МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДНІПРОВСЬКИЙ ДЕРЖАВНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

КОНСПЕКТ ЛЕКЦІЙ

з дисципліни

«ТЕОРІЯ ПЕРЕДАВАННЯ ІНФОРМАЦІЇ»

для здобувачів вищої освіти другого (магістерського) рівня
зі спеціальності – 172 «Електронні комунікації та радіотехніка»
за освітньо-професійною програмою «Телекомунікації та радіотехніка»

Затверджено:

науково-методичною радою ДДТУ
від «20» 06 2024 рік
Протокол № 6

Кам’янське
2024

Конспект лекцій з дисципліни «Теорія передавання інформації» для здобувачів вищої освіти другого (магістерського) рівня зі спеціальності - 172 «Електронні комунікації та радіотехніка». /Укл.: Литвиненко В.А., - Кам'янське; ДДТУ, 2024 р. – 77 с.

Укладачі: к.т.н., доц. Литвиненко В.А.

Рецензент: канд. техн. наук, доцент Багрій В.В.

Відповідальний за випуск: в.о. завідувача кафедри ЕЕК,
канд. техн. наук, доцент Багрій В.В.

Затверджено на засіданні кафедри ЕЕК
від « 10 » 06 2024 року
протокол № 9

Коротка анотація. У посібнику розглянуто основи теорії передачі інформації. Основну увагу звернено на визначення середньої швидкості передавання інформації та розв'язання задачі максимізації швидкості застосуванням відповідного кодування. Висвітлено характеристики ймовірностей і втрат інформації на виході нелінійного елемента та ефективність систем передавання інформації. Викладено основні поняття кодування джерел інформації та каналів зв'язку. Розглянуто коректувальні, циклічні, несистематичні, ітеративні коди; коди Боуза–Чоудхурі та ефективність систематичних коректувальних кодів.Конспект лекцій складено на основі [1].

ЗМІСТ

ВСТУП

Конспект лекцій “Теорія передачі інформації ” (ТПІ) призначено для роботи здобувачів вищої освіти другого (магістерського) рівня зі спеціальності – 172 «Електронні комунікації та радіотехніка»

Навчальний посібник підготовлено для роботи за кредитно-модульною системою, відповідає програмі дисципліни ТПІ.

Предметом навчальної дисципліни є:

- основи сучасної теорії передачі інформації з акцентом на фізичне тлумачення процесів, які відбуваються під час передавання повідомлень та сигналів у системах зв'язку; математичний опис основних фізичних процесів передавання сигналів та методи забезпечення граничних характеристик систем зв’язку як за достовірністю, так і за швидкістю передачі інформації; загальні принципи модуляції; методи цифрової модуляції; теорема Котельникова;
- процеси передавання сигналів каналами зв’язку при наявності завад з математичної точки зору. Методи ефективного кодування. Теорема Шенона для каналу з завадами;
- оптимальний прийом сигналів. Принципи побудови багатоканальних модемів, багатопозиційні сигнали і їх застосування у високошвидкісних модемах;
- принципи побудови телекомунікаційних мереж; цифрові методи передачі неперервних повідомлень. Основи теорії лінійного розділу сигналів.

ТЕМА 1. КІЛЬКІСНА МІРА ІНФОРМАЦІЇ. ЕНТРОПІЯ ДЖЕРЕЛА ДИСКРЕТНИХ ПОВІДОМЛЕНЬ

В теорії інформації вивчаються кількісні закономірності передавання, збереження й обробляння інформації. Зокрема, основна увага приділяється визначенню середньої швидкості передавання інформації та розв'язанню задачі максимізації цієї швидкості застосуванням відповідного кодування. Граничні співвідношення теорії інформації дають можливість оцінити ефективність різних систем зв'язку та встановити умови узгодження в інформаційному відношенні джерела з каналом і каналу з одержувачем (споживачем) інформації.

Для дослідження цих питань з обох позицій необхідно насамперед встановити універсальну кількісну міру інформації, що не залежить від конкретної фізичної природи повідомлень, які передаються.

Статистичні методи визначення кількісної міри інформації. При визначенні одиниці кількості інформації слід мати на увазі, що кількість інформації $I(a)$ в повідомленні про деяку подію істотно залежить від імовірності цієї події:

$$I(a) = \log_k \frac{1}{P(a)} = -\log_k P(a). \qquad (1)$$

Оскільки $0 < P(a) < 1$, то $I(a)$ – величина додатна і скінченна. При $P(a) = 1$ кількість інформації дорівнює нулеві, тобто повідомлення про відому подію ніякої інформації не несе.

Логарифмічна міра у даному випадку має природну *властивість адитивності*, відповідно до якої кількість інформації, що міститься в кількох незалежних повідомленнях, дорівнює сумі кількості інформації, що міститься в усіх повідомленнях.Дійсно, якщо загальна ймовірність n незалежних повідомлень $P(a_1 \text{ і } a_2 \text{ і} \dots \text{ і } a_n) = P(a_1)P(a_2)\dots P(a_n)$, то кількість інформації в цих повідомленнях дорівнює

$$I(a_1 \text{ і } a_2 \text{ і} \dots \text{ і } a_n) = -\log_k P(a_1 \text{ і } a_2 \text{ і} \dots \text{ і } a_n) = -\sum_{i=1}^{n} \log_k P(a_i) = \sum_{i=1}^{n} I(a_i). \qquad (2)$$

При $k = 2$ кількість інформації виражається в двійкових одиницях (дв. од.):

$$I(a_i) = \log P(a_i).$$

Визначити кількість інформації можна по-іншому. Розглянемо як повідомлення не окрему букву, а ціле число. Якщо всі букви рівноймовірні і незалежні, то всі слова також рівноймовірні:

$$P_{\text{сл}} = 1/N,$$

де $N = m^n$ – кількість слів.

Тоді можна записати:

$$I_{\text{сл}} = -\log P_{\text{сл}} = -\log(1/m^2) = n \log m. \qquad (3)$$

Для двійкового каналу $m = 2$. Тоді $n \log 2 = n$ дв. од.

У загальному випадку при передаванні повідомлень невизначеність зменшується. Так, у каналі з шумами можливі помилки. За прийнятим повідомленням тільки з деякою ймовірністю $P(u/\text{v}) < 1$ можна зрозуміти, що було передано повідомлення u при прийнятому повідомленні v. Тому після одержання повідомлення залишається деяка невизначеність, що характеризується величиною апостеріорної ймовірності $P(u/\text{v})$, а кількість інформації, яка міститься в сигналі v, визначається ступенем зменшення невизначеності при його прийманні.

Якщо $P(u)$ – апріорна ймовірність, то кількість інформації в прийнятому повідомленні v щодо переданого повідомлення u дорівнює

$$I(u,\text{v}) = \log \frac{1}{P(u)} - \log \frac{1}{P(u/\text{v})} = -\log P(u) + \log P(\text{v}) = \log \frac{P(u/\text{v})}{P(u)}. \qquad (4)$$

Цей вираз можна розглядати також як різницю між кількістю інформації, що надійшла від джерела повідомлень, і тією кількістю інформації, що втрачена в каналі внаслідок дії шумів.

Наведена вище кількісна міра інформації стосується тільки статистичної властивості повідомлень, що передаються. Для придбання, збереження, перероблення і передавання повідомлень (інформації, знань) використовуються *знаки* (символи) – матеріальні предмети, явища, події, що виступають як представники деяких інших предметів, властивостей або відносин. Розрізняють два види знаків: *мовні,* що входять у деяку знакову систему (природні або штучні мови), і *немовні*. Прикладами останніх можуть бути відбитки пальців, симптоми захворювань, зображення маски як символу театрального мистецтва і т. д. Наука про властивості знаків називається *семіотикою*. Семіотика вивчає знаки в трьох аспектах. *Синтактика* вивчає знаки і співвідношення між ними (правила побудови складених знаків у межах знакової системи), наприклад, правила побудови слів з букв, речень зі слів безвідносно до інтерпретації. *Семантика* розглядає співвідношення між знаками та їх інтерпретаціями незалежно від того, хто є інтерпретатором (адресатом). Так, знак у вигляді червоного хреста, розташований на будинку, означає пункт медичної допомоги. *Прагматика* досліджує зв'язок знаків з тими, хто використовує їх, тобто проблеми тлумачення знаку "адресатами", його корисності та цінності для інтерпретатора. Очевидно, корисність і цінність знаку істотно різні для здорової й хворої людини. Всі три аспекти вивчення знаків взаємопов'язані, і кожен наступний містить у собі попередній. Оскільки знак є носієм інформації, то відповідно до наведеної структури семіотики виділяють синтактичний, семантичний і прагматичний аспекти теорії інформації.

Структурні міри інформації. Дискретні повідомлення складаються зі зліченної множини символів, створюваних джерелом інформації. Набір символів, якими можуть бути цифри, букви, імпульси й інші знаки, називають *алфавітом джерела*. Кількість символів у алфавіті визначає *обсяг алфавіту*. Типовими дискретними повідомленнями є текст, записаний за допомогою якого-небудь алфавіту, та послідовність чисел. Символи можуть мати різні фізичні властивості, які дають змогу однозначно відрізняти їх один від одного. Ці властивості називають *якісними ознаками*. При використанні структурних мір інформації враховується кількість символів, зв'язків між ними або комбінацій з них. *Кількість інформації* в комбінаторній мірі обчислюється як кількість комбінацій символів, що входять до алфавіту. Отже, оцінюється *комбінаторна властивість* потенціальної структурної різноманітності джерела інформації.

Комбінування можливе при двох і більше неоднакових символах, наявності змінних або зв'язків, коли якісною ознакою символу є його положення (позиція) у послідовності. В останньому випадку місце розташування символу впливає на ціле число (наприклад, у позиційній системі числення). Нехай є двійкові числа 11110 і 01111 або 00001 і 10000. У першому випадку змінює положення нуль, у другому – одиниця. Відповідно число 30 перетворюється в 15, а одиниця – в 16.

У комбінаториці розглядаються різні види сукупності (сполуки), що утворилася з елементів деякої множини *M*, яка містить n різних елементів. Найпростіші з таких сполук – перестановки, розміщення, комбінації.

Кількість різних сполук з n елементів (символів) по m різних елементах множини *M*

$$N = C_m^n = m!/\left[n!(m-n)!\right].$$

Якщо, наприклад, символами є a, b, c, d , то число сполук по двох символах дорівнює 6. Це будуть сполуки ab, ac, ad, bc, bd, cd.

Сполуки з повтореннями також відрізняються складом елементів, але елементи в них можуть повторюватися до n разів. При цьому

$$N = C_{m(\text{повт})}^n = (m+n-1)!/\left[n!(m-1)!\right].$$

При вищевказаних чотирьох символах і $n=2$ матимемо 10 сполук. До наведених раніше шести додадуться комбінації aa, bb, cc, dd .

Перестановки – це сполуки з m елементів, які різняться між собою тільки порядком елементів у послідовності. Число можливих перестановок m символів $N=m$. Як випливає з наведеного виразу, при визначенні числа перестановок завжди передбачається, що $n=m$. В

усіх попередніх і наступних прикладах довжина n сполук дорівнюватиме двом, тому для одержання порівнянних результатів покладемо $n = m = 2$. Тоді матимемо тільки дві сполуки: *ab* і *ba*.

Розміщення – це сполуки з *m* символів по *n* елементах, що відрізняються одна від одної або складом елементів, або їхнім порядком. Можливе число розміщень з *m* символів по *n*

$$N = C^n_{m(\text{повт})} = m!/(m-n)!.$$

При $m = 4$ і $n = 2$ маємо 12 сполук. Шість із них ідентичні тим, що були отримані при сполученні символів, у шести інших символи в кожній сполуці міняються місцями (достатньо записати їх у зворотному порядку).

Можливе число *розміщень з повтореннями* по *n* символах з *m*

$$N = C^n_{m(\text{повт})} = m^n. \qquad (5)$$

При $m = 4$ і $n = 2$ маємо 16 сполук. До наведених вище 12 сполук додаються ще чотири: *aa, bb, cc, dd* . Вираз (5) визначає максимальне число сполук, які можна одержати, розміщуючи *m* різних символів по *n* позиціях.

Комбінації – це сполуки з *n* елементів по *m*, що різняться між собою принаймні одним елементом.

При застосуванні комбінаторної міри кількість інформації збігається з числом можливих поєднань. Отже, визначення кількості інформації в комбінаторній мірі полягає в пошуку кількості можливих або справді здійснених комбінацій, тобто в оцінці *структурної різноманітності*. Внаслідок показникового закону залежності числа можливих з'єднань від *n* – довжини послідовності символів у поєднанні (5) – число можливих комбінацій не є зручною мірою кількості інформації. У статті Р.Хартлі запропонував за міру кількості інформації двійковий логарифм числа можливих комбінацій символів:

$$I = \log_2 N = \log_2 m^2 = n \log m. \qquad (6)$$

З виразу (6) випливає, що кількість інформації пропорційна довжині *n* послідовності символів у сполуці. Логічно за одиницю взяти кількість інформації, що припадає на одну позицію.

Первинним і неподільним елементом інформації є *елементарна двійкова подія А* – вибір з твердження або заперечення, наявності або відсутності якого-небудь явища, істини і неправди тощо. Двійковість події дає можливість зобразити її умовно в геометричній символіці точкою і пробілом, в арифметичній – одиницею і нулем, у сигнальній – імпульсом і паузою (відсутністю імпульсу). Тому дуже часто інформацію подають послідовністю, складеною тільки з двох символів (0 і 1), тобто $m = 2$. Якщо тепер у (6) покласти $n = 1,\ m = 2$, то кількість інформації $I = 1 \cdot \log_2 2 = 1$ біт.

Одиниця кількості інформації, отримана при зазначених вище умовах, називається *двійковою* або одиницею *біт*. Уведене визначення кількості інформації еквівалентне числу двійкових знаків 0 і 1 при поданні повідомлень числами двійкової системи числення. Одному біту відповідає один двійковий розряд.

Особливості семантичної оцінки інформації. Якщо статистична теорія інформації будується на понятті математичної ймовірності, то семантична теорія інформації Р. Карнапа і Бар-Хілела ґрунтується на семантичному визначенні ймовірності, що називається *логічною,* або *індуктивною ймовірністю*. Під терміном логічна, або індуктивна ймовірність Р. Карнап розуміє вид імовірності, який мається на увазі щоразу, коли роблять індуктивний висновок, що не випливає з логічною необхідністю (з імовірністю "одиниця") зі свідчень, істинність яких гарантується. При цьому вважатимемо, що стосовно даних свідчень висновок має деякий ступінь імовірності. Висновок між істиною і неправдою називається у ймовірній логіці *гіпотезою,* тобто припущенням, здогадом.

Коли ваш співрозмовник-філателіст стверджує, що цього року, ймовірно, буде випущена поштова марка, присвячена Т. Г. Шевченку, то він не виходить за межі статистичної

ймовірності і виражає свою впевненість у цьому. Така логічна ймовірність поширюється як на користь висловленої гіпотези, так і на спростування її. Нехай такими взаємними виключеннями є висловлення:

A – Т. Г. Шевченко – видатний український поет, $\bar{A}$ – ні;

B – цього року відзначатиметься 150-річчя з дня народження поета, $\bar{B}$ – ні;

C – раніше випускалися марки, присвячені поетам, $\bar{C}$ – ні.

Логічна ймовірність – це поняття, застосовне до пар тверджень – гіпотези (буде випущена поштова марка, присвячена Т. Г. Шевченку) і свідчення, яке виражає щось, що має відношення до гіпотези.

Нехай обговорюється можливість випуску зазначеної поштової марки. Потрібно оцінити апріорну логічну ймовірність того, що така марка буде випущена. Спочатку припустимо, що коли висловлюється гіпотеза H – буде випущена поштова марка, присвячена Т. Г. Шевченку, немає ніяких свідчень щодо висловленої гіпотези. Тоді, відповідно до *принципу байдужості* (однакової ймовірності), передбачається, що з імовірністю 0,5 марка буде або не буде випущена. Перед тим, як висловити наступну гіпотезу, мабуть, варто оцінити ймовірність того, що марка буде випущена залежно від співвідношення свідчень на користь висловлюваної гіпотези і проти неї. Дійсно, якщо першим свідченням є A – висловлення на користь гіпотези H, то логічно вважати гіпотезу H більш імовірною. Аналогічно можна висловитись, якщо першим свідченням було $\bar{A}$ – висловлення на користь гіпотези $\bar{H}$. В обох варіантах першого свідчення логічна ймовірність гіпотези, що суперечить висловленому свідченню, оцінюватиметься меншою величиною (табл. 1).

Як же встановити числове значення ймовірності одних висловлень щодо інших? Однозначної відповіді на це питання поки немає. Зрозуміло лише, що ступінь ймовірності гіпотези залежить від стану накопичених свідчень – знань. Розглянемо спосіб присвоювання величини логічної ймовірності, запропонований Карнапом. Щоразу, коли змінюється кількість свідчень, переоцінюються присвоєні послідовності апріорних імовірностей за таким правилом. Усі можливі послідовності з деякою кількістю розглянутих висловлень поєднуються в групи; в межах кожної зібрані комбінації з однаковим відношенням висловлювань одного виду до висловлювань іншого виду; кожній групі присвоюється та сама ймовірність. Потім імовірності груп поділяються нарівно між окремими послідовностями, тобто визначаються апріорні логічні ймовірності кожної можливої послідовності.

За таким методом присвоювання апріорних імовірностей можна враховувати накопичений досвід. Як випливає з табл. 1.1, зі збільшенням у послідовності відносного числа висловлювань логічні ймовірності гіпотези H зростають. Припустимо, що три свідчення дали послідовність: Т. Г. Шевченко – український поет; у цьому році відзначатиметься 150-річчя з дня народження; раніше не випускалися марки, присвячені поетам (варіанти 3 і 10 табл. 1).

Подібна послідовність показує перевагу випуску поштової марки, присвяченої Т. Г. Шевченку, над тим, що такої марки не буде. Дійсно, ймовірність того, що марка, присвячена Т. Г. Шевченку, буде випущена, дорівнює 3/60 проти 2/60 для гіпотези $\bar{H}$, тобто становить 3/2. У цьому варіанті розподілу апріорних імовірностей не враховується "вага" свідчення, тобто припускається, наприклад, що висловлювання: Т. Г. Шевченко – видатний український поет, цього року буде відзначатися 150-річчя з дня його народження – рівноцінні стосовно гіпотези.

Розглянута теорія встановлює міру кількості семантичної інформації, що міститься в простих твердженнях або пропозиціях. Кількість семантичної інформації $\mathrm{cont}(i)$, що міститься в якому-небудь елементарному твердженні i (гіпотезі i), пропонується розглядати як функцію логічної ймовірності $L(i)$. Позначення cont є скороченням англійського слова content – зміст, суть. Семантична інформація, як і в статистичній теорії інформації, розглядається тут як знята, усунута невизначеність при переході від відносного знання (гіпотези) до достовірного знання. Якщо i й j – два висловлення, і $L(i) > L(j)$, то слід вважати, що гіпотеза i містить меншу кількість інформації, ніж гіпотеза j, тобто $\mathrm{cont}(i) < \mathrm{cont}(j)$.

Подібна послідовність показує перевагу випуску поштової марки, присвяченої Т. Г. Шевченку, над тим, що такої марки не буде. Дійсно, ймовірність того, що марка, присвячена Т. Г. Шевченку, буде випущена, дорівнює 3/60 проти 2/60 для гіпотези $\bar{H}$, тобто становить 3/2. У цьому варіанті розподілу апріорних імовірностей не враховується "вага" свідчення, тобто припускається, наприклад, що висловлювання: Т. Г. Шевченко – видатний український поет, цього року буде відзначатися 150-річчя з дня його народження – рівноцінні стосовно гіпотези.

Таблиця 1. **Апріорна ймовірність залежно від кількості свідчень**

Номер варіанта	Ймовірна послідовність	Апріорно присвоєна ймовірність	
		групі	послідовності
Кількість свідчень – 0			
1	H	½	1/2
2	$\bar{H}$	½	1/2
Кількість свідчень – 1			
1	AH	1/3	2/6
2	$A\bar{H}$	1/3	1/6
3	$\bar{A}H$	1/3	1/6
4	$\bar{A}\bar{H}$	1/3	2/6
Кількість свідчень – 2			
1	ABH	1/3	6/18
2	$AB\bar{H}$	1/3	1/18
3	$A\bar{B}H$	1/3	1/18
4	$\bar{A}BH$	1/3	1/18
5	$\bar{A}\bar{B}H$	1/3	1/18
6	$\bar{A}B\bar{H}$	1/3	1/18
7	$A\bar{B}\bar{H}$	1/3	1/18
8	$\bar{A}\bar{B}\bar{H}$	1/3	6/18
Кількість свідчень – 3			
1	$ABCH$	1/5	12/60
2	$ABC\bar{H}$	1/5	3/60
3	$AB\bar{C}H$	1/5	3/60
4	$A\bar{B}CH$	1/5	3/60
5	$\bar{A}BCH$	1/5	3/60
6	$A\bar{B}\bar{C}H$	1/5	2/60
7	$\bar{A}\bar{B}CH$	1/5	2/60
8	$\bar{A}B\bar{C}H$	1/5	2/60
9	$A\bar{B}C\bar{H}$	1/5	2/60
10	$AB\bar{C}\bar{H}$	1/5	2/60
11	$\bar{A}BC\bar{H}$	1/5	2/60
12	$A\bar{B}\bar{C}\bar{H}$	1/5	3/60
13	$\bar{A}B\bar{C}\bar{H}$	1/5	3/60
14	$\bar{A}\bar{B}C\bar{H}$	1/5	3/60
15	$\bar{A}\bar{B}\bar{C}H$	1/5	3/60
16	$\bar{A}\bar{B}\bar{C}\bar{H}$	1/5	12/60

Розглянута теорія встановлює міру кількості семантичної інформації, що міститься в простих твердженнях або пропозиціях. Кількість семантичної інформації $\mathrm{cont}(i)$, що міститься в якому-небудь елементарному твердженні i (гіпотезі i), пропонується розглядати як функцію логічної ймовірності $L(i)$. Позначення cont є скороченням англійського слова content – зміст,

суть. Семантична інформація, як і в статистичній теорії інформації, розглядається тут як знята, усунута невизначеність при переході від відносного знання (гіпотези) до достовірного знання. Якщо i й j – два висловлення, і $L(i) > L(j)$, то слід вважати, що гіпотеза i містить меншу кількість інформації, ніж гіпотеза j, тобто $\text{cont}(i) < \text{cont}(j)$.

Чим більша логічна ймовірність гіпотези "буде випущена поштова марка, присвячена Т. Г. Шевченку", тим менша кількість семантичної інформації міститься в ній щодо достовірного знання "поштова марка, присвячена Т. Г. Шевченку, випущена". Дійсно, якщо Ваше припущення про ймовірний випуск марки підтверджується, то усунута при цьому невизначеність невелика, а кількість інформації мала. Якщо марка не вийде, логічна ймовірність такого результату передбачалася малою, результат для філателістів сенсаційний, і вони мають велику кількість інформації.

Бажано, щоб у випадку, коли i та j логічно незалежні, мало місце співвідношення $\text{cont}(i \cap j) = \text{cont}(i) + \text{cont}(j)$.Прийняте тлумачення логічної незалежності можна показати на прикладі висловлення: книга художня і книга (та сама) не українська. Ці висловлення незалежні, але не виключають одне одного.

Вищевказані умови аналогічні прийнятим при розгляді статистичної теорії інформації, тому кількість семантичної інформації $\text{cont}(i) = \log[1/L(i)]$. Цей вираз формально подібний розглянутому в статистичній теорії, але він базується на логічній імовірності гіпотези, що дає змогу наблизитися до оцінки змісту інформації.

В іншому варіанті оцінки кількості семантичної інформації використовується поняття *тезауруса* (від грецького "скарбниця"), під яким розуміється запас знань або словник, використовуваний приймачем інформації. Відомості, отримані під час приймання повідомлення, змінюють вихідний тезаурус. Величина цієї зміни береться за характеристику кількості семантичної інформації, яка міститься в даному повідомленні щодо даного приймача. Недостатньо розвинений тезаурус одержує нульову або дуже малу інформацію з даного повідомлення (не може його зрозуміти). Але й надто повний тезаурус також не може одержати багато інформації, оскільки вона не буде для нього новою. Кількість інформації, яку одержує приймач з деякого повідомлення, залежить від кількості наявної в тезаурусі (приймачі) інформації $I_т$.

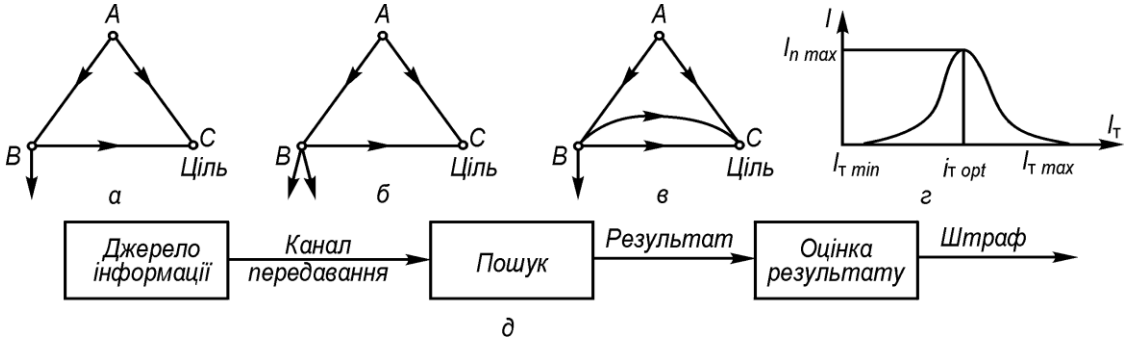

Рис. 1. До визначення цінності інформації: *а*, *б*, *в* – шляхи досягнення цілі; *г* – кількість інформації в тезаурусі; *д* – система визначення цінності за Р. Л. Стратоновичем

Цінність інформації. Визначаючи цінність інформації, звичайно розглядають таку ситуацію. Хтось прагне до поставленої цілі, маючи можливість наблизитися до неї, затративши певні зусилля. Завдяки одержуваним ззовні відомостям він може точніше наблизитися до цілі або заощадити витрати на її досягнення. Цінність отриманої інформації вимірюється ступенем наближення до цілі або величиною економії на витратах.

Якщо інформація використовується для управління досягненням цілі, то її цінність можна виразити через зміну ймовірності її досягнення. Нехай до одержання інформації ймовірність досягнення цілі дорівнювала P_0, а після її одержання – P_1, тоді за А. А. Харкевичем цінність інформації визначається як

$$\text{Ц} = \log_2 P_1 - \log_2 P_0 = \log_2 (P_1 / P_0). \qquad (7)$$

Тут імовірність обчислюють відношенням числа сприятливих результатів до їх загального числа.

Розглянемо випадок, наведений на рис. 1, *а, б*. Пункти A, B і C сполучені шляхами. Мандрівник знаходиться у точці *A*, точка *C* – ціль подорожі. Перед ним два шляхи. Який з них веде до *C*, він не знає. Не маючи інформації для вибору, він з імовірністю 0,5 може бути у *B* або *C*, тобто ймовірність досягнення цілі до одержання інформації $P_0 = 0,5$. Нехай за отриманою інформацією $I = \log[1/P_0] = 1$ біт (наприклад, хтось сказав, що треба йти лівим шляхом) він переходить у пункт *B*. Яка цінність отриманої інформації? Вона залежить від імовірності досягнення цілі з пункту *B*. Визначимо цінності отриманої інформації для трьох випадків:

1. Число хороших випадків дорівнює одному з двох можливих: $P_1 = 0,5$ (рис. 1, *а*). Отже, $\text{Ц} = \log_2[P_1/P_0] = \log_2 1 = 0$. Тут імовірність досягнення цілі після одержання інформації не змінилася і, як наслідок, її цінність дорівнює нулеві.

2. Є один успішний результат із трьох: $P_1 = 1/3$ (рис. 1, *б*) і $\text{Ц} = \log_2[P_1/P_0] = \log_2(2/3) = -0,58$. Враховуючи, що мандрівник був посланий у напрямку, протилежному цілі, негативна цінність дає змогу назвати таку пораду *дезінформацією*. Дезінформація, збільшуючи вихідну невизначеність (від вибору одного з двох шляхів до одного з трьох), зменшує ймовірність досягнення цілі і має негативну цінність.

3. Є два успішних результати з трьох: $P_1 = 2/3$ (рис. 1, *в*) і $\text{Ц} = \log_2[P_1/P_0] = \log_2(4/3) = 0,42$.

Найбільшу цінність має порада “йдіть правим шляхом” ($I = 1$ біт), що направляє мандрівника з *A* до *C*; тоді $P_1 = 1$ і $\text{Ц} = \log_2[P_1/P_0] = \log_2 2 = 1$. Наведений приклад показує, що цінність тієї самої кількості інформації може змінюватися в широких межах.

Інший підхід до визначення цінності інформації пов’язаний з працями Р. Л. Стратоновича. Розглянемо систему (рис. 1, *д*), у котрій деякий спостерігач робить пошук, здійснює вибір між гіпотезами або оцінює невідому величину. Підсумок своєї діяльності він подає в блок, де оцінюється результат і призначається штраф. Розмір штрафу чи втрат обчислюється відповідно до функції штрафу за величиною помилки, зробленої спостерігачем. *Функцією штрафу* називають залежність між розміром штрафу і величиною помилки, за яку він призначається. Якщо спостерігач нічого не знає про об’єкт і дію, наприклад, або навмання, або методом спроб і помилок, йому можна допомогти, вказавши, як звузити область пошуку. Завдяки обраному методу він діятиме ефективніше. Це проявиться в тому, що середнє значення його штрафів зменшиться.

Указівки дає джерело інформації, їх повідомляють спостерігачу. Вважається, що кількість інформації, яка міститься в кожному отриманому спостерігачем повідомленні, відома. До одержання інформації спостерігач діє в умовах невизначеності, що оцінюється ентропією H і втратами $R(H)$. Отримана і використана інформація спричинює нову, меншу, невизначеність H_1 і нові втрати $R(H_1)$. Різниця втрат $R(H) - R(H_1) = \Delta R$ кількісно показує користь, принесену інформацією, або *кількісну міру цінності інформації*. Отже, цінність інформації можна визначити за розміром штрафу.

Розмір втрат (штрафу) залежить не тільки від корисності інформації. Спостерігач може поводитися “нерозумно” і не користуватися отриманою цінною інформацією. Тому передбачається, що спостерігач отриману інформацію використовує для себе за допомогою найкращих доступних йому методів. Потрібна така функція штрафів, щоб максимально зменшити втрати внаслідок надходження інформації.

У багатьох випадках, коли потрібно узгодити канал із джерелом повідомлень, знати кількість інформації, що міститься в окремих повідомленнях, недостатньо. Виникає потреба у характеристиках, які б давали можливість оцінювати інформаційні властивості джерела повідомлень у цілому. Однією з таких важливих характеристик є середня кількість інформації одного повідомлення.

У найпростішому випадку, коли всі повідомлення рівноймовірні, тобто $P(a_i) = 1/m$, середня кількість інформації дорівнює $\log m$. Отже, при рівноймовірних незалежних повідомленнях

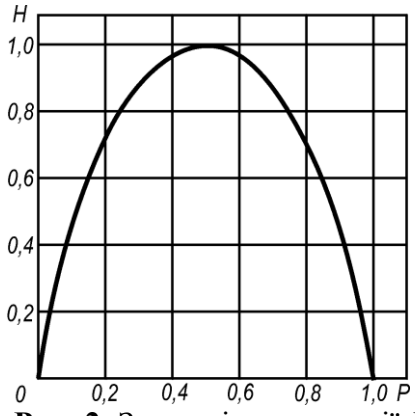

Рис. 2. Залежність ентропії H від імовірності P

інформаційні властивості джерела залежать тільки від числа повідомлень в ансамблі m. Однак у реальних умовах повідомлення, як правило, мають різну ймовірність. Так, деякі букви алфавіту зустрічаються частіше за інші. Тому крім знання
числа повідомлень m в ансамблі треба мати відомості про ймовірності кожного повідомлення $P(a_i),\ i=\overline{1,m}$.

Оскільки ймовірності повідомлень неоднакові, вони несуть різну кількість інформації. Менш імовірні повідомлення несуть більшу кількість інформації, і навпаки.

Середня кількість інформації одного повідомлення (біт/повідомлення) визначається як математичне сподівання

$$H(a)=\overline{I(a_i)}=\sum_{i=1}^{m}P(a_i)I(a_i)=-\sum_{i=1}^{m}P(a_i)\log P(a_i).$$

Величина $H(a)$ називається *ентропією*. Цей термін запозичений з термодинаміки, де є аналогічний за своєю формою вираз, який характеризує невизначеність стану фізичної системи.

В теорії інформації ентропія $H(a)$ також характеризується невизначеністю ситуації до передавання повідомлення, оскільки заздалегідь невідомо, яке саме повідомлення з ансамблю повідомлень джерела буде передано. Чим більша ентропія, тим сильніша невизначеність, і тим більшу інформацію несе одне повідомлення джерела.

Властивості ентропії джерела незалежних дискретних повідомлень. Ентропія як міра невизначеності скінченної системи подій (повідомлень) має такі основні властивості.

1. У тому разі, коли події (повідомлення) є рівноймовірними, зі збільшенням числа можливих подій m, що складають скінченну систему подій (повну групу), невизначеність цієї системи подій збільшується.

Коли всі повідомлення рівноймовірні $[P(a_i)=1/m]$, ентропія системи дорівнює

$$H(a)=-\sum_{i=1}^{m}P(a_i)\log P(a_i)=-m\frac{1}{m}\log(1/m)=\log m$$

і збільшується за логарифмічним законом.

2. Якщо ймовірність якого-небудь повідомлення дорівнює одиниці $[P(a_i)=1]$, то ентропія скінченної системи дорівнює нулеві.

Враховуючи, що $\sum_{i=1}^{m}P(a_i)=1,$ одержуємо

$$H(a)=-[1\cdot\log 1+(m+1)\cdot 0\cdot\log 0].$$

Очевидно, що $1\cdot\log 1=0$, оскільки $\log 1=0$. Добуток виду $0\cdot\log 0$ є невизначеністю типу $0\cdot(\infty)$, тому що $\log 0=-\infty$. Ця невизначеність дорівнює нулеві:

$$\lim_{x\to 0}x\ \log x=\lim_{x\to 0}\frac{\log x}{x^{-1}}=\lim_{x\to 0}\frac{\frac{d(\log x)}{dx}}{\frac{d(x^{-1})}{dx}}=\lim_{x\to 0}\frac{\frac{d(c\ln x)}{dx}}{x^{-2}}=\lim_{x\to 0}cx=0.$$

Отже, при $P(a_i)=0$ ентропія системи повідомлень $H(a)$ дорівнює нулеві.

3. Невизначеність скінченної системи повідомлень при даному m максимальна при рівноймовірних повідомленнях. Доведення цієї властивості в загальному вигляді є досить складним. Тому обмежимося прикладом, коли загальне число подій $m=2$:

$$\sum_{i=1}^{2}P(a_i)=1;\quad P(a_i)=P;\quad P(a_2)=1-P;$$

$$H(a)=-P\log P-(1-P)\log(1-P).$$

Ентропія скінченної системи, що складається з двох подій, досягає свого максимального

значення, яке дорівнює одиниці, при $P=0{,}5$, тобто при однаковій імовірності подій, оскільки при цьому $1-P=1-0{,}5=0{,}5$.

Залежність $H(a)$ від P показана на рис. 2. Максимум має місце при $P=0{,}5$, тобто коли ситуація є найбільш невизначеною. При $P=0$ або $P=1$, що відповідає передаванню одного з повідомлень (a_1 чи a_2), невизначеність відсутня. В цих випадках ентропія $H(a)$ дорівнює нулю.

4. Ентропія має властивість адитивності, тобто невизначеність сполученої системи подій дорівнює сумі ентропій незалежних скінченних систем подій, що складають сполучену систему.

Розглянемо такі дві незалежні скінченні системи подій:

$$a=\begin{pmatrix} a_1, a_2, \ldots, a_n \\ P_1, P_2, \ldots, P_n \end{pmatrix}; \quad b=\begin{pmatrix} b_1, b_2, \ldots, b_m \\ q_1, q_2, \ldots, q_m \end{pmatrix},$$

де P_i, q_i – імовірності подій a_i і b_i відповідно.

Ентропія цих скінченних систем подій відповідно

$$H(a)=-\sum_{i=1}^{n} P_i \log P_i; \quad H(b)=-\sum_{i=1}^{m} q_i \log q_i.$$

Припустимо, що кожна з подій a_i настає одночасно з однією з подій b_i. Тоді можна розглядати нову систему подій, в яку входить повна група сполучених подій

$$ab=\begin{pmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_m \\ a_2b_1 & a_2b_2 & \cdots & a_2b_m \\ \cdots & \cdots & \cdots & \cdots \\ a_nb_1 & a_nb_2 & \cdots & a_nb_m \end{pmatrix} \tag{9}$$

з їх імовірностями

$$\begin{pmatrix} P_1q_1 & P_1q_2 & \cdots & P_1q_m \\ \cdots & \cdots & \cdots & \cdots \\ P_nq_1 & P_nq_2 & \cdots & P_nq_m \end{pmatrix}.$$

Відповідно до загальної формули (8) запишемо значення ентропії для сполученої системи (9). Очевидно, при цьому необхідно скласти добутки всіх імовірностей сполучених подій на логарифми цих імовірностей, тобто

$$H(ab)=-\sum_{i=1}^{n}\sum_{i=1}^{m} P_iq_i \log P_iq_i.$$

Враховуючи, що

$$\log P_iq_i=\log P_i+\log P_j; \quad \sum_{i=1}^{n} P_i=1; \quad \sum_{j=1}^{m} q_j=1,$$

після простих перетворень одержимо:

$$H(ab)=-\sum_{j=1}^{m} q_j \sum_{i=1}^{n} P_i \log P_i-\sum_{i=1}^{n} q_j \sum_{j=1}^{m} q_j \log q_j=$$

$$=-\sum_{i=1}^{n} P_i \log P_i-\sum_{j=1}^{m} q_j \log q_j=H(a)+H(b), \tag{10}$$

що і потрібно було довести.

Ентропія джерела залежних дискретних повідомлень. Джерела незалежних повідомлень є найпростішим типом джерел. У реальних умовах все значно ускладнюється через наявність статистичних зв'язків між повідомленнями.

Статистичний зв'язок очікуваного повідомлення з попереднім повідомленням кількісно оцінюється спільною $P(a_i, a_k)$ або умовною ймовірністю $P(a_i / a_k)$.

Кількість інформації, що міститься в повідомленні a_i за умови, що попереднє повідомлення a_k відоме,

$$I(a_i / a_k) = -\log P(a_i / a_k).$$

Середня кількість інформації визначається умовною ентропією, яка обчислюється як математичне сподівання:

$$H_2(a) = H(a_i / a_k) = -\sum_{i=1}^{n}\sum_{k=1}^{m} P(a_i, a_k)\log P(a_i / a_k).$$

У загальному випадку для *n* залежних повідомлень

$$H(a) = H(a_i / a_k, \ldots, a_i) = -\sum_{i=1}^{n}\sum_{k=1}^{m}\ldots\sum_{l=1}^{m} P(a_i, a_k, \ldots, a_l)\log P(a_i / a_k, \ldots, a_l).$$

Враховуючи, що при рівноймовірних повідомленнях

$$H_0(a) = \log m = H_{\max}, \tag{11}$$

а при нерівноймовірних, але незалежних,

$$H_1(a) = -\sum_{i=1}^{n} P_i \log P_i, \tag{12}$$

запишемо значення ентропії в порядку її зменшення при збільшенні статистичних зв'язків між повідомленнями, включаючи також формули (11) і (12), так:

$$H_0(a) = \log m = H_{\max};$$

$$H_1(a) = -\sum_{i=1}^{n} P(a_i)\log P(a_i);$$

$$H_2(a) = H(a_i / a_k) = -\sum_{i=1}^{n}\sum_{k=1}^{m} P(a_i, a_k)\log P(a_i / a_k);$$

$$H_3(a) = -\sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} P(a_i, a_k, a_l)\log P(a_i / a_k, a_l);$$

……

$$H_n(a) = -\sum_{i=1}^{n}\cdots\sum_{k=1}^{m}\sum_{l=1}^{m} P(a_i, a_k, \ldots, a_l)\log P(a_i / a_z, \ldots, a_l).$$

$$H_0 > H_1 > H_2 > \ldots > H_n.$$

Надлишковість джерела повідомлення. Важливою властивістю умовної ентропії джерела залежних повідомлень є те, що при незмінній кількості повідомлень в ансамблі джерела його ентропія зменшується зі збільшенням частоти повідомлень, між якими існує статистичний взаємозв'язок. Відповідно до цієї властивості наявність статистичних зв'язків між повідомленнями завжди спричинює зменшення кількості інформації одного повідомлення.

Зменшення ентропії джерела зі збільшенням статистичного взаємозв'язку можна розглядати як зниження інформаційної ємності повідомлень. Те саме повідомлення за наявності взаємозв'язку містить в середньому менше інформації, ніж за його відсутності. Інакше кажучи, якщо джерело створює послідовність повідомлень, які мають статистичний зв'язок, і характер цього зв'язку відомий, то частина повідомлень, що видається джерелом, є надлишковою, тобто її можна відновити за відомими статистичними зв'язками. З'являється можливість передавання повідомлень у скороченому вигляді без втрати інформації, що міститься в них. Наприклад, передаючи телеграму, ми виключаємо з тексту сполучники, розділові знаки, оскільки вони легко відновлюються при читанні телеграми за відомими правилами побудови слів і фраз.

Отже, будь-яке джерело має надлишок, кількісне визначення якого можна отримати з наступних міркувань. Щоб передати кількість інформації *I*, джерело без надлишковості має видати в середньому повідомлень

$$k_0 = I / H_0,$$

а з надлишковістю

$$k_n = I / H_n.$$

Оскільки $H_0 > H_n$, то $k_n > k_0$, тобто джерело з надлишковістю для передавання тієї самої кількості інформації має використовувати більшу кількість повідомлень.

Надлишкова кількість повідомлень

$$\Delta k = k_n - k_0,$$

а надлишок джерела

$$\chi = \frac{k_n - k_0}{k_n} = \frac{\Delta k}{k_n} = 1 - H_n / H_0. \qquad (13)$$

Величина надлишку лежить у межах $0 \le \chi < 1$ і є неспадною функцією *n*.

Для української мови:

$$H_0 = 5 \text{ дв.од./літеру;}$$
$$H_1 = 0,5 \text{ дв.од./літеру;}$$
$$H_2 = 3,52 \text{ дв.од./літеру;}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$H_8 = 2 \text{ дв.од./літеру.}$$

Звідси надлишок української мови має порядок 60 %.

Коефіцієнт

$$r = H_n / H_0 \qquad (14)$$

називається *коефіцієнтом стискання*. Він показує, до якої величини можна стиснути повідомлення, що передаються, якщо усунути надлишок.

Оскільки джерело з надлишковістю передає зайву кількість повідомлень, тривалість передавання зростає і ефективність використання каналу зв'язку знижується. Стискання повідомлень можна здійснити за допомогою відповідного кодування.

Але надлишок джерела не завжди є негативною властивістю. Наявність взаємозв'язку між літерами тексту дає можливість відновлювати його при спотворенні окремих літер, тобто використовувати для підвищення вірогідності.

Статистичні властивості джерел повідомлення. Використання ентропії як усередненої величини, що кількісно характеризує інформаційні властивості джерела, котре видає послідовності дискретних повідомлень, є доцільним за умови, що імовірнісні співвідношення для цих послідовностей зберігаються незмінними. Джерело називають *стаціонарним*, коли розподіл імовірностей повідомлень не залежить від їх місця в послідовності повідомлень, що створюються цим джерелом, тобто

$$P(a_l / a_i, \ldots, a_k) = P(a_{l+n} / a_{i+n}, \ldots, a_{k+n}), \qquad (15)$$

де – будь-яке ціле число.

За аналогією зі стаціонарним випадковим процесом статистичні характеристики послідовності повідомлень стаціонарного джерела не залежать від вибору початку відліку.

Серед стаціонарних джерел повідомлень важливе місце займають *ергодичні* джерела, які відрізняються тим, що з імовірністю, близькою до одиниці, будь-яка достатньо довга послідовність повідомлень такого джерела повністю характеризує його статистичні властивості. Важливою особливістю ергодичних джерел є те, що статистичний взаємозв'язок між повідомленнями завжди розповсюджується тільки на скінченне число попередніх повідомлень.

Існують стаціонарні джерела, які можуть працювати в різних за своїми статистичними характеристиками режимах. У цьому випадку джерело не є ергодичним, оскільки при роботі в одному режимі навіть тривала послідовність повідомлень вже не може в цілому характеризувати властивості джерела.

Умовна ентропія стаціонарного джерела знаходиться як результат усереднення в усіх режимах роботи:

$$H_n(a) = \sum_j P_j H_{nj}(a), \qquad (16)$$

де P_j і $H_{nj}(a)$ – імовірність й умовна ентропія j-го режиму роботи відповідно.

Розглянемо умовну ентропію при заданій послідовності попередніх повідомлень:

$$H_n(a/h_a) = -\sum_{l=1}^{m} P(a_l/h_a)\log(a_l/h_a). \quad (17)$$

Тут символом h_a позначена послідовність $n-1$ попередніх повідомлень $a_i,\ldots,a_k$, причому індексом a нумеруються сполуки з m повідомлень по $n-1$, тобто всього $N = m^{n-1}$ послідовностей h_a.

Послідовність h_a можна трактувати як стан джерела, в якому воно знаходиться після передавання повідомлення. Подібні випадкові послідовності (які мають ергодичні властивості) відомі в математиці як дискретні ланки А. А. Маркова. В марковському ергодичному джерелі ймовірність передавання того чи іншого повідомлення однозначно визначається станом джерела. Після передавання повідомлення джерело переходить в новий стан, який залежить від попереднього стану і переданого повідомлення.

Достатньо довгі ергодичні послідовності повідомлень з високим ступенем імовірності, які містять всі відомості про статистичні характеристики джерела, називаються *типовими*. Чим довша послідовність, тим більша ймовірність того, що вона є типовою. В *типових послідовностях частота появи окремих повідомлень або груп повідомлень як завгодно мало відрізняється від їх імовірності.* Звідси випливає важлива властивість типових послідовностей: *типові послідовності однакової довжини приблизно рівноймовірні.* Це легко показати для послідовностей незалежних повідомлень.

Припустимо, що ансамбль повідомлень складається з m повідомлень $a_1, a_2, \ldots, a_m$, і нас цікавить імовірність того, що в послідовності, довжина якої повідомлень, число повідомлень a_1 дорівнює k_1, число повідомлень a_2 дорівнює $k_2, \ldots$; a_m відповідає k_m, причому $k_1 + k_2 + \ldots + k_m = n$. Якщо повідомлення незалежні, ця імовірність, очевидно, дорівнює

$$P_1^{k_1} P_2^{k_2} \ldots P_m^{k_m}, \quad (18)$$

де P_i – імовірність повідомлення a_i. Оскільки в усіх типових послідовностях за означенням виконується умова $k_1 \approx nP_i$, то ймовірності типових послідовностей приблизно однакові й дорівнюють

$$P_n = (P_1^{k_1} P_2^{k_2} \ldots P_m^{k_m})^n. \quad (19)$$

Тоді кількість інформації в будь-якій з типових послідовностей

$$I_n \approx \log\frac{1}{P_n}. \quad (20)$$

З іншого боку, величину I_n можна виразити через ентропію джерела $H(a)$:

$$I_n \approx nH(a). \quad (21)$$

Використовуючи вирази (20) і (21), ентропію джерела можна визначити так:

$$H(a) \approx \frac{1}{n}\log\frac{1}{P_n}. \quad (22)$$

У загальному випадку, в тому числі й для залежних повідомлень, у теорії інформації доводиться теорема, згідно з якою всі ергодичні послідовності, що містять достатньо велике число повідомлень n, можуть бути розбиті на дві групи:

- типові послідовності з імовірностями P_n, для яких задовольняється нерівність

$$\left| H(a) - \frac{1}{n}\log\frac{1}{P_n} \right| < \varepsilon, \quad (23)$$

де $H(a)$ – ентропія джерела; ε – як завгодно мала величина;

- нетипові послідовності, сумарна ймовірність яких δ як завгодно мала.

Величини ε і δ необмежено зменшуються з ростом , що дозволяє завжди вибрати таке значення , при якому всі послідовності джерела, за винятком дуже малої їх частини, можуть бути віднесені до рівноймовірних типових послідовностей. Важливим наслідком теореми є можливість встановлення залежності між числом варіантів всіх можливих типових

послідовностей M_n і ентропією джерела. Для достатньо довгих послідовностей величини ε і δ малі. Тоді на підставі (23)

$$M_n \approx \frac{1}{P_n} \approx 2^{nH(a)}. \quad (24)$$

Що стосується нетипових послідовностей, то внаслідок їх малої ймовірності при великому вони в багатьох випадках взагалі не враховуються.

Контрольні питання до Теми 1:

1. Що таке кількісна міра інформації?
2. Як визначається ентропія джерела дискретних повідомлень?
3. Які властивості має ентропія системи дискретних повідомлень?
4. Як залежить ентропія від кількості можливих повідомлень?
5. Які основні властивості адитивності ентропії?

ТЕМА 2 ВЗАЄМНА ІНФОРМАЦІЯ. ШВИДКІСТЬ І ПРОПУСКНА ЗДАТНІСТЬ ДИСКРЕТНОГО КАНАЛУ БЕЗ ЗАВАД

Ентропія ансамблю характеризує середню кількість повної інформації, що міститься в

повідомленні. Визначимо тепер інформацію, що міститься в одному ансамблі відносно іншого (наприклад, в прийнятому сигналі відносно переданого повідомлення). Для цього розглянемо об'єднання двох залежних дискретних ансамблів A і B. Його можна інтерпретувати або як пару ансамблів повідомлень, або як ансамблі повідомлення і сигналу, за допомогою якого повідомлення передається, або як ансамблі сигналів на вході і виході каналу зв'язку і т. д. Нехай $P(a_k, b_l)$ – спільна ймовірність реалізацій a_k і b_l. *Спільною ентропією ансамблів A* і *B* будемо називати

$$H(A,B) = M\left\{\log\frac{1}{P(a_k,b_l)}\right\}. \tag{25}$$

Введемо також поняття *умовної ентропії*:

$$H(A/B) = M\left\{\log\frac{1}{P(a_k/b_l)}\right\}, \tag{26}$$

де $P(a_k/b_l)$ – умовна імовірність a_k, якщо має місце b_l; тут математичні сподівання беруться за об'єднаним ансамблем AB. Зокрема, для джерел без пам'яті

$$H(A/B) = \sum_k \sum_l P(a_k,b_l)\log\frac{1}{P(a_k/b_l)}. \tag{27}$$

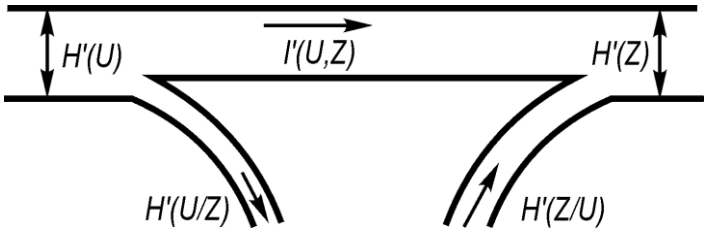

Рис. 3. Передавання інформації по каналу з завадами

Із теореми множення ймовірностей $P(a,b) = P(a)P(b/a) = P(b)P(a/b)$ випливає, що

$$H(A,B) = H(A) + H(B/A) = H(B) + H(A/B). \tag{28}$$

Для умовної ентропії слушна подвійна нерівність

$$0 \le H(A/B) \le H(A). \tag{29}$$

Рівність $H(A/B) = 0$, як видно з (26), виконується, коли при кожному значенні b_l умовна ймовірність однієї реалізації $P(a_k/b_l) = 1$, а для решти реалізацій $P(a_k/b_l) = 0$. Це означає, що за відомою реалізацію B можна точно встановити реалізацію A. Іншими словами, B містить повну інформацію про A.

Рівність $H(A/B) = H(A)$ має місце, якщо $P(a_k/b_l) = P(a_k)$ при всіх a і b, тобто коли події A і B незалежні. У такому разі знання реалізації B не зменшує невизначеності A, тобто B не містить ніякої інформації про A.

У загальному випадку ентропія $H(A/B)$ менша від безумовної $H(A)$, і знання реалізації B знижує в середньому початкову невизначеність A. Природно назвати різницю $H(A) - H(A/B)$

кількістю інформації, що міститься в B відносно A. Її називають також *взаємною інформацією* між *A* і *B* та позначають $I(A,B)$:

$$I(A,B) = H(A) - H(A/B). \tag{30}$$

Підставивши в (30) вираз (26) і $H(A) = M\left\{\log\frac{1}{P(a)}\right\}$, зобразимо взаємну інформацію через розподіл імовірностей:

$$I(A,B) = M\left\{\log\frac{1}{P(a_k)}\right\} - M\left\{\log\frac{1}{P(a_k/b_l)}\right\} = M\left\{\log\frac{P(a_k/b_l)}{P(a_k)}\right\}. \tag{31}$$

Якщо скористатися теоремою множення $P(a_k,b_l) = P(b_l)P(a_k/b_l)$, то можна записати $I(A,B)$ у симетричній формі:

$$I(A,B) = M\left\{\log\frac{P(a_k,b_l)}{P(a_k)P(b_l)}\right\}. \tag{32}$$

Взаємна інформація вимірюється в тих самих одиницях, що й ентропія, наприклад в бітах. Величина $I(A,B)$ показує, яку в середньому отримуємо кількість інформації про реалізацію ансамблю *A*, спостерігаючи реалізацію ансамблю *B*.

Сформулюємо основні властивості взаємної інформації:

$$I(A,B) \geq 0, \tag{33}$$

причому рівність має місце тоді і тільки тоді, коли *A* і *B* незалежні між собою. Це випливає з (30) та (29);

$$I(A,B) = I(B,A), \tag{34}$$

тобто *B* містить стільки ж інформації відносно *A*, скільки *A* містить відносно *B*. Ця властивість випливає з симетрії виразу (32). Тому можна також записати

$$I(A,B) = H(B) - H(B/A); \tag{35}$$

$$I(A,B) \leq H(A), \tag{36}$$

причому рівність має місце, коли за реалізацією *B* можна однозначно відновити *A*. Це випливає з (29) і (30);

$$I(A,B) \leq H(B), \tag{37}$$

причому рівність має місце, коли за реалізацією *A* можна точно встановити реалізацію *B*. Це випливає з (34) та (36).

Взявши в (30) *B*=*A* і врахувавши, що $H(A/A) = 0$, дістанемо

$$I(A,A) = H(A),$$

що дозволяє інтерпретувати ентропію джерела як його *власну інформацію*, тобто інформацію, що міститься в ансамблі *A* про саму себе.

Отримані співвідношення дозволяють поглянути на сутність ентропії з інших точок зору. Так, із (37) видно, що ентропія ансамблю *A* є максимальною кількістю інформації, яка може міститися в *A* відносно будь-якого іншого ансамблю *B*. Із (30) випливає, що для того, щоб відновити реалізацію ансамблю *A* в точності, необхідно передати в середньому кількість інформації, рівну ентропії *A*.

Нехай *A* – ансамбль дискетних повідомлень, а *B* – ансамбль дискретних сигналів, в які перетворюються повідомлення *A*. Тоді $I(A,B) = H(A)$ в тому і тільки в тому випадку, коли перетворення $A \to B$ оборотне. При необоротному перетворенні $I(A,B) < H(A)$, і різницю $H(A) - I(A,B) = = H(A/B)$ можна назвати *втратою (ненадійністю) інформації* при перетворенні $A \to B$. Таким чином, інформація не втрачається тільки при оборотних перетвореннях.

Якщо *T* – середній час передавання одного повідомлення, то поділивши формули (25) – (37) на *T* і позначивши

$$H'(A/B) = \frac{1}{T}H(A/B), \quad I'(A,B) = \frac{1}{T}I(A,B), \ldots,$$

отримаємо відповідні рівності для ентропій і кількості інформації, розраховані не на одне повідомлення, а на одиницю часу. Величина $I'(A,B)$ називається *швидкістю передавання інформації від A до B* (чи навпаки).

Якщо, наприклад, U – ансамбль сигналів на вході дискретного каналу, а Z – ансамбль сигналів на його виході, то швидкість передавання інформації по каналу

$$I'(U,Z) = H'(U) - H'(U/Z) = H'(Z) - H'(Z/U). \tag{38}$$

Ці співвідношення наглядно ілюструє рис. 3. Тут $H'(U)$ – продуктивність джерела сигналу U, який передається, а $H'(Z)$ – продуктивність каналу, тобто повна власна інформація в прийнятому сигналі за одиницю часу. Величина $H'(U/Z)$ – це швидкість "витоку" інформації при проходженні через канал, а $H'(Z/U)$ – швидкість передавання побічної інформації, що не має відношення до U і створюється присутніми в каналі завадами. Співвідношення між $H'(U/Z)$ і $H'(Z/U)$ залежать від властивостей каналу. Так, при передаванні телефонного сигналу по каналу з вузькою смугою пропускання, недостатньою для задовільного відтворення сигналу, та з низьким рівнем завад втрачається частина корисної інформації, але майже не отримується некорисної. В цьому випадку $H'(U/Z) >> H'(Z/U)$. Якщо ж сигнал відтворюється точно, але в паузах ясно прослуховуються "наводки" від сусіднього телефонного каналу, то, майже не втрачаючи корисної інформації, можна отримати багато додаткової, як правило, некорисної інформації, і $H'(U/Z) << H'(Z/U)$.

Швидкість передавання інформації R дорівнює кількості інформації, переданої в середньому за одиницю часу, і вимірюється в двійкових одиницях за загальну тривалість T (дв. од./с).

Для ергодичних послідовностей повідомлень, коли допускається усереднення за часом, швидкість

$$R = \lim_{T\to\infty} \frac{I(a_T)}{T}, \tag{39}$$

де $I(a_T)$ – кількість інформації в послідовності повідомлень a_T, загальна тривалість яких дорівнює T.

Кількість інформації, що створюється джерелом повідомлень у середньому за одиницю часу, називається *продуктивністю джерела* $R_д$. Враховуючи, що при $T \to \infty$

$$I(a_T) = nH(a), \quad T = n\bar{\tau},$$

де n – число повідомлень; $\bar{\tau}$ – середня тривалість повідомлення, одержимо

$$R_д = \lim_{n\to\infty} \frac{nH(a)}{n\bar{\tau}} = \frac{H(a)}{\bar{\tau}}, \tag{40}$$

де $\bar{\tau} = \sum_{i=1}^{n} \tau_i P(\tau_i) = \sum_{i=1}^{n} \tau_i P(a_i)$ – математичне сподівання τ_i; $P(\tau_i) = P(a_i)$ – імовірність повідомлення a_i тривалістю τ_i.

Для повідомлень однакової тривалості найбільшу продуктивність має джерело з максимальною ентропією $H_0 = \log m$, тобто

$$R_{д\max} = \frac{1}{\tau}\log m. \tag{41}$$

Видана джерелом інформація з каналів зв'язку передається за допомогою сигналів $u(t)$, які мають іншу природу та, у загальному випадку, інші

статистичні характеристики. При цьому швидкість визначається так:

$$R = \lim_{T\to\infty} \frac{I(u_T)}{T} = \frac{H(u_T)}{\bar{\tau}}. \tag{42}$$

Максимально можлива швидкість передавання інформації в каналах зв'язку при фіксованих обмеженнях називається *пропускною здатністю каналу*, дв. од./с:

$$C = \max R = \max \frac{H(u)}{\bar{\tau}} = \frac{\max H(u)}{\min \bar{\tau}} = \frac{\log m}{\tau}, \tag{43}$$

де $\overline{\tau} = \tau$ (тривалості сигналів однакові).

Для двійкових дискретних каналів (при $m = 2$) пропускна здатність

$$C = \frac{\log m}{\tau} = \frac{\log_2 2}{\tau} = \frac{1}{\tau}, \tag{44}$$

що збігається зі швидкістю телеграфування, бод:

$$C = \frac{1}{\tau} = \text{v}.$$

При телеграфуванні мінімальна смуга пропускання

$$F = F_m,$$

де $F_m = 1/(2\tau)$ – частота маніпуляції; $\tau = 1/(2F_m)$.

При цьому пропускна здатність (межа Найквіста)

$$C = \frac{1}{\tau} = \frac{1}{0{,}5F_m} = 2F_m. \tag{45}$$

Контрольні питання до Теми 2:

1. Що таке взаємна інформація?
2. Як визначається швидкість передачі інформації в дискретному каналі без завад?
3. Що таке пропускна здатність дискретного каналу і як вона розраховується?
4. Що таке теорема Шеннона для безпомилкової передачі інформації?
5. Як взаємна інформація допомагає оцінити ефективність системи передачі інформації?

ТЕМА 3 ОПТИМАЛЬНІ СТАТИСТИЧНІ КОДУВАННЯ ПОВІДОМЛЕНЬ. ДИСКРЕТНІ КАНАЛИ З ЗАВАДАМИ

Статистичним оптимальним називається кодування, при якому найкраще використовується пропускна здатність каналу без завад.

Для дискретних каналів без завад К. Е. Шенноном була доведена така теорема:

якщо продуктивність джерела $R_д \le C - \varepsilon$, де ε – як завгодно мала величина, то завжди існує спосіб кодування, який дозволяє передавати по каналу всі повідомлення джерела. Передати всі повідомлення при $R_д > C$ неможливо.

Суть теореми зводиться до того, що якою б великою не була надлишковість джерела, всі його повідомлення можуть бути передані по каналу, що легко доводиться від протилежного. Припустимо, що $R_д > C$, тоді для передавання всіх повідомлень джерела по каналу необхідно, щоб швидкість передавання інформації R була не меншою від $R_д$. Тоді маємо $R \ge R_д > C$, що неможливо, оскільки за означенням пропускна здатність $C = R_{\max}$.

Для раціонального використання пропускної здатності каналу необхідно застосовувати відповідні способи кодування. При оптимальному кодуванні фактична швидкість передавання інформації по каналу R наближається до пропускної здатності C, що досягається шляхом узгодження джерела з каналом. Кодування повідомлення має найкраще відповідати обмеженням, які накладаються на сигнали, що передаються каналом зв'язку. Тому структура оптимального коду залежить як від статистичних характеристик джерела, так і від особливостей каналу.

Розглянемо основні принципи оптимального кодування на практиці узгодження джерела незалежних повідомлень з двійковим каналом без завад. При цьому процес кодування полягає в однозначному перетворенні повідомлень джерела в двійкові кодові комбінації. Тоді ентропія кодових комбінацій дорівнюватиме ентропії джерела:

$$H(u) = H(a),$$

а швидкість

$$R = \frac{H(u)}{\overline{\tau}_k} = \frac{H(a)}{\overline{\tau}_k}, \tag{46}$$

де $\overline{\tau}_k$ – середня тривалість кодової комбінації, котра визначається як математичне сподівання:

$$\overline{\tau}_k = \sum_{i=1}^{m} \overline{\tau}_{ki} P(a_i) = \tau_0 \sum_{i=1}^{m} n_i P(a_i);$$

τ_0 – тривалість одного елемента коду; n_i – кількість елементів у комбінації, що присвоюється повідомленню a_i.

Враховуючи, що при незалежних нерівноймовірних повідомленнях

$$H(a) = -\sum_{i=1}^{m} P(a_i) \log P(a_i),$$

Отримуємо

$$R = \frac{\sum_{i=1}^{m} P(a_i) \log P(a_i)}{\tau_0 \sum_{i=1}^{m} n_i P(a_i)}. \qquad (47)$$

Чисельник цього виразу визначається виключно статистичними властивостями джерела, а величина τ_0 – характеристиками каналу.

Як видно з (47), для того щоб швидкість передавання набула свого максимального значення $C = \frac{1}{\tau_0}$ (межі Найквіста), необхідно виконати умову

$$n_i = -\log P(a_i) = I(a_i), \qquad (48)$$

що відповідає мінімуму $\overline{\tau}_k$ і максимуму R. Очевидно, вибір $n_i < I(a_i)$ не має сенсу, бо в такому разі $R > C$, що суперечить теоремі Шеннона для сигналів без завад.

Код Шеннона–Фано. Одним із кодів, які задовольняють умову (48), є код Шеннона–Фано. Сутність кодування полягає в наступному. Нехай маємо джерело повідомлення, яке виробляє чотири повідомлення: a_1, a_2, a_3, a_4 з імовірностями $P(a_1) = 0{,}5$; $P(a_2) = 0{,}25$; $P(a_3) = P(a_4) = 0{,}125$.

Усі повідомлення записуються до кодової таблиці в порядку зменшення їх імовірностей. Ці повідомлення спочатку поділяють на дві групи так, щоб суми їх імовірностей були однаковими. В даному випадку в першу групу входить повідомлення з імовірністю $P(a_1) = 0{,}5$, а в другу – повідомлення a_2, a_3 та a_4, сумарна ймовірність яких також дорівнює 0,5.

a_i	$P(a_i)$	Групи			Комбінації	n_i	$I(a_i)$
		I	II	III			
a_1	0,5	0			0	1	1
a_2	0,25	1	0		10	2	2
a_3	0,125	1	1	0	110	3	3
a_4	0,125	1	1	1	111	3	3

Комбінаціям, що відповідають повідомленням першої групи, надається за перший символ код 0, а комбінаціям другої групи – 1. Кожна з двох груп знову поділяється на дві групи за тим же правилом надання символів 0 і 1.

В ідеальному випадку після першого поділу ймовірності кожної групи мають дорівнювати 0,5, а після другого – 0,25 і т. д. Процес триває доти, поки в групі не залишиться по одному повідомленню.

При заданому розподілі ймовірностей повідомлень код стає нерівномірним – його комбінації мають різну кількість елементів n_i.

Швидкість передавання дорівнює пропускній здатності:

$$R = C = 1/\tau_0.$$

В цьому легко переконатися, підставивши числові значення:

$$R = \frac{-0,5\log 0,5 - 0,25\log 0,25 - 2\cdot 0,125\log 0,125}{(0,5\cdot 1 + 0,25\cdot 2 + 2\cdot 0,125\cdot 3)\tau_0} = \frac{1,75}{1,75\tau_0} = \frac{1}{\tau_0} = C.$$

Важливою властивістю коду Шеннона–Фано є те, що незважаючи на його нерівномірність, він *не потребує розмежувальних знаків*. Це зумовлено тим, що короткі комбінації не є початком більш довгих комбінацій. Вказану властивість легко перевірити на прикладі будь-якої послідовності:

$$\underset{a_2}{10}\quad \underset{a_1}{0}\quad \underset{a_2}{10}\quad \underset{a_3}{110}\quad \underset{a_4}{111}\ .$$

Отже, *основний принцип оптимального кодування* полягає в тому, що найбільш імовірним повідомленням мають надаватися короткі комбінації, а повідомленням з малою ймовірністю – довгі комбінації.

Основний недолік, зумовлений повною відсутністю надлишковості, полягає в тому, що оптимальні коди прийнятні тільки для каналів, в яких вплив завад незначний.

Наявність завад у каналах призводить до руйнування та незворотної втрати частини інформації, що надходить від джерела повідомлення.

Оскільки в дискретному каналі з завадами отриманому сигналу v_j може відповідати передавання одного з декількох сигналів u_i, то після отримання v_j залишається деяка невизначеність відносно переданого сигналу. Відповідність між u та v має випадковий характер, тому ступінь невизначеності характеризується умовною апостеріорною ймовірністю $P(u_i/\mathrm{v}_j)$, причому

$$P(u_i/\mathrm{v}_j) < 1.$$

Кількість інформації, потрібної для усунення невизначеності, що залишилася, дорівнює тій частині інформації, яку знищено внаслідок дії завад. Кількість отриманої інформації

$$I(u_i,\mathrm{v}_j) = \log\frac{1}{P(u_i)} - \log\frac{1}{P(u_i/\mathrm{v}_j)} = \log\frac{P(u_i/\mathrm{v}_j)}{P(u_i)}. \qquad (49)$$

Для оцінки середньої кількості отриманої інформації при передаванні одного повідомлення значення $I(u_i,\mathrm{v}_j)$ слід усереднити по всьому ансамблю u та v:

$$I(u,\mathrm{v}) = \bar{I}(u_i,\mathrm{v}_j)\sum_{i=1}^{m_u}\sum_{j=1}^{m_\mathrm{v}} P(u_i,\mathrm{v}_j)\log\frac{P(u_i/\mathrm{v}_j)}{P(u_i)} =$$

$$= \sum_{j=1}^{m_\mathrm{v}} P(\mathrm{v}_j)\sum_{i=1}^{m_u} P(u_i,\mathrm{v}_j)\log\frac{P(u_i/\mathrm{v}_j)}{P(u_i)}, \qquad (50)$$

де $P(u_i,\mathrm{v}_j) = P(\mathrm{v}_j)P(u_i/\mathrm{v}_j) = P(u_i)P(\mathrm{v}_j/u_i)$ – спільна ймовірність переданого та отриманого сигналів; m_u – кількість сигналів в ансамблі на вході; m_v – кількість сигналів в ансамблі на виході каналу. У загальному випадку $m_\mathrm{v} \neq m_u$.

Величина $I(u,\mathrm{v})$ характеризує середню кількість інформації, яку містить отриманий сигнал v відносно переданого сигналу u, і тому $I(u,\mathrm{v})$ називають також *середньою взаємною інформацією між* u та v.

Звичайно вираз для $I(u,\mathrm{v})$ подається у вигляді формули

$$I(u,\mathrm{v}) = H(u) - H(u/\mathrm{v}),$$

де $H(u) = -\sum_{i=1}^{m_u}\sum_{j=1}^{m_\mathrm{v}} P(u_i,\mathrm{v}_j)\log P(u_i) = -\sum_{i=1}^{m_u} P(u_i)\log P(u_i)$ – ентропія джерела сигналів;

$$H(u/v) = -\sum_{i=1}^{m_u}\sum_{j=1}^{m_\mathrm{v}} P(u_i,\mathrm{v}_j)\log P(u_i/\mathrm{v}_j) = -\sum_{j=1}^{m_\mathrm{v}} P(\mathrm{v}_j)\sum_{i=1}^{m_u} P(u_i/\mathrm{v}_j)\log P(u_i/\mathrm{v}_j) -$$

умовна ентропія або *ненадійність*.

Вираз для $I(u,\mathrm{v})$ показує, що середня кількість отриманої інформації дорівнює середній кількості переданої інформації $H(u)$ за мінусом середньої кількості інформації $H(u/\mathrm{v})$, втраченої в каналі внаслідок дії завад.

Інша форма запису:

$$I(u,\mathrm{v}) = H(\mathrm{v}) - H(u/\mathrm{v}),$$

де $H(\mathrm{v}) = -\sum_{j=1}^{m_\mathrm{v}} P(\mathrm{v}_i)\log P(\mathrm{v}_i)$ – ентропія каналу виходу;

$H(\mathrm{v}/u) = -\sum_{i=1}^{m_u} P(u_i)\sum_{j=1}^{m_\mathrm{v}} P(\mathrm{v}_j/u_i)\log P(u_i/\mathrm{v}_j)$ – умовна ентропія або ентропія шуму (визначає непотрібну частину інформації, яка міститься в отриманих сигналах внаслідок дії завад).

Поняття швидкості передавання та пропускної здатності, отримані для дискретного каналу без завад, можна поширити і на канали з завадами:

$$R = \lim_{T\to\infty} \frac{I(u_T,\mathrm{v}_T)}{T},$$

де u_T, v_T – відповідно послідовності сигналів тривалістю T, що передаються і отримуються.

Швидкість передавання можна подати й у вигляді

$$R = \frac{I(u,\mathrm{v})}{\overline{\tau}} = \frac{1}{\tau}\big[H(u) - H(u/\mathrm{v})\big] = \frac{1}{\tau}\big[H(\mathrm{v}) - H(\mathrm{v}/u)\big].$$

Пропускна здатність

$$C = \max R.$$

Для каналів з сигналами однакової тривалості пропускна здатність

$$C = \frac{\max I(u,\mathrm{v})}{\tau} = \frac{\max\big[H(u) - H(u/\mathrm{v})\big]}{\tau},$$

де максимум є по всіх можливих ансамблях сигналів u.

Для двійкового симетричного каналу

$$C = \frac{1}{\tau}\big[1 + P_0\log P_0 + (1-P_0)\log(1-P_0)\big] \qquad (51)$$

або

$$C = \mathrm{v}\big[1 + P_0\log P_0 + (1-P_0)\log(1-P_0)\big],$$

де P_0 – повна ймовірність помилки.

Для симетричного каналу при $m_u = m_\mathrm{v} > 2$

$$C = \frac{1}{\tau}\left[\log m + P_0\log\frac{m-1}{P_0} - (1-P_0)\log\frac{1}{1-P_0}\right].$$

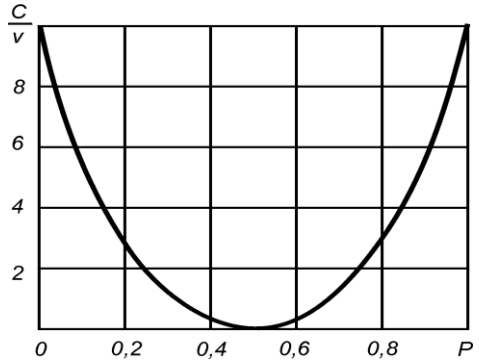

Рис. 4. Залежність пропускної здатності двійкового симетричного каналу без пам'яті від імовірності помилкового приймання символу

Залежність C/v від P відповідно до (51) зображена на рис. 4. При $P = 0{,}5$ пропускна здатність двійкового каналу $C = 0$, оскільки з такою ймовірністю помилки послідовність вихідних двійкових символів можна одержати, не передаючи сигнали по каналу, а вибираючи

їх навмання, тобто при $P = 0,5$ послідовності на вході і на виході каналу незалежні. Випадок $C = 0$ називають *обривом каналу*. Те, що пропускна здатність при $P = 1$ в двійковому каналі така ж сама, як і при $P = 0$ (канал без завад), пояснюється тим, що при $P = 1$ достатньо всі вихідні символи інвертувати (тобто замінити 0 на 1 і 1 на 0), щоб правильно відновити вхідний сигнал.

Потенціальні можливості передавання інформації, що характеризуються пропускною здатністю каналу, розкриваються в фундаментальній теоремі теорії інформації, відомій як основна теорема кодування К. Шеннона. Щодо дискретного джерела вона формулюється так: *якщо індуктивність джерела повідомлень $R_д$ менша від пропускної здатності каналу C, тобто $R_д < C$, то існує спосіб кодування (перетворення повідомлення в сигнал на вході) і декодування (перетворення сигналу в повідомлення на виході каналу), при якому ймовірність помилкового декодування та ненадійність можуть бути як завгодно малі. Якщо ж $R_д > C$, то таких способів не існує.*

Контрольні питання до Теми 3:

1. Які основні принципи оптимального кодування для дискретних каналів?
2. Як завади в каналі впливають на процес кодування?
3. Що таке умовна ентропія для каналу з завадами?
4. Які методи використовуються для покращення передачі в зашумлених каналах?

Тема 4 Інформація в неперервних повідомленнях. Епсилон-ентропія

Еквівалентність повідомлень. Джерело неперервних повідомлень за скінченний час T може видати будь-яку інформацію з нескінченної множини реалізацій деякого ансамблю повідомлень. Імовірність появи деякої конкретної реалізації дорівнює нулю. Якщо спробувати визначити ентропію та продуктивність такого джерела шляхом граничного переходу для неперервних повідомлень, то вони виявляться нескінченними.

Пояснення цього результату полягає в тому, що для передавання неперервного повідомлення з абсолютною точністю треба було б передати нескінченну кількість інформації, що, зрозуміло, неможливо зробити за скінченний час, користуючись каналом зі скінченною пропускною здатністю. Так само неперервне повідомлення не можна абсолютно точно запам’ятати (записати) за наявності як завгодно слабкої завади.

Тим не менш, як відомо, неперервні повідомлення (наприклад, телефонні, телевізійні) успішно передаються каналами зв’язку з завадами та записуються (наприклад, на магнітну плівку) завдяки тому, що на практиці ніколи не ставиться вимога точного відтворення переданого чи записаного повідомлення. А для передавання навіть з дуже високою, але обмеженою точністю, потрібна скінченна кількість інформації, як і для передавання дискретних повідомлень. Зрозуміло, ця кількість інформації тим більша, чим вища точність, з якою треба передати неперервне повідомлення. Нехай допустима точність вимірюється деяким параметром ε. Ту мінімальну кількість інформації, яку необхідно передати каналом зв’язку для відтворення неперервного повідомлення з *неточністю, не більшою за допустиму*, Колмогоров запропонував називати *ε-ентропією (епсилон-ентропією)*.

Критерій ε, який визначає необхідну точність, може бути яким завгодно. Називатимемо два варіанти повідомлення, які розрізняються не більше, ніж на ε, *еквівалентними*. Це означає, що у разі послання одного повідомлення, а прийняття іншого, еквівалентного йому за переданим критерієм, передане повідомлення вважається прийнятим правильно. Так, у системі телефонного зв’язку, якщо необхідно передати лише зміст розмови, однакові тексти, розбірливо прочитані двома різними дикторами (наприклад, чоловіком та жінкою), є еквівалентними повідомленнями, хоча вони різко відрізняються одне від одного навіть за спектром. Критерієм еквівалентності повідомлень тут є розбірливість мови. Для художніх

мовних передач такий критерій не прийнятний, оскільки в цьому разі істотні й більш тонкі характеристики повідомлення.

Ентропія неперервних повідомлень. При передаванні неперервних повідомлень передані сигнали $s(t)$ є функціями часу, які належать до деякої множини, а отримані сигнали $x(t)$ будуть їх реалізаціями.

Всі дійсні сигнали мають спектри, загальна енергія яких зосереджена в обмеженій смузі частот F. Згідно з теоремою В. О. Котельникова такі сигнали визначаються своїми значеннями в точках відліку, які вибирають через інтервали $\Delta t = 1/(2F)$.

У каналі на сигнал накладаються завади, внаслідок чого кількість різних рівнів сигналу в точках відліку буде скінченною. Отже, сукупність значень неперервного сигналу еквівалентна деякій дискретній скінченній сукупності. Це дає змогу визначити необхідну кількість інформації та пропускну здатність при передаванні неперервних повідомлень, що базуються на результатах, отриманих для дискретних повідомлень.

Визначимо спочатку кількість інформації, яка міститься в одному відліку сигналу x_j відносно переданого сигналу s_i:

$$I(s_i, x_j) = \lim_{\Delta s \to 0} \log \frac{p(s_i / x_j)}{p(s_i)\Delta s} = \log \frac{p(s_i / x_j)}{p(s_i)},$$

де $p(s_i)$, $p(s_i / x_j)$ – щільність імовірності.

Середня кількість інформації

$$\overline{I(s,x)} = \int_s \int_x p(s,x) I(s,x) ds\, dx = \int_s \int_x p(s,x) \log \frac{p(s/x)}{p(s)} dx\, ds,$$

де $p(s,x)$ – спільна щільність імовірності; s, x – області можливих значень s_i та x_j.

Вираз для $\overline{I(s,x)}$ можна подати так:

$$\left.\begin{aligned} \overline{I(s,x)} &= H(s) - H(s/x), \\ \overline{I(s,x)} &= H(x) - H(x/s), \end{aligned}\right\} \qquad (52)$$

де $H(s) = -\int_s p(s) \log p(s) ds$ – диференціальна ентропія сигналу s;

$H(x) = -\int_x p(x) \log p(x) dx$ – диференціальна ентропія сигналу x;

$H(s/x) = -\int_s \int_x p(x,s) \log p(s/x) ds\, dx$ – умовна диференціальна ентропія сигналу s;

$H(x/s) = -\int_s \int_x p(s,x) \log p(x/s) ds\, dx$ – умовна диференціальна ентропія сигналу або ентропія шуму.

Величина $H(s)$ характеризує інформаційні властивості сигналів; за формою запису вона аналогічна ентропії дискретних повідомлень. Оскільки у вирази (1.52) входить диференціальний розподіл імовірностей $p(s)$, то $H(s)$ називають диференціальною ентропією сигналу s. Вираз для $H(s/x)$ є умовною диференціальною ентропією сигналу s.

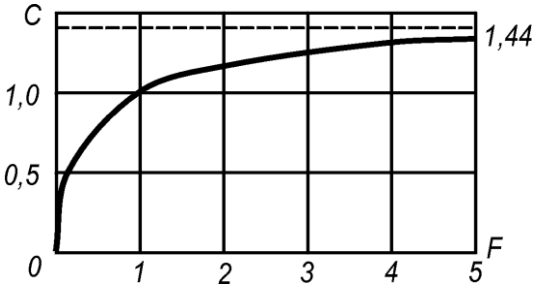

Рис. 1.5. Залежність пропускної здатності C від смуги пропускання F в неперервному каналі

Багато властивостей ентропії неперервного розподілу аналогічні властивостям ентропії дискретного сигналу.

Якщо неперервний сигнал s обмежений інтервалом $s_1 \le s < s_2$, то ентропія $H(s)$ максимальна і дорівнює $\log(s_2 - s_1)$, коли сигнал має рівномірний розподіл:

$$p(s) = 1/(s_2 - s_1).$$

Якщо середньоквадратичне значення неперервного сигналу

$$\int_{-\infty}^{\infty} s^2 p(s) ds = \sigma^2,$$

то ентропія максимальна при нормальному розподілі щільності ймовірностей, тобто при

$$p(s) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-s^2/(2\sigma^2)},$$

і дорівнює

$$H(s)_{\max} = \int_{-\infty}^{\infty} p(s) \log \frac{1}{P(s)} ds = \int_{-\infty}^{\infty} p(s) \log\left(\sqrt{2\pi} \cdot \sigma\right) ds +$$

$$+ \int_{-\infty}^{\infty} p(s) \frac{s^2 \log e}{2\sigma^2} ds = \log\left(\sqrt{2\pi e} \cdot \sigma\right), \qquad (53)$$

де $e = 2{,}71$.

Диференціальна ентропія (на відміну від ентропії дискретних сигналів) залежить від розмірності неперервного сигналу, завдяки чому вона не є мірою кількості інформації, хоч і характеризує ступінь невизначеності, яка властива джерелу. Тільки різниця ентропій кількісно визначає середню кількість інформації.

Формула Шеннона. Для визначення середньої кількості інформації $\overline{I(s,x)}$, яка передається сигналом на інтервалі T, необхідно розглянути $n = 2FT$ відліків неперервного сигналу на вході каналу $(s_1, s_2, \ldots, s_n)$ і на виході $(x_1, x_2, \ldots, x_n)$. При цьому

$$\overline{I(s,x)} = \begin{cases} H_T(s) - H_T(s/x), \\ H_T(x) - H_T(s), \end{cases} \qquad (54)$$

де

$$H_T(x) = -\int_{x_1}\int_{x_2}\cdots\int_{x_n} p(x_1, x_2, \ldots, x_n) \log p(x_1, x_2, \ldots, x_n) dx_1 \ldots dx_n,$$

$$H_T(x/s) = -\int_{s_1}\int_{s_2}\cdots\int_{s_n}\int_{x_1}\int_{x_2}\cdots\int_{x_n} p(s_1, s_2, \ldots, s_n; x_1, x_2, \ldots x_n) \times$$

$$\times \log p(s_1, s_2, \ldots, s_n; x_1, x_2, \ldots x_n) ds_1, ds_2 \ldots ds_n dx_1 dx_2 \ldots dx_n.$$

Вирази для ентропії $H_T(s)$ та $H_T(s/x)$ аналогічні. Швидкість передавання неперервним каналом

$$R = \lim_{T\to\infty} \frac{\overline{I(s,\, x)}}{T};$$

$$C = \max R = \max \lim_{T\to\infty} \frac{1}{T} \overline{I(s,\, x)},$$

де максимум обчислюють по всіх можливих ансамблях вхідних сигналів s.

Для сигналу s і завади ω за формулою Шеннона

$$C = F \log\left(\frac{P_c}{P_з} + 1\right), \qquad (55)$$

де F – ширина спектра сигналу; P_c – потужність сигналу; $P_з$ – потужність завади (шуму).

Формулу отримано в припущенні, що x і ω мають нормальний розподіл, тому сигнал s також повинен мати нормальний розподіл щільності ймовірності. Отже, максимальну швидкість передавання можна забезпечити сигналами з нормальним розподілом щільності

ймовірності та рівномірним спектром. Формула Шеннона відіграє важливу роль у теорії та техніці зв'язку. Оскільки при *рівномірному спектрі* $P_з = N_0 F$, то

$$C = F \log\left(\frac{P_с}{N_0 F} + 1\right).$$

Зі збільшенням F пропускна здатність монотонно зростає і наближається (рис.5) до величини, дв. од./с,

$$C_{\max} = \frac{P_с}{N_0} \log e = 1,44 \frac{P_с}{N_0}.$$

Формулу Шеннона (55), яку виведено для рівномірних спектрів сигналу і шуму, можна поширити також на випадок *нерівномірних спектрів*. З цією метою в межах деякої частоти f виберемо достатньо вузьку смугу Δf , в якій енергетичні спектри сигналу $G_с(t)$ та завади $G_з(t)$ будуть постійними. При цьому для такої смуги

$$\Delta C = \Delta f \log\left(\frac{\Delta P_с}{\Delta P_з} + 1\right) = \Delta f \log\left(\frac{G_с(f)}{G_з(f)} + 1\right),$$

де $\Delta P_с = \Delta f G_с(f)$; $\Delta P_з = \Delta f G_з(f)$.

Повна пропускна здатність обчислюється як інтеграл за всіма частотами спектра сигналу:

$$C = \int_{f_1}^{f_2} \log\left(\frac{G_с(f)}{G_з(f)} + 1\right) df.$$

Можна показати, що при заданому спектрі шуму $G_з(t)$ та обмеженій потужності сигналу максимум C має місце при виконанні умови

$$G_с(f) + G_з(f) = \text{const}, \tag{56}$$

тобто потужність сигналу має збільшуватись на тих частотах, де зменшується потужність шуму, і навпаки.

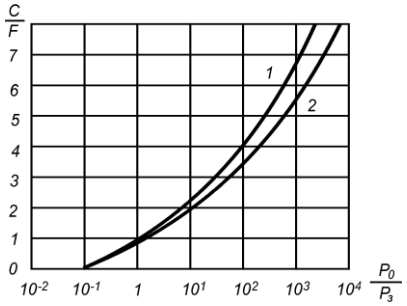

Рис. 6. Залежність пропускної здатності від середнього відношення сигналу до завади для каналів з постійними параметрами (*1*) і релеївським завмиранням (*2*)

У разі виконання умови (56) пропускна здатність зменшується найбільше, якщо спектр шуму рівномірний, тобто є спектром білого шуму. Таким чином, білий шум, що найбільш зменшує пропускну здатність, є найнебезпечніший вид завад.

Продуктивність джерела неперервних повідомлень. За відсутності будь-яких обмежень, накладених на неперервні повідомлення, кількість інформації, яка в них міститься, нескінченна:

$$I(u) = -\log P(u) = \lim_{\Delta u \to 0} \left[-\log \Delta u P(u)\right] = \infty. \tag{57}$$

Тому джерело таких повідомлень має нескінченну продуктивність. Для того щоб кількість інформації та продуктивність джерела набули певного змісту і стали скінченними величинами, слід розглядати неперервне повідомлення $u(t)$ з урахуванням точності його оцінки. Остання, зокрема, визначається похибкою приладів, за допомогою яких вимірюється чи реєструється неперервне повідомлення. Як правило, похибка кількісно оцінюється середньоквадратичним відхиленням наближеного неперервного повідомлення $u^*(t)$ від його точного значення $u(t)$:

$$\overline{\varepsilon_0^2} = \overline{\left[u^*(t) - u(t)\right]}^2 = \int\limits_u \int\limits_{u^*} p(u,u^*)(u^* - u)^2 \, du \, du^*. \qquad (58)$$

Чим менше $\overline{\varepsilon_0^2}$, тим менша кількість інформації в середньому міститься в $u^*(t)$ відносно $u(t)$ і тим вища продуктивність джерела.

Кількість інформації на виході джерела при $\overline{\varepsilon_0^2} > 0$ визначається так:

$$I(u,u^*) = \int\limits_u \int\limits_{u^*} p(u,u^*) \log \frac{p(u/u^*)}{p(u)} \, du \, du^*. \qquad (59)$$

Для обмеженої похибки $\overline{\varepsilon_0^2} \le \varepsilon$ завжди можна знайти такий спосіб відтворення $u(t)$ через $u^*(t)$, а отже, такий розподіл $p'(u,u^*)$, при якому вираз (59) набуває найменшого значення. Розподіл $p'(u,u^*)$ є найбільш вигідним, оскільки він дозволяє при заданій похибці відтворювати $u(t)$, використовуючи мінімальну кількість інформації. Найменше значення $I(u,u^*)$ при $\overline{\varepsilon_0^2} \le \varepsilon$ називається ***епсилон-ентропією***:

$$H_\varepsilon(u) = \min_{\overline{\varepsilon_0^2} \le \varepsilon} I(u,u^*). \qquad (60)$$

Тоді продуктивність джерела неперервних повідомлень

$$R_д = \lim_{T \to \infty} \frac{H_\varepsilon(u)}{T}.$$

Для неперервного каналу з пропускною здатністю C, на вхід якого підключається джерело продуктивністю $R_д$, Шенноном була доведена наступна теорема.

Якщо при заданій похибці оцінки повідомлень джерела $\overline{\varepsilon_0^2}$ його продуктивність $R_д < C$, то існує спосіб кодування, який дозволяє передавати всі неперервні повідомлення джерела з похибкою у відтворенні на виході каналу, що як завгодно мало відрізняється від $\overline{\varepsilon_0^2}$.

Інакше кажучи, вибравши певний спосіб кодування, можна зробити так, щоб додаткова неточність у відтворенні повідомлення $\mathrm{v}(t)$ на виході каналу, зумовлена дією завад, була дуже незначною, тобто $\left[\overline{\mathrm{v}(t) - u(t)}\right]^2 = = \overline{\varepsilon_0^2} + \overline{\delta^2}$, де $\overline{\delta^2}$ – як завгодно мала величина.

Швидкість передавання інформації по каналу, нарешті, визначається швидкістю потоку інформації на виході приймача:

$$R = \lim_{T \to \infty} \frac{1}{T} \left[H_T(\mathrm{v}) - H_T(\omega^*) \right],$$

де $H_T(\mathrm{v})$ – ентропія прийнятого повідомлення $\mathrm{v}(t)$; $H_T(\omega^*)$ – ентропія шуму на виході приймача. Якщо вважати, що повідомлення $\mathrm{v}(t)$ і завада $\omega^*(t)$ на виході приймача мають нормальний розподіл і рівномірний спектр, то

$$R = F_m \log\left(\frac{P_\text{с}^*}{P_\text{з}^*} + 1\right), \qquad (61)$$

де F_m – ширина спектра частот повідомлення, яке передається (звичайно дорівнює смузі пропускання приймача на низькій частоті); $P_\text{с}^*$ – середня потужність прийнятого повідомлення $\mathrm{v}(t)$; $P_\text{з}^*$ – середня потужність завади (шуму) на виході приймача.

Пропускна здатність каналів зі змінними параметрами. Характеристики системи зв'язку значною мірою залежать від параметрів каналу зв'язку, який використовується для передавання повідомлень. Досліджуючи пропускну здатність каналів, ми вважали їх параметри постійними. Однак більшість реальних каналів мають параметри, що, як правило, випадково змінюються в часі. Випадкові зміни коефіцієнта передачі каналу μ спричинюють завмирання сигналу, що еквівалентно дії мультиплікативної завади з еквівалентною потужністю $P_\text{с} D[\mu]$. Тоді сумарна потужність завад в каналі зі змінними параметрами при наявності адитивної завади

$$P_{\text{з}\Sigma} = P_\text{з} + P_\text{с} D[\mu]. \qquad (62)$$

Звідси випливає, що завмирання сигналу призводять до збільшення потужності завад, а отже, до зниження пропускної здатності каналу (1.55).

Розглянемо, як обчислюється пропускна здатність каналу із завмираннями при передаванні неперервних повідомлень. Для цього необхідно знайти такий розподіл імовірностей сигналу, який при заданих статистичних властивостях коефіцієнта передачі μ забезпечував би максимальну швидкість передавання інформації.

Розв'язання задачі в загальному вигляді викликає значні труднощі. Проте у випадку повільних завмирань, коли швидкість їх набагато менша від швидкості зміни неперервного повідомлення, можна з достатньою точністю передбачити за поточними значеннями коефіцієнта μ його наступні значення. При такому допущенні максимум швидкості передавання, як і раніше, має місце для повідомлень з нормальним розподілом, що дає можливість користуватися формулою (55). Підставляючи в (55) потужність $\mu^2 P_0$ сигналу, який приймається, отримаємо пропускну здатність при фіксованому μ:

$$C(\mu) = F \log\left(\frac{\mu^2 P_0}{P_з} + 1\right). \tag{63}$$

$$C = \int_0^\infty p(\mu) C(\mu) d\mu = \int_0^\infty p(\mu) F \log\left(\frac{\mu^2 P_0}{P_з} + 1\right) d\mu, \tag{64}$$

Контрольні питання до Теми 4:

1. Що таке епсілон-ентропія?
2. Як визначається кількість інформації в неперервних повідомленнях?
3. Як неперервні повідомлення відрізняються від дискретних з точки зору ентропії?
4. Які методи використовуються для кодування неперервних повідомлень?
5. Назвіть основні характеристики неперервних каналів без завад.

ТЕМА 5 ВИЗНАЧЕННЯ ЩІЛЬНОСТІ РОЗПОДІЛУ ЙМОВІРНОСТЕЙ І ВТРАТ ІНФОРМАЦІЇ НА ВИХОДІ НЕЛІНІЙНОГО ЕЛЕМЕНТА. ЕФЕКТИВНІСТЬ СИСТЕМ ПЕРЕДАВАННЯ ІНФОРМАЦІЇ

Визначення щільності розподілу ймовірностей. Багато елементів каналів зв'язку є нелінійними, тобто мають нелінійні статистичні характеристики різних типів (зона нечутливості, насичення тощо). Наявність нелінійностей у каналі зв'язку призводить до ускладнень синтезу та до втрат інформації.

Визначимо щільність розподілу ймовірностей випадкового сигналу $\eta(t)$ на виході нелінійного елемента з заданою статичною характеристикою. При цьому вважатимемо, що випадковий сигнал $\xi(t)$ на вході нелінійності $F[\xi(t)]$ має щільність $\omega_\xi(x)$ та функцію розподілу $W_\xi(x)$.

1. *Статична характеристика нелінійності типу зони нечутливості* (рис. 1.7, *а*):

$$F[\xi(t)] = \begin{cases} 0, & |\xi(t)| < b_н; \\ k\xi(t) - kb_н \operatorname{sgn}\xi(t); & |\xi(t)| > b_н, \end{cases}$$

де $b_н$ – половина ширини зони нечутливості.

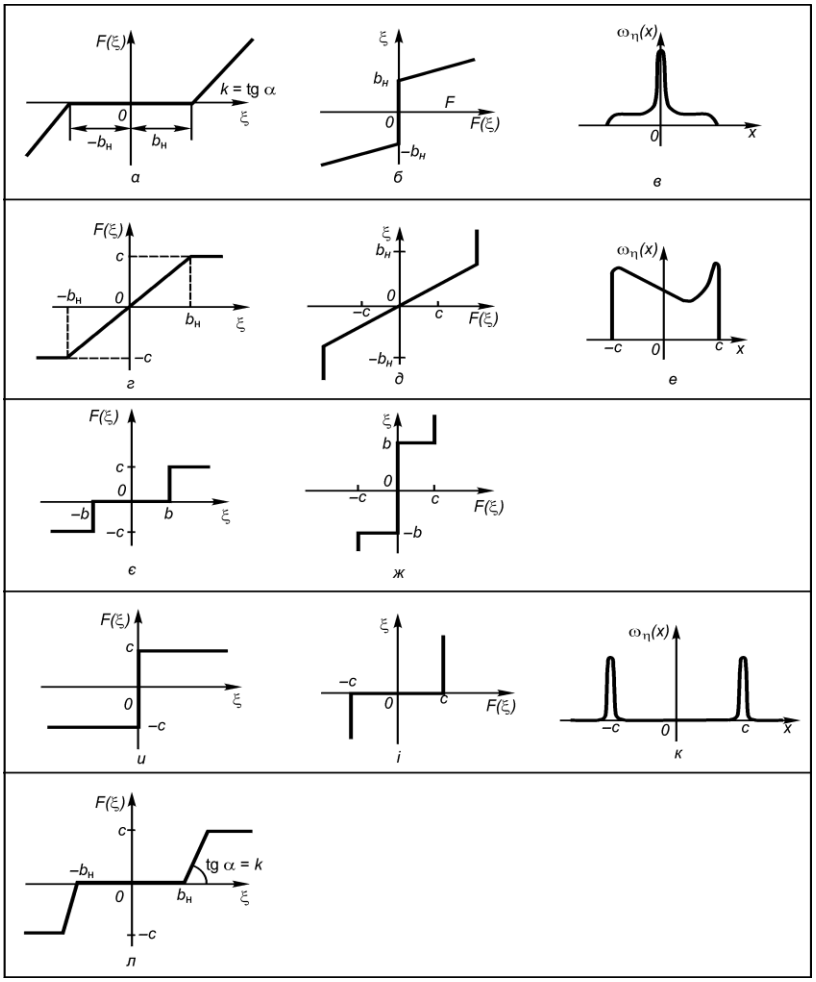

Рис. 7. Нелінійні статичні характеристики; характеристики, обернені до них; щільності вихідного сигналу

Обернена статична характеристика зони нечутливості наведена на рис. 7, *б*. Випадковий процес $\eta(t)$ на виході нелінійності розподілений у цьому випадку з щільністю (рис. 7, *в*)

$$\omega_\eta(x) = \begin{cases} \dfrac{1}{k}\omega_\xi\left(\dfrac{x - kh_n}{k}\right), -\infty < x < 0; \\ S_0\delta(x), \qquad\qquad x = 0; \\ \dfrac{1}{k}\omega_\xi\left(\dfrac{x + kh_n}{k}\right), 0 < x < \infty, \end{cases} \tag{65}$$

де стала S_0 визначається через задану щільність на вході за формулою

$$S_0 = \int_{-b_н}^{b_н} \omega_\xi(x)dx. \tag{66}$$

2. *Статична характеристика нелінійності типу зони насичення* (рис. 7, *г*):

$$F\left[\xi(t)\right] = \begin{cases} k\xi(t), & \left|\xi(t)\right| \le b_н; \\ c\,\mathrm{sgn}\,\xi(t), & \left|\xi(t)\right| > b_н, \ b_н = \dfrac{c}{k}. \end{cases}$$

Для визначення щільності вихідного сигналу на рис. 7, *д* зображена характеристика, обернена до зони насичення.

Випадковий процес $\eta(t)$ на виході нелінійності має щільність (рис. 7, *е*)

$$\omega_\eta(x)=\begin{cases}0, & |x|>c;\\ S_{-c}\delta(x+c), & x=-c;\\ \dfrac{1}{k}\omega_\xi\left(\dfrac{x}{k}\right), & -c<x<c;\\ S_c\delta(x-c), & x=c,\end{cases} \tag{67}$$

де $\delta(x)$ – дельта-функція; сталі S_{-c} і S_c визначаються через задану щільність на вході за формулами

$$S_{-c}=\int\limits_{-\infty}^{-c/k}\omega_\xi(x)dx=W_\xi(-c/k),\ c/k=b;$$
$$S_c=\int\limits_{c/k}^{\infty}\omega_\xi(x)dx=1-W_\xi(c/k),\ \text{оскільки}\int\limits_{-\infty}^{c/k}+\int\limits_{c/k}^{\infty}\equiv 1. \tag{68}$$

3. *Статична характеристика нелінійностей, що поєднують зони нечутливості та насичення, за відсутності лінійної ділянки* (рис. 7, *є*)*:*

$$F\left[\xi(t)\right]=\begin{cases}0, & |\xi(t)|\le b;\\ c\,\mathrm{sgn}\,\xi(t), & |\xi(t)|>b.\end{cases}$$

Статична характеристика, обернена до характеристики нелінійності, зображеної на рис. 7, *є*, наведена на рис. 7, *ж*.

Випадковий процес $\eta(t)$ на виході нелінійності розподілений зі щільністю

$$\omega_\eta(x)=\begin{cases}0, & x\ne 0, x\ne\pm c;\\ S_{-c}\delta(x+c), & x=-c;\\ S_0\delta(x), & x=0;\\ S_c\delta(x-c), & x=c,\end{cases} \tag{69}$$

де $S_{-c}=\text{const},\ S_0=\text{const},\ S_c=\text{const},$ причому $S_{-c}=\int\limits_{-\infty}^{-b}\omega_\xi(x)dx,\ S_0=\int\limits_{b}^{-b}\omega_\xi(x)dx,\ S_c=\int\limits_{b}^{\infty}\omega_\xi(x)dx,$ $S_{-c}+S_0+S_c=1.$

4. *Статична релейна характеристика*. Нелінійність та обернена до неї статична характеристика наведені на рис. 7, *и* та рис. 7, *і* відповідно.

Випадковий процес $\eta(t)$ на виході нелінійності має щільність (рис. 7, *к*)

$$\omega_\eta(x)=\begin{cases}S_{-c}\delta(x+c), & x=-c;\\ S_c\delta(x-c), & x=c;\\ 0, & |x|\ne c,\end{cases} \tag{70}$$

де $S_{-c}=\int\limits_{-\infty}^{0}\omega_\xi(x)dx;\quad S_c=\int\limits_{0}^{\infty}\omega_\xi(x)dx;\quad S_{-c}+S_c=1.$

5. *Статична характеристика нелінійності загального вигляду, яка поєднує зони нечутливості і насичення* (рис. 7, *л*):

$$F\left[\xi(t)\right]=\begin{cases}0, & |\xi(t)|\le b_i;\\ k\xi(t)-kb_{\text{н}}\,\mathrm{sgn}\,\xi(t), & b_{\text{н}}<|\xi(t)|\le b_i+c/k;\\ c\,\mathrm{sgn}\,\xi(t), & |\xi(t)|>b_i+c/k.\end{cases} \tag{71}$$

Як бачимо, попередні чотири розглянуті нелінійності є окремими випадками цієї нелінійності: 1) $b_i=0$; 2) $c=\infty$; 3) $k=\infty$; 4) $b_i=0$, $k=\infty$. Тоді в загальному випадку $\eta(t)$ на вході нелінійності має щільність

$$\omega_\eta(x)=\begin{cases}0, & |x|>c;\\ S_{-c}\delta(x+c), & x=-c;\\ \dfrac{1}{k}\omega_\xi\left(\dfrac{x-kb_н}{k}\right), & -c<x<0;\\ S_0\delta(x), & x=0;\\ \dfrac{1}{k}\omega_\xi\left(\dfrac{x+kb_н}{k}\right), & 0<x<c;\\ S_c\delta(x-c), & x=c,\end{cases} \tag{72}$$

де S_i – сталі, які визначаються так:

$$S_{-c}=\int\limits_{-\infty}^{-b_i-c/k}\omega_\xi(x)dx;\quad S_0=\int\limits_{-b_i}^{b_i}\omega_\xi(x)dx;\quad S_c=\int\limits_{b_i+c/k}^{\infty}\omega_\xi(x)dx.$$

Визначення втрат інформації. Визначимо ентропію випадкового сигналу $\eta(t)$ та втрати інформації на виході нелінійності з заданою статичною характеристикою.

1. Щільність розподілу ймовірностей випадкового сигналу нелінійності *типу зони нечутливості* визначається виразами (65) і (66). Тому ентропія сигналу $\eta(t)$ на виході нелінійності

$$h(\eta)=-\int\limits_{-\infty}^{\infty}\omega_\eta(x)\log_2\omega_\eta(x)=-\int\limits_{-\infty}^{0}\frac{1}{k}\omega_\xi\left(\frac{x-kb_i}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x-kb_i}{k}\right)dx-$$
$$-\int\limits_{-\infty}^{0}S_0\delta(x)\log_2 S_0\delta(x)dx-\int\limits_{0}^{\infty}\frac{1}{k}\omega_\xi\left(\frac{x-kb_i}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x-kb_i}{k}\right)dx. \tag{73}$$

У виразі (1.73) $S_0=\int\limits_{-b_н}^{b_н}\omega_\xi(x)dx=W_\xi(b_н)-W_\xi(-b_н);$ введемо такі позначення:

$$I_1=\int\limits_{-\infty}^{0}\frac{1}{k}\omega_\xi\left(\frac{x-kb_i}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x-kb_i}{k}\right)dx=\left|\begin{array}{l}\dfrac{x}{k}-b_i=u\\ dx=k\,du\\ x=0,\ u=-b_i\end{array}\right|=$$
$$=\int\limits_{-\infty}^{-b_i}\omega_\xi(x)\log_2\frac{1}{k}\omega_\xi(x)dx; \tag{74}$$

$$I_2=\int\limits_{0}^{\infty}S_0\delta(x)\log_2 S_0\delta(x)dx=$$
$$=S_0\int\limits_{0}^{\infty}\delta(x)\log_2 S_0dx+S_0\int\limits_{0}^{\infty}\delta(x)\log_2\delta(x)dx=$$
$$=S_0\log_2 S_0+0=S_0\log_2 S_0; \tag{75}$$

$$I_3=\int\limits_{0}^{\infty}\frac{1}{k}\omega_\xi\left(\frac{x+kb_i}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x+kb_i}{k}\right)dx=\left|\begin{array}{l}\dfrac{x}{k}+b_i=u\\ dx=kdu\\ x=0,\ u=+b_i\end{array}\right|=$$
$$=\int\limits_{b_i}^{\infty}\omega_\xi(x)\log_2\frac{1}{k}\omega_\xi(x)dx. \tag{76}$$

Тоді ентропія з урахуванням (74) – (76) набирає вигляду

$$h(\eta) = -(I_1 + I_2 + I_3) =$$

$$= -\int_{-\infty}^{-b_i} \omega_\xi(x) \log_2 \frac{1}{k} \omega_\xi(x)\, dx - S_0 \log_2 S_0 -$$

$$-\int_{b_i}^{\infty} \omega_\xi(x) \log_2 \frac{1}{k} \omega_\xi(x)\, dx =$$

$$= -\int_{-\infty}^{-b_i} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx - \int_{b_i}^{\infty} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx -$$

$$-\log_2 \frac{1}{k} \left[\int_{-\infty}^{-b_i} \omega_\xi(x)\, dx + \int_{b_i}^{\infty} \omega_\xi(x)\, dx \right] - S_0 \log_2 S_0,$$

звідки ентропія сигналу $\eta(t)$, перерахована до входу нелінійності,

$$h(\eta) = -\int_{-\infty}^{-b_i} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx - \int_{b_i}^{\infty} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx -$$

$$-\left[W_\xi(b_i) - W_\xi(-b_i) \right] \log_2 \left[W_\xi(b_i) - W_\xi(-b_i) \right]. \qquad (77)$$

В той же час ентропія сигналу $\xi(t)$ на вході нелінійності без урахування зони нечутливості

$$h(\xi) = -\int_{-\infty}^{\infty} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx. \qquad (78)$$

З урахуванням (77), 78) *втрати інформації в системі синхронізації, зумовлені наявністю зони нечутливості*:

$$\Delta I = h(\xi) - h(\eta) = -\int_{-\infty}^{\infty} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx +$$

$$+\int_{-\infty}^{-b_i} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx + \int_{b_i}^{\infty} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx +$$

$$+\left[W_\xi(b_i) - W_\xi(-b_i) \right] \log_2 \left[W_\xi(b_i) - W_\xi(-b_i) \right] =$$

$$= \int_{-b_i}^{b_i} \omega_\xi(x) \log_2 \omega_\xi(x)\, dx +$$

$$+\left[W_\xi(b_i) - W_\xi(-b_i) \right] \log_2 \left[W_\xi(b_i) - W_\xi(-b_i) \right]. \qquad (79)$$

2. Щільність розподілу ймовірностей випадкового сигналу *нелінійності типу насичення* визначається виразами (.67) і (68). Тоді ентропія сигналу $\eta(t)$

$$h(\eta) = -\int_{-\infty}^{\infty} \omega_\eta(x) \log_2 \omega_\eta(x) dx = -\int_{-c} S_{-c} \delta(x+c) \log_2 S_{-c} \delta(x+c) dx -$$

$$-\int_{-c} \frac{1}{k} \omega_\xi\left(\frac{x}{k}\right) \log_2 \frac{1}{k} \omega_\xi\left(\frac{x}{k}\right) dx - \int_{-c} S_c \delta(x-c) \log_2 S_c \delta(x-c) dx. \qquad (80)$$

Оскільки перший інтеграл у (1.80)

$$I_1 = \int_{-c} S_{-c} \delta(x+c) \log_2 S_{-c} \delta(x+c) dx =$$

$$= S_{-c} \int_{-c} \delta(x+c) \left[\log_2 S_{-c} + \log_2 \delta(x+c) \right] dx =$$

$$= S_{-c} \int_{-c} \delta(x+c) \log_2 S_{-c} dx = S_{-c} \log_2 S_{-c},$$

то за аналогією третій інтеграл

$$I_3 = \int_{-c} S_c \delta(x-c) \log_2 S_c \delta(x-c) dx = S_c \log_2 S_c .$$

Підставивши в другий інтеграл виразу (80) $\frac{x}{k} = u,\ dx = k\,du$, матимемо

$$I_2 = \log_2 \frac{1}{k} \int_{-c/k}^{c/k} \omega_\xi(u)du + \int_{-c/k}^{c/k} \omega_\xi(u)\log_2\omega_\xi(u)du.$$

Тоді

$$h(\eta) = (-I_1 + I_2 + I_3) =$$

$$= S_{-c}\log_2 S_{-c} - \log_2 \frac{1}{k} \int_{-c/k}^{c/k} \omega_\xi(x)dx + \int_{-c/k}^{c/k} \omega_\xi(x)\log_2\omega_\xi(x)dx - S_c \log_2 S_c$$

і ентропія сигналу $\eta(t)$, перерахована до входу нелінійності,

$$h(\eta) = -S_{-c}\log_2 S_{-c} - \int_{-c/k}^{c/k} \omega_\xi(x)\log_2\omega_\xi(x)dx - S_c \log_2 S_c, \qquad (81)$$

де $S_{-c} = W_\xi\left(-\frac{c}{k}\right);\ S_c = 1 - W_\xi\left(-\frac{c}{k}\right);\ W_\xi(x)$ – інтегральна функція розподілу сигналу $\xi(t)$.

Перепишемо вираз (81) у такому вигляді:

$$h(\eta) = -W_\xi\left(-\frac{c}{k}\right)\log_2 W_\xi\left(-\frac{c}{k}\right) - \int_{-c/k}^{c/k} \omega_\xi(x)\log_2 \omega_\xi(x)dx -$$

$$-\left[1 - W_\xi\left(\frac{c}{k}\right)\right]\log_2\left[1 - W_\xi\left(\frac{c}{k}\right)\right].$$

Втрати інформації в системі синхронізації, зумовлені наявністю нелінійності типу зони насичення,

$$\Delta I = -\int_{-\infty}^{\infty} \omega_\xi(x)\log_2 \omega_\xi(x)\,dx + W_\xi\left(-\frac{c}{k}\right)\log_2 W_\xi\left(-\frac{c}{k}\right) +$$

$$+ \int_{-c/k}^{c/k} \omega_\xi(x)\log_2\omega_\xi(x)dx + \left[1 - W_\xi\left(\frac{c}{k}\right)\right]\log_2\left[1 - W_\xi\left(\frac{c}{k}\right)\right]$$

або з урахуванням того, що $\int_{-\infty}^{\infty} \omega_\xi(x)\log_2 \omega_\xi(x)dx = \int_{-\infty}^{-c/k} + \int_{-c/k}^{c/k} + \int_{c/k}^{\infty}$, матимемо

$$\Delta I = -\int_{-\infty}^{-c/k} \omega_\xi(x)\log_2 \omega_\xi(x)\,dx + W_\xi\left(-\frac{c}{k}\right)\log_2 W_\xi\left(-\frac{c}{k}\right) +$$

$$+ \int_{c/k}^{\infty} \omega_\xi(x)\log_2\omega_\xi(x)dx + \left[1 - W_\xi\left(\frac{c}{k}\right)\right]\log_2\left[1 - W_\xi\left(\frac{c}{k}\right)\right]. \qquad (82)$$

3. Щільність розподілу ймовірностей з статичною *релейною характеристикою* на виході нелінійності визначається виразом (70). З урахуванням (70) ентропія сигналу

$$h(\eta) = -\int_{-\infty}^{\infty} \omega_\eta(x)\log_2 \omega_\eta(x)\,dx = \int_{-c} S_{-c}\delta(x+c)\log_2 S_{-c}\delta(x+c)dx -$$

$$-\int_{c}^{-c} S_c\delta(x-c)\log_2 S_c\delta(x-c)dx = -S_{-c}\log_2 S_{-c} - S_c\log_2 S_c,$$

де $S_{-c} = \int_{-\infty}^{0} \omega_\xi(x)\,dx = W_\xi(0),\ S_c = \int_{0}^{\infty} \omega_\xi(x)\,dx = 1 - \int_{-\infty}^{0} \omega_\xi(x)\,dx = 1 - W_\xi(0)$. Тоді

$$h(\eta) = -S_{-c}\log_2 S_{-c} - S_c\log_2 S_c =$$

$$= -W_\xi(0)\log_2 W_\xi(0) - \left[1 - W_\xi(0)\right]\log_2\left[1 - W_\xi(0)\right].$$

Втрати інформації, зумовлені наявністю релейної характеристики нелінійності,

$$\Delta I = h(\xi) - h(\eta) = -\int_{-\infty}^{\infty} \omega_\xi(x)\log_2 \omega_\xi(x)\,dx +$$

$$+W_\xi(0)\log_2 W_\xi(0)+\left[1-W_\xi(0)\right]\log_2\left[1-W_\xi(0)\right]. \tag{83}$$

4. Щільність розподілу ймовірностей випадкового сигналу $\eta(t)$ на виході нелінійності *загального вигляду, що поєднує зону нечутливості і насичення*, визначається виразом (72).

З урахуванням (1.72) ентропія сигналу

$$h(\eta)=-\int_{-\infty}^{\infty}\omega_\eta(x)\log_2\omega_\eta(x)dx=\int_{-c}^{}S_{-c}\delta(x+c)\log_2 S_{-c}\delta(x+c)dx-$$

$$-\int_{-c}^{0}\frac{1}{k}\omega_\xi\left(\frac{x-kb_{\text{н}}}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x-kb_{\text{н}}}{k}\right)dx-\int_{0}^{}S_0\delta(x)\log_2 S_0\delta(x)dx-$$

$$-\int_{0}^{c}\frac{1}{k}\omega_\xi\left(\frac{x+kb_{\text{н}}}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x+kb_{\text{н}}}{k}\right)dx-\int_{c}^{}S_c\delta(x-c)\log_2 S_c\delta(x-c)dx,$$

(84)

де $S_{-c}=\int_{-\infty}^{-b_{\text{н}}-c/k}\omega_\xi(x)dx$; $S_0=\int_{-b_{\text{н}}}^{b_{\text{н}}}\omega_\xi(x)dx$; $S_c=\int_{b_{\text{н}}+c/k}^{\infty}\omega_\xi(x)dx$. Тоді

$$I_1=\int_{-c}^{}S_{-c}\delta(x+c)\log_2 S_{-c}\delta(x+c)dx=S_{-c}\log_2 S_{-c};$$

$$I_2=\int_{-c}^{0}\frac{1}{k}\omega_\xi\left(\frac{x-kb_{\text{н}}}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x-kb_{\text{н}}}{k}\right)dx=$$

$$=\left|\frac{x}{k}-b_{\text{н}}=u;\ dx=k\,du\right|=\int_{-b_{\text{н}}-c/k}^{-b_{\text{н}}}\omega_\xi(u)\log_2\frac{1}{k}\omega_\xi(u)du;$$

$$I_3=\int_{0}^{}S_0\delta(x)\log_2 S_0\delta(x)dx=S_0\log_2 S_0; \tag{85}$$

$$I_4=\int_{0}^{c}\frac{1}{k}\omega_\xi\left(\frac{x+kb_{\text{н}}}{k}\right)\log_2\frac{1}{k}\omega_\xi\left(\frac{x-kb_{\text{н}}}{k}\right)dx=$$

$$=\left|\frac{x}{k}+b_{\text{н}}=u,dx=k\,du\right|=\int_{b_{\text{н}}}^{b_{\text{н}}+c/k}\omega_\xi(u)\log_2\frac{1}{k}\omega_\xi(u)du;$$

$$I_5=\int_{-c}^{}S_c\delta(x-c)\log_2 S_c\delta(x-c)dx=S_c\log_2 S_c.$$

З урахуванням (85)

$$h(\eta)=-(I_1+I_2+I_3+I_4+I_5)=-S_{-c}\log_2 S_{-c}-\int_{-b_{\text{н}}-c/k}^{-b_{\text{н}}}\omega_\xi(u)\log_2\frac{1}{k}\omega_\xi(u)du-$$

$$-S_0\log_2 S_0-\int_{b_{\text{н}}}^{b_{\text{н}}+c/k}\omega_\xi(u)\log_2\frac{1}{k}\omega_\xi(u)du-S_c\log_2 S_c. \tag{86}$$

Перетворивши інтеграли у виразі (86), дістанемо

$$h(\eta)=-S_{-c}\log_2 S_{-c}-\int_{-b_{\text{н}}-c/k}^{-b_{\text{н}}}\omega_\xi(x)\log_2\omega_\xi(x)dx-S_0\log_2 S_0-$$

$$-\int_{b_{\text{н}}}^{b_{\text{н}}+c/k}\omega_\xi(x)\log_2\omega_\xi(x)dx-S_c\log_2 S_c, \tag{87}$$

звідки

$$S_{-c}=W_\xi\left(-b_{\text{н}}-\frac{c}{k}\right),\ S_0=W_\xi(b_{\text{н}})-W_\xi(-b_{\text{н}}),\ S_c=1-W_\xi\left(b_{\text{н}}+\frac{c}{k}\right).$$

Перерахована ентропія вихідного сигналу набирає вигляду

$$h(\eta) = -W_\xi\left(-b_н - \frac{c}{k}\right)\log_2 W_\xi\left(-b_н - \frac{c}{k}\right) - \int\limits_{-b_н - c/k}^{-b_н} \omega_\xi(x)\log_2 \omega_\xi(x)dx -$$

$$-\left[W_\xi(b_н) - W_\xi(-b_н)\right]\log_2\left[W_\xi(b_н) - W_\xi(-b_н)\right] - \int\limits_{b_н}^{b_н + c/k} \omega_\xi(x)\log_2 \omega_\xi(x)dx -$$

$$-\left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right]\log_2\left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right]. \qquad (88)$$

Визначимо тепер *втрати інформації в пристрої зв'язку, зумовлені наявністю нелінійності з характеристикою загального вигляду* (рис. 7, *л*).

$$\Delta I = h(\xi) - h(\eta) = -\int\limits_{-\infty}^{\infty} \omega_\xi(x)\log_2(x)dx + W_\xi\left(-b_н - \frac{c}{k}\right)\log_2 W_\xi\left(-b_н - \frac{c}{k}\right) +$$

$$+ \int\limits_{-b_н - c/k}^{-b_н} \omega_\xi(x)\log_2 \omega_\xi(x)dx + \left[W_\xi(b_н) - W_\xi(-b_н)\right]\log_2 + \left[W_\xi(b_н) - W_\xi(-b_н)\right] +$$

$$+ \int\limits_{b_н}^{b_н + c/k} \omega_\xi(x)\log_2 \omega_\xi(x)dx + \left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right]\log_2\left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right]$$

Або

$$\Delta I = -\int\limits_{-\infty}^{c/k - b_н} - \int\limits_{-c/k - b_н}^{-b_н} - \int\limits_{-b_н}^{b_н} - \int\limits_{b_н}^{c/k + b_н} - \int\limits_{c/k + b_н}^{\infty} + W_\xi\left(-b_н - \frac{c}{k}\right)\log_2 W_\xi\left(-b_н - \frac{c}{k}\right) +$$

$$+ \int\limits_{-b_н - c/k}^{-b_н} \left[W_\xi(b_н) - W_\xi(-b_н)\right]\log_2\left[W_\xi(b_н) - W_\xi(-b_н)\right] +$$

$$+ \int\limits_{b_н}^{b_н + c/k} \left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right]\log_2\left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right] =$$

$$= \int\limits_{-\infty}^{-c/k - b_н} \omega_\xi(x)\log_2\omega_\xi(x)dx - \int\limits_{-b_н}^{b_н} \omega_\xi(x)\log_2\omega_\xi(x)dx -$$

$$- \int\limits_{b_н + c/k}^{\infty} \omega_\xi(x)\log_2\omega_\xi(x)dx + W_\xi\left(-b_н - \frac{c}{k}\right)\log_2 W_\xi\left(-b_н - \frac{c}{k}\right) +$$

$$+ \left[W_\xi(b_н) - W_\xi(-b_н)\right]\log_2\left[W_\xi(b_н) - W_\xi(-b_н)\right] +$$

$$+ \left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right]\log_2\left[1 - W_\xi\left(b_н + \frac{c}{k}\right)\right]. \qquad (89)$$

Як окремі випадки, з виразу (89) можна отримати співвідношення (79) і (82), що визначають втрати інформації, зумовлені наявністю насичення і зони нечутливості відповідно.

Пропускна здатність каналу зв'язку C визначає максимальну швидкість передавання інформації, тобто є тією границею, якої можна досягти при ідеальному кодуванні. Природно, що в реальних каналах швидкість передавання R завжди буде меншою за C. Ступінь залежності R від C визначається тим, наскільки раціонально вибрана і ефективно використовується система зв'язку. Найбільш загальною оцінкою ефективності системи зв'язку є *коефіцієнт використання каналу*:

$$\eta = \frac{R}{C}. \qquad (90)$$

Для *дискретних* систем зв'язку $\eta = \eta_1\eta_2$, де η_1 і η_2 – ефективність системи кодування і ефективність системи модуляції. Вводячи надлишковість повідомлення $x_1 = 1 - \eta_1$ та надлишковість сигналу $x_2 = 1 - \eta_2$, отримаємо

$$\eta = 1 - x, \qquad (91)$$

де $x = x_1 + x_2 - x_1x_2$ – повна надлишковість системи.

При передаванні *неперервних* повідомлень

$$\eta = \frac{F_m \log\left(\frac{P_c^*}{P_0^*}+1\right)}{F \log\left(\frac{P_c}{P_0}+1\right)}, \tag{92}$$

де F_m – ширина смуги модульованого сигналу; P_c^*/P_0^* – відношення сигнал–шум для модульованого сигналу; F – ширина смуги первинного сигналу; P_c/P_0 – відношення сигнал–шум для первинного сигналу.

В ряді практичних випадків зручними оцінками ефективності зв'язку є *коефіцієнт використання потужності сигналу*

$$\eta_P = \frac{RN_0}{P_c}, \tag{93}$$

де N_0 – інтенсивність завади, і *коефіцієнт використання смуги частот каналу*

$$\eta_F = \frac{R}{F}. \tag{94}$$

Коефіцієнт η_P є основним показником ефективності для тих систем, в яких потужність сигналу жорстко обмежена (наприклад, космічних). У системах проводового зв'язку більш важливим показником є коефіцієнт η_F. Згідно з (91), ефективність системи зв'язку повністю визначається її надлишковістю. Тому задача підвищення ефективності зв'язку зводиться до зменшення надлишковості повідомлення та сигналу.

Надлишковість повідомлення зумовлена тим, що елементи повідомлення не є рівноймовірними і між ними існує статистичний зв'язок. При кодуванні можна перерозподілити ймовірності вихідного повідомлення так, щоб розподіл імовірностей символів коду наближався до оптимального (рівномірного для передавання дискретних повідомлень або нормального для неперервних). Такий перерозподіл дозволяє усунути надлишковість, що залежить від розподілу ймовірностей елементів повідомлення. Прикладом подібного кодування є код Шеннона–Фано, розглянутий раніше. Якщо перейти від кодування окремих символів повідомлення до кодування цілих груп символів, то можна усунути взаємозв'язок між ними і тим самим зменшити надлишковість. Загальна ідея такого методу кодування, який називають *методом збільшення*, полягає в наступному. Вихідне повідомлення розбивається на відрізки по k символів у кожному. Такі відрізки можуть розглядатися як збільшені елементи повідомлення. Можна показати, що ймовірнісні зв'язки між такими збільшеними елементами слабкіші, ніж між елементами вихідного повідомлення. Очевидно, що чим більше k (більші відрізки), тим слабшим буде зв'язок між ними. Далі збільшені елементи кодуються з урахуванням їх розподілу ймовірностей. В разі збільшення елементів відбувається перетворення, котре полягає в переході до коду з більш високою основою: $m_1 = m^k$, де m – початковий стан.

Своєрідним прикладом методу збільшення повідомлень є стенографічний текст. Кожний стенографічний знак в цьому тексті зображує ціле слово або навіть групу слів.

Що стосується сигналу, то його надлишковість залежить від способу модуляції та виду переносника. Процес модуляції звичайно супроводжується розширенням смуги частот сигналу в порівнянні зі смугою частот повідомлення, яке передається. Це розширення смуги і є надлишковим. Частотна надлишковість також збільшується при переході від синусоїдального переносника до переносника імпульсного чи шумоподібного.

З точки зору підвищення ефективності передачі слід було б вибирати такі способи модуляції, які мають малу надлишковість. До таких систем, частково, належить односмугова передача, при якій сигнали, що передаються, не містять частотної надлишковості – вони є просто копіями повідомлень, які передаються. Однак, говорячи про ефективність системи зв'язку, не можна забувати про її завадостійкість. Усунення надлишковості підвищує ефективність передавання, але знижує при цьому вірогідність (завадостійкість) і, навпаки,

збереження чи введення надлишковості дозволяє забезпечити високу вірогідність передавання. Наприклад, усунення надлишковості при телеграфному передаванні тексту призводить до ускладнення виправляння помилок у повідомленні і, врешті-решт, до зниження завадостійкості. Зі збереженням надлишковості в тексті завадостійкість буде вищою.

Контрольні питання до Теми 5:

1. Як визначаються втрати інформації на виході нелінійного елемента?
2. Які основні характеристики втрат інформації в системах з нелінійними елементами?
3. Які методи використовуються для мінімізації втрат інформації в нелінійних системах?
4. Що таке апостеріорна ймовірність і як вона використовується в аналізі втрат?
5. Які фактори впливають на ефективність передачі інформації через нелінійні елементи?

Тема 6 Основні поняття та означення кодування. Коректувальні та циклічні коди. Коди Боуза–Чоудхурі–Хоквінгема

Перетворення дискретного повідомлення в сигнал складається з двох операцій: кодування і модуляції. *Кодування* визначає закон побудови сигналу, а *модуляція* – вид формованого сигналу, що має передаватися каналом зв'язку.

Найпростішим прикладом дискретного повідомлення є текст. Будь-який текст складається зі скінченного числа елементів: букв, цифр, розділових знаків. Для європейських мов число елементів коливається від 52 до 55, для східних – може обчислюватися сотнями і навіть тисячами. Оскільки число елементів у дискретному повідомленні скінченне, то їх можна пронумерувати і тим самим звести передавання повідомлень до передавання послідовності чисел.

Множина знаків (символів) і система правил побудови складених знаків називається *кодом*. Кінцева послідовність кодових знаків називається *словом* або *кодовою комбінацією*. З цього погляду алфавіт української або іншої мови і сукупність правил побудови слів є кодом, за допомогою якого усне мовлення перетворюється в текст. Слова (кодові комбінації) можуть розглядатися як символи деякої іншої множини; разом із установленими правилами утворення речень вони теж утворюють код, яким ми користуємося для обміну інформацією. Коди поділяються:

- за числом m символів у множині – на двійкові ($m = 2$) і недвійкові ($m \neq 2$). Залежно від числа символів у множині коди називають, наприклад, трійковими, вісімковими. Кількість символів у множині є *основою коду*;
- за довжиною кодових комбінацій – на нерівномірні, у яких слова мають нерівну кількість символів, і рівномірні – з однаковою кількістю символів у словах;
- за принципом використання кодових комбінацій – на коди з використанням усіх дозволених комбінацій заданої довжини n для подання повідомлень і коди з використанням частини можливих комбінацій. Останні у свою чергу поділяються на коди, що дають змогу виявляти або виправляти спотворені символи (помилки), які виникають через завади, – коректувальні коди, і на коди, що не дають такої можливості.

У загальному випадку процес перетворення повідомлення в кодову комбінацію прийнято називати кодуванням. При цьому мається на увазі взаємооднозначне перетворення. В окремому випадку кодування можна визначити як операцію встановлення однозначної відповідності між символами групи символів деякого коду з символами групи символів іншого коду. Таке кодування є переведенням з однієї системи числення в іншу (наприклад, десяткових чисел у двійкові). Слова "код" і "кодування" походять від латинського слова *codex* – кодекс, тобто книга, що містить знаки, систему правил.

Повідомлення, які є текстом природною мовою, мають значний надлишок, тому для його зменшення використовують кодування, що враховує статистичні особливості повідомлення. Для зменшення надлишку доцільно кодувати повідомлення так, щоб середня довжина *L* кодових комбінацій була найменшою, тобто здійснювати *статистичне кодування.* Одним з кодів, що використовують статистичні властивості повідомлень, є *код Шеннона – Фано*.

Зазначимо, що поліпшення якісних характеристик каналів для підвищення надійності приймання завжди пов'язано зі значними матеріальними витратами, а іноді й дуже високими. Тому величезне значення мають широко застосовувані в техніці передавання методи підвищення надійності приймання, які не потребують поліпшення якості каналу. Ці методи засновані на внесенні в переданий сигнал значного надлишку. Надлишок, що вводиться в переданий сигнал, накладає на нього певні додаткові умови, перевірка дотримання яких при прийманні дає змогу встановити факт спотворення сигналу, а також ототожнити прийнятий спотворений сигнал з відповідним неспотвореним.

Будь-які методи внесення надлишку в переданий сигнал пов'язані зі збільшенням об'єму сигналу, тобто зі збільшенням або потужності сигналу, або ширини спектра, або часу передавання. Можливості підвищення надійності приймання збільшенням потужності і ширини спектра при передаванні дискретної інформації зі стандартних каналів зв'язку досить обмежені, тому переважне застосування дістав метод уведення надлишку *збільшенням часу передавання сигналу*. Цей метод можна реалізувати у двох різновидах – використанням для передавання дискретної інформації, зниженої стосовно номінального значення швидкості, і застосуванням коректувальних кодів. *Зниження швидкості передавання інформації* по каналах невисокої якості набуло широкого застосування. Так, у багатьох типах апаратури передавання даних (АПД), що працюють по каналах тональної частоти (ТЧ), передбачаються дві і більше швидкості модуляції (наприклад, 600 і 1200 бод), причому з меншою швидкістю звичайно ведеться передавання на великі відстані, а також робота по каналах зі звуженою смугою пропускання (300–2700 Гц).

Застосування коректувальних кодів є ефективнішим методом підвищення надійності, ніж застосування зниженої швидкості. Зазначимо, що обидва методи не слід протиставляти один одному: зниження швидкості передавання можна вважати окремим випадком застосування коректувального коду.

Розглянемо деякі поняття, пов'язані з коректувальними кодами.

Звичайний (простий) код характеризується тим, що окремі його кодові комбінації відрізняються одна від одної лише одним розрядом. Тому навіть один помилково прийнятий розряд спричинює заміну однієї кодової комбінації іншою і, отже, неправильне приймання повідомлення. *Коректувальні (надлишкові, завадостійкі) коди* будують так, що для передавання інформації використовується лише частина кодових комбінацій (дозволені комбінації), які відрізняються одна від одної більше ніж одним розрядом. Усі інші комбінації для передавання не використовуються і належать до числа недозволених (заборонених). Таким чином, при застосуванні коректувальних кодів помилка в одному розряді спричинює заміну дозволеної кодової комбінації недозволеною, що дає змогу знайти помилку. При досить великій відмінності дозволених комбінацій одна від одної можливе виявлення дворазової, триразової і т. д. помилки, оскільки вони зумовлять утворення недозволених комбінацій, а перехід однієї дозволеної комбінації в іншу відбуватиметься під впливом помилок більшої кратності, що є результатом найбільш інтенсивних завад.

Пояснимо це на прикладі. Використаємо для передавання інформації чотирирозрядні кодові комбінації, що відрізняються одна від одної не менш ніж двома розрядами: 0011, 0110, 1001, 1010, 1100, 1111, 0101, 0000. Нехай при передаванні кожної з цих комбінацій (наприклад, 0011) сталася одинична помилка, внаслідок чого спотворився перший розряд, і прийнято комбінацію 1011. Ця комбінація є недозволеною, що свідчить про наявність помилки. Підберемо далі чотирирозрядні комбінації, що відрізняються всіма чотирма

розрядами: 0011 і 1100. Легко переконатися, що при використанні цих комбінацій виявляються одно-, дво- і триразові помилки, а не виявляється лише чотириразова помилка.

Цей самий код, що складається з двох кодових комбінацій, може використовуватися і для виправлення одиничних помилок. Нехай, наприклад, прийнято комбінацію 1011. Ця комбінація відрізняється від дозволеної комбінації 0011 одним розрядом, а від іншої дозволеної 1100 – трьома. Отже, прийнята комбінація "ближча" до комбінації 0011, ніж до комбінації 1100; це дає підставу вважати, що була передана комбінація 0011.

Підвищена завадостійкість двох розглянутих кодів пов'язана з наявним в них надлишком. Так, перший код містить вісім комбінацій, що становлять чотири розряди кожна. В той самий час у простому коді для утворення восьми комбінацій достатньо трьох розрядів, а не чотирьох. Отже, підвищення завадостійкості потребує введення додаткового розряду. Другий код має ще більшу завадостійкість, і це потребує ще більшого надлишку – трьох додаткових розрядів.

Коректувальні коди так само, як і прості, можуть бути *рівномірними* або *нерівномірними*, *двійковими* або *багатопозиційними*. Використання нерівномірних або багатопозиційних кодів спричинює значне ускладнення апаратури передавання даних, тому застосовуються вони дуже рідко. У зв'язку з цим надалі розглядатимемо лише двійкові рівномірні коректувальні коди. Останні поділяються на два класи – блокові і неперервні (рис.8). При використанні блокових кодів передана інформаційна послідовність розбивається на окремі кодові комбінації (блоки), що кодуються і декодуються незалежно одна від одної. Неперервні коди – це неперервна послідовність розрядів, і поділ її на окремі блоки неможливий. Блокові коди, у свою чергу, поділяються на роздільні та нероздільні.

Рис. 8 Класифікація коректувальних кодів

Роздільними називаються коди, в яких одні розряди є інформаційними, інші – перевірними. Останні і вносять у код надлишок, необхідний для виявлення або виправлення помилок. У роздільних кодах інформаційні і перевірні розряди займають завжди одні й ті самі позиції в кодовій комбінації. Роздільні коди позначаються як (n, k)-коди, де n – довжина або число розрядів коду; k – число інформаційних розрядів.

Нероздільні коди утворюють зараз нечисленну групу. До них, зокрема, належать рекомендований МСЕ стандартний телеграфний код № 3 – семирозрядний код, кожна кодова комбінація якого містить три одиниці і чотири нулі.

Серед роздільних кодів розрізняють коди систематичні і несистематичні. *Систематичними* називаються такі блокові роздільні (n, k)-коди, в яких перевірні розряди – це лінійні комбінації інформаційних. Систематичні коди утворюють велику групу кодів і дуже

широко застосовуються на практиці. Тому головну увагу приділяють саме цим кодам і, зокрема, їх найбільш відомим різновидам – кодам Хеммінга і циклічним.

Кількість розрядів, якими різняться дві кодові комбінації, називається *кодовою відстанню* між двома комбінаціями. Так, кодова відстань між комбінаціями 11011 і 00010 дорівнює трьом, оскільки вони різняться трьома розрядами – першим, другим і п'ятим. Найменша з кодових відстаней у коді називається *мінімальною кодовою* або *хеммінговою відстанню*. Наприклад, у трирозрядному коді з дозволеними комбінаціями 101, 110, 011, 000 мінімальна кодова відстань d_0 дорівнює двом, для простих кодів – одиниці.

Мінімальна кодова відстань d_0 пов'язана з кількістю помилок, що виявляються, і кількістю помилок, що виправляються, так:

$$d_0 \geq \sigma + 1, \tag{95}$$

$$d_0 \geq 2t + 1, \tag{96}$$

де σ – кількість помилок, що виявляються; t – кількість помилок, що виправляються.

Коректувальні коди можна використовувати для виправлення помилок і одночасного виявлення помилок більшої кратності. Можна показати, що при цьому (коли $\sigma > t$)

$$d_0 \geq \sigma + t + 1. \tag{97}$$

Мінімальна кодова відстань лише частково характеризує коректувальні властивості коду, оскільки звичайно забезпечується виправлення і виявлення помилок і більш високої кратності, ніж обумовлене співвідношеннями (95) – (97).

Загальна кількість комбінацій коду завдовжки n дорівнює 2^n. Число дозволених кодових комбінацій визначається числом інформаційних розрядів і дорівнює $M = 2^k = 2^{n-r}$, де r – число перевірних розрядів. Отже, число дозволених кодових комбінацій у 2^r рази менше загального числа комбінацій.

Надлишком коду називають відношення r/n. Питання про мінімально необхідний надлишок коду при заданих мінімальних кодовій відстані і довжині коду в загальному випадку не вирішено. Існує лише ряд верхніх і нижніх оцінок. Найбільше наближення звичайно забезпечує критерій Варламова

$$r \geq \log_2\left(1 + \sum_{i=1}^{d_0-2} C_{n-1}^{i}\right), \tag{98}$$

де C_{n-1}^{i} – число сполучень з $n-1$ елементів по i елементах.

Для деяких кодів отримано точні залежності між числом перевірних і інформаційних розрядів при заданій мінімальній кодовій відстані. Так, при $d_0 = 3$ має місце співвідношення

$$r \geq \log_2(n+1), \tag{99}$$

причому r – найменше ціле число, при якому задовольняється нерівність (99).

Принцип побудови систематичних кодів. Нехай комбінація $a_1a_2...a_kb_1b_2...b_r$, де $a_1a_2...a_k$ – інформаційні, а $b_1b_2...b_r$ – перевірні розряди, є дозволеною комбінацією систематичного (n, k)-коду. В систематичних кодах перевірні розряди є лінійною комбінацією інформаційних. Це означає, що значення будь-якого перевірного розряду $b_i = c_{i1}a_1 \oplus c_{i2}a_2 \oplus ... \oplus c_{ik}a_k$, де c_{i1}, c_{i2}, ..., c_{ik} – числа, які дорівнюють 0 або 1. Нульова (тобто така, що складається з одних нулів) комбінація в будь-якому систематичному коді є дозволеною, оскільки лінійною комбінацією нулів є нуль.

Складемо за модулем 2 дві дозволені кодові комбінації систематичного коду:

$$a_1a_2 \ldots a_kb_1b_2 \ldots b_r \oplus a_1'a_2' \ldots a_k'b_1'b_2' \ldots b_r' =$$
$$= (a_1 \oplus a_1')(a_2 \oplus a_2') \ldots (a_k \oplus a_k')(b_1 \oplus b_1')(b_2 \oplus b_2') \ldots (b_r \oplus b_r'),$$

де

$$b_i \oplus b_i' = c_{i1}a_1 \oplus c_{i2}a_2 \oplus ... \oplus c_{ir}a_r \oplus c_{i1}a_1' \oplus c_{i1}a_2' \oplus ... \oplus c_{ik}a_k' =$$
$$= c_{i1}(a_1 \oplus a_1') \oplus c_{i2}(a_2 \oplus a_2') \oplus ... \oplus c_{ik}(a_k \oplus a_k').$$

Як бачимо, перевірні розряди суми за модулем 2 двох дозволених комбінацій утворюються за тим самим правилом, що і для кожної дозволеної комбінації. Звідси сума двох дозволених комбінацій систематичного коду також є дозволеною комбінацією.

Таке положення є правдивим і при підсумовуванні за модулем 2 будь-яких кількостей дозволених комбінацій систематичного коду, що дає можливість визначити всі дозволені кодові комбінації, користуючись лише їх обмеженою кількістю.

Усі дозволені комбінації систематичного коду можна визначити так. Виберемо кілька (g) ненульових дозволених кодових комбінацій. Складемо їх у деяких сполученнях: по дві, по три, по чотири, ..., по g комбінацій. Кожне додавання дає нам нову дозволену комбінацію.

Процес побудови множини кодових комбінацій можна зобразити як $x_i = c_1 x_1 \oplus c_2 x_2 \oplus ... \oplus c_g x_g$, де $x_1, x_2, ..., x_g$ – початкові кодові комбінації; $c_1, c_2, ..., c_g$ – коефіцієнти, що набувають значення 1 або 0. Надаючи різним коефіцієнтам c_i значення 1 або 0, можна скласти вихідні комбінації в різних сполученнях. Загальна кількість отриманих комбінацій становитиме $c_g^2 + c_g^3 + ... + c_g^g$. Враховуючи g вихідних кодових комбінацій і одну нульову, можна прийти до висновку, що загальна кількість дозволених комбінацій дорівнює

$$c_g^0 + c_g^1 + c_g^2 + ... + c_g^g = 2^k, \qquad (100)$$

оскільки систематичний код містить 2^k дозволених кодових комбінацій. Ця умова задовольняється при $g = k$.

Співвідношення (100) є слушним тоді, коли комбінації, отримані внаслідок додавань, не збігаються ні з початковими комбінаціями, ні з нульовою. Для цього початкові комбінації треба вибирати так:

- усі початкові комбінації мають бути різні; у противному випадку при додаванні виходитиме нульова комбінація;
- нульова комбінація не повинна входити в число початкових, оскільки внаслідок додавання нульової комбінації і якої-небудь початкової виходитиме та сама початкова комбінація;
- усі початкові кодові комбінації мають бути лінійно незалежні, тобто має виконуватися рівність $c_1 x_1 \oplus c_2 x_2 \oplus ... \oplus c_k x_k = 0$ при всіх значеннях c_i, за винятком $c_1 = c_2 = ... = c_k = 0$. У противному разі похідні кодові комбінації збігатимуться з початковими;
- кожна початкова кодова комбінація, як і будь-яка ненульова дозволена комбінація, повинна мати вагу не меншу ніж d_0. (Нагадаємо, що *вагою* кодової комбінації називається кількість одиниць у ній.) Слушність цієї умови легко зрозуміти, якщо врахувати, що нульова комбінація також є дозволеною, а кодова відстань між парою будь-яких кодових комбінацій коду не повинна бути меншою, ніж d_0;
- кодова відстань між будь-якими парами початкових комбінацій не повинна бути меншою, ніж d_0.

Підібрані певним чином k початкових кодових комбінацій однозначно визначають систематичний код. Початкові комбінації прийнято записувати одна під одною у вигляді таблиці (матриці) із k рядків і n стовпців, яка називається *породжувальною матрицею*:

$$G = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1k} & b_{11} & b_{12} & \dots & b_{1r} \\ a_{21} & a_{22} & \dots & a_{21} & b_{21} & b_{22} & \dots & b_{2r} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} & b_{k1} & b_{k2} & \dots & b_{kr} \end{vmatrix}.$$

Звичайно інформаційні розряди займають перші k позицій у кодовій комбінації. При цьому породжувальну матрицю зручно будувати, починаючи з одиничної матриці E_k, що має k стовпців і k рядків:

$$E_k = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{vmatrix};$$

праворуч до неї приписується матриця $C_{r;k}$, що має r стовпців і k рядків:

$$G = \left| E_k,\ C_{r;k} \right| = \begin{vmatrix} 1\,0\,\dots\,0\ b_{11}\ b_{12}\ \dots\ b_{1r} \\ 0\,1\,\dots\,0\ b_{21}\ b_{22}\ \dots\ b_{2r} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ 0\,0\,\dots\,1\ b_{k1}\ b_{k2}\ \dots\ b_{kr} \end{vmatrix}. \qquad (101)$$

Матриця G при належним чином підібраній матриці $C_{r;k}$ також є твірною. Рядки приписаної матриці $C_{r;k}$ знаходять перебором різних r-розрядних комбінацій, що містять не менше ніж $d_0 - 1$ одиниць, при цьому сума за модулем 2 двох будь-яких рядків матриці $C_{r;k}$ не повинна мати менше, ніж $d_0 - 2$ одиниць. Матриця G записана в канонічній формі.

Розглянемо тепер приклад побудови систематичного коду. Нехай треба побудувати код $n = 7$, що забезпечує виправлення одиничної помилки. Відповідно до формули (96) $d_0 = 3$. Користуючись нерівністю (99), знайдемо $r = 3$. Відповідно $k = 4$. Побудуємо твірну матрицю коду (7, 4):

$$G = \left| \begin{array}{cccc:ccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array} \right|.$$

Рядки приписаної частини твірної матриці (обведено штриховою лінією) містять не менше, ніж дві одиниці. Сума за модулем 2 двох будь-яких рядків приписаної матриці не менша, ніж 1.

Оскільки $k = 4$, то код (7, 4) має 16 дозволених комбінацій: перші чотири комбінації є рядками породжувальної матриці, п'ята – нульової, інші одинадцять комбінацій знайдемо підсумовуванням за модулем 2 різних сполучень рядків породжувальної матриці:

$1 \oplus 2$	1	1	0	0	1	1	0
$1 \oplus 3$	1	0	1	0	1	0	1
$1 \oplus 4$	1	0	0	1	1	0	0
$2 \oplus 3$	0	1	1	0	0	1	1
$2 \oplus 4$	0	1	0	1	0	1	0
$3 \oplus 4$	0	0	1	1	0	0	1
$1 \oplus 2 \oplus 3$	1	1	1	0	0	0	0
$2 \oplus 3 \oplus 4$	0	1	1	1	1	0	0
$1 \oplus 3 \oplus 4$	1	0	1	1	0	1	0
$1 \oplus 2 \oplus 4$	1	1	0	1	0	0	1
$1 \oplus 2 \oplus 3 \oplus 4$	1	1	1	1	1	1	1

Неважко переконатися, що побудований код має мінімальну кодову відстань, яка дорівнює 3.

У систематичному коді процес кодування зводиться до визначення r перевірних розрядів на основі відомих k інформаційних. Кожен перевірний розряд визначається за допомогою *перевірного співвідношення*, а визначення всіх r перевірних розрядів вимагає r перевірних співвідношень. Перевірні співвідношення прийнято записувати одне під одним у вигляді таблиці (матриці), яку називають *перевірною матрицею*. Перевірна матриця містить r рядків і n стовпців, причому кожен рядок є перевірним співвідношенням для знаходження

значення одного з перевірних розрядів. Перевірна матриця утворюється так: будується одинична матриця

$$E_r = \begin{vmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots\ldots\ldots\ldots \\ 0 & 0 & \ldots & 1 \end{vmatrix},$$

після чого до неї ліворуч приписується матриця $D_{k;r}$, що містить k стовпців і r рядків, причому кожен її рядок відповідає стовпцю перевірних розрядів породжувальної матриці в канонічній формі:

$$H = \left| D_{k;r}, E_r \right| = \begin{vmatrix} b_{11} & b_{21} & \ldots & b_{k1} & 1 & 0 & \ldots & 0 \\ b_{12} & b_{22} & \ldots & b_{k2} & 0 & 1 & \ldots & 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_{1r} & b_{2r} & \ldots & b_{kr} & 0 & 0 & \ldots & 1 \end{vmatrix}. \qquad (102)$$

Матриця (102) є перевірною. За її допомогою операція кодування здійснюється дуже просто. Позиції, займані одиницями в i-му рядку приписаної частини перевірної матриці, визначають ті інформаційні розряди, що мають брати участь у формуванні i-го перевірного розряду. Наприклад, якщо перший рядок перевірної матриці має вигляд 1011100, то перевірний розряд визначиться зі співвідношення $a_1 \oplus a_3 \oplus a_4 \oplus a_5 = 0$.

Декодування також зручно здійснювати за допомогою перевірних матриць; при цьому виконується r перевірок на парність відповідно до (102). Якщо хоча б одна з перевірок не дорівнює 0, то це означає, що в прийнятій комбінації є помилки.

Перевірку кодової комбінації при прийманні можна виконати зіставленням прийнятих перевірних розрядів і перевірних розрядів, обчислених на підставі прийнятих інформаційних. Їх сума за модулем 2 називається *синдромом*. Характерною рисою синдрому є те, що він не залежить від виду переданої комбінації, а цілком визначається помилками прийнятої комбінації. Між синдромом і комбінацією, що спричинила помилки, немає взаємно однозначної відповідності – тому самому синдрому відповідають 2^k різних комбінацій помилок. Так, нульовому синдрому відповідає нульова комбінація помилок (тобто безпомилкове приймання), а також інші $2^k - 1$ комбінації помилок, що збігаються з дозволеними кодовими комбінаціями (невиявлені помилки). Тільки одну з комбінацій помилок, що відповідають ненульовому синдрому, можна виправити кодом. При цьому за кожним синдромом закріплюється така виправна комбінація помилок, поява якої в каналі найбільш імовірна. Зазначимо, що синдром збігається з комбінацією результатів перевірки на парність. Це легко перевірити практично.

Для побудованого раніше коду (7, 4) перевірна матриця

$$H = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{vmatrix}. \qquad (103)$$

Тому

$$a_5 = a_2 \oplus a_3 \oplus a_4;$$
$$a_6 = a_1 \oplus a_3 \oplus a_4;$$
$$a_7 = a_1 \oplus a_2 \oplus a_4,$$

де $a_1, \ldots, a_4$ – інформаційні розряди; a_5, a_6, a_7 – перевірні.

Декодування відбувається обчисленням перевірних співвідношень:

$$\begin{matrix} a_2 \oplus a_3 \oplus a_4 \oplus a_5 \\ a_1 \oplus a_3 \oplus a_4 \oplus a_6; \\ a_1 \oplus a_2 \oplus a_4 \oplus a_7. \end{matrix} \qquad (104)$$

Якщо код використовується для виправлення помилок (а даний код може виправити одну помилку), то на приймальному боці має бути заздалегідь визначена відповідність між видом синдрому і видом однократної помилки, що виправляється. Встановимо цю

відповідність. Нехай помилка є в першому розряді (комбінація помилки 1000000). Перевіримо цю комбінацію з урахуванням (104):

$$a_2 \oplus a_3 \oplus a_4 \oplus a_5 = 0 \oplus 0 \oplus 0 \oplus 0 = 0,$$
$$a_1 \oplus a_3 \oplus a_4 \oplus a_6 = 1 \oplus 0 \oplus 0 \oplus 0 = 1,$$
$$a_1 \oplus a_2 \oplus a_4 \oplus a_7 = 1 \oplus 0 \oplus 0 \oplus 0 = 1.$$

Отже, при наявності помилки в 1-му розряді синдром дорівнює 011. Аналогічно можна одержати види синдромів і у разі наявності всіх інших можливих однократних помилок:

Помилка в розряді							
Синдром	01	01	10	11	00	10	01

Нехай, наприклад, у кодувальний пристрій надійшла інформаційна послідовність 1101. Згідно з (2.9) ця послідовність кодується так: 1101001. Тепер припустимо, що в процесі передавання сталася помилка в 2-му розряді і комбінація 1101001 прийнята як 1001001. При декодуванні відповідно до (2.10) будуть отримані такі результати:

1-ша перевірка $0 \oplus 0 \oplus 1 \oplus 0 = 1$,
2-га » $1 \oplus 0 \oplus 1 \oplus 0 = 1$,
3-тя » $1 \oplus 0 \oplus 1 \oplus 0 = 1$.

Отже, синдром дорівнює 101, що свідчить про те, що помилковим є саме 2-й розряд у прийнятій комбінації.

Використання систематичних кодів. Одним із найпростіших систематичних кодів, що одержали практичне застосування, є *код з повторенням*, який має два різновиди. В одному з них є S-кратне повторення комбінації простого коду $a_1, a_2, \ldots, a_k$:

$$\underbrace{a_1 a_2 \ldots a_k}_{1}\ \underbrace{a_1 a_2 \ldots a_k}_{2}\ \ldots\ldots\ \underbrace{a_1 a_2 \ldots a_k}_{S}.$$

Інший різновид коду з повторенням характеризується S-кратним передаванням кожного розряду:

$$\underbrace{a_1 a_1 \ldots a_1}_{S_{\text{разів}}}\ \underbrace{a_2 a_2 \ldots a_2}_{S_{\text{разів}}} \ldots \underbrace{a_k a_k \ldots a_k}_{S_{\text{разів}}}.$$

Код з повторенням має довжину $n = S_k$, число перевірних розрядів $r = k(S-1)$, мінімальну кодову відстань $d_0 = S$. Надлишок цих кодів дорівнює $(S-1)/S$. Звичайно застосовується перший різновид коду з повторенням, що має в умовах корельованих помилок підвищену завадостійкість. Це зумовлено тим, що вхідні в одну перевірку на парність розряди досить далеко стоять один від одного і з малою ймовірністю вражаються одним пакетом помилок. Число повторень звичайно дорівнює 2 ($d_0 = 2$) і набагато рідше 3 ($d_0 = 3$). Велика кратність повторень практично не використовується.

Код з повторенням характеризується досить високими властивостями, що проявляються при дії пакетів помилок. Так, при $S = 2$ завжди виявляються пакети помилок завдовжки до $n/2$, а також усі помилки непарної кратності. При $S = 2$ перевірна матриця має вигляд

$$H = \left| E_k, E_k \right|. \qquad (105)$$

Недоліком кодів з повторенням є дуже великий надлишок. Навіть при дворазовому повторенні він становить 0,5.

Існує особливий різновид коду з дворазовим повторенням, що забезпечує вдвічі більшу мінімальну кодову відстань ($d_0 = 4$). Це так званий *код Бауера* (інверсний код). При використанні даного коду комбінації з парним числом одиниць повторюються в незмінному вигляді, а комбінації з непарним числом одиниць – в інвертованому. Перевірна матриця:

$$H = \left| \overline{E_k}, E_k \right|, \qquad (106)$$

де $\overline{E_k}$ – матриця, отримана з одиничної матриці E_k заміною одиниць нулями, а нулів – одиницями.

При невеликих довжинах кодових комбінацій (до 10 –14) інверсний код за завадостійкістю має такий самий надлишок, що і складніші коди.

Код з парним числом одиниць (код з однією перевіркою на парність, паритетний код), так само, як і код із дворазовим повторенням, забезпечує $d_0 = 2$, але має набагато менший надлишок. Незалежно від довжини кодової комбінації цей код має один перевірний розряд, значення якого вибирають з умови одержання парного числа одиниць у кодовій комбінації, тобто він визначається зі співвідношення

$$b_1 = a_1 \oplus a_2 \oplus ... \oplus a_k .$$

Перевірна матриця коду з парним числом одиниць містить лише один рядок:

$$H = |\, 1\,1\,1 ... 1\,| . \qquad (107)$$

Як і код з дворазовим повторенням, розглянутий код виявляє не тільки однократні помилки, але і взагалі всі помилки непарної кратності. Не слід, однак, думати, що обидва коди мають однакову завадостійкість. У разі дії пакетних помилок завадостійкість коду з дворазовим повторенням набагато вища.

Існує також *код із двома перевірками на парність*. Незалежно від довжини кодової комбінації цей код має два перевірних розряди, один із яких вибирають з умови парності всіх інформаційних розрядів, а другий – з умови парності всіх непарних (або парних) за номером інформаційних розрядів. Так, при $k = 5$ значення перевірних розрядів

$$b_1 = a_1 \oplus a_2 \oplus a_3 \oplus a_4 \oplus a_5 ,\ b_2 = a_1 \oplus a_3 \oplus a_5 .$$

Як і код з парним числом одиниць, розглянутий код має мінімальну кодову відстань $d_0 = 2$, але більшу завадостійкість, оскільки виявляє частину помилок парної кратності – суміжних, тобто розташованих поруч.

Перевірна матриця коду з двома перевірками на парність має вигляд:

$$H = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & \dots & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & \dots & 0 & 1 \end{vmatrix} .$$

Коди Хеммінга. Кодами Хеммінга називають коди з мінімальною кодовою відстанню $d_0 = 3$, що виправляють усі одиничні помилки, і коди з відстанню $d_0 = 4$, що виправляють усі одиничні й усі подвійні помилки. Коди Хеммінга з мінімальною кодовою відстанню $d_0 = 3$ мають довжину $n \le 2^r - 1$. Відповідно

$$2^k \le \frac{2^n}{n+1} . \qquad (108)$$

Характерною рисою перевірної матриці цього коду з $d_0 = 3$ є те, що її стовпці – це будь-які різні ненульові комбінації завдовжки *r*. Наприклад, для коду Хеммінга (7, 4) можлива така перевірна матриця:

$$H = \begin{vmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{vmatrix} .$$

Стовпці перевірної матриці можна довільно переставляти, при цьому коректувальні властивості коду не зміняться.

Коди Хеммінга з мінімальною кодовою відстанню $d_0 = 4$ утворюються на основі кодів з $d_0 = 3$ уведенням ще одного перевірного розряду, що доповнює кодову комбінацію до парного числа одиниць. Перевірна матриця даного коду також утвориться з перевірної матриці коду Хеммінга з $d_0 = 3$ приписуванням праворуч нульового стовпця і приписуванням додаткового рядка, що складається тільки з одиниць. Перевірна матриця, наприклад коду (8, 4), має вигляд

$$H = \begin{vmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{vmatrix}.$$

При декодуванні кодової комбінації, переданої кодом Хеммінга з $d_0 = 4$, можливі три випадки:

- помилок немає (всі перевірки означають відсутність помилок);
- одинична помилка (всі перевірки означають наявність помилки);
- подвійна помилка (остання перевірка означає відсутність помилок, а інші – їх наявність).

Як видно, подвійна помилка, яка не виправляється кодом, має місце тільки тоді, коли результат хоча б однієї з перших чотирьох перевірок ненульовий, а результат останньої перевірки дорівнює нулеві. Це дає змогу запобігти видаванню одержувачу спотвореного повідомлення, тобто знайти подвійні помилки.

Іноді під кодами Хеммінга розуміють особливий різновид розглянутих кодів з $d_0 = 3$, який відрізняється тим, що комбінація синдрому збігається з номером спотвореного розряду, записаним у двійковій формі. Так, при спотворенні, наприклад, третього розряду синдром має вигляд 011, четвертого – 100 і т. д. Ці коди характеризуються особливим порядком розташування перевірних розрядів – на першому, другому, четвертому, восьмому, шістнадцятому і т. д. місцях. Великого поширення ці коди не одержали, тому докладно їх не розглядатимемо.

Циклічні коди є різновидом систематичних кодів і мають усі їх властивості. Спочатку вони були створені з метою спрощення схем кодування і декодування. Згодом виявилися їх високі коректувальні властивості, що й забезпечило їм широке використання на практиці.

При побудові циклічних кодів кодові комбінації подають у вигляді поліномів

$$G(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1x + a_0, \qquad (109)$$

де $a_0, a_1, \ldots, a_{n-1}$ – коефіцієнти, що набувають значень 0 або 1. Наприклад, комбінацію 1100101 можна записати як $G(x) = x^6 + x^5 + x^2 + 1$.

Основна властивість розглянутих кодів полягає в тому, що циклічний зсув дозволеної кодової комбінації також є дозволеною кодовою комбінацією. Отже, якщо комбінація 1000111 є дозволеною, то комбінації 0001111, 0011110 і т. д. теж дозволені. Нульова комбінація є дозволеною, оскільки циклічний код належить до класу систематичних. Зазначимо, що циклічний зсув нульової комбінації є також нульовою комбінацією.

Можна показати, що циклічний зсув є еквівалентним множенню на x кодової комбінації, записаної у вигляді полінома. Дійсно,

$$xG(x) = a_{n-1}x^n + a_{n-2}x^{n-1} + \ldots + a_1x^2 + a_0x.$$

Оскільки в кодовій комбінації, що має довжину n, степінь полінома не може перевищувати $n-1$ (у противному разі довжина кодової комбінації перевищить n), то x^n замінюють на 1. При цьому

$$xG(x) = a_{n-2}x^{n-1} + \ldots + a_1x^2 + a_0x + a_{n-1}.$$

Отже, $xG(x)$ є циклічним зсувом комбінації $G(x)$.

Циклічні коди прийнято визначати за допомогою твірних поліномів $P(x)$ степеня n. Твірну матрицю циклічного коду можна утворити з твірного полінома циклічним зсувом останнього (або, що те саме, множенням його на $x, x^2, \ldots, x^{n-1}$):

$$G = \begin{vmatrix} P(x) \\ xP(x) \\ x^2P(x) \\ x^{k-1}P(x) \end{vmatrix}. \qquad (110)$$

Безпосередньо з твірної матриці (2.16) випливає, що всі дозволені кодові комбінації циклічного коду без остачі діляться на похідний поліном. Нагадаємо, що тут і далі мається на увазі ділення за модулем 2. Останнє виконується майже так само, як звичайне ділення, з тією різницею, що в процесі ділення операція віднімання замінюється додаванням за модулем 2. Поліном $x^6 + x^5 + x^3 + 1$ на поліном $x^2 + x + 1$ (відповідні двійкові форми 1101001 і 111) ділиться так:

$$
\begin{array}{r|l}
1101001 & 111 \\
\underline{111} & \overline{10101} \\
110 & \\
\underline{111} & \\
101 & \\
\underline{111} & \\
10 &
\end{array}
$$

Двочлен 10 є остачею від ділення полінома 1101001 на поліном 111.

Оскільки кожна дозволена комбінація без остачі ділиться на твірний поліном, то ця обставина стає основою дуже зручного способу визначення приналежності прийнятої кодової комбінації до дозволених.

Розглянемо принцип побудови циклічних кодів. Кожну кодову комбінацію $G(x)$ простого k-елементного коду помножимо на x^r, а потім поділимо на твірний поліном степеня r. Внаслідок множення степінь кожного члена x_i, що входить у поліном $G(x)$, підвищується на r. При діленні добутку $x^r G(x)$ на $P(x)$ виходить частка $Q(x)$ такого самого степеня, що $G(x)$. Крім того, якщо добуток $x^r G(x)$ не ділиться на $P(x)$ без остачі, то остачею є $R(x)$:

$$\frac{x^r G(x)}{P(x)} = Q(x) \oplus \frac{R(x)}{P(x)}. \tag{111}$$

Оскільки частка $Q(x)$ має той самий степінь, що $G(x)$, то вона також є комбінацією простого k-елементного коду.

Помноживши обидві частини рівності (2.17) на $P(x)$, маємо

$$F(x) = Q(x)P(x) = x^r G(x) \oplus R(x). \tag{112}$$

Отже, кодову комбінацію циклічного коду можна отримати двома способами:

- множенням k-елементної комбінації простого коду на твірний поліном $P(x)$;
- множенням кодової комбінації простого коду на одночлен x^r і додаванням до цього добутку остачі від ділення добутку $x^r G(x)$ на $P(x)$.

Зазначимо, що перший спосіб спричинює утворення *нероздільного* коду. Нероздільність значно ускладнює процес декодування, тому на практиці використовується *інший спосіб* побудови кодових комбінацій. Дуже важливо, що цей спосіб дає можливість одержати твірну матрицю відразу в канонічній формі:

$$G = \left| \, E_k; \quad C_{r;k} \, \right|,$$

де $C_{r;k}$ – матриця, що складається з r стовпців і k рядків, причому кожен рядок є остачею від ділення рядка одиничної матриці, доповненої r-нулями, на твірний поліном.

Кодують і декодують циклічні коди не на основі обчислення перевірок на парність, а на основі ділення на твірний поліном. Проте за необхідності перевірну матрицю можна побудувати обчисленням перевірного полінома

$$h(x) = \frac{x^n + 1}{P^{-1}(x)}, \tag{113}$$

де $P^{-1}(x)$ – поліном, об'єднаний з твірним поліномом $P(x)$. Нагадаємо, що в об'єднаних поліномах члени розташовані в зворотному порядку. Так, поліноми 100111 і 111001 є об'єднаними. Перший рядок перевірної матриці циклічного коду є перевірним поліномом $h(x)$,

помноженим на x^{r-1} (тобто доповнений праворуч $r-1$ нулями). Наступні рядки перевірної матриці є циклічним зсувом першої.

Перевірна матриця може бути також побудована звичайним способом, виходячи з твірної матриці. У такому разі перевірна матриця зовні може відрізнятися від побудованої за допомогою перевірного полінома, однак обидві матриці завжди можуть бути зведені до одного вигляду.

Наведемо приклад побудови семирозрядного циклічного коду з $d_0 = 3$. Для цього потрібні три перевірних розряди ($d_0 = 3$), а твірний поліном має бути третього степеня. Нехай твірний поліном $P(x) = x^3 + x + 1 = 1011$. (Принципи вибору твірного полінома розглянемо пізніше.) Твірна матриця має вигляд

$$G = \begin{vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{vmatrix}.$$

Нагадаємо, що перевірні розряди 101 маємо внаслідок ділення комбінації 1000000 на твірній поліном 1011, а 111 – внаслідок ділення 100000 на той самий твірній поліном, і т. д.

Перевірний поліном

$$h(x) = \frac{x^7 + 1}{(x^3 + x + 1)^{-1}} = \frac{10000001}{1101} = 11101,$$

отже, перевірна матриця

$$H = \begin{vmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{vmatrix}.$$

Якщо будувати твірну матрицю з породжувальної, то одержимо матрицю

$$H' = \begin{vmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{vmatrix},$$

яка відрізняється останнім рядком від раніше отриманої матриці H. Однак цю відмінність можна ліквідувати, якщо до останнього рядка матриці H' додати перший рядок; при цьому матриця H' збігається з матрицею H.

Коректувальна здатність циклічного коду цілком визначається виглядом твірного полінома. Для деяких циклічних кодів можна сформулювати досить прості принципи вибору вигляду твірного полінома.

Циклічний код з $d_0 = 2$. Твірний поліном має вигляд $x + 1$. Цей поліном дає змогу будувати код будь-якої довжини. Циклічний код з $d_0 = 2$ виявляє будь-яку непарну кількість помилок і повністю тотожний коду з парним числом одиниць.

Степінь полінома	Вигляд полінома
1	$1+x$
2	x^2+x+1
3	x^3+x+1
	x^3+x^2+1
4	x^4+x+1
	x^4+x^3+1
	$x^4+x^3+x^2+x+1$
5	x^5+x^2+1
	x^5+x^3+1
	$x^5+x^3+x^2+x+1$
	$x^5+x^4+x^2+x+1$
	$x^5+x^4+x^3+x+1$
	$x^5+x^4+x^3+x^2+1$

Твірним поліномом для циклічного коду з $d_0=2$ може бути також поліном x^2+x+1. Код при цьому має підвищену завадостійкість – виявляються не тільки будь-які помилки непарної кратності, але й будь-які парні суміжні помилки (тобто пакети помилок завдовжки 2), а також усі парні помилки, поділені одним неспотвореним розрядом.

Циклічний код з $d_0=3$ є різновидом кодів Хеммінга. Довжину кодової комбінації вибирають з умови $n=2^r-1$. Твірним поліномом може бути будь-який *незвідний поліном* степеня *r*. (Поліном називається незвідним, якщо він ділиться без остачі тільки на одиницю і на самого себе.) Незвідні поліноми до п'ятого степеня включно наведені поруч на цій самій сторінці.

Циклічний код з $d_0=4$ є також різновидом кодів Хеммінга і будується на основі твірних поліномів для кодів з $d_0=3$. Твірний поліном циклічного коду з $d_0=4$ є добутком двочлена $x+1$ на незвідний поліном, що придатний як твірний для коду з $d_0=3$. Довжину кодової комбінації вибирають з умови $n=2^m-1$; число перевірних розрядів $r=m+1$. Наприклад, при $n=7$ твірний поліном має вигляд $(x+1)(x^3+x+1)=x^4+x^3+\ +x^2+1$.

Дуже потужними кодами, що мають високу коректувальну здатність, є коди Боуза–Чоудхурі–Хоквінгема. Довжина кодової комбінації становить $n=2^m-1$. Твірний поліном знаходять як найменше спільне кратне *мінімальних поліномів* $a_i(x)$:

$$P(x)=\text{НСК}\left\{a_1(x)a_3(x)\ldots a_i(x)\ldots a_{d_0-2}(x)\right\}. \qquad (114)$$

Мінімальні поліноми для $m\le 7$ наведені в табл. 10.1, причому всі вони є незвідними.

Таблиця 2. **Мінімальні поліноми**

	Значення *m*					
	2	3	4	5	6	7
	x^2+x+1	x^3+x+1	x^4+x+1	x^5+x^2+1	x^6+x+1	x^7+x^3+1
	–	–	$x^4+x^3+x^2+x+1$	$x^5+x^4+x^3+x^2+1$	$x^6+x^4+x^2+x+1$	$x^7+x^3+x^2+x+1$
	–	–	–	$x^5+x^4+x^2+x+1$	$x^6+x^5+x^2+x+1$	$x^5+x^4+x^3+x^2+1$
	–	–	–	–	x^6+x^3+1	$x^7+x^6+x^5+x^4+x^2+x+1$

Нехай потрібно побудувати код Боуза–Чоудхурі–Хоквінгема з $n=31$ і $d_0=5$. Маємо

$$P(x)=\text{НСК}\left\{a_1(x)a_3(x)\right\}=a_1(x)a_3(x).$$

Оскільки в даному випадку *m* = 5, то значення мінімальних поліномів випишемо з 5-го стовпця табл. 2.

$$a_1(x)=x^5+x^2+1,$$

$$a_3(x)=x^5+x^4+x^3+x^2+1,$$

$$P(x)=x^{10}+x^9+x^8+x^6+x^5+x^3+1.$$

Коди Боуза–Чоудхурі–Хоквінгема мають непарні значення мінімальної кодової відстані d_0. При бажанні мінімальну кодову відстань можна збільшити на одиницю, застосувавши твірний поліном, що дорівнює добутку твірного полінома коду Боуза–Чоудхурі–Хоквінгема на двочлен $x+1$. Так, у розглянутому коді з $d_0=5$ мінімальну кодову відстань можна підвищити до 6, якщо використовувати твірний поліном

$$P(x) = = (x+1)(x^{10}+x^9+x^8+x^6+x^5+x^3+1) = x^{11}+x^8+x^7+x^5+x^4+x^3+x+1.$$

Зазначимо, що такий спосіб збільшення мінімальної кодової відстані можна застосувати до будь-яких систематичних кодів з непарною мінімальною кодовою відстанню. Для цього в циклічних кодах змінюється описаним чином твірний поліном, а в нециклічних кодах уводиться додаткова перевірка на парність, що охоплює всі інформаційні розряди аналогічно тому, як це було зроблено в циклічних і нециклічних кодах Хеммінга з $d_0 = 4$.

Циклічні коди поширені в системах передавання даних і використовуються у різній за призначенням апаратурі передавання даних (АПД). Незважаючи на те, що умови застосування кожної конкретної АПД різні, а використовувані канали можуть сильно відрізнятися за своїми характеристиками, є тенденція до стандартизації методів підвищення надійності приймання за допомогою коректувальних кодів і, зокрема, стандартизації твірних поліномів. Це дає можливість у ряді випадків спільно використовувати апаратуру різного виробництва, комплектувати АПД різного призначення тими самими функціональними вузлами і т. д. Зараз для середньошвидкісних систем передавання даних існує рекомендація МСЕ V.41, відповідно до якої для підвищення надійності приймання пропонується використовувати виявлення помилок за допомогою циклічного коду, що має довжину кодової комбінації 260, 500 і 980 розрядів, причому у всіх випадках береться твірний поліном $x^{16}+x^{12}+x^5+1$. Це код Боуза–Чоудхурі–Хоквінгема з мінімальною кодовою відстанню $d_0 = 4$. Численні випробування коду з зазначеним твірним поліномом підтвердили його високу ефективність. Навіть при використанні цього коду для передавання даних по комутованій телефонній мережі загального користування (коефіцієнт помилок по одиничних елементах порядку 10^{-3} і більше) частота помилкового приймання восьмирозрядних знаків, з яких складена інформаційна частина кодової комбінації, не перевищує 10^{-6}.

Тепер розглянемо питання про вибір довжини інформаційної частини кодової комбінації циклічного коду. ЕОМ, що є джерелом або споживачем інформації, переданої АПД, обмінюються з зовнішніми пристроями або безпосередньо машинними словами (16, 24, 32 і т. д. розрядів) або складами (байтами), що містять 8 розрядів. Тому інформаційна частина кодової комбінації має дорівнювати або бути кратною довжині машинного слова, якщо ЕОМ обмінюється з зовнішніми пристроями машинними словами, або дорівнювати довжині байта, якщо ЕОМ обмінюється з зовнішніми пристроями байтами. Проте можливості циклічних кодів не забезпечують вільного варіювання числом інформаційних розрядів при заданій мінімальній кодовій відстані або заданому числі перевірних розрядів. У зв'язку з цим на практиці часто скорочують циклічний код.

Укорочені циклічні коди отримують з повних циклічних кодів, що мають необхідну мінімальну кодову відстань (число перевірних розрядів), але більше, ніж потрібно, число інфор- маційних розрядів. Укорочення повного циклічного коду полягає в тому, що для передавання інформації використовуються не всі комбінації повного коду, а тільки ті, які містять ліворуч нулі, причому ці нулі в канал зв'язку зовсім не передаються. Практично, щоб укоротити код на l розряд, треба викреслити з відповідної матриці однин верхній рядок і однин лівий стовпець.

Наведемо приклад побудови укороченого циклічного коду. Нехай необхідно побудувати код, що має 5 інформаційних розрядів і мінімальну кодову відстань, яка дорівнює 3. Як відомо, при $d_0 = 3$ довжина кодової комбінації становить $n = 2^r - 1$, що дає змогу будувати коди зі значеннями *n*, *k* і *r*, наведеними нижче:

5	1	3
1	6	7

Отже, не існує циклічних кодів з $d_0 = 3$, що містять 5 інформаційних розрядів. Однак код з необхідними параметрами можна утворити з коду (15, 11), укоротивши його на 6

розрядів. Твірним поліномом цього коду може бути будь-який поліном четвертого степеня. Твірна матриця циклічного (15, 11)-коду має вигляд

$$G(15,11)=\begin{vmatrix}1&0&0&0&0&0&0&0&0&0&0&1&0&0&1\\0&1&0&0&0&0&0&0&0&0&0&1&1&0&1\\0&0&1&0&0&0&0&0&0&0&0&1&1&1&1\\0&0&0&1&0&0&0&0&0&0&0&1&1&1&0\\0&0&0&0&1&0&0&0&0&0&0&0&1&1&1\\0&0&0&0&0&1&0&0&0&0&0&1&0&1&0\\0&0&0&0&0&0&1&0&0&0&0&0&1&0&1\\0&0&0&0&0&0&0&1&0&0&0&1&0&1&1\\0&0&0&0&0&0&0&0&1&0&0&1&1&0&0\\0&0&0&0&0&0&0&0&0&1&0&0&1&1&0\\0&0&0&0&0&0&0&0&0&0&1&0&0&1&1\end{vmatrix}.$$

Коректувальна здатність укороченого циклічного коду буде не нижча за коректувальну здатність початкового повного циклічного коду. Техніка кодування і декодування в обох випадках та сама. Однак циклічний зсув кодової комбінації укороченого циклічного коду не завжди спричинює утворення дозволеної комбінації, тому вкорочені коди належать до псевдоциклічних.

Контрольні питання до Теми 6:

1. Що таке коректувальні коди і для чого вони використовуються?
2. Які основні відмінності між циклічними і несистематичними кодами?
3. Що таке код Боуза–Чоудхурі–Хоквінгема (BCH), і як він застосовується?
4. Які основні властивості циклічних кодів?
5. Які переваги та недоліки має систематичне кодування?

ТЕМА 7 НЕСИСТЕМАТИЧНІ ТА ІТЕРАТИВНІ КОДИ. ЕФЕКТИВНІСТЬ СИСТЕМАТИЧНИХ КОРЕКТУВАЛЬНИХ КОДІВ

Несистематичні коди в техніці передавання даних використовуються рідше, ніж систематичні. Одним з найбільш відомих кодів цього типу є код з постійною вагою, що називають також кодом з постійним відношенням числа одиниць і нулів. У коді з постійною вагою дозволеними є тільки ті комбінації, що містять визначене число одиниць, однакове у всіх дозволених комбінаціях. Серед кодів з постійною вагою найбільш поширений на практиці семирозрядний код з постійною вагою 3, кожна комбінація якого містить три одиниці і чотири нулі. Цей код рекомендовано МСЕ для використання при передаванні телеграфних повідомлень по короткохвильових радіоканалах і відомий як *міжнародний телеграфний код № 3.*

Міжнародний телеграфний код № 3 містить 35 дозволених кодових комбінацій, мінімальна кодова відстань $d_0 = 2$. Код виявляє всі одиничні помилки, а також усі помилки непарної кратності. З різних помилок парної кратності не виявляються тільки помилки типу зсувів (транспозицій), при яких від однієї або кількох одиниць у кодовій комбінації переходять у нулі, а така сама кількість нулів – в одиниці. У цілком асиметричних каналах (тобто каналах, у яких переходять або тільки одиниці у нулі, або тільки нулі в одиниці) транспозиції не можуть мати місця, тому коди з постійною вагою в таких каналах виявляють усі помилки. Саме з цієї причини коди з постійною вагою поширені в апаратурі, що працює на короткохвильових радіоканалах, які є асиметричними.

Істотним недоліком кодів з постійною вагою є те, що вони належать до числа нероздільних кодів, у комбінаціях яких неможливо виділити інформаційні і перевірні розряди.

Це значно ускладнює кодування і декодування, внаслідок чого ускладнюється і дорожчає апаратура. Тому останнім часом дані коди поступово витісняються систематичними і, зокрема, циклічними. Цьому сприяє наявна тенденція до створення досить універсальних пристроїв захисту від помилок, придатних для роботи на різних каналах, як провідних, так і короткохвильових.

Коди Бергера (коди з підсумовуванням) призначені для використання в асиметричних каналах. Мінімальна кодова відстань у цих кодах $d_0 = 2$. Існує ряд варіантів побудови кодів Бергера. У найпростішому варіанті кодування відбувається так: в інформаційній частині кодової комбінації підраховується число одиниць, після чого формуються перевірні розряди, що являють собою запис цього числа в двійковій формі. Так само формуються перевірні розряди під час приймання і порівнюються з прийнятими перевірними розрядами. Підвищення надійності за допомогою кодів Бергера дає такі самі результати, як і використання коду № 3, однак найважливішою перевагою коду Бергера є його *роздільність*, що різко спрощує побудову кодувальних і декодувальних пристроїв.

Для передавання даних поширені коди з інвертованими перевірками на парність. Це несистематичні коди, одержувані із систематичних інвертуванням одного або декількох перевірних розрядів. Так, широко застосовується *код з непарним числом одиниць,* що відрізняється від коду з парним числом одиниць інвертованим перевірним розрядом. Коректувальна здатність кодів з інвертованими перевірками цілком збігається з коректувальною здатністю систематичних кодів. Значною перевагою розглянутих кодів є те, що вони не містять кодових комбінацій, які складаються з одних нулів. Ця обставина дуже сприятлива для побудови систем передавання даних, оскільки відсутність у коді нульових комбінацій підвищує стійкість синхронізації і тим самим поліпшує роботу систем передавання даних. Крім того, у ряді випадків застосування кодів з інвертованими перевірками, отриманих з циклічних кодів, дає змогу набагато ефективніше виявляти порушення циклової синхронізації, ніж при використанні звичайних циклічних кодів. Така властивість кодів з інвертованими перевірками значною мірою підвищує стійкість роботи системи передавання даних.

Ітеративні (матричні) коди характеризуються наявністю двох перевірок усередині кожної кодової комбінації. Розглянемо принцип побудови ітеративного коду на конкретному прикладі. Запишемо всі інформаційні розряди блоку, що підлягає передаванню, у вигляді таблиці, що, наприклад, може мати такий вигляд:

1 0 1 1 1
0 0 1 0 0
1 1 1 0 0
0 1 0 0 1
1 1 1 1 0

Закодуємо спочатку кожен рядок таблиці яким-небудь кодом, а потім (не обов'язково тим самим кодом) – кожен стовпець. За перший код візьмемо код з парним числом одиниць, а за другий – з непарним. Тоді одержимо

1 0 1 1 1 0
0 0 1 0 0 1
1 1 1 0 0 1
0 1 0 0 1 0
1 1 1 1 0 0
0 0 1 1 1 1

Отримана комбінація є кодовою комбінацією найпростішого двовимірного ітеративного коду, перевірні розряди якого зосереджені в правому стовпці і нижньому рядку. Кожен інформаційний розряд цього коду входить у комбінацію двох ітеративних кодів – коду з парним числом одиниць і коду з непарним числом одиниць. Передавання комбінації ітеративного коду звичайно відбувається по рядках послідовно – від першого рядка до останнього.

Можуть бути утворені також багатовимірні ітеративні коди, в яких кожен інформаційний розряд входить у комбінації трьох, чотирьох і т. д. ітеративних кодів, однак багатовимірні ітеративні коди менш поширені.

Властивості ітеративного коду визначаються його параметрами, залежно від яких код може бути як систематичним, так і несистематичним, як роздільним, так і нероздільним. Довжина кодової комбінації, число інформаційних розрядів і мінімальна кодова відстань ітеративного коду дуже просто виражаються через відповідні параметри цих кодів:

$$n = \prod_{i=1}^{S} n_i,\ k = \prod_{i=1}^{S} k_i,\ d_0 = \prod_{i=1}^{S} d_{0i}, \tag{115}$$

де n_i, k_i, d_{0i} – параметри ітеративних кодів; S – кратність ітерування.

Отже, найпростіший ітеративний код, утворений перевіркою на парність (непарність) рядків і стовпців, має мінімальну кодову відстань $d_{0i} = 4$ і тому дає можливість виявляти всі помилки кратністю до 3. Крім того, виявляються всі помилки непарної кратності. Не виявляються чотириразові помилки, що розташовуються у вершинах правильного чотирикутника, а також деякі шестиразові, восьмиразові і т. д. помилки (рис. 9).

Найпростіший ітеративний код має досить високу виявляючу здатність – при дії пакетних помилок виявляється будь-який пакет помилок завдовжки $l+1$ і менше, де l – довжина рядка.

На практиці найпоширеніші *двовимірні ітеративні коди.* Довжину рядка звичайно вибирають такою, що дорівнює довжині одного знака первинного коду. Як ітеративні коди найчастіше використовуються коди з однією і двома перевірками на парність і набагато рідше коди Хеммінга. При використанні досить довгих блоків (завдовжки в кілька десятків знаків і більше) проста перевірка на парність по рядках і стовпцях забезпечує на реальних каналах зв'язку коефіцієнт помилок на знак порядку 10^{-6}, а застосування кодів із двома перевірками на парність і кодів Хеммінга – порядку $10^{-8} - 10^{-10}$.

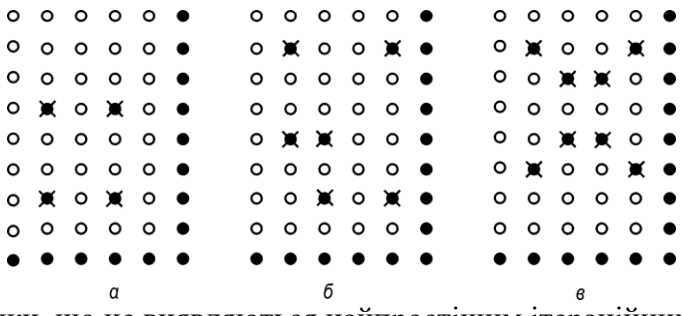

Рис. 9. Деякі помилки, що не виявляються найпростішим ітераційним кодом:
● – чотириразові помилки; ✖ – шести-, восьмиразові і т. д. помилки

Істотним недоліком ітеративних кодів, які використовують для перевірок по рядках і стовпцях кодів з однією або двома перевірками на парність, є їх порівняно високий надлишок, що звичайно становить 15–20 % і значно перевищує за інших рівних умов надлишок циклічних кодів. Однак кодування і декодування за допомогою ЕОМ таких ітеративних кодів звичайно набагато простіше, ніж циклічних. Тому найпростіші ітеративні коди, незважаючи на їх високий надлишок, застосовують у системах передавання даних, що використовують програмні способи підвищення надійності (зокрема, у системах з комутацією повідомлень).

Розглянемо ефективність використання систематичних коректувальних кодів, що виявляють помилки. Найважливішою характеристикою таких кодів є коефіцієнт підвищення правильності

$$K_{\text{п.п}} = \frac{P_{\text{пом}}}{P_{\text{н.пом}}}, \tag{116}$$

де $P_{\text{пом}}$ – частота появи помилкових кодових комбінацій у дискретному каналі (тобто перед пристроєм захисту від помилок); $P_{\text{н.пом}}$ – частота появи кодових комбінацій з невиявленими помилками в каналі передавання даних (тобто після пристрою захисту від помилок).

Характерною рисою використання коректувальних кодів є те, що їх ефективність залежить від особливостей розподілу помилок у каналах. Код, що виявляє якісь певні сполучення помилок, буде досить ефективним при його використанні в каналах, де ці сполучення є переважаючими. Так, код з однією перевіркою на парність виявляє всі помилки непарної кратності і може при нежорстких вимогах до надійності приймання бути цілком придатним для роботи з каналами з частотною модуляцією (ЧМ). Проте цей код практично непридатний під час роботи з каналами з відносною фазовою модуляцією (ВФМ), характерною рисою яких є помилки парної кратності. Причина неоднакового поводження коду в різних каналах – в різних закономірностях розподілу помилок у них.

Відомо, що коректувальний код з мінімальною кодовою відстанню d_0 виявляє всі сполучення помилок кратністю до $d_0 - 1$ включно. Тому ймовірність невиявлення помилки кодом можна визначити як

$$P_{\text{н.пом}} = \frac{B(d_0)}{C_n^{d_0}} P(d_0, n) + \frac{B(d_0+1)}{C_n^{d_0+1}} P(d_0+1, n) + \ldots = \sum_{i=d_0}^{n} \frac{B(i)}{C_n^i} P(i, n), \qquad (117)$$

де $B(i)$ – кількість i-кратних помилок, що не виявляються кодом; C_n^i – загальна кількість можливих i-кратних помилок у кодовій комбінації зав-довжки n (дорівнює числу сполучень з n по i); $P(i, n)$ – імовірність появи i-кратних помилок у кодовій комбінації завдовжки n. Ця ймовірність цілком залежить від властивостей використовуваного каналу.

Визначення відношення B_i / C_n^i у багатьох випадках є дуже трудомісткою задачею, особливо при великих значеннях n. Доведено, що для багатьох систематичних кодів це відношення порівняно мало залежить від конкретного різновиду коду. Існує наближене співвідношення

$$\frac{B(i)}{C_n^i} \approx 2^{-r}, \qquad (118)$$

яке визначає відношення кількості комбінацій помилок, що не виявляються, кратності k до загальної кількості можливих комбінацій помилок цієї кратності для коду в середньому. Зазначимо, що дане співвідношення дуже близьке до відношення кількості комбінацій помилок, які не виявляються кодом, до загальної кількості можливих помилок без урахування їх кратності. Дійсно, загальне число можливих комбінацій помилок різної кратності від 1 до n включно дорівнює $2^n - 1$, причому з них не виявляються тільки ті комбінації, що збігаються з ненульовими кодовими комбінаціями, число яких дорівнює $2^n - 1$. Їх відношення

$$\frac{2^k - 1}{2^n - 1} \approx \frac{2^k}{2^n} = 2^{-r}.$$

Отже,

$$P_{\text{н.пом}} \approx 2^{-r} \sum_{i=d_0}^{n} P(i, n) = 2^{-r} P(\geq d_0, n), \qquad (119)$$

де $P(\geq d_0, n)$ – імовірність того, що в кодовій комбінації завдовжки n буде помилок більше, ніж d_0, або їх кількість дорівнюватиме d_0. Нагадаємо, що ймовірність появи в кодовій комбінації $P_{\text{пом}}$ помилок записується як $P_{\text{пом}} = P(\geq 1, n)$. Тоді коефіцієнт підвищення правильності коду можна визначати як

$$K_{\text{п.п}} = \frac{P_{\text{пом}}}{P_{\text{н.пом}}} = \frac{P(\geq 1, n)}{2^{-r} P(\geq d_0, n)} = 2^r \frac{P(\geq 1, n)}{P(\geq d_0, n)}. \qquad (120)$$

Відношення $P(\geq 1, n)/P(\geq d_0, n)$ завжди більше за 1. Тому коефіцієнт підвищення правильності коду $K_{\text{п.п}}$ майже завжди перевищує 2^r. Для більшості систематичних кодів величина 2^r є досить надійною нижньою оцінкою коефіцієнта підвищення правильності $K_{\text{п.п}}$.

Співвідношення (120) визначає коефіцієнт підвищення надійності коду за допомогою ймовірності $P(\geq d_0, n)$, тобто ймовірності того, що в кодовій комбінації буде помилок більше, ніж d_0, або їх кількість дорівнюватиме d_0. Іноді зручніше користуватися не ймовірностями появи помилок визначеної кратності, а ймовірностями появи пакетів помилок визначеної довжини. При цьому нагадаємо, що пакетом помилок називається комбінація помилок, котра починається і закінчується помилковими розрядами, між якими можуть розташовуватися як помилкові, так і безпомилкові розряди. Циклічний код, утворений поліномом степеня r, виявляє будь-який пакет помилок завдовжки r і менше. Правдивість цього положення легко зрозуміти, якщо врахувати, що пакету помилок завдовжки r і менше відповідає поліном помилок степеня, меншого за r, оскільки він не ділиться без остачі на твірний поліном $P(x)$. Отже, кодова комбінація, що є сумою за модулем 2, передані кодові комбінації і комбінації завади не ділитимуться без остачі на твірний поліном. Унаслідок цього будь-який пакет помилок завдовжки r і менше завжди виявляється.

Наведемо без доведення положення, що характеризують властивості циклічних кодів:

- кількість пакетів завдовжки $r+1$, що виявляють циклічним кодом, становить $1/2^{r-1}$ частини всіх пакетів завдовжки $r+1$;
- кількість пакетів завдовжки більше $r+1$, що виявляють циклічним кодом, становить $1/2^r$ частини всіх пакетів помилок завдовжки від $r+2$ до n включно.

Отже, для циклічних кодів можна записати, що

$$P_{\text{н.пом}} = \frac{1}{2^{r-1}} P(r+1, n) + \frac{1}{2^r} P(\geq r+2, n) \approx \frac{1}{2^r} P(> r, n), \qquad (121)$$

де $P(> r, n)$ – імовірність появи в кодовій комбінації пакету помилок завдовжки більше r розрядів. Імовірність появи в кодовій комбінації пакету помилок будь-якої довжини $P(\geq 1, n)$ дорівнює, і в цьому легко переконатися, ймовірності появи в кодовій комбінації помилки. Тому коефіцієнт підвищення правильності при використанні циклічного коду

$$K_{\text{п.п}} \approx 2^r \frac{P(\geq 1, n)}{P(> r, n)} . \qquad (122)$$

Величина 2^r для більшості випадків є досить надійною нижньою оцінкою коефіцієнта підвищення правильності циклічних кодів.

Отже, при використанні систематичних кодів можна вважати, що ймовірність невиявлення кодом помилкового блоку в 2^r рази менша, ніж імовірність появи помилкових блоків на вході декодувального пристрою. Властивості виявляючих систематичних кодів цілком визначаються кількістю перевірних розрядів. Досить істотне значення має вигляд перевірних співвідношень, а для циклічних кодів – вигляд твірного полінома. Серед систематичних кодів з однаковим числом перевірних розрядів існують так звані “сильні” коди, що забезпечують на даних каналах коефіцієнт підвищення правильності, який значно перевищує 2^r. Є також “слабкі” коди, коефіцієнт підвищення правильності яких менше ніж 2^r. Тому при створенні систем передавання даних завжди відбирають “сильні” коди, які щонайкраще відповідають розподілу помилок у дискретних каналах, і найбільші значення коефіцієнта, що забезпечують підвищення правильності. Саме таким “сильним” кодом є згаданий вище циклічний код з твірним поліномом $x^{16} + x^{12} + x^5 + 1$.

Контрольні питання до Теми 7:

1. Що таке несистематичні коди, і як вони відрізняються від систематичних?
2. Як оцінюється ефективність коректувальних кодів?
3. Як визначається ймовірність помилки при використанні ітеративних кодів?
4. Які методи оптимізації ітеративного декодування існують?
5. Як ітеративні методи впливають на швидкість і точність передачі даних?

ТЕМА 8 МАЖОРИТАРНЕ ДЕКОДУВАННЯ. УЗАГАЛЬНЕННЯ ТЕОРІЇ КОДУВАННЯ ТА НЕДВІЙКОВІ КОДИ

В основі мажоритарного способу виправлення помилок – визначення, який розряд кодової комбінації, що декодує, за більшістю результатів перевірок на парність значно спрощує декодери. Мажоритарне декодування застосовують не для всіх кодів, а для тих, структура яких має певні особливості. До таких кодів належать, зокрема, деякі циклічні коди, а також більш прості коди, у тому числі коди з повторенням при $S > 2$. Розглянемо дуже простий випадок мажоритарного декодування стосовно використання кодів з розділеними перевірками.

Як відомо, кожен рядок перевірної матриці будь-якого систематичного коду

$$H = \begin{vmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1n} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2n} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ h_{r1} & h_{r2} & h_{r3} & \dots & h_{rn} \end{vmatrix}$$

складається з коефіцієнтів, що входять у співвідношення перевірки на парність: $h_{i1}x_1 \oplus h_{i2}x_2 \oplus \dots \oplus h_{in}x_n = 0$. Лінійні комбінації рядків перевірної матриці також утворюють перевірні співвідношення. Виконавши μ лінійних операцій над рядками перевірної матриці, можна (для деяких кодів) одержати нову матрицю

$$L = \begin{vmatrix} l_{11} & l_{12} & l_{13} & \dots & l_{1n} \\ l_{21} & l_{22} & l_{23} & \dots & h_{2n} \\ \dots\dots\dots\dots\dots\dots\dots \\ l_{\mu 1} & l_{\mu 2} & l_{\mu 3} & \dots & l_{\mu n} \end{vmatrix},$$

що характеризується двома важливими властивостями:

- один із стовпців містить тільки одиничні елементи;
- всі інші стовпці містять не більше ніж по одному одиничному елементу.

Матриця L визначає μ перевірок на парність для розряду, що відповідає одиничному стовпцю. Додавши до цієї сукупності перевірок тривіальну перевірку $x_i = x_i$, одержимо $\mu + 1$ незалежних перевірних співвідношень для одного розряду x_i, причому властивості матриці L такі, що кожен розряд кодової комбінації входить тільки в одну перевірку. Така сукупність перевірок називається *системою поділених (ортогональних) перевірок* щодо розряду x_i.

Мажоритарне декодування здійснюється так. Якщо в прийнятій кодовій комбінації помилки відсутні, то при визначенні розряду x усі μ + 1 перевірки вкажуть одне й те саме значення (або 1, або 0). Одинична помилка в кодовій комбінації може спричинити спотворення лише однієї перевірки, подвійна помилка – двох і т. д. Значення розряду x_i вибирають за більшістю (тобто мажоритарно) однойменних результатів перевірок. При цьому декодування безпомилкове, якщо число помилок у кодовій комбінації не перевищує μ/2, тобто спотворено не більше ніж μ/2 перевірок. Якщо всі системи поділених перевірок для кожного розряду кодової комбінації містять не менше ніж $\mu + 1$ поділених перевірок, то реалізована мінімальна кодова відстань $d_0 = \mu + 1$.

Зазначимо, що далеко не всі коди допускають мажоритарне декодування, оскільки вимоги до структури перевірної матриці для такого декодування задовольняють не всі коди.

Пояснимо принцип мажоритарного декодування на конкретному прикладі. Нехай необхідно побудувати декодувальний пристрій для циклічного (15, 7)-коду Боуза–Чоудхурі–Хоквінгема, що допускає мажоритарне декодування. Код має твірний поліном

$P(x) = x^8 + x^7 + x^6 + \quad + x^4 + x \approx 111010001.$ Мінімальна кодова відстань $d_0 = 5$, і код може виправити дворазову помилку.

Обчислимо перевірний поліном:

$$h(x) = \frac{x^n + 1}{p^{-1}(x)} = x^7 + x^3 + x + 1 \approx 10001011.$$

Побудуємо перевірну матрицю. При цьому як перший рядок використаємо перевірний поліном, помножений на x^{r-1}, а інші рядки одержимо циклічним зсувом першого:

$$H = \begin{vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{vmatrix}.$$

Перетворимо перевірну матрицю так. Додамо 1, 5, 7 і 8-й рядки матриці:

$$1 \oplus 5 \oplus 7 \oplus 8 = 100000001000101.$$

Аналогічно

$$2 \oplus 3 \oplus 6 \oplus 7 \oplus 8 = 011000000010001,$$

$$4 \oplus 6 \oplus 7 \oplus 8 = 000101100000001.$$

Складемо матрицю L, використавши для її побудови три отримані суми і 8-й рядок перевірної матриці H:

$$L = \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}.$$

Як бачимо, в цій матриці один із стовпців складається тільки з одиниць, а всі інші стовпці містять не більше однієї одиниці. Матриця L дає чотири незалежних перевірних співвідношення з поділеними щодо члена a_0 перевірками:

$$a_0 = a_1 \oplus a_3 \oplus a_7;$$

$$a_0 = a_2 \oplus a_6 \oplus a_{14};$$

$$a_0 = a_4 \oplus a_{12} \oplus a_{13};$$

$$a_0 = a_8 \oplus a_9 \oplus a_{11}.$$

Додавши до цих співвідношень тривіальну перевірку $a_0 = a_0$, одержимо систему поділених відносно a_0 перевірок:

$$a_0 = a_0;$$

$$a_0 = a_2 \oplus a_6 \oplus a_{14};$$

$$a_0 = a_4 \oplus a_{12} \oplus a_{13};$$

$$a_0 = a_8 \oplus a_9 \oplus a_{11};$$

$$a_0 = a_1 \oplus a_3 \oplus a_7.$$

Нехай при передаванні спотворено розряд a_6, який входить тільки в другу перевірку, тому чотири перевірки дадуть правильний результат, а друга перевірка – неправильний. Значення розряду не вибирають за критерієм більшості і тому воно буде правильним. Помилкова реєстрація розряду відбудеться при дії трьох і більше помилок, що спричинюють неправильні результати трьох і більше перевірок.

Ми одержали систему поділених перевірок щодо розряду a_0. Системи поділених перевірок для інших розрядів отримують циклічним зсувом рядків матриці L. Зробивши, наприклад, зсув на один розряд, одержимо

$$L' = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix},$$

звідки випливає

$$a_{14} = a_0 \oplus a_2 \oplus a_6;$$
$$a_{14} = a_1 \oplus a_5 \oplus a_{13};$$
$$a_{14} = a_3 \oplus a_{11} \oplus a_{12};$$
$$a_{14} = a_7 \oplus a_8 \oplus a_{10}.$$

Крім того, $a_{14} = a_{14}$.

Аналогічно можна одержати системи поділених перевірок і для всіх інших розрядів кодової комбінації. Крім того, системи перевірок для циклічних кодів можна отримати циклічним зсувом якої-небудь однієї системи перевірок.

Схема мажоритарного декодера циклічного (15, 7)-коду показана на рис. 10. Декодер складається з *n*-розрядного регістра зсуву, набору двовхідних суматорів за модулем 2 і так званих мажоритарних органів *M*, завдяки яким вибирають значення декодованого розряду відповідно до критерію більшості. Підключають суматори до чарунок регістра зсуву згідно з системою поділених щодо розряду a_0 перевірок.

Діє декодер так. На виходах суматорів формуються результати поділених перевірок щодо розряду a_0, і мажоритарний орган вибирає значення розряду a_0. Далі в регістр подається ще один тактовий імпульс, і мажоритарний орган вибирає значення розряду $a_1 a_1$, і т. д. до декодування розряду a_{14}. Отже, декодування кодової комбінації здійснюється за 2^n такти: протягом перших тактів заповнюється регістр зсуву, а протягом наступних визначається кожний розряд.

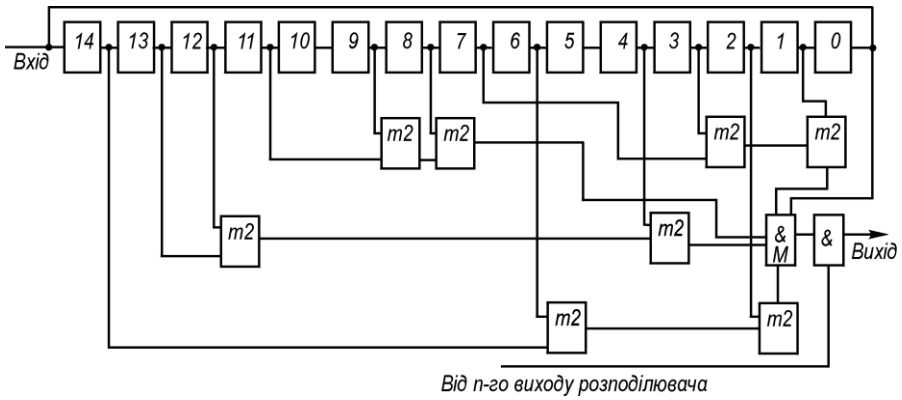

Рис. 10. Побудова мажоритарного декодера циклічного (15, 7)-коду з подільними перевірками

Дотепер ми розглядали тільки двійкові лінійні коди. Однак це робилося лише для простоти. Насправді теорія лінійних кодів звичайно викладається відразу для *m*-кових кодів, де $m = p^l$ (p – просте число; l – натуральне число), тобто для випадку, коли символи коду утворять кінцеве поле Галуа $GF(q)$ і над ними можуть бути здійснені всі арифметичні дії, що існують над дійсними чи комплексними числами. Для таких *m*-кових кодів можуть бути визначені всі поняття і доведені всі властивості, отримані раніше для двійкових кодів, а саме – породжувальна і перевірна матриці, систематичні та дуальні коди, границі мінімальних відстаней, стандартні розташування, синдроми, вагові спектри і границі для ймовірностей невиявлених і виявлених помилок, циклічні і БЧХ-коди, алгебраїчні і мажоритарні алгоритми декодування. Найбільш важливим класом *m*-кових кодів є коди Ріда–Соломона (скорочено РС-коди). Вони можуть бути побудовані як систематичні циклічні (n, k)-коди при

$n = q-1, n-k = 2t,$ де t – число помилок, що виправляються. Коди РС є частиною стандарту цифрового запису на компакт-диски.

За означенням вектор **x** є словом *m*-кового (*n*, *k*)-коду РС, якщо відповідний йому многочлен $f_x(D)$ має корені, рівні елементам поля $GF(q): \alpha, \alpha^2, ..., \alpha^{n-k}$, де α – примітивний елемент цього поля. Породжувальний многочлен коду РС має вигляд

$$g(D) = (D-\alpha)(D-\alpha^2)...(D-\alpha^{n-k}). \quad (123)$$

Як бачимо з означення РС-коду, він є частковим випадком *m*-кових БЧХ-кодів, і, відповідно до доведеного раніше, мінімальна кодова відстань таких кодів буде в точності дорівнюти $d = n-k+1$.

Легко показати, що жодний лінійний систематичний *m*-ковий ($m \geq 2$) код не може мати $d > n - k + 1$. Дійсно, якщо вибрати значення *k* мінус одного інформаційного символу рівними нулю, то це дасть ненульове кодове слово вагою не більше, ніж $n - k + 1$, що за властивістю лінійного коду і визначає верхню границю для *d* як $n - k + 1$. Оскільки РС-код реалізує верхню границю для мінімальної кодової відстані, то він є оптимальним серед усіх *m*-кових (*n*, *k*)-кодів щодо виправлення і виявлення помилок гарантованої кратності.

Можна дати простий опис РС-коду в несистематичному зображенні. Тоді кодовий вектор визначається як

$$\mathbf{x} = [F(1), F(\alpha), F(\alpha^2), ..., F(\alpha^{q-2})], \quad (124)$$

де $F(\alpha) = b_0 + b_1 D + ... + b_{k-1} D^{k-1}$, b_0, b_1, ..., b_{k-1} – значення інформаційних *m*-кових символів.

Вибір довжини коду $n = d-1$ є досить сильним обмеженням, особливо при великому порядку поля. Тому будують так звані *скорочені* коди РС, що мають довільну довжину $n \leq d-1$. Їх можна одержати з повних РС-кодів, що мають довжину $n = d-1$, якщо покласти частину інформаційних символів рівними нулю і викинути їх з кодових блоків. Оскільки укорочення коду не може зменшити мінімальної кодової відстані, то (*n*, *k*)-код при $n \leq d-1$ буде, як і раніше, мати $d = n - k + 1$.

Коди РС, як окремий випадок БЧХ-кодів, мають алгебраїчний алгоритм виправлення помилок з поліноміальною складністю. Коди *m*-кові можуть бути використані разом із двійковими кодами для побудови каскадних кодів.

Ітеративні і каскадні коди. Потужні коди (тобто коди з довгими блоками і великою кодовою відстанню *d*) при порівняно простій процедурі декодування можна будувати, поєднуючи кілька коротких кодів. Так будується, наприклад, *ітеративний* код із двох лінійних систематичних кодів (n_1, k_1) і (n_2, k_2), табл. 10.2. Спочатку повідомлення кодується кодом 1-го ступеня (n_1, k_1). Нехай k_2 блоків коду 1-го ступеня записані у вигляді рядків матриці. Її стовпці містять по k_2 символів, які вважатимемо інформаційними для коду 2-го ступеня (n_2, k_2), і допишемо до них $n_2 - k_2$ перевірних символів. У результаті вийде блок (матриця $n_1 \times n_2$), що містить $n_1 n_2$ символів, з яких $k_1 k_2$ є інформаційними. Процес побудови коду можна продовжити в 3-му вимірі і т. д.

Таблиця 3. **Побудова ітеративного коду**

	Інформаційні символи 1-го ступеня	Перевірні символи 1-го ступеня
Інформаційні символи 2-го ступеня	$b_{1,1}, b_{1,2}, ..., b_{1,k1}$ $b_{2,1}, b_{2,2}, ..., b_{2,k2}$ ..... $b_{k2,1}, b_{k2,2}, ..., b_{k2,kk}$	$b_{1,k1+1}, b_{1,k1+2}, ..., b_{1,n1}$ $b_{2,k1+1}, b_{2,k1+2}, ..., b_{2,n1}$ ..... $b_{k2,k1+1}, b_{k2,k1+2}, ..., b_{k2,n1}$
Перевірні символи 2-го ступеня	$b_{k2+1,1}, b_{k2+1,2}, ..., b_{k2+1,k1}$ ..... $b_{n2,1}, b_{n2,2}, ..., b_{n2,k1}$	$b_{k2+1}, b_{k2+1,k1+1}, ..., b_{k2+1,n1}$ ..... $b_{n2,k1+1}, b_{n2,k1+2}, ..., b_{n2,n1}$

При декодуванні кожного блоку 1-го ступеня виявляють і виправляють помилки. Після того як прийнятий весь двовимірний блок, знову виправляють помилки і витирання, але вже по стовпцях, кодом 2-го ступеня, причому доводиться виправляти тільки ті помилки, що не були виправлені (чи були "виправлені" неправильно) кодом 1-го ступеня. Легко переконатися, що мінімальна кодова відстань для двовимірного ітеративного коду $d = d_1 d_2$, де d_1 і d_2 – відповідно мінімальні кодові відстані для кодів 1-го і 2-го ступенів.

Дуже ефективний різновид потужних кодів – *каскадні* коди. Двокаскадний код (р11ис.) будується так: спочатку k_1 двійкових символів джерела розглядаються як збільшений символ багатопозиційнного коду з основою $m = 2^{k_1}$. Потім до послідовності з k_2 таких збільшених символів додається $n_2 - k_2$ перевірних символів *m*-кового коду (кожен перевірний символ – це послідовність із k_1 двійкових символів). На цьому завершується утворення зовнішнього коду. Після цього формується внутрішній код з кодовою відстанню d_1: до кожних k_1 елементарних двійкових символів зовнішнього коду додається $n_1 - k_1$ перевірних двійкових символів. Результуюча кодова комбінація містить $n_1 n_2$ двійкових символів, з яких $k_1 k_2$ є інформаційними. Цим каскадний код схожий на ітеративний. Однак декодування каскадного коду виконується в такий спосіб: спочатку послідовно здійснюється декодування всіх блоків внутрішнього коду (з виявленням або виправленням помилок), потім декодується блок зовнішнього *m*-кового коду (n_2, k_2), причому виправляються помилки і стирання, що залишилися після декодування внутрішнього коду. Внутрішній код звичайно розрахований на виправлення одиничних помилок, зовнішній – на виправлення груп помилок, які є одиничними помилками в збільшених *m*-кових символах (рис. 11). Як зовнішній код звичайно використовується *m*-ковий код Ріда–Соломона (див. вище), який забезпечує найбільше можливе d_2 при заданих значеннях n_2 і k_2, якщо $n_2 < m$.

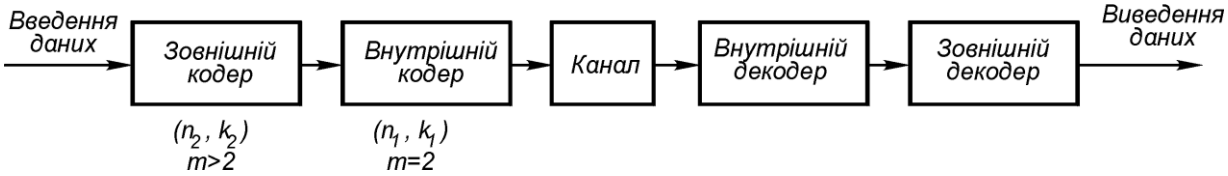

Рис. 11. Схема каскадного кодування та декодування

Побудований каскадний код еквівалентний лінійному двійковому коду з мінімальною відстанню $d \geq d_1 d_2$. Фактично розглянутий вище алгоритм декодування каскадного коду є більш ефективним і простим, ніж алгоритм декодування двійкового коду з еквівалентною мінімальною відстанню *d.*

Використання каскадних кодів у каналах без пам'яті дозволяє забезпечити експонентне спадання ймовірності помилкового декодування при збільшенні повної довжини блока $n = n_1 n_2$, при цьому швидкість передавання $R = \frac{k_1 k_2}{n_1 n_2}$ може бути як завгодно близькою до пропускної здатності каналу, а складність декодування поліноміально залежить від повної довжини блока *n.* Ще ефективнішим є застосування каскадних кодів у каналах з пам'яттю. Процес нарощування ступенів каскадного коду можна продовжити. Каскадні коди в багатьох випадках найперспективніші серед відомих блокових завадостійких кодів.

Кодування в каналах з пам'яттю. Хоча пам'ять і збільшує пропускну здатність каналів зв'язку, однак це не означає, що використання тих самих кодів у каналах з пам'яттю і без пам'яті дає меншу ймовірність помилкового декодування для каналів з пам'яттю. Більше того, навіть найкращі коди для каналів з пам'яттю можуть виявитися значно гіршими, ніж посередні коди в каналах без пам'яті.

Реальні фізичні канали зв'язку володіють, як правило, пам'яттю, що може пояснюватися кореляцією випадкових завад і параметрів каналів зв'язку. Типовими прикладами каналів з пам'яттю є канали з завмираннями. Універсальним методом зведення

каналів з пам'яттю до каналів без пам'яті є переміжність символів на передачі та їх деперемiжність на прийомі. Після переміжності символів можна використовувати алгоритми виправлення помилок для каналів без пам'яті. Платою за дану перевагу є затримки декодування, додаткова пам'ять, а також зменшення пропускної здатності каналу зв'язку. Остання обставина може бути "пом'якшена" переходом до алгоритму кодування-декодування з переміжністю і оцінкою ймовірності помилок. У цьому випадку пам'ять каналу використовується для оцінки ймовірностей помилок символів у кодових блоках, які потім враховуються при виправленні помилок.

Якщо пам'ять каналу проявляється в появі чітко виражених *пачок* помилок, тобто групи помилкових символів, що йдуть один за одним, то можна застосувати спеціальні коди, орієнтовані на виправлення помилок саме такої конфігурації, а не незалежних помилок. Ефективними в каналах з пам'яттю є й каскадні коди, якщо внутрішні коди в них використовуються в основному для виявлення пачок помилок і стирання блоків, на яких ці пачки виявлені, а зовнішній код використовується для виправлення стертих блоків.

Контрольні питання до Теми 8:

1. Як мажоритарне декодування використовується для виправлення помилок?
2. Як відрізняються двійкові і недвійкові коди за своїми властивостями?
3. Що таке недвійкові коди, і які приклади їх застосування?
4. Як мажоритарні алгоритми покращують точність передачі інформації?

Тема 9 Системи зі зворотним зв'язком

Ми розглянули системи, в яких передавання інформації здійснювалося в одному напрямку: від передавача до приймача. Але існують системи, де між кінцевими пунктами можливий двосторонній обмін інформацією, що, природно, припускає розміщення в кожному з них передавача і приймача. Зворотний канал зв'язку в таких системах може використовуватися для передавання не тільки звичайної інформації, але й спеціальних повідомлень, призначених для підвищення завадостійкості прямого каналу. Системи зв'язку, в яких по зворотному каналу передаються сигнали, що автоматично коректують помилки в прямому каналі, називаються *системами зі зворотним зв'язком*.

Коректування помилок може здійснюватися двома способами. Перший з них характеризується тим, що помилки виявляються на приймальному кінці. У разі виявлення помилок по зворотному каналу передається сигнал запиту для повторення спотвореного повідомлення. Якщо на передавальному кінці сигнал запиту прийнятий правильно, то поточне передавання переривається й автоматично повторюється помилково прийняте повідомлення. Така система зветься *системою з керуючим зворотним зв'язком* або *системою з автозапитом* (перезапитом).

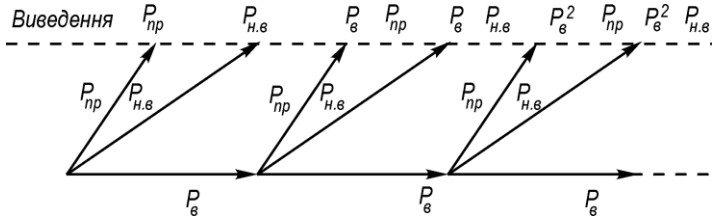

Рис. 12. Векторна діаграма станів системи з авто запитом

При іншому способі коректування помилок по зворотному каналу на передавач передаються дані про кожне прийняте від нього повідомлення,і виявлення помилок, на відміну від системи з автозапитом, відбувається на передавальному кінці. Якщо в результаті аналізу виявляється помилка, то також відбувається повторення спотвореного повідомлення. Ця система називається *системою з інформаційним зворотним зв'язком* чи *системою з порівнянням.*

Зупинимося спочатку більш докладно на системах з автозапитом. Одним з методів виявлення спотворень у прийнятому повідомленні є застосування коректувальних кодів, що виявляють помилки. Визначимо ймовірність помилкових комбінацій у системі з автозапитом і виявляючим кодом. При прийманні кодової комбінації можливі три несумісні події, сумарна ймовірність яких дорівнює одиниці:

$$P_{\text{пр}} + P_{\text{н.в}} + P_{\text{в}} = 1,$$

де $P_{\text{пр}}$ – імовірність правильного прийому; $P_{\text{н.в}}$ – імовірність помилок, що не виявляються даним кодом; $P_{\text{в}}$ – імовірність помилок, що виявляються.

У разі відсутності чи невиявлення помилок комбінація надходить на вихід системи. Якщо помилки виявлені, то прийнята комбінація стирається і посилається запит для її повторення. Тому при правильному прийманні сигналу запиту ймовірність повторення комбінації дорівнює $P_{\text{в}}$. У результаті вторинного передавання комбінації знову виникає первісне положення з трьома можливими наслідками, і т. д. Процес передавання окремої комбінації в такій системі можна зобразити графічно за допомогою векторної діаграми. На рис. 12 векторами показані ймовірні переходи системи з одного стану в інший при прийманні деякої комбінації.

Надалі вважатимемо, що в системі відсутнє обмеження кількості повторних передач однієї і тієї ж комбінації, а переходи системи в різні стани незалежні. За цих умов неважко визначити ймовірність помилкового переходу, що складається з l повторень комбінації, з наступним поданням її на вихід у випадку наявності невиявлених помилок. Як видно з рис. 12, ця ймовірність дорівнює $P_{\text{в}}^{l} P_{\text{н.в}}$. Повна ймовірність помилки дорівнює сумарній імовірності всіх помилкових переходів:

$$P_{\text{пом}} = P_{\text{н.в}} + P_{\text{в}} P_{\text{н.в}} + P_{\text{в}}^{2} P_{\text{н.в}} + \ldots = P_{\text{н.в}} \sum_{l=0}^{\infty} P_{\text{в}}^{l}.$$

Обчисливши за відомою формулою суму геометричної прогресії, одержимо

$$P_{\text{пом}} = \frac{P_{\text{н.в}}}{1 - P_{\text{в}}}. \qquad (124)$$

Виводячи цей вираз, ми не враховували ймовірність помилки в сигналі запиту, що в багатьох випадках допустимо, оскільки за такі сигнали вибираються спеціальні кодові послідовності, котрі мають високу завадостійкість.

Якщо $P_{\text{в}} \ll 1$, то $P_{\text{пом}} \approx P_{\text{н.в}}$, і застосування системи з автозапитом дасть підвищення завадостійкості, обумовлене величиною $P_{\text{н.в}}$.

У розглянутих системах передавання повідомлень відбувається з деякою надлишковістю, обумовленою надлишковістю самого коду і повторним передаванням комбінацій, що еквівалентно збільшенню їхньої тривалості. Визначимо величину цієї надлишковості. Звичайно при виявленні помилки приймач переходить у режим очікування запитуваної комбінації, тривалість якого $\tau_{\text{оч}}$ не менша за сумарну тривалість проходження сигналів в обидва кінці. Тоді при l повтореннях час, затрачуваний на приймання однієї комбінації, дорівнює $\tau_{\text{к}} + l\tau_{\text{оч}}$, де $\tau_{\text{к}}$ – номінальна тривалість комбінації. Ймовірність того, що комбінація надійшла на вихід після l повторень (див. рис. 12), визначається виразом

$$P_l = P_{\text{в}}^{l} \left(P_{\text{пр}} + P_{\text{н.в}} \right) = P_{\text{в}}^{l} \left(1 - P_{\text{в}} \right).$$

Звідси еквівалентна середня тривалість комбінації з урахуванням втрат на повторне передавання може бути зображена сумою

$$\overline{\tau}_{\text{к}} = \sum_{l=0}^{\infty} \left(\tau_{\text{к}} + l\tau_{\text{оч}} \right) P_l = \sum_{l=0}^{\infty} \left(\tau_k + l\tau_{\text{оч}} \right) P_{\text{в}}^{l} \left(1 - P_{\text{в}} \right) =$$

$$= \tau_{\text{к}} \left[1 + N \left(P_{\text{в}} + P_{\text{в}}^{2} + P_{\text{в}}^{3} + \ldots \right) \right],$$

де $N = \dfrac{\tau_{\text{оч}}}{\tau_{\text{к}}}$ – число додатково переданих комбінацій.

Використовуючи формулу суми геометричної прогресії, одержимо

$$\bar{\tau}_к = \tau_к \left[1 + N \left(\frac{1}{1-P_в} - 1 \right) \right] = \tau_к \frac{1+(N-1)P_в}{1-P_в}.$$

Отже, повна надлишковість у системі з автозапитом при використанні *n*-значного виявляючого коду, буде дорівнювати

$$x = 1 - \frac{k}{n\bar{\tau}_к / \tau_к} = 1 - \frac{k(1-P_в)}{n\left[1+(N-1)P_в\right]},$$

де k – значність первинного коду; $n\bar{\tau}_к / \tau_к$ – еквівалентна значність виявляючого коду при наявності втрат, завданих повторним передаванням.

Розглянемо тепер характеристики систем з інформаційним зворотним зв'язком. У найпростішому випадку прийняті приймачем повідомлення повністю передаються по зворотному каналу на передавач, де відбувається порівняння з переданими повідомленнями. За наявності розбіжностей спотворені повідомлення повторюються. Цей різновид інформаційного зворотного зв'язку іноді називають *ретрансляційним зворотним зв'язком*. У подібній системі помилка не буде виявлена, якщо вона з'явиться в одному і тому ж символі двічі: при передаванні по прямому і зворотному каналах. Якщо ймовірності помилок у цих каналах відповідно дорівнюють P_0 і P_0^*, то можна показати, що ймовірність невиявленої помилки в одному символі при $P_0 \ll 1$ і $P_0^* \ll 1$ приблизно дорівнює їхньому добутку $P_0 P_0^*$. Оскільки P_0^* звичайно величина досить мала, виграш у завадостійкості може бути значним.

Перевага ретрансляційного зворотного зв'язку полягає в достатньо простій конструкції. Однак для ретрансляції необхідно повністю займати зворотний канал.

У більш складних системах інформаційного зворотного зв'язку застосовуються коди, що виявляють помилки. Відмінність від систем з автозапитом полягає в тому, що по прямому каналу передається тільки та частина кодової комбінації, що містить інформаційні символи, а по зворотному каналу передаються відповідні їм контрольні символи. На передавальному кінці зіставляються передані інформаційні та прийняті контрольні символи. Ті комбінації, в яких порушується відповідність між цими двома групами символів, вважаються помилковими, і вони знову повторюються. Підбираючи коди з різною надлишковістю і коректувальною здатністю, тут легше, ніж у ретрансляційній системі, забезпечити найкращі умови передавання повідомлень. Аналіз показує, що застосування кодів з надлишковістю, рівною 0,5, тобто кодів, контрольні елементи яких також повністю займають зворотний канал, дає можливість одержати більш високу завадостійкість, ніж у ретрансляційній системі.

Якщо ймовірності помилок символів у прямому і зворотному каналах однакові, повна ймовірність помилки кодової комбінації при інформаційному зворотному зв'язку виражається тією ж формулою, що й у системі з автозапитом. Що стосується надлишковості (з урахуванням прямої і зворотної передач), ці системи також рівноцінні. Проте у випадку, коли ймовірність помилки в зворотному каналі значно менша, ніж у прямому, системи з інформаційним зворотним зв'язком мають більшу завадостійкість. У межі при нескінченно великому відношенні сигнал–завада в зворотному каналі $\left(P^*_{пом} = 0\right)$ і використанні кодів з достатньою надлишковістю інформаційний зв'язок може забезпечити безпомилкове передавання в прямому каналі. У цьому легко переконатися на прикладі ретрансляційного зв'язку. Дійсно, при $P^*_{пом} = 0$ всі помилки в прямому каналі будуть виявлені, а отже, й виправлені.

Контрольні питання до Теми 9:

1. Що таке системи зі зворотним зв'язком у теорії передачі інформації?
2. Які основні типи систем зі зворотним зв'язком існують?
3. Як впливає зворотний зв'язок на швидкість і точність передачі інформації?
4. Що таке адаптивні системи з використанням зворотного зв'язку?

ТЕМА 10 ПОЄДНАННЯ ПРОЦЕДУР ДЕМОДУЛЯЦІЇ І ДЕКОДУВАННЯ. ЗГОРТКОВІ (ҐРАТЧАСТІ) КОДИ

Вище були розглянуті двохетапні процедури визначення на прийомі переданого кодового слова. На першому етапі в схемі, яку можна назвати першою розв'язувальною (1РС), визначався вид переданого елемента (символу) кодової комбінації, а на другому етапі в другій розв'язувальній схемі (2РС) за послідовністю $\hat{a}_1, \hat{a}_2, ..., \hat{a}_n$ визначалося кодове слово. Друга розв'язувальна схема при двохетапній процедурі винесення рішення називається також *жорстким* (*дискретним*) *декодером*, а найчастіше просто декодером. Двохетапне декодування називають *поелементним* (*посимвольним*).

Дискретний декодер виносить рішення про передане кодове слово тільки за послідовністю "0" і "1", не використовуючи інформацію про якість прийому кожного з елементів. Це, зрозуміло, не сприяє одержанню максимальної правильності рішення і, крім того, обмежує можливості декодера. Для пояснення даної тези розглянемо декодування для коду (7, 6), який при поелементному прийманні здатний тільки виявляти помилки. Якщо декодер матиме інформацію про якість прийому кожного з елементів, то можна і виправляти одиничні помилки. Якість (надійність) прийому *i*-го елемента визначається умовною ймовірністю його неправильного прийому $p_i(H/Y)$. Чим більше $p_i(H/Y)$, тим гірша якість прийнятого елемента (символу). Припустимо, що в послідовності двійкових елементів $\hat{a}_1, \hat{a}_2, ..., \hat{a}_7$ мала місце однократна помилка, яку необхідно виправити. Найбільш імовірно, що місцеположення помилки відповідає місцеположенню найменш надійного символу. Змінюючи його на протилежний, зробимо виправлення одиничної помилки. Декодування, коли використовується інформація про надійність прийнятих символів, називається *м'яким* або *аналоговим*. Чим повніша інформація про надійність символів на декодері, тим менша ймовірність неправильного декодування. Найбільш повне використання інформації про надійність прийнятих символів відповідає *прийому в цілому*. Свою назву такий метод прийому одержав унаслідок розгляду тракту приймання як єдиної розв'язувальної схеми, що виносить рішення про передане кодове слово на підставі аналізу сигналу $Y(t) = \{y_1(t), y_2(t), ..., y_n(t)\}$ без попереднього визначення виду переданих елементів, тобто аналізу всього сигналу в цілому, з використанням всієї інформації про сигнал на вході приймача $Y(t)$. Реалізувати прийом у цілому в умовах гауссових завад і дискретного каналу без пам'яті можна, використовуючи схему, аналогічну схемі оптимального приймача. Ця схема дозволяє здійснити розділення двох сигналів, що відповідають символам "0" і "1". У нашому випадку число сигналів, які потрібно розділити, дорівнює 2^k, де k – число інформаційних елементів у кодовому слові. Якщо $k = 20$, то треба розділити 1 048 576 сигналів і, отже, таке ж число гілок прийому. Реалізувати такий приймач, навіть на сучасній елементній базі, неможливо.

Застосування прийому в цілому порівняно з поелементним прийомом по гауссовому каналу без пам'яті дає енергетичний виграш приблизно в 3 дБ при фіксованій імовірності помилкового декодування. Для каналів зі змінними параметрами і пам'яттю цей виграш буде значно більшим і становитиме вже 10...20 дБ.

Складність прийому в цілому змушує шукати такі двохетапні процедури винесення рішення про передане кодове слово, які забезпечують більш просту реалізацію тракту приймання, ніж при прийомі в цілому, і в той же час використовують інформацію про надійність прийнятих елементів, тобто тією чи іншою мірою поєднують процедури демодуляції і декодування. Залежно від ступеня використання інформації про надійність елементів буде забезпечене більше чи менше наближення за завадостійкістю до прийому в цілому.

Розрізняють два підходи до проблеми м'якого декодування. При першому мінімізується ймовірність неправильного декодування кодового слова, при другому мінімізується середня ймовірність помилки на елемент. В останньому випадку декодована послідовність може і не бути кодовим словом (дозволеною кодовою комбінацією). Характеристики декодерів, що реалізують ці два правила, збігаються, оскільки мінімізація ймовірності помилки на елемент веде до мінімізації ймовірності помилки в послідовності елементів, і навпаки.

Найбільш часто на практиці розв'язується задача мінімізації ймовірності неправильного декодування. При цьому враховуються обмеження на складність реалізації приймального тракту. В усіх цих випадках робиться спроба породити відносно невелике число кодових слів, серед яких з великою ймовірністю міститься передане кодове слово, що знаходиться на найменшій відстані від прийнятої послідовності елементів (символів). Методи мінімізації ймовірності неправильного декодування розрізняються способом породження цих кодових слів. Нижче розглядаються деякі з них.

Приймання за найбільш надійними символами. Ідея методу декодування за найбільш надійними символами (елементами) запропонована Л. Ф. Бородіним і може бути пояснена в такий спосіб. Визначимо вид двійкових символів $a_1, a_2, \ldots a_n$ та їх надійність γ_i, $i = \overline{1, N}$. Нехай найменш надійні символи розташовані на перевірних позиціях. Зітремо ці символи. Тоді, якщо інші k символів, що залишилися, прийняті без помилок, то послідовність $\hat{a}_1, \hat{a}_2, ..., \hat{a}_k$ буде правильно декодована, тобто буде виправлено r стирань. Відомо, що будь-який код здатний виправити до $(d_0 - 1)$ стирань, якщо на нестертих позиціях не було помилок. Звідси випливає, що число символів n_1, за якими може бути винесене рішення про передане кодове слово, визначається нерівністю $k \le n_1 \le n - (d_0 - 1)$. На кодувальний пристрій надходить послідовність одиниць і нулів, $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n$ а також інформація про надійність кожного із символів прийнятої послідовності .

Декодування починається з виділення k найбільш надійних символів і перевірки можливості однозначного декодування за цими символами. Якщо однозначне декодування неможливе, то додається ще один найбільш надійний з $n - k$ символів, що залишилися, і всі операції повторюються. При прийманні за методом Бородіна можуть бути виправлені всі помилки (стирання) кратності $t \le d_0 - 1$, якщо кожний з $n - d_0 + 1$ безпомилково прийнятих символів має міру надійності більшу, ніж у кожного з неправильно прийнятих.

Для гауссових каналів метод Бородіна забезпечує в порівнянні з поелементним прийманням виграш приблизно в 1,4 дБ, а для коду з $k = n - 1$ таку ж саму завадостійкість, як і метод прийому в цілому.

Більш складними в порівнянні з методом Бородіна є алгоритми декодування, запропоновані Чейзом.

Клас алгоритмів Чейза. Запропонована Чейзом процедура полягає в наступному.

1. Визначається значення кожного елемента (символу), що входить у кодову комбінацію. В результаті цього одержуємо вектор $\hat{\mathbf{A}} = \{\hat{a}_1, \hat{a}_2, ..., \hat{a}_n\}$, який складається з 0 і 1.

2. На прийняту кодову комбінацію $\hat{\mathbf{A}}$ накладаються сформовані певним чином тестові послідовності **T.** Отримані послідовності $\hat{\mathbf{A}}^* = \hat{\mathbf{A}} \oplus \mathbf{T}$ декодуються жорстким (дискретним) декодером за раніше розглянутими правилами. У результаті одержуємо кодові слова $\hat{\mathbf{A}}'_1, \hat{\mathbf{A}}'_2, ..., \hat{\mathbf{A}}'_j, ..., \hat{\mathbf{A}}'_N$, де N – число тестових послідовностей.

3. З урахуванням вірогідності (надійності) кожного з елементів визначаємо відстань послідовності $\hat{\mathbf{A}}$ до кожного з кодових слів $\hat{\mathbf{A}}'_i$, де $i = 1, 2, \ldots, N$. Прийнятим вважається те кодове слово, відстань до якого буде мінімальною. Відстань L_i до j-го кодового слова можна обчислити за формулою

$$L_j = \sum_{i=1}^{n} \gamma_i (\hat{a}_i \oplus \hat{a}'_{ij}); \quad j = \overline{1,\ N}, \tag{126}$$

де γ_i – додатне число, що характеризує ступінь надійності прийнятого *i*-го елемента. Наприклад, при восьмирівневому квантуванні сигналів на виході демодулятора це можуть бути числа 0...7; $\hat{a}'_{ij}$ – *i*-й символ *j*-го кодового слова $\hat{\mathbf{A}}'_j$; послідовність $\hat{a}_i \oplus \hat{a}_{ij}$, $i = \overline{1, n}$ визначає вектор помилки $\mathbf{E}'_j$.

На схемі, що реалізує декодування за Чейзом (рис. 13), подвійними лініями зображені векторні зв'язки. На вхід схеми з виходу неперервного каналу подається сигнал $V(t)$, який демодулюється; рівень сигналу на виході демодулятора визначає надійність прийнятих елементів. Вид елемента визначається в першій розв'язувальній схемі (1РС).

Три запропонованих Чейзом алгоритми відрізняються методом формування тестових послідовностей **Т.** Алгоритм 1 потребує формування найбільшого числа послідовностей і забезпечує найкращу завадостійкість. В алгоритмах 2 і 3 формується менше число тестових послідовностей і відповідно забезпечується гірша завадостійкість.

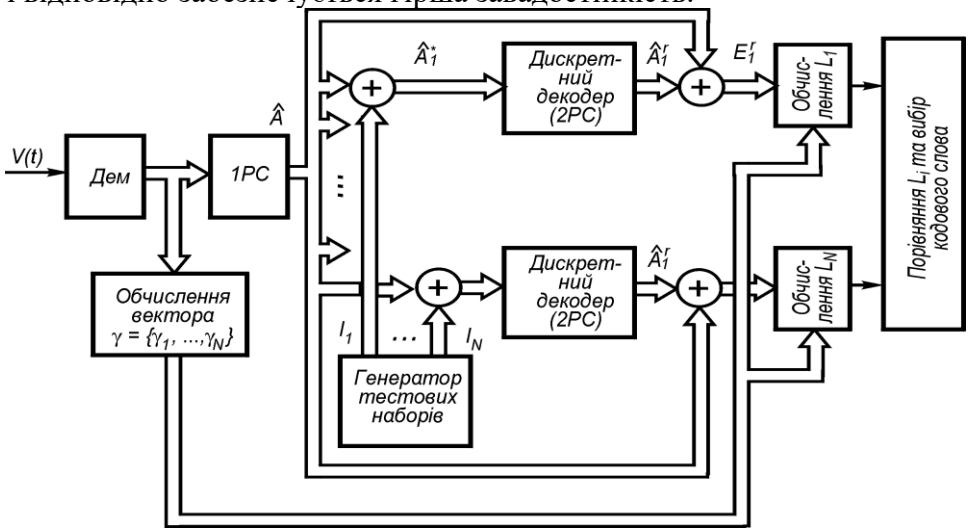

Рис. 13 Декодер Чейза

Для алгоритму 1 потрібно взяти як тестові послідовності нульовий вектор і всі комбінації ваги $[d_0/2]$; кількість таких тестових комбінацій $C_n^{[d_0/2]}$. Для алгоритму 2 використовується $2^{[d_0/2]}$ тестові послідовності, що мають різні символи на $[d_0/2]$ найменш надійних позиціях і нульові символи – на інших. Для алгоритму 3 визначається $d_0 - 1$ найменш надійних символів. Тестові послідовності мають одиниці на *i* найменш надійних позиціях і нулі на інших ($i = 0, 2, 4, \ldots, d_0 - 1$ для непарного d_0; $i = 0, 1, 3, \ldots, d_0 - 1$ для парного d_0). Загальне число можливих тестових послідовностей $[d_0/2+1]$.

Приклад 2.1. Визначити необхідне число тестових послідовностей для коду (7, 3) з кодовою відстанню $d_0 = 4$.

Для алгоритму 1 число тестових послідовностей $N = C_7^2 = 21$; для алгоритму 2 $N = 2^{[d_0/2]} = 2^2 = 4$; для алгоритму 3 $N = [d_0/2+1] = 2+1 = 3$. Отже, найменше число тестових послідовностей потрібно при використанні алгоритму 3.

Розглянемо на прикладі процедуру декодування за Чейзом для алгоритму 3.

Приклад 2.2. Код (6, 3) заданий твірною матрицею. Передано кодову комбінацію 000000. Внаслідок дії завад прийнято кодову комбінацію з помилками на позиціях 2 і 3 $\hat{\mathbf{A}} = 011000$. Оскільки $d_0 = 3$, то необхідно сформувати дві тестові послідовності, перша з яких 000000, а друга містить 2 одиниці на позиціях найменш надійних символів. Нехай надійність символів визначається числами 732656, тобто найменш надійними є символи, що розміщені на позиціях 2 і 3. Тоді друга тестова послідовність має вигляд 011000, і в результаті складання прийнятої послідовності з тестовими маємо

$$\begin{array}{r} 011000 \\ \oplus \\ 000000 \\ \hline 011000 = \hat{\mathbf{A}}_1^* \end{array} \qquad \begin{array}{r} 011000 \\ \oplus \\ 011000 \\ \hline 000000 = \hat{\mathbf{A}}_2^* \end{array}$$

Після декодування жорстким декодером одержимо кодові слова $\mathbf{A}_1' = 011100$ і $\mathbf{A}_2' = 000000$. Обчисливши L за формулою (126) для першого і другого кодових слів, одержимо відповідно $L_1 = 6$ і $L_2 = 5$. Таким чином, робиться висновок про те, що передавалося кодове слово 000000. У розглянутій вище ситуації вдалося виправити дворазову помилку.

При розробці систем передавання дискретних повідомлень доводиться розв'язувати не тільки задачі поєднання процедур демодуляції і декодування, але й модуляції та кодування, тому що характеристики дискретного каналу залежать від виду модуляції. Найбільш загальний підхід до розв'язання цих задач зводиться до того, що кодування і модуляція розглядаються як єдиний процес формування найкращого сигналу, а демодуляція і декодування – як процес найкращої обробки прийнятого сигналу.

Вище розглядалися блокові коди, коли значення елементів, що входять у різні блоки, не залежали одне від одного.

Для систематичних двійкових блокових (n, k)-кодів послідовність інформаційних символів джерела розбивається на блоки довжиною k біт, потім у кодері до кожного такого блока додається $r = n - k$ перевірних символів, після чого блоки довжиною n символів передаються в канал зв'язку. Декодування блоків також здійснюється незалежно одне від одного.

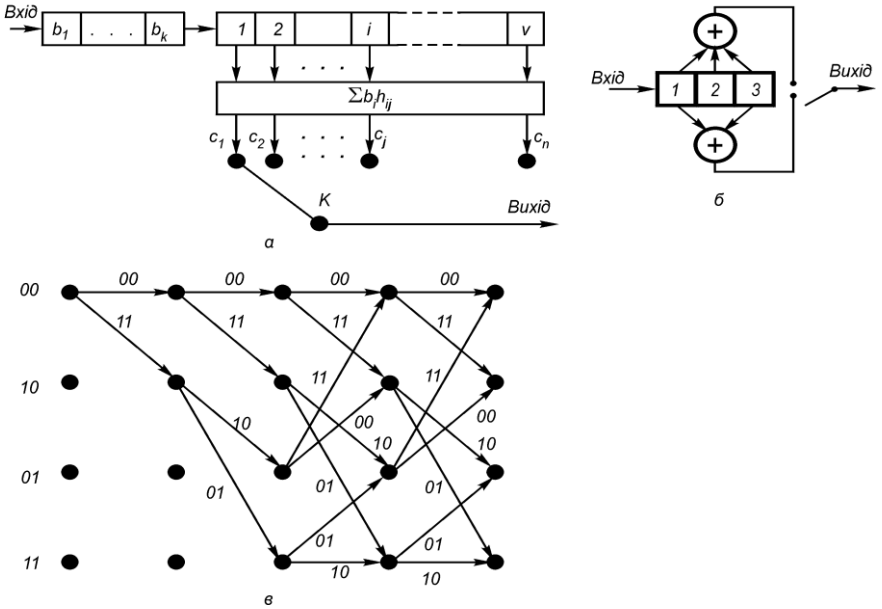

Рис. 14. Структури згорткових кодів зі швидкістю $R = k/n$ (*а*), зі швидкістю $R = 1/2$ (*б*) та ґратка кодера зі швидкістю $R = 1/2$ (*в*)

Однак можливий і інший, *неперервний* принцип кодування і декодування,

В 1967 р. з'явився алгоритм Вітербі для оптимального декодування згорткових кодів.

Структура двійкового згорткового коду зі швидкістю $R = \frac{k}{n}$ показана на рис. 2.7, *а*. Згортковий кодер складається зі зсувного регістра, що містить ν чарунок пам'яті, і блока суматорів за mod 2, входи кожного з яких зв'язані з деякими виходами чарунок пам'яті регістра, обумовленими коефіцієнтами $h_{ij} = (0, 1)$. Виходи суматорів зчитуються за допомогою комутатора K та подаються в канал зв'язку. Таким чином, на кожному такті в регістр зсуву послідовно надходить черговий блок з k інформаційних символів джерела й одночасно він звільняється від k символів, що містяться в його крайніх правих чарунках пам'яті. На цьому ж такті формуються n вихідних символів, що послідовно зчитуються в канал зв'язку. Отже, якщо ν_i – швидкість надходження символів у кодер, то для відсутності зростаючих затримок у часі

швидкість передавання символів по каналу зв'язку має бути не меншою, ніж $\nu_k = \frac{n}{k}\nu_i$, звідки випливає, що відношення *k/n* дійсно визначає швидкість згорткового коду. Величина ν (чи довжина регістра) звичайно називається *довжиною кодового обмеження.* У деяких роботах довжина кодового обмеження визначається числом чарунок зсувного регістру мінус 1.

Можливе й інше, більш загальне зображення згорткового кодера у вигляді схеми з *k* регістрами зсуву з кодовими обмеженнями ν_i ($i = 1, 2, ..., k$). На вхід кожного з них подається один інформаційний символ за час одного такту.

На рис. 14, *б* наведено окремий випадок згорткового коду зі швидкістю $R = 1/2$ і довжиною кодового обмеження $\nu = 3$. При нульовій інформаційній послідовності вихідна кодова послідовність також дорівнює нулю. Нижче розглянуто приклад формування вихідної послідовності для кодера, показаного на рис. 14, *б*:

Вхід

Вихід 1 0 1 0 1 1 1 1 0

Вихідна послідовність кодера може бути подана як цифрова згортка вхідної інформаційної послідовності й імпульсного відклику кодера (звідси назва кодів – *згорткові*).

Згортковий код характеризується такими параметрами: *відносною швидкістю* коду $R = k/n$ і *надлишковістю* $\iota = 1 - R$, де *k* та *n* – число інформаційних і кодових символів, що відповідають одному такту роботи кодера (для кодера на рис. 14, *б* *R* = 1/2); *довжиною кодового обмеження* ν (довжина регістра кодера); *породжувальним поліномом коду*, коефіцієнти яких описують зв'язки суматорів з чарунками регістра кодера (для верхнього суматора $g^{(1)} = 1 + D + D^2$, для нижнього суматора $g^{(2)} = 1 + D^2$). Поліноми звичайно записують скорочено, позначаючи кожні три відводи (двійкові коефіцієнти) як одну вісімкову цифру.

Крім названих параметрів згортковий код характеризується *вільною відстанню* $d_в$, під якою розуміють відстань за Хеммінгом між двома напівнескінченними кодовими послідовностями. Якщо дві однакові інформаційні послідовності кодувати за допомогою кодера, зображеного на рис. 14, *б*, то відповідні їм кодові послідовності збігатимуться одна з одною. Якщо в деякий момент в одній інформаційній послідовності опиниться символ 0, а в іншій 1, то з цього моменту кодові послідовності будуть відрізнятися одна від одної незалежно від подальшого змісту інформаційних послідовностей. Мінімальна відстань за Хеммінгом між будь-якими двома напівнескінченними кодовими послідовностями з того моменту, як відповідні їм інформаційні послідовності починають розрізнятися, називається вільною відстанню згорткового коду $d_в$.

Вільна відстань $d_в$ характеризує завадозахисні властивості згорткового коду (аналогічно тому, як мінімальна відстань *d* характеризує завадозахисні властивості блокових кодів). Вона показує, яке найменше число помилок має статися в каналі, щоб одна кодова послідовність перейшла в іншу і помилки не були виявлені. Для коду, наведеного в нашому прикладі, вільна відстань $d_в = 5$.

Пошук гарних згорткових кодів (з найбільшим $d_в$ при заданих *R* і ν) звичайно здійснюється методом перебору всіх породжувальних поліномів на ЕОМ.

Згорткові коди є окремим випадком (лінійною реалізацією) *ґратчастих* кодів. Можна також припустити, що ґратки є просто іншим (іноді більш зручним) способом подання і звичайних згорткових кодів.

Ґратками називається орієнтований граф з періодично повторюваною структурою "чарунок". Кожна чарунка містить стовпчики з однакового числа вершин (вузлів), з'єднаних ребрами. Між процедурою кодування згортковим кодом і ґратками існує взаємно однозначна відповідність, яка задається такими правилами:

кожна вершина (вузол) відповідає внутрішньому стану кодера;

ребро, що виходить з кожної вершини, відповідає одному з можливих символів джерела (для двійкового джерела з кожної вершини виходять два ребра – верхнє для 0 і нижнє для 1);

над кожним ребром проставлені значення символів, переданих у канал зв'язку, якщо кодер знаходився в стані, який відповідає даній вершині, і джерело видало символ, що відповідає даному ребру;

послідовність ребер (шлях на ґратках) – це послідовність символів, виданих джерелом.

Так, якщо під станом кодера розуміти вміст двох останніх чарунок пам'яті (2, 3) у регістрі зсуву на рис. 14, *б*, то ґратка з чотирма станами, що відповідає даному кодеру, матиме вигляд, показаний на рис. 14, *в* (ґратка може відображати і нелінійний кодер, коли вихідні символи не є лінійною функцією вхідних). Так само як і блокові коди, згорткові допускають зображення у вигляді напівнескінченних породжувальних чи перевірних матриць, однак зображення у вигляді ґратки є зручнішим для опису алгоритмів декодування.

Згорткові коди мають такі основні переваги над блоковими при їхньому використанні для виправлення помилок.

1. Вони не потребують синхронізації по блоках; необхідна лише синхронізація комутаторів *К* (на передачі та прийомі).

2. Якщо кодове обмеження ν вибрати рівним довжині блокового коду, то виправна здатність згорткового коду буде більшою, ніж виправна здатність такого блокового коду (при найкращому виборі обох кодів).

3. Алгоритм декодування згорткових кодів допускає просте узагальнення на випадок м'якого декодування, що забезпечує додатковий енергетичний виграш.

4. Згорткові коди допускають просте об'єднання кодування і модуляції (так звана *кодована модуляція* чи *сигнально-кодові конструкції*), що особливо важливо при побудові енергетично ефективних систем зв'язку для каналів з обмеженою смугою частот.

Для оптимального декодування згорткових кодів у каналах без пам'яті часто використовується рекурентний *алгоритм декодування Вітербі* (АВ). Розглянемо його на прикладі м'якого декодування в постійному каналі з адитивним білим гауссовим шумом.

Оскільки сигнал, що приймається на *k*-му тактовому інтервалі, нам відомий, то можна обчислити евклідові (або гільбертові) відстані між прийнятим сигналом і всіма можливими сигналами:

$$\Delta_{ki} = \int_0^T \left[z_k(t) - s_k^{(i)}(t) \right]^2 dt,\ i = 0,\ 1, ...,\ m-1,\ k = 0,\ 1, ...,$$

де $s_k^{(i)}(t)-$ очікуваний у місці прийому сигнал, що відповідає *i*-му символу (для двійкових сигналів $i = 0,\ 1$); $z_k(t)$ – сигнал, прийнятий на *k*-му тактовому інтервалі.

Тепер можна кожному ребру ґратки послідовно приписувати на *k*-х її ланках значення Δ_{ki}. Оптимальне за правилом максимальної правдоподібності декодування тоді відповідатиме вибору такого шляху на ґратках (тобто послідовності неперервно продовжуваних ребер), що $\sum_k \Delta_{ki}$ буде мінімальною. Здавалося б, для ґратки завдовжки *n* (тобто послідовності переданих символів завдовжки *n*) потрібно перебрати 2^n можливі варіанти, але насправді це не так. Ключовий момент АВ полягає в тому, що для кожної вершини на даному кроці (такті) існує безліч метрик, які відповідають з'єднаним з нею ребрами вершинам. На попередньому кроці можна залишити тільки одне ребро, що мінімізує суму метрик на всіх попередніх кроках.

Найпростіше можна пояснити даний алгоритм на такому прикладі. Нехай ґратки мають усього два стани і структуру, показану на рис. 15, *а*, де над ребрами поставлені відповідні метрики. Вважаємо, що перший інформаційний символ 0. Тоді шляхи, залишені (тобто такі, що "вижили") на різних кроках, показані на рис. 2.8, *б*. Видно, що на 4-му кроці одержуємо "вижившій" шлях, котрий в умовах наших позначень (орієнтація ребра вниз – 1, вгору – 0) відповідає інформаційній послідовності 0100.

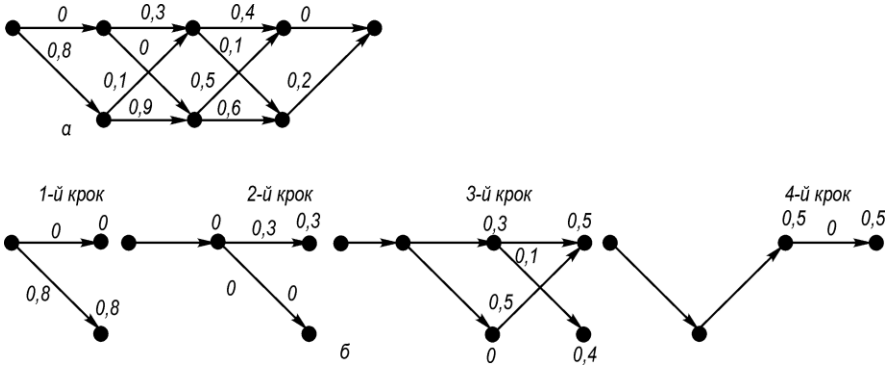

Рис. 15 Ґратка з метриками (*а*) та побудова шляху, що "вижив", за алгоритмом Вітербі (*б*)

Складність АВ визначається на кожному кроці числом порівнянь метрик, що з'єднують усі вершини, і воно обмежене величиною M^2, де M – число станів ґратки. Оскільки зі схеми згорткового кодера одержуємо, що $M = 2^{\nu-1}$, де ν – число чарунок пам'яті регістра зсуву кодера, то бачимо, що складність АВ експоненціально залежить від довжини кодових обмежень, але лінійно залежить від довжини послідовності, що передається. Тому довжина кодових обмежень ν при використанні АВ як алгоритму декодування звичайно вибирається не більшою за 10...15, що, утім, цілком достатньо для одержання великого енергетичного виграшу. АВ вимагає обробки всієї послідовності сигналів для оптимального декодування навіть першого інформаційного символу. Така процедура потребує значної пам'яті на прийомі та затримки для декодування елементів повідомлення. Для усунення цих недоліків використовується модифікація АВ у вигляді *зрізаного алгоритму,* коли рішення про інформаційний символ на *i*-му такті приймається за результатами обробки по АВ послідовності символів на даному *i*-му і L наступних тактових інтервалах. Теорія й експеримент показують, що коли L вибрати порядку декількох довжин кодових обмежень, то енергетичні втрати при використанні такої модифікації будуть невеликими.

Зазначимо, що АВ є ефективним методом розв'язання значно загальнішої оптимізаційної задачі. Нехай потрібно знайти такий дискретний вектор $\mathbf{x}_N = (x_1,...,x_N)$, $x_i \in \mathbf{X}$, $|\mathbf{X}| = r$, що максимізує (мінімізує) функцію $\Lambda(\mathbf{x}_N)$. Тоді прямий метод вимагає перебору r^N різних векторів. Однак, якщо функція $\Lambda(\mathbf{x}_N)$ допускає зображення

$$\Lambda(\mathbf{x}_N) = \sum_{k=1}^{N} \lambda_k(\sigma_k),$$

де $\lambda_k(\sigma_k)$ – довільні функції векторних аргументів вигляду

$$\sigma_k = \begin{cases} (x_{k-\nu}, x_{k-\nu+1},..., x_k), & k > \nu, \\ (x_1,..., x_k), & k \le \nu, \end{cases}$$

то можна довести, що максимум такої функції знаходиться за допомогою наступного *узагальненого* АВ.

Крок 1. Знайти $\hat{x}_1(x_2,..., x_{\nu+1}) = \operatorname{Arg}\max_{x_1} \Lambda_{\nu+1}$ (при $\hat{x}_{\nu+1}$). (Оскільки $|\mathbf{X}| = r$, то для кожного значення аргументу $(x_2,..., x_{\nu+1})$ це потребує не більше ніж *r* обчислень.)

Крок 2. Знайти $\hat{x}_2(x_3,..., x_{\nu+2}) = \operatorname{Arg}\max_{x_2} \Lambda_{\nu+2}$ (при $\hat{x}_{\nu+2}$), де $\hat{x}_{\nu+2} = = (\hat{x}_1, \hat{x}_2,..., x_{\nu+2})$, а $\hat{x}_1$ було знайдено на 1-му кроці, і т. д.

Крок *S*. Знайти $\hat{x}_S(x_{S+1},..., x_{\nu+S}) = \operatorname{Arg}\max_{x_S} \Lambda_{\nu+S}$ (при $\hat{x}_{\nu+S}$), де $\hat{x}_{\nu+S} = (\hat{x}_1, \hat{x}_2,..., x_{S-1}, x_S,..., x_{\nu+S})$, а $\hat{x}_1, \hat{x}_2,..., \hat{x}_{S-1}$ були знайдені на попередніх кроках, і т. д.

Крок $N - \nu$. Знайти $N - \nu$ $(\hat{x}_{N-\nu},..., \hat{x}_N) = \operatorname{Arg}\max_{(x_{N-\nu},..., x_N)} \Lambda_N$ (при $\hat{x}_N$), де $\hat{x}_N = (\hat{x}_1, \hat{x}_2,..., \hat{x}_{N-\nu-1}, x_{N-\nu},..., x_N)$, а $\hat{x}_1, \hat{x}_2,..., \hat{x}_{N-\nu-1}$ були знайдені на всіх попередніх кроках.

Таким чином, узагальнений АВ має поліноміальну залежність від довжини послідовності N на відміну від експоненціальної залежності при прямому методі перебору всіх альтернативних гіпотез.

Вибираючи різні форми функцій $\lambda_k(\sigma_k)$, можна одержати досить прості методи розв'язання оптимізаційних задач, а отже, і оптимальні алгоритми обробки сигналів не тільки для згорткових кодів, але й для інших моделей каналів, наприклад для каналу з міжсимвольною інтерференцією.

Універсальність АВ полягає в тому, що він може бути використаний для різних розподілів сигналів і завад, неоднорідних каналів, для поєднання декодування й демодуляції і не тільки для незалежних розподілів, але й для випадків залежності, описуваних марковськими послідовностями.

У каналах із МСІ *алгоритм Кловського–Ніколаєва* (АКН) зі зворотним зв'язком за рішенням при фіксованій затримці прийняття рішення $D = Q$ (Q – пам'ять каналу) практично не поступається за завадостійкістю АВ з тією ж затримкою рішення, проте реалізується простіше. АКН, отриманий спочатку для субоптимального поелементного приймання дискретних повідомлень в каналах з розсіюванням (МСІ), можна використовувати і для спільної демодуляції–декодування при згортковому кодуванні, тому що згортковий кодер є аналогом деякого каналу з розсіюванням. Природно, що в цьому випадку затримка в прийнятті рішення L має враховувати не тільки пам'ять каналу (Q), але і довжину кодового обмеження.

Згорткові коди можуть декодуватися й іншими алгоритмами (наприклад, послідовного декодування і синдромного декодування), що не є, взагалі кажучи, оптимальними.

Послідовне декодування було ведене Возенкрафтом, однак найбільш широко використовуваний алгоритм належить *Фано* [42].

У той час, коли відповідно до алгоритму Вітербі виконується просування і відновлення метрики для всіх шляхів, які можуть здаватися нам кращими, послідовний декодер істотно обмежує число шляхів, які фактично відновляються. Основна ідея послідовного декодування полягає в тому, що використовуватися має лише той шлях, який виглядає найбільш імовірним. Через обмеженість пошуку при декодуванні ніколи не можна бути цілком упевненим, що цей шлях є найкращим. Цей підхід може розглядатися як метод спроб і помилок для пошуку правильного шляху на кодовому дереві. Такий пошук здійснюється послідовно, так, що в кожен момент відбувається обробка лише одного шляху. Однак декодер має можливість повернутися назад при пошуку найкращого рішення.

Два найбільш часто використовуваних методи синдромного декодування – *декодування шляхом табличного пошуку* і *порогове декодування* – застосовуються при декодуванні як згорткових, так і блокових кодів. Про табличний метод синдромного декодування ми вже говорили. Порогове декодування – це метод, в основі якого лежить спеціальний вибір породжувальних многочленів коду, що допускають розв'язання перевірних рівнянь за допомогою мажоритарного вибору символу помилки за деякими оцінками. Останні отримуються додаванням одного чи кількох символів синдрому.

Для оцінки якості м'якого декодування за допомогою АВ скористаємося означенням: *мінімальною евклідовою відстанню d_e згорткового коду при обраному методі модуляції називається мінімальна сума евклідових метрик помилкових шляхів на гратках, що починаються і закінчуються на правильному шляху*.

Тоді для ймовірності помилки p_e в першому символі в детермінованому каналі без пам'яті з білим шумом при декодуванні по АВ буде слушна така границя:

$$p_e \le N(d_e)e^{-\frac{d_e^2}{4N_0}},$$

де $N(d_e)-$ число шляхів з евклідовою відстанню d_e, які починаються і закінчуються на правильному шляху; N_0 – спектральна щільність потужності білого шуму.

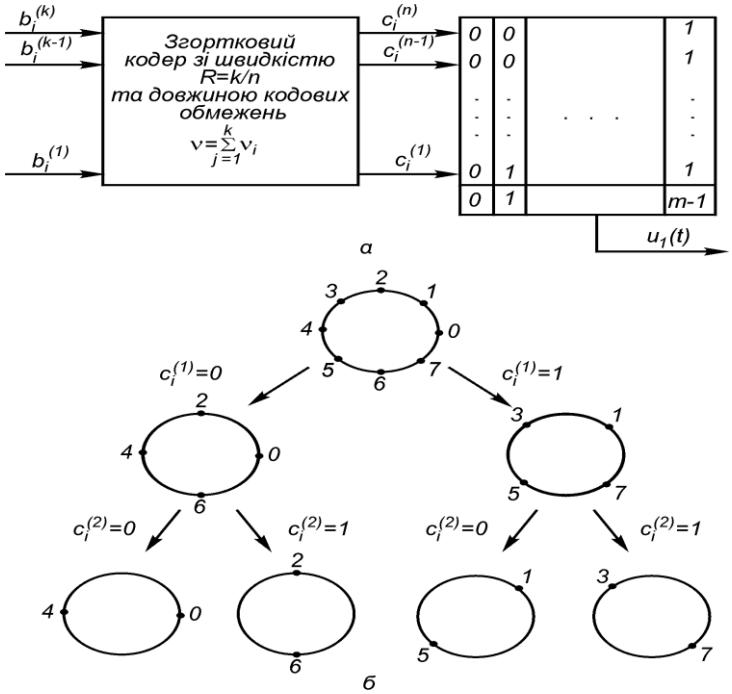

Рис. 16. Система з кодованою модуляцією (*а*) та розбиття вісімкових ФМ-сигналів на сузір'я (*б*)

Використовуючи більш повно структуру ґратки згорткового коду, можна оцінити також і середню ймовірність помилки на біт p_b, причому ця величина буде також експоненціально залежати від d_e^2.

Для одержання найбільшої енергетичної ефективності, особливо в каналах з пам'яттю, доцільно використовувати каскадні коди з внутрішніми згортковими кодами і м'яким декодуванням по АВ і зовнішніми РС-кодами з використанням алгебраїчних методів декодування. Така конструкція дозволяє одержати енергетичний виграш, що досягає 5 дБ при еквівалентній імовірності помилки $p_e = 10^{-5}$ і прийнятній складності декодування. Зазначимо, що АВ можна використовувати і для декодування блокових кодів, якщо ці коди можуть бути описані за допомогою ґраток.

Для того щоб одержати одночасно найкращу енергетичну і частотну ефективність, використовується *кодована модуляція*, або за іншою термінологією – певні *сигнально-кодові конструкції* (СКК). Загальну ідею такого методу ілюструє рис. 2.9, *а*, де індексом *i* позначається номер такту.

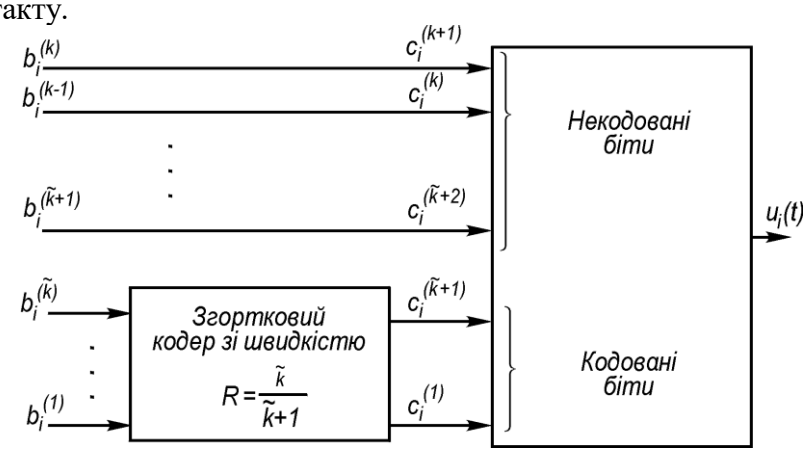

Рис. 17. Кодер Унгербоєка

В цьому випадку кодер згорткового коду зі швидкістю *k/n* зручніше подавати як паралельний набір *k* регістрів зсуву з різними довжинами кодових обмежень ν_i. На виході такого кодера у кожен момент часу з'являється *n*-вимірний двійковий вектор, який відображається в один з 2^n неперервних сигналів, що передаються в канал зв'язку. Для каналів з обмеженою смугою типовим є використання сигналів з багаторазовою фазовою або амплітудно-фазовою модуляцією, причому ключовим моментом тут є *розбиття множини сигналів на сузір'я.*

Приклад такого розбиття для вісімкової ФМ показаний на рис. 16, *б*. Сузір'я знаходяться в нижньому ряді рисунка. На рис. 17 наведена схема кодування Унгербоєка, в якій сигнали зображені у вигляді сузір'їв.

Тут двійкова послідовність символів джерела розбивається на блоки по *k* біт, і перші $\tilde{k}$ біт цих блоків подаються на вхід згорткового кодера, а ті, що залишилися, надходять на модулятор у некодованому вигляді. Загальний принцип кодування полягає в тому, що некодовані біти вибирають сигнал у сузір'ї, а кодовані біти визначають вибір сузір'я. (Наприклад, для схеми розбиття, показаної на рис. 16, *б*, кодер Унгербоєка, зображений на рис. 17, має параметри: $k = 2,\ \tilde{k} = 1,\ R = 1/2$).

Оскільки, як видно з рис. 16, *б*, сигнали в кожнім із сузір'їв віддалені на значну евклідову відстань, а мінімальна евклідова відстань між сигналами різних сузір'їв може бути мала, то використання розглянутого вище принципу дозволяє максимізувати евклідову відстань між послідовностями кодованих сигналів при збереженні високої спектральної ефективності.

Розроблено спеціальну техніку, що дозволяє об'єднати згорткові коди й амплітудно-фазову модуляцію сигналу для забезпечення високої енергетичної та спектральної ефективності.

При практичному використанні завадостійких кодів головним обмеженням є складність пристрою декодування, яка може бути виражена або числом логічних схем у декодері, або кількістю обчислювальних операцій, необхідних для декодування. Тому серед кодів, що забезпечують заданий виграш, слід вибирати ті, котрі допускають менш складну реалізацію, або навпаки, при заданій складності декодування слід вибирати коди, що забезпечують найбільший виграш.

Контрольні питання до Теми 10:

1. Як поєднуються процедури демодуляції і декодування?
2. Що таке згорткові (ґратчасті) коди, і як вони використовуються в телекомунікаціях?
3. Які алгоритми використовуються для декодування згорткових кодів?
4. Як поєднання демодуляції і декодування впливає на ефективність системи?
5. Які проблеми виникають при використанні згорткових кодів у реальних системах?

ПЕРЕЛІК ПОСИЛАНЬ

1. Болюх В. Ф., Данько В. Г. Основи електроніки і мікропроцесорної техніки: Навч. посібник. – Харків: НТУ «ХПІ», 2011. – 257 с.
2. Колонтаєвський Ю., Сосков А. Промислова електроніка та мікросхемотехніка, теорія і практикум : Навч. посібник. 2-ге вид. К. : Каравела, 2004. 432 с.
3. Ben G. Streetman, Sanjay Kumar Banerjee. Solid State Electronic Devices. Pearson College Div; 6th edition. 2005 - 581 p.
4. Aminghasem Safarian, Payam Heydari. Silicon-Based RF Front-Ends for Ultra Wideband Radios (Analog Circuits and Signal Processing). Springer. 2007 - 105 p.
5. Eugene I. Nefyodov, Sergey M. Smolskiy. Electromagnetic Fields and Waves: Microwave and mmWave Engineering with Generalized Macroscopic Electrodynamics. Springer. 2019 - 346 p.
6. Волочій Б.Ю.Передавання сигналів у інформаційних системах.Ч.1.:Нав. Посібник. – Львів: Видавництво Національного університету «Львівська політехніка», 2005. -196 с.
7. Зайцев Г. Ф., Стеклов В. К., Бріцький О. І. Теорія автоматичного управління. – К.: Техніка, 2002. – 688 с.
8. Огірко О. І., Галайко Н. В. Теорія ймовірностей та математична статистика: навчальний посібник / О. І. Огірко, Н. В. Галайко. – Львів: ЛьвДУВС, 2017. – 292 с.
9. Батаєв О.П., Ковтун І.В., Корольова Н.А. Теорія електричного зв’язку: Навч. посібник. Харків: УкрДАЗТ, 2010. - 630 с.
10. Стеклов В.К., Беркман Л.Н. Теорія електричного зв'язку. К.:Техніка. 2006. - 552 с.

НАВЧАЛЬНЕ ВИДАННЯ

Конспект лекцій з дисципліни «Теорія передавання інформації» для здобувачів вищої освіти другого (магістерського) рівня зі спеціальності - 172 «Електронні комунікації та радіотехніка». /Укл.: Литвиненко В.А., - Кам’янське; ДДТУ, 2024 р. – 77 с.

Укладачі: к.т.н., доц. Литвиненко В.А.

Підписано до друку 20.06. 2024 року.
Формат А5 Обсяг 3,5 др. арк.
Тираж 40 прим. Замовлення 360
51918 , м. Кам’янське, вул. Дніпробудівська, 2