(z_i) = - \int_{-\infty}^{\infty} \omega(z_i) \log_2 \omega(z_i) dz_i$ – диференціальна ентропія i -го відліку сигналу Z ; $m = T/\Delta T$ – загальна кількість відліків сигналу Z тривалістю T ; ΔT – інтервал часового квантування.

Обмежуючись розглядом стаціонарних процесів, отримаємо:

$$\omega(z_1) = \omega(z_2) = \dots = \omega(z_m);$$

$$h(z_1) = h(z_2) = \dots = h(z_m) = h(Z).$$

Тоді $h_T(Z) = mh(Z)$, де $h(Z)$ – диференціальна ентропія одного відліку.

Аналогічно можна показати, що умовна диференціальна ентропія дорівнює $h_T(Z/Y) = mh(Z/Y)$, де $h(Z/Y)$ – умовна ентропія одного відліку.

Таким чином, вираз для кількості інформації прийме вигляд:

$$I_T(Z, Y) = m[h(Z) - h(Z/Y)].$$

Швидкість передачі інформації:

$$\bar{I}_T(Z, Y) = \frac{m}{T} [h(Z) - h(Z/Y)] = F_0 [h(Z) - h(Z/Y)], \quad (8.17)$$

де $F_0 = \frac{m}{T} = \frac{1}{\Delta T}$ – частота часового квантування (відліку).

Пропускна здатність каналу:

$$C = \max [\bar{I}_T(Z, Y)] = F_0 \max [h(Z) - h(Z/Y)]. \quad (8.18)$$

Максимальне значення першої частини виразу (8.18) можна досягнути, варіюючи $\omega(Y)$, оскільки інші параметри відносяться до каналу зв'язку, які ми змінювати не можемо.

Розглянемо деякі окремі випадки.

1. Сигнал обмеженої потужності передається по каналу, в якому діє адаптивна завада обмеженої потужності типу білого гаусового шуму.

При адаптивній заваді сигнал $Z(t)$ на виході каналу буде рівний:

$$Z(t) = Y(t) + \xi(t),$$

де $\xi(t)$ – завада, що діє в каналі передачі інформації.

Середні потужності сигналу і завади відповідно рівні P_Y і $P_\xi = \sigma_\xi^2$. Смуга пропускання каналу обмежена межами 0 і F_c .

Ширина спектра сигналу і завади обмежуються смугою пропускання каналу.

Частота часового квантування обмежена за спектром сигналу відповідно до теореми Котельникова $F_0 = 2F_c$.

Тоді вираз (8.18) для пропускної здатності каналу прийме вигляд:

$$C = 2F_c \max [h(Z) - h(Z/Y)]. \quad (8.19)$$

При взаємно незалежних сигналі Y і заваді ξ ймовірність того, що при передачі сигналу Y відповідний сигнал буде рівний $Z = Y + \xi$, повинна визначатися ймовірністю того, що завада прийме дане значення $\xi = Z - Y$, тобто $\omega(Z/Y)dY = \omega(\xi)d\xi$, а

$$\omega(Z/Y) = \omega(Y/Y + \xi/Y) = \omega(Y/Y)\omega(\xi/Y) = \omega(\xi). \quad (8.20)$$

Враховуючи формулу (8.20), вираз для умов диференціальної ентропії $h(Z/Y)$ можна перетворити таким чином:

$$\begin{aligned} h(Z/Y) &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(Y)\omega(Z/Y) \log_2 \omega(Z/Y) dY dZ = \\ &= - \int_{-\infty}^{\infty} \omega(Y) \left[\int_{-\infty}^{\infty} \omega(\xi) \log_2 \omega(\xi) d\xi \right] dY = h(\xi) \int_{-\infty}^{\infty} \omega(Y) dY = h(\xi), \end{aligned} \quad (8.21)$$

де $h(\xi)$ – диференціальна ентропія завади.

Таким чином, у разі адаптивної завади умовна диференціальна ентропія $h(Z/Y)$ повністю визначається властивостями завади. Раніше було встановлено (див. приклад 7.3), що диференціальна ентропія сигналу, розподіленого згідно з нормальним законом становить:

$$h(\xi) = \log_2(\sqrt{2\pi e}\sigma_\xi). \quad (8.22)$$

Підставивши вираз (8.22) в (8.19), отримаємо наступний вираз для пропускну здатності каналу:

$$C = 2F_k \max \left\{ h(Z) - \log_2 \sqrt{2\pi e}\sigma_\xi \right\}. \quad (8.23)$$

Оскільки δ_ξ задано, то максимальне значення виразу (8.23) буде забезпечено при умові максимізації диференціальної ентропії вихідного сигналу $h(Z)$. Оскільки середні потужності вхідного сигналу $Y(t)$ і завади $\xi(t)$ обмежені, то середня потужність вихідного сигналу $Z(t)$ також обмежена. При такій умові величина диференціальної ентропії $h(Z)$ буде максимальною у випадку, якщо $Z(t)$ характеризується нормальним законом розподілу. Якщо ж сумарний сигнал $Z(t)$ і одна з його складових $\xi(t)$ розподілені згідно з нормальним законом, то і друга складова, тобто вхідний сигнал $Y(t)$, також повинна мати нормальний закон розподілу.

Таким чином, диференціальна ентропія вихідного сигналу:

$$h(Z) = \log_2 \sqrt{2\pi e}\sigma_z = \log_2 \sqrt{(\sigma_y^2 + \sigma_\xi^2)2\pi e}. \quad (8.24)$$

Підставляючи вираз (8.24) в (8.23), остаточно отримаємо:

$$C = 2F_k \left[\log_2 \sqrt{(\sigma_y^2 + \sigma_\xi^2)2\pi e} - \log_2 \sqrt{\sigma_\xi^2 2\pi e} \right] = \\ F_k \log_2 \frac{\sigma_y^2 + \sigma_\xi^2}{\sigma_\xi^2} = F_k \log_2 \left(1 + \frac{P_y}{P_\xi} \right), \quad (8.25)$$

де $P_y = \sigma_y^2$ - середня потужність корисного сигналу; $P_\xi = \sigma_\xi^2$ - середня потужність завади.

Отже, середня швидкість передачі інформації сигналами з обмеженою середньою потужністю через канал, в якому діє білий гаусів шум, виявляється максимальною при повній подібності між сигналом і завадою. Отже, максимальна швидкість передачі інформації буде забезпечена, якщо за фізичне перенесення інформації прийняти стаціонарний випадковий процес у вигляді гаусового шуму.

Як видно з виразу (8.25), пропускну здатність каналу можна регулювати шляхом зміни F_k або P_v . Причому практична залежність пропускну здатності каналу від F_k при постійній потужності сигналу нелінійна. Це зумовлене тим, що потужність завади P_z також залежить від ширини частотного спектра. Дійсно, енергетичний спектр білого шуму рівномірний, тому потужність такої завади можна представити у вигляді:

$$P_z = P_0 F_k, \quad (8.26)$$

де P_0 – потужність завади, що припадає на смугу в 1 Гц (спектральна щільність потужності завади).

Підставивши вираз (8.26) в (8.25), отримуємо вираз, що визначає дійсний характер залежності пропускну здатності каналу від ширини його смуги пропускання:

$$C = F_k \log_2 \left(1 + \frac{P_v}{P_0 F_k} \right). \quad (8.27)$$

Характер залежності $C = f(F_k)$ наведений на рис. 8.6.

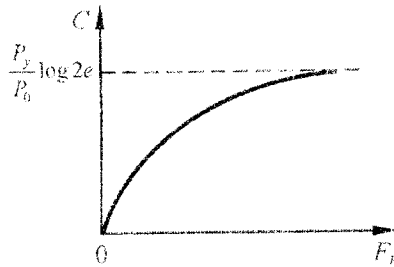


Рис. 8.6

Визначимо межу, до якої наближається пропускну здатність каналу при необмеженому збільшенні його смуги пропускання:

$$\lim_{F_k \rightarrow \infty} C = \lim_{F_k \rightarrow \infty} \frac{\log_2 \left(1 + \frac{P_v}{P_0 F_k} \right)}{\frac{1}{F_k}}$$

Ввівши позначення $\alpha = \frac{1}{F_k}$, отримуємо:

$$\lim_{F_k \rightarrow \infty} C = \lim_{\alpha \rightarrow 0} \frac{\log_2 \left(1 + \frac{P_y}{P_0} \alpha \right)}{\alpha}.$$

Розкриваючи невизначеність, отримуємо граничне значення пропускну́ї здатності каналу:

$$\lim_{F_k \rightarrow \infty} C = \frac{P_y}{P_0} \log_2 e. \quad (8.28)$$

З формули (8.28) видно, що максимальне значення, до якого прагне пропускну́а здатність каналу, із зростанням його ширини смуги пропускання, пропорційна відношенню середньої потужності сигналу до спектральної щільності потужності завади.

З проведеного аналізу можна також визначити, що немає сенсу значно збільшувати смугу пропускання каналу, оскільки по мірі розширення смуги пропускання зростання пропускну́ї здатності каналу сповільнюється і при $F_k \rightarrow \infty$ пропускну́а здатність наближається до постійної величини. Доцільно збільшувати смугу пропускання приблизно до значення, рівного відношенню P_y/P_0 . Гранично можливе значення пропускну́ї здатності може бути збільшено за рахунок збільшення відношення P_y/P_0 .

2. Сигнал обмеженої потужності передається по каналу, в якому діє адаптивна завада у вигляді довільного шуму. Як вже встановлено, при певному середньому квадратичному значенні (при певній потужності) завади найбільшою ентропією володіє завада з нормальним законом розподілу ймовірностей. При іншому законі розподілу ймовірностей завади її ентропія при тому ж середньому квадратичному значенні виявляється меншою. Залежність ентропії від вигляду закону розподілу спонукала К. Шеннона характеризувати дію завади не її дійсною потужністю, а так званою ентропійною потужністю завади. Під ентропійною потужністю К. Шеннон розумів потужність еквівалентного білого шуму, який, маючи ті самі тривалість і ширину спектра, володіє такою ж, як дана завада, ентропією.

Таким чином, якщо довільна завада $x(t)$ характеризується ентропією на один відлік $h(\xi)$, то потужність еквівалентного білого шуму, тобто ентропійна потужність $\sigma_{\xi e}^2$ може бути визначена з умови:

$$h(\xi) = \log_2 \sqrt{2\pi e \sigma_{\xi e}^2}, \text{ звідки:}$$

$$\sigma_{\xi e}^2 = \frac{2^{h(\xi)}}{2\pi e} = \frac{2^{h(\xi)-1}}{\pi e}. \quad (8.29)$$

Величина відношення середньої потужності завади σ_{ξ}^2 та її ентропійної потужності $\sigma_{\xi e}^2$ визначається законом розподілу завади. Визначивши це відношення через k_e , отримаємо:

$$\sigma_{\xi}^2 = k_e \sigma_{\xi e}^2. \quad (8.30)$$

Зокрема, для закону рівної ймовірності $k_e = 1,3$.

Використовуючи поняття ентропійної потужності, можна вираз для пропускної здатності каналу, в якому діє довільна завада, представити в наступному вигляді:

$$C = 2F_k \left[\log_2 \sqrt{(\sigma_v^2 + \sigma_{\xi}^2) 2\pi e} - \log_2 \sqrt{\sigma_{\xi e}^2 2\pi e} \right] = F_k \log_2 \frac{\sigma_v^2 + \sigma_{\xi}^2}{\sigma_{\xi e}^2} = F_k \log_2 \left(k_e \frac{\sigma_v^2 + \sigma_{\xi}^2}{\sigma_v^2} \right). \quad (8.31)$$

Для всіх законів розподілу крім нормального, коефіцієнт $k_e > 1$, тому пропускна здатність каналу, в якому діє довільна завада, буде завжди більша пропускної здатності каналу, в якому діє білий шум з такою ж середньою потужністю, що і довільна завада.

Для неперервного каналу з завадами К. Шенноном сформульована наступна *теорема*.

Якщо ієpsilon-продуктивність $\bar{H}_\epsilon(X)$ джерела неперервних повідомлень, що визначає кількість інформації, яка виробляється в одиницю часу при заданій оцінці g правильності відтворення, як завгодно близька до пропускної здатності каналу, тобто справедливе співвідношення: $\bar{H}_\epsilon(X) = C - \alpha$, де α – як завгодно мале, то існує метод передачі, при якому всі повідомлення, які виробляються джерелом, можуть бути передані, а правильність відтворення при цьому як завгодно близька до g .

8.5 Приклади

Приклад 8.1

В інформаційному каналі без завад для передачі повідомлень використовується алфавіт з чотирма різними символами. Тривалості

усіх символів однакові і рівні $\tau = 1$ мс. Визначити пропускну здатність каналу передачі інформації.

Розв'язування. Для розрахунку пропускну здатності дискретного каналу без завад скористаємося формулою:

$$C = \bar{U}_y \max \{H(Y)\} = \frac{\log_2 N}{T_y},$$

де N – загальна кількість повідомлень; T_y – середня тривалість сигналу.

Оскільки код кожного повідомлення містить чотири символи, то тривалість всіх сигналів буде постійною і рівною $T_y = 4\tau = 4$ мс.

Оскільки для передачі повідомлень використовується алфавіт з чотирма символами, то $N = 4^4$.

$$\text{Тому } C = 250 \log_2 4^4 = 250 \log_2 2^8 = 2 \cdot 10^3 \frac{\text{дв. од.}}{\text{с}}.$$

Приклад 8.2

Джерело виробляє символи з ймовірностями:

$$p_1 = 0,2; \quad p_2 = 0,7; \quad p_3 = 0,1.$$

Кореляційні зв'язки між повідомленнями відсутні. Передача інформації здійснюється двійковим кодом, тривалість символів якого дорівнює $\tau = 1$ мс. Визначити швидкість передачі інформації по каналу без завад при використанні рівномірного коду.

Розв'язування. Швидкість передачі інформації визначаємо за формулою (8.11).

Середня ентропія повідомлень на один символ:

$$H(Y) = -\sum_{i=1}^3 p_i \log_2 p_i = -0,2 \log_2 0,2 - \\ -0,7 \log_2 0,7 - 0,1 \log_2 0,1 = 1,16 \frac{\text{дв. од.}}{\text{символ}}.$$

Для передачі трьох повідомлень двійковим кодом необхідні два розряди. Як наслідок, довжина кодових комбінацій дорівнює 2τ і швидкість передачі сигналів:

$$U_y = \frac{1}{2\tau} = 500 \frac{\text{повід.}}{\text{с}}.$$

Таким чином, швидкість передачі інформації:

$$\bar{T}(Y) = U_y H(Y) = 500 \cdot 1,16 = 580 \frac{\text{дв. од.}}{\text{с}}$$

Приклад 8.3

Джерело, яке виробляє чотири символи з апіорними ймовірностями $P_1 = 0,4$; $P_2 = 0,3$; $P_3 = 0,2$ та $P_4 = 0,1$ підключено до каналу передачі інформації, який має пропускну здатність $C = 1000 \frac{\text{дв. од.}}{\text{с}}$.

Передача інформації здійснюється рівномірним двійковим кодом. З якою швидкістю буде здійснюватися передача інформації при відсутності завад?

Розв'язання. Швидкість передачі інформації визначаємо за формулою (8.11).

Середня ентропія повідомлень на один символ джерела:

$$H(Y) = -0,4 \log_2 0,4 - 0,3 \log_2 0,3 - \\ -0,2 \log_2 0,2 - 0,1 \log_2 0,1 = 1,848 \frac{\text{дв. од.}}{\text{символ}}$$

Швидкість передачі символів U_y може бути визначена з виразу для пропускну здатності каналу. Дійсно, максимальна ентропія буде мати місце при рівній ймовірності і статистичній незалежності повідомлень:

$$\max \{H(Y)\} \log_2 4 = 2 \frac{\text{дв. од.}}{\text{повід.}}$$

Тоді з виразу (8.12) отримаємо:

$$\bar{U}_y = \frac{C}{\max \{H(Y)\}} = \frac{1000}{2} = 500 \frac{\text{символ}}{\text{с}}$$

Звідси, швидкість передачі інформації:

$$\bar{T}(\bar{Y}) = \bar{U}_y H(Y) = 500 \cdot 1,848 = 924 \frac{\text{дв. од.}}{\text{с}}$$

Приклад 8.4

Скільки в середньому можна передати букв російського тексту в секунду по каналу без завад з пропускну здатністю $C = 1000 \frac{\text{дв. од.}}{\text{с}}$.

Розв'язування. У п. 7.7 встановлено, що середня ентропія російської мови на одну букву з врахуванням всіх корелятивних зв'язків:

$$H(Y) = 2 \frac{\text{дв. од.}}{\text{букву}}$$

по каналу:
$$\bar{U}_y = \frac{1000}{2} = 500 \frac{\text{букв}}{\text{с}}$$

Приклад 8.5

При передачі неперервних сигналів по каналу зв'язку здійснюється кодоімпульсна модуляція. Квантування за часом здійснюється з частотою $F_k = 100$ Гц. При квантуванні за рівнем шкала рівнів сигналу поділяється на 64 дискретних рівня.

Визначити пропускну здатність каналу без завад.

Розв'язування. Кількість дискретних рівнів визначає кількість повідомлень, які передаються. Звідси максимально можлива ентропія повідомлень: $\max\{H(Y)\} = \log_2 64 = 6 \frac{\text{дв. од.}}{\text{повід.}}$, а пропускна здатність

каналу:
$$C = U_y \max\{H(Y)\} = F_k \max\{H(Y)\} = 100 \cdot 6 = 600 \frac{\text{дв. од.}}{\text{с}}$$

Приклад 8.6

По каналу зв'язку з завадами передаються три повідомлення, яким відповідають сигнали y_1, y_2 та y_3 . При цьому канал такий, що сигнал y_1 завжди передається безпомилково, а сигнали y_2 та y_3 передаються або правильно з ймовірностями p , або зі спотвореннями, коли один сигнал переходить в інший з ймовірністю q . Априорні ймовірності посилення сигналів $p(y_1) = p, p(y_2) = p(y_3) = Q$. Схема передачі сигналів показана на рис. 8.7.

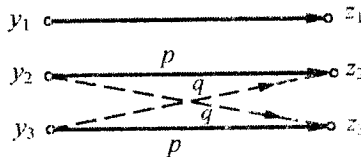


Рис. 8.7

Визначити пропускну здатність каналу, якщо швидкість передачі сигналів дорівнює U_y .

Розв'язування. Пропускна здатність дискретного каналу з завадами може бути визначена за формулою: $C = \bar{U}_y \max \{H(Y) - H(Y/X)\}$.

Початкова ентропія сигналів:

$$H(Y) = -\sum_{i=1}^3 p(y_i) \log_2 p(y_i) = -[p \log_2 p + 2Q \log_2 Q].$$

Залишкова ентропія визначається виразом:

$$\begin{aligned} H(Y/Z) &= -\sum_{i=1}^3 \sum_{j=1}^3 p(z_i) p(y_j/z_i) \log_2 p(y_j/z_i) = \\ &= -p(z_1) p(y_1/z_1) \log_2 p(y_1/z_1) - p(z_2) [p(y_2/z_2) \log_2 p(y_2/z_2) + \\ &+ p(y_3/z_2) \log_2 p(y_3/z_2)] - p(z_3) [p(y_2/z_3) \log_2 p(y_2/z_3) + \\ &+ p(y_3/z_3) \log_2 p(y_3/z_3)]. \end{aligned}$$

Оскільки за умовою задачі $p(y_1/z_1) = 1$, то перший доданок у нашому виразі буде дорівнювати нулю.

Крім того, за умовою задачі:

$$p(y_2/z_2) = p(y_3/z_3) = p; \quad p(y_2/z_3) = p(y_3/z_2) = q.$$

Ймовірності $p(z_2)$ і $p(z_3)$ знаходимо із формули повної ймовірності:

$$p(z_2) = p(y_2) p(z_2/y_2) + p(y_3) p(z_2/y_3) = Q [p(z_2/y_2) + p(z_2/y_3)] = Q,$$

$$p(z_3) = p(y_2) p(z_3/y_2) + p(y_3) p(z_3/y_3) = Q [p(z_3/y_2) + p(z_3/y_3)] = Q.$$

Отже, залишкова ентропія:

$$H(Y/Z) = -2Q(p \log_2 p + q \log_2 q).$$

Таким чином, пропускна здатність каналу визначається виразом:

$$C = \bar{U}_y \max \{-p \log_2 p - 2Q \log_2 Q + 2Q(p \log_2 p + q \log_2 q)\}.$$

Шукана величина C є функцією апріорних ймовірностей p і Q . Отже, знаходження максимуму правої частини отриманого виразу зводиться до відповідного вибору p і Q при умові, що $p + 2Q = 1$. Як відомо, такі задачі на максимум вирішуються за допомогою невизначених коефіцієнтів Лагранжа. Згідно з цим методом складемо нову функцію виду:

$$F = p \log_2 p - 2Q \log_2 Q + 2Q(p \log_2 p + q \log_2 q) + \lambda(p + 2Q),$$

де λ – коефіцієнт Лагранжа.

Для зручності математичних викладок перейдемо до натуральних логарифмів:

$$F = -p \ln p - 2Q \ln Q + 2Q(p \ln p + q \ln q) + \lambda(p + 2Q).$$

Шукаємо максимум даної функції, для чого спочатку знайдемо її часткові похідні по p і Q і прирівняємо їх до нуля:

$$\frac{\partial F}{\partial p} = -1 - \ln p + \lambda = 0;$$

$$\frac{\partial F}{\partial Q} = -2 - 2 \ln Q + 2(p \ln p + q \ln q) + 2\lambda = 0.$$

Із отриманих рівнянь знайдемо: $p = e^{\lambda-1}$; $Q = e^{\lambda-1} e^{(p \ln p + q \ln q)}$.

Із умови $p + 2Q = 1$ знайдемо: $e^{\lambda-1} (1 + 2e^{(p \ln p + q \ln q)}) = 1$, звідки:

$$e^{\alpha-1} = \frac{1}{1 + 2e^{(p \ln p + q \ln q)}} = \frac{\delta}{\delta + 2}, \text{ де } \delta = e^{-(p \ln p + q \ln q)}.$$

Таким чином, оптимальні значення апріорних ймовірностей:

$$p = \frac{\delta}{\delta + 2}, \quad Q = \frac{1}{\delta + 2},$$

Звідси, шукане значення пропускнуої здатності каналу:

$$\begin{aligned} C &= \bar{U}_y \left[-\frac{\delta}{\delta + 2} \ln \frac{\delta}{\delta + 2} - \frac{2}{\delta + 2} \ln \frac{1}{\delta + 2} - \frac{2}{\delta + 2} \ln \delta \right] = \\ &= \bar{U}_y \left[-\ln \delta \left(\frac{\delta}{\delta + 2} + \frac{2}{\delta + 2} \right) + \ln(\delta + 2) \left(\frac{\delta}{\delta + 2} + \frac{2}{\delta + 2} \right) \right] = \\ &= \bar{U}_y [\ln(\delta + 2) - \ln \delta] = \bar{U}_y \ln \frac{\delta + 2}{\delta}. \end{aligned}$$

Приклад 8.7

Визначити пропускну здатність неперервного каналу з завадами типу білого гаусового шуму. Смуга пропускання каналу $F = 1000$ Гц.

Дисперсія корисного сигналу $\delta_y^2 = P_v = 4000 \text{ В}^2$, спектральна щільність потужності завади $P_0 = 0,002 \text{ В}^2/\text{с}$.

Розв'язування. Для визначення пропускної здатності неперервного каналу з завадами скористаємося формулою (8.27):

$$C = F_v \log_2 \left(1 + \frac{P_v}{P_0 F_v} \right) = 1000 \log_2 \left(1 + \frac{4 \cdot 10^3}{0,002 \cdot 10^3} \right) = 10970 \frac{\text{дв. од.}}{\text{с}}$$

Приклад 8.8

Визначити необхідну пропускну здатність та смугу пропускання каналу, який призначений для передачі телевізійного зображення. Можна вважати, що телевізійне зображення складається із 300 000 дрібних елементів зображення. Кожен із цих елементів може приймати 10 різних градаций яскравості. Всі градации можна вважати рівномірними. За одну секунду передається 30 кадрів зображення. Крім того, відомо, що для достатнього відтворення зображення необхідне відношення сигнал/шум дорівнює 1000.

Розв'язування. Оскільки кожен елемент зображення може приймати 10 рівнів з рівною ймовірністю, то один елемент зображення містить кількість інформації: $i = \log_2 10 = 3,32 \frac{\text{дв. од.}}{\text{елемент}}$. Один кадр мі-

стить кількість інформації, яка дорівнює 300000 i . Оскільки за одну секунду повинно бути передано 30 кадрів, то пропускна здатність каналу:

$$C = 30 \cdot 300000 i = 29,9 \cdot 10^6 \frac{\text{дв. од.}}{\text{с}}$$

З іншого боку, при відомому відношенні сигнал/шум пропускна здатність каналу може бути визначена за формулою (8.25) звідки смуга

$$\text{пропускання каналу: } F_v \cong \frac{C}{\log_2 \frac{P_v}{P_z}} = \frac{29,9 \cdot 10^6}{9,95} \cong 3 \text{ МГц.}$$

Приклад 8.9

Визначити пропускну здатність телефонного каналу, який забезпечує передачу розмови з швидкістю 100 слів/хв.

Вважати, що середня довжина слова складається з п'яти букв.

Розв'язування. Вважаючи, що на кожну букву в середньому припадає дві двійкові одиниці інформації (див. п. 7.7), одержимо наступне значення пропускної здатності каналу:

$$C = \frac{100 \cdot 5 \cdot 2}{60} = 16,6 \frac{\text{дв. од.}}{\text{с}}$$

Контрольні питання

1. Якими узагальненими характеристиками визначаються сигнал та канал передачі інформації?
2. Що розуміють під об'ємом сигналу та смістю каналу передачі інформації?
3. Сформулюйте необхідні і достатні умови узгодження сигналу з каналом передачі інформації?
4. Чим спричиняється необхідність узгодження сигналу з каналом передачі інформації?
5. Поясніть, як можна передати широкосмуговий сигнал через вузькосмуговий канал без спотворень?
6. Що розуміють під швидкістю передачі інформації та пропускну здатністю каналу?
7. Якими факторами визначаються швидкість передачі інформації та пропускну здатність каналу?
8. Що розуміють під дискретним каналом і неперервним каналом?
9. Теорема Шеннона для дискретного каналу без завад
10. Що розуміють під каналом без пам'яті?
11. Яким чином визначається швидкість передачі інформації по дискретному каналу з завадами і без завад?
12. Яким чином визначається пропускну здатність дискретного каналу з завадами та без завад?
13. Теорема Шеннона для дискретного каналу з завадами?
14. Яким чином визначається швидкість передачі інформації по неперервному каналу з завадами?
15. Яким чином визначається пропускну здатність неперервного каналу з завадами при передачі сигналів з обмеженою потужністю?
16. Пояснити характер залежності пропускну здатності неперервного каналу з завадами від смуги пропускання каналу.
17. Теорема Шеннона для неперервного каналу з завадами.

ОПТИМАЛЬНИЙ ПРИЙОМ ТА ОБРОБКА ІНФОРМАЦІЇ

9

Задача оптимального прийому полягає у використанні надлишковості, а також у наявних даних про властивості корисного сигналу, завади і каналу для збільшення ймовірності правильного прийому.

Внаслідок того, що на вхід приймача надходить сума корисного сигналу і завад, ймовірність правильного прийому буде визначатись відношенням корисного сигналу до завади. Для підвищення ймовірності правильного прийому повинна бути здійснена попередня обробка прийнятого сигналу, яка забезпечує збільшення відношення сигнал/завада. Таким чином, приймач повинен містити два основні елементи (рис. 9.1): фільтр Φ , що забезпечує покращення відношення сигнал/завада, та розв'язувальний пристрій РП, які виконують головні функції прийому (виявлення, розпізнавання або відновлення сигналів).



Рис. 9.1

Відомі наступні методи фільтрації, що забезпечують покращення відношення сигнал/завада:

- частотна фільтрація;
- метод накопичування;
- кореляційний метод;
- узгоджена фільтрація.

Всі ці методи основані на використанні різних властивостей корисного сигналу та завади.

9.1 Частотна фільтрація

Ідея частотної фільтрації заснована на відмінності спектрів корисного сигналу та завади. При цьому використовуються лінійні частотні фільтри, які дозволяють заглушати заваду та покращувати тим самим співвідношення сигнал/завада. Параметри фільтра визначаються спектральними характеристиками сигналу і завади.

На практиці найбільш часто зустрічаються наступні випадки:

а) на вхід приймального пристрою надходять вузькосмуговий сигнал і широкосмугова завада. Цей випадок показаний на рис. 9.2, де $S_x(\omega)$ та $S_z(\omega)$ – спектральна щільність сигналу і завади відповідно. У цьому випадку в тракт приймального пристрою ввімкнено вузькосмуговий фільтр зі смугою пропускання $\Delta\omega_f$;

б) на вхід приймача надходять широкосмуговий сигнал і вузькосмугова завада (рис. 9.3). У таких випадках в тракт приймача вмикається фільтр, що забезпечує заглушення завади у смузі $\Delta\omega_f$;

в) на вхід приймача надходять періодичний сигнал та широкосмугова завада.

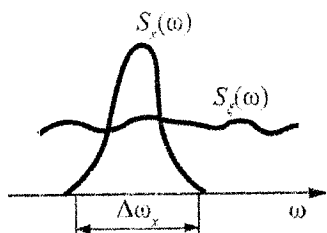


Рис. 9.2

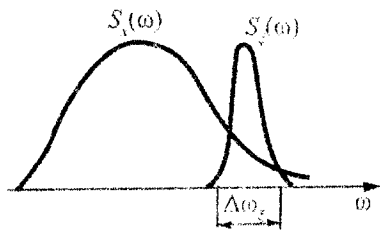


Рис. 9.3

Як відомо, періодичний сигнал має дискретний частотний спектр. У такому випадку в приймальній пристрій потрібно ввімкнути набір фільтрів (гребінчасті фільтри), які пропускають лише дискретні частоти періодичного сигналу.

Розглянемо більш детально випадок, коли корисний сигнал є гармонічним, а завада – типу білого шуму. Для виділення корисного сигналу в цьому випадку повинен бути використаний вузькосмуговий фільтр, який настроєно на частоту сигналу. Відношення потужності сигналу до потужності завади на вході фільтра при цьому:

$$\left(\frac{P_x^*}{P_z} \right)_{\text{вхл}} = \frac{P_{x\text{вх}}}{P_0 \Delta\omega_f} = \frac{P_{x\text{вх}}}{2\pi \Delta f_f P_0}, \quad (9.1)$$

де P_0 – середня потужність завади, що припадає на одиницю смуги; $\Delta\omega_f$ – смуга пропускання фільтра.

Як видно із виразу (9.1), відношення P_x^*/P_z можна зробити як завгодно великим за рахунок зменшення смуги пропускання фільтра Δf_f .

У реальних умовах корисний сигнал надходить лише під час визначеного часу T_x . Тому корисний сигнал в дійсності являє собою від-

різок синусоїди, спектр якої необмежений. Відомо, що практична ширина спектра такого сигналу пов'язана з його тривалістю співвідношенням:

$$\Delta f_x T_x = \mu, \quad (9.2)$$

де μ – постійна, яка залежить від форми сигналу; зазвичай приймається $\mu \cong 1$.

Тривалість сигналу T_x повинна бути вибрана такою, щоб його спектр був не ширшим смуги пропускання фільтра: $\Delta f_x \leq \Delta f_\phi$.

Підставивши в (9.1) замість Δf_ϕ величину Δf_x , отримаємо:

$$\left(\frac{P_x^*}{P_\xi} \right)_{\text{вих}} = \frac{P_{\text{max}}}{2\pi P_0 |t|} T_x. \quad (9.3)$$

Формула (9.3) показує, що збільшення відношення сигнал/завада досягається за рахунок збільшення тривалості сигналу T_x , тобто часу спостереження. Таким чином, при частотній фільтрації поліпшення відношення сигнал/завада реалізується ціною збільшення часу передачі.

9.2 Метод накопичування

Метод накопичування застосовується в тому випадку, якщо корисний сигнал протягом часу прийому постійний або є періодичною функцією. Метод полягає в багатократному повторенні сигналу та підсумовуванні окремих його реалізацій в приймальному пристрої. Такий спосіб обробки сигналу забезпечує при визначених умовах поліпшення відношення сигнал/завада.

Нехай протягом часу передачі T_x сигнал постійний і дорівнює a . На сигнал діє адаптивна завада ξ .

Послідовність відліків можна представити у вигляді: $y_1 = a + \xi_1$;
 $y_2 = a + \xi_2$; ... $y_n = a + \xi_n$; де ξ_i – значення завади в момент i -го відліку.

На виході накопичувача буде сигнал $Y = \sum_{i=1}^n (a + \xi_i) = na + \sum_{i=1}^n \xi_i$,
де n – кількість відліків за час передачі.

Відношення потужностей сигналу і завади на виході накопичувача:

$$\left(\frac{P_x}{P_\xi} \right)_{\text{вих}} = \frac{(na)^2}{D \left[\sum_{i=1}^n \xi_i \right]},$$

де $D\left[\sum_{i=1}^n \xi_i\right]$ – дисперсія завади на виході накопичувача.

При припущенні, що значення ξ_i некорельовані, дисперсія суми підрахунків ξ_i дорівнює сумі дисперсії відліків: $D\left[\sum_{i=1}^n \xi_i\right] = \sum_{i=1}^n D(\xi_i)$.

Вважаючи заваду стандартним випадковим процесом, отримуємо:

$$D\left[\sum_{i=1}^n \xi_i\right] = nD(\xi).$$

Отже, відношення потужностей сигналу і завади на виході накопичувача можна представити у вигляді:

$$\left(\frac{P_x}{P_\xi}\right)_{\text{вих}} = \frac{(na)^2}{nD(\xi)} = n \frac{a^2}{D(\xi)} = n \left(\frac{P_x}{P_\xi}\right)_{\text{вх}}. \quad (9.4)$$

Таким чином, при перерахованих умовах, в результаті n -кратного відліку, відношення потужностей сигналу і завади збільшується в n раз. Часовий інтервал між окремими підрахунками повинен бути більший інтервалу кореляції завади τ_0^ξ . В іншому випадку виграш за рахунок накопичування буде менше значення, що дається виразом (9.4).

За рахунок збільшення кількості відліків n , тобто часу передачі T_x , можна скільки завгодно збільшувати відношення сигнал/завада.

Якщо сигнал є періодичною функцією часу, то відліки слід проводити через інтервали, які рівні або кратні періоду цієї функції. У таких випадках метод носить назву методу синхронного накопичування. Ефект накопичування такий самий, як і у випадку постійного сигналу.

Ефект накопичування можна здійснювати також за рахунок інтегрування вхідного сигналу впродовж визначеного часу T_x . Такий метод отримав назву інтегрального прийому. Інтегральний прийом доцільно застосовувати у випадку, коли корисний сигнал постійний (або квазіпостійний).

У випадку надходження на вхід інтегратора суми постійного сигналу a та завади ξ вихідний сигнал становить:

$$Y_a(t) = \int_0^{T_x} [a + \xi(t)] dt = aT_x + \int_0^{T_x} \xi(t) dt.$$

Відношення потужностей сигналу і завади на виході інтегратора:

$$\left(\frac{P_x}{P_\xi} \right)_{\text{вих}} = \frac{(aT_x)^2}{D \left[\int_0^{T_x} \xi(t) dt \right]}$$

Згідно з теоремою Котельникова, неперервний випадковий сигнал $\xi(t)$ можна замінити сукупністю відліків у моменти часу, що віддалені один від одного на інтервали, рівні інтервалу кореляції завади τ_0 .

Отже, можна записати, що: $\int_0^{T_x} \xi(t) dt \cong \sum_{i=1}^n \xi(t_i) \tau_0$, де $n = T_x / \tau_0$.

Тоді дисперсія завади на виході накопичувача:

$$D \left[\int_0^{T_x} \xi(t) dt \right] \cong D \left[\sum_{i=1}^n \xi(t_i) \tau_0 \right] = \tau_0^2 D \left[\sum_{i=1}^n \xi(t_i) \right] = \tau_0^2 n D(\xi) = T_x \tau_0 D(\xi).$$

Отже, відношення потужностей сигналу і завади на виході інтегратора визначається співвідношенням:

$$\left(\frac{P_x}{P_\xi} \right)_{\text{вих}} \cong \frac{(aT_x)^2}{T_x \tau_0 D(\xi)} = \frac{T_x a^2}{\tau_0 D(\xi)} = \frac{T_x}{\tau_0} \left(\frac{P_x}{P_\xi} \right)_{\text{вх}}. \quad (9.5)$$

Таким чином, інтегральний прийом забезпечує збільшення потужностей сигналу та завади в T_x / τ_0 раз. Завадостійкість інтегрального прийому буде тим вища, чим більший час інтегрування T_x і чим більш високочастотною є завада (тобто чим менший інтервал кореляції завади τ_0).

9.3 Кореляційний метод

Сутність методу полягає у використанні різниці між кореляційними функціями сигналу та завади. Даний метод ефективний лише у випадку прийому періодичних або квазіперіодичних сигналів.

Розглянемо сутність методу на прикладі, коли корисний сигнал є періодичним, а завада – типу білого шуму Гауса.

У приймальному пристрої визначається кореляційна функція суми корисного сигналу та завади, що надходить на вхід:

$$\begin{aligned}
K_y(\tau) &= \frac{1}{T_1} \int_{-1/2T_1}^{1/2T_1} [X(t) + \xi(t)][X(t+\tau) + \xi(t+\tau)] dt = \\
&= \frac{1}{T_1} \int_{-1/2T_1}^{1/2T_1} X(t)X(t+\tau) dt + \frac{1}{T_1} \int_{-1/2T_1}^{1/2T_1} X(t)\xi(t+\tau) dt + \\
&+ \frac{1}{T_1} \int_{-1/2T_1}^{1/2T_1} \xi(t)X(t+\tau) dt + \frac{1}{T_1} \int_{-1/2T_1}^{1/2T_1} \xi(t)\xi(t+\tau) dt = \\
&= K_{xx}(\tau) + K_{x\xi}(\tau) + K_{\xi x}(\tau) + K_{\xi\xi}(\tau).
\end{aligned} \tag{9.6}$$

В отриманому виразі $K_{x\xi}(\tau)$ і $K_{\xi x}(\tau)$ є взаємними кореляційними функціями сигналу та завади, а $K_{xx}(\tau)$ і $K_{\xi\xi}(\tau)$ – автокореляційними функціями сигналу та завади відповідно.

Оскільки сигнал, що передається, і завада статистично незалежні, то $K_{x\xi}(\tau) = K_{\xi x}(\tau) = 0$. Отже, вираз (9.6) прийме вигляд:

$$K_y(\tau) = K_{xx}(\tau) + K_{\xi\xi}(\tau). \tag{9.7}$$

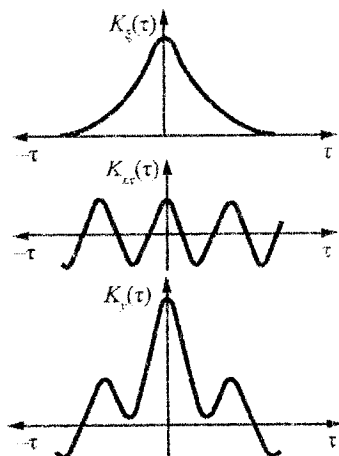


Рис. 9.4

З виразу (9.7) видно, що кореляційна функція суміші сигналу і завади дорівнює сумі автокореляційних функцій сигналу і завади. Як відомо, кореляційна функція періодичного сигналу є періодичною функцією аргументу. Функція $K_{\xi\xi}(\tau)$ зі збільшенням τ наближається до нуля і при $\tau \geq \tau_0$ практично дорівнює нулю (рис. 9.4). Отже, вибираючи такий

час τ , при якому значенням $K_{\xi\xi}(\tau)$ можна знехтувати, ми забезпечимо тим самим отримання функції $K_y(\tau)$, яка відображає корисний сигнал, тобто виділення корисного сигналу із суміші корисного сигналу з завадою.

Для уточнення того, від яких факторів залежить час, який витрачається для виділення корисного сигналу при кореляційному прийомі, виразимо кореляційні функції $K_{xx}(\tau)$ та $K_{\xi\xi}(\tau)$ через дисперсії й нормовані кореляційні функції $r_{xx}(\tau)$ та $r_{\xi\xi}(\tau)$:

$$K_{xx}(\tau) = D(x) \cdot r_{xx}(\tau); \quad K_{\xi\xi}(\tau) = D(\xi)r_{\xi\xi}(\tau). \quad (9.8)$$

Підставивши вираз (9.8) у (9.7), знайдемо:

$$\begin{aligned} K_y(\tau) &= D(x)r_{xx}(\tau) + D(\xi)r_{\xi\xi}(\tau) = \\ &= D(x) \left[r_{xx}(\tau) + \frac{D(\xi)}{D(x)} r_{\xi\xi}(\tau) \right] = D(x) \left[r_{xx}(\tau) + \frac{P_\xi}{P_x} r_{\xi\xi}(\tau) \right]. \end{aligned} \quad (9.9)$$

З виразу (9.9) видно, що вибір часу прийому залежить від інтервалу кореляції завади τ_0 і відношення сигнал/завада. Дійсно, при $P_\xi / P_x \ll 1$ другим членом можна знехтувати і для виділення корисного сигналу теоретично часу не потрібно. При $P_\xi / P_x \geq 1$ необхідно збільшувати τ , щоб отримати $r_{\xi\xi}(\tau) \cong 0$. Отже, в останньому випадку для виділення корисного сигналу необхідний додатковий час, який повинен зростати зі збільшенням відношення P_ξ / P_x та інтервалу кореляції завади τ_0 .

Загальний час, який витрачається на кореляційний прийом, визначається не тільки τ , але і часом інтегрування T_i , який вибирається достатньо великим. Практично обмежуються значенням $T_i \geq 10\tau_0$.

Схема кореляційного приймача наведена на рис. 9.5.

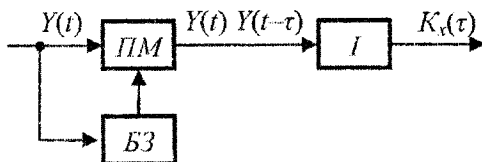


Рис. 9.5

Схема містить блок затримки $БЗ$, пристрій множення $ПМ$ та інтегратор $І$. Блок затримки здійснює затримку вхідного сигналу на час τ . Пристрій множення здійснює перемноження функції $Y(t)$, що надійшла

на вхід приймача і функції $Y(t + \tau)$, яка знімається з блоку затримки. Інтегратор усереднює отриманий добуток за час T_1 .

9.4 Узгоджена фільтрація

Узгоджені фільтри призначені для виділення сигналів відомої форми на фоні шумів. Критерієм оптимальності таких фільтрів є отримання на виході максимально можливого відношення амплітудного значення сигналу до діючого значення завади.

Функція передачі таких фільтрів визначається виразом:

$$K(j\omega) = C \frac{S_x^*(j\omega)}{G_z(\omega)} \exp\{-j\omega T\}, \quad (9.10)$$

де $S_x^*(j\omega)$ – комплексно-спряжена функція спектра сигналу; $G_z(\omega)$ – енергетичний спектр завади на вході фільтра; T – момент часу спостереження за сигналом на виході фільтра, в який досягається максимум відношення сигнал/завада; C – коефіцієнт пропорційності.

Для випадку, коли завада є білим шумом, формула (9.10) приймає вигляд:

$$K(j\omega) = B S_x^*(j\omega) \exp\{-j\omega T\}, \quad (9.11)$$

де $B = \frac{C}{G_{z_0}}$; G_{z_0} – енергетичний спектр білого шуму.

Якщо корисний сигнал описується функцією $x(t)$, то перехідна функція оптимального фільтра має вигляд:

$$h(t) = Bx(T-t), \quad (9.12)$$

тобто являє собою дзеркальне відображення сигналу відносно часової точки $T/2$ (рис. 9.6).

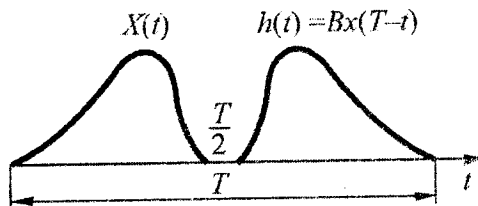


Рис. 9.6

Реакція на виході фільтра визначається інтегралом згортки:

$$\begin{aligned}
 Y_{\phi}(t) &= \int_{-\infty}^{\infty} [x(\tau) + \xi(\tau)]h(t - \tau)d\tau = \\
 &= B \int_{-\infty}^{\infty} x(\tau)x(T - t + \tau)d\tau + B \int_{-\infty}^{\infty} \xi(\tau)x(T - t + \tau)d\tau.
 \end{aligned}
 \tag{9.13}$$

Формула (9.13) показує, що вихідна реакція фільтра з точністю до постійного множника виражає суму автокореляційної функції сигналу і взаємно кореляційної функції сигналу і завади.

Отже, реакція узгодженого фільтра еквівалентна дії кореляційного приймача, який можливо реалізувати.

9.5 Сутність основної задачі прийому сигналів при наявності завад

Результатом дії завад є часткова або повна втрата інформації, яка переноситься корисним сигналом. Приймальний пристрій, здійснюючи обробку вхідного сигналу, який є сумою корисного сигналу і завади, повинен забезпечити вилучення з прийнятого сигналу якнайбільшої кількості необхідної інформації.

Основна задача приймача полягає у тому, щоб на основі прийнятої реалізації вирішити, найкращим в певному розумінні способом, чи наявний даний сигнал в даній реалізації (задача знаходження або розпізнавання), або якими є параметри корисного сигналу (задача відтворення). У зв'язку з цим повинні бути вироблені критерії, які дозволяють за прийнятим сигналом оптимальним способом вирішити поставлену задачу.

Задача вибору оптимального способу обробки сигналів і вироблення при цьому відповідних критеріїв становлять зміст теорії статистичних рішень. Деякі положення цієї теорії наводяться нижче. З метою наочного представлення положень теорії статистичних рішень введені геометричні поняття простору сигналу, який приймається (простору спостережень).

Нехай відліки сигналу, що приймається, є сумою корисного сигналу і завади, здійснюються в дискретні моменти часу t_1, t_2, \dots, t_n . Відлікові значення прийнятого сигналу y_1, y_2, \dots, y_n називають вибірковими значеннями, а їх сукупність – вибіркою. Число n вибіркових значень називають розміром або об'ємом вибірки.

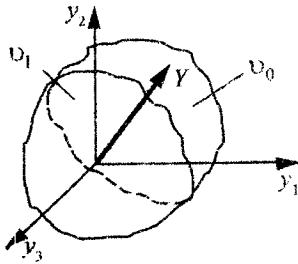


Рис. 9.7

Сукупність вибірових значень представляють геометрично у вигляді радіус-вектора Y в n -вимірному просторі, де y_1, y_2, \dots, y_n – координати кінця вектора. Оскільки величини y_1, y_2, \dots, y_n випадкові, то вектор Y також є випадковим вектором. Множина можливих значень вектора Y складає простір спостережень V . На рис. 9.7 відображений випадок тривимірного простору ($n = 3$).

Загальна ймовірність попадання кінця вектора Y в будь-яку точку простору:

$$\int_V \omega(y) dy = 1. \quad (9.14)$$

Аналогічно вводять поняття вектора корисного сигналу і вектора завад і відповідно до них поняття простору корисного сигналу і простору завад. Після знаходження вектора прийнятого сигналу Y можна однозначно судити про вектор корисного сигналу X . Мова може йти тільки про апостеріорну щільність ймовірності $\omega(X/Y) = \omega(x_1, x_2, \dots, x_n / y_1, y_2, \dots, y_n)$, тобто умовну щільність ймовірності X , якщо заданий вектор Y .

Розрахунок апостеріорної щільності ймовірності можна виконати за допомогою формули Байеса:

$$\omega(X/Y) = \frac{\omega(X)\omega(Y/X)}{\omega(Y)}, \quad (9.15)$$

де $\omega(X)$ – апіорна щільність ймовірності вектора X ; $\omega(Y)$ – безумовна щільність ймовірності вектора Y ; $\omega(X/Y)$ – умовна щільність ймовірності Y , якщо заданий X .

Безумовна щільність ймовірності $\omega(Y)$ визначається виразом:

$$\omega(Y) = \int_{V_x} \omega(X)\omega(Y/X) dX, \quad (9.16)$$

де V_x означає, що інтегрування здійснюється у просторі сигналу X . Якщо підставити значення $\omega(Y)$ з виразу (9.16) в (9.15), то отримаємо:

$$\omega(X, Y) = \frac{\omega(X)\omega(Y/X)}{\int_{V_x} \omega(X)\omega(Y/X) dx}. \quad (9.17)$$

Якщо вектор X може мати скінченне число можливих значень x_1, x_2, \dots, x_r з апіорними ймовірностями $p(x_1), p(x_2), \dots, p(x_r)$, то формула (9.17) приймає вигляд:

$$p(X/Y) = \frac{p(X)\omega(Y/X)}{\omega(Y)} = \frac{p(X)\omega(Y/X)}{\sum_{j=1}^r p(x_j)\omega(Y/x_j)}, \quad (9.18)$$

де $p(X/Y)$ – апостеріорна ймовірність вектора X , якщо заданий вектор Y ; $p(X)$ – апіорна ймовірність вектора X .

Отже для знаходження ймовірності (або щільності ймовірності) необхідно знати $p(X)$ або $\omega(X)$, тобто апіорні характеристики корисного сигналу, і $\omega(Y/X)$, які визначаються апіорними характеристиками корисного сигналу завади, а також характером їх композиції.

Для визначення апостеріорних ймовірностей $p(X/Y)$ або щільності ймовірності $\omega(X/Y)$ необхідно знати $\omega(Y/X)$, яка при заданому значенні Y буде залежати тільки від X : $\omega(Y/X) = L(X)$.

Функція $L(X)$ називається функцією правдоподібності. Залежно від того, є X дискретною чи неперервною величиною, функція правдоподібності $L(X)$ може приймати скінченну або нескінченну множину значень.

Розглянемо основні критерії, які використовуються при вирішенні задачі оптимального прийому. Почнемо з найпростішої задачі – виявлення сигналів.

9.6 Виявлення сигналу

Задача виявлення полягає в тому, щоб в результаті обробки прийнятого сигналу Y встановити, міститься в ньому корисний сигнал X чи ні.

Нехай сигнал, що приймається, є сумою корисного сигналу і завад:

$$y(t) = x(t) + \xi(t).$$

Корисний сигнал може приймати два значення: x_1 і x_0 з апіорними ймовірностями $p(x_1)$ і $p(x_0)$ відповідно. Оскільки сигнал X напевно має одне з цих двох значень, то справедливе співвідношення:

$$p(x_1) + p(x_0) = 1. \quad (9.20)$$

Таким чином, можливі дві взаємовиключні альтернативні гіпотези: в прийнятому сигналі міститься корисний сигнал (гіпотеза H_1) і

відеутній корисний сигнал (гіпотеза H_0). Розв'язувальний пристрій приймача за даними вибірки повинен встановити, яка з цих гіпотез є істинною.

У геометричній інтерпретації поставлена задача може бути сформульована наступним чином. Простір прийнятих сигналів V умовно розбивається на дві частини (див. рис. 9.7): область v_1 , яка відповідає прийняттю гіпотези H_1 про те, що $X = x_1$, і область v_0 , яка відповідає прийняттю гіпотези H_0 про те, що $X = x_0$. Це означає, що якщо вектор прийнятого сигналу з'явиться в межах області v_1 , то приймається гіпотеза H_1 . Якщо ж вектор сигналу Y з'явиться в області v_0 , то приймається гіпотеза H_0 .

У цих умовах можуть мати місце два значення апостеріорної ймовірності $p(X/Y)$: $p(X_1/Y)$ – умовна ймовірність наявності корисного сигналу X при даному значенні вибірки Y ; $p(X_0/Y)$ – умовна ймовірність відсутності X при даному значенні вибірки Y .

Аналогічно можна розглядати два значення функції правдоподібності $L(X)$: $L(x_1) = \omega(Y/x_1)$ – умовна щільність ймовірності вибірки Y при наявності корисного сигналу X ; $L(x_0) = \omega(Y/x_0)$ – умовна щільність ймовірності вибірки Y при відсутності X .

Відношення функцій правдоподібності

$$\lambda = \frac{L(x_1)}{L(x_0)} = \frac{\omega(Y/x_1)}{\omega(Y/x_0)} \quad (9.21)$$

прийнято називати відношенням правдоподібності.

Для вибору гіпотези H_1 або H_0 необхідно взяти за основу визначене правило прийняття рішень.

Вибір правила прийняття рішення у математичному відношенні зводиться до оптимального розбиття простору V сигналів, що приймаються, на області v_1 та v_0 . Для того, щоб вибрати те або інше правило прийняття рішення, необхідно керуватися певними критеріями.

Критерій максимуму правдоподібності. Цей критерій формулюється наступним чином: найбільш правдоподібне те значення параметра X , для якого функція правдоподібності $L(X)$ максимальна.

Згідно з цим критерієм у випадку двоальтернативної ситуації знаходження сигналу порівнюються два значення функції правдоподібності $L(x_1)$ та $L(x_0)$ і приймається та гіпотеза, якій відповідає більше значення функції правдоподібності. Якщо, наприклад, $L(x_1) > L(x_0)$, то приймається гіпотеза H_1 . Якщо $L(x_1) \leq L(x_0)$, то приймається гіпотеза H_0 .

Цей критерій можна записати в наступному вигляді через відношення правдоподібності:

$$\begin{aligned} \text{якщо } \lambda = \frac{L(x_1)}{L(x_0)} > 1, \text{ то } x = x_1; \\ \text{якщо } \lambda = \frac{L(x_1)}{L(x_0)} \leq 1, \text{ то } x = x_0. \end{aligned} \quad (9.22)$$

Таким чином, згідно з даним критерієм методика прийняття рішення зводиться до наступного: обчислюються функції правдоподібності $L(x_1)$ і $L(x_0)$, визначається відношення правдоподібності λ і залежно від того, більше, рівно або менше λ одиниці, приймається відповідна гіпотеза.

Практична перевага даного критерію полягає в тому, що при його застосуванні не потрібні знання апіорних ймовірностей $p(x_1)$ і $p(x_0)$ сигналу X .

Критерій максимуму апостеріорної ймовірності. За цим критерієм при отриманому значенні вибірки Y приймається та гіпотеза, при якій апостеріорна ймовірність $p(X/Y)$ максимальна.

Для випадку двохальтернативної ситуації порівнюються два значення апостеріорної ймовірності: $p(x_1/Y)$ і $p(x_0/Y)$. Зазвичай розглядається відношення цих величин і правило прийняття рішення записується у вигляді:

$$\begin{aligned} \text{якщо } \frac{p(x_1/Y)}{p(x_0/Y)} > 1, \text{ то } X = x_1; \\ \text{якщо } \frac{p(x_1/Y)}{p(x_0/Y)} \leq 1, \text{ то } X = x_0. \end{aligned} \quad (9.23)$$

Використовуючи формулу Байєса (9.18), виразимо відношення апостеріорної ймовірності через відношення функцій правдоподібності:

$$\frac{p(x_1/Y)}{p(x_0/Y)} = \frac{p(x_1)L(X_1)}{p(x_0)L(X_0)} = \frac{p(x_1)}{p(x_0)} \lambda. \quad (9.24)$$

Тоді критерій максимуму апостеріорної ймовірності (9.23) може бути наступним чином виражений через відношення правдоподібності:

$$\begin{aligned} \text{якщо } \frac{p(x_1)}{p(x_0)} \lambda > 1, \text{ то } X = x_1; \\ \text{якщо } \frac{p(x_1)}{p(x_0)} \lambda \leq 1, \text{ то } X = x_0. \end{aligned} \quad (9.25)$$

Співвідношення (9.25) можна представити у вигляді:

$$\begin{aligned} \text{якщо } \lambda > \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_1; \\ \text{якщо } \lambda \leq \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_0. \end{aligned} \quad (9.26)$$

де λ_0 – порогове значення відношення правдоподібності.

Таким чином, процедура прийняття рішення відповідно до критерію максимуму апостеріорної ймовірності така сама, як і відповідно до критерію максимуму правдоподібності. Відмінність полягає лише в тому, що в першому випадку відношення правдоподібності порівнюється з одиницею, а в другому – з відношенням апіорних ймовірностей $p(x_1)/p(x_0)$. За наявності апіорних даних $p(x_1)$ і $p(x_0)$ доцільно застосувати критерій максимуму апостеріорної ймовірності, тому що при цьому є можливість користуватися допоміжною інформацією, яка дозволяє точніше вирішувати задачу виявлення сигналу.

Слід зазначити, що критерій максимуму правдоподібності є оптимальним з інформаційної точки зору. Дійсно, з точки зору теорії інформації найбільш ймовірним слід вважати те значення параметра X , відносно якого в прийнятому сигналі Y міститься найбільша кількість інформації. Взвзявши різницю кількості інформації, яка міститься в сигналі Y відносно x_1 та x_0 , отримаємо:

$$\begin{aligned} I(Y, x_1) - I(Y, x_0) &= [-\log_2 p(x_1) + \log_2 p(x_1/Y)] - [-\log_2 p(x_0/Y) - \\ &-\log_2 p(x_0)] = \log_2 \frac{p(x_1/Y)p(x_0)}{p(x_0/Y)p(x_1)} = \log_2 \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \log_2 \lambda. \end{aligned}$$

Таким чином, інформаційний критерій прийняття рішення зводиться до визначення двійкового логарифма відношення правдоподібності. Якщо цей логарифм додатний, то приймається гіпотеза H_1 про те, що $X = x_1$; якщо він від'ємний або рівний нулю, то приймається гіпотеза H_0 про те, що $X = x_0$.

Критерій ідеального спостерігача (критерій Котельникова). Згідно з даним критерієм приймається та гіпотеза, при якій забезпечується мінімум загальної помилки прийняття рішення.

При рішенні задачі виявлення сигналу можуть мати місце помилки двох видів:

1) при відсутності корисного сигналу вектор прийнятого сигналу Y з'являється у області v_1 і згідно з цим приймається гіпотеза H_1 ;

2) при наявності корисного сигналу вектор Y з'являється в області v_0 і згідно з цим приймається гіпотеза H_0 . Перша помилка називається помилкою першого роду, або "хибною тривоною". Друга помилка називається помилкою другого роду, або "пропуском сигналу". Кількісно помилки першого та другого родів оцінюються умовними ймовірностями α та β помилкових рішень про наявність корисного сигналу, коли в дійсності він відсутній, і про відсутність сигналу, коли в дійсності він є:

$$\alpha = p(Y \in v_1 / x_0) = \int_{v_1} \omega(Y / x_0) dY; \quad (9.27)$$

$$\beta = p(Y \in v_0 / x_1) = \int_{v_0} \omega(Y / x_1) dY.$$

Загальна безумовна ймовірність помилкового рішення визначається виразом:

$$p_{\text{ном}} = p(x_0)\alpha + p(x_1)\beta. \quad (9.28)$$

Критерій ідеального спостерігача мінімізує загальну помилку, яка визначається виразом (9.28). Отже, умова оптимального рішення за критерієм ідеального спостерігача має вигляд:

$$p_{\text{ном}} = p(x_0)\alpha + p(x_1)\beta = \min. \quad (9.29)$$

Підставимо у вираз (9.28) з виразу (9.27) значення помилок першого та другого роду:

$$p_{\text{ном}} = p(x_0) \int_{v_1} \omega(Y / x_0) dY + p(x_1) \int_{v_0} \omega(Y / x_1) dY. \quad (9.30)$$

Помилку другого роду можна представити у вигляді:

$$\beta = p(Y \in v_0 / x_1) = 1 - p(Y \in v_1 / x_1) = 1 - \int_{v_1} \omega(Y / x_1) dY. \quad (9.31)$$

Підставивши з виразу (9.31) в (9.30) значення β , отримаємо:

$$p_{\text{ном}} = p(x_1) - \int_{v_1} [p(x_1)\omega(Y / x_1) - p(x_0)\omega(Y / x_0)] dY. \quad (9.32)$$

Умова (9.29) буде забезпечена, якщо інтеграл в (9.32) буде максимальним. А для цього потрібно так вибрати область v_1 , щоб підінтегральна функція була додатною, тобто:

$$p(x_1)\omega(Y/x_1) - p(x_0)\omega(Y/x_0) > 0 \quad (9.33)$$

Умова (9.33) визначає належність вектора Y області v_1 , тобто вибір гіпотези H_1 . Перепишемо вираз (9.33) у вигляді:

$$\text{якщо } \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda > \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_1;$$

$$\text{якщо } \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda \leq \frac{p(x_0)}{p(x_1)} = \lambda_0, \text{ то } X = x_0.$$

Таким чином, правила рішення, що відповідають критеріям ідеального спостерігача та максимуму апостеріорної ймовірності, збігаються. Відмінність полягає в початкових умовах.

Критерій Неймана–Пірсона. Цей критерій заснований на тому, що помилки першого і другого родів не однаково небезпечні, причому помилка першого роду приводить до таких наслідків, що її ймовірність необхідно обмежити деякою дуже малою величиною. Другу помилку бажано при цьому забезпечити мінімальною.

Виходячи з цього, критерій Неймана–Пірсона можна сформулювати наступним чином: найкращим рішенням є таке, при якому забезпечується найменша ймовірність помилки другого роду при заданій допустимій ймовірності помилки першого роду. І тому, згідно з критерієм Неймана–Пірсона повинно бути забезпечено:

$$\beta = \int_{v_0} \omega(Y(x_1)) dY = \min \quad (9.34)$$

при

$$\alpha = \int_{v_1} \omega(Y/x_0) dY = \epsilon, \quad (9.35)$$

де ϵ — наперед задана величина.

Задача може бути вирішена методом Лагранжа відшукання умовного екстремуму. Для спрощення задачі доцільно перейти від багатовимірної змінної Y до одновимірної змінної λ , що можна здійснити за допомогою рівностей:

$$\begin{aligned} \omega(Y/x_1) dY &= \omega(\lambda/x_1) d\lambda; \\ \omega(Y/x_0) dY &= \omega(\lambda/x_0) d\lambda. \end{aligned} \quad (9.36)$$

При такому переході областям v_1 і v_0 у просторі Y відповідають області $[0, \lambda_0]$ і $[\lambda_0, \infty)$ значень λ .

Умовні ймовірності помилок першого і другого роду будуть при цьому представлені у вигляді:

$$\alpha = \int_{\nu_1}^{\infty} \omega(Y/x_0) dY = \int_{\lambda_0}^{\infty} \omega(\lambda/x_0) d\lambda; \quad (9.37)$$

$$\beta = \int_{\nu_0}^{\infty} \omega(Y/x_1) dY = \int_0^{\lambda_0} \omega(\lambda/x_1) d\lambda. \quad (9.38)$$

Тоді для відшукування умовного екстремуму повинна бути складена допоміжна функція:

$$F = \lambda \int_{\lambda_0}^{\infty} \omega(\lambda/x_0) d\lambda + \int_0^{\lambda_0} \omega(\lambda/x_1) d\lambda.$$

Взявши похідну $\frac{\partial F}{\partial \lambda_0}$ і прирівнявши її до нуля, отримаємо

$$\omega(\lambda_0/x_1) - \lambda \omega(\lambda_0/x_0) = 0 \text{ або } \frac{\omega(\lambda_0/x_1)}{\omega(\lambda_0/x_0)} = \lambda.$$

Але з виразу (9.36): $\frac{\omega(\lambda/x_1)}{\omega(\lambda/x_0)} = \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda$ і відповідно

$$\frac{\omega(\lambda_0/x_1)}{\omega(\lambda_0/x_0)} = \lambda_0.$$

Таким чином, даний критерій буде справедливий при $\lambda = \lambda_0$, де порогове значення λ_0 визначається з рівності:

$$\alpha = \int_{\lambda_0}^{\infty} \omega(\lambda/x_0) d\lambda = \varepsilon. \quad (9.39)$$

Отже, правило прийняття рішення згідно з критерієм Неймана-Пірсона може бути записане у вигляді:

$$\text{якщо } \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda > \lambda_0, \text{ то } X = x_1; \quad (9.40)$$

$$\text{якщо } \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda \leq \lambda_0, \text{ то } X = x_0.$$

Критерій мінімального ризику (критерій Байеса). Цей критерій враховує не тільки нерівномірність помилок першого і другого роду, але і ті наслідки, до яких призводять ці помилки. Для врахування цих наслідків введені вагові коефіцієнти (коефіцієнти ціни помилок) r_{10} і r_{01} , які приписують помилкам першого і другого роду відповідно.

Усереднена величина:

$$r = r_{10}p(x_0)\alpha + r_{01}p(x_1)\beta \quad (9.41)$$

отримала назву ризику.

Згідно з критерієм мінімального ризику правило вибору рішення формулюється наступним чином: приймається та гіпотеза, при якій забезпечується мінімальний ризик:

$$r = \min. \quad (9.42)$$

Представимо вираз (9.41) у вигляді:

$$\begin{aligned} r &= r_{10}p(x_0) \int_{v_1} \omega(Y/x_0)dY + r_{01}p(x_1) \int_{v_0} \omega(Y/x_1)dY = \\ &= r_{01}p(x_1) - \int_{v_1} [r_{01}p(x_1)\omega(Y/x_1) - r_{10}p(x_0)\omega(Y/x_0)]dY. \end{aligned} \quad (9.43)$$

Мінімум виразу (9.43) буде при умові, якщо підінтегральна функція додатна:

$$r_{01}p(x_1)\omega(Y/x_1) - r_{10}p(x_0)\omega(Y/x_0) > 0.$$

Звідси отримуємо наступне правило прийняття рішення:

$$\begin{aligned} \text{якщо } \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda > \frac{r_{10}p(x_0)}{r_{01}p(x_1)} = \lambda_0, \text{ то } X = x_1; \\ \text{якщо } \frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda \leq \frac{r_{10}p(x_0)}{r_{01}p(x_1)} = \lambda_0, \text{ то } X = x_0. \end{aligned} \quad (9.44)$$

Критерій, що розглядається, є найбільш економічно доцільним, оскільки забезпечує мінімізацію витрат, обумовлених помилками в прийнятті рішень. Але він вимагає максимальної апріорної інформації, тому що крім функцій розподілу $\omega(Y/X)$ і апріорних ймовірностей $p(X)$ необхідно також знати вагові коефіцієнти r_{10} і r_{01} .

Мінімаксний критерій. Мінімаксний критерій являє собою спеціальний випадок критерію мінімального ризику, коли апріорні ймовірності $p(x_1)$ і $p(x_0)$ не задані.

Справа в тому, що ризик r , який приймає найменше значення при умові (9.44), залежить від апіорних ймовірностей $p(x_1)$ і $p(x_0)$. При деякому співвідношенні цих ймовірностей, яке ми називаємо найгіршим, ризик r буде максимальним.

Ідея мінімаксного критерію полягає в тому, що забезпечується мінімум ризику при найгіршому співвідношенні апіорних ймовірностей. Для визначення найгіршого співвідношення між $p(x_1)$ і $p(x_0)$ необхідно привіяти до нуля похідну від правої частини виразу (9.43) по $p(x_1)$ (або по $p(x_0)$). У результаті отримаємо трансцендентне рівняння, яке забезпечує максимум ризику. Потім визначається порогове значення відношення правдоподібності:

$$\lambda_0^* = \frac{r_{10} P^*(x_0)}{r_{01} P^*(x_1)}, \quad (9.45)$$

де $p^*(x_0)$ і $p^*(x_1)$ – найбільш несприятливі значення апіорних ймовірностей $p(x_1)$ і $p(x_0)$, отримані з умови $\frac{\partial r}{\partial p(x_1)} = 0$.

Таблиця 9.1

| Критерій | Порогове значення відношення правдоподібності |
|-------------------------------------|---|
| Максимум правдоподібності | $\lambda_0 = 1$ |
| Максимум апіостеріорної ймовірності | $\lambda_0 = \frac{p(x_0)}{p(x_1)}$ |
| Ідеальний спостерегач | $\lambda_0 = \frac{p(x_0)}{p(x_1)}$ |
| Неймана - Пірсона | λ_0 знаходиться з умови $\int_{\lambda_0}^{\infty} \omega(\lambda/x_0) d\lambda = \epsilon$ |
| Мінімальний ризик | $\lambda_0 = \frac{r_{10} P(x_0)}{r_{01} P(x_1)}$ |
| Мінімакс | $\lambda_0^* = \frac{r_{10} P^*(x_0)}{r_{01} P^*(x_1)}$, де $p^*(x_0)$ і $p^*(x_1)$ знаходяться з умови $\frac{\partial r}{\partial p(x_1)} = 0$ |

Таким чином, правило прийняття рішення для всіх розглянутих критеріїв однакове і зводиться до порівняння відношення правдоподібності λ з пороговим значенням λ_0 . Різниця полягає лише у величині λ_0 . Для наочності значення λ_0 для окремих критеріїв зведені у таблицю 9.1.

Оскільки величина λ_0 визначає межу між областями v_1 і v_0 простору V , то кожний критерій визначає спосіб розбиття простору прийнятого сигналу на області v_1 і v_0 . Рівність $\frac{\omega(Y/x_1)}{\omega(Y/x_0)} = \lambda_0$ визначає рівняння поверхні поділу областей v_1 і v_0 .

9.7 Розпізнавання сигналів

При розпізнаванні сигналів має місце багатоальтернативна ситуація, коли корисний сигнал X може мати багато значень і приймач повинен визначити, яке саме значення з цієї множини наявне в дійсності. Розпізнавання багатьох сигналів у принциповому відношенні мало чим відрізняється від випадку виявлення сигналу, тобто випадку розпізнавання двох сигналів. Згідно з цим методи багатоальтернативних рішень є узагальненням відповідних методів двоальтернативних рішень.

Нехай сигнал X може мати m можливих значень x_1, x_2, \dots, x_m з апіорними ймовірностями $p(x_1), p(x_2), \dots, p(x_m)$ відповідно:

$$X = \begin{cases} x_1 \rightarrow p(x_1); \\ x_2 \rightarrow p(x_2); \\ \dots\dots\dots \\ x_m \rightarrow p(x_m). \end{cases}$$

При цьому простір сигналу V розбивається на m областей v_1, v_2, \dots, v_m , що відповідають прийняттю гіпотез H_1, H_2, \dots, H_m про те, що $X = x_1, X = x_2, \dots, X = x_m$ відповідно. Правила прийняття рішень і розбиття простору V на області v_1, v_2, \dots, v_m можуть проводитись відповідно до будь-якого із критеріїв, розглянутих для випадку двоальтернативної ситуації.

Процедура роботи розв'язувального пристрою приймача при розпізнаванні сигналів наступна. За даними вибірки Y визначаються функції правдоподібності $L(x_1) = \omega(Y/x_1), L(x_2) = \omega(Y/x_2), \dots, L(x_m) = \omega(Y/x_m)$, і обчислюються відношення правдоподібності $\lambda_{ji} = \frac{\omega(Y/x_j)}{\omega(Y/x_i)}$ для всіх

можливих сполучень пар x_j і v_i . Порівнюються отримані значення відношень правдоподібності з пороговим значенням і вибирається таке значення сигналу x_j , для якого всі $\lambda_{ji} > \lambda_0$ ($i = 1, 2, \dots, m$).

Розглянемо як приклад випадок, коли використовується критерій мінімального ризику. У випадку багатоальтернативної ситуації помилка прийняття рішення полягає в тому, що вибірка, опиняється в області v_k , в той час як в дійсності сигнал X має значення x_j . Ціна помилкових рішень враховується шляхом введення вагових коефіцієнтів r_{jk} .

Для заданого значення сигналу x_j середня величина втрат за рахунок неправильних рішень може бути оцінена коефіцієнтом:

$$r_j = \sum_{k=1}^m r_{jk} p(Y \in v_k / x_j) = \sum_{k=1}^m r_{jk} \int_{v_k} \omega(Y / x_j) dY, \quad (9.46)$$

де $p(Y \in v_k / x_j)$ – умовна ймовірність попадання вибірки Y в область v_k , якщо в дійсності сигнал X дорівнює x_j .

Величини r_j носять назву умовного ризику.

Усереднюючи умовний ризик за всіма можливими значеннями X , отримаємо середній ризик:

$$\begin{aligned} r &= \sum_{j=1}^m p(x_j) r_j = \sum_{j=1}^m \sum_{k=1}^m p(x_j) r_{jk} p(Y \in v_k / x_j) = \\ &= \sum_{j=1}^m \sum_{k=1}^m p(x_j) r_{jk} \int_{v_k} \omega(Y / x_j) dY. \end{aligned} \quad (9.47)$$

Критерій мінімального ризику для випадку багатоальтернативної ситуації зводиться до мінімізації функції (9.47):

$$r = \min. \quad (9.48)$$

Умова (9.48) визначає правила прийняття рішення, а також спосіб розбиття простору прийнятого сигналу на області v_1, v_2, \dots, v_m .

Розмірковуючи, як при виборі співвідношення (9.44), можна показати, що реалізація умови (9.48) дає наступну систему m нерівностей, які забезпечують прийняття гіпотези H_k , що $X = x_k$:

$$\sum_{i=1}^m (c_i - r_k) \frac{p(x_i) \omega(Y / x_i)}{p(x_k) \omega(Y / x_k)} > 0 \quad (j = 1, 2, \dots, m, j \neq k). \quad (9.49)$$

9.8 Відновлення сигналів

Відновлення сигналів зводиться до оцінювання деякого числа невідомих параметрів корисного сигналу. Обмежимося розглядом випадку оцінювання одного з параметрів сигналу, наприклад, амплітуди B , при заданій формі сигналу. При цьому завжди будемо вважати адитивними типу білого гаусового шуму. Представимо корисний сигнал у вигляді:

$$x(t) = Bf(t),$$

де $f(t)$ – відома функція часу, B – параметр сигналу.

Задача полягає в тому, щоб за прийнятою вибіркою Y визначити, яким є значення параметра B у корисному сигналі X .

На відміну від випадків виявлення і розпізнавання сигналів має місце нескінченна множина гіпотез. Методи, розглянуті у випадку двох- та багатоальтернативних ситуацій, можуть застосовуватись і для задачі відновлення сигналу. Проведемо оцінку параметра B методом максимуму правдоподібності. Якщо відлік прийнятого сигналу проводиться в дискретні моменти часу, то функція правдоподібності для параметра B буде дорівнювати:

$$L(B) = \omega(Y/B) = \left(\frac{1}{\sqrt{2\pi\sigma_\xi^2}} \right)^n \exp \left\{ - \frac{\sum_{j=1}^n [y(t_j) - Bf(t_j)]^2}{2\sigma_\xi^2} \right\}. \quad (9.50)$$

Задача полягає в тому, щоб знайти таке значення параметра B , для якого функція правдоподібності максимальна. Максимуму функції правдоподібності відповідає мінімальне значення показника степеня у виразі (9.50).

$$\begin{aligned} \Phi(B) &= \frac{1}{2\sigma_\xi^2} \sum_{i=1}^n [y(t_i) - Bf(t_i)]^2 = \\ &= \frac{1}{2\sigma_\xi^2} \left\{ \sum_{j=1}^n y[t_j]^2 - 2B \sum_{i=1}^n y(t_i)f(t_i) + B^2 \sum_{i=1}^n [f(t_i)]^2 \right\} = \min. \end{aligned}$$

З умови мінімуму $\frac{\partial \Phi(B)}{\partial B} = 0$ маємо:

$$-2 \sum_{i=1}^n y(t_i)f(t_i) + 2B \sum_{i=1}^n [f(t_i)]^2 = 0,$$

звідки отримуємо оціночне значення параметра:

$$B^* = \frac{\sum_{i=1}^n y(t_i) f(t_i)}{\sum_{i=1}^n [f(t_i)]^2}. \quad (9.51)$$

Здійснивши перехід до неперервного прикладу, отримаємо:

$$B^* = \frac{\int_0^T y(t) f(t) dt}{\int_0^T [f(t)]^2 dt} = \frac{\int_0^T y(t) f(t) dt}{E_0}, \quad (9.52)$$

де $\int_0^T [f(t)]^2 dt$ — питома енергія сигналу (енергія сигналу при амплітуді $B = 1$).

На рис. 9.8 наведена схема розв'язувального пристрою, який здійснює операцію оцінки параметра сигналу. Пристрій містить генератор сигналу, множучою ланку ML , що здійснює множення $y(t)$ на $f(t)$, та інтегратор, що здійснює інтегрування добутку $f(t)y(t)$.

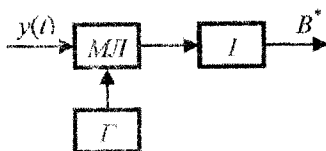


Рис. 9.8

Для оцінювання точності відновлення сигналу використовуємо критерій середнього квадратичного відхилення. З цією метою у виразі (8.52) сигнал, що приймається, виразимо у вигляді суми $y(t) = f(t) + \xi(t)$.

Тоді:

$$B^* = \frac{B}{E_0} \int_0^T [f(t)]^2 dt + \frac{1}{E_0} \int_0^T \xi(t) f(t) dt = B + \frac{1}{E_0} \int_0^T \xi(t) f(t) dt.$$

Похибка відновлення:

$$B^* - B = \frac{1}{E_0} \int_0^T \xi(t) f(t) dt.$$

Дисперсія похибки:

$$\begin{aligned}\sigma^2 &= m_1 \left\{ \frac{1}{E_0^2} \left[\int_0^T \xi(t) f(t) dt \right]^2 \right\} = m_1 \left\{ \frac{1}{E_0^2} \int_0^T \xi(t_1) f(t_1) dt_1 \int_0^T \xi(t_2) f(t_2) dt_2 \right\} = \\ &= \frac{1}{E_0^2} \int_0^T \int_0^T m_1 \{ \xi(t_1) \xi(t_2) \} f(t_1) f(t_2) dt_1 dt_2.\end{aligned}$$

Середнє від добутку $\xi(t_1)\xi(t_2)$ становить кореляційну функцію завади: $m_1 \{ \xi(t_1)\xi(t_2) \} = K\xi(t) = G_0\delta(\tau)$, де G_0 – спектральна щільність завади; $\delta(\tau)$ – дельта-функція.

$$\tau = t_2 - t_1.$$

Тоді:

$$\begin{aligned}\sigma^2 &= \frac{G_0}{E_0^2} \int_0^T \int_0^T \delta(t_2 - t_1) f(t_1) f(t_2) dt_1 dt_2 = \\ &= \frac{G_0}{E_0^2} \int_0^T f(t_1) \left[\int_0^T \delta(t_2 - t_1) f(t_2) dt_2 \right] dt_1 = \frac{G_0}{E_0^2} \int_0^T [f(t_1)]^2 dt = \frac{G_0}{E_0}.\end{aligned}$$

Отже, середнє квадратичне значення похибки відновлення:

$$\sigma = \sqrt{\frac{G_0}{E_0}}. \quad (9.53)$$

Задача відновлення сигналу може бути також вирішена методом оптимальної фільтрації. У п. 9.5 розглядалися оптимальні фільтри, які забезпечують при відомій формі сигналу максимально можливе відношення амплітудного значення сигналу до діючого значення завади. Такі фільтри отримали найменування узгоджених (відносно форми сигналу) або фільтрів Норса. Фільтри цього типу доцільно використовувати при рішенні задач розпізнавання або виявлення сигналів.

Максимізуючи відношення сигнал/завада навіть ціною суттєвого спотворення форми сигналу, ці фільтри забезпечують зменшення помилок при розпізнаванні або виявленні сигналу.

Однак при рішенні задачі відновлення важливе значення має забезпечення мінімальної помилки оцінки параметрів корисного сигналу. Ця задача вирішується за допомогою фільтра Калманова-Вінера. Критерієм оптимальності даних фільтрів є отримання мінімаль-

ної середньої квадратичної помилки відтворення сигналів. Ці фільтри не тільки очищають сигнал від завади, але і передбачають його значення на деякий час вперед. У зв'язку з цим їх часто називають фільтрами згладжування та передбачення. При реалізації таких фільтрів вважається, що корисний сигнал є випадковим процесом з відомими ймовірнісними характеристиками.

Показано, що якщо на вхід фільтра надходить сума сигналів та завад, які є стаціонарними випадковими функціями часу, то мінімальна середня квадратична помилка відтворення сигналу буде при умові, що функція передачі фільтра виражається співвідношенням:

$$K(j\omega) = \frac{S_x(j\omega)}{S_x(j\omega) + S_\xi(j\omega)} \exp\{j\omega T\}, \quad (9.54)$$

де $S_x(j\omega)$, $S_\xi(j\omega)$ – спектральні щільності корисного сигналу та завади відповідно; T – час передбачення.

При цьому середня квадратична помилка відтворення сигналу:

$$\sigma = \sqrt{\int_0^\infty \frac{S_x(j\omega)S_\xi(j\omega)}{S_x(j\omega) + S_\xi(j\omega)} d\omega}. \quad (9.55)$$

Звідси видно, що ідеальна фільтрація ($\sigma = 0$) можлива тільки у тому випадку, коли $S_x(j\omega) \cdot S_\xi(j\omega) = 0$, тобто коли спектри сигналу та завади не перекриваються.

9.9 Приклади

Приклад 9.1

По каналу зв'язку, в якому діє адитивна стаціонарна завада, передається періодична послідовність прямокутних імпульсів. Параметри корисного сигналу: величина $A_x = 2$ В, період проходження $T_n = 100$ мс. Завада має нормальний закон розподілу. Середнє квадратичне значення завади $\sigma_\xi = 5$ В, математичне очікування $m_\xi = 0$. Обробка сигналів на приймальній стороні здійснюється методом синхронного накопичення.

Визначити час обробки сигналів, необхідний для забезпечення підвищення сигналу над завадою в чотири рази.

Розв'язування. Спрощена схема приймача з синхронним накопичуванням наведена на рис. 9.9. Підсилювач приймача в початковому

положенні закритий. Строб-сигнал надходить синхронно з корисним сигналом і забезпечує відмикання приймача на час подачі корисного сигналу.

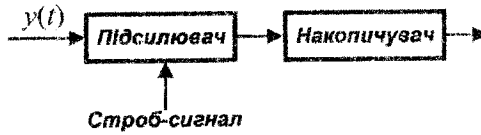


Рис. 9.9

Відношення сигнал/завада на виході накопичувача у випадку стаціонарної завади становить:

$$\left(\frac{A_x}{A_\xi} \right)_{\text{вих}} = \sqrt{n} \sqrt{\left(\frac{P_x}{P_\xi} \right)_{\text{вх}}},$$

де $\left(\frac{P_x}{P_\xi} \right)_{\text{вх}}$ – відношення потужностей та завад на вході приймача; n – кількість відліків за час приймання.

У нашому випадку:

$$\left(\frac{P_x}{P_\xi} \right)_{\text{вх}} = \left(\frac{A_x}{\sigma_\xi} \right)^2 = \left(\frac{2}{5} \right)^2.$$

Тривалість обробки сигналів у приймачі:

$$T_{np} = nT_n = \frac{\left(\frac{A_x}{A_\xi} \right)_{\text{вих}}^2}{\left(\frac{A_x}{\sigma_\xi} \right)_{\text{вх}}^2} T_n.$$

За умовою задачі: $\left(\frac{A_x}{A_\xi} \right)_{\text{вих}} = 4$.

Тоді необхідний час спостереження:

$$T_{np} = \frac{4^2}{\left(\frac{2}{5} \right)^2} \cdot 0,1 = 10 \text{ с.}$$

Приклад 9.2

Необхідно здійснити виявлення постійного сигналу величиною $a = 2$ В на фоні адитивної завади з нормальним розподілом й середнім значенням, рівним нулю. Метод прийому – однократний відлік. Здійснити синтез пристрою приймання, який працює на основі критерію максимуму правдоподібності, й визначити пороговий рівень.

Розв'язування. Оскільки за умовою задачі завада адитивна і вибірка Y становить одновимірну величину, то функції правдоподібності $L(a)$ та $L(0)$ визначаються законом розподілу завод:

$$L(a) = \omega(Y/a) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left\{-\frac{(Y-a)^2}{2\sigma_\xi^2}\right\}; \quad (9.56)$$

$$L(0) = \omega(Y/0) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left\{-\frac{Y^2}{2\sigma_\xi^2}\right\}. \quad (9.57)$$

Відношення правдоподібності при цьому:

$$\lambda = \frac{L(a)}{L(0)} = \exp\left\{-\frac{(Y-a)^2 - Y^2}{2\sigma_\xi^2}\right\} = \exp\left\{\gamma\left[\frac{Y}{a} - \frac{1}{2}\right]\right\}, \quad (9.58)$$

$$\text{де } \gamma = \frac{\alpha^2}{\sigma_\xi^2}.$$

Графіки функцій (9.56–9.58) наведені на рис. 9.10.

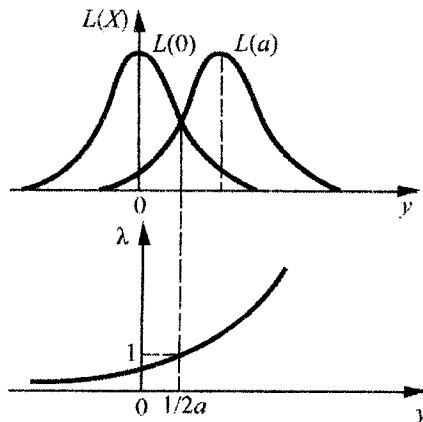


Рис. 9.10

При використанні критерію максимуму правдоподібності порогове значення відношення правдоподібності $\lambda_0 = 1$. Тоді, порівнюючи вираз (9.58) з одиницею, отримаємо умову для порогового значення вхідного сигналу.

$$\exp \left\{ \gamma \left[\frac{Y_{\text{пор}}}{a} - \frac{1}{2} \right] \right\} = 1,$$

звідки отримуємо $\frac{Y_{\text{пор}}}{a} - \frac{1}{2} = 0$, або $Y_{\text{пор}} = \frac{a}{2} = 1 \text{ В}$.

Таким чином, пристрій приймання являє собою пристрій порівняння, який порівнює вхідний сигнал з пороговим рівнем $Y_{\text{пор}}$, рівним половині величини корисного сигналу.

Якщо вхідний сигнал перевищує пороговий рівень в 1 В, то виноситься рішення, що у вхідному сигналі є корисний сигнал. Якщо ж вхідний сигнал рівний або менший $Y_{\text{пор}}$, то виноситься рішення про відсутність корисного сигналу.

Приклад 9.3

Вирішити попередню задачу для випадку, коли використовується критерій ідеального спостерігача і додатково відомо, що відношення апріорних ймовірностей відсутності й присутності корисного сигналу в прийнятому $\frac{p(0)}{p(a)} = 10$ і середнє квадратичне значення завад ξ дорівнює $\sigma_\xi = 0,5 \text{ В}$.

Розв'язування. При використанні критерію ідеального спостерігача порогове значення відношення правдоподібності $\lambda_0 = \frac{p(0)}{p(a)}$.

Тоді, порівнюючи праву частину (9.58) до порогового значення λ_0 , отримаємо наступний вираз для порогового рівня:

$$Y_{\text{пор}} = a \left(\frac{1}{\gamma} \ln \lambda_0 + \frac{1}{2} \right) = 2 \left[\left(\frac{0,5}{2} \right)^2 \ln 10 + \frac{1}{2} \right] = 1,88 \text{ В}.$$

Приклад 9.4

Вирішити задачу, наведену в прикладі 9.2, для випадку, коли використовується критерій Неймана-Пірсона і відомо, що середнє квадратичне значення завад $\sigma_\xi = 1,25 \text{ В}$ і ймовірність помилки першого роду прийняття рішення не повинна перевищувати величини $\epsilon = 0,05$.

Розв'язування. Шукане порогове значення $Y_{\text{пор}}$ можна визначити із співвідношення (9.39), якщо перейти від змінних λ до змінних Y :

$$\alpha = \int_{Y_{\text{пор}}}^{\infty} \omega(Y/0) dY = \varepsilon.$$
 Підставляючи з (9.57) вираз для щільності

$$\omega(Y/0), \text{ отримаємо } \alpha = \int_{Y_{\text{пор}}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{\xi}^2}} \exp\left\{-\frac{Y^2}{2\sigma_{\xi}^2}\right\} dY = \varepsilon.$$

Після заміни $\frac{Y}{\sqrt{\sigma_{\xi}^2}} = z$ отримуємо рівняння для знаходження $Y_{\text{пор}}$:

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{Y_{\text{пор}}}{\sigma_{\xi}}}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = \frac{1}{2} \left[1 - 2\Phi\left(\frac{Y_{\text{пор}}}{\sigma_{\xi}}\right) \right] = \varepsilon, \text{ де } \Phi\left(\frac{Y_{\text{пор}}}{\sigma_{\xi}}\right) - \text{функція Лапласа.}$$

ція Лапласа.

$$\text{За умовою задачі } \Phi\left(\frac{Y_{\text{пор}}}{\sigma_{\xi}}\right) = \frac{1}{2} - \varepsilon = 0,45.$$

За таблицями функцій Лапласа знаходимо $\frac{Y_{\text{пор}}}{\sigma_{\xi}} = 1,65$, звідки

шуканий пороговий рівень $Y_{\text{пор}} = 1,65\sigma_{\xi} = 1,65 \cdot 1,25 = 2,06$ В.

Приклад 9.5

На вхід приймального пристрою надходить суміш корисного гармонічного сигналу $x(t) = A \sin \omega_0 t$ з амплітудою $A = 0,4$ В і адитивної завади, розподіленої за нормальним законом з нульовим середнім і середнім квадратичним значенням $\sigma_{\xi} = 0,2$ В.

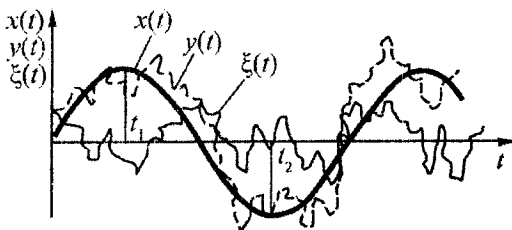


Рис. 9.11

Проведено два заміри вхідного сигналу $y(t) = x(t) + \xi(t)$ в моменти часу $t_1 = T_0/4$ та $t_2 = 3T_0/4$ (рис. 9.11), де $T_0 = 2\pi/\omega$ – період сигналу $x(t)$. Апостеріорні ймовірності значень $x_1 = 0$ і $x_2 = A = 0,4$ В становлять $p(x_1) = p(x_2) = 0,5$.

Знайти апостеріорні ймовірності значень $x_1 = 0$ і $x_2 = A$ після вказаних замірів.

Розв'язування. Шукані апостеріорні ймовірності значень можуть бути визначені за формулою Байєса (див. вираз (9.18):

$$p(x_1 / Y) = \frac{p(x_1)\omega(Y / x_1)}{p(x_1)\omega(Y / x_1) + p(x_2)\omega(Y / x_2)};$$

$$p(x_2 / Y) = \frac{p(x_2)\omega(Y / x_2)}{p(x_1)\omega(Y / x_1) + p(x_2)\omega(Y / x_2)}.$$

Умовна щільність ймовірності вибірки $y_1 = y(t_1)$, $y_2 = y(t_2)$ при $x = x_1 = 0$ та при $x = x_2 = A$ в припущенні, що відліки завад $\xi(t_1)$ та $\xi(t_2)$ статистично незалежні відповідно:

$$\omega\left(\frac{y_1 y_2}{x_1}\right) = \omega\left(\frac{y_1}{x_1}\right)\omega\left(\frac{y_2}{x_1}\right) \frac{1}{\left(\sqrt{2\pi\sigma_\xi^2}\right)^2} \exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\};$$

$$\omega\left(\frac{y_1 y_2}{x_2}\right) = \omega\left(\frac{y_1}{x_2}\right)\omega\left(\frac{y_2}{x_2}\right) \frac{1}{\left(\sqrt{2\pi\sigma_\xi^2}\right)^2} \exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}.$$

Отже, шукані апостеріорні ймовірності:

$$\begin{aligned} P\left(\frac{0}{Y}\right) &= \frac{\exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\}}{\exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\} + \exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}} = \\ &= \frac{\exp(-1,63)}{\exp(-1,63) + \exp(-0,626)} = 0,268. \end{aligned}$$

$$\begin{aligned}
 P\left(\frac{A}{Y}\right) &= \frac{\exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}}{\exp\left\{-\frac{y_1^2 + y_2^2}{2\sigma_\xi^2}\right\} + \exp\left\{-\frac{(y_1 - A)^2 + (y_2 - A)^2}{2\sigma_\xi^2}\right\}} = \\
 &= \frac{\exp(-0,626)}{\exp(-1,63) + \exp(-0,626)} = 0,732.
 \end{aligned}$$

Приклад 9.6

Виконується оцінка амплітуди A сигналу $x(t) = A \sin \omega_0 t$. Час спостереження $T_{np} = 2c \gg T_0 = 2\pi/\omega_0$. Визначити, яким повинно бути середнє значення параметра $A_{ср}$, щоб відносна похибка оцінки A становила $\gamma_\sigma = \sigma / A_{ср} = 0,01$.

Наявні завади типу білого шуму зі спектральною щільністю $C_0 = 0,2 \text{ мВ}^2\text{с}$.

Розв'язування. При $T_{np} \gg T_0$, $T_{np} = nT_0$, де $n \gg 1$.
Тоді енергія сигналу:

$$\begin{aligned}
 E_x &= n \int_0^{T_0} [A \sin \omega_0 t]^2 dt = A^2 n \int_0^{T_0} \sin^2 \omega_0 t dt = \\
 &= A^2 n \int_0^{T_0} \left(\frac{1}{2} - \frac{1}{2} \cos 2\omega_0 t \right) dt = A^2 n \frac{1}{2} T_0 = \frac{A^2 T_{np}}{2}.
 \end{aligned}$$

Абсолютна похибка оцінки параметра визначається за формулою (9.53):

$$\sigma = \sqrt{\frac{G_0}{E_0}}, \text{ де } E_0 = \frac{E_x}{A^2} = \frac{T_{np}}{2}.$$

$$\text{Тоді } \sigma = \sqrt{\frac{2G_{10}}{T_{np}}} = \sqrt{\frac{2-0,2}{2}} = 0,447 \text{ мВ}.$$

$$\text{Отже } A_{ср} = \frac{\sigma}{\gamma_\sigma} = \frac{0,447}{0,01} \cong 0,45 \text{ мВ}.$$

Приклад 9.7

На вхід оптимального фільтра Норса надходить сигнал прямокутної форми тривалістю δ_0 з амплітудою A (рис. 9.12). Необхідно ви-

значити передаточну функцію фільтра при умові, що в каналі діє білий гаусів шум.

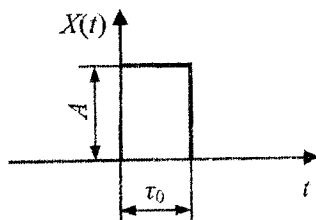


Рис. 9.12

Розв'язування. Спектральна щільність сигналу:

$$\begin{aligned} S_x(j\omega) &= \int_{-\infty}^{\infty} x(t) \exp\{-j\omega t\} dt = A \int_0^{\tau_0} \exp\{-j\omega t\} dt = \\ &= -\frac{A}{j\omega} (1 - \exp\{-j\omega\tau_0\}). \end{aligned}$$

Комплексно-сполучена спектральна щільність:

$$S_x^*(j\omega) = \frac{A}{j\omega} (\exp\{-j\omega\tau_0\} - 1).$$

Передаточна функція фільтра відповідно до (8.11)

$$\begin{aligned} K(j\omega) &= bS_x^*(j\omega) \exp\{-j\omega T\} = \frac{bA}{j\omega} (1 - \exp\{-j\omega\tau_0\}) = \\ &= \frac{B}{j\omega} (1 - \exp\{-j\omega\tau_0\}), \end{aligned}$$

де $B = bA$, $T = \tau_0$.

Така передаточна функція реалізується схемою, що складається з інтегруючої комірки з коефіцієнтом передачі, який дорівнює $\frac{B}{j\omega}$, лінії затримки, що затримує сигнал на час ω_0 , і різницевого каскаду (рис. 9.13).

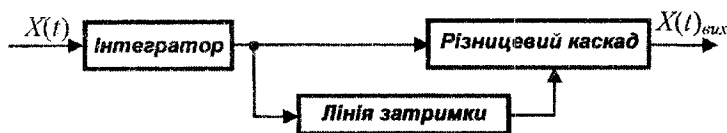


Рис. 9.13

Контрольні питання

1. У чому полягає сутність частотної фільтрації сигналів?
2. Як пов'язано поліпшення відношення сигнал/завада при частотній фільтрації з часом обробки сигналів?
3. У чому полягає сутність методу накопичення і у яких випадках доцільно його використовувати?
4. З яких міркувань потрібно вибирати інтервали між відліками при накопичувальному прийомі?
5. Чим відрізняється метод інтегрального прийому від методу накопичування, та в яких випадках він доцільний?
6. За рахунок чого досягається збільшення відношення сигнал/завада при використанні методів накопичування та інтегрального прийому?
7. У чому полягає сутність кореляційного методу прийому та в яких випадках доцільно його використовувати?
8. З яких міркувань вибирається час зсуву при кореляційному методі прийому?
9. Які фільтри називаються узгодженими?
10. Для виділення яких сигналів використовуються фільтри Норса?
11. Як пов'язана перехідна функція фільтра Норса з вхідним сигналом?
12. У чому полягає сутність задачі перевірки гіпотез?
13. Що розуміється під простором спостереження, простором корисного сигналу та простором завал?
14. Що розуміється під вибіркою та об'ємом вибірки?
15. Що таке функція правдоподібності і відношення правдоподібності?
16. Що таке помилки першого та другого родів та як вони кількісно оцінюються?
17. Як кількісно оцінюється повна помилка прийняття рішення?
18. Як залежать помилки першого і другого родів від порогового рівня?
19. У чому полягає сутність задачі виявлення сигналу?

20. У чому полягає сутність критерію максимуму правдоподібності і які його переваги?

21. У чому полягає сутність критерію максимуму апостеріорної ймовірності і які його переваги перед критерієм максимуму правдоподібності?

22. У чому полягає сутність критерію ідеального спостерігача?

23. У чому полягає відмінність критерію ідеального спостерігача від критерію максимуму апостеріорної ймовірності; що спільного у цих критеріїв?

24. У чому полягає сутність критерію Неймана–Пірсона та в яких випадках доцільно цей критерій використовувати?

25. Що розуміється під ризиком?

26. У чому полягає сутність критерію мінімального ризику?

27. У чому полягає сутність мінімального критерію?

28. У чому полягає сутність задачі розрізнення сигналів?

29. Що розуміється під ідеальним приймачем?

30. У чому полягає сутність задачі відновлення сигналів?

31. Що розуміється під фільтрами Колмогорова–Вінера?

32. У яких випадках можлива ідеальна фільтрація за допомогою фільтрів Колмогорова–Вінера?

ЕФЕКТИВНІСТЬ ІНФОРМАЦІЙНИХ СИСТЕМ

10

10.1 Критерії оцінювання ефективності інформаційних систем

Ефективність інформаційної системи характеризує її здатність забезпечити дану кількість інформації з найменшими витратами часу, смуги частот і потужності сигналу, витратами на обслуговування тощо.

Одним із показників ефективності системи може слугувати її швидкодія, тобто швидкість передачі інформації. Для порівняльної оцінки ефективності різних систем використовується критерій “питомої швидкості” передачі інформації, тобто відношення швидкості передачі інформації до смуги частот, що має місце у каналі зв’язку:

$$R = \frac{\bar{I}_T}{F_c} = \frac{I_T}{F_c T_c}, \quad (10.1)$$

де T_c – час передачі сигналу; I_T – кількість інформації, що передається за час T_c .

У тих випадках, коли потужність передавача жорстко обмежена (наприклад, у системах космічного зв’язку), як критерій ефективності застосовується коефіцієнт використання потужності сигналу:

$$\beta = \frac{I_T}{\frac{P_y}{P_\xi} F_c T_c}, \quad (10.2)$$

де P_y/P_ξ – відношення потужності сигналу до потужності завади в каналі передачі інформації.

Коефіцієнт β відіграє значну роль при порівняльному оцінюванні та енергетичному розрахунку радіоканалів з дуже великими потоками інформації. Він показує, в скільки разів середня потужність сигналу повинна перевищувати середню потужність завад для забезпечення заданої швидкості передачі інформації на 1 Гц смуги частот, що має місце у каналі зв’язку.

Часто ефективність інформаційних систем характеризують коефіцієнтом:

$$\nu = \frac{I_T}{F_c T_c \log_2 \frac{P_v}{P_s}}, \quad (10.3)$$

який отримав назву питомої змістовності сигналу. Він визначає кількість інформації в одиниці об'єму сигналу. У таблиці 10.1 наведені типові значення швидкості передачі інформації та питомої змістовності сигналу для найбільш розповсюджених видів зв'язку.

Таблиця 10.1

| Вид зв'язку | Швидкість передачі інформації \bar{T} , дв.од./с | Питома змістовність сигналу V |
|--------------------|--|---------------------------------|
| Телеграф Морзе | $8 \cdot 10^2$ | 0,45 |
| Телеграф Бодо | 80 | 0,45 |
| Фототелеграф | $2,2 \cdot 10^4$ | 0,68 |
| Телефон імпульсний | $5,6 \cdot 10^4$ | 0,79 |
| Телебачення | $5,9 \cdot 10^7$ | 0,72 |

Як видно з таблиці, найбільшу швидкодію з наведених в ній систем передачі інформації має телебачення, але найбільш економічною системою є телефон.

10.2 Способи підвищення ефективності інформаційних систем

Як видно з виразів (10.1)–(10.3), при накладанні обмежень на фізичні параметри сигналу (потужність, частоту, тривалість) ефективність систем може бути підвищена за рахунок збільшення швидкості передачі інформації, тобто для підвищення ефективності системи передачі інформації необхідно підвищувати ентропію повідомлень.

Відомо також, що при наявності кореляційних зв'язків між елементами повідомлень ентропія останніх зменшується. Тому підвищення ефективності інформаційних систем можна також отримати за рахунок усунення чи послаблення взаємозв'язків між елементами повідомлень, тобто за рахунок декореляції повідомлень. Наявність, підвищення ефективності систем можна отримати за рахунок відповідного вибору способу кодування, що забезпечує економію в часі при передачі повідомлень. Отже, найбільшу ефективність системи дає код, при якому

середня кількість кодових символів, що припадає на один елемент повідомлення, буде мінімальною. Такий код називають ефективним.

Взаємні зв'язки, що існують між окремими повідомленнями, дозволяють за даними спостережень за попередніми повідомленнями передбачити наступні повідомлення. Тоді, якщо відняти від передбаченого повідомлення дійсне, можна в лінію надіслати отриману різницю (сигнал помилки). Різницевий сигнал несе по суті ті нові відомості, що не могли бути отриманні раніше за відомими кореляційними зв'язками між повідомленнями.

Оскільки середнє значення сигналу помилки менше середнього значення сигналу, то такий спосіб дозволяє зменшити об'єм сигналу, а отже, збільшити ефективність системи.

Вибір того чи іншого способу підвищення ефективності систем серед відомих повинен робитися з урахуванням складності його технічної реалізації, а також з урахуванням забезпечення необхідної завадостійкості систем.

Розглянемо більш детально деякі найбільш загальновідомі способи підвищення ефективності систем передачі інформації.

10.3 Перерозподіл щільності ймовірностей елементів повідомлення

Раніше зазначалось, що найбільшу ентропію мають повідомлення при рівноймовірному і незалежному виникненні їх елементів. Проте висновок було зроблено без урахування енергетичних характеристик сигналу. Більш повний розгляд цього питання показав, що при накладанні обмежень на потужність сигналу максимум ентропії дає симетричний нормальний розподіл. Тому в реальних системах передачі інформації доцільно використовувати розподіл, що наближається до нормального.

Отже, якщо джерело виробляє повідомлення з розподілом елементів, що відрізняється від нормального, то з метою кращого використання каналу передачі інформації необхідно цей розподіл змінити так, щоб він наблизився до нормального закону. У загальному випадку зміна закону розподілу можлива лише за рахунок нелінійних перетворень повідомлення, що передається. Суть такого перетворення розглянемо на конкретному прикладі, який ілюструє рис. 10.1.

Сигнал x , що характеризується щільністю розподілу ймовірностей $\omega(x)$ і підлягає перетворенню, подається на відхиляючі пластини електронно-променевої трубки.

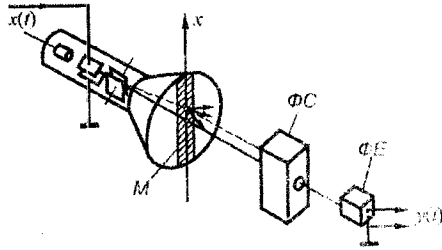


Рис. 10.1

Функції нелінійного перетворювача виконує маска M зі змінною прозорістю, що розташовується перед екраном трубки. Закон зміни прозорості (рис. 10.2) описується нелінійною функцією $f(x)$.

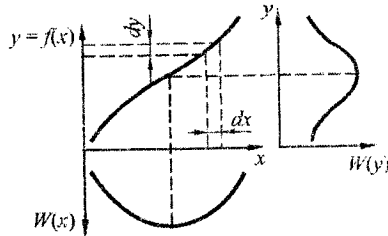


Рис. 10.2

Світловий потік Φ , що проходить через маску, знаходиться в нелінійній залежності від x :

$$\Phi = k_1 f(x), \quad (10.4)$$

де k_1 — коефіцієнт пропорційності.

За допомогою спеціальної фокусувальної системи Φ_C потік Φ попадає на фотоелемент Φ_E , струм якого пропорційний цьому потоку:

$$I = k_2 \Phi = k_1 k_2 f(x) = k f(x). \quad (10.5)$$

Отже, якщо раніше ймовірність рівня x визначалася щільністю $\omega(x)$, то тепер та сама щільність відповідає рівню $y = f(x)$. Таким чином, струм на виході фотоелемента являє собою новий сигнал з перетвореним розподілом ймовірностей $\omega(y)$.

Встановимо залежність між щільністю розподілу ймовірностей вхідного і вихідного сигналів перетворювача. З рис. 10.2 видно, що ймовірність попадання вхідного сигналу x в інтервал dx дорівнює ймовірності попадання вихідного сигналу y в інтервал dy , тобто:

$$\omega(y)dy = \omega(x)dx, \quad (10.6)$$

де $\omega(y)$ – щільність розподілу ймовірностей вихідного сигналу y .
Функції (10.5) відповідає обернена функція:

$$x = f^{-1}(y), \quad (10.7)$$

тоді

$$dx = \frac{d[f^{-1}(y)]}{dy} dy. \quad (10.8)$$

Підставляючи вирази (10.7) і (10.8) у формулу (10.6), знайдемо:

$$\omega(y)dy = \omega[f^{-1}(y)] \frac{d[f^{-1}(y)]}{dy} dy,$$

звідки:

$$\omega(y) = \omega[f^{-1}(y)] \frac{d[f^{-1}(y)]}{dy}. \quad (10.9)$$

Очевидно, сигнал на виході передавального пристрою буде спотворений внаслідок проходження його через нелінійний чотириполюсник, тому в приймальному пристрої отриманий сигнал необхідно піддати зворотній операції.

10.4 Декореляція повідомлень

Відомі декілька способів декореляції повідомлень: метод укрупнення, метод прогнозу і т.д. Ці способи детально викладені в літературі.

Розглянемо як приклад суть методу укрупнення.

Ідея декореляції методом укрупнення полягає в наступному. Сигнал розбивається не на окремі елементи, а на відрізки (поліграми), кожний з яких містить групу елементів. Ці відрізки розглядаються як елементи нового сигналу. При цьому виявляється, що взаємозв'язок між укрупненими елементами буде слабший, ніж між елементами у початкового сигналу.

Декореляція сигналів тим значніше, чим більша кількість елементів включена в поліграми.

При укрупненні сигналу кодуються не окремі елементи, а їх групи, тобто здійснюється перехід до коду з більш високою основою:

$$m_2 = m_1^r, \quad (10.10)$$

де m_1 – первинна основа коду; r – кількість елементів у поліграмі; m_2 – нова основа коду.

Ентропія, як відомо, володіє важливою властивістю адитивності. Перехід до укрупнених елементів коду забезпечує збільшення середнього значення ентропії на елемент повідомлення в r раз, оскільки середнє значення ентропії на поліграму дорівнює сумі середніх значень ентропії на елемент поліграми. Як при цьому зміниться надмірність повідомлень?

Якщо середню ентропію на елемент початкового повідомлення прийняти рівною h_1 , то після укрупнення середня ентропія на елемент буде рівна $h_2 = rh_1$. Максимальна ентропія повідомлення з укрупненими елементами $H_{2\max} = \log_2 N_2 = r \log_2 N_1 = rH_{1\max}$; де N_2 – об'єм укрупненого алфавіту; N_1 – об'єм первинного алфавіту; $H_{1\max}$ – максимальна ентропія первинного повідомлення.

Надмірність повідомлення з укрупненими елементами:

$$k_{надл2} = 1 - \frac{H_2}{H_{2\max}} = 1 - \frac{H_1}{H_{1\max}} = k_{надл1}, \quad (10.11)$$

де $k_{надл1}$ – надмірність первинного повідомлення.

Вираз (10.11) показує, що укрупнення елементів не забезпечує зменшення надмірності повідомлень. Це пояснюється тим, що при укрупненні елементів зменшення надмірності за рахунок ослаблення зв'язків між елементами компенсується збільшенням надмірності за рахунок більш нерівномірного розподілу ймовірностей елементів. Тому для повного усунення надмірності укрупнення елементів повинно поєднуватися з використанням оптимального статистичного коду для кодування укрупненого алфавіту або інших заходів, що забезпечують перерозподіл ймовірностей елементів повідомлень.

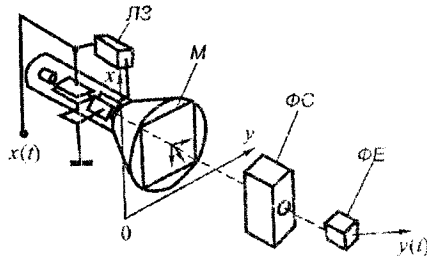


Рис. 10.3

На рис. 10.3 наведена схема пристрою, що суміщає в собі функції формування укрупненого сигналу і перерозподілу ймовірностей. У пристрої враховуються лише взаємозв'язки двох сусідніх елементів початкового сигналу.

Початковий сигнал подається на вертикально відхиляючі пластини електроіно-променевої трубки, а на горизонтально відхиляючі пластини падає той же сигнал, затриманий лінійною затримки $LЗ$ на час, рівний тривалості одного елемента. Промінь трубки в цьому випадку відхиляється по горизонталі на величину, пропорційну значенню попереднього елемента, а по вертикалі – на величину, пропорційну значенню даного елемента. Положення променя на екрані трубки визначається, таким чином, парою елементів l_i та l_j .

Закон розподілу прозорості маски M , розташованої перед екраном трубки, вибраний так, щоб забезпечувався нормальний закон розподілу ймовірностей світлового потоку, що надходить через маску і фокусувальну систему $ФС$ на фотоеlement $ФЕ$. Отже, струм фотоеlementа буде відповідати укрупненому сигналу з нормальним законом розподілу ймовірностей.

10.5 Оптимальне статистичне кодування

Оптимальне статистичне кодування забезпечує мінімізацію середньої кількості кодових символів на один елемент повідомлення. Цим забезпечується отримання максимально можливої кількості інформації, що передається кодовими комбінаціями при заданій тривалості роботи каналу, а отже, і пропускну здатності каналу.

Припустимо, що кодувальний пристрій може формувати m різних кодових комбінацій з тривалістю t_1, t_2, \dots, t_m .

Задача полягає в тому, щоб встановити таку залежність між тривалістю кодових комбінацій t_i і ймовірностями їх надходження p_i , при якій забезпечується максимум швидкості передачі інформації.

Нехай для передачі повідомлення тривалістю T необхідно N кодових комбінацій. Тоді повинні бути справедливі наступні рівності:

$$N = \sum_{i=1}^m n_i; \quad (10.12)$$

$$T = \sum_{i=1}^m n_i t_i; \quad (10.13)$$

$$\sum_{i=1}^m p_i = 1; \quad (10.14)$$

$$T = N \sum_{i=1}^m p_i t_i, \quad (10.15)$$

де n_i – середня кількість i -х кодових комбінацій, що використовуються при передачі повідомлень.

Повна кількість інформації, що міститься в повідомленні тривалістю T при незалежності появи його елементів:

$$I = -N \sum_{i=1}^m p_i \log_2 p_i. \quad (10.16)$$

Поставлена задача зводиться до забезпечення максимуму кількості інформації I відповідним вибором величин N та p_i при дотриманні умов (10.14) і (10.15).

Задача може бути вирішена методом невизначених множників Лагранжа. Відповідно до цього методу складемо нову функцію:

$$F = -N \sum_{i=1}^m p_i \log_2 p_i + \lambda_1 \left(1 - \sum_{i=1}^m p_i \right) + \lambda_2 \left(T - N \sum_{i=1}^m p_i t_i \right).$$

Визначимо максимум цієї функції, для чого знайдемо заздалегідь її частинні похідні по N та p_i і прирівняємо їх до нуля:

$$\frac{\partial F}{\partial N} = -\sum_{i=1}^m p_i \log_2 p_i - \lambda_2 \sum_{i=1}^m p_i t_i = 0$$

$$\frac{\partial F}{\partial p_i} = -N \left(\log_2 p_i + \frac{1}{\ln 2} \right) - \lambda_1 - \lambda_2 N t_i = 0, \quad (i = 1, 2, \dots, m).$$

З другого рівняння знайдемо:

$$\log_2 p_i = -\frac{\lambda_1}{N} - \lambda_2 t_i - \frac{1}{\ln 2}, \quad (i = 1, 2, \dots, m), \quad (10.17)$$

і, підставивши цей вираз в перше рівняння, будемо мати $\sum_{i=1}^m p_i \left(-\frac{\lambda_1}{N} - \frac{1}{\ln 2} \right) = 0$, звідки $\lambda_1 = -\frac{N}{\ln 2}$.

Підставивши знайдене значення λ_1 в (10.17), отримаємо: $\log_2 p_i = -\lambda_2 t_i$, ($i = 1, 2, \dots, m$), звідки:

$$p_i = 2^{-\lambda_2 t_i} \quad (i = 1, 2, \dots, m), \quad (10.18)$$

Вираз (10.18) показує характер залежності між тривалістю кодових комбінацій і ймовірностями їх появи при оптимальному кодуванні. З виразу видно, що кодові комбінації з малою ймовірністю появи повинні мати велику тривалість і навпаки.

Для з'ясування фізичного значення коефіцієнта λ_2 підставимо вираз (10.18) у (10.16). Оскільки умова (10.18) визначає умову максимізації кількості інформації, що передається за заданий час T , то отримаємо

$$-N \sum_{i=1}^m p_i (-\lambda_2 t_i) N \lambda_2 \sum_{i=1}^m p_i t_i = \lambda_2 T = I_{\max}, \text{ звідки } \lambda_2 = I_{\max} / T = c, \text{ тобто}$$

коефіцієнт λ_2 чисельно рівний пропускній здатності каналу передачі інформації. Таким чином,

$$p_i = 2^{-ct_i} \quad (i = 1, 2, \dots, m), \quad (10.19)$$

де

$$t_i = -\frac{\log_2 p_i}{c} \quad (i = 1, 2, \dots, m). \quad (10.20)$$

Вирази (10.19) і (10.20) визначають умову узгодження роботи кодувального пристрою зі статистичними властивостями повідомлень, що передаються, і забезпечує передачу інформації по каналу з швидкістю, практично рівною його пропускній здатності.

На сьогодні розроблена велика кількість різних способів оптимального статистичного кодування. Всі вони повинні забезпечувати рішення двох основних задач:

- 1) при заданій статистиці джерела повідомлень формування кодових комбінацій зі статистичними характеристиками, при яких досягається наближення швидкості передачі інформації до пропускної здатності каналу;
- 2) можливість однозначного декодування сигналів на приймальній стороні.

Для двійкового каналу з відсутністю статистичних зв'язків між символами цим вимогам задовольняє код Шеннона–Фано.

Відомо, що при відсутності статистичних зв'язків між символами, швидкість передачі інформації буде максимальною при умові рівної передачі символів 0 і 1. Згідно з цим побудова коду Шеннона–Фано здійснюється методом дихотомій (послідовного розділення навпіл). Усі символи повідомлення, що підлягають кодуванню, розбиваються на дві групи так, щоб суми ймовірностей появи елементів повідомлень у кожній групі були б по можливості однакові.

У результаті такого розбиття ніби створено нове повідомлення, що складається лише з двох елементів, ймовірності появи яких приблизно однакові. Всім символам першої групи приписується 0 і всім символам другої групи – 1. Кожна з отриманих груп розбивається на дві підгрупи

з однаковими сумарними ймовірностями і т.д. Процес поділу повторюється до тих пір, поки в кожній підгрупі залишиться по одному символу.

Методика побудови коду Шеннона-Фано для випадку передачі чотирьох символів повідомлення з ймовірностями $p(x_1) = 0,5$; $p(x_2) = 0,25$; $p(x_3) = 0,125$; $p(x_4) = 0,125$ ілюструється таблицею 10.2. Для зручності побудови всі символи повідомлення виписуються в таблицю в порядку зменшення ймовірностей. При розбитті верхнім половинам груп приписується символ 0 і нижнім – символ 1.

Таблиця 10.2

| Символ повідомлення | Ймовірність | Етап поділу на підгрупи | | | | Символ кодівих комбінацій | | | |
|---------------------|-------------|-------------------------|------|------|---|---------------------------|---|---|---|
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| x_1 | 0,5 | } I | | | | 0 | | | |
| x_2 | 0,25 | } II | } I | | | 1 | 0 | | |
| x_3 | 0,125 | | } II | } I | | 1 | 1 | 0 | |
| x_4 | 0,125 | | } II | } II | | 1 | 1 | 1 | |

Як видно з таблиці 10.2, отриманий код є нерівномірним, тому що довжина кодівих комбінацій знаходиться у зворотній залежності від їх ймовірностей. Для будь-якої позиції всієї сукупності кодівих комбінацій ймовірності передачі 0 і 1 однакові. Крім того, жодна з кодівих комбінацій не є початком іншої. Цим забезпечується вимога поділу кодівих комбінацій, тобто можливість однозначного декодування сигналів.

Пірачуємо швидкість передачі інформації, яка забезпечується отриманим кодом. Нехай тривалість символів кодівих комбінацій дорівнює τ . Тоді середня тривалість кодівих комбінацій:

$$\bar{T} = p(x_1)\tau + p(x_2)2\tau + p(x_3)3\tau + p(x_4)3\tau = 1,75\tau.$$

Середня ентропія на символ повідомлення:

$$H(X) = -\sum_{i=1}^4 p(x_i) \log_2 p(x_i) = 1,75 \frac{\text{дв.од.}}{\text{повідом.}}$$

Таким чином, швидкість передачі інформації:

$$\bar{I}(X) = \frac{1}{\tau} = C \frac{\text{дв.од.}}{\text{повідом.}}$$

Отже, отриманий код у даному випадку, дозволив отримати максимально можливе значення швидкості передачі інформації, тобто, забезпечити повне узгодження статистичних характеристик джерела

повідомлення з властивостями каналу. Це вдалось завдяки тому, що в розглянутому прикладі значення ймовірностей $p(x_i)$ вибрані такими, що умови поділу на підгрупи вдається виконати точно. У реальних умовах це, як правило, не забезпечується і швидкість передачі інформації буде менше пропускну здатності каналу.

Ефективність кодування може бути при необхідності збільшена шляхом переходу від кодування одиничних символів повідомлення до кодування груп символів, причому зі збільшенням груп ефективність буде зростати. Збільшення ефективності відбувається при цьому за рахунок того, що при збільшенні груп набір отриманих ймовірностей можна поділити на більш близькі за сумарною ймовірністю підгрупи.

10.6 Приклади

Приклад 10.1

Густина розподілу ймовірностей сигналу описується наступним законом:

$$\omega(x) = \begin{cases} \exp\{-x\}, & \text{при } x \geq 0; \\ 0, & \text{при } x < 0. \end{cases}$$

Сигнал пропускається крізь нелінійний перетворювач з характеристикою виду $y = f(x) = \sqrt[4]{x}$.

Знайти щільність розподілу ймовірностей вихідного сигналу перетворювача.

Розв'язування. Для визначення закону розподілу вихідного сигналу скористаємось співвідношенням (10.9).

Зворотня функція перетворювача $x = f^{-1}(y) = y^4$, а похідна

цієї функції
$$-\frac{d[f^{-1}(y)]}{dy} = 4y^3.$$

Тоді функція шуканого розподілу:

$$\omega(y) = \omega[f^{-1}(y)] \frac{d[f^{-1}(y)]}{dy} = 4 \exp\{-y^4\} y^3.$$

Приклад 10.2

По каналу необхідно передавати повідомлення, що формується з незалежних символів з ймовірностями появи, що визначаються таблицею 10.3.

Таблиця 10.3

| Символ | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| Ймовірність | 0,25 | 0,20 | 0,15 | 0,15 | 0,10 | 0,05 | 0,05 | 0,03 | 0,01 | 0,01 |

Канал розрахований на передачу елементів повідомлення з середньою тривалістю $\bar{\tau} = 1$ мс. Необхідно побудувати оптимальний код з час-імпульсною модуляцією.

Розв'язування. Код будемо будувати на основі залежності (10.20). Пропускнну здатність каналу визначасмо з (8.12). Середня швидкість передачі символів сигналу за умовою задачі $\bar{U}_v = \frac{1}{\tau} = 1000$ символ/с.

Тоді пропускна здатність каналу:

$$C = \bar{U}_v \log_2 n = 1000 \log_2 10 = 3320 \text{ дв.од./с.}$$

Дані розрахунку оптимального коду зведені у таблицю 10.4.

Таблиця 10.4

| Символ | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| $-\log p(x_i)$ | 2,00 | 2,32 | 2,75 | 2,74 | 3,32 | 4,32 | 4,32 | 5,06 | 6,64 | 6,64 |
| T_i , мс | 0,602 | 0,698 | 0,827 | 0,827 | 1,0 | 1,3 | 1,3 | 1,52 | 2,0 | 2,0 |

Приклад 10.3

По каналу необхідно передати повідомлення, що формується з трьох незалежних символів з ймовірностями появи, що визначаються за таблицею 10.5.

Таблиця 10.5

| Символ | x_1 | x_2 | x_3 |
|-------------|-------|-------|-------|
| Ймовірність | 0,2 | 0,7 | 0,1 |

Канал має смугу пропускання, що допускає передачу кодових комбінацій з тривалістю символів $\tau = 1$ мс.

Визначити швидкість передачі інформації по каналу при використанні коду:

- 1) рівномірного двійкового;
- 2) Шеннона-Фано при кодуванні окремих символів повідомлень;
- 3) Шеннона-Фано при кодуванні груп з двох символів повідомлення.

Розв'язування

1. Довжина (кількість розрядів) n -рівномірного двійкового коду визначається з відношення $n = \log_2 m$, де m - кількість елементів повідомлення. При $m = 3$ довжина коду повинна дорівнювати $n = 2$.

Можливі варіанти кодових комбінацій наведені у таблиці 10.6.

Таблиця 10.6

| | | | |
|--------------------|-------|-------|-------|
| Символ повідомлень | x_1 | x_2 | x_3 |
| Код | 00 | 01 | 10 |

Тривалість кодових комбінацій: $T = 2\tau = 2$ мс.

Ентропія повідомлення:

$$H(X) = -\sum_{i=1}^3 p(x_i) \log_2 p(x_i) =$$

$$= -0,2 \log_2 0,2 - 0,7 \log_2 0,7 - 0,1 \log_2 0,1 = 1,16 \frac{\text{дв. од.}}{\text{с}}.$$

Швидкість передачі інформації при використанні рівномірного двійкового коду $\bar{I}(X) = \frac{H(X)}{T} = \frac{1,16}{2 \cdot 10^{-3}} = 580 \frac{\text{дв. од.}}{\text{с}}$.

2. Побудова коду Шеннона–Фано з кодуванням окремих символів повідомлення відбувається відповідності до методики, що викладена в п. 10.5, та ілюструється таблицею 10.7.

Таблиця 10.7

| Символ повідомлень | Ймовірність | Поділ на групи | Код | Тривалість кодових комбінацій |
|--------------------|-------------|----------------|------------|-------------------------------|
| x_1 | 0,7 | } I | 00 | τ |
| x_2 | 0,2 | } II | } I 01 | 2τ |
| x_3 | 0,1 | | } II 10 | 2τ |

Середня тривалість кодових комбінацій при цьому:

$$\bar{T} = \sum_{i=1}^3 p(x_i) T_i = 0,7 \cdot 10^{-3} + 0,22 \cdot 10^{-3} + 0,1 \cdot 2 \cdot 10^{-3} = 1300 \text{ мкс.}$$

$$\text{Швидкість передачі інформації } \bar{I}(X) = \frac{1,16}{1300 \cdot 10^{-6}} = 890 \frac{\text{дв. од.}}{\text{с}}.$$

3. Для побудови коду Шеннона–Фано з кодуванням груп із двох символів повідомлення обчислюємо ймовірності окремих груп за формулою $p(x_i x_j) = p(x_i) p(x_j)$, розглядаємо в подальшому отримані групи як символи нового повідомлення і проводимо кодування за методикою п. 2. Побудова коду ілюструється таблицею 10.8.

Середня тривалість кодових комбінацій:

$$\bar{T} = 0,49\tau + 2 \cdot 0,14 \cdot 3\tau + 2 \cdot 0,07 \cdot 4\tau + 0,04 \cdot 4\tau +$$

$$+ 0,02 \cdot 5\tau + 0,02 \cdot 6\tau + 0,01 \cdot 6\tau = 1,165 \text{ мс.}$$

Таблиця 10.8

| Символ повідомлень | Ймовірність | Поділ на групи | | | | | | Код | Тривалість кодової комбінації |
|--------------------|-------------|----------------|------|------|------|------|------|-------|-------------------------------|
| x_2x_2 | 0,49 | } I | | | | | | 00 | τ |
| x_1x_2 | 0,14 | } II | } I | } I | | | | 100 | 3τ |
| x_2x_1 | 0,14 | | | } II | | | | 101 | 3τ |
| x_2x_3 | 0,07 | | } I | | | | | 1100 | 4τ |
| x_3x_2 | 0,07 | | | } I | | | | 1101 | 4τ |
| x_1x_1 | 0,04 | | | } II | | | | 1110 | 4τ |
| x_1x_3 | 0,02 | | } II | } II | } I | | | 11110 | 5τ |
| x_3x_1 | 0,02 | | | | } II | | | | 111110 |
| x_3x_3 | 0,01 | | | | | } II | } II | | 111111 |

Швидкість передачі інформації:

$$\bar{I}(X) = \frac{1,16}{1,165 \cdot 10^{-3}} = 995 \frac{\text{дв. од.}}{\text{с}}$$

Пропускна здатність двійкового каналу:

$$C = \frac{1}{\tau} = 1000 \frac{\text{дв. од.}}{\text{с}}$$

Таким чином, у третьому випадку забезпечується найбільше наближення швидкості передачі інформації до пропускної здатності каналу.

Контрольні питання

1. Що розуміється під ефективністю системи передачі інформації і якими показниками вона характеризується?
2. Як залежить ефективність системи передачі інформації від надлишковості повідомлень?
3. У чому сутність методу підвищення ефективності систем за рахунок перерозподілу щільності ймовірностей елементів повідомлення?
4. Як впливає кореляція між елементами повідомлень на ефективність системи передачі інформації?
5. У чому сутність методу збільшення?
6. У чому сутність оптимального статистичного кодування?
7. Яким основним вимогам повинні відповідати оптимальні статистичні коди?
8. Пояснити методику побудови коду Шеннона-Фано.

ЗАВАДОСТІЙКІСТЬ ІНФОРМАЦІЙНИХ СИСТЕМ

11

11.1 Загальна характеристика завад у системах передачі інформації

Під **завадами** розуміють будь-які збурення в каналі передачі інформації, які призводять до випадкових відхилень прийнятого повідомлення від переданого. Завади зазвичай класифікуються за місцем їх виникнення, за статистичними властивостями і за характером впливу на корисний сигнал.

За характером виникнення завади можна поділити на зовнішні і внутрішні. До **зовнішніх завад** відносяться завади, джерела яких знаходяться ззовні системи передачі інформації. Сюди можна віднести:

- атмосферні завади (викликані грозовими розрядами);
- космічні завади викликані радіовипромінюванням Сонця та інших небесних тіл;
- промислові завади, обумовлені роботою різноманітних електричних пристроїв і агрегатів.

Внутрішні завади виникають у самій апаратурі системи передачі інформації. До них можна віднести завади у вигляді теплових шумів електронних ламп напівпровідникових приборів, опорів та інших елементів; завади, що викликані зміною параметрів ліній зв'язку, впливом ліній одна на іншу, а також за рахунок короточасних розривів зв'язку; завади, що виникають при перетворенні сигналів в окремих елементах системи (шуми квантування, спотворення сигналів за рахунок обмеженого значення смуги пропускання елементів, за рахунок нелінійності характеристик перетворення); завади, що обумовлені нестабільністю елементів апаратури, а також апаратурні спотворення, викликані технічною несправністю або недостатньо точним налагодженням апаратури.

За своїми властивостями завади можуть бути **детермінованими** і **випадковими**. Захист від детермінованих завад не викликає особливих труднощів.

Всі випадкові завади можна об'єднати в три групи: імпульсні; флуктуаційні та синусоїдальні.

Імпульсні завади являють собою послідовність імпульсів довільної форми з випадковими амплітудою, тривалістю і моментом появи.

Характерною особливістю імпульсних завад є те, що перехідні процеси, викликані в апаратурі яким-небудь імпульсом встигають практично затухнути до появи наступного імпульсу.

Характерними ознаками імпульсних завад є завади від грозових розрядів, від системи запалювання двигунів внутрішнього згорання, завади, пов'язані з комутаційними процесами тощо.

Флюктуаційні завади являють собою сукупність великої кількості короткочасних нерегулярних імпульсів з випадковими параметрами. Перехідні процеси від впливу окремих імпульсів, накладаючись один на одного, створюють неперервний випадковий процес. Характерною особливістю цих завад є відсутність викидів, які перевищують середній рівень більш ніж в три-чотири рази.

Оскільки тривалість перехідного процесу визначається смугою пропускання каналу передачі інформації, то і характер завад, залежить від ширини смуги каналу. Одна і та сама завада може бути імпульсною для широкосмугової і флюктуаційною для вузькосмугової системи.

Флюктуаційні завади являють собою білий шум, шум Гауса або білий шум Гауса. Останній характерний як своєю поширеністю, так і тим що він принципово не може бути усунений.

До таких завад можна віднести:

- теплові шуми опорів і напівпровідникових приборів, дробовий ефект електронних ламп;
- космічні завади;
- атмосферні завади в діапазоні коротких хвиль та ін.

Синусоїдальні завади являють собою синусоїдальні коливання з амплітудою, фазою, частотою, що інколи змінюються. Ці завади характеризуються повільною зміною параметрів, внаслідок чого ширина спектра модулюючої функції синусоїдальної завади виявляється практично малою у порівнянні зі смугою пропускання каналу.

Джерелами синусоїдальних завад можуть бути сторонні радіо-установки, генератори змінного струму і т. д.

За характером впливу на корисний сигнал завади поділяють на адитивні і мультиплікативні. **Адитивна завада** – це завада, яка представляється незалежним від сигналу випадковим доданком. Адитивну заваду інколи називають “шумом”. **Мультиплікативна завада** – це завада, що представляється незалежним від сигналу випадковим множителем. Більша частина завад, що зустрічаються на практиці, відносяться до групи адитивних завад. Характерним прикладом мультиплікативної завади є спотворення сигналу за рахунок випадкових змін характеристик каналу передачі інформації.

Всі випадкові завади являють собою випадковий процес і описуються за допомогою функцій розподілу ймовірностей або числових характеристик у вигляді моментів розподілу.

11.2 Критерії оцінювання завадостійкості інформаційних систем

Завадостійкість визначена в першому розділі як здатність інформаційної системи протистояти шкідливому впливу завад. У результаті дії завад прийняте повідомлення буде деякою мірою відрізнятися від переданого. Тому завадостійкість можна характеризувати як міру відповідності прийнятого повідомлення переданому при заданій заваді. При порівнянні декількох систем та з них буде більш завадостійкою, яка при однаковій заваді забезпечить меншу різницю між прийнятими і переданими повідомленнями.

Для характеристики міри відповідності прийнятого повідомлення переданому введений термін "достовірність". Кількісна міра цієї відповідності вибирається по-різному залежно від характеру повідомлень. При передачі неперервних повідомлень часто як критерій достовірності використовується критерій середнього квадратичного відхилення прийнятого повідомлення $Y(t)$ відносно переданого $X(t)$.

$$\sigma = \sqrt{|Y(t) - X(t)|^2}. \quad (11.1)$$

Застосовуються також критерій абсолютного відхилення:

$$\delta_{abc} = |Y(t) - X(t)| \quad (11.2)$$

і критерій найбільшого відхилення:

$$\delta_{\max} = \max |Y(t) - X(t)|. \quad (11.3)$$

Зазвичай для зручності порівняльної оцінки завадостійкості різних систем розглядаються відносні відхилення:

$$\gamma_{\sigma} = \frac{\sigma}{X_e}; \quad \gamma_{abc} = \frac{\delta_{abc}}{X_e}; \quad \gamma_{\max} = \frac{\delta_{\max}}{X_e}; \quad (11.4)$$

або приведені відхилення:

$$\alpha_{\sigma} = \frac{\sigma}{L_x}; \quad \alpha_{abc} = \frac{\delta_{abc}}{L_x}; \quad \alpha_{\max} = \frac{\delta_{\max}}{L_x}, \quad (11.5)$$

де X_e – ефективне значення повідомлення; L_x – динамічний діапазон повідомлень, що передаються.

Якщо припустити, що канал передачі інформації має ідеальну П-подібну амплітудно-частотну і лінійну фазочастотну характеристики,

то при наявності флуктуаційної завади типу білого гаусового шуму середнє квадратичне значення відхилення прийнятого повідомлення відносно переданого дорівнює кореню квадратному з середньої потужності завади на виході приймача, а відносна величина цього відхилення виражається через корінь квадратний відношення середніх потужностей завади і сигналу на виході приймача:

$$\gamma_{\sigma} = \sqrt{\left(\frac{P_{\xi}}{P_x}\right)_{\text{вих}}} \quad (11.6)$$

Для порівняльної оцінки систем і практичних розрахунків часто за критерій завадостійкості приймають величину "виграшу системи":

$$B = \frac{\left(\frac{P_x}{P_{\xi}}\right)_{\text{вих}}}{\left(\frac{P_x}{P_{\xi}}\right)_{\text{вх}}} \quad (11.7)$$

де $\left(\frac{P_x}{P_{\xi}}\right)_{\text{вих}}$, $\left(\frac{P_x}{P_{\xi}}\right)_{\text{вх}}$ – відношення середніх потужностей сигналу і завади на виході і вході пристрою відповідно.

У випадку передачі дискретних повідомлень, а також неперервних повідомлень з кодімппульсною модуляцією сигналів як критерій достовірності доцільно використовувати ймовірність правильного прийому:

$$P_{\text{пр}} = 1 - P_{\text{пом}}, \quad (11.8)$$

де $P_{\text{пом}}$ – ймовірність помилки у відтворенні повідомлення.

У реальних умовах ймовірність помилки $P_{\text{пом}}$ дуже мала і значно менше одиниці. Тому дуже часто для оцінювання завадостійкості використовують логарифмічну величину:

$$S = \lg \frac{1}{P_{\text{пом}}} = \lg \frac{1}{1 - P_{\text{пр}}} \quad (11.9)$$

За змістом визначення завадостійкість розуміється як власливість передачі інформації в цілому. Проте оцінювання завадостійкості системи в цілому – доволі складна задача. Тому говорять про завадо-

стійкість окремих ланок системи: про завадостійкість передачі (а саме, про завадостійкість коду або виду модуляції) і про завадостійкість прийому. Завадостійкість коду може бути оцінена величиною (11.9), де $P_{\text{пом}}$ – ймовірність спотворення кодової комбінації під дією завад.

При оцінюванні впливу виду модуляції на завадостійкість системи проводять порівняння різних видів модуляції з амплітудою. З цією метою часто застосовується коефіцієнт:

$$R_M = \frac{\left(\frac{X_e}{\sigma_\xi}\right)_{\text{вих}}}{\left(\frac{X_e}{\sigma_\xi}\right)_{\text{вихАМ}}} \quad (11.10)$$

або

$$R_M = \frac{\left(\frac{P_x}{P_\xi}\right)_{\text{вих}}}{\left(\frac{P_x}{P_\xi}\right)_{\text{вихАМ}}}, \quad (11.11)$$

де $\left(\frac{X_e}{\sigma_\xi}\right)_{\text{вих}}$, $\left(\frac{P_x}{P_\xi}\right)_{\text{вих}}$ – відношення ефективного значення сигналу до середнього квадратичного значення завади і відношення середньої потужності сигналу до середньої потужності завади відповідно на виході приймального пристрою при довільному виді модуляції;

$\left(\frac{X_e}{\sigma_\xi}\right)_{\text{вихАМ}}$ і $\left(\frac{P_x}{P_\xi}\right)_{\text{вихАМ}}$ – аналогічні відношення на виході приймального пристрою при амплітудній модуляції.

При прийомі залежно від призначення сигналів можуть мати місце три види задач: виявлення, розпізнавання та відтворення сигналів.

Задача виявлення полягає в тому, щоб за результатами обробки прийнятого сигналу, який може бути або тільки завадою, або сумою корисного сигналу і завади, вирішити, входить до складу прийнятого сигналу корисний чи не входить. При цьому можуть бути помилки двоякого роду: при відсутності корисного сигналу приймається помилкове рішення про наявність сигналу; при наявності сигналу приймається

помилкове рішення про відсутність сигналу. Перша помилка називається помилкою першого роду або “удаваною тривоого”. Друга помилка називається помилкою другого роду або “пропуском сигналу”. Кількісно помилки першого і другого родів оцінюються умовними ймовірностями α і β помилкових рішень про наявність корисного сигналу, якщо в дійсності він відсутній, і про відсутність сигналу, якщо насправді він є. Повна ймовірність помилкового рішення визначається виразом:

$$P_{\text{пом}} = q\alpha + p\beta, \quad (11.12)$$

де q і p – апіорні ймовірності відсутності і присутності корисного сигналу.

Завадостійкість приймача, який вирішує задачу виявлення сигналу, може бути оцінена за допомогою виразу (11.9), де $P_{\text{пом}}$ визначається формулою (11.12).

Коли апіорні відомо, що переданий один з двох сигналів x_1 і x_2 , то ставиться задача розпізнавання двох сигналів, тобто задача визначення, чи є на вході приймача сигнал x_1 “плюс” завада або сигнал x_2 “мінус” завада. Очевидно, завадостійкість приймача в цьому випадку також визначається виразами (11.9) і (11.12), причому величини α і β у формулі (11.12) є умовними ймовірностями помилкових рішень про наявність сигналів x_1 і x_2 відповідно, а q і p є апіорними ймовірностями надходження сигналів x_1 і x_2 .

Випадок розпізнавання багатьох сигналів у принциповому відношенні мало відрізняється від випадку розпізнавання двох сигналів. Завадостійкість приймача в цьому випадку також може бути оцінена формулою (11.9), причому ймовірність помилки $P_{\text{пом}}$ визначається за допомогою виразу:

$$P_{\text{пом}} = \sum_{i=1}^n p_i \alpha_i, \quad (11.13)$$

де p_i – апіорна ймовірність надходження i -го сигналу; α_i – умовна ймовірність помилкового рішення про наявність i -го сигналу, якщо насправді на вхід приймача надходить будь-який інший з n сигналів.

Задача відновлення повідомлення значно відрізняється від задач виявлення і розпізнавання сигналів. Вона зводиться до отримання вихідного повідомлення $Y(t)$, найменш відмінного від переданого з точки зору вибраного критерію достовірності, що передається з точки. Завадостійкість приймача в цьому випадку може бути оцінена за допомогою критеріїв відхилення, що визначаються виразами (11.1) – (11.6).

Завадостійкість такого приймача може бути також оцінена за допомогою критерію (11.7). Однак треба мати на увазі, що ця величина не завжди може бути визначена однозначно і не дозволяє об'єктивно порівнювати між собою різні системи, якщо завада має більш широкий спектр, ніж сигнал і, зокрема, якщо вона типу білого шуму. Потужність завади на вході приймача в цьому випадку визначається смугою пропускання каналу зв'язку. Різні системи можуть істотно відрізнятися смугою пропускання каналів зв'язку. При одному і тому ж каналі зв'язку можуть застосовуватися приймачі з різними смугами пропускання.

Для усунення неоднозначності можна порівнювати на вході і виході приймача відношення потужностей сигналу не з потужністю завади, а з питомою потужністю завад. У цьому випадку вираш приймача оцінюється співвідношенням:

$$B' = \frac{\frac{P_x}{P'_{\xi_{\text{вих}}}}}{\frac{P_x}{P'_{\xi_{\text{вх}}}}}, \quad (11.14)$$

де $P'_{\xi_{\text{вих}}} = \frac{P_{\xi_{\text{вих}}}}{F_{\text{вих}}}$ — питома потужність завад на виході приймача;

$P'_{\xi_{\text{вх}}} = \frac{P_{\xi_{\text{вх}}}}{F_{\text{вх}}}$ — питома потужність завади на вході приймача; $F_{\text{вх}}$ і $F_{\text{вих}}$ — смуги частот, в яких вимірюється потужність завади на вході і виході приймача.

Очевидно:

$$B' = \beta / \rho, \quad (11.15)$$

де $\rho = F_{\text{вх}} / F_{\text{вих}}$.

11.3 Засоби підвищення завадостійкості інформаційних систем

В основі всіх способів підвищення завадостійкості інформаційних систем лежить використання певних відмінностей між корисним сигналом і завадою. Тому для боротьби з завадами необхідні апріорні відомості про властивості завади і сигналу.

На сьогоднішній день відома велика кількість способів підвищення завадостійкості систем. Всі ці способи зручніше розбити на дві групи.

Перша група способів заснована на виборі методу передачі повідомлень. Друга група способів пов'язана з побудовою завадостійких приймачів.

Простим і часто застосовуваним способом підвищення завадостійкості передачі є збільшення відношення сигнал/завада за рахунок збільшення потужності передавача. Однак цей метод, незважаючи на свою простоту, може виявитися економічно не вигідним, оскільки пов'язаний з істотним зростанням складності і вартості обладнання. Крім того, збільшення потужності передачі супроводжується посиленням заважаючої дії даного каналу на інші.

Важливим способом підвищення завадостійкості передачі безперервних сигналів є раціональний вибір типу модуляції сигналів. Застосовуючи види модуляції, що забезпечують значне розширення смуги частот сигналу, можна досягти істотного підвищення завадостійкості передачі. Оцінка завадостійкості окремих видів модуляції зроблена в п. 11.4.

Радикальним способом підвищення завадостійкості передачі дискретних сигналів є використання спеціальних завадостійких кодів. При цьому є два шляхи підвищення завадостійкості кодів. Перший полягає у виборі таких способів передачі, які забезпечують меншу ймовірність спотворення коду, другий — у збільшенні коригувальних властивостей кодових комбінацій. Дослідження першого шляху показали високу завадостійкість кодів з великою кількістю частотних ознак.

Другий шлях пов'язаний з використанням кодів, що дозволяють виявляти і усувати спотворення в кодових комбінаціях. Такий спосіб кодування пов'язаний з введенням у код додаткових, надлишкових символів, що супроводжується збільшенням часу передачі або частоти передачі символів коду. Це приводить до розширення спектра сигналу.

Підвищення завадостійкості передачі може бути також досягнуте шляхом повторної передачі одного і того ж повідомлення. На приймальній стороні порівнюються отримані повідомлення і як істинні приймаються ті, які мають найбільшу кількість збігів. Щоб виключити невизначеність при обробці прийнятої інформації та забезпечити відбір за критерієм більшості, повідомлення повинно повторюватися не менш трьох разів. Очевидно, цей спосіб підвищення завадостійкості пов'язаний із збільшенням часу передачі.

Системи з повторенням передачі дискретної інформації поділяються на системи з груповим підсумовуванням, у яких порівняння проводиться за кодовими комбінаціями, і на системи з посимвольним підсумовуванням, у яких порівняння здійснюється за символами кодових комбінацій. Дослідження показали, що посимвольна перевірка є більш ефективною, ніж групова (див. приклад 11.2).

Різновидом систем, у яких підвищення завадостійкості досягається за рахунок збільшення часу передачі, є системи із зворотним зв'язком. При наявності спотворень у повідомленнях, що передаються,

інформація, що надходить по зворотному каналу, забезпечує повторення передачі. Наявність зворотного каналу приводить до ускладнення системи. Однак на відміну від систем з повторенням передачі в системах із зворотним зв'язком повторення передачі буде мати місце лише у випадку виявлення спотворень в сигналі, що передається, тобто надлишковість загалом виявляється меншою.

Характеристика систем із зворотним зв'язком наведена в п. 11.5.

Завадостійкий прийом полягає у використанні надлишковості, а також апріорних відомостей про сигнали і перешкоди для розв'язання оптимальним способом задачі прийому: виявлення сигналу, розрізнення сигналів або відновлення повідомлень. На сьогодні для синтезу оптимальних приймачів широко використовується апарат теорії статистичних рішень (див. п. 9.5–9.8).

Кількість помилок приймача зменшується із збільшенням відношення сигнал/завада на вході приймача. У зв'язку з цим часто проводять попередню обробку прийнятого сигналу з метою збільшення відношень корисної складової до перешкоди. До таких методів попередньої обробки сигналів відноситься метод ШОБ (поєднання ширококутового підсилювача, обмежувача і вузькокутового підсилювача), селекція сигналів за тривалістю, метод компенсації завади, метод фільтрації, кореляційний метод, метод накопичення та інші. Опис суті деяких з цих методів і їх характеристика наведені у розділі 9 (див. пп. 9.1–9.4).

11.4 Завадостійкість різних видів модуляції

Переносником інформації, як відомо, є сигнал, який в загальному випадку може бути представлений у вигляді:

$$x(t) = f(a_1, a_2, \dots, a_n, t),$$

де a_1, a_2, \dots, a_n – параметри сигналу.

Модуляція сигналу полягає в тому, що один або декілька його параметрів змінюються відповідно до повідомлення, що передається. Вплив завади на носій приводить до додаткової (паразитної) модуляції його параметрів.

При накладенні завади $\xi(t)$ на корисний сигнал $x(t)$ одержимо складний сигнал:

$$y(t) = x(t) + \xi(t) = f_1(a_1 + \delta a_1, a_2 + \delta a_2, \dots, a_n + \delta a_n, t),$$

де $\delta a_1, \delta a_2, \dots, \delta a_n$ – приріст параметрів сигналу під впливом завади (паразитна модуляція).

Очевидно, різні параметри сигналу будуть по-різному реагувати на вплив завади, тобто завадостійкість різних видів модуляції повинна бути неоднаковою. Оцінку завадостійкості модуляції будемо проводити за допомогою критеріїв (11.7) та (11.11).

11.4.1 Амплітудна і частотна модуляції

Передбачимо, що сигнал, що передається, і завада змінюються згідно з гармонічним законом:

$$x(t) = A \sin \omega_0 t; \quad \xi(t) = B \sin \omega t.$$

Сумарний сигнал, отриманий внаслідок накладання завади на корисний сигнал, можна представити у вигляді:

$$y(t) = A \sin \omega_0 t + B \sin \omega t = (A + \delta A)(\omega_0 + \delta \omega)t,$$

де δA і $\delta \omega$ – паразитна модуляція амплітуди і частоти сигналу.

На рис. 11.1 наведена векторна діаграма, що ілюструє утворення сумарного сигналу. На цьому малюнку \bar{X} – вектор сигналу, що обертається навколо точки 0 з частотою ω_0 ; \bar{Z} – вектор завади, що обертається навколо точки 1 з частотою ω ; \bar{Y} – вектор результуючого сигналу, що обертається навколо точки 0.

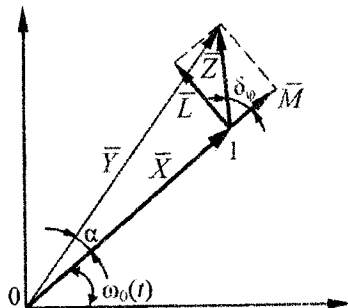


Рис. 11.1

Вектор \bar{Y} здійснює складний рух з швидкістю, що періодично змінюється. При цьому також буде періодично змінюватися його амплітуда. Періодичні зміни швидкості обертання вектора \bar{Y} і його амплітуди характеризують частотну і амплітудну паразитні модуляції сигналу.

Для оцінювання паразитної модуляції розкладемо вектор \bar{Y} на складові \bar{M} і \bar{L} (див. рис. 11.1).

Проекція вектора \bar{Z} на вектор \bar{X} виражає амплітудну паразитну модуляцію. Ця проекція:

$$M = B \cos(\omega_0 - \omega)t = B \cos \delta\varphi = \delta A \leq B, \quad (11.16)$$

$$\text{де } \delta\varphi = (\omega_0 - \omega)t = \delta_1 \omega t.$$

Щоб знайти паразитну частотну модуляцію, необхідно визначити кутову швидкість обертання вектора. Остання рівна:

$$\omega' = \frac{d\varphi}{dt},$$

де $\varphi = \omega_0 t + \alpha$ – миттєва фаза вектора \bar{Y} .

Таким чином,

$$\omega' = \omega_0 + \frac{d\alpha}{dt} = \omega_0 + \delta\omega,$$

де $\delta\omega = \frac{d\alpha}{dt}$ – паразитна частотна модуляція.

При $M \ll A$:

$$\alpha = \operatorname{tg} \alpha = \frac{B \sin \delta\varphi}{A}.$$

Отже,

$$\delta\omega = \frac{d\alpha}{dt} = \frac{B}{A} \cos \delta\varphi \frac{d(\delta\varphi)}{dt} = \frac{B}{A} \delta_1 \omega \cos \delta\varphi, \quad (11.17)$$

$$\text{де } \delta_1 \omega = \frac{d(\delta\varphi)}{dt}.$$

Остаточно отримаємо наступний вираз для кутової швидкості вектора:

$$\omega' = \omega_0 + \frac{B}{A} \delta_1 \omega \cos \delta\varphi. \quad (11.18)$$

Оцінимо величину відношення сигнал/завада на виході приймача при амплітудній модуляції. Оскільки на виході приймача стоїть

амплітудний детектор, то вихідний сигнал приймача пропорційний змінам параметра, що модулюється.

Отже, відношення потужностей сигналу і завади на виході приймача при амплітудній модуляції:

$$\gamma_{вихАМ} = \frac{(\Delta A)^2}{(\delta A)^2} = \frac{(\Delta A)^2}{M_{\max}^2},$$

де ΔA , M_{\max} – максимальні значення зміни амплітуди сигналу при корисній та паразитній модуляції відповідно.

При стовідсотковій амплітудній модуляції $\Delta A = A$.

Тоді

$$\gamma_{вих} = \frac{A^2}{B^2} = \gamma_{вихАМ},$$

де $\gamma_{вхАМ} = \frac{A^2}{B^2}$ – відношення потужності корисного сигналу до потужності завади на вході приймача.

Таким чином, вигравш амплітудної модуляції:

$$B_{АМ} = \frac{\gamma_{вихАМ}}{\gamma_{вхАМ}} = 1.$$

Отже, при амплітудній модуляції відношення сигнал/завада на виході і вході приймача однакові.

При частотній модуляції використовуються приймачі з частотними детекторами. При цьому величина сигналу на виході приймача буде пропорційна девіації частоти вхідного сигналу. Тоді відношення потужностей сигналу і завади на виході приймача:

$$\gamma_{вихЧМ} = \frac{(\Delta \omega)^2}{(\delta \omega)_{\max}^2} = \frac{A^2 (\Delta \omega)^2}{B^2 (\delta_1 \omega)^2} = \left(\frac{\Delta \omega}{\delta_1 \omega} \right)^2 \gamma_{вхЧМ},$$

де $\Delta \omega$ і $(\delta \omega)_{\max}$ – максимальні зміни частоти сигналу при корисній і паразитній частотній модуляції відповідно.

Отже, вигравш, що забезпечується частотною модуляцією:

$$B_{ЧМ} = \frac{\gamma_{вихЧМ}}{\gamma_{вхЧМ}} = \left(\frac{\Delta \omega}{\delta_1 \omega} \right)^2. \quad (11.19)$$

При $\Delta\omega > \delta\omega$ частотна модуляція заглушує заваду і це придушення буде тим ефективніше, чим більша девіація частоти при корисній модуляції. Таким чином, частотна модуляція у порівнянні з амплітудною забезпечує виграш в $\left(\frac{\Delta\omega}{\delta_1\omega}\right)^2$ раз.

Загальний випадок – по каналу передається довільний сигнал. Нехай сигнал характеризується енергетичною спектральною щільністю $G(\omega)$. У каналі діє адитивна завада типу білий шум з питомою енергетичною спектральною щільністю P_0 .

Відношення потужностей сигналу і завади на виході приймача для випадку амплітудної модуляції:

$$\gamma_{вихАМ} = \frac{\int_0^{\Omega} G(\omega) d\omega}{P_0 \Omega}, \quad (11.20)$$

де Ω – смуга частот, в якій діють сигнал і завада.

Це саме відношення для випадку частотної модуляції:

$$\begin{aligned} \gamma_{вихЧМ} &= \frac{\int_0^{\Omega} (\Delta\omega)^2 G(\omega) d\omega}{\int_0^{\Omega} (\delta_1\omega)^2 P_0 d(\delta_1\omega)} = \frac{3(\Delta\omega)^2 \int_0^{\Omega} G(\omega) d\omega}{P_0 \Omega^3} = \\ &= \frac{3(\Delta\omega)^2}{\Omega^2} \cdot \frac{\int_0^{\Omega} G(\omega) d\omega}{P_0 \Omega} = 3\beta^2 \gamma_{вихАМ}, \end{aligned} \quad (11.21)$$

де $\beta = \frac{\Delta\omega}{\Omega}$ – індекс частотної модуляції.

Таким чином, при довільному спектрі сигналу частотна модуляція у порівнянні з амплітудною забезпечує виграш у $3\beta^2$ раз. Цей виграш отримують ціною розширення спектра модульованого сигналу.

11.4.2 Кодоімпульсна модуляція

Кодоімпульсна модуляція (КІМ) широко використовується при передачі безперервних повідомлень. КІМ складається з операції квантування сигналів (за часом і рівнем) і кодування. Операції кванту-

вання супроводжуються похибками квантування (оцінка цих похибок зроблена у розділі 4). Зокрема, при квантуванні сигналів за рівнем, коли крок квантування Δ_x значно менше динамічного діапазону сигналу, дисперсія похибки квантування рівна:

$$D(\delta_x) = \frac{\Delta_x^2}{12}.$$

Нехай величина сигналу розподілена рівномірно в межах від 0 до A . Тоді дисперсія сигналу:

$$D(x) = \int_0^A x^2 \omega(x) dx = \frac{A^2}{3}.$$

Відношення потужностей сигналу і шуму квантування:

$$\gamma_{КІМ} = \frac{D(X)}{D(\delta_x)} = 4 \frac{A^2}{\Delta_x^2}. \quad (11.22)$$

Кількість рівнів квантування:

$$m = \frac{A}{\Delta_x}. \quad (11.23)$$

Якщо квантовані сигнали представляються двійковим кодом, то:

$$m = \frac{A}{\Delta_x} = 2^n,$$

де n – кількість розрядів в кодовій комбінації.

Тоді:

$$\gamma_{КІМ} = 4m^2 = 4 \cdot 2^{2n}. \quad (11.24)$$

Якщо сигнал, що модулюється має смугу частот $0 \dots F_c$, то відповідно до критерію В.А. Котельникова квантування сигналу за часом здійснюється з частотою $2F_c$. Нехай тривалість кодових комбінацій дорівнює інтервалу часового квантування сигналу $T_k = \frac{1}{2F_c}$. Тоді при

умові, що тривалість одного імпульсу в кодовій комбінації дорівнює тривалості паузи між імпульсами, отримаємо наступний вираз для ширини частотного спектра сигналу, що модулюється:

$$\Delta f_k \approx \frac{1}{\tau} = \frac{2n}{T_k} = 4F_c n,$$

звідки

$$n = \frac{\Delta f_k}{4F_c}. \quad (11.25)$$

Як видно з (11.25), число n виражає відношення спектра модульованого сигналу до спектра початкового немодульованого сигналу, тобто відіграє таку ж роль, як і індекс модуляції β при частотній модуляції.

Отже, величина $\gamma_{\text{КІМ}}$, що визначає завадостійкість КІМ, пропорційна 2^β , у той час як завадостійкість частотної модуляції пропорційна β^2 .

Висока завадостійкість КІМ досягнута як за рахунок розширення спектра сигналу, що модулюється, так і за рахунок збільшення часу передачі окремих повідомлень.

Незначне розширення спектра модульованого сигналу (збільшення n) супроводжується різким збільшенням завадостійкості КІМ. Однак, як видно з виразу (11.23), зростання n (при заданому A) може бути здійснене за рахунок істотного зменшення кроку квантування Δ_x . Останнє, в свою чергу, може призвести до пониження завадостійкості КІМ за рахунок впливу зовнішніх завад.

Дійсно, нехай при квантуванні діє адитивна завада $\xi(x)$. Дія завади $\xi(x)$ приводить до додаткових похибок при квантуванні сигналу, якщо рівень завад перевищує половину кроку квантування. Ймовірність такої похибки:

$$P_{\text{ном}} = \int_{\frac{\Delta_x}{2}}^{\infty} \omega(\xi) d\xi + \int_{-\infty}^{-\frac{\Delta_x}{2}} \omega(\xi) d\xi.$$

При симетричному розподілі завад:

$$P_{\text{ном}} = 2 \int_{\frac{\Delta_x}{2}}^{\infty} \omega(\xi) d\xi.$$

При нормальному законі розподілу завад:

$$P_{\text{ном}} = \frac{2}{\sqrt{2\pi\sigma_\xi^2}} \int_{\frac{\Delta_x}{2}}^{\infty} \exp\left\{-\frac{\xi^2}{2\sigma_\xi^2}\right\} d\xi = 1 - \Phi\left(\frac{\Delta_x}{2\sigma_\xi}\right), \quad (11.26)$$

де $\Phi\left(\frac{\Delta_x}{2\sigma_\xi}\right)$ – функція Лапласа; σ_ξ – середнє квадратичне значення завади.

На рис. 11.2 показані графіки зміни функції Лапласа та ймовірності помилок залежно від відношення $\frac{\Delta_x}{2\sigma_\xi}$. Як видно, при збільшенні

кроку квантування Δ_x , відбувається різке зменшення ймовірності помилок за рахунок дії зовнішніх завад.

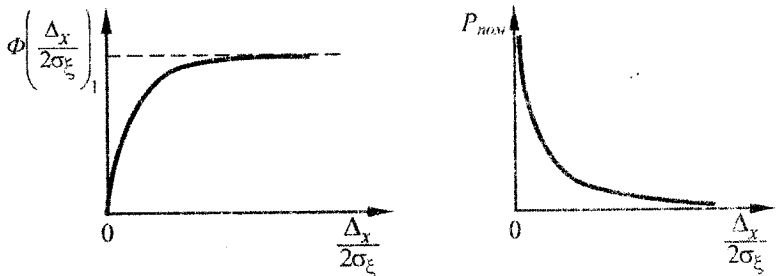


Рис. 11.2

При передачі кодового сигналу можуть виникнути додаткові похибки внаслідок спетворення кодових комбінацій за рахунок дії завад в каналі зв'язку. Ці похибки можуть мати лише дискретні значення:

$$\xi_{mi} = \pm \Delta_x, \pm 2\Delta_x, \dots, \pm (m-1)\Delta_x,$$

тобто

$$\xi_{mi} = \pm i\Delta_x,$$

при цьому: $-(m-1) \leq i \leq (m-1)$.

Дисперсія похибки для випадку симетричного розподілу завад:

$$D(\xi_{mi}) = 2 \sum_{i=1}^{m-1} (i\Delta_x)^2 P_{mi}, \quad (11.27)$$

де P_{mi} – ймовірність того, що похибка за рахунок дії зовнішніх завад дорівнює $i\Delta_x$.

11.5 Системи із зворотним зв'язком

Одним з ефективних способів підвищення завадостійкості передачі є використання зворотного каналу, по якому інформація може передаватися від приймача до передавача.

Системи, в яких застосовується передача по зворотному каналу, носять загальну назву систем із зворотним зв'язком.

Існує велика кількість способів побудови зворотного каналу. Основні варіанти схем зворотного зв'язку показані на рис. 11.3. У першому варіанті зворотній зв'язок охоплює тільки лінію зв'язку. У варіантах II та III зворотний зв'язок підключений після розв'язувального пристрою. Варіант III відрізняється тим, що зворотний зв'язок охоплює всю систему.

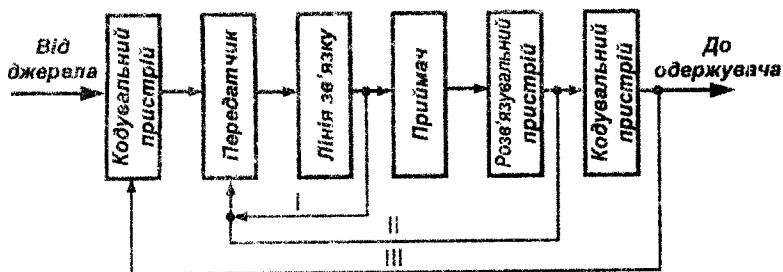


Рис. 11.3

Залежно від способу використання зворотного каналу системи можуть бути поділені на два основні типи: з інформаційним зворотним зв'язком (системи з порівнянням) і з розв'язувальним зворотним зв'язком (системи з перениткуванням).

У системах з інформаційним зворотним зв'язком по зворотному каналу передаються всі прийняті сигнали. На передавальній стороні ці сигнали порівнюються з переданими. У випадку наявності розбіжностей здійснюється повторна передача сигналу або передача даних про необхідні виправлення.

У системах з розв'язувальним зворотним зв'язком приймач сам перевіряє правильність сигналу, що приймається. При виявленні спотворень приймач посилає по зворотному каналу сигнал перепитивання на повторення передачі сигналу, що спотворюється.

Таким чином, в системах з інформаційним зворотним зв'язком рішення про необхідність повторення передачі приймається на переда-

вальній стороні, а в системах з розв'язувальним зворотним зв'язком на приймальній стороні.

Зворотні канали в системах з інформаційним зворотним зв'язком і в системах з розв'язувальним зворотним зв'язком використовуються не в однаковій мірі. Зворотний канал в системі з розв'язувальним зворотним зв'язком завантажений значно менше, ніж в системах з інформаційним зворотним зв'язком, тому що в системах з розв'язувальним зворотним зв'язком сигнал по зворотному каналу надходить лише у випадку виявлення спотворень у переданому сигналі, в той час як в системах з інформаційним зворотним зв'язком кожний прийнятий сигнал пересилається по зворотному каналу на передавальну сторону.

Структура сигналів у системах з розв'язувальним зворотним зв'язком повинна бути такою, щоб помилка при передачі могла бути виявлена на приймальній стороні. Це може бути здійснено лише при використанні кодів, що дозволяють виявляти спотворення. Дослідження показали, що системи з розв'язувальним зворотним зв'язком, які використовують коди з виявленням помилок, ефективніші за системи без зворотного зв'язку, що використовують коди з виправленням помилок. Це пояснюється тим, що в системах з розв'язувальним зворотним зв'язком повторення передачі здійснюється лише при виявленні помилок у переданому сигналі, в той час як у системах з автоматичним коригуванням помилок на приймальній стороні незалежно від наявності спотворень у сигналах постійно вводиться надлишковість, необхідна для виправлення помилок.

11.6 Приклади

Приклад 11.1

Повідомлення передаються примітивним двійковим кодом при впливі завад. Довжина кодових комбінацій $n = 7$. Ймовірність спотворення символів кодових комбінацій під впливом перешкод $p_e = 10^{-3}$. Спотворення символів незалежні. Оцінити завадостійкість передачі при трикратній передачі з груповою перевіркою.

Розв'язування. При трикратній передачі з груповою перевіркою, згідно з критерієм більшості, умови правильного прийому будуть наступними:

а) всі три кодові посилки прийняті правильно; ймовірність такої події дорівнює $(1 - p_e)^{3n}$.

б) з трьох посилок дві прийняті правильно; ймовірність такої події рівна $C_3^1 (1 - p_e)^{2n} [1 - (1 - p_e)^n]$, де C_3^1 — кількість випадків, коли з трьох кодових комбінацій одна прийнята з помилкою; $(1 - p_e)^{2n}$ —

ймовірність правильного прийому двох посилок; $[1 - (1 - p_e)^n]$ – ймовірність неправильного прийому однієї посліжки.

Таким чином, ймовірність помилкового прийому при трикратному повторенні з груповим контролем:

$$\begin{aligned} P_{e,3} &= 1 - \left\{ (1 - p_e)^{3n} + C_3^1 (1 - p_e)^{2n} [1 - (1 - p_e)^n] \right\} = \\ &= 1 - (1 - p_e)^{2n} \left\{ (1 - p_e)^n + C_3^1 [1 - (1 - p_e)^n] \right\} \approx \\ &\approx 1 - [(1 - 2np_e)(1 + 2np_e)] = 1 - (1 - 4n^2 p_e^2) = \\ &= 4n^2 p_e^2 = 4 \cdot 7^2 \cdot 10^{-6} = 2 \cdot 10^{-4}. \end{aligned}$$

Завадостійкість передачі:

$$S = \lg \frac{1}{P_{e,3}} = \lg \frac{1}{2 \cdot 10^{-4}} = 3,7.$$

Приклад 11.2

Вирішити попередній приклад для випадку трикратної передачі з посимвольною перевіркою.

Розв'язування. При трикратній передачі і посимвольному контролі можливі наступні варіанти правильного прийому одного визначеного символу кодової комбінації:

а) символ прийнятий правильно в усіх трьох посліжках, ймовірність такої події дорівнює $(1 - p_e)^3$;

б) символ прийнято правильно в двох з трьох посилок, ймовірність такої події дорівнює $C_3^1 (1 - p_e)^2 p_e$.

Отже, ймовірність правильного прийому символу кодової комбінації при трьох посліжках буде дорівнювати:

$$(1 - p_e)^3 + C_3^1 (1 - p_e)^2 p_e.$$

Ймовірність правильного прийому n -розрядної кодової комбінації становить:

$$[(1 - p_e)^3 + C_3^1 (1 - p_e)^2 p_e]^n.$$

Отже, ймовірність помилкового прийому:

$$P_{e,3} = 1 - [(1 - p_e)^3 + C_3^1 (1 - p_e)^2 p_e]^n \approx 1 - (1 - 6np_e^2)^n \approx 6np_e^2 = 42 \cdot 10^{-6}.$$

Завадостійкість передачі:

$$S = \lg \frac{1}{P_{н.о.}} = \lg \frac{1}{42 \cdot 10^{-6}} = 4,38.$$

Отже, як видно з розрахунків, при повторній передачі посимвольна перевірка більш ефективна, ніж групова.

Приклад 11.3

Частотно-модульований сигнал з індексом частотної модуляції β та енергетичною спектральною щільністю $G(\omega) = G_0 - b\omega$ проходить через інформаційний канал зі смугою пропускання Ω . У каналі діє завада типу адитивного білого шуму з питомою енергетичною спектральною щільністю P_0 . Оцінити відношення потужності корисного сигналу до потужності завади на виході приймача.

Розв'язування. Відношення потужностей корисного сигналу і завади на виході приймача для випадку частотної модуляції згідно з виразом (11.21) дорівнює:

$$\gamma_{вихУМ} = 3\beta^2 \frac{\int_0^{\Omega} G(\omega) d\omega}{P_0 \Omega} = 3\beta^2 \frac{\int_0^{\Omega} (G_0 - b\omega) d\omega}{P_0 \Omega} = 3 \frac{\beta^2}{P_0} \left(G_0 - \frac{b\Omega}{2} \right).$$

Приклад 11.4

Сигнал, який ізолюється в межах від 0 до 100 Б, рівномірно квантується за рівнем з кроком квантування $\Delta_x = 1$ В і передається по каналу за допомогою примітивного двійкового коду. У каналі діють симетричні шумові завади. Під дією завад відбуваються спотворення кодових комбінацій. Спотворення кодових комбінацій однократні з ймовірністю $p_e = 10^{-3}$. Оцінити дисперсію помилок, які виникають внаслідок дії завад, а також дисперсію сумарної помилки за рахунок дії завад і квантування сигналів за рівнем.

Розв'язування. Помилки, які обумовлені спотворенням кодових комбінацій внаслідок дії шумових завад, носять дискретний характер. Оскільки помилки можуть бути тільки однократними, то їх величина може приймати значення $\epsilon_{ном_i} = 2^i \Delta_x$, $i = 0, 1, \dots, n-1$, де n — довжина кодових комбінацій.

Дисперсія помилок для випадку симетричного розподілу завад згідно з виразом (11.27) дорівнює:

$$D(\varepsilon_{\text{ном}}) = 2 \sum_{i=0}^{n-1} (2^i \Delta_x)^2 p_e.$$

Кількість повідомлень, що передаються, визначається кількістю рівнів квантування сигналу і дорівнює:

$$N = \frac{100}{\Delta_x} = 100.$$

Виходячи з співвідношення $N = 2^n$, визначаємо довжину кодових комбінацій:

$$n = \frac{\lg 100}{\lg 2} = 6,6.$$

Якщо округлити до найближчого цілого, то отримаємо $n = 7$. Тоді дисперсія помилки:

$$D(\varepsilon_{\text{ном}}) = 2(1 + 2^2 + 2^4 + 2^6 + 2^8 + 2^{10} + 2^{12}) \cdot 10^{-3}.$$

Якщо розглядати суму в дужках як геометричну прогресію з знаменником 2^2 і використовувати формулу для суми членів геометричної прогресії, отримуємо:

$$D(\varepsilon_{\text{ном}}) = 2 \frac{2^{14} - 1}{2^2 - 1} 10^{-3} = 10,9 \text{ В}^2.$$

Дисперсія сумарної помилки за рахунок дії завад в каналі і квантування сигналів за рівнем:

$$D_{\text{ном}\Sigma} = D(\varepsilon_{\text{ном}}) + \sigma_k^2,$$

де σ_k – середнє квадратичне значення помилки квантування за рівнем.

При рівномірному квантуванні за рівнем середнє квадратичне значення помилки квантування:

$$\sigma_k = \frac{\Delta_x}{2\sqrt{3}} = \frac{1}{2\sqrt{3}} = 0,29 \text{ В}.$$

Таким чином, дисперсія сумарної помилки:

$$D_{\text{ном}\Sigma} = 10,9 + 0,29^2 \approx 10,98 \text{ В}^2.$$

Контрольні питання

1. Які є критерії для оцінки завадостійкості систем передачі інформації з неперервними сигналами?
2. Які є критерії для оцінки завадостійкості систем передачі інформації з дискретними сигналами?
3. Перерахувати відомі способи підвищення завадостійкості передачі.
4. Що таке системи з повторенням передачі? Які відомі різновиди цих систем?
5. Що розуміється під системою із зворотним зв'язком?
6. Що розуміється під системою із інформаційним зворотним зв'язком?
7. Що розуміється під системою з розв'язувальним зворотним зв'язком?
8. Порівняйте ступінь завантаженості зворотного каналу в системах з інформаційним і розв'язувальним зворотним зв'язком.
9. Порівняйте ефективність систем з розв'язувальним зворотним зв'язком, які використовують коди з виявленням помилки, з системами без зворотного зв'язку, які використовують коди з виправленням помилок.
10. Порівняйте за завадостійкістю амплітудну, частотну і кодоімпульсну модуляції.
11. Якими факторами визначається висока завадостійкість кодоімпульсної модуляції?

ІНФОРМАЦІЙНА ОЦІНКА АВТОМАТИЗОВАНИХ СИСТЕМ КОНТРОЛЮ ТА УПРАВЛІННЯ

12

12.1 Загальна характеристика автоматизованих систем контролю і управління

Останнім часом для отримання інформації про положення і стан об'єктів у просторі і часі все більш широкое використання знаходять складні системи контролю і управління. Окремі, порівняно прості автомати на базі обчислювальної техніки об'єднуються в більш автоматизовані системи контролю і управління.

До таких систем, в першу чергу, слід віднести автоматизовані системи стабілізації літальних апаратів, технологічні і організаційно-економічні автоматизовані системи управління енергомережею країни, цехами, підприємствами, фірмами, галузями народного господарства, країною в цілому. Всі ці системи не тільки мають багато складних пристроїв і органів, але і високий рівень організації, складні функціональні взаємозв'язки пристроїв і органів.

Для того щоб правильно судити про хід процесу контролю і управління, а також про стан і положення об'єкта, який контролюється і управляється, необхідно контролювати більше число параметрів різної фізичної природи. Технічні процеси контролю і управління складними об'єктами тісно пов'язані між собою і не завжди контроль і управління ними можна доручити декільком діючим незалежно операторам і автоматам. Тому велику кількість вимірювальних і управляючих пристроїв доводиться сконцентрувати в одному місці. Використання централізованих автоматизованих систем контролю і управління (АСКУ) позбавляє необхідності користування показниками численних пристроїв і транспарантів.

Основне призначення АСКУ – правильно визначити протягом заданого інтервалу часу стан і положення об'єкта та управляти ним з урахуванням стану. У випадку появи відхилення або неполадок АСКУ повинна забезпечувати виявлення відхилення або відмови місця неполадки з заданою точністю.

Все частіше останнім часом АСКУ використовується також для прогнозування працездатності і положення об'єкта в цілому або його функціональних елементів. У таких випадках вона повинна передбачати поведінку параметрів або об'єкта в цілому на заданий час вперед.

АСКУ дозволить крім ефективного контролю і управління вести оптимізацію цих процесів, а також накопичувати інформацію для прогнозування відмов системи контролю і управління об'єктом.

Процес контролю і управління складними об'єктами можна розглядати як процес формування і регулювання фізично незалежних або взаємопов'язаних процесів, які відбуваються одночасно або з деяким зсувом у часі. Технічну систему контролю і управління складними об'єктами необхідно в цьому випадку розглядати як багатовимірну систему.

За цільовим призначенням АСКУ можна розділити на системи: стабілізації, програмного управління, слідкуючі, оптимальні і змішані.

Багатовимірні АСКУ стабілізації призначені для одночасної підтримки з визначеною точністю багатьох вихідних параметрів. Багатовимірні АСКУ програмного управління призначені для автоматичного управління декількома фізичними процесами за заданою програмою з визначеною точністю. Багатовимірні слідкуючі АСКУ призначені для одночасного відтворення всіма вихідними параметрами невідомих наперед різних законів управління. Оптимальні багатовимірні АСКУ мають оптимізатор, який дозволяє отримувати оптимальні значення вихідних параметрів. Багатовимірні змішані АСКУ можуть одночасно слідкувати, стабілізувати, управляти за програмою і оптимізувати множину параметрів.

Багатовимірні АСКУ – це складні інформаційні системи, які характеризуються інтенсивними потоками інформації.

У зв'язку з цим становить інтерес проведення оцінювання стану і функціонування АСКУ з використанням положень теорії інформації.

12.2 Оцінка ступеня невизначеності стану об'єкта контролю

При контролі і управлінні об'єкт можна розглядати як багатовимірну динамічну систему, на яку водночас із закономірними і випадковими управляючими впливами або сигналами контролю впливають різні випадкові завади у вигляді зовнішніх і внутрішніх збурень. Стан такої системи визначається деякими вихідними параметрами, певним чином пов'язаними з впливами на систему через вектор-оператор системи.

У зв'язку з випадковим характером різних впливів і збурень вихідні параметри об'єкта будуть також випадковими функціями часу.

Повними ймовірнісними характеристиками як вихідних параметрів, так і самого об'єкта, є диференціальні багатовимірні закони розподілу, а також рівняння для визначеності ймовірності стану виходів системи при різних ймовірностях стану вхідних сигналів. Однак вони не дають інтегральної якісної і кількісної оцінки невизначеності об'єкта при конт-

ролі і управління, а також інтегральної оцінки зміни невизначеності об'єкта в процесі контролю і управління. Цей недолік усувається при використанні інформаційних характеристик об'єкта в процесі контролю і управління.

Для інтегральної оцінки невизначеності об'єкта з неперервною множиною станів у процесі контролю і управління зручно застосовувати диференціальну ентропію стану об'єкта:

$$h(X) = - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX, \quad (12.1)$$

де $X(x_1, x_2, \dots, x_m)$ – вектор вихідних параметрів (координат) об'єкта.

При незалежних вихідних координатах ентропія об'єкта дорівнює сумі часткових ентропій, обумовлених невизначеністю окремих координат:

$$h(X) = \sum_{i=1}^m h(x_i). \quad (12.2)$$

Ентропія об'єкта з двома дискретними станами (справним і несправним) може бути представлена у вигляді:

$$H(X) = -\{P_{o.c.} \log_2 P_{o.c.} + (1 - P_{o.c.}) \log_2 (1 - P_{o.c.})\}, \quad (12.3)$$

де $P_{o.c.}$ – ймовірність справного стану об'єкта.

Ймовірність справного стану об'єкта пов'язана з багатовимірною щільністю розподілу ймовірності $\omega(X)$ співвідношенням:

$$P_{o.c.} = \int \dots \int_{V_x} \omega(X) dX, \quad (12.4)$$

де V_x – область допустимих значень вектора X .

Випадковий розкид параметрів об'єкта обумовлено великою кількістю факторів технологічного і експлуатаційного характеру. Серед цих факторів важко виділити переважаючий. У такій ситуації, згідно з центральною граничною теоремою Ляпунова, закон розподілу параметрів дуже близький до нормального. Спираючись на це положення, закон розподілу вихідних параметрів $\omega(x_i)$ у більшості випадків можна прийняти нормальним:

$$\omega(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ -\frac{(x_i - a_i)^2}{2\sigma_i^2} \right\}, \quad (12.5)$$

де a_i , σ_i – математичне сподівання і середнє квадратичне відхилення параметра.

Тоді багатовимірна щільність розподілу вектора \vec{X} :

$$\omega(X) = \frac{1}{\sigma_1 \sigma_2 \dots \sigma_m \sqrt{(2\pi)^m D}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m \sum_{k=1}^m D_{ik} \frac{x_i - a_i}{\sigma_i} \frac{x_k - a_k}{\sigma_k} \right\}, \quad (11.6)$$

$$\text{де } D = \begin{vmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{vmatrix} \text{ -- визначник } m\text{-го порядку коефіцієнта}$$

кореляції; D_{ik} -- алгебраїчне доповнення елемента r_{ik} у визначнику D .

При статистичній незалежності координат x_i вираз (12.6) приводиться до вигляду:

$$\omega(X) = \frac{1}{\sigma_1 \sigma_2 \dots \sigma_m \sqrt{(2\pi)^m}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^m \frac{(x_k - a_k)^2}{\sigma_k^2} \right\}. \quad (12.7)$$

Ентропія вихідних параметрів при нормальному законі розподілу останніх:

$$h(x_i) = \log_2 \sigma_i \sqrt{2\pi e},$$

звідки середнє квадратичне відхилення параметра:

$$\sigma_i = \frac{2^{h(x_i)}}{\sqrt{2\pi e}}. \quad (12.8)$$

При статистичній незалежності вихідних параметрів об'єктів:

$$P_{o.c} = \prod_{i=1}^m P_{ci}, \quad (12.9)$$

де P_{ci} -- ймовірність знаходження у межах допуску i -го вихідного параметра.

Тоді з урахуванням виразів (12.5), (12.8) і (12.9) отримаємо:

$$P_{c.o} = 2^{\sum_{i=1}^m h(x_i)} \sqrt{e^m} \int_{V_x} \dots \int \exp \left\{ -\pi e \sum_{i=1}^m \frac{(x_i - m_i)^2}{2^{h(x_i)}} \right\} dx_1 dx_2 \dots dx_m. \quad (12.10)$$

Вираз (12.10) показує залежність ймовірності справного стану об'єкта від ентропії вихідних параметрів. При наявності статистичних залежностей між вихідними параметрами даний вираз виявляється достатньо складним і тут не наводиться.

12.3 Інформаційна оцінка точності результатів контролю

Контроль – це один з різновидів процесу отримання перетворення, передачі і накопичення інформації. Під контролем в усіх випадках практики розуміється процес отримання людиною або машиною інформації про дійсний стан об'єкта контролю.

Інформація, отримана в результаті контролю, кількісно оцінюється зменшенням ентропії від значення $h(X)$, яке характеризує невизначеність об'єкта контролю перед контролем, до значення $h(X/Y)$, яке залишається після отримання результату контролю Y , тобто:

$$I = h(X) - h(X/Y). \quad (12.11)$$

Апостеріорна ентропія $h(X/Y)$ визначається розподілом ймовірностей похибок контролю. Оскільки апостеріорна ентропія $h(X/Y)$ обумовлює зменшення кількості інформації, отриманої при контролі, то говорять, що похибки контролю здійснюють дезінформаційну дію.

Дезінформаційна дія похибки у загальному випадку визначається законом розподілу останньої, причому при різних законах розподілу отримують різні значення $h(X/Y)$ навіть у тому випадку, якщо дисперсії похибок однакові.

Так, якщо похибка розподілена рівномірно в межах від $+\Delta_z$ до $-\Delta_z$, то ентропія похибки (див. розділ 7, приклад 7.4):

$$h(X/Y) = \log_2 2\Delta_z. \quad (12.12)$$

Дисперсія похибки з рівномірним розподілом:

$$\sigma_z^2 = \int_{-\Delta_z}^{\Delta_z} z^2 \frac{1}{2\Delta_z} dz = \frac{\Delta_z^2}{3}. \quad (12.13)$$

Підставивши значення Δ_z із виразу (12.13) у (12.12), отримаємо:

$$h(z) = h(X/Y) = \log_2 2\sqrt{3\sigma_z^2}. \quad (12.14)$$

Якщо похибка розподілена за нормальним законом, то ентропія, як відомо (див. розділ 7, приклад 7.3), становить:

$$h(z) = h(X/Y) = \log_2 \sqrt{2\pi e\sigma_z^2}. \quad (12.15)$$

Із порівнянь виразів (12.14) і (12.15) видно, що при рівних σ_z дезінформаційна дія нормально розподілених похибок більша, ніж рівно-

мірно розподілених. У зв'язку з цим вказано на недосконалість характеристик точності контролю використання і введено поняття ентропійного значення похибки контролю. Як ентропійне значення похибки прийнято значення похибки з рівномірним законом розподілу, яке вносить таку ж дезінформаційну дію, що і похибка з даним законом розподілу ймовірностей.

Ентропійне значення похибки Δ_e з довільним законом розподілу $\omega(z)$ може бути отримано із рівності:

$$\log_2 2\Delta_e = - \int_{-\infty}^{\infty} \omega(z) \log_2 \omega(z) dz = h(z),$$

звідки

$$\Delta_e = \frac{1}{2} 2^{h(z)} = 2^{h(z)-1}. \quad (12.16)$$

Ентропійне значення похибки пов'язано з середнім квадратичним значенням співвідношення:

$$\Delta_e = k_e \sigma_z, \quad (12.17)$$

де k_e – ентропійний коефіцієнт, величина якого визначається законом розподілу похибок.

Найбільшу ентропію при заданому значенні дисперсії має нормальний розподіл, тому для нього ентропійний коефіцієнт k_e найбільший і дорівнює 2,07. Будь-який інший розподіл має менше значення коефіцієнта k_e . Зокрема, для рівномірного розподілу $k_e = 1,73$.

Потрібно зауважити, що викладене відноситься до випадку, коли контроль параметрів здійснюється за критерієм кількісної оцінки, тобто по суті зводиться тільки до вимірювання величини параметрів. У тих випадках, коли контроль здійснюється за допусковим критерієм, тобто коли ціллю контролю є встановлення, знаходиться або не знаходиться величина параметра, що контролюється, в межах допуску, ентропія похибки контролю повинна визначатися співвідношенням:

$$H(X/Y) = - \{ p_{ном} \log_2 p_{ном} + (1 - p_{ном}) \log_2 (1 - p_{ном}) \}, \quad (12.18)$$

де $p_{ном}$ – загальна безумовна ймовірність помилкового розв'язку відносно стану контрольованого параметра, що визначається виразом (9.28).

Помилки першого і другого роду, які входять у вираз (9.28), можна визначити за формулами:

$$\alpha = \int_{-\infty}^{x_1} \omega(x) \left[\int_{x_1-x}^{x_2-x} \omega(\xi) d\xi \right] dx + \int_{x_2}^{\infty} \omega(x) \left[\int_{x_1-x}^{x_2-x} \omega(\xi) d\xi \right] dx;$$

$$\beta = \int_{x_1}^{x_2} \omega(x) \left[\int_{-\infty}^{x_1-x} \omega(\xi) d\xi + \int_{x_2-x}^{\infty} \omega(\xi) d\xi \right] d\xi dx,$$

де x_1 і x_2 – допуски на контрольований параметр; $\omega(x)$ і $\omega(\xi)$ – закони розподілу ймовірностей контрольованого параметра x та похибок ξ контролю.

Очевидно, і в цьому випадку може бути введено ентропійне значення похибки контролю, яке за дезінформаційною дією еквівалентне похибці $P_{ном}$. Причому, величина цієї похибки може бути визначена за формулою (12.16), якщо в показник степеня числа 2 підставити ентропію, яка визначається виразом (12.18).

12.4 Інформаційна здатність пристрою контролю та ККД процесу контролю

Під інформаційною здатністю пристрою контролю можна розуміти еквівалентне число різних інтервалів, яке визначає кількість інформації, отримуваної від пристрою контролю.

Таким чином, згідно з визначенням інформаційна здатність пристрою контролю N може бути визначена із співвідношення:

$$I = \log_2 N, \quad (12.19)$$

де I – кількість інформації, отримуваної при контролі.

Із формули (12.19) отримаємо вираз для інформаційної здатності:

$$N = 2^I. \quad (12.20)$$

Якщо у всьому діапазоні вимірювань ентропійна похибка постійна, то кількість інформації I визначається формулою (12.11). Якщо ж Δ_c змінна, то в формулу (12.11) як другий член повинно бути підставлено середньозважене згідно з розподілом щільності $\omega(x)$ значення умовної ентропії:

$$h_{сер}(x/y) = - \int_{-\infty}^{\infty} \omega(x) h(x/y) dx. \quad (12.21)$$

У реальних умовах контрольований параметр x зазнає різних зовнішніх випадкових впливів. Флуктуації параметра x , які з'являються в результаті цього, характеризують звичайно середнім квадратичним значенням σ_ξ .

За аналогією з ентропійною похибкою пристрою контролю можна користуватися поняттям ентропійної похибки параметра:

$$\Delta_\xi = k_\xi \sigma_\xi, \quad (12.22)$$

де k_ξ – ентропійний коефіцієнт, величина якого визначається законом розподілу флуктуацій $\omega(\xi)$.

Отже, можна стверджувати, що контрольований параметр x , який надходить на вхід пристрою контролю, несе кількість інформації:

$$I_x = h(x) = - \int_{-\infty}^{\infty} \omega(\xi) \log_2 \omega(\xi) d\xi \quad (12.23)$$

і характеризується інформаційною здатністю:

$$N_x = 2^{I_x} \quad (12.24)$$

Внаслідок того, що пристрій контролю має похибки, кількість інформації I_1 , отримувана на виході пристрою контролю, менше інформації I_x , яку несе параметр x . Внаслідок цього процес контролю супроводжується втратою інформації:

$$\Delta I = I_x - I_1. \quad (12.25)$$

Величина ΔI є інформаційною оцінкою удосконалення процесу контролю. Звичайно користуються відносною величиною:

$$\eta_i = \frac{I_1}{I_x} = \frac{\log_2 N_1}{\log_2 N_x}, \quad (12.26)$$

яка отримала назву інформаційного ККД процесу контролю.

У формулах (12.25) і (12.26) кількість інформації I_1 , реально отримувана в результаті контролю, визначається сумарною похибкою контрольованого параметра і пристрою контролю. Однак у переважній більшості практичних випадків похибка, яка вноситься пристроєм контролю, суттєво перевищує флуктуації параметра, що контролюється. Тому в сумарній похибці доданком, обумовленим флуктуаціями контрольованої величини, можна знехтувати зі збереженням дуже високого ступеня точності.

12.5 Пропускна здатність процесу контролю

Пропускна здатність процесу контролю V_k є динамічною характеристикою і виражає максимально можливу кількість інформації, яку можна отримати у процесі контролю за одиницю часу:

$$V_k = \frac{I_{\max}}{\bar{T}_k}, \quad (12.27)$$

де I_{\max} – максимальна кількість інформації, яку можна отримати в середньому при контролі одного об'єкта; \bar{T}_k – математичне сподівання часу, який витрачається на контроль одного об'єкта.

Якщо об'єкт контролю має один параметр або декілька параметрів, які контролюються паралельно, то \bar{T}_k визначається часом контролю одного параметра. При послідовному контролі декількох параметрів об'єкта контроль припиняється після виявлення першої відмови.

У зв'язку з цим при послідовному контролі час контролю буде визначатися ймовірністю виявлення відмов окремих параметрів об'єкта:

$$\bar{T}_k = \sum_{i=1}^m p_{i-1} t_{ki}, \quad (12.28)$$

де t_{ki} – час контролю i -го параметра об'єкта; p_{i-1} – апіорна ймовірність того, що при контролі $(i-1)$ -го параметра відмова об'єкта не буде виявлена.

При заданому законі розподілу похибок контролю кількість інформації, отримувана в процесі контролю, буде максимальною у випадку статистичної незалежності і при рівномірному законі розподілу контрольованих параметрів:

$$I_{\max} = \sum_{i=1}^m [h_{\max}(x_i) - h(\xi_i)] = \sum_{i=1}^m \log_2 \frac{(x_{i2} - x_{i1})}{2\Delta_{ei}}, \quad (12.29)$$

де x_{i1} і x_{i2} – границі зміни i -го контрольованого параметра.

Підставивши з виразів (12.28) і (12.29) в (12.27) значення \bar{T}_k і I_{\max} , отримаємо:

$$V_k = \frac{\sum_{i=1}^m \log_2 \frac{(x_{i2} - x_{i1})}{2\Delta_{ei}}}{\sum_{i=1}^m p_{i-1} t_{ki}}. \quad (12.30)$$

Якщо при контролі вирішується двоальтернативна задача, то максимальну кількість інформації буде отримано при однаковій ймовірності “норма” і “не норма” контрольованих параметрів. При цьому апіорна ймовірність контрольованих параметрів дорівнює 1 дв. од. і вираз (12.30) приймає вигляд:

$$V_k = \frac{m \log_2 \frac{1}{\Delta_e}}{\sum_{i=1}^m P_{i-1} f_{ki}} \quad (12.31)$$

Із виразів (12.30) і (12.31) видно, що параметр V_k є характеристикою швидкодії та точності процесу контролю. При заданій пропускній здатності контролю точність контролю може бути підвищена в обмін на зниження швидкодії, і навпаки.

12.6 Ентропія та інформація у системах автоматичного управління

Поняття ентропії та інформації є зручними узагальненими характеристиками при описуванні складних систем автоматичного управління.

Будь-який процес, який належить автоматичному управлінню, можна охарактеризувати сукупністю координат x_1, x_2, \dots, x_m . Ці координати практично завжди мають розкид відносно номінальних значень. Цей розкид обумовлений численними факторами і має випадковий характер. Для оцінювання невизначеності процесу може бути використана ентропія розподілу ймовірностей координат керованого процесу, яка виражається такою ж формулою, що й ентропія, розглянута раніше в системах передачі інформації:

$$h(X) = - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \omega(X) \log_2 \omega(X) dX \quad (12.32)$$

Незважаючи на спільність математичних виразів, ентропія процесу суттєво відрізняється від інформаційної ентропії. У системах передачі інформації після отримання інформації ентропія об'єкта зменшується. Ентропія ж процесу, після отримання спостерігачем інформації про його стан, не змінюється. Отже, ентропію розподілу ймовірностей координат процесу не можна змінити вимірюванням цих координат. Для зміни ентропії процесу необхідно управління або відновлення.

Якщо проводити аналогію з процесом контролю параметрів як процесом передачі інформації, то введене поняття ентропії процесу відповідає при контролі ентропії параметрів за рахунок їх флуктуації внаслідок зовнішніх збурюючих впливів.

Ентропія процесу може бути змінена тільки за рахунок впливу на цей процес. Основним призначенням автоматизованого управління процесом є зменшення відхилень процесу від заданого, зменшення невизначеності перебігу процесу, тобто зменшення ентропії процесу.

Щоб проілюструвати викладене, розглянемо наступний приклад.

Генератор змінного струму має при деяких умовах розкид напруги $\sigma_{0v} = 50$ В і частоти $\sigma_{0f} = 50$ Гц. При включенні системи регулювання напруги і частоти розкид зменшується до $\sigma_{1v} = 2$ В і $\sigma_{1f} = 1$ Гц. Знайти зменшення ентропії координат процесу при переході від некерowanego процесу до керованого при умові незалежності і нормального розподілу ймовірностей координат v і f .

Загальне зменшення ентропії буде дорівнювати сумі зменшень ентропії координат:

$$h_0 - h_1 = h_0(v) - h_1(v) + h_0(f) - h_1(f) = \\ = \log_2 \frac{\sigma_{0v}}{\sigma_{1v}} + \log_2 \frac{\sigma_{0f}}{\sigma_{1f}} = \log_2 25 + \log_2 10 = 7,97 \text{ дв. од.}$$

Для процесів управління характерні замкнені контури циркуляції інформації. Схема такого контуру зображена на рис. 12.1.

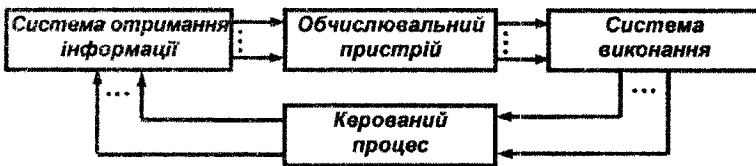


Рис. 12.1

Параметри процесу x_1, x_2, \dots, x_m вимірюються системою отримання інформації. Отримана інформація опрацьовується обчислювальним пристроєм. Опрацьована інформація надходить в систему виконання, яка формує управляючі впливи z_i , які впливають на керований процес в напрямку зменшення його ентропії.

Управління процесом може бути неперервним або дискретним в часі. Обмежимося розглядом дискретного управління, характерного для контурів з цифровими управляючими машинами.

Управляючі впливи надходять на керований процес через інтервали часу τ .

Нехай до передачі управляючих впливів початкова ентропія процесу дорівнює $h_0(X)$. Після включення системи управління ентропія процесу зменшується до значення $h_{\xi}(X)$, яке визначається впливом різного роду збурень в контурі управління (похибок вимірювачів, перетворювачів та пристроїв виконання, зовнішніх збурень тощо). Кількість інформації, внесеної при цьому в контур управління:

$$I(X) = h_0(X) - h_{\xi}(X). \quad (12.33)$$

Цей процес буде відбуватися за певну кількість тактів роботи пристрою управління. Після подачі кожного управляючого впливу ентропія процесу зменшується на певну величину. В інтервалах між впливами ентропія процесу зростає.

Кількість інформації $I_k(X)$, яка вноситься в контур управління в k -му такті роботи:

$$I_k X = h_{k-1}(X) - h_k(X) + \Delta h_k(X), \quad (12.34)$$

де $h_{k-1}(X)$ – ентропія процесу в кінці $(k-1)$ -го такту роботи; $h_k(X)$ – ентропія процесу в кінці k -го такту роботи; $\Delta h_k(X)$ – приріст ентропії процесу за інтервал дискретності τ внаслідок дії збурень.

Якщо за час перехідного періоду, протягом якого ентропія процесу зменшується від значення $h_0(X)$ до значення $h_{\xi}(X)$, здійснено n тактів роботи пристрою управління, то загальна кількість інформації, яка внесена в контур управління за цей час:

$$I_{пер}(X) = \sum_{k=1}^n [h_{k-1}(X) - h_k(X)] + \sum_{k=1}^n \Delta h_k(X). \quad (12.35)$$

В усталеному режимі буде дотримуватися умова:

$$h_{k-1}(X) = h_k(X); \quad I_k(X) = \Delta h_k(X), \quad (12.36)$$

тобто кількість інформації, яка передається через контур управління за інтервал дискретності, буде дорівнювати приросту ентропії процесу за рахунок дії збурень.

Сумарний приріст кількості інформації за час перехідного періоду дорівнює сумарному зменшенню ентропії процесу:

$$I(x) = \sum_{k=1}^n [h_{k-1}(X) - h_k(X)], \quad (12.37)$$

Характер зміни ентропії процесу X в перехідному і в усталеному режимах зображений на рис. 12.2.

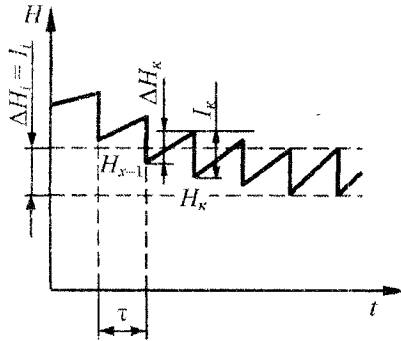


Рис. 12.2

Розглянемо наступний приклад. Припустимо, що необхідно побудувати систему стабілізації літального апарату, в контур якого включена ЦЕОМ, яка забезпечує в стаціонарному режимі точність стабілізації:

- кута тангажу $\sigma_\psi = 1^\circ$;
- кута ристання $\sigma_\beta = 0,5^\circ$;
- кута обертання $\sigma_\alpha = 0,3^\circ$;
- дальності польоту $\sigma_n = 50$ м;
- по нормалі $\sigma_x = 30$ м.

Потрібно, щоб система стабілізації після її включення виводила об'єкт у вказаний стаціонарний режим не більше ніж за 30 с при наступному початковому розкиді координат (до включення системи):

$$\sigma_{0\psi} = 8^\circ; \quad \sigma_{0\beta} = 10^\circ; \quad \sigma_{0\alpha} = 20^\circ; \quad \sigma_{0n} = 3 \text{ км}; \quad \sigma_{0x} = 33 \text{ км}.$$

У випадку раптового припинення управління в стаціонарному режимі на 1 с, відхилення зростають до наступних значень:

$$\sigma'_\psi = 1,5^\circ; \quad \sigma'_\beta = 5^\circ; \quad \sigma'_\alpha = 10^\circ; \quad \sigma'_n = 55 \text{ м}; \quad \sigma'_x = 33 \text{ м}.$$

Для простоти закон розподілу ймовірностей вихідних координат приймається нормальним, координати – незалежними, інтервал дискретності управління – однаковим для всіх координат і дорівнює $\tau = 1$ с, а також те, що самі координати зберігаються в спеціальних запам'ятовуючих пристроях і в прикладі не розглядаються.

1. Визначити кількість інформації, яку необхідно передавати через повний багатовимірний контур управління в стаціонарному режимі протягом інтервалу дискретності.

2. Оцінити загальну кількість інформації, яку необхідно визначити через контур управління в перехідному режимі.

3. Оцінити нижню границю ємності пам'яті ЦЕОМ.

Оскільки кількість інформації, яка передається через контур управління за інтервал дискретності в усталеному режимі, дорівнює приросту ентропії процесу за рахунок дії збурень, то відповідь на перше питання буде наступною:

$$I_k = \log_2 \frac{\sigma'_\psi}{\sigma_\psi} + \log_2 \frac{\sigma'_\beta}{\sigma_\beta} + \log_2 \frac{\sigma'_\alpha}{\sigma_\alpha} + \log_2 \frac{\sigma'_\eta}{\sigma_\eta} + \log_2 \frac{\sigma'_x}{\sigma_x} =$$

$$= \log_2 1,5 + \log_2 10 + \log_2 33,3 + \log_2 1,1 + \log_2 1,1 = 9,239 \text{ дв. од.}$$

Загальна кількість інформації, яка передається через контур управління в перехідному режимі, дорівнює сумі загальної зміни ентропії процесу за рахунок введення управління і приросту ентропії за рахунок дії збурень. Оскільки за час перехідного періоду здійснюється $m = 30$ тактів управління, то:

$$I_{\text{пер}} = h_0 + h_1 + 30I_k = \log_2 \frac{\sigma_{0\psi}}{\sigma_\psi} + \log_2 \frac{\sigma_{0\beta}}{\sigma_\beta} + \log_2 \frac{\sigma_{0\alpha}}{\sigma_\alpha} + \log_2 \frac{\sigma_{0\eta}}{\sigma_\eta} + \log_2 \frac{\sigma_{0x}}{\sigma_x} =$$

$$= \log_2 8 + \log_2 20 + \log_2 66,6 + \log_2 15 + \log_2 20 + 30I_k = 298,87 \text{ дв. од.}$$

Нижня границя ємності пам'яті ЦЕОМ:

$$Q > \frac{I_{\text{пер}}}{m} = \frac{298,87}{30} = 9,9 \text{ дв. од./такт.}$$

12.7 Оцінка ефективності АСКУ

З інформаційної точки зору ефективність АСКУ доцільно оцінювати кількістю інформації, отриманої системою в певний інтервал часу з врахуванням економічних затрат. У зв'язку з цим ефективність АСКУ потрібно оцінювати величиною:

$$K = \frac{I(X)}{C}, \quad (12.38)$$

де $I(X)$ – кількість інформації, яку отримує АСКУ в певний інтервал часу; C – вартість АСКУ.

Незручність критерію ефективності, який виражається формулою (12.38), полягає в його ненормованості. Критерієм, який не має даного недоліку, є узагальнений статистичний критерій, який виражає відношення ефективності реальної і потенціальної АСКУ:

$$E = \frac{K_p}{K_n}, \quad (12.39)$$

де K_p – ефективність реальної АСКУ; K_n – ефективність потенціальної (ідеальної) АСКУ.

Потенціально АСКУ забезпечує отримання максимально можливої кількості інформації. Це буде мати місце при максимумі апіорної і мінімумі апостеріорної ентропії об'єкта контролю. Отже, для потенціальної системи повинна бути рівність апіорних ймовірностей контрольованих параметрів:

$$p(x_i) = q(x_i),$$

де $p(x_i)$ – ймовірність стану “норма” i -го параметра¹; $q(x_i)$ – ймовірність стану “не норма” i -го параметра.

При цьому апіорна ентропія об'єкта буде дорівнювати:

$$H_0(x) = -\sum_{i=1}^m [p(x_i) \log_2 p(x_i) + q(x_i) \log_2 q(x_i)] = m \text{ дв. од.} \quad (12.40)$$

Розв'язувальний пристрій потенціальної системи прагне за критерієм ідеального спостерігача, який мінімізує сумарну помилку прийняття рішення:

$$q(x_i)\alpha_i + p(x_i)\beta_i = \min, \quad (12.41)$$

де α_i і β_i – помилки першого і другого роду.

У результаті здійснення контролю потенціальна система переводить стан об'єкта в деякий більш визначений стан, який характеризується згідно з формулою Байєса апостеріорною ймовірністю справного стану:

$$p(x_i / y_i) = \frac{p(x_i)[1 - \alpha_i]}{p(x_i)[1 - \alpha_i] + q(x_i)\beta_i} \quad (12.42)$$

і ентропією

$$H_1(x_i) = -p(x_i / y_i) \log_2 p(x_i / y_i) - [1 - p(x_i / y_i)] \log_2 [1 - p(x_i / y_i)] \quad (12.43)$$

¹Мається на увазі, що практично $p(x_i) \geq 0,5$.

Загальна кількість інформації, отримувана потенціальною системою:

$$I_n(x) = I_{\max} = \sum_{i=1}^m H_0(x_i) - \sum_{i=1}^m H_1(x_i).$$

Потенціальна система є ідеальною не тільки в сенсі максимуму кількості отримуваної інформації, вона є ідеальною також в сенсі простоти, оскільки в ній не передбачені резервування, доопрацювання для отримання потрібної швидкодії, об'єму, маси тощо.

Отже, потенціальна система при виконанні вказаних умов буде мати мінімальну вартість C_{\min} .

Ефективність потенціальної системи оцінюється коефіцієнтом:

$$K_n = \frac{I_{\max}}{C_{\min}}. \quad (12.44)$$

Математична модель реальної АСКУ будується аналогічно математичній моделі потенціальної системи. Однак при цьому приймаються реальні закони розподілу ймовірностей станів контрольованих параметрів і різні алгоритми роботи розв'язувальних пристроїв, обумовлені вибраним критерієм оцінювання параметрів.

Загальна вартість C інформації I_p з врахуванням затрат на отримання потрібної надійності, швидкодії, об'єму, ваги і т. п. в реальній АСКУ:

$$C = \sum_{i=1}^m C_i > C_{\min}. \quad (12.45)$$

Ефективність роботи реальної АСКУ:

$$K_p = \frac{I_p}{C} \leq K_n. \quad (12.46)$$

Достоїнством узагальненого статистичного критерію (12.38), отриманого на основі потенціальної і реальної математичних моделей АСКУ, є повна наочність, порівняна простота і загальність, що дозволяють одним числом характеризувати систему як загалом, так і по частинам, які включають складні і прості пристрої. При цьому дає змогу змінити критерію для практичних систем:

$$0 \leq E \leq 1. \quad (12.47)$$

Недосконала АСКУ має E , близький до нуля, досконала – до одиниці.

12.8 Інформаційне обґрунтування оптимального алгоритму контролю і пошуку несправностей

Кожний складний об'єкт можна представити у вигляді системи, яка складається із декількох динамічних послідовно з'єднаних ланок, які, в свою чергу, складаються із елементів.

При експлуатації таких систем розрізняють два специфічних алгоритми:

— контролю, призначений для встановлення факту справності чи несправності системи;

— пошуку несправностей у випадку встановлення факту несправності системи.

Визначення оптимальних алгоритмів контролю та пошуку несправностей в системі можна виконати на основі теорії інформації, теорії ігор, теорії масового обслуговування та теорії послідовного аналізу Вальда.

При визначенні вказаних алгоритмів на основі теорії інформації можна використати два різних критерії:

— максимальної кількості інформації за одиницю часу, з застосуванням якого контроль та пошук несправності розпочинаються з дослідження, який дає максимальну кількість інформації за одиницю часу;

— максимальної цінності інформації, з застосуванням якого контроль та пошук несправності розпочинаються з дослідження, який дає максимальну цінність інформації.

Критерій максимальної кількості інформації доцільно використовувати при автоматичному контролі та пошуку несправностей.

Критерій максимальної цінності інформації слід використовувати при розробці інструкцій з експлуатації устаткування обслуговуючим персоналом.

Розглянемо методику вибору оптимальних алгоритмів контролю та пошуку несправностей в системі, які ґрунтуються на застосуванні критерію максимальної кількості інформації. При цьому вважається, що система контролю працює без помилок.

Невизначеність стану системи, що контролюється, до початку контролю та пошуку несправностей характеризується її повною ентропією $H(X)$. Процес контролю являє собою складний дослід A_k , що складається не більш, ніж з k дослідів a .

Середня кількість інформації, отримувана при кожному досліді a_i відносно стану системи X :

$$I(a_i, X) = H(X) - H_{a_i}(X),$$

де $H_{a_i}(X) = p(A_{i1})H_{A_{i1}}(X) + \dots + p(A_{ij})H_{A_{ij}}(X)$ – середня умовна ентропія стану системи після здійснення дослідів a_i ; $P(A_{ij})$ – ймовірність окремих результатів A_{ij} дослідів a_i .

Якщо результат складного дослідів A_k однозначно визначає стан системи, що характеризується вектором X , то ентропія системи після здійснення цього дослідів $H(X) = 0$, а середня кількість інформації про стан системи, яка отримана при здійсненні дослідів A_k :

$$I(A_k, X) = H(X).$$

Якщо результат складного дослідів не повністю розкриває працездатність стану системи, то:

$$H_{A_k}(X) \neq 0 \text{ і } I(A_k, X) < H(X).$$

Основа оптимального алгоритму контролю полягає у виборі такої мінімальної кількості дослідів a , що є складовими складного дослідів A_k , кожен з яких давав би максимальну кількість інформації.

Процес контролю повинен вестися за алгоритмом, який складений на основі критерію:

$$I(a_i, X) = H(X) - H_{in}(X) = I_{\max},$$

тобто процес контролю повинен починатися з дослідів, який дає максимальну кількість інформації.

Після здійснення дослідів a_1 система буде зводити ентропією деякого нового стану $H_{a_1}(X)$.

Припустимо, що стан системи до початку процесу контролю характеризується станами: система справна з ймовірністю P ; система несправна з ймовірністю $1 - P$. Крім того, для спрощення задачі, а також з врахуванням досвіду експлуатації припустимо, що система не працює через відмову тільки одного елемента із m "з'єднаних" ланцюгом, який має умовну ймовірність відмови q_k , яка визначається на основі статистичних відомостей про надійність елементів.

Таким чином, відмови елементів системи передбачаються несутісними, а для повної групи елементів буде справедливою рівність (при наявності відмови):

$$\sum_{k=1}^m q_k = 1.$$

Ентропія системи:

$$H(X) = - \left\{ P \log_2 P + (1-P) \log_2 (1-P) + (1-P) \sum_{k=1}^m q_k \log_2 q_k \right\}. \quad (12.48)$$

Припустимо, що при проведенні дослідів a_1 контролюється n елементів системи, а також, що маємо два результати дослідів:

– частина системи, що контролюється, справна – ймовірність цього виходу $\left[1 - (1-P) \sum_{k=1}^n q_k \right]$;

– частина системи, що контролюється, несправна – ймовірність такого виходу $\left[(1-P) \sum_{k=1}^n q_k \right]$.

Середня умовна ентропія стану системи після здійснення першого дослідів:

$$\begin{aligned} H_{a_1}(X) = & - \left\{ P \log_2 P + (1-P) \log_2 (1-P) + \right. \\ & \left. + (1-P) \sum_{k=1}^n q_k \log_2 q_k - \left[1 - \sum_{k=1}^n q_k (1-P) \right] \times \right. \\ & \left. \times \log_2 \left[1 - \sum_{k=1}^n q_k (1-P) \right] - \sum_{k=1}^n q_k (1-P) \log_2 (1-P) \sum_{k=1}^n q_k \right\}. \end{aligned} \quad (12.49)$$

Кількість інформації про стан системи, отримуваної при проведенні дослідів:

$$\begin{aligned} I(a_1, X) = & - \left\{ \left[1 - \sum_{k=1}^n q_k (1-P) \right] \log \left[1 - \sum_{k=1}^n (1-P) \right] + \right. \\ & \left. + \left[\sum_{k=1}^n q_k (1-P) \right] \log_2 \left[\sum_{k=1}^n q_k (1-P) \right] \right\}. \end{aligned} \quad (12.50)$$

Задача полягає в тому, щоб на першому етапі контролю вибрати таку кількість контрольованих елементів, при якій кількість отриманої інформації була б максимальною.

Із формули (12.50) видно, що $I(a_1, X)$ залежить від $P \sum_{k=1}^n q_k$ і є максимальним при виконанні умови:

$$1 - \sum_{k=1}^n q_k (1-P) = \frac{1}{2},$$

або

$$\sum_{k=1}^n q_k = \frac{1}{2(1-P)}, \quad (12.51)$$

Розглянемо, як змінюється n залежно від ймовірності справного стану об'єкта P при виконанні умови (12.51).

При $P = 0$ повинна виконуватись рівність $\sum_{k=1}^n q_k = 0,5$.

При рівності ймовірностей відмов елементів

$$q_1 = q_2 = \dots = q_m = \frac{1}{m}$$

отримаємо:

$$n = \frac{m}{2}, \quad (12.52)$$

тобто, першим етапом контролю повинна бути охоплена половина елементів об'єкта.

Зі збільшенням P збільшується n . При $P = 0,5$, $n = m$, тобто контролем на першому етапі повинні бути охоплені всі елементи.

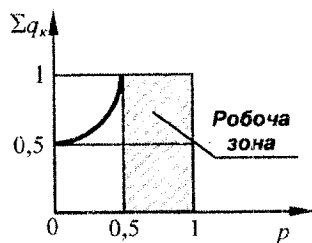


Рис. 12.3

З подальшим збільшенням P збільшується і сума $\sum_{k=1}^n q_k$. Але оскільки в реальних умовах сума $\sum_{k=1}^n q_k$ не може бути більшою за одиницю, то при $P > 0,5$ приймаємо $\sum_{k=1}^n q_k = 1$ (рис. 12.3) або $n = m$.

Оскільки для реальних систем ймовірність справного стану системи $P > 0,5$, то при виборі першого дослідження необхідно алгоритм побудувати так, щоб охопити контролем $n = m$ елементів системи.

Якщо $P < 0,5$ (що практично малоімовірно), то першим етапом контролю охоплюються не всі елементи об'єкта. Тоді наступний дослід вибирається на основі того ж критерію, але з врахуванням того, що стан системи характеризується ентропією (12.49). Досліди ведуться до тих пір, поки ентропія системи не буде дорівнювати нулю.

Алгоритм пошуку несправностей в системі складається при наступних передумовах:

– відомо, що система несправна;
 – у системі відмовив тільки один елемент із m , “з’єднаних” ланцюгом;

– відомі апріорні ймовірності відмови всіх елементів системи.

Методика визначення оптимального алгоритму пошуку несправностей така ж, як і при контролі, тобто полягає у виборі такої кількості дослідів, щоб кожний із них давав максимальну кількість інформації.

Нехай на першому етапі пошуку несправностей a_1 розглядається n елементів об’єкта із загальної кількості m . При цьому можливі два результати дослідів:

– у частині об’єкта, що розглядається, виявлено відмову (ймовірність цього виходу $\sum_{k=1}^n q_k$);

– у частині об’єкта, що розглядається, не виявлено відмов (ймовірність цього виходу $1 - \sum_{k=1}^n q_k$).

Кількість інформації, отримуваної при проведенні дослідів a_1 :

$$I(a_1, x) = - \left\{ \sum_{k=1}^n q_k \log_2 \sum_{k=1}^n q_k + \left(1 - \sum_{k=1}^n q_k\right) \log_2 \left(1 - \sum_{k=1}^n q_k\right) \right\}. \quad (12.53)$$

Із формули (12.53) видно, що для отримання максимальної кількості інформації $I(a_1, x)$ при проведенні першого дослідів пошуку несправності необхідно, щоб сума $\sum_{k=1}^n q_k$ дорівнювала $1/2$. Наступний дослід вибирається на основі того ж критерію, але з врахуванням ентропії стану системи після здійснення першого дослідів a_1 .

Так, наприклад, при пошуку відмови у системі, яка складається із m елементів, які мають однакову ймовірність відмови, необхідно в першому досліді перевірити $n = m/2$ елементів, в другому – $n/2$ і т.д.

Таким чином, алгоритм пошуку несправності відрізняється від алгоритму контролю внаслідок того, що критерії вибору першого дослідів різні. Крім того, після проведення дослідів a_1 дослід a_2 у випадку пошуку несправності проводиться, виходячи з дослідів a_1 , який відповідає непрацездатності системи, а у випадку контролю – виходячи із дослідів a_1 , який відповідає працездатності частини системи.

Контрольні питання

1. Які системи відносяться до автоматизованих систем контролю та управління?
2. Який вираз описує невизначеність стану об'єкта контролю?
3. Яких значень може досягати ентропійне значення похибки контролю?
4. Як можна підвищити точність контролю?
5. Наведіть структурну схему замкненого контуру циркуляції інформації для процесів управління.
6. Як оцінюється ефективність АСКУ?
7. Дайте визначення оптимального алгоритму контролю.
8. При яких ймовірностях реальна система вважається справною?

13.1 Алгоритми архівації зображень

13.1.1 Загальні характеристики

Протягом останніх десяти років у рамках комп'ютерної графіки бурхливо розвивається зовсім нова область – алгоритми архівації зображень. Поява цієї області обумовлена тим, що зображення – це своєрідний тип даних, що характеризуються трьома особливостями.

1. Зображення (як і відео) займають набагато більше місця в пам'яті, чим текст. Так, скромна, не дуже якісна ілюстрація на обкладинці книги розміром 500×800 точок займає 1,2 Мб – стільки ж, скільки художня книга з 400 сторінок (60 знаків у рядку, 42 рядки на сторінці). Як приклад можна розглянути також, скільки тисяч сторінок тексту ми зможемо помістити на CD-ROM, і як мало там уміститься якісних нестиснутих фотографій. Ця особливість зображень визначає актуальність алгоритмів архівації графіки.

2. Другою особливістю зображень є те, що людський зір при аналізі зображення оперує контурами, загальним переходом кольорів і порівняно невідчутний до малих змін у зображенні. Таким чином, ми можемо створити ефективні алгоритми архівації зображень, у яких декомпресоване зображення не буде збігатися з оригіналом, однак людина цього не помітить. Ця особливість людського зору дозволила створити спеціальні алгоритми стиснення, орієнтовані тільки на зображення. Ці алгоритми мають дуже високі характеристики.

3. Ми можемо легко помітити, що зображення, на відміну, наприклад, від тексту, має надлишковість у двох вимірах. Тобто, як правило, сусідні точки як по горизонталі, так і по вертикалі у зображенні близькі за кольором. Крім того, ми можемо використовувати подібність між кольоровими площинами R, G і B у алгоритмах, що дає можливість створити ще більш ефективні алгоритми. Таким чином, при створенні алгоритму компресії графіки ми використовуємо особливості структури зображення.

Усього на даний момент відомі мінімум три сімейства алгоритмів, що розроблені винятково для стиснення зображень, і методи, які в них застосовуються, практично неможливо застосувати до архівації ще яких-небудь видів даних.

Для того, щоб говорити про алгоритми стиснення зображень, необхідно визначитися з декількома важливими питаннями:

1. Які критерії ми можемо запропонувати для порівняння різних алгоритмів?

2. Які класи зображень існують?

3. Які класи додатків, що використовують алгоритми компресії графіки, існують і які вимоги вони висувають до алгоритмів?

Розглянемо ці питання докладніше.

13.1.2 Класи зображень

Статичні растрові зображення являють собою двовимірний масив чисел. Елементи цього масиву називають пікселами (англ. *pixel* – picture element). Усі зображення можна поділити на дві групи – з палітрою і без неї. У зображень з палітрою в пікселі зберігається число-індекс у деякому одновимірному векторі кольорів, що називається палітрою. Найчастіше зустрічаються палітри з 16 і 256 кольорів.

Зображення без палітри бувають у будь-якій системі кольорового представлення і у градаціях сірого. Для останніх значення кожного піксела інтерпретується як яскравість відповідної точки. Зустрічаються зображення з 2, 16 і 256 рівнями сірого. Одна з цікавих практичних задач полягає в приведенні кольорового чи чорно-білого зображення до двох градацій яскравості, наприклад, для друку на лазерному принтері. При використанні деякої системи представлення кольору кожен піксел являє собою запис (структуру), полями якого є компоненти кольору. Найпоширенішою є система RGB, у якій колір представлений значеннями інтенсивності червоної (R), зеленої (G) і синьої (B) компонент. Існують і інші системи представлення кольору, такі, як CMYK, CIE XYZcig60-1 тощо.

Для того, щоб коректніше оцінювати ступінь стиснення, потрібно ввести поняття класу зображень. Під класом буде розумітися деяка сукупність зображень, застосування до яких алгоритму архівації дає якісно однакові результати. Наприклад, для одного класу алгоритм дає дуже високий ступінь стиснення, для іншого – майже не стискає, для третього – збільшує файл у розмірі, (відомо, що багато алгоритмів у гіршому випадку збільшують файл).

Розглянемо наступні приклади неформального визначення класів зображень.

Клас 1. Зображення з невеликою кількістю кольорів (4–16) і великими областями, заловненими одним кольором. Плавні переходи кольорів відсутні. Приклади: ділова графіка – гістограми, діаграми, графіки тощо.

Клас 2. Зображення із плавними переходами кольорів, побудовані на комп'ютері. Приклади: графіка презентацій, ескізи моделі в САПР, зображення, побудовані за методом Гуро.

Клас 3. Фотореалістичні зображення. Наприклад, скановані фотографії.

Клас 4. Фотореалістичні зображення з накладанням ділової графіки. Наприклад, реклама.

Розвиваючи дану класифікацію, як окремі класи можуть бути запропоновані неякісно скановані в 256 градациях сірого кольору сторінки книг, растрові зображення топографічних карт (цей клас не тотожний класу 4). Формально будучи 8- чи 24-бітовими, вони несуть навіть не растрову, а чисто векторну інформацію.

Окремі класи можуть утворювати і зовсім специфічні зображення: рентгенівські знімки чи фотографії в профіль і фас з електронного носія.

Досить складною і цікавою задачею є пошук найкращого алгоритму для конкретного класу зображень.

13.1.3 Класи програм архівації (розархівації)

Розглянемо наступну просту класифікацію програм архівації (розархівації), що використовують алгоритми компресії.

Клас 1. Характеризується високими вимогами до часу архівації та розархівації. Нерідко потрібно переглядати зменшені копії зображення і шукати їх у базі даних зображень. Приклади: видавничі системи в широкому розумінні цього слова, причому, як ті, що готують якісні публікації (журнали) з високою якістю зображень і використанням алгоритмів архівації без втрат, так і газети, і інформаційні вузли у www, де є можливість оперувати зображеннями меншої якості і використовувати алгоритми стиснення з втратами. У подібних системах доводиться мати справу з повнокольоровими зображеннями самого різного розміру: від 640×480 – формат цифрового фотоапарата до 3000×2000 і з великими двоколірними зображеннями. Оскільки ілюстрації займають значну частку від загального обсягу матеріалу в документі, проблема збереження стоїть дуже гостро. Проблеми також створює велика різноманітність ілюстрацій (доводиться використовувати універсальні алгоритми). Єдانه, що можна сказати заздалегідь, це те, що будуть переважати фотореалістичні зображення і ділова графіка.

Клас 2. Характеризується високими вимогами до ступеня архівації та часу розархівації. Час архівації ролі не грає. Іноді подібні програми архівації (розархівації) також вимагають від алгоритму компресії легкості масштабування зображення під конкретну роздільну здатність монітора користувача. Наприклад, довідники і енциклопедії на CD-ROM.

З появою великої кількості комп'ютерів, оснащених цим приводом, досить швидко сформувався ринок програм, що випускаються на лазерних дисках. Незважаючи на те, що ємність одного диска досить велика (приблизно 650 Мб), її, як правило, не вистачає. При створенні енциклопедій та ігор велику частину диска займають статичні зображення і відео. Таким чином, для цього класу програм архівації (розархівації) актуальними стають істотно асиметричні за часом алгоритми (симетричність за часом – відношення часу компресії до часу декомпресії).

Клас 3. Характеризується дуже високими видами до ступеня архівації. Програма архівації (розархівації) клієнта одержує від сервера інформацію з мережі. Приклад: всесвітня мережа Internet. У цій гіпертекстовій системі досить активно використовуються ілюстрації. При оформленні інформаційних чи рекламних сторінок хочеться зробити їх більш яскравими і барвистими, що природно позначається на розмірі зображень. Більше всього при цьому страждають користувачі, підключені до мережі за допомогою повільних каналів зв'язку. Якщо сторінка www перенасичена графікою, то очікування її повної появи на екрані може затягтися. Оскільки при цьому навантаження на процесор мале, то тут можуть знайти застосування ефективні стискаючі складні алгоритми з порівняно великим часом розархівації. Крім того, ми можемо видозмінити алгоритм і формат даних так, щоб переглядати примітивне зображення файла до його повного одержання.

Можна навести безліч більш вузьких класів програм архівації (розархівації). Так, своє застосування машинна графіка знаходить і в різних інформаційних системах. Наприклад, уже стає звичним досліджувати ультразвукові та рентгенівські знімки не на папері, а на екрані монітора. Поступово в електронний вигляд переводять і історії хвороб. Зрозуміло, що зберігати ці матеріали логічніше в єдиній картотеці. При цьому без використання спеціальних алгоритмів велику частину архівів займають фотографії. Тому при створенні ефективних алгоритмів розв'язку цієї задачі потрібно врахувати специфіку рентгенівських знімків – переваги розмитих ділянок.

У геоінформаційних системах – при збереженні аерофотознімків місцевості – специфічними проблемами є великий розмір зображення і необхідність вибірки лише частини зображення за вимогою. Крім того, може знадобитися масштабування. Це неминуче накладає свої обмеження на алгоритм компресії.

В електронних картотеках і досє різних служб для зображень характерна подібність між фотографіями в профіль і в анфас, що також необхідно враховувати при створенні алгоритму архівації. Подібність між фотографіями спостерігається й у будь-яких інших спеціалізованих довідниках. Як приклад можна навести енциклопедії птахів чи квітів.

Характер використання зображень задає ступінь важливості наведених спірних вимог до алгоритму:

1. Високий ступінь компресії. Зауважимо, що далеко не для всіх програм актуальний високий ступінь компресії. Крім того, деякі алгоритми дають краще співвідношення якості до розміру файла при високих ступенях компресії, однак програють іншим алгоритмам при низьких ступенях.

2. Висока якість зображень. Виконання цієї вимоги прямо суперечить виконанню попередньої вимоги.

3. Висока швидкість компресії. Ця вимога для деяких алгоритмів із втратою інформації є взаємовиключною з першими двома. Інтуїтивно зрозуміло, що чим більше часу ми будемо аналізувати зображення, намагаючись одержати найвищий ступінь компресії, тим краще буде результат. І, відповідно, чим менше часу ми витратимо на компресію (аналіз), тим нижче буде якість зображення і більше його розмір.

4. Висока швидкість декомпресії. Досить універсальна вимога, актуальна для багатьох додатків. Однак можна навести приклади додатків, де час декомпресії далеко не критичний.

5. Масштабування зображень. Ця вимога має на увазі легкість зміни розмірів зображення до розмірів вікна активного додатка. Справа у тому, що одні алгоритми дозволяють легко масштабувати зображення прямо під час декомпресії, у той час як інші не тільки не дозволяють легко масштабувати, але і збільшують ймовірність появи спотворень після застосування стандартних алгоритмів масштабування до декомпресованого зображення. Можна навести приклад “поганого” зображення для алгоритму JPEG – це зображення з досить дрібним регулярним малюнком (піджак у дрібну клітку). Характер внесених алгоритмом JPEG спотворень такий, що зменшення (збільшення) зображення може дати негативні ефекти.

6. Можливість показати примітивне зображення (низької роздільної здатності), **використавши тільки початок файла.** Дана можливість актуальна для різного роду мережних програм, де перекачування зображень може тривати досить довго, і бажано, одержавши початок файла, коректно показати попереднє зображення. Примітивна реалізація зазначеної вимоги шляхом записування в початок зображення його зменшеної копії помітно погіршить ступінь компресії.

7. Стьйкість до помилок. Ця вимога означає локальність порушень у зображенні при втраті фрагмента переданого файла. Ця можливість використовується при передачі за багатьма адресами зображень по мережі, тобто в тих випадках, коли неможливо використовувати протокол передачі, повторно запитуючи дані в сервера при помилках. Наприклад, якщо передається відеоряд, то було б неправильно викорис-

товувати алгоритм, у якого збій приводив би до припинення правильного показу всіх наступних кадрів. Дана вимога суперечить високому ступеню архівації, оскільки інтуїтивно зрозуміло, що ми повинні вводити в потік надлишкову інформацію. Однак для різних алгоритмів обсяг цієї надлишкової інформації може істотно відрізнятись.

8. Облік специфіки зображення. Більш високий ступінь архівації для класу зображень, що статистично частіше будуть застосовуватися у додатку. У попередніх розділах ця вимога вже обговорювалася.

9. Редагованість. Під редагованістю розуміється мінімальний ступінь погіршення якості зображення при його повторному збереженні після редагування. Багато алгоритмів із втратою інформації можуть істотно зіпсувати зображення за кілька ітерацій редагування.

10. Невелика вартість апаратної реалізації. Ефективність програмної реалізації. Дані вимоги до алгоритму реально висувують не тільки виробники ігрових приставок, але і виробники багатьох інформаційних систем. Так, декомпресор фрактального алгоритму досить ефективно і з невеликими витратами об'єму можна реалізувати з використанням технології MMX і розпаралелювання обчислень, а стиснення у стандарті CCITT Group 3 легко реалізується апаратно.

13.1.4 Критерії порівняння алгоритмів

Сформульовані характеристики алгоритму щодо деяких вимог додатків залежать від конкретних умов, в які буде поставлений алгоритм. Виділимо декілька найбільш важливих критеріїв порівняння алгоритмів компресії, які й будемо використовувати надалі.

Найгірший, середній і кращий коефіцієнти стиснення. Тобто частка, на яку зросте зображення, якщо вихідні дані будуть найгіршими; деякий середньостатистичний коефіцієнт для того класу зображень, на який орієнтований алгоритм; і, нарешті, кращий коефіцієнт. Останній необхідний лише теоретично, оскільки показує ступінь стиснення найкращого (як правило, абсолютно чорного) зображення, іноді фіксованого розміру.

Клас зображень, на який орієнтований алгоритм. Іноді зазначається гажож, чому на інших класах зображень виходять гірші результати.

Симетричність. Відношення характеристики алгоритму кодування до аналогічної характеристики при декодуванні. Характеризує ресурсоемність процесів кодування і декодування. Для нас найбільш важливою є симетричність за часом: відношення часу кодування до часу декодування. Іноді нам буде потрібна симетричність за пам'яттю.

Чи є втрати якості? І якщо є, те за рахунок чого змінюється коефіцієнт архівації? Справа в тім, що в більшості алгоритмів стиснення з втратою інформації існує можливість зміни коефіцієнта стиснення.

Характерні риси алгоритму і зображень, до яких його застосовують. Тут можуть вказуватися найбільш важливі для алгоритму властивості, що можуть стати визначальними при його виборі.

Використовуючи дані критерії, розглянемо алгоритми архівації зображень.

Перш, ніж безпосередньо почати розмову про алгоритми, хотілося б зробити застереження. Той самий алгоритм часто можна реалізувати різними способами. Багато відомих алгоритмів, такі як RLE, LZW чи JPEG, мають десятки різних реалізацій. Крім того, у алгоритмів буває кілька явних параметрів, варіюючи які, можна змінювати характеристики процесів архівації та розархівації. При конкретній реалізації ці параметри фіксуються, виходячи з найбільш ймовірних характеристик вхідних зображень, вимог до економії пам'яті, вимог до часу архівації тощо. Тому в алгоритмів одного сімейства кращий і гірший коефіцієнти можуть відрізнитися, але якісно картина не зміниться.

13.2 Алгоритми архівації без втрат

13.2.1 Алгоритм RLE

Алгоритм надзвичайно простий у реалізації. Групове кодування (від англ. *Run Length Encoding* – RLE) одне із самих старих і найпростіше для алгоритмів архівації графіки. Зображення в ньому (як і в декількох алгоритмах, описаних нижче) витягується в ланцюжок байтів за рядками растра. Саме стиснення в RLE відбувається за рахунок того, що у вихідному зображенні зустрічаються ланцюжки однакових байт. Заміна їх на пари лічильних повторень, значень зменшує надлишковість даних. У кращому випадку цей алгоритм стискає файл у 64 рази, у гіршому – збільшує на 1/128. Середні показники ступеня компресії цього алгоритму знаходяться на рівні показників першого варіанта.

Схожі схеми компресії використані як один із алгоритмів, підтримуваних форматом TIFF, а також у форматі TGA.

Характеристики алгоритму RLE:

- коефіцієнти компресії: 64, 3, 128/129 (кращий, середній, гірший);
- клас зображень: алгоритм орієнтований на зображення з великою кількістю кольорів, ділову та наукову графіку;
- симетричність: приблизно одиниця;

– характерні риси: до позитивних сторін алгоритму, відноситься те, що він не вимагає додаткової пам'яті при архівації та розархівації, а також швидко працює. Цікава особливість групового кодування полягає в тому, що ступінь архівації для деяких зображень може бути істотно підвищений всього лише за рахунок зміни порядку кольорів у палітрі зображення.

13.2.2 Алгоритм LZW

Назву алгоритм одержав за першими буквами прізвищ його розробників – Lempel, Ziv і Welch. Стиснення у ньому, на відміну від RLE, здійснюється за рахунок однакових ланцюжків байт.

Існує досить велике сімейство LZ-подібних алгоритмів, що розрізняються, наприклад, за методом пошуку повторюваних ланцюжків. Один з досить простих варіантів цього алгоритму, наприклад, припускає, що у вхідному потоці йде або пара “лічильник – зсув щодо точної позиції”, або просто “лічильник” “пропуску” байт і самі значення байтів (як у алгоритмі RLE). При розархіваванні для пари “лічильник – зсув” копіюються “лічильник” байт із вихідного масиву, отриманого в результаті розархіваванні, на “зсув” байт раніше, а “лічильник” (тобто число рівне лічильнику) значень “пропуску” байт просто копіюється у вихідний масив із вхідного потоку. Даний алгоритм є несиметричним за часом, оскільки вимагає повного перебору буфера при пошуку однакових підрядків. У результаті нам складно задати великий буфер через різке зростання часу компресії. Однак потенційна побудова алгоритму, у якому на “лічильник” і на “зсув” буде виділено по 2 байти (старший біт старшого байта лічильника – ознака повтору рядка (копіювання) потоку), дасть нам можливість стискати всі повторювані підрядки розміром до 32 Кб у буфері розміром 64 Кб.

При цьому ми одержимо збільшення розміру файлу в гіршому випадку на 32770/32768 (у двох байтах записано, що потрібно переписати у вихідний потік наступні 215 байт), що зовсім непогано. Максимальний коефіцієнт стиснення складе приблизно 8192 рази, оскільки максимальне стиснення ми одержуємо, перетворюючи 32 Кб буфера в 4 байти, а буфер такого розміру ми заповнимо не відразу. Однак мінімальний підрядок, для якого нам вигідно проводити стиснення, повинен складатися в загальному випадку мінімум з 5 байт, що і визначає малу цінність даного алгоритму. До переваг LZ-алгоритму можна віднести надзвичайно простий алгоритм декомпресії.

Розглянутий нижче варіант алгоритму буде використовувати кодове дерево для представлення і збереження ланцюжків. Очевидно, що це досить сильне обмеження на вид ланцюжків, і далеко не всі однакові підланки в нашому зображенні будуть використані при стисненні. Однак у пропонованому алгоритмі вигідно стискати навіть ланцюжки, що складаються з 2 байт.

Процес стиснення виглядає досить просто. Ми зчитуємо послідовно символи вхідного потоку і перевіряємо, чи є в створеній нами таблиці рядків такий рядок. Якщо рядок є, то ми зчитуємо наступний символ, а якщо рядка немає, то заносимо в потік код для попереднього знайденого рядка, заносимо рядок у таблицю і починаємо пошук знову.

У розглянутому варіанті алгоритму використовується 12-бітовий код, і, відповідно, підкоди для рядків нам залишають значення від 258 до 4095. Рядки, що додаються, записуються в таблицю послідовно, при цьому індекс рядка в таблиці стає її кодом.

Розглянемо приклад. Нехай ми стискаємо послідовність 45, 55, 55, 151, 55, 55, 55. Тоді, відповідно до наведеного вище алгоритму, ми помістимо у вихідний потік спочатку код очищення <256>, потім додамо до порожнього рядка “45” і перевіримо, чи є рядок “45” у таблиці. Оскільки ми при ініціалізації занесли в таблицю всі рядки з одного символу, то рядок “45” є в таблиці. Далі ми читаємо наступний символ 55 із вхідного потоку і перевіряємо, чи є рядок “45, 55” у таблиці. Такого рядка у таблиці поки немає. Ми заносимо в таблицю рядок “45, 55” (з першим вільним кодом 258) і записуємо в потік код <45>. Можна коротко представити архівацію так:

“45” – є в таблиці; “45, 55” – немає. Додаємо в таблицю <258>“45, 55”. У потік: <45>; “55, 55” – немає. У таблицю: <259>“55, 55”. У потік: <55>; “55, 151” – немає. У таблицю: <260>“55, 151”. У потік: <55>; “151, 55” – немає. У таблицю: <261>“151, 55”. У потік: <151>; “55, 55” – є в таблиці; “55, 55, 55” – немає. У таблицю: “55, 55, 55” <262>. У потік: <259>;

Послідовність кодів для даного прикладу, що попадають у вихідний потік: <256>, <45>, <55>, <55>, <151>, <259>.

Особливість LZW полягає в тому, що для декомпресії нам не потрібно зберігати таблицю рядків у файл для розпакування. Алгоритм побудований таким чином, що ми можемо відновити таблицю рядків, користуючись тільки потоком кодів.

Ми знаємо, що для кожного коду треба додавати в таблицю рядок, що складається з уже наявного там рядка і символу, з якого починається наступний рядок у потоці.

Підрахуємо кращий і гірший коефіцієнти компресії для даного алгоритму. Кращий коефіцієнт, мабуть, буде отриманий для ланцюжка однакових байтів великої довжини (тобто для 8-бітового зображення, усі точки якого мають, для визначеності, колір 0). При цьому в 258-й рядок таблиці ми запишемо рядок “0, 0”, у 259-й – “0, 0, 0”,...у 4095-й – рядок з 3839 (4095 – 256) нулів. При цьому в потік потрапить 3840 кодів, включаючи код очищення. Отже, поразивши суму арифметичної прогресії від 2 до 3839 (тобто довжину стиснутого ланцюжка) і поділивши її на 3840 × 12/8 (у потік записуються 12-бітові коди), одержимо кращий коефіцієнт компресії.

Гірший коефіцієнт буде отриманий, якщо ми жодного разу не зустрінемо підрядок, що уже є в таблиці (у ній не повинно бути ні однієї однакової пари символів).

У випадку, якщо ми постійно будемо зустрічати новий підрядок, ми запишемо у вихідний потік 3840 кодів, яким буде відповідати рядок з 3838 символів, це складе збільшення файла майже в 1,5 рази.

LZW-алгоритм реалізований у форматах GIF і TIFF.

Характеристики алгоритму LZW:

- коефіцієнти компресії: приблизно 1000, 4, 5/7 (кращий, середній, гірший коефіцієнти); стиснення у 1000 разів досягається тільки на одноколірних зображеннях розміром, кратним приблизно 7 Мб;

- клас зображень: орієнтований LZW-алгоритм на 8-бітові зображення, побудовані на комп'ютері; стискає за рахунок однакових підланок в потоці;

- симетричність: майже симетричний, за умови оптимальної реалізації операції пошуку рядків в таблиці;

- характерні риси: ситуація, коли алгоритм збільшує зображення, зустрічається вкрай рідко; LZW універсальний – саме його варіанти використовуються в звичайних архіваторах.

13.2.3 Алгоритм Хаффмана

Один із класичних алгоритмів, відомих з 60-х років ХХ ст. Використовує тільки частоту появи однакових байт у зображенні. Заміняє символи вхідного потоку, що зустрічаються більше число раз, ланцюжками біт меншої довжини. І, навпаки, ті, що зустрічаються рідко – ланцюжками більшої довжини. Для збору статистики вимагає двох проходів по зображенню.

Алгоритм Хаффмана вимагає запису у файл таблиці відповідності кодованих символів і ланцюжків, що кодують.

На практиці використовуються його різновиди. Так, у деяких випадках резонно або використовувати фіксовану таблицю, або будувати її "адаптивно" тобто в процесі архівації (розархівації). Ці прийоми звільняють нас від двох проходів по зображенню і необхідності збереження таблиці разом з файлом. Кодування з фіксованою таблицею застосовується як останній етап архівації в деяких алгоритмах стиснення зображень.

Характеристики класичного алгоритму Хаффмана:

- коефіцієнти компресії: 8, 1,5, 1 (кращий, середній, гірший);

- клас зображень: практично не застосовується до зображень в чистому вигляді; зазвичай використовується як один з етапів компресії в більш складних алгоритмах;

- симетричність: 2 (за рахунок того, що вимагає двох проходів по масиву стиснутих даних);

- характерні риси: єдиний алгоритм, що не збільшує розміру вихідних даних у гіршому випадку (якщо не вважати на необхідність зберігати таблицю перекодування разом з файлом).

13.2.4 Алгоритм Хаффмана з фіксованою таблицею

Близька модифікація алгоритму використовується при стисненні чорно-білих зображень (один біт на піксел). Повна назва даного алгоритму – CCITT Group 3. Це означає, що даний алгоритм був запропонований третьою групою з стандартизації Міжнародного Консультативного Комітету з Телеграфії і Телефонії (Consultative Committee International Telegraph and Telephone). Послідовності чорних і білих точок у ньому замінюються числом, рівним їхній кількості. А цей ряд, вже у свою чергу, стискається за методом Хаффмана з фіксованою таблицею.

Набір точок одного кольору, що йдуть підряд, називається **серією**. Довжина цього набору точок називається **довжиною серії**.

Нижче наведені два види кодів:

- завершення серій – задані з 0 до 63 із кроком 1;
- складені (додаткові) – задані з 64 до 2560 із кроком 64.

Кожен рядок зображення стискається незалежно. Ми вважаємо, що в нашому зображенні істотно переважає білий колір, і всі рядки зображення починаються з білої точки. Якщо рядок починається з чорної точки, то ми вважаємо, що рядок починається білою серією з довжиною 0. Наприклад, послідовність довжин серій 0, 3, 556, 10, ... означає, що в цьому рядку зображення йдуть спочатку три чорні точки, потім 556 білі, потім 10 чорних і т.д.

Оскільки чорні і білі серії чергуються, то реально код для білої і код для чорної серії буде працювати поперемінно.

У термінах регулярних виразів ми одержимо для кожного рядка нашого зображення (досить довгого, що починається з білої точки) вихідний бітовий потік виду:

```
((<Б-2560>)*[<Б-сст.>]<Б-зв.>(<Ч-2560>)*[<Ч-сст.>]<Ч-зв.>)+  
[<Б-2560>]*[<Б-сст.>]<Б-зв.>]
```

де (*) – повтор 0 чи більш раз, (+) – повтор 1 чи більш раз, [] – включення 1 чи 0 разів.

Для наведеного прикладу: 0, 3, 556, 10... алгоритм сформує наступний код: <Б-0><Ч-3><Б-512><Б-44><Ч-10>, чи, відповідно до таблиці, 001101011001100101001011010000100 (різні коди в потоці виділені для зручності). Цей код має властивість префіксних кодів і легко може згорнутись назад у послідовність довжин серій. Легко підрахувати, що для приведення рядка в 569 біт ми одержали код довжиною в 33 біта, тобто коефіцієнт стиснення складає приблизно 17 разів.

Зображення, для яких дуже вигідне застосування алгоритму CCITT-3 – великі області заповнені одним кольором.

Зображення, для яких менш вигідне застосування алгоритму CCITT-3 – невелика кількість областей, заповнених одним кольором. Багато коротких “чорних” і “білих” серій).

Характеристики алгоритму CCITT Group 3:

- коефіцієнти компресії: кращий коефіцієнт знаходиться в межах до 213.(3), середній – 2, у гіршому випадку збільшує файл у 5 разів;
- клас зображень: двоколірні чорно-білі зображення, в яких переважають великі області, заповнені білим кольором;
- симетричність: близька до 1;
- характерні риси: алгоритм надзвичайно простий у реалізації, швидкий і може бути легко реалізований апаратно.

13.2.5 Алгоритм JBIG

Алгоритм розроблений групою експертів ISO (Joint Bi-level Experts Group) спеціально для стиснення однобітових чорно-білих зображень, наприклад, факсів сканованих документів. Може застосовуватися і до дво-, і до чотирибітових картинок. При цьому алгоритм розбиває їх на окремі бітові площини: JBIG дозволяє керувати такими параметрами, як порядок розбивки зображення на бітові площини, ширина смуг у зображенні, рівні масштабування. Остання можливість дозволяє легко орієнтуватися в базі великих за розмірами зображень, переглядаючи спочатку їхні зменшені копії. Надбудовуючи ці параметри, можна використовувати описаний ефект “примітивного зображення” при одержанні зображення по мережі або будь-якому іншому каналу, пропускна здатність якого мала в порівнянні з можливостями процесора. Розпаковуватися зображення на екрані буде поступово, ніби повільно “виявляючись”. При цьому людина починає аналізувати картинку задовго до кінця процесу розархівзації.

Алгоритм побудований на базі Q-кодера, патентом на який володіє IBM. Q-кодер, так само як і алгоритм Хаффмана, використовується частіше для символів з великими значеннями ймовірностей появи – короткі ланцюжки, а рідше – для символів з малими ймовірностями появи – довгі ланцюжки. Однак, на відміну від нього, в алгоритмі використовуються і послідовності символів.

13.2.6 Алгоритм Lossless JPEG

Цей алгоритм розроблений групою експертів в області фотографії (Joint Photographic Expert Group). На відміну від JBIG, Lossless JPEG орієнтований на повнокольорові 24-бітові чи 8-бітові в градациях сірого зображення без палітри. Він являє собою спеціальну реалізацію JPEG без втрат. Коефіцієнти стиснення: 20, 2, 1. Lossless JPEG рекомендується застосовувати в тих додатках, де необхідна побітова відповідність вихідного і декомпресованого зображень. Докладніше про алгоритм стиснення JPEG див. п. 13.3.2.

З одного боку, наведені алгоритми досить універсальні і покривають усі типи зображень, з іншого – у них, за сьогоднішніми мірками, занадто маленький коефіцієнт архівації. Використовуючи один з алгоритмів стиснення без втрат, можна забезпечити архівацію зображення приблизно в два рази. У той же час алгоритми стиснення з втратами оперують з коефіцієнтами 10–200. Крім можливості модифікації зображення, одна з основних причин подібної різниці полягає в тому, що традиційні алгоритми орієнтовані на роботу з ланцюжком. Вони не враховують, так звану, “когерентність областей” у зображеннях. Ідея когерентності областей полягає в невеликій зміні кольору і структури на невеликій ділянці зображення. Всі алгоритми, про які мова йтиме нижче, були створені пізніше спеціально для стиснення графіки і використовують цю ідею.

Задля справедливості слід зазначити, що й у класичних алгоритмах можна використовувати ідею когерентності. Існують алгоритми обходу зображення по “фрактальній” кривій, при роботі яких вона також витягається в ланцюжок; але за рахунок того, що крива обходить області зображення по складній траєкторії, ділянки близьких кольорів у ланцюжку, що виходить, подовжуються.

13.3 Алгоритми архівації з втратами

13.3.1 Загальні характеристики

Першими для архівації зображень стали застосовуватися звичайні алгоритми – ті, що використовувалися і використовуються в системах резервного копіювання, при створенні дистрибутивів і т.п. Ці алгоритми архівували інформацію без змін. Однак основною тенденцією останнім часом стало використання нових класів зображень. Старі алгоритми перестали задовольняти вимогам, що пропонуються для архівації. Багато зображень практично не стискалося, хоча “на погляд” мали явну надлишковість. Це привело до створення нового типу алгоритмів – стискаючих із втратою інформації. Як правило, коефіцієнт архівації a , отже, ступінь втрат якості в них, можна задавати. При цьому досягається компроміс між розміром і якістю зображень.

Одна із серйозних проблем машинної графіки полягає в тому що дотепер не знайдено адекватний критерій оцінювання втрат якості зображення. А губиться воно постійно – при оцифруванні, при переведенні в обмежену палітру кольорів, при переведенні в іншу систему кольорового представлення для друку, і, що для нас особливо важливо, при архівації з втратами. Можна навести приклад простого критерію: середньоквадратичне відхилення значень пікселів (L_2 міра, чи root mean square – RMS):

$$d(x, y) = \sqrt{\frac{\sum_{i=1, j=1}^{n, n} (x_{ij} - y_{ij})^2}{n^2}}$$

За цим критерієм зображення буде сильно зіпсовано при зниженні яскравості усього на 5 % (око цього не помітить – у різних моніторів настроювання яскравості варіюється набагато сильніше). У той же час зображення зі “снігом” – різкою зміною кольору окремих точок, слабкими смугами – “муаром” – будуть визнані такими, що “що майже не змінилися”. Свої негативні сторони є й в інших критеріях.

Розглянемо, наприклад, максимальне відхилення:

$$d(x, y) = \max_{i, j} |x_{ij} - y_{ij}|.$$

Ця міра вкрай чутлива до биття окремих пікселів. Тобто в зображенні може істотно змінитися тільки значення одного пікселя (що практично непомітно для ока), однак відповідно до цієї міри зображення буде сильно зіпсоване.

Міра, що тепер використовують на практиці, називається мірою відношення сигналу до шуму (peak-to-peak signal-to-noise ratio – PSNR).

$$d(x, y) = 10 \cdot \log_{10} \frac{255^2 \cdot n^2}{\sum_{i=1, j=1}^{n, n} (x_{ij} - y_{ij})^2}.$$

Ця міра, по суті, аналогічна середньоквадратичному відхиленню, однак користуватися нею трохи зручніше за рахунок логарифмічного масштабу шкали. Їй властиві ті самі недоліки, що і середньоквадратичному відхиленню.

Найкраще втрати якості зображень оцінюють наші очі. Відмінною вважається архівація, при якій неможливо на око розрізнити первісне і розархівоване зображення. Гарною вважається архівація, коли можна сказати, яке з зображень піддавалося архівації, тільки порівнюючи дві картинки, що знаходяться поруч. При подальшому збільшенні стиснення, як правило, стають помітні побічні ефекти, характерні для даного алгоритму. На практиці, навіть при відмінному збереженні якості, у зображення можуть бути внесені регулярні специфічні зміни. Тому алгоритми архівації з втратами не рекомендується використовувати при стисненні зображень, що надалі будуть або друкувати з високою якістю, або обробляти програмами розпізнавання образів. Негативні ефекти з такими зображеннями, як ми вже говорили, можуть виникнути навіть при простому масштабуванні зображення.

13.3.2 Алгоритм JPEG

JPEG – один із самих нових і досить могутніх алгоритмів. Практично він є стандартом де-факто для повнокольорових зображень. Оперує алгоритм областями 8×8 , на яких яскравість і колір міняються порівняно плавно. Внаслідок цього, при розкладанні матриці такої області в подвійний ряд по косинусах значимими виявляються тільки перші коефіцієнти. Таким чином, стиснення в JPEG здійснюється за рахунок плавності зміни кольорів у зображенні.

Алгоритм розроблений групою експертів в області фотографії спеціально для стиснення 24-бітових зображень. JPEG – Joint Photographic Expert Group – підрозділ у рамках ISO – Міжнародної організації по стандартизації. У цілому алгоритм заснований на дискретному косинусоїдальному перетворенні (ДКП), застосовуваному до матриці зображення для одержання деякої нової матриці коефіцієнтів. Для одержання вихідного зображення застосовується зворотнє перетворення.

ДКП розкладає зображення по амплітудах деяких частот. Таким чином, при перетворенні ми одержуємо матрицю, у якій багато коефіцієнтів або близькі, або дорівнюють нулю. Крім того, завдяки недосконалості людського зору, можна апроксимувати коефіцієнти більш грубо без помітної втрати якості зображення.

Для цього використовується квантування коефіцієнтів. У найпростішому випадку – це арифметичний побітовий зсув вправо. При цьому перетворенні губиться частина інформації, але можуть досягатися великі коефіцієнти стиснення.

Отже, розглянемо алгоритм докладніше. Нехай ми стискаємо 24-бітове зображення.

Крок 1. Переводимо зображення з кольорового формату RGB, з компонентами, що відповідають за червону (Red), зелену (Green) і синю (Blue) складові кольору точок, у кольоровий формат $YCrCb$.

У цьому Y – складова яскравості, а Cr , Cb – компоненти, що відповідають за колір (хроматичний червоний і хроматичний синій). За рахунок того, що людське око менш чутливе до кольору, ніж до яскравості, з'являється можливість архівувати масиви для Cr і Cb компонент із великими втратами і, відповідно, великими коефіцієнтами стиснення. Подібне перетворення вже давно використовується в телебаченні. На сигналі, що відповідають за колір, там виділяється більш вузька смуга частот.

Спрощено переведення з кольорового формату RGB у кольоровий формат $YCrCb$ можна представити за допомогою матриці переходу:

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0,299 & 0,587 & 0,144 \\ 0,5 & -0,4187 & -0,0813 \\ 0,1687 & -0,3313 & 0,5 \end{pmatrix} \cdot \left(\begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix} \right)$$

Зворотне перетворення здійснюється множенням вектора $YCrCb$ на зворотну матрицю.

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1,402 \\ 1 & -0,34414 & -0,71414 \\ 1 & 1,772 & 0 \end{pmatrix} \cdot \begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix}.$$

Крок 2. Розбиваємо вихідне зображення на матриці 8×8 . Формуємо з кожної три робочі матриці ДКП – по 8 біт окремо для кожного компонента. При великих коефіцієнтах стиснення цей крок може виконуватися складніше. Зображення поділяється по компоненті Y – як і в першому випадку, а для компонентів Cr і Cb матриці набираються через рядок і через стовпчик. Тобто з вихідної матриці розміром 16×16 виходить тільки одна робоча матриця ДКП. При цьому, як неважко помітити, ми втрачаємо $3/4$ корисної інформації про кольорові складові зображення й одержуємо відразу стиснення у два рази. Ми можемо робити так завдяки роботі в просторі $YCrCb$. На результуючому RGB зображенні, як показала практика, це суттєво не позначається.

Крок 3. Застосовуємо ДКП до кожної робочої матриці. При цьому ми одержуємо матрицю, у якій коефіцієнти в лівому верхньому куті відповідають низькочастотній складовій зображення, а в правому нижньому – високочастотній.

У спрощеному вигляді це перетворення можна представити так:

$$Y[u, v] = \frac{1}{4} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} C(i, u) \cdot C(j, v) \cdot y[i, j],$$

$$C(i, u) = A(u) \cdot \cos\left(\frac{(2x_i + 1) \cdot u \cdot \pi}{2\pi}\right),$$

де

$$A(u) = \begin{cases} \frac{1}{\sqrt{2}} & \text{для } u = 0; \\ 1 & \text{для } u \neq 0. \end{cases}$$

Крок 4. Робимо квантування. У принципі, це просто розподіл робочої матриці на матрицю квантування поелементно. Для кожного компонента (Y , U і V) задається своя матриця квантування $q[u, v]$ (далі МК):

$$Y_q[u, v] = \text{IntegerRound}\left(\frac{Y[u, v]}{q[u, v]}\right).$$

На цьому кроці здійснюється керування ступенем стиснення і відбуваються найбільші втрати. Задаючи МК із великими коефіцієнтами, ми одержимо більше нулів і, отже, великий ступінь стиснення.

У стандарт JPEG включені рекомендовані МК, побудовані дослідним шляхом. Матриці для більшого чи меншого коефіцієнту стиснення одержують шляхом множення вихідної матриці на деяке число gamma.

З квантуванням пов'язані і специфічні ефекти алгоритму. При великих значеннях коефіцієнта gamma втрати на низьких частотах можуть бути настільки великі, що зображення розпадеться на квадрати 8×8 . Втрати на високих частотах можуть проявитися у так званому "ефекті Гіббса", коли навколо контурів з різким переходом кольору утвориться своєрідний "німб".

Крок 5. Переводимо матрицю 8×8 у 64-елементний вектор за допомогою "зигзаг"-сканування (рис. 13.1), тобто беремо елементи з індексами (0,0), (0,1), (1,0), (2,0)... .

| | | | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $a_{0,0}$ | $a_{0,1}$ | $a_{0,2}$ | $a_{0,3}$ | $a_{0,4}$ | $a_{0,5}$ | $a_{0,6}$ | $a_{0,7}$ |
| $a_{1,0}$ | $a_{1,1}$ | $a_{1,2}$ | $a_{1,3}$ | $a_{1,4}$ | $a_{1,5}$ | $a_{1,6}$ | |
| $a_{2,0}$ | $a_{2,1}$ | $a_{2,2}$ | $a_{2,3}$ | $a_{2,4}$ | | | |
| $a_{3,0}$ | $a_{3,1}$ | $a_{3,2}$ | $a_{3,3}$ | | | | |
| $a_{4,0}$ | $a_{4,1}$ | $a_{4,2}$ | | | | | |
| $a_{5,0}$ | $a_{5,1}$ | | | | | | |
| $a_{6,0}$ | $a_{6,1}$ | | | | | | |
| $a_{7,0}$ | | | | | | | |

Рис. 13.1

Таким чином, на початку вектора ми одержуємо коефіцієнти матриці, що відповідають низьким частотам, а наприкінці – високим.

Крок 6. Згортаємо вектор за допомогою алгоритму групового кодування. При цьому одержуємо пари типу (пропустити, число), де "пропустити" є лічильником нулів, що пропускаються, а "число" – значення, які необхідно поставити в наступний осередок. Так, вектор 42 3 0 0 0 – 2 0 0 0 0 1 ... буде згорнутий у пари (0,42) (0,3) (3,-2) (4,1)... .

Крок 7. Згортаємо пари, що вийшли, кодуванням за Хаффманом з фіксованою таблицею.

Процес відновлення зображення в цьому алгоритмі цілком симетричний. Метод дозволяє стискати деякі зображення у 10–15 разів без серйозних втрат.

Позитивні сторони алгоритму:

- задається ступінь стиснення;
- вихідне кольорове зображення може мати 24 біта на точку.

Негативні сторони алгоритму:

– при підвищенні ступеня стиснення зображення розпадається на окремі квадрати (8×8). Це пов'язано з тим, що відбуваються великі втрати в низьких частотах при квантуванні, і відновити вихідні дані стає неможливо;

– виявляється ефект Гіббса – ореоли по границях різких переходів кольорів.

Характеристики алгоритму JPEG:

- коефіцієнти компресії: 2...200 (задається користувачем);
- клас зображень: повнокольорові 24-бітові зображення без різких переходів кольорів (фотографії) чи зображення в градаціях сірого;
- симетричність: 1;
- характерні риси: у деяких випадках, алгоритм створює “ореол” навколо різких горизонтальних і вертикальних границь у зображенні (ефект Гіббса); крім того, при високому ступені стиснення зображення розпадається на блоки 8×8 пікселів.

13.3.3 Фрактальний алгоритм

Фрактальна архівація заснована на тому, що ми представляємо зображення в більш компактній формі – за допомогою коефіцієнтів системи функцій, що інтегруються (Iterated Function System – далі IFS). Перш, ніж розглядати сам процес архівації, розглянемо, як IFS буде зображення, тобто процес декомпресії.

Строго говорячи, IFS являє собою набір тривимірних афінних перетворень, у нашому випадку тих, що перетворюють одне зображення в інше. Перетворенню піддаються точки у тривимірному просторі (x-координата, у-координата, яскравість).

Найбільше наочно цей процес продемонстрував Барнслі у книзі “Fractal Image Compression”. Там введено поняття фотокопіювальної машини, що складається з екрана, на якому зображена вихідна картинка, і системи лінз, що проєктують зображення на інший екран.

Лінзи можуть проєктувати частини зображення довільної форми в будь-яке інше місце нового зображення. Області, в які проєктуються зображення, не перетинаються. Лінза може змінювати яскравість і зменшувати контрастність. Лінза може дзеркально відбивати і повертати свій фрагмент зображення. Лінза повинна масштабувати (зменшувати) свій фрагмент зображення.

Розставляючи лінзи і змінюючи їхні характеристики, можна керувати одержуваням зображенням. Одна ітерація роботи машини

полягає в тому, що за вихідним зображенням за допомогою проектування будується нове, після чого нове береться за вихідне. Стверджується, що в процесі ітерацій ми одержимо зображення, що перестане змінюватися. Воно буде залежати тільки від розташування та характеристик ліній і не буде залежати від вихідної картинки. Це зображення називається “нерухомою точкою” чи аттрактором даної IFS. Відповідна теорія гарантує наявність рівно однієї нерухомої точки для кожної IFS.

Оскільки відображення ліній є стискаючим, кожна лінійка в явному вигляді задає самоподібні області в нашому зображенні. Завдяки самоподібності ми одержуємо складну структуру зображення при будь-якому збільшенні. Таким чином, інтуїтивно зрозуміло, що система функцій, що інтегрується задає фрактал (нестрого – самоподібний математичний об’єкт).

У віршому випадку, якщо не буде застосовуватися оптимізуючий алгоритм, будуть потрібні перебір і порівняння всіх можливих фрагментів зображення різного розміру. Навіть для невеликих зображень при обліку дискретності ми одержимо астрономічне число варіантів, що перебираються. Причому, навіть різке звуження класів перетворень, наприклад, за рахунок масштабування тільки у визначену кількість разів, не дає помітного виграшу в часі. Крім того, при цьому зникає якість зображення. Переважна більшість досліджень в області фрактальної компресії тепер спрямовані на зменшення часу архівації, необхідного для одержання якісного зображення.

Побудова алгоритму. Основною задачею при компресії фрактальним алгоритмом є наявність відповідних афінних перетворень. І взагалі ми можемо переводити будь-які за розмірами і формою області зображення, однак у цьому випадку виходить астрономічна кількість варіантів різних фрагментів, що перебираються, які на даний момент неможливо обробити навіть на суперкомп’ютері.

Для побудови алгоритмів зроблені наступні обмеження на області:

- всі області є квадратами зі сторонами, кратними сторонам зображення;
- при переведенні доменної області в рангову зменшення розмірів відбувається рівно в два рази;
- усі доменні блоки – квадрати і мають фіксований розмір; зображення рівномірною сіткою розбивається на набір доменних блоків;
- доменні області беруться “через точку” і по X , і по Y , що відразу зменшує перебір у 4 рази;
- при переведенні доменної області в рангову поворот квадрата можливий тільки на 0° , 90° , 180° чи 270° . Також допускається дзеркальне відображення. Загальна кількість можливих перетворень (враховуючи порожнє) – 8;

– масштабування (стиснення) по вертикалі (яскравості) здійснюється у фіксоване число раз – 0,75.

Ці обмеження дозволяють:

1) побудувати алгоритм, для якого потрібно порівняно мале число операцій навіть на досить великих зображеннях.

2) дуже компактно представити дані для запису у файл. Нам потрібно на кожне афінне перетворення в IFS:

– два числа для того, щоб задати зсув доменного блоку. Якщо ми обмежимо вхідні зображення розміром 512×512 , то досить буде по 8 біт на кожне число;

– 3 біт для того, щоб задати перетворення симетрії при переведенні доменного блоку в ранговий;

– 7–9 біт для того, щоб задати зсуви за яскравістю при переведенні.

Інформацію про розмір блоків можна зберігати в заголовку файла. Таким чином, ми затратили менше чотирьох байт на одне афінне перетворення. Залежно від розміру блоку, можна вирахувати, скільки блоків буде в зображенні. Таким чином, ми можемо одержати оцінку ступеня компресії.

Наприклад, для файла в градациях сірого 256 квітів 512×512 пікселів при розмірі блоку 8 пікселів афінних перетворень буде 4096 ($512/8 \times 512/8$). На кожне буде потрібно 3,5 байта. Отже, якщо вихідний файл займав 262144 (512×512) байт (без обліку заголовка), то файл із коефіцієнтами буде займати 14336 байт. Коефіцієнт архівації – 18 разів. При цьому ми не враховуємо, що файл із коефіцієнтами теж може мати надлишковість і архівуватися методом архівації без втрат, наприклад, LZW.

Негативні сторони запропонованих обмежень:

– оскільки всі області є квадратами, неможливо скористатися подобою об'єктів, за формою далеких від квадратів (які зустрічаються в реальних зображеннях досить часто);

– аналогічно ми не зможемо скористатися подібністю об'єктів у зображенні, коефіцієнт подібності між якими сильно відрізняється від 2;

– алгоритм не зможе скористатися подібністю об'єктів у зображенні, кут між якими не кратний 90° .

Така плата за швидкість компресії та за простоту увакування коефіцієнтів у файл.

У цьому алгоритмі, для кожного рангового блоку робимо його перевірку з усіма можливими домешними блоками (у тому числі з попередніми перетвореннями симетрії), знаходимо варіант із найменшою мірою L_2 (найменшим середньоквадратичним відхиленням) і зберігаємо коефіцієнти цього перетворення у файл. Коефіцієнти – це (1) координати знайденого блоку, (2) число від 0 до 7, що характеризує перетво-

рення симетрії (поворот, відображення блоку), і (3) зсув по яскравості для цієї пари блоків. Зсув по яскравості обчислюється як:

$$q = \frac{\sum_{i=1}^n \sum_{j=1}^n d_{i,j} - \sum_{i=1}^n \sum_{j=1}^n r_{i,j}}{n^2},$$

де $r_{i,j}$ – значення пікселів рангового блоку R , а $d_{i,j}$ – значення пікселів доменного блоку D . При цьому міра знаходиться як:

$$d(R, D) = \sum_{i=1}^n \sum_{j=1}^n (0,75 \cdot r_{i,j} + q - d_{i,j})^2.$$

Ми не обчислюємо квадратного кореня з L_2 міри і не ділимо її на n , оскільки дані перетворення монотонні і не перешкоджають нам знайти екстремум, однак ми зможемо виконувати на дві операцій менше для кожного блоку.

Для декомпресії алгоритму фрактального стиснення необхідно провести декілька ітерацій тривимірних афінних перетворень, коефіцієнти яких були отримані на етапі компресії.

Характеристики фрактального алгоритму:

- коефіцієнти компресії: 2–2000 (задається користувачем);
- клас зображень: повнокольорові 24-бітові зображення без різких переходів кольорів (фотографії) чи зображення в градациях сірого; бажано, щоб області більшої значимості (для сприйняття) були більш контрастними і різкими, а області меншої значимості – неконтрастними і розмитими;
- симетричність: 100–100000;
- характерні риси: може вільно масштабувати зображення при розархівачії, збільшуючи його в 2–4 рази без проявлення “сходинкового ефекту”; при збільшенні ступеня компресії з’являється “блочний” ефект на границях блоків у зображенні.

13.3.4 Рекурсивний (хвильовий) алгоритм

Англійська назва рекурсивного стиснення – wavelet. Українською мовою вона перекладається як “хвильове стиснення”, і як “стиснення із використанням сплесків”. Цей вид архівації відомий досить давно і прямо виходить з ідеї використання когерентності областей. Орієнтовано алгоритм на кольорові і чорно-білі зображення з плавними переходами. Ідеальний для картинок типу рентгенівських знімків. Коефіцієнт стиснення задається і варіюється в межах 5–100. При спробі задати більший коефіцієнт на різких границях, що особливо проходять по діагоналі, виявляється “сходинковий ефект” – сходинки різної яскравості розміром у декілька пікселів.

Ідея алгоритму полягає в тому, що ми зберігаємо у файлі різницю – число між середніми значеннями сусідніх блоків у зображенні, що звичайно приймає значення, близькі до 0.

Так два числа a_{2i} та a_{2i+1} завжди можна представити у вигляді:

$$b_i^1 = \frac{a_{2i} + a_{2i+1}}{2} \quad \text{та} \quad b_i^2 = \frac{a_{2i} - a_{2i+1}}{2}.$$

Аналогічно послідовність a_i може бути попарно переведена в послідовність $b_i^{1,2}$.

Розглянемо конкретний приклад: нехай ми стискаємо рядок з 8 значень яскравості пікселів (a_i): (220, 211, 212, 218, 217, 214, 210, 202). Ми одержимо наступні послідовності b_i^1 і b_i^2 : (215.5, 215, 215.5, 206) і (4.5, -3, 1.5, 4). Значення b_i^2 досить близькі до 0. Повторимо операцію, розглядаючи b_i^1 як a_i . Дана дія виконується ніби рекурсивно, звідки і пішла назва алгоритму. Ми одержимо з (215.5, 215, 215.5, 206): (215.25, 210.75) (0.25, 4.75). Отримані коефіцієнти, округливши їх до цілих і стиснувши, наприклад, за допомогою алгоритму Хаффмана з фіксованими таблицями, ми можемо помістити у файл.

Ми застосовували перетворення до ланцюжка тільки два рази. Реально ми можемо дозволити собі застосування wavelet-перетворення 4–6 разів. Більш того, додаткове стиснення можна одержати, використовуючи таблиці алгоритму Хаффмана з нерівномірним кроком (тобто доведеться зберігати код Хаффмана для найближчого в таблиці значення). Ці прийоми дозволяють досягти помітних коефіцієнтів стиснення.

Алгоритм для двовимірних даних реалізується аналогічно. Якщо в нас є квадрат з чотирьох точок з яскравістю $a_{2i,2j}$, $a_{2i+1,2j}$, $a_{2i,2j+1}$ і $a_{2i+1,2j+1}$, то:

$$b_{i,j}^1 = \frac{a_{2i,2j} + a_{2i+1,2j} + a_{2i,2j+1} + a_{2i+1,2j+1}}{4},$$

$$b_{i,j}^2 = \frac{a_{2i,2j} + a_{2i+1,2j} - a_{2i,2j+1} - a_{2i+1,2j+1}}{4},$$

$$b_{i,j}^3 = \frac{a_{2i,2j} - a_{2i+1,2j} + a_{2i,2j+1} - a_{2i+1,2j+1}}{4},$$

$$b_{i,j}^4 = \frac{a_{2i,2j} - a_{2i+1,2j} - a_{2i,2j+1} + a_{2i+1,2j+1}}{4}.$$

Використовуючи ці формули, ми для зображення 512×512 пікселів одержимо після першого перетворення чотири матриці розміром 256×256 елементів.

У першій зберігається зменшена копія зображення. У другій – усереднені різниці пар значень пікселів по горизонталі. У третій – усереднені різниці пар значень пікселів по вертикалі. У четвертій – усереднені різниці значень пікселів по діагоналі. За аналогією з двовимірним випадком ми можемо повторити наше перетворення й одержати замість першої матриці чотири матриці розміром 128×128 . Повторивши наше перетворення втретє, ми одержимо в підсумку: чотири матриці 64×64 , три матриці 128×128 і три матриці 256×256 . На практиці при записі у файл, значеннями, одержуваними в останньому рядку (b_{ij}^4), зазвичай нехтують (відразу одержуючи виграти приблизно на третину розміру файла – $1-1/4-1/16-1/64\dots$).

До переваг цього алгоритму можна віднести те, що він дуже легко дозволяє реалізувати можливість поступового “проявлення” зображення при передачі зображення по мережі. Крім того, оскільки на початку зображення ми фактично зберігаємо його зменшену копію, спрощується показ “примітивного” зображення по заголовку.

На відміну від JPEG і фрактального алгоритму даний метод не оперує блоками, наприклад, 8×8 пікселів. Точніше, ми оперуємо блоками 2×2 , 4×4 , 8×8 і т.д. Однак за рахунок того, що коефіцієнти для цих блоків ми зберігаємо незалежно, ми можемо досить легко уникнути подрібнення зображення на “мозайчні” квадрати.

Характеристики хвильового алгоритму:

- коефіцієнти компресії: 2–200 (задається користувачем);
- клас зображень: як у фрактального і JPEG;
- симетричність: приблизно 1,5;
- характерні риси: при високому ступені стиснення зображення розпадається на окремі блоки.

13.4 MPEG: загальна інформація

Стандарт стиснення MPEG розроблений Експертною групою кінематографії (Moving Picture Experts Group – MPEG). MPEG – це стандарт на стиснення звукових і відеофайлів у формат, більш зручний для завантаження чи пересилання, наприклад, через Internet. Існують різні стандарти MPEG: MPEG-1, MPEG-2, MPEG-3, MPEG-4, MPEG-7. MPEG складається з трьох частин: Audio, Video, System (об'єднання і синхронізація двох інших).

Таблиця 13.1

| Вид формату | Відношення розрізнення | |
|-------------|------------------------|---------------------|
| | по горизонталі (Cb/Y) | по вертикалі (Cr/Y) |
| 4 : 4 : 4 | 1 : 1 | 1 : 1 |
| 4 : 2 : 2 | 1 : 2 | 1 : 1 |
| 4 : 2 : 0 | 1 : 2 | 1 : 2 |
| 4 : 1 : 1 | 1 : 4 | 1 : 1 |
| 4 : 1 : 0 | 1 : 4 | 1 : 4 |

Для більшого стиснення у В- і Р-кадрах використовується алгоритм передбачення руху (що дозволяє сильно зменшити розмір Р- і В-кадрів, табл. 13.2) на виході якого виходить:

– вектор зсуву (вектор руху) блоку, який потрібно передбачити щодо базового блоку.

– різниця між блоками (яка потім і кодується).

Оскільки не всякий блок можна передбачити на підставі інформації про попередній, то в Р- і В-кадрах можуть знаходитися І-блоки (блоки без пророкування руху).

Таблиця 13.2

| Вид кадру | I | P | B | Середній розмір |
|---|-----|----|----|-----------------|
| Розмір кадру для стандарту SIF(kilobit) | 150 | 50 | 20 | 38 |

Метод кодування блоків (або різниці, одержуваної при методі передбачення руху) містить у собі:

– Discrete Cosine Transforms (DCT – дискретне перетворення косинусів);

– Quantization (перетворення даних з безупинної форми в дискретну);

– кодування отриманого блоку в послідовність.

DCT використовує той факт, що піксели в блоці і самі блоки зв'язані між собою (тобто корельовані), тому відбувається розбиття на частотні Фур'є-компоненти (у підсумку виходить quantization matrix – матриця перетворень даних з безупинної в дискретну форму, числа в якій є величиною амплітуди відповідної частоти), потім алгоритм Quantization розбиває частотні коефіцієнти на визначену кількість значень. Encoder (кодер) вибирає quantization matrix яка визначає те, як кожен частотний коефіцієнт у блоці буде розбитий (людина більш чутлива до дискретності розбиття для малих частот, ніж для великих). Оскільки в процесі quantization багато коефіцієнтів виходять нульовими, то застосовується алгоритм зигзага для одержання довгих послідовностей нулів (рис. 13.3).

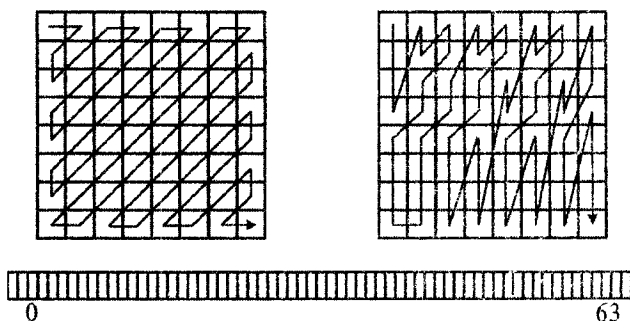


Рис. 13.3

13.4.2 Звук у MPEG

Формати кодування звуку діляться на три частини: Layer I, Layer II, Layer III (прообразом для Layer I і Layer II став стандарт MUSICAM, цим ім'ям тепер іноді називають Layer II). Layer III досягає самого великого стиснення, але, відповідно, вимагає більше ресурсів на кодування.

Принципи кодування засновані на тому факті, що людське вухо недосконале і насправді в нестиснутому звуці (CD-audio) передається багато надлишкової інформації. Принцип стиснення ґрунтується на ефектах маскування деяких звуків для людини (наприклад, якщо йде сильний звук на частоті 1000 Гц, то більш слабкий звук на частоті 1100 Гц уже не буде чутний людині, також буде ослаблена чутливість людського вуха на період у 100 мс після і 5 мс до виникнення сильного звуку).

Psycoacoustic (психоакустична) модель, використовувана в MPEG, розбиває весь частотний спектр на частини, у яких рівень звуку вважається однаковим, а потім видаляє звуки, які не сприймаються людиною, завдяки описаним вище ефектам.

У Layer III частини розбитого спектра самі маленькі, що забезпечує найкраще стиснення. MPEG Audio підтримує сумісність шарів знизу вгору, тобто декодер для Layer II буде також розпізнавати Layer I.

Синхронізація та об'єднання звуку і відео, здійснюється за допомогою системного потоку (рис. 13.4), що містить у собі шари:

- системний, з тимчасовою та іншою інформацією, яка необхідна для розділення і синхронізації відео та аудіо;
- компресійний, з відео- та аудіопотоками.

Відеопотік (рис. 13.5) має заголовки, потім кілька груп картинок (заголовки і кілька картинок необхідні для того, щоб забезпечити довільний доступ до картинок у групі незалежно від їхнього порядку).

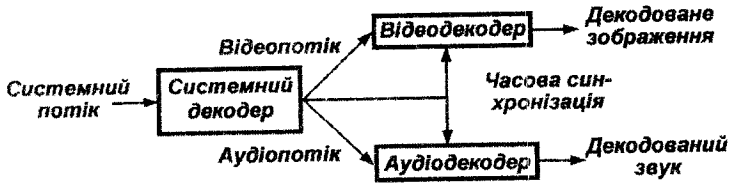


Рис. 13.4

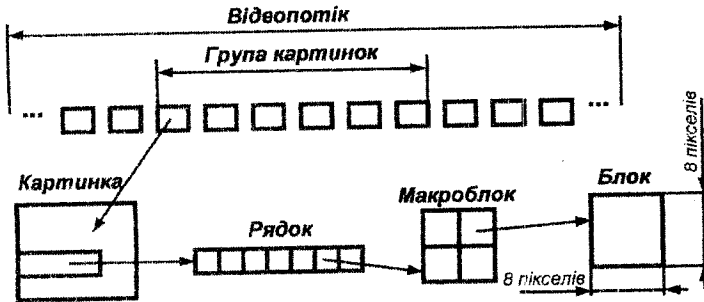


Рис. 13.5

Аудіопотік складається з пакетів, кожний з яких складається з заголовка і декількох звукових кадрів (audio-frame).

Для синхронізації аудіо- і відеопотоків у системний потік вбудовується таймер, що працює з частотою 90 КГц (System Clock Reference – SCR – мітка, за якою відбувається збільшення часового лічильника в декодері) і Presentation Data Stamp (PDS, мітка початку відтворення, вставляються в картинку чи в звуковий кадр, щоб вказати декодеру, коли їх відтворювати). Розмір PDS складає 33 біта, що забезпечує можливість представлення будь-якого тимчасового циклу довжиною до 24 годин).

Компресія за стандартом MPEG-2 кардинально змінює положення речей. Більш 97 % цифрових даних, що представляють відеосигнал, дублюються, тобто є надлишковими і можуть бути стиснуті без втрати якості зображення. Алгоритм MPEG-2 аналізує відеозображення в пошуках повторень, що називаються надлишковістю. У результаті процесу видалення надлишковості, забезпечується якість відеозображення у форматі MPEG-2 при більш низькій швидкості передачі даних.

MPEG-4 – стандарт для низькошвидкісної передачі даних (64 kbps). Перша версія випущена в 1999 р.

Короткий опис:

– розділяє картинку на різні елементи, називані media objects (медіа об'єкти);

– описує структуру цих об'єктів та їхнього взаємозв'язку, щоб потім зібрати їх у відеозвукову сцену;

– дозволяє змінювати сцену, що забезпечує високий рівень інтерактивності для кінцевого користувача.

Відеозвукова сцена складається з медіа-об'єктів, що об'єднані в ієрархічну структуру:

– нерухомі картинки (наприклад, фон);

– відеооб'єкти (людина, що говорить);

– аудіооб'єкти (голос, зв'язаний з цією людиною);

– текст, зв'язаний з даною сценою;

– синтетичні об'єкти – об'єкти, яких спочатку не було в записуваній сцені, але які туди додаються при демонстрації кінцевому користувачу (наприклад, синтезується голова, що говорить),

– текст, зв'язаний з головою, з якою наприкінці синтезується голос.

Такий спосіб представлення даних дозволяє:

– переміщати і поміщати медіа-об'єкти в будь-яке місце сцени;

– трансформувати об'єкти, змінювати геометричні розміри;

– збирати з окремих об'єктів складений об'єкт і проводити над ним будь-які операції;

– змінювати текстуру об'єкта (наприклад, колір), маніпулювати об'єктом (змусити шухляду пересуватися по сцені);

– змінювати точку спостереження за сценою.

MPEG-7 – не є продовженням MPEG як такого – розроблений порівняно недавно. MPEG-7 забезпечує стандарт для опису різних типів мультимедійної інформації (а не для її кодування), щоб забезпечити ефективний і швидкий її пошук. MPEG-7 офіційно називають “Multimedia Content Description Interface” (Інтерфейс опису мультимедіа даних). MPEG-7 визначає стандартний набір дескрипторів для різних типів мультимедіа інформації, так само він стандартизує спосіб визначення своїх дескрипторів і їхнього взаємозв'язку (description schemes). Для цієї мети MPEG-7 вводить DDL (Description Definition Language – мова опису визначень). Основна мета застосування MPEG-7 – це пошук мультимедіа інформації (так само ми можемо знайти текст за якою-небудь пропозицією), наприклад:

– музика: зігравши кілька нот на клавіатурі, можна одержати список музичних творів, що містять дану послідовність;

– графіка: намалювавши кілька ліній на екрані, одержимо набір малюнків, які містять даний фрагмент;

- картини: визначивши об'єкт (задавши його форму і текстуру), одержимо список картин, що містять його;
- відео: задавши об'єкт і рух, одержимо набір чи відео анімації;
- голос: задавши фрагмент голосу співака, одержимо набір пісень і відеороликів, де він співає.

MHEG (Multimedia & Hypermedia Expert Group – експертна група з мультимедіа і гіпермедіа) визначає стандарт для обміну мультимедійними об'єктами (відео, звук, текст та інші довільні дані) між додатками і передачі їх різними способами (локальна мережа, мережі телекомунікацій і передач) з використанням MHEG object classes. Він дозволяє програмним об'єктам містити в собі будь-яку систему кодування (наприклад, MPEG), що визначена в базовому додатку. MHEG був прийнятий DAVIC (Digital Audio-Visual Council – Радою з цифрового відео і звуку). MHEG-об'єкти робляться мультимедіа додатками, використовуючи multimedia scripting languages.

Стверджується, що MHEG – майбутній міжнародний стандарт для інтерактивного TV, тому що він працює на будь-яких платформах і його документація вільно розповсюджується.

Контрольні питання

1. Які класи зображень ви знаєте?
2. У чому особливість програм архівації (розархівації)?
3. Які ви знаєте критерії порівняння програм архівації (розархівації)?
4. Наведіть порівняльну характеристику алгоритмів RLE і LZV?
5. Наведіть порівняльну характеристику алгоритмів архівації зображень із втратами інформації?
6. Які переваги фрактального алгоритму архівації?
7. Наведіть загальні риси стандарту MPEG?

ЛІТЕРАТУРА

1. Кузьмин И. В. Основы теории информации и кодирования / И. В. Кузьмин, В. А. Келрус. – 2-е изд., перераб. и доп. – К. : Вища школа, 1986. – 238 с.
2. Финк Л. М. Сигналы, помехи, ошибки. Заметки о некоторых неожиданностях, парадоксах и заблуждениях в теории связи / Л. М. Финк. – 2-е изд., перераб. и доп. – М. : Радио и связь, 1984. – 256 с.
3. Основы цифровой обработки сигналов : курс лекций / А. И. Солонина [и др.]. – СПб. : БХВ-Петербург, 2003.
4. Троцишин І. В. Вимірювання та перетворення фазочастотних параметрів радіосигналів / І. В. Троцишин. – Хмельницький : ПП Ковальський В.В., 2002. – 382 с.
5. Метрологическое обеспечение систем передачи : учебное пособие для вузов / Б. П. Хромой [и др.] ; под ред. проф. Б. П. Хромого. – М. : Радио и связь, 1991. – 392 с.
6. Сиберт У. М. Цепи, сигналы системы : в 2 ч. / У. М. Сиберт ; пер. с англ. – М. : Мир, 1988.
7. Гоноровский С. И. Радиотехнические цепи и сигналы : учебное пособие для вузов / С. И. Гоноровский, М. П. Демин. – М. : Радио и связь, 1994.
8. Голд Б. Цифровая обработка сигналов / Б. Голд, Ч. Райдер ; под. ред. А. М. Трахтмана ; пер. с англ. – М. : Сов. радио, 1973. – 363 с.
9. Сергиенко А. Б. Цифровая обработка сигналов : учебник для вузов / А. Б. Сергиенко. – 2-е изд. – СПб. : Питер, 2006. – 751 с.
10. Чуи К. Введение в вейвлеты / К. Чуи ; пер. с англ. – М. : Мир, 2001. – 412 с.
11. Айфичер Э. Цифровая обработка сигналов : практический подход / Э. Айфичер, Б. Джервис ; пер. с англ. – 2-е изд. – М. : Вильямс, 2004.
12. Когновицкий О. С. Основы циклических кодов : учебное пособие / О.С. Когновицкий. – Л. : ЛЭИС, 1990.
13. Питерсон У. Коды, исправляющие ошибки / У. Питерсон, Э. Уэлдон ; пер с англ. – М. : Мир, 1976.
14. Ватолин Д. С. Алгоритмы сжатия изображений : методическое пособие / Д. С. Ватолин. – М. : Издательский отдел факультета вычислительной математики и кибернетики МГУ им. М.В. Ломоносова, 1999. – 76 с.

ДОДАТОК А

Значення функції $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$.

Таблиця А.1

| x | Φ(x) | x | Φ(x) | x | Φ(x) | x | Φ(x) |
|------|--------|------|--------|------|--------|------|--------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0,00 | 0,0000 | | | | | | |
| 0,01 | 0,0040 | 0,31 | 0,1217 | 0,61 | 0,2291 | 0,91 | 0,3186 |
| 0,02 | 0,0080 | 0,32 | 0,1255 | 0,62 | 0,2324 | 0,92 | 0,3212 |
| 0,03 | 0,0120 | 0,33 | 0,1293 | 0,63 | 0,2357 | 0,93 | 0,3238 |
| 0,04 | 0,0160 | 0,34 | 0,1331 | 0,64 | 0,2389 | 0,94 | 0,3264 |
| 0,05 | 0,0199 | 0,35 | 0,1368 | 0,65 | 0,2422 | 0,95 | 0,3289 |
| 0,06 | 0,0239 | 0,36 | 0,1406 | 0,66 | 0,2454 | 0,96 | 0,3315 |
| 0,07 | 0,0279 | 0,37 | 0,1443 | 0,67 | 0,2486 | 0,97 | 0,3340 |
| 0,08 | 0,0319 | 0,38 | 0,1480 | 0,68 | 0,2517 | 0,98 | 0,3365 |
| 0,09 | 0,0359 | 0,39 | 0,1517 | 0,69 | 0,2549 | 0,99 | 0,3389 |
| 0,10 | 0,0398 | 0,40 | 0,1554 | 0,70 | 0,2580 | 1,00 | 0,3413 |
| 0,11 | 0,0438 | 0,41 | 0,1591 | 0,71 | 0,2611 | 1,01 | 0,3438 |
| 0,12 | 0,0478 | 0,42 | 0,1628 | 0,72 | 0,2642 | 1,02 | 0,3461 |
| 0,13 | 0,0517 | 0,43 | 0,1664 | 0,73 | 0,2673 | 1,03 | 0,3485 |
| 0,14 | 0,0557 | 0,44 | 0,1700 | 0,74 | 0,2703 | 1,04 | 0,3508 |
| 0,15 | 0,0596 | 0,45 | 0,1736 | 0,75 | 0,2734 | 1,05 | 0,3531 |
| 0,16 | 0,0636 | 0,46 | 0,1772 | 0,76 | 0,2764 | 1,06 | 0,3554 |
| 0,17 | 0,0675 | 0,47 | 0,1808 | 0,77 | 0,2794 | 1,07 | 0,3577 |
| 0,18 | 0,0714 | 0,48 | 0,1844 | 0,78 | 0,2823 | 1,08 | 0,3599 |
| 0,19 | 0,0753 | 0,49 | 0,1879 | 0,79 | 0,2852 | 1,09 | 0,3621 |
| 0,20 | 0,0793 | 0,50 | 0,1915 | 0,80 | 0,2881 | 1,10 | 0,3643 |
| 0,21 | 0,0832 | 0,51 | 0,1950 | 0,81 | 0,2910 | 1,11 | 0,3665 |
| 0,22 | 0,0871 | 0,52 | 0,1985 | 0,82 | 0,2939 | 1,12 | 0,3686 |
| 0,23 | 0,0910 | 0,53 | 0,2019 | 0,83 | 0,2967 | 1,13 | 0,3708 |
| 0,24 | 0,0948 | 0,54 | 0,2054 | 0,84 | 0,2995 | 1,14 | 0,3729 |
| 0,25 | 0,0987 | 0,55 | 0,2088 | 0,85 | 0,3023 | 1,15 | 0,3749 |
| 0,26 | 0,1026 | 0,56 | 0,2123 | 0,86 | 0,3051 | 1,16 | 0,3770 |
| 0,27 | 0,1064 | 0,57 | 0,2157 | 0,87 | 0,3078 | 1,17 | 0,3790 |
| 0,28 | 0,1103 | 0,58 | 0,2190 | 0,88 | 0,3106 | 1,18 | 0,3810 |
| 0,29 | 0,1141 | 0,59 | 0,2224 | 0,89 | 0,3133 | 1,19 | 0,3830 |
| 0,30 | 0,1179 | 0,60 | 0,2257 | 0,90 | 0,3159 | 1,20 | 0,3849 |

Продовження таблиця А.1

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------|--------|------|--------|------|--------|------|----------|
| 1,21 | 0,3869 | 1,55 | 0,4394 | 1,89 | 0,4706 | 2,46 | 0,4931 |
| 1,22 | 0,3888 | 1,56 | 0,4406 | 1,90 | 0,4718 | 2,48 | 0,4934 |
| 1,23 | 0,3907 | 1,57 | 0,4418 | 1,91 | 0,4719 | 2,50 | 0,4938 |
| 1,24 | 0,3925 | 1,58 | 0,4429 | 1,92 | 0,4726 | 2,52 | 0,4941 |
| 1,25 | 0,3944 | 1,59 | 0,4441 | 1,93 | 0,4732 | 2,54 | 0,4945 |
| 1,26 | 0,3962 | 1,60 | 0,4452 | 1,94 | 0,4738 | 2,56 | 0,4948 |
| 1,27 | 0,3980 | 1,61 | 0,4463 | 1,95 | 0,4744 | 2,58 | 0,4951 |
| 1,28 | 0,3997 | 1,62 | 0,4474 | 1,96 | 0,4750 | 2,60 | 0,4953 |
| 1,29 | 0,4015 | 1,63 | 0,4484 | 1,97 | 0,4756 | 2,62 | 0,4956 |
| 1,30 | 0,4032 | 1,64 | 0,4495 | 1,98 | 0,4761 | 2,64 | 0,4959 |
| 1,31 | 0,4049 | 1,65 | 0,4505 | 1,99 | 0,4767 | 2,66 | 0,4961 |
| 1,32 | 0,4066 | 1,66 | 0,4515 | 2,00 | 0,4772 | 2,68 | 0,4963 |
| 1,33 | 0,4082 | 1,67 | 0,4525 | 2,02 | 0,4783 | 2,70 | 0,4965 |
| 1,34 | 0,4099 | 1,68 | 0,4535 | 2,04 | 0,4793 | 2,72 | 0,4967 |
| 1,35 | 0,4115 | 1,69 | 0,4545 | 2,06 | 0,4803 | 2,74 | 0,4969 |
| 1,36 | 0,4131 | 1,70 | 0,4554 | 2,08 | 0,4812 | 2,76 | 0,4971 |
| 1,37 | 0,4147 | 1,71 | 0,4564 | 2,10 | 0,4821 | 2,78 | 0,4973 |
| 1,38 | 0,4162 | 1,72 | 0,4573 | 2,12 | 0,4830 | 2,80 | 0,4974 |
| 1,39 | 0,4177 | 1,73 | 0,4582 | 2,14 | 0,4838 | 2,82 | 0,4976 |
| 1,40 | 0,4192 | 1,74 | 0,4591 | 2,16 | 0,4846 | 2,84 | 0,4977 |
| 1,41 | 0,4207 | 1,75 | 0,4599 | 2,18 | 0,4854 | 2,86 | 0,4979 |
| 1,42 | 0,4222 | 1,76 | 0,4608 | 2,20 | 0,4861 | 2,88 | 0,4980 |
| 1,43 | 0,4236 | 1,77 | 0,4616 | 2,22 | 0,4868 | 2,90 | 0,4981 |
| 1,44 | 0,4251 | 1,78 | 0,4625 | 2,24 | 0,4875 | 2,92 | 0,4982 |
| 1,45 | 0,4265 | 1,79 | 0,4633 | 2,26 | 0,4881 | 2,94 | 0,4984 |
| 1,46 | 0,4279 | 1,80 | 0,4641 | 2,28 | 0,4887 | 2,96 | 0,4985 |
| 1,47 | 0,4292 | 1,81 | 0,4649 | 2,30 | 0,4893 | 2,98 | 0,4986 |
| 1,48 | 0,4306 | 1,82 | 0,4656 | 2,32 | 0,4898 | 3,00 | 0,49865 |
| 1,49 | 0,4319 | 1,83 | 0,4664 | 2,34 | 0,4904 | 3,20 | 0,49931 |
| 1,50 | 0,4332 | 1,84 | 0,4671 | 2,36 | 0,4909 | 3,40 | 0,49966 |
| 1,51 | 0,4345 | 1,85 | 0,4678 | 2,38 | 0,4913 | 3,60 | 0,499841 |
| 1,52 | 0,4357 | 1,86 | 0,4686 | 2,40 | 0,4918 | 3,80 | 0,499928 |
| 1,53 | 0,4370 | 1,87 | 0,4693 | 2,42 | 0,4922 | 4,00 | 0,499968 |
| 1,54 | 0,4382 | 1,88 | 0,4699 | 2,44 | 0,4927 | 4,50 | 0,499997 |
| | | | | | | 5,00 | 0,499999 |

| | |
|---|-----|
| Вступ | 3 |
| 1 Інформація та інформаційні системи | |
| 1.1 Основні поняття і визначення | 5 |
| 1.2 Критерії оцінки ефективності якості | 9 |
| 1.3 Предмет і метод теорії інформації | 11 |
| Контрольні питання | 12 |
| 2 Математичні моделі детермінованих сигналів | |
| 2.1 Загальна характеристика сигналів | 13 |
| 2.2 Частотне представлення детермінованих сигналів | 15 |
| 2.3 Приклади | 27 |
| Контрольні питання | 34 |
| 3 Математичні моделі випадкових сигналів | |
| 3.1 Випадкові сигнали та їх ймовірнісні характеристики | 35 |
| 3.2 Числові характеристики випадкового процесу | 38 |
| 3.3 Стационарні випадкові процеси | 42 |
| 3.4 Властивості кореляційної функції стационарного випадкового процесу | 43 |
| 3.5 Ергодичність стационарних процесів | 46 |
| 3.6 Спектральна щільність випадкового процесу | 47 |
| 3.7 Широко- та вузькосмугові процеси | 49 |
| 3.8 Ефективна ширина спектра випадкового процесу | 52 |
| 3.9 Приклади | 53 |
| Контрольні питання | 61 |
| 4 Квантування сигналів | |
| 4.1 Засоби квантування сигналів | 63 |
| 4.2 Квантування за часом | 64 |
| 4.3 Рівномірне квантування за часом | 65 |
| 4.4 Квантування за рівнем | 72 |
| 4.5 Рівномірне квантування за рівнем | 73 |
| 4.6 Нерівномірне квантування за рівнем | 78 |
| 4.7 Приклади | 80 |
| Контрольні питання | 86 |
| 5 Модуляція і демодуляція | |
| 5.1 Амплітудна модуляція | 88 |
| 5.2 Різновиди амплітудної модуляції | 98 |
| 5.3 Кутова модуляція | 106 |
| 5.4 Методи здійснення зовнішньої частотної модуляції радіосигналів | 116 |
| 5.5 Квадратурна модуляція | 124 |
| 5.6 Способи модуляції, які використовуються при передачі цифрової інформації | 127 |
| Контрольні питання | 140 |

| | |
|---|-----|
| 6 Кодування сигналів | |
| 6.1 Мета кодування. Основні поняття і визначення | 141 |
| 6.2 Рівномірні прості цифрові коди | 145 |
| 6.3 Складені коди | 147 |
| 6.4 Рефлексні (відображені) коди | 148 |
| 6.5 Завадостійке кодування. Класифікація завадостійких кодів | 149 |
| 6.6 Основні принципи завадостійкого кодування | 151 |
| 6.7 Зв'язок коригувальної здатності коду з кодовою відстанню | 154 |
| 6.8 Побудова кодів з заданою виправною здатністю | 156 |
| 6.9 Показники якості коригувального коду | 158 |
| 6.10 Систематичні коди | 159 |
| 6.11 Коди з виявленням помилок | 164 |
| 6.12 Коди з виявленням і виправленням помилок | 166 |
| 6.13 Приклади | 180 |
| Контрольні питання | 193 |
| 7 Інформаційні моделі сигналів | |
| 7.1 Кількість інформації і невизначеність. Ентропія як міра невизначеності | 195 |
| 7.2 Властивості ентропії дискретних повідомлень | 199 |
| 7.3 Ентропія неперервних подій | 200 |
| 7.4 Ентропія складних повідомлень | 202 |
| 7.5 Кількість інформації при неповній достовірності повідомлень | 205 |
| 7.6 Ентропія і кількість інформації при статистичній залежності елементів повідомлень | 211 |
| 7.7 Надлишковість повідомлення | 213 |
| 7.8 Приклади | 216 |
| Контрольні питання | 224 |
| 8 Передача інформації | |
| 8.1 Узагальнені характеристики інформаційних сигналів | 225 |
| 8.2 Швидкість передачі інформації та пропускна здатність дискретного каналу без завад | 228 |
| 8.3 Швидкість передачі інформації та пропускна здатність дискретного каналу із завадами | 231 |
| 8.4 Швидкість передачі інформації та пропускна здатність безперервного каналу із завадами | 233 |
| 8.5 Приклади | 239 |
| Контрольні питання | 246 |
| 9 Оптимальний прийом та обробка інформації | |
| 9.1 Частотна фільтрація | 247 |
| 9.2 Метод накопичування | 249 |
| 9.3 Кореляційний метод | 251 |
| 9.4 Узгоджена фільтрація | 254 |
| 9.5 Сутність основної задачі прийому сигналів при наявності завад | 255 |
| 9.6 Виявлення сигналу | 257 |
| 9.7 Розпізнавання сигналів | 266 |
| 9.8 Відновлення сигналів | 268 |
| 9.9 Приклади | 271 |
| Контрольні питання | 279 |

| | |
|--|-----|
| 10 Ефективність інформаційних систем | |
| 10.1 Критерії оцінювання ефективності інформаційних систем | 281 |
| 10.2 Способи підвищення ефективності інформаційних систем | 282 |
| 10.3 Перерозподіл щільності ймовірностей елементів повідомлення | 283 |
| 10.4 Декореляція повідомлень | 285 |
| 10.5 Оптимальне статистичне кодування | 287 |
| 10.6 Приклади | 291 |
| Контрольні питання | 294 |
| 11 Завадостійкість інформаційних систем | |
| 11.1 Загальна характеристика завад у системах передачі інформації | 295 |
| 11.2 Критерії оцінювання завадостійкості інформаційних систем | 297 |
| 11.3 Засоби підвищення завадостійкості інформаційних систем | 301 |
| 11.4 Завадостійкість різних видів модуляції | 303 |
| 11.5 Системи із зворотним зв'язком | 311 |
| 11.6 Приклади | 312 |
| Контрольні питання | 316 |
| 12 Інформаційна оцінка автоматизованих систем контролю та управління | |
| 12.1 Загальна характеристика автоматизованих систем контролю і управління | 317 |
| 12.2 Оцінка ступеня невизначеності стану об'єкта контролю | 318 |
| 12.3 Інформаційна оцінка точності результатів контролю | 321 |
| 12.4 Інформаційна здатність пристрою контролю та ККД процесу контролю | 323 |
| 12.5 Пропускна здатність процесу контролю | 325 |
| 12.6 Ентропія та інформація у системах автоматичного управління | 326 |
| 12.7 Оцінка ефективності АСКУ | 330 |
| 12.8 Інформаційне обґрунтування оптимального алгоритму контролю і пошуку несправностей | 333 |
| Контрольні питання | 338 |
| 13 Стиснення зображень | |
| 13.1 Алгоритми архівації зображень | 339 |
| 13.2 Алгоритми архівації без втрат | 345 |
| 13.3 Алгоритми архівації з втратами | 351 |
| 13.4 MPEG: загальна інформація | 361 |
| Контрольні питання | 367 |
| Література | 368 |
| Додаток | 369 |

Навчальне видання

*Кузьмін Іван Васильович
Троцишин Іван Васильович
Кузьмін Андрій Іванович
та ін.*

ОСНОВИ ТЕОРІЇ ІНФОРМАЦІЇ ТА КОДУВАННЯ

Підручник для студентів вищих навчальних закладів

3-тє видання, перероблене та доповнене

Відповідальний за випуск: **В.С. Яремчук**

Художнє оформлення обкладинки: **О.С. Ванчурова**

Редактор-коректор: **Г.М. Соколова**

Комп'ютерна верстка: **В.П. Романюк**

Підписано до друку 29.04.2009

Формат 30×42/2. Папір офс. Гарн. Times New Roman
Друк різнографією. Ум. друк. арк. – 21,93. Обл.-вид. арк. – 20,50
Тираж 500. Зам. № 96/09