

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ ТА ІНФОРМАТИЗАЦІЇ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОКОМУНІКАЦІЙ

Беркман Л.Н.
Жураковський Б.Ю.
Макаренко А.О.

Навчальний посібник

“ТЕОРІЯ ПЕРЕДАЧІ ТА ОБРОБКИ ДАНИХ В
ІНФОКОМУНІКАЦІЯХ”

Київ – 2015

ЗМІСТ

ПЕРЕДМОВА	6
РОЗДІЛ 1 ОПТИМАЛЬНЕ ПРИЙМАННЯ ДИСКРЕТНИХ СИГНАЛІВ. ОСНОВИ ТЕОРІЇ ЗАВАДОСТІЙКОСТІ	7
1.1. Загальні відомості про завадостійкість	7
1.2. Критерії оптимального приймання сигналів	8
1.3. Оптимальне приймання дискретних сигналів	11
1.4. Імовірність помилки при когерентному прийманні двійкових сигналів ...	15
1.5. Імовірність помилки при когерентному прийманні багатопозиційних сигналів	20
1.6. Некогерентне приймання дискретних сигналів.....	23
1.7. Передавання дискретних сигналів по каналах із змінними параметрами .	25
РОЗДІЛ 2 ОПТИМАЛЬНЕ ПРИЙМАННЯ НЕПЕРЕРВНИХ ПОВІДОМЛЕНЬ ОПТИМАЛЬНА ФІЛЬТРАЦІЯ	29
2.1. Оптимальне приймання неперервних повідомлень	29
2.2. Синтез пристроїв (систем) зв'язку за заданою структурною схемою.....	32
2.3. Фізичний зміст оптимальної смуги пропускання.....	33
2.4. Синтез оптимальних пристроїв (систем) зв'язку	34
2.5. Синтез оптимальних фільтрів Калмана методом простору станів	39
2.6. Приклади синтезу оптимальних фільтрів.....	47
2.7. Оптимальний фільтр для повідомлення, описуваного степеневим поліномом	54
2.8. Поняття про оптимальну нелінійну фільтрацію	58
РОЗДІЛ 3 ВИКОРИСТАННЯ ЗАВАДОСТІЙКОГО КОДУВАННЯ В КАНАЛАХ СИСТЕМИ УПРАВЛІННЯ ІНФОКОМУНІКАЦІЙНИОЮ МЕРЕЖЕЮ	62
3.1. Математична модель дискретного каналу	62
3.2. Аналіз моделей дискретного каналу.....	64
3.3. Інтенсивність помилок та характер їх розподілу у часі.....	68
3.4 Двійкові групові коди.....	75
3.5. Рекурентні коди.....	81
3.6. Недвійкові коди.....	82
3.7. Завдостійке кодування в каналах передачі	83
3.8 Оцінка ефективності використання коду Файра	90
3.9. Розроблення блок-схем кодування та декодування кодом Файра.....	92
3.10. Програмна реалізація кодера Файра	93
РОЗДІЛ 4 СТИСНЕННЯ ПЕРВИННИХ ПОВІДОМЛЕНЬ В ІНФОРМАЦІЙНИХ КАНАЛАХ СИСТЕМ УПРАВЛІННЯ	96
4.1. Характеристики джерел повідомлень в системах управління телекомунікаційними мережами	96

4.2. Оцінки ефективності методів стиснення.....	100
4.3. Класифікація методів стиснення.....	101
4.4. Способи стиснення даних при передачі.....	102
4.5. Класифікація алгоритмів стиснення при передачі та архівації.....	117
4.6. Кодо-матричне стиснення.....	130
4.7. Алгоритм повздовжнього та матричного стиснення.....	133
4.8. Комбінований метод підвищення ефективності передачі інформації.....	133
РОЗДІЛ 5 ВИКОРИСТАННЯ ШУМОПОДІБНИХ СИГНАЛІВ У СИСТЕМАХ ЗВ'ЯЗКУ.....	134
5.1. Основні поняття і визначення.....	134
5.2. Кодовий поділ абонентів.....	137
5.3. Основні структурні схеми широкосмугових систем зв'язку.....	142
5.4. Шумоподібні сигнали та їхні спектри.....	145
5.5. Оптимальна обробка сигналів.....	148
5.6. Кореляційні функції сигналів.....	150
5.7. Основні типи ШПС.....	154
СПИСОК ЛІТЕРАТУРИ, РЕКОМЕНДОВАНОЇ ДЛЯ ПОГЛИБЛЕНОГО ВИВЧЕННЯ ДИСЦИПЛІНИ.....	158

ПЕРЕДМОВА

Навчальний посібник призначений для вивчення дисципліни “Теорія передачі та обробки даних в інфокомунікаціях” (ТПОДІ) з циклу дисциплін професійної та практичної підготовки для студентів вищих навчальних закладів усіх форм навчання за напрямом 6.050102 Комп’ютерна інженерія.

Навчальний посібник відповідає програмі дисципліни “Теорія передачі та обробки даних в інфокомунікаціях”.

Посібник призначено для студентів, які прагнуть поглиблено вивчити процеси передавання, кодування та стиснення сигналів в інформаційних системах.

Предметом навчальної дисципліни є:

- основи сучасної теорії електрозв'язку 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 і завад та методи забезпечення граничних характеристик систем зв'язку як за достовірністю, так і за швидкістю передачі інформації;
- процеси передавання сигналів каналами зв'язку при наявності завад з математичної точки зору.

Метою вивчення навчальної дисципліни є:

- з'ясування фундаментальних понять інформаційної інфраструктури для спеціалістів з телекомунікацій;
- опанування основними термінами, категоріями, базовими знаннями із сучасної теорії електричного зв'язку, використання і оцінювання у своїй практичній діяльності математичних моделей процесів (у тому числі сигналів, каналів зв'язку) для розв'язання виробничих, проектних та наукових задач з телекомунікацій;
- здатність застосовувати правила, методи, принципи, закони у конкретних ситуаціях, своєчасно адаптуватися до зростаючого потоку інформації, проблем розвитку галузі зв'язку та новітніх науково-технічних досягнень в галузі телекомунікацій;
- сформувати у випускників активну позицію (за вимогами до сучасних спеціалістів), спрямовану на практичну реалізацію важливих завдань - інформатизації держави та входження до глобальної інфраструктури.

Завданнями навчальної дисципліни є формування наступних умінь:

- уміти характеризувати системи електрозв'язку, типові сигнали та завади, характеризувати та оцінювати основні характеристики сигналів, аналізувати переваги і недоліки конкретних видів модуляції, проводити розрахунки характеристик випадкових сигналів на виході каналів, чітко характеризувати процес дискретизації сигналів за теоремою Котельникова,
- уміти розуміти сутність та оцінювати значення, здійснювати класифікацію каналів електрозв'язку та характеризувати математичні моделі дискретних та неперервних каналів, диференціювати, інтегрувати, уніфікувати, оцінювати зміст теорії передачі інформації каналами телекомунікаційних мереж;
- уміти узгоджувати параметри джерела та каналу зв'язку, застосовувати завадостійкі коди та методи оптимальної демодуляції, реалізувати алгоритми оптимального прийому багатопозиційних сигналів багатоканальних модемів;
- уміти аналізувати завадостійкість систем електрозв'язку з різними видами модуляції та методами приймання, застосовувати теорію лінійного розподілу сигналів, проводити аналіз архітектури сучасних телекомунікаційних мереж, окремих технічних рішень на мережах зв'язку і прогнозувати очікувані результати.

РОЗДІЛ 1

ОПТИМАЛЬНЕ ПРИЙМАННЯ ДИСКРЕТНИХ СИГНАЛІВ. ОСНОВИ ТЕОРІЇ ЗАВАДОСТІЙКОСТІ

1.1. Загальні відомості про завадостійкість

Завадостійкістю системи зв'язку називається здатність системи розрізняти (відновлювати) сигнали з заданою вірогідністю.

Звичайно спосіб передавання (спосіб кодування і модуляції) заданий, і потрібно визначити завадостійкість, яку забезпечують різні способи приймання сигналів. Який з можливих способів приймання є оптимальним? Це питання є предметом розгляду теорії завадостійкості, основи якої розроблені академіком В. О. Котельниковим.

Задача визначення завадостійкості всієї системи в цілому дуже складна. Тому часто визначають завадостійкість окремих ланок системи: приймача при заданому способі передавання, системи кодування або системи модуляції при заданому способі приймання тощо.

Гранично досяжна завадостійкість називається, за Котельниковим, *потенційною завадостійкістю*. Порівняння потенційної та реальної завадостійкості пристрою дозволяє дати оцінку якості реального пристрою і знайти ще невикористані резерви. Знаючи, наприклад, потенційну завадостійкість приймача, можна судити, наскільки близька до неї реальна завадостійкість існуючих способів приймання і наскільки доцільне подальше удосконалення їх при заданому способі передавання. Відомості про потенційну завадостійкість приймача при різних способах передавання дозволяють порівняти їх між собою і визначити найбільш досконалі.

У разі відсутності завад кожному прийнятому сигналу x відповідає цілком визначений сигнал s . За наявності завад ця однозначна відповідність порушується. Завада, впливаючи на переданий сигнал, вносить невизначеність щодо того, яке з можливих повідомлень було передано; за прийнятим сигналом x тільки з деякою ймовірністю можна судити про те, що був переданий той чи інший сигнал s . Ця невизначеність описується апостеріорним розподілом імовірностей $P(s/x)$.

Якщо статистичні властивості сигналу s і завади ω відомі, то можна створити приймач, який на підставі аналізу сигналу x знаходитиме апостеріорний розподіл $P(s/x)$. Потім за видом цього розподілу оператором чи самим приймачем за правилом, що визначається заданим критерієм, приймається рішення про те, яке з можливих повідомлень було передано.

Завдання полягає в тому, щоб відтворити за заданим критерієм передане повідомлення щонайкраще. Такий приймач називається *оптимальним*; його завадостійкість буде максимальною при заданому способі передавання.

Незважаючи на випадковий характер сигналів x , у більшості випадків існує можливість виділити множину найбільш імовірних сигналів $\{x_i\}$, що відповідають передаванню деякого сигналу s_i . Геометричне зображення дозволяє множину сигналів замінити областю багатовимірного простору.

Нехай область X сигналів, що приймаються, розбита на неперекривні області X_j ($j=1, 2, \dots, m$), причому кожному сигналу s_i відповідає область X_i . Якщо прийнятий сигнал попав саме в цю область, то приймач вирішує, що передавався сигнал s_i . З деякою ймовірністю сигнал x_i може попасти в будь-яку іншу область X_j ($j \neq i$), і тоді приймається помилкове рішення: замість сигналу s_i відтворюється сигнал s_j . Імовірність того, що переданий сигнал прийнятий правильно, дорівнює $P(x_i/s_i)$, а ймовірність того, що він прийнятий помилково, дорівнює

$$1 - P(x_i / s_i) = \sum_{j \neq i} P(x_j / s_i). \quad (1.1)$$

Умовна ймовірність $P(x_j / s_i)$ залежить від способу формування сигналу, від завад в каналі і від вибраної розв'язувальної схеми приймача.

Повна ймовірність помилкового приймання елемента сигналу, очевидно, дорівнюватиме

$$P_{\text{п}} = \sum_{i=1}^m P(s_i) 1 - P(x_i / s_i), \quad (1.2)$$

де $P(s_i)$ – апіорні ймовірності сигналів, що передаються.

У випадку двійкового каналу область сигналів, які приймаються, розбивається на дві області: X_1 та X_2 . Якщо сигнал x попадає в область X_1 , то відтворюється сигнал s_1 , а якщо в область X_2 , то s_2 .

1.2. Критерії оптимального приймання сигналів

Для того щоб визначити, яка з розв'язувальних схем є оптимальною, необхідно насамперед установити, в якому сенсі розуміється оптимальність. Вибір критерію оптимальності не є універсальним, він залежить від поставленої задачі та умов роботи системи.

Нехай на вхід приймача надходить сума сигналу і завади $x(t) = s_k(t) + \omega(t)$, де $s_k(t)$ – сигнал, якому відповідає кодовий символ a_k ; $\omega(t)$ – адитивна завада з відомим законом розподілу. Сигнал s_k у місці приймання є випадковим з апіорним розподілом $P(s_k)$. На підставі аналізу коливання $x(t)$ приймач відтворює сигнал s_i . У разі наявності завад це відтворення не може бути абсолютно точним. За прийнятою реалізацією сигналу приймач обчислює апостеріорний розподіл $P(s_i / x)$, що містить усі відомості, які можна здобути з прийнятої реалізації сигналу $x(t)$. Тепер треба установити критерій, за яким приймач видаватиме на підставі апостеріорного розподілу $P(s_i / x)$ рішення щодо переданого сигналу s_k .

Для передавання дискретних повідомлень широко використовується **критерій В. О. Котельникова (критерій ідеального спостерігача)**. Відповідно до цього критерію приймається рішення, що передано сигнал s_i , для якого апостеріорна ймовірність $P(s_i / x)$ має найбільше значення, тобто реєструється сигнал s_i , якщо виконуються нерівності

$$P(s_i / x) > P(s_j / x), \quad j \neq i. \quad (1.3)$$

При використанні такого критерію повна ймовірність помилкового рішення буде мінімальною. Дійсно, якщо за сигналом x приймається рішення про те, що був переданий сигнал s_i , то ймовірність правильного рішення дорівнюватиме $P(s_i / x)$, а ймовірність помилки $1 - P(s_i / x)$. Звідси випливає, що максимуму апостеріорної ймовірності $P(s_i / x)$ відповідає мінімум повної ймовірності помилки (1.2).

На підставі формули Бейеса

$$P(s_i / x) = \frac{P(s_i)P(x / s_i)}{P(x)}. \quad (1.4)$$

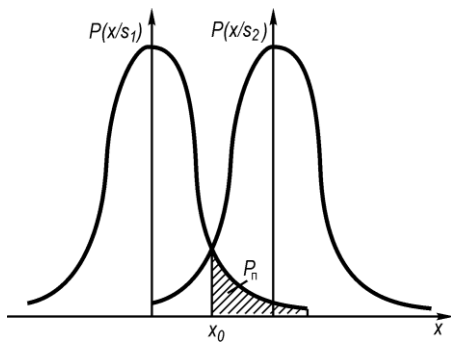


Рис. 1.1. Графік щільності умовного розподілу ймовірностей при передаванні сигналів s_1 і s_2

Тоді нерівність (1.3) можна записати в іншому вигляді:

$$P(s_i)P(x/s_i) > P(s_j)P(x/s_j)$$

Або

$$\frac{P(x/s_i)}{P(x/s_j)} > \frac{P(s_j)}{P(s_i)}$$

Функцію $P(x/s)$ часто називають *функцією правдоподібності*. Чим більше значення цієї функції при даній реалізації сигналу x , тим правдоподібніше, що передавався сигнал s . Відношення, що входить у нерівність (1.5),

$$\Lambda = \frac{P(x/s_i)}{P(x/s_j)} \quad (1.7)$$

називається *відношенням правдоподібності*. Користуючись цим поняттям, правило розв'язання (1.6), що відповідає критерію Котельникова, можна записати так:

$$\Lambda > \frac{P(s_j)}{P(s_i)} \quad (1.8)$$

Якщо передані сигнали рівноймовірні, тобто $P(s_i) = P(s_j) = \frac{1}{m}$, то це правило розв'язання набирає простішого вигляду:

$$\Lambda > 1 \quad (1.9)$$

Отже, критерій ідеального спостерігача зводиться до порівняння відношень правдоподібності (1.8). Цей критерій є більш загальним і називається *критерієм максимальної правдоподібності*.

Розглянемо бінарну систему, в якій передавання повідомлень здійснюється за допомогою сигналів $s_1(t)$ і $s_2(t)$, що відповідають кодовим символам a_1 і a_2 . Рішення приймається за результатом обробки прийнятого коливання $x(t)$ пороговим методом: реєструється s_1 , якщо $x < x_0$, і s_2 , якщо $x \geq x_0$, де x_0 – деякий пороговий рівень x . Тут можуть бути помилки двох видів: відтворюється сигнал s_1 , коли передавався сигнал s_2 , і сигнал s_2 , коли передавався сигнал s_1 . Умовні ймовірності цих помилок (ймовірності переходів) відповідно дорівнюють

$$P_{12} = P(s_1/s_2) = \int_{-\infty}^{x_0} P(x/s_2) dx, \quad (1.10)$$

$$P_{21} = P(s_2/s_1) = \int_{x_0}^{\infty} P(x/s_1) dx. \quad (1.11)$$

Значення інтегралів, які входять у (1.10) та (1.11), можуть бути обчислені як відповідні площі, обмежені графіком щільності умовного розподілу ймовірностей (рис. 1.1). Ймовірності помилок першого і другого виду відповідно

$$P_I = P(s_2)P(s_1 / s_2) = P_2 P_{12}, \quad (1.12)$$

$$P_{II} = P(s_1)P(s_2 / s_1) = P_1 P_{21}. \quad (1.13)$$

Повна ймовірність помилки при цьому

$$P_{\Pi} = P_I + P_{II} = P_2 P_{12} + P_1 P_{21}. \quad (1.14)$$

Нехай $P_1 = P_2$, тоді

$$P_{\Pi} = \frac{1}{2}(P_{12} + P_{21}).$$

Неважко переконатися, що в цьому випадку мінімум P_{Π} має місце при $P_{12} = P_{21}$, тобто при виборі порога відповідно до рис. 7.1. Для такого порога $P_{\Pi} = P_{12} = P_{21}$. На рис. 1.1 значення P_{Π} визначається заштрихованою площею. При будь-якому іншому значенні порога величина P_{Π} буде більшою.

Незважаючи на природність і простоту, критерій Котельникова має недоліки. Перший полягає в тому, що для побудови розв'язувальної схеми, як впливає зі співвідношення (1.5), необхідно знати апіорні ймовірності передавання різних символів коду. Другим недоліком цього критерію є те, що всі помилки вважаються однаково небажаними (мають однакову вагу). Проте в деяких випадках таке припущення не є правильним. Наприклад, при передаванні чисел помилка в перших значущих цифрах більш небезпечна, ніж помилка в останніх цифрах. Пропуск команди чи помилкова тривога в різних системах оповіщення можуть мати різні наслідки.

Отже, у загальному випадку, вибираючи критерій оптимального приймання, необхідно враховувати ті втрати, які несе одержувач повідомлення при різних видах помилок. Ці втрати можна виразити деякими *ваговими коефіцієнтами*, приписуваними кожному з помилкових рішень. Позначимо втрати помилкових рішень першого і другого видів відповідно L_{12} і L_{21} . Тоді можна визначити середні очікувані втрати, чи *середній ризик*

$$r = L_{12}P_I + L_{21}P_{II} = L_{12}P_2P_{12} + L_{21}P_1P_{21}. \quad (1.15)$$

Оптимальною розв'язувальною схемою буде та, що забезпечує мінімум середнього ризику. Критерій мінімального ризику належить до класу *бейєсових критеріїв*.

У радіолокації широко використовується *критерій Неймана–Пірсона*. При виборі цього критерію враховується, по-перше, що помилкова тривога і пропуск цілі не є рівноцінними за своїми наслідками, і, по-друге, що апіорна ймовірність сигналу, який передається, є невідомою. Якщо пропуск цілі є більш небажаним, то можна задати деяку величину β допустимої ймовірності помилкової тривоги і поставити вимогу, щоб розв'язувальна схема максимізувала ймовірність правильного виявлення $P_{\text{вияв}}$ (або, що те ж саме, мінімізувала ймовірність пропуску $P_{\text{пр}}$).

Відповідно до критерію Неймана–Пірсона приймач є оптимальним у тому випадку, якщо при заданій ймовірності помилкової тривоги

$$P_{\text{п.т}} = \int_{x_0}^{\infty} P(x/0) dx = \beta \quad (1.16)$$

він забезпечує найбільшу ймовірність правильного виявлення:

$$P_{\text{вияв}} = 1 - P_{\text{пр}} = 1 - \int_{-\infty}^{x_0} P(x/s) dx. \quad (1.17)$$

Можна показати, що критерій Неймана–Пірсона приводить до наступного правила розв'язку: ціль вважається виявленою, якщо

$$\Lambda = \frac{P(x/s)}{P(x/0)} > \lambda, \quad (1.18)$$

де λ – деяке число, зумовлене допустимою ймовірністю помилкової тривоги β .

1.3. Оптимальне приймання дискретних сигналів

Джерело дискретних повідомлень характеризується сукупністю можливих елементів повідомлення u_1, u_2, \dots, u_m та ймовірностями появи цих елементів на виході джерела $P(u_1), P(u_2), \dots, P(u_m)$. У передавальному пристрої повідомлення перетворюється в сигнал так, що кожному елементу повідомлення відповідає певний сигнал. Позначимо ці сигнали через s_1, s_2, \dots, s_m , а ймовірності появи їх на виході передавачів (апріорні ймовірності) відповідно через $P(s_1), P(s_2), \dots, P(s_m)$. Очевидно, апріорні ймовірності сигналів $P(s_i)$ дорівнюють апріорним ймовірностям $P(u_i)$ відповідних повідомлень: $P(s_i) = P(u_i)$. У процесі передавання на сигнал накладається завада. Нехай ця завада має рівномірний спектр потужності з інтенсивністю $\frac{1}{2}N_0$. Тоді сигнал на вході можна подати як суму переданого сигналу $s_i(t)$ і завади $\omega(t)$:

$$x(t) = s_i(t) + \omega(t) \quad (i = \overline{1, m}). \quad (1.19)$$

Оскільки сигнали $s_i(t)$, $x(t)$ і завада $\omega(t)$ задані на скінченному інтервалі ($0 < t < T$), то їх можна зобразити у вигляді розкладань по ортогональних функціях:

$$s_i(t) = \sum_l s_{il} \phi_l(t), \quad \omega(t) = \sum_l \omega_l \phi_l(t), \quad x(t) = \sum_l x_l \phi_l(t), \quad (1.20)$$

Де

$$x_l = s_{il} + \omega_l, \quad s_{il} = \int_0^T s_i(t) \phi_l(t) dt, \quad \omega_l = \int_0^T \omega(t) \phi_l(t) dt. \quad (1.21)$$

Оскільки ми припускаємо, що завада має нормальний розподіл, то і коефіцієнти Фур'є в останньому виразі (1.21) матимуть нормальний розподіл з дисперсією $\sigma_l^2 = \bar{\omega}_l^2 = \frac{1}{2}N_0$ і середнім значенням, рівним нулю:

$$P(\omega_l) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{\bar{\omega}_l^2}{N_0}\right). \quad (1.22)$$

Коефіцієнти x_l також мають нормальний розподіл з тією ж дисперсією $\frac{1}{2}N_0$ і середнім значенням s_{il} :

$$P(x_l / s_{il}) = \frac{1}{\sqrt{\pi N_0}} \exp \left[-\frac{(x_l - s_{il})^2}{N_0} \right]. \quad (1.23)$$

На підставі незалежності коефіцієнтів ω_l багатовимірний розподіл коефіцієнтів x_l , тобто умовний розподіл $P(x/s_i)$, дорівнюватиме добутку одновимірних розподілів [див. третє рівняння (1.21)]:

$$P(x/s_i) = \prod_{l=l_1}^{l_2} P(x_l / s_{il}) = (\pi N_0)^{-n/2} \exp \left[-\frac{1}{N_0} \sum_{l=l_1}^{l_2} (x_l - s_{il})^2 \right]. \quad (1.24)$$

Підставляючи цей вираз у (1.6), одержимо нерівність, яка визначає умову оптимального приймання за Котельниковим:

$$\frac{\prod_{l=l_1}^{l_2} P(x/s_i)}{\prod_{l=l_1}^{l_2} P(x/s_j)} = \frac{\exp \left[-\frac{1}{N_0} \sum_{l=l_1}^{l_2} (x_l - s_{il})^2 \right]}{\exp \left[-\frac{1}{N_0} \sum_{l=l_1}^{l_2} (x_l - s_{jl})^2 \right]} > \frac{P(s_j)}{P(s_i)}. \quad (1.25)$$

Логарифмуючи обидві частини нерівності, дістаємо еквівалентний вираз

$$\sum_{l=l_1}^{l_2} (x_l - s_{il})^2 - \sum_{l=l_1}^{l_2} (x_l - s_{jl})^2 < N_0 \ln \frac{P(s_i)}{P(s_j)}. \quad (1.26)$$

Відповідно до (1.22) і (1.20) маємо

$$x(t) - s_i(t) = \sum_{l=l_1}^{l_2} (x_l - s_{il}) \phi_l(t). \quad (1.27)$$

Після піднесення в квадрат і усереднення за часом виразу (1.27) з урахуванням властивостей ортогональних функцій $\phi_l(t)$ одержуємо

$$\int_0^{T_0} [x(t) - s_i(t)]^2 dt = \sum_{l=l_1}^{l_2} (x_l - s_{il})^2. \quad (1.28)$$

Тоді умову оптимального приймання (1.26) можна записати в іншому вигляді:

$$\int_0^T [x(t) - s_i(t)]^2 dt - \int_0^T [x(t) - s_j(t)]^2 dt < N_0 \ln \frac{P(s_i)}{P(s_j)}. \quad (1.29)$$

Нерівності (1.25) або еквівалентні їм нерівності (1.26) і (1.29) визначають умови правильного приймання сигналу $s_i(t)$.

У випадку, коли апіорні ймовірності сигналів однакові, тобто $P(s_1) = P(s_2) = \dots = P(s_m) = \frac{1}{m}$, критерій Котельникова набирає простішого вигляду:

$$\int_0^T x(t) - s_i(t) \, dt < \int_0^T [x(t) - s_j(t)]^2 \, dt, \quad j \neq i. \quad (1.30)$$

Звідси випливає, що при рівноймовірних сигналах оптимальний приймач відтворює повідомлення, яке відповідає переданому сигналу з найменшим середньоквадратичним відхиленням від прийнятого сигналу.

Розкривши дужки в (1.30), матимемо

$$\int_0^T x^2(t) \, dt + \int_0^T s_i^2(t) \, dt - 2 \int_0^T x(t) s_i(t) \, dt < \int_0^T x^2(t) \, dt + \int_0^T s_j^2(t) \, dt - 2 \int_0^T x(t) s_j(t) \, dt.$$

Для сигналів, енергії яких однакові, ця нерівність для всіх $j \neq i$ набуває простішої форми:

$$\int_0^T x(t) s_i(t) \, dt > \int_0^T x(t) s_j(t) \, dt. \quad (1.31)$$

У цьому випадку умову оптимального приймання можна сформулювати так: *якщо всі можливі сигнали рівноймовірні і мають однакову енергію, оптимальний приймач відтворює повідомлення, що відповідає тому переданому сигналу, взаємна кореляція якого з прийнятим сигналом максимальна.*

Для двійкової системи отриманим результатом можна дати дуже наочне геометричне трактування. Нехай передаються сигнали s_1 та s_2 . Першому сигналу відповідає вектор s_1 у n -вимірному просторі, а другому – вектор s_2 . Прийнятому сигналу відповідає вектор x , рівний сумі векторів сигналу s і завади ω . Простір можливих значень сигналу можна розбити на дві області так, щоб при попаданні кінця вектора x в одну область відтворювався сигнал s_1 (область сигналу s_1), а при попаданні в іншу область – сигнал s_2 (область сигналу s_2). Якщо вектор x , який відповідає даному сигналу, попадає в область іншого сигналу, то має місце помилка (замість s_1 відтворюється s_2 , чи навпаки, s_2 замість s_1). Імовірність помилки, очевидно, залежить від конфігурації областей сигналу. В оптимальному приймачі Котельникова простір сигналів розбивається на області сигналу s_1 і сигналу s_2 так, щоб повна ймовірність помилки (1.14) була мінімальною.

Якщо сигнали рівноймовірні, а завади мають рівномірний розподіл, то оптимальним розбиттям простору буде таке, при котрому будь-яка точка x належить області того сигналу s , кінець вектора якого знаходиться найближче до точки x . У двовимірній моделі (рис. 1.2, а) для двійкової системи границею областей сигналів s_1 і s_2 є геометричне місце точок, рівновіддалених від s_1 і s_2 , тобто гіперплощина, яка перпендикулярна до вектора різниці $\Delta s = s_1 - s_2$ і розділяє його навпіл.

Якщо, наприклад, передавався сигнал s_1 , то помилка станеться в тому випадку, коли виконується нерівність

$$r_2 < r_1 \quad \text{або} \quad \mu > \frac{1}{2} d,$$

де $r_1 = \|x - s_1\|$; $r_2 = \|x - s_2\|$; μ – проекція ω на вектор, колінеарний Δs , тобто $\mu = \frac{\omega \Delta s}{\|\Delta s\|}$.

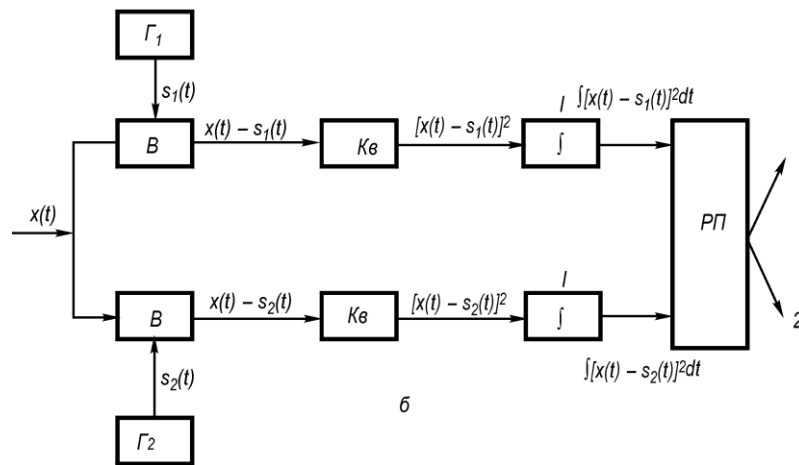
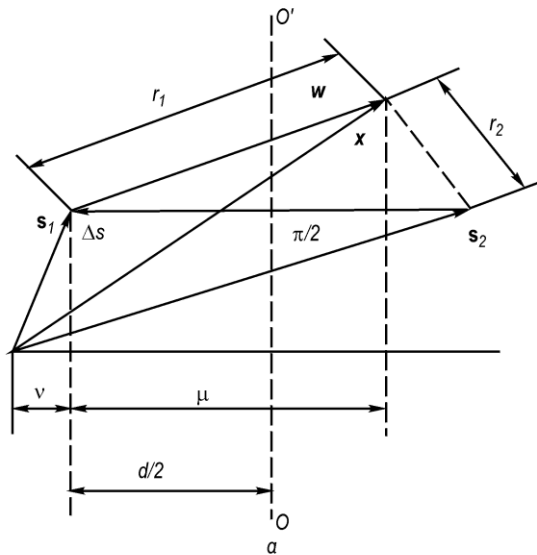


Рис. 1.2. Геометричне зображення роботи оптимального приймача (а) і функціональна схема оптимального приймача В. О. Котельникова (б)

Замість нерівності (1.30) можна записати $\|x - s_2\|^2 < \|x - s_1\|^2$, або в евклідовій метриці

$$\int_0^T x(t) - s_2(t) \cdot x(t) - s_1(t) \, dt < \int_0^T x(t) - s_1(t) \cdot x(t) - s_1(t) \, dt. \quad (1.32)$$

Умова (1.32) цілком збігається з умовою (1.30).

Структурна схема приймача, що реалізує операції (1.32), наведена на рис. 1.2, б. Тут Γ_1 і Γ_2 – генератори опорних сигналів, що формують точні копії переданих сигналів s_1 і s_2 , B – віднімальний пристрій, $Kв$ – квадратуєчий пристрій, I – інтегратор, $ПП$ – схема порівняння і вибору (розв’язувальний пристрій).

Нерівність (1.32) можна записати в іншому вигляді, розкривши дужки під інтегралами:

$$-2 \int_0^T x(t) s_2(t) \, dt + \int_0^T s_2^2(t) \, dt < -2 \int_0^T x(t) s_1(t) \, dt + \int_0^T s_1^2(t) \, dt$$

Або

$$\int_0^T x(t) \cdot s_1(t) - s_2(t) \, dt > \frac{1}{2}(E_2 - E_1). \quad (1.33)$$

Ця нерівність повністю еквівалентна нерівності (1.32), але вимагає іншої схемної реалізації оптимального приймача.

На рис. 1.3, *a* наведена структурна схема приймача, що реалізує умови роботи (1.33). За цією схемою після операцій перемножування (*II*) та інтегрування (*I*) виконується порівняння отриманого результату з постійним порогом, рівним різниці енергій сигналів $\frac{1}{2}(E_2 - E_1)$.

Ця схема простіша за схему на рис.1.2, *б*, проте вона потребує в разі зміни рівня сигналів автоматичного регулювання порога. Цей недолік усувається, якщо сигнали мають однакову енергію ($E_2 = E_1$); тоді поріг дорівнює нулю і розв'язувальна схема визначає тільки знак сигналу на виході.

При $E_2 = E_1$ спрощується і схема приймача на рис. 1.2, *б*. Розкривши дужки в (1.32), одержуємо умови оптимального приймання у вигляді

$$\int_0^T x(t)s_1(t)dt < \int_0^T x(t)s_2(t)dt, \quad (1.34)$$

що збігається з умовою (1.29). Отже, при $E_2 = E_1$ приймач Котельникова перетворюється в кореляційний (когерентний) (рис. 1.3, *б*).

Оптимальне приймання можна також реалізувати за схемою з погодженими лінійними фільтрами (рис. 1.3, *в*), імпульсні реакції яких мають бути $g_1(t) = cs_i(T-t)$, де c – сталий коефіцієнт.

Розглянуті схеми оптимальних приймачів належать до типу когерентних, вони враховують не тільки амплітуду, але й фазу високочастотного сигналу.

Зазначимо, що в схемах оптимальних приймачів фільтри на вході відсутні, а в реальних приймачах вони завжди є. Це означає, що оптимальний приймач при флуктуаційних завадах не потребує фільтрації на вході. Його завадостійкість, як ми побачимо далі, не залежить від ширини смуги пропускання приймача.

1.4. Імовірність помилки при когерентному прийманні двійкових сигналів

Визначимо ймовірність помилки в системі передавання двійкових сигналів на *оптимальний приймач*. Ця ймовірність, очевидно, буде мінімально можливою і характеризуватиме потенційну завадостійкість при даному способі передавання. У разі приймання сигналів на реальний приймач завадостійкість може дорівнювати потенційній, але не може бути більшою за неї.

Нехай сигнал набуває значень $s_1(t)$ з імовірністю P_1 і значень $s_2(t)$ з імовірністю P_2 . Якщо передавався сигнал s_1 , то відповідно до умови (1.29) помилка станеться в тому випадку, коли

$$\int_0^T x(t) - s_1(t) \quad ^2 dt - \int_0^T x(t) - s_2(t) \quad ^2 > N_0 \ln \frac{P_1}{P_2}. \quad (1.35)$$

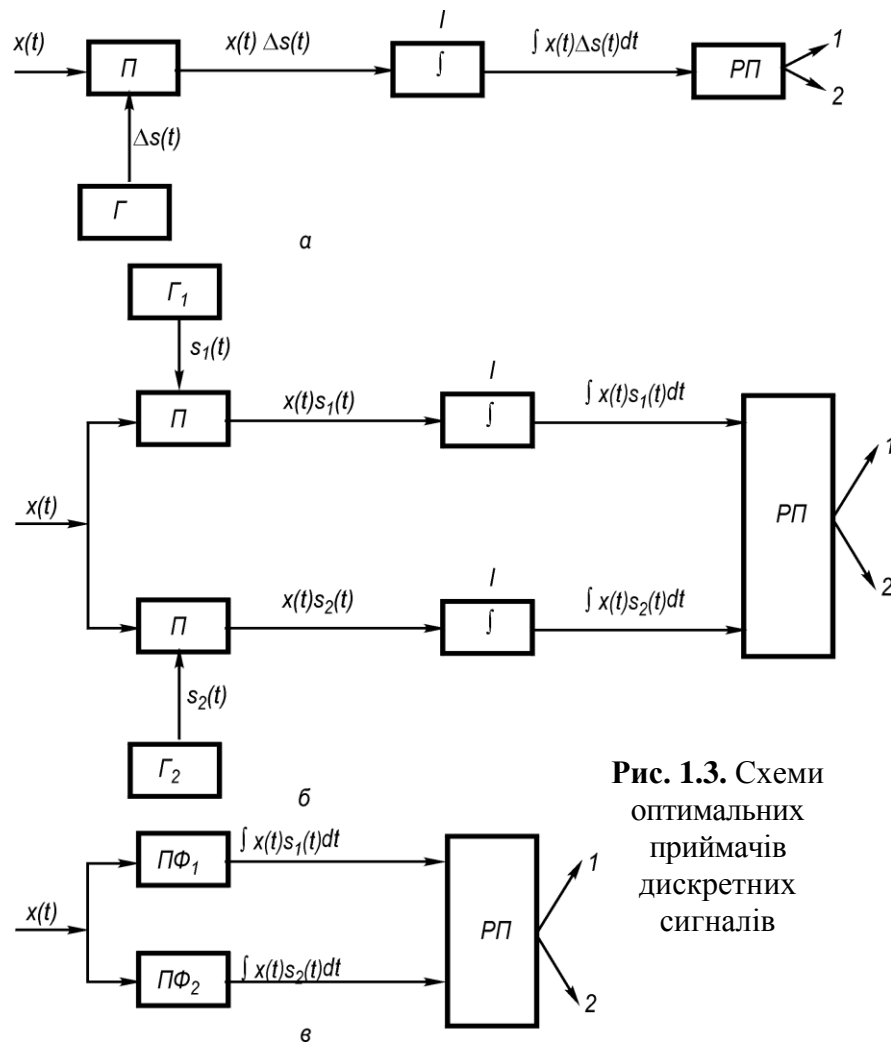


Рис. 1.3. Схеми оптимальних приймачів дискретних сигналів

Оскільки $x(t) = s_1(t) + \omega(t)$, то нерівність (1.35) можна звести до вигляду

$$\int_0^T \omega^2(t) dt - \int_0^T s_1(t) - s_2(t) + \omega(t) \, dt > N_0 \ln \frac{P_1}{P_2}$$

або

$$\int_0^T \omega(t) s_1(t) - s_2(t) \, dt < \frac{1}{2} N_0 \ln \frac{P_2}{P_1} - \frac{1}{2} \int_0^T s_1(t) - s_2(t) \, dt. \quad (1.36)$$

Відповідно до першого і другого виразів (1.20) ліву частину нерівності (1.36) можна записати так:

$$\xi = \int_0^T \omega(t) s_1(t) - s_2(t) \, dt = \sum_l \omega_l (s_{1l} - s_{2l}). \quad (1.37)$$

Кожен коефіцієнт ω_l має нормальний розподіл із середнім значенням, рівним нулю, тому сума (1.37) також являтиме нормальну випадкову величину ξ з нульовим середнім значенням і дисперсією

$$D\xi = \overline{\xi^2} = \sum_l \overline{\omega_l^2} (s_{1l} - s_{2l})^2 = \frac{1}{2} N_0 \int_0^T [s_1(t) - s_2(t)]^2 dt = \sigma_\xi^2. \quad (1.38)$$

Щільність імовірності випадкової величини ξ

$$p(\xi) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} e^{-\frac{\xi^2}{2\sigma_\xi^2}}.$$

Згідно з (1.36) помилка станеться при передаванні сигналу s_1 , якщо

$$\xi < A = \frac{1}{2} N_0 \ln \frac{P_2}{P_1} - \frac{1}{2} \int_0^T s_1(t) - s_2(t)^2 dt,$$

а її ймовірність дорівнюватиме

$$P_{21} = P(\xi < A) = \int_{-\infty}^A p(\xi) d\xi = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \int_{-\infty}^A e^{-\frac{\xi^2}{2\sigma_\xi^2}} d\xi. \quad (1.39)$$

Ввівши нову змінну $u = \xi/\sigma_\xi$, на підставі співвідношення, наведеного в [42], маємо

$$P_{21} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{A}{\sigma_\xi}} e^{-\frac{u^2}{2}} du = \frac{1}{2} \left[1 + \Phi\left(\frac{A}{\sigma_\xi}\right) \right],$$

де

$$\frac{A}{\sigma_\xi} = \frac{\frac{N_0}{2} \ln \frac{P_2}{P_1} - \frac{1}{2} \int_0^T s_1(t) - s_2(t)^2 dt}{\sqrt{\frac{N_0}{2} \int_0^T s_1(t) - s_2(t)^2 dt}}.$$

Після нескладних перетворень остаточно дістаємо [41]:

$$P_{21} = \frac{1}{2} [1 - \Phi(a_{21})], \quad (1.40)$$

де

$$a_{21} = a + \frac{1}{2a} \ln \frac{P_1}{P_2}; \quad (1.41)$$

$$a^2 = \frac{1}{2N_0} \int_0^T s_1(t) - s_2(t)^2 dt. \quad (1.42)$$

Аналогічно визначається ймовірність помилки передавання сигналу $s_2(t)$:

$$P_{12} = \frac{1}{2} [1 - \Phi(a_{12})], \quad (1.43)$$

де

$$a_{12} = a + \frac{1}{2a} \ln \frac{P_2}{P_1}. \quad (1.44)$$

Повна ймовірність помилки при оптимальному прийманні бінарних сигналів $s_1(t)$ і $s_2(t)$

$$P_{\text{п}} = P_1 P_{21} + P_2 P_{12}, \quad (1.45)$$

або, згідно з (1.40) і (1.43),

$$P_{\text{п}} = \frac{1}{2} P_1 [1 - \Phi(a_{21})] + \frac{1}{2} P_2 [1 - \Phi(a_{12})]. \quad (1.46)$$

З отриманих формул випливає, що ймовірність помилки, яка визначає потенційну завадостійкість, залежить від двох величин: a^2 і P_2/P_1 . Перша величина визначається відношенням питомої енергії різниці сигналів до інтенсивності завади N_0 . Чим більше це відношення, тим більша потенційна завадостійкість. Відношення апіорних імовірностей P_2/P_1 залежить від статистичних властивостей переданих повідомлень.

Якщо сигнали, які передаються, рівноймовірні, тобто $P_1 = P_2 = 0,5$, то (1.46) спрощується і набуває вигляду

$$P_{\text{п}} = \frac{1}{2} [1 - \Phi(a)]. \quad (1.47)$$

Формулу (1.47) легко одержати з геометричних зображень. Як видно з рис. 1.2, при передаванні сигналу s_1 помилка станеться у разі виконання нерівностей $r_1 > r_2$ чи $\mu > \frac{1}{2}d$. Отже, ймовірність помилки можна визначити як імовірність виконання однієї з цих нерівностей, тобто

$$P_{\text{п}} = P(r_1 > r_2) = P\left(\mu > \frac{1}{2}d\right).$$

Помноживши обидві частини нерівності на d , одержуємо

$$P_{\text{п}} = P\left(\xi > \frac{1}{2}d^2\right),$$

де $\xi = \mu d \omega \Delta s$ – випадкова величина (1.37), що має нормальний розподіл з дисперсією $D_{\xi} = \frac{1}{2} N_0 d^2$ (1.38). Тоді маємо [45]

$$P_{\text{п}} = P\left(\xi > \frac{1}{2}d^2\right) = \frac{1}{2} [1 - \Phi(a)],$$

Де

$$a^2 = \frac{1}{4} \left(\frac{d^2}{\sqrt{D_{\xi}}} \right)^2 = \frac{1}{2} \frac{d^2}{N_0} = \frac{1}{2N_0} \int_0^T [s_1(t) - s_2(t)]^2 dt,$$

що збігається з (1.47) і (1.42).

У разі малої інтенсивності завад, коли $a \gg 1$, другим членом формул (1.41) та (1.44) можна знехтувати. У цьому випадку формула (1.46) також зводиться до формули (1.47), причому ймовірність помилки практично не залежить від P_1 і P_2 . При великому рівні завад, коли a мале, залежність імовірності помилки від відношення апіорних імовірностей P_1/P_2 стає помітною. Зі збільшенням цього відношення ймовірність помилки збільшується [41].

Отже, для рівноймовірних сигналів імовірність помилки цілком визначається величиною a . Значення цієї величини залежить від спектральної щільності завад N_0 і сигналів $s_1(t)$ і $s_2(t)$, які передаються.

Для систем з активною паузою, в яких сигнали мають однакову енергію $\overline{Ts_1^2(t)} = \overline{Ts_2^2(t)} = E$, величину a^2 можна подати в наступному вигляді:

$$a^2 = \frac{E(1-P_{12})}{N_0} = q_0(1-P_{12}), \quad (1.48)$$

де $P_{12} = \frac{1}{E} \int_0^T s_1(t)s_2(t)dt$ – коефіцієнт взаємної кореляції між сигналами; $q_0 = E/N_0$ – відношення енергії сигналу до питомої потужності завади.

Імовірність помилки для таких систем визначається формулою

$$P_n = \frac{1}{2} \left[1 - \Phi \left[\sqrt{q_0(1-P_{12})} \right] \right]. \quad (1.49)$$

Звідси випливає, що при $P_{12} = -1$, тобто $s_1(t) = -s_2(t)$, система забезпечує найбільшу потенційну завадостійкість. Це система з протилежними сигналами. Для неї $a_{\max}^2 = 2q_0$. Практичною реалізацією системи з протилежними сигналами є система з фазовою маніпуляцією.

Порівнювати різні системи передавання дискретних повідомлень зручно за параметром a^2 , який є зведеним відношенням сигнал-завада на виході оптимального приймача при заданому способі передавання:

$$a^2 = q_{\text{вих}} = q_0(1-P_{12}) \approx TF(1-P_{12})q_{\text{вх}}, \quad (1.50)$$

де $q_0 = TFq_{\text{вх}}$, або за величиною виграшу

$$\gamma = \frac{q_{\text{вих}}}{q_{\text{вх}}} = TF(1-P_{12}).$$

Множник TF визначає виграш за рахунок оптимальної обробки сигналу на прийомі, а множник $(1-P_{12})$ – за рахунок способу передавання.

У загальному вигляді радіотелеграфний сигнал можна записати так:

$$s_i(t) = A_i(t) \cos(\omega_i t + \varphi_i), \quad 0 < t < \tau_0, \quad i = 1, 2, \quad (1.51)$$

де параметри коливання A_i , ω_i , φ_i набувають певних значень в залежності від виду маніпуляції. Згідно з (1.42) для сигналів (1.51) маємо:

$$a^2 = \left\{ \frac{\tau_0}{4N_0} \overline{A_1^2(t)} + \overline{A_2^2(t)} - \right.$$

$$-\frac{2A_1(t)A_2(t)}{\tau_0(\omega_1 - \omega_2)} \left[\sin(\omega_1 - \omega_2)\tau_0 + \varphi_1 - \varphi_2 - \sin(\varphi_1 - \varphi_2) \right] \}. \quad (1.52)$$

Для амплітудної маніпуляції $A_1(t) = A_0$, $A_2(t) = 0$, $\omega_1 = \omega_2 = \omega_0$; тоді

$$a_{\text{AM}}^2 = \frac{A_0^2 \tau_0}{4N_0} = \frac{1}{2} q_0.$$

Для частотної маніпуляції $A_1(t) = A_2(t) = A_0$, $\omega_1 \neq \omega_2$. При оптимальному виборі розносу частот $(\omega_1 - \omega_2) \cdot 2\pi = k / \tau_0$, де k – ціле число; $\varphi_1 = \varphi_2$. Тоді на підставі (1.50) одержуємо

$$a_{\text{ЧМ}}^2 = q_0.$$

Для фазової маніпуляції $A_1(t) = A_2(t) = A_0$, $\omega_1 = \omega_2 = \omega_0$, $\varphi_1 - \varphi_2 = \pi$; тоді

$$a_{\text{ФМ}}^2 = 2q_0.$$

Порівняння отриманих формул показує, що з усіх систем передавання бінарних сигналів найбільшу потенційну завадостійкість забезпечує система з фазовою маніпуляцією. У порівнянні з системою ЧМ вона дозволяє одержати дворазовий, а в порівнянні з системою АМ – чотириразовий вигравш за потужністю.

У системах зв'язку сигнал звичайно складається з послідовності простих сигналів. Так, у телеграфії кожній букві відповідає кодова комбінація з п'яти елементарних посилок. Можливі й складніші комбінації. Якщо елементарні сигнали, що складають кодову комбінацію, незалежні, то ймовірність помилкового приймання кодової комбінації визначається наступною формулою:

$$P_{\text{п.к}} = 1 - (1 - P_{\text{п}})^n, \quad (1.53)$$

де $P_{\text{п}}$ – ймовірність помилки елементарного сигналу; n – число елементарних сигналів у кодовій комбінації (значність коду).

Слід звернути увагу, що ймовірність помилки в розглянутих вище випадках цілком визначається відношенням енергії сигналу до спектральної щільності завади і не залежить від форми сигналу. В загальному випадку, коли спектр завади відрізняється від рівномірного, ймовірність помилки можна зменшити, змінивши спектр сигналу, тобто його форму.

1.5. Ймовірність помилки при когерентному прийманні багатопозиційних сигналів

Робота приймача в багатопозиційних системах зводиться до розрізнення m сигналів, що відповідають m позиціям коду. Схему приймача можна уявити собі як таку, що складається з m каналів (гілок), кожний з яких розрахований на приймання одного визначеного сигналу. Одним із прикладів багатопозиційної системи є система з частотною маніпуляцією, сигналами якої є гармонічні коливання різних частот. Приймач у цій системі містить m фільтрів, настроєних на частоти сигналів, що передаються. За допомогою цих фільтрів і здійснюється поділ (розрізнення) сигналів.

Нехай $s_1(t)$, $s_2(t)$, ..., $s_m(t)$ – сигнали, використовувані для передачі; $\omega(t)$ – адитивна завада, що впливає на приймач; x_1 , x_2 , ..., x_m – прийняті сигнали. Якщо передавався сигнал $s_1(t)$, то на перший фільтр надійдуть сигнал і завада, тобто $x_1 = s_1 + \omega$, а на інші фільтри – тільки завада. Приймач порівнює прийняті сигнали і відтворює найбільший з них, тобто виносить рішення про те, що передано k -тий сигнал, якщо $x_k = \max(x_1, x_2, \dots, x_m)$. Оскільки ми припустили, що передавався сигнал s_1 , то ймовірність правильного рішення дорівнюватиме

$$P_{\text{прав}} = P(x_1 > x_2, x_3, \dots, x_m),$$

а ймовірність помилки

$$P_{\text{п}} = 1 - P_{\text{прав}}.$$

Оптимальний когерентний приймач у m -позиційній системі являє собою багатоканальний корелятор чи систему з m погоджених фільтрів. Структурна схема такого приймача аналогічна схемам на рис. 1.3, б або рис. 1.3, в для двійкових сигналів (різниця лише в числі каналів). У цьому разі приймач відповідно до умов (1.31) обчислює функцію взаємної кореляції прийнятого сигналу $x(t)$ з усіма m опорними сигналами:

$$y_k = \int_0^T s_k(t)x(t)dt = \int_0^T s_k(t) s_1(t) + \omega(t) dt, \quad k = \overline{1, m}, \quad (1.54)$$

і видає рішення про те, що був переданий той сигнал, кореляція якого найбільша. Ймовірність правильного рішення (якщо передається сигнал s_1) дорівнюватиме

$$P_{\text{прав}} = P(y_1 > y_2, y_3, \dots, y_m). \quad (1.55)$$

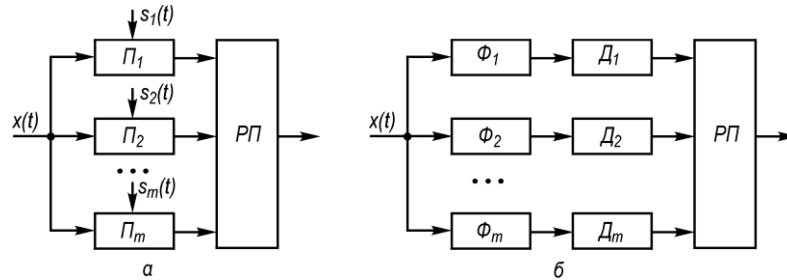


Рис. 1.4. Функціональні схеми приймачів багатопозиційних сигналів

На рис. 1.4, а наведена структурна схема когерентного приймача багатопозиційних сигналів, побудована на базі кореляційної техніки. У кожному каналі цієї схеми виконується синхронне детектування сигналів. Отримані відліки надходять на схему порівняння (розв'язувальний пристрій ПП), після чого видається рішення про те, який з m сигналів був переданий.

Визначимо ймовірність помилки при оптимальному когерентному прийманні ортогональних m -позиційних сигналів. Вважатимемо, що всі можливі сигнали рівноймовірні і мають однакову енергію E . Для цього випадку умови правильного приймання сигналу s_1 згідно з (1.31)

$$\int_0^T x(t)s_1(t)dt > \int_0^T x(t)s_j(t)dt, \quad j = \overline{2, m}.$$

Оскільки $x(t) = s_1(t) + \omega(t)$, $\int_0^T s_1(t)s_j(t)dt = 0$ і $\int_0^T s_1^2(t)dt = E$, то ця нерівність набирає вигляду

$$\int_0^T \omega(t)[s_1(t) - s_j(t)]dt > -E \quad \text{або} \quad \int_0^T s_j(t)\omega(t)dt < \int_0^T s_1(t)\omega(t)dt + E.$$

Розглянемо функцію $\xi_j = \int_0^T s_j(t)\omega(t)dt$, що є нормальною випадковою величиною (завадою на виході j -го каналу) з дисперсією

$$D\xi_j = \sigma_\xi^2 = \overline{\xi_j^2} = \frac{1}{2} N_0 E.$$

Щільність імовірності величини ξ_j

$$p(\xi_j) = \frac{1}{\sqrt{\pi N_0 E}} \exp\left(-\frac{\xi_j^2}{N_0 E}\right).$$

Імовірність того, що завада в j -му каналі не перевищить сумарного значення сигналу і завади в першому каналі, тобто ймовірність того, що $\xi_j < \xi_1 + E$, дорівнюватиме

$$\int_{-\infty}^{\xi_1 + E} \frac{\exp\left[-\xi_j^2 / (N_0 E)\right]}{\sqrt{\pi N_0 E}} d\xi_j = \frac{1}{2} \left[1 + \Phi(t + \sqrt{2q_0}) \right],$$

$$\text{де } \Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt, \quad t = \frac{\xi_1}{\sqrt{\frac{N_0 E}{2}}} = \frac{\xi_1}{\sigma_\xi}, \quad q_0 = \frac{E}{N_0}.$$

Імовірність того, що завада ξ_j у всіх $m-1$ каналах без сигналу не перевищить сумарного значення сигналу і завади в першому каналі,

$$P_{\xi_j} = \left(\frac{1}{2}\right)^{m-1} \left[1 + \Phi(t + \sqrt{2q_0}) \right]^{m-1}.$$

Інтегрування цього виразу за всіма можливими значеннями завади ξ_j дає ймовірність правильного приймання

$$P_{\text{прав}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \left\{ \frac{1}{2} \left[1 + \Phi(t + \sqrt{2q_0}) \right] \right\}^{m-1} dt. \quad (1.56)$$

Імовірність помилки

$$P_{\text{п.кг}} = 1 - P_{\text{прав}} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \left[\frac{1}{2} + \frac{1}{2} \Phi(t + \sqrt{2q_0}) \right]^{m-1} dt. \quad (1.57)$$

В окремому випадку при $m=2$ вираз (1.57) перетворюється у формулу, аналогічну (1.47) для двійкових систем:

$$P_{\text{п}} = \frac{1}{2} \left[1 - \Phi(\sqrt{q_0}) \right].$$

При $m > 2$ інтеграл у правій частині (1.57) може бути обчислений наближеними методами. Якщо відношення сигнал-завада відносно велике $q_0 > 1$, то має місце асимптотичний вираз

$$P_{\text{п.кг}} \approx \frac{m-1}{2} \left[1 - \Phi(\sqrt{q_0}) \right]. \quad (1.58)$$

При неоптимальному когерентному прийманні дискретних сигналів у схемі інтегратор відсутній. У цьому випадку після синхронного детектора ставиться фільтр нижніх частот і береться відлік (стробування) на виході фільтра в середині посилки. Можна показати, що

ймовірність помилки у разі неоптимального когерентного приймання визначається отриманими вище виразами (1.57) і (1.58), якщо в останні замість q_0 підставити $q = q_0 / TF = P_c / P_s$.

1.6. Некогерентне приймання дискретних сигналів

При некогерентному прийманні інформація про фазу прийнятих сигналів не використовується. Такий спосіб застосовується в каналах зі змінними параметрами, коли фаза сигналу випадково змінюється і її визначення викликає значні труднощі, а також у каналах зі сталими параметрами з метою спрощення схеми приймача.

Оптимальний некогерентний приймач обчислює модуль (обвідну) функції взаємної кореляції $z_k = \left| \int_0^T \dot{x}(t) \dot{s}_k^*(t) dt \right|$ і вирішує, що був переданий той сигнал, для якого значення z у деякий момент часу $t = t_0$ найбільше. Нехай передавався сигнал $s_1(t)$; тоді умову правильного приймання цього сигналу можна записати в наступному вигляді:

$$z_k < z_1 \text{ або } \left| \int_0^T \dot{x}(t) \dot{s}_k^*(t) dt \right| < \left| \int_0^T \dot{x}(t) \dot{s}_1^*(t) dt \right| \quad (k=2,3,\dots,m) \quad (1.59)$$

Схема приймача, що реалізує умову (1.59), наведена на рис. 1.4, б. Ця схема містить m погоджених фільтрів Φ , що відповідають m окремим сигналам. На виході кожного фільтра напруга пропорційна функції взаємної кореляції $y_k = \int_0^T \dot{x}(t) \dot{s}_k(t) dt$. Амплітудний детектор D виділяє обвідну (модуль) цієї функції. Потім робиться відлік сигналу і приймається рішення.

Припустимо, що сигнали рівноймовірні, мають однакову енергію і є ортогональними. За цих умов маємо

$$z_k^2 = \left[2 \int_0^T \omega(t) s_k(t) dt \right]^2 + \left[2 \int_0^T \omega(t) \xi_k(t) dt \right]^2 = \xi_k^2 + \eta_k^2, \\ z_1^2 = (2E + \xi_1)^2 + \eta_1^2, \quad (1.60)$$

де $\xi_k = 2 \int_0^T \omega(t) s_k(t) dt$; $\eta_k = 2 \int_0^T \omega(t) \xi_k(t) dt$.

Випадкові величини ξ і η мають нормальний розподіл з нульовим середнім значенням і дисперсією $\sigma^2 = \sigma_\xi^2 = \sigma_\eta^2 = 2N_0E$. У цьому легко переконатися так само, як це було зроблено при виведенні формули (1.38).

Випадкова величина z_k^2 є сумою квадратів двох незалежних випадкових величин ξ_k і η_k з нормальним розподілом, нульовим середнім значенням і однаковими дисперсіями, рівними $2N_0E$. Така величина, як відомо, має розподіл Релея [42]. У нашому випадку

$$p(z_k) = \frac{z_k}{2N_0E} e^{-\frac{z_k^2}{4N_0E}} \quad (1.61)$$

Випадкову величину z_1^2 можна розглядати як квадрат векторної суми постійного вектора, довжина якого $L = 2E$, і випадкового вектора з нормально розподіленими незалежними

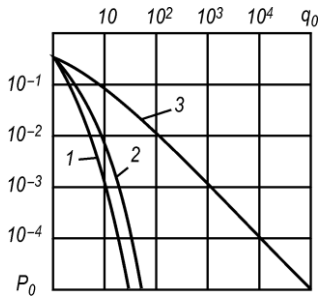


Рис. 1.5. Залежність імовірності помилки в біполярній системі з активною паузою від відношення сигнал–завада:

складовими, що мають дисперсію $\sigma_1^2 = 2N_0E$. Тому величина z_1 підпорядковується узагальненому розподілу Релея з щільністю ймовірностей

$$p(z_1) = \frac{z_1}{2N_0E} \exp\left(-\frac{z_1^2 + L^2}{4N_0E}\right) I_0\left(\frac{z_1L}{2N_0E}\right). \quad (1.62)$$

Випадкові величини z_k є не що інше, як обвідні напруги в каналах без сигналу, тобто обвідні завад. Оскільки завади ми вважаємо гауссовими, то цим і пояснюється, що z_k має релеївський розподіл. Випадкова величина z_1 – це обвідна сумарного коливання сигналу і завади в каналі з сигналом.

Тепер можна визначити ймовірність помилки при некогерентному прийманні. У загальному випадку вона дорівнюватиме

$$P_{\text{п.нкг}} = 1 - P(z_1 > z_2, z_3, \dots, z_m). \quad (1.63)$$

При бінарному передаванні $m = 2$

$$P_{\text{п.нкг}} = 1 - P(z_1 > z_2) = P(z_2 > z_1).$$

Для обчислення ймовірності помилки спочатку треба знайти при деякому фіксованому значенні z_1 імовірність того, що $z_2 > z_1$. Ця ймовірність виражається інтегралом

$$I(z_1) = \int_{z_1}^{\infty} P(z_2) dz_2,$$

який має різні значення при різних z_1 . Щоб знайти повну ймовірність того, що $z_2 > z_1$, необхідно $I(z_1)$ усереднити по всіх можливих значеннях z_1 відповідно до розподілу $P(z_1)$. Таким чином,

$$\begin{aligned} P_{\text{п.нкг}} &= P(z_2 > z_1) = \int_0^{\infty} I(z_1) P(z_1) dz_1 = \\ &= \int_0^{\infty} P(z_1) dz_1 \int_{z_1}^{\infty} P(z_2) dz_2 \end{aligned} \quad (1.64)$$

Після підстановки в (1.64) виразів $P(z_2)$ і $P(z_1)$ відповідно до (1.61) і (1.62) та інтегрування одержуємо наступний вираз для ймовірності помилки при оптимальному некогерентному прийманні двійкових сигналів:

$$P_{\text{п2}} = \frac{1}{2} e^{-\frac{q_0}{2}}, \quad (1.65)$$

де $q_0 = E/N_0$.

Для m -позиційних систем слушним є наближене співвідношення [45]

$$P_{\text{п.нкг}} \approx \frac{m-1}{2} e^{-\frac{q_0}{2}}. \quad (1.66)$$

Із порівняння (1.58) і (1.66) випливає, що ймовірність помилки в багатопозиційних системах P_{pm} приблизно визначається через ймовірність помилки у відповідній двійковій системі $P_{п2}$. Їхнє співвідношення виглядає так:

$$P_{pm} \approx (m-1)P_{п2}. \quad (1.67)$$

На рис. 1.5 наведені графіки залежності ймовірності помилки в двійковій системі з активною паузою від відношення сигнал–завада при когерентному і некогерентному прийманні. Порівняння кривих показує, що оптимальне когерентне приймання несуттєво відрізняється за завадостійкістю від оптимального некогерентного приймання, а при неоптимальному прийманні і великому рівні завад $q < 1$ розбіжності можуть бути значними [45].

1.7. Передавання дискретних сигналів по каналах із змінними параметрами

У системах передавання інформації, параметри каналу яких незмінні, ймовірність помилки обумовлена лише наявністю адитивних завад. Оптимальний приймач таких систем використовує всі параметри сигналу і забезпечує максимально можливу завадостійкість. За реальних умов параметри каналу і відповідно параметри сигналу можуть випадково змінюватися, тому ефективне використання їх на прийомі ускладнюється, а завадостійкість неминуче погіршується.

У реальних каналах сигнали на вхід приймача можуть надходити різними шляхами з різними загасаннями μ_k і різними запізненнями τ_k . Прийнятий сигнал $x(t)$ у цьому випадку можна подати у вигляді суми:

$$x(t) = \sum_{k=1}^n \mu_k s(t - \tau_k) + \omega(t), \quad (1.68)$$

де μ_k і τ_k – випадкові процеси, що в загальному випадку залежать від часу; $\omega(t)$ – адитивна завада.

Канали, описувані виразом (1.68), називаються *багатопроменевими*.

У багатьох випадках має місце тільки один шлях поширення сигналу; в такому разі канали зв'язку описуються виразом

$$x(t) = \mu s(t - \tau) + \omega(t) \quad (1.69)$$

і називаються *однопроменевими*. Якщо величини μ і τ фіксовані в часі, то маємо канали зі сталими параметрами. Це, наприклад, канали проводового зв'язку та ультракороткохвильові канали при передаванні в межах прямої видимості.

Майже усі види радіоканалів є каналами з випадково змінними параметрами. До цього класу каналів належать короткохвильові лінії зв'язку, в яких завдяки зміні стану іоносфери відбуваються неперервні коливання амплітуди сигналу в точці приймання. Випадкові зміни умов поширення радіохвиль мають місце на ультракоротких хвилях, а також на середніх і навіть довгих хвилях.

Будь-які зміни коефіцієнта передачі і часу поширення сигналів спричинюють флуктуації сигналів на виході каналів і можуть розглядатися як завади. Зміни коефіцієнта μ проявляються у вигляді флуктуацій амплітуди сигналу і називаються *мультиплікативною завадою*. Випадкові затримки променів спричинюють фазові і тимчасові флуктуації сигналів.

Вплив мультиплікативної завади на переданий сигнал s можна розглядати як модуляцію цього сигналу випадковим процесом μ . Таку заваду можна звести до адитивної:

$$x = \mu s = \mu_0 s + \omega_e, \quad (1.70)$$

де $\omega_e = s(\mu - \mu_0) = s\xi$ – еквівалентна адитивна завада.

Еквівалентна завада ω_e є добутком випадкового процесу $\xi = \mu - \mu_0$ на детерміновану функцію часу s . Теоретично це означає, що ω_e – нестационарний процес і що всі його розподіли та їх моменти залежать від часу, а практично – що при аналізі потрібно після усереднення за множиною виконувати усереднення за часом. Зазначимо також, що μ є випадковим процесом з ненульовим середнім: $\bar{\mu} = \mu_0 \neq 0$, і що $\mu > 0$. Процес $\xi = \mu - \mu_0$ має вже нульове середнє і відповідно зміщений розподіл. Величини μ і ξ безрозмірні, ω_e має розмірність сигналу.

Еквівалентна потужність мультиплікативної завади, очевидно, дорівнюватиме її дисперсії:

$$P_{z,e} = \widetilde{D\omega_e} = \widetilde{Ds\xi} = s^2 \widetilde{D\xi} = s^2 D\mu.$$

Відношення сигнал–еквівалентна завада при цьому

$$q_e = \frac{\widetilde{\mu_0 s^2}}{\widetilde{D\omega_e}} = \frac{\mu_0^2}{D\xi} = \frac{\mu_0^2}{D\mu}, \quad (1.71)$$

тобто визначається тільки середнім значенням і дисперсією процесу $\mu(t)$, що характеризує мультиплікативну заваду.

Розповсюдженням явищем, що призводить до випадкових коливань параметрів сигналу в точці приймання, є багатопроменеве поширення радіохвиль. Багатопроменевість – основна причина завмирання сигналу. Внаслідок різниці ходу променів, що приходять від передавача до приймача, сигнал у приймальній антені являє собою суму окремих коливань з різними фазами й амплітудами. Інтерференція цих коливань в умовах, коли різниці ходу променів не лишаються постійними, і є основною причиною флуктуацій як амплітуд, так і фаз складових сигналу. В залежності від ширини спектра сигналу F і властивостей каналу розрізняють загальні (чи гладкі) і селективні завмирання, які, в свою чергу, можуть бути швидкими і повільними.

Загальні завмирання мають місце, коли час запізнювання променів $\tau \ll 1/F$. При цьому коефіцієнт передачі каналу μ (чи амплітуда сигналу) і фаза сигналу φ для всіх частотних складових змінюються однаково. Коли число променів велике, то можна вважати, що випадкова величина φ має рівномірну щільність імовірності на інтервалі від 0 до 2π , а μ – розподіл Релея:

$$p(\mu) = \frac{2\mu}{\mu_0^2} e^{-\frac{\mu^2}{\mu_0^2}}, \quad (1.72)$$

де $\mu_0 = \sqrt{\overline{\mu^2}}$ – середньоквадратичне значення коефіцієнта передачі μ .

У ряді випадків сигнал у точці приймання має дві складові: регулярну з повільнозмінними параметрами і розсіяну швидкофлуктуючу. В такому разі коефіцієнт передачі μ описується узагальненим розподілом Релея

$$p(\mu) = \frac{2\mu}{\overline{\mu_\phi^2}} \exp\left(-\frac{\mu^2 + \mu_p^2}{\overline{\mu_\phi^2}}\right) I_0\left(\frac{2\mu\mu_p}{\overline{\mu_\phi^2}}\right), \quad (1.73)$$

де μ_p – регулярна складова коефіцієнта передачі; $\overline{\mu_\phi^2} = \mu_0^2 - \mu_p^2$ – середнє значення квадрата флуктуючої складової; $\mu_0 = \sqrt{\overline{\mu^2}}$ – середньоквадратичне значення μ . Фаза сигналу ϕ в цьому випадку розподілена нерівномірно.

У разі сильних завмирань, коли $\overline{\mu_\phi^2} \gg \mu_p^2$, розподіл (1.73) наближається до розподілу Релея (1.72). Тому такі завмирання іноді називають *релеївськими*, а завмирання, що підпорядковуються узагальненому закону Релея, – *квазірелеївськими*.

Для слабких завмирань, коли $\overline{\mu_\phi^2} \ll \mu_p^2$, на підставі (1.73) одержуємо розподіл, близький до нормального:

$$p(\mu) \approx \frac{1}{\sqrt{2\pi\overline{\mu_\phi^2}}} \sqrt{\frac{\mu}{\mu_p}} e^{-\frac{(\mu-\mu_p)^2}{2\overline{\mu_\phi^2}}}.$$

Слабкі завмирання називають *гауссовими*.

Згідно з експериментальними даними в діапазонах середніх і коротких хвиль релеївські та квазірелеївські завмирання зустрічаються приблизно однаково часто. В ультракороткохвильовому діапазоні при далекому іоносферному чи тропосферному поширенні переважають релеївські завмирання, при ближньому поширенні – квазірелеївські.

Селективні завмирання спостерігаються тоді, коли час запізнювання променів порівняний з величиною $1/F$. У цьому випадку амплітуди і фази частотних складових сигналу змінюються незалежно один від одного. При швидких завмираннях амплітуди і фази суміжних елементів сигналу некорельовані між собою, а при повільних завмираннях вони змінюються однаково.

У випадку повільних завмирань, що ми і розглянемо, коефіцієнт передачі каналу μ і фаза сигналу ϕ практично не змінюються протягом тривалості декількох елементів сигналу. Аналіз раніше прийнятих елементів сигналу дозволяє з достатнім ступенем точності передбачити очікувані параметри наступного елемента. За цих умов приймання може бути здійснено так само, як і у разі відсутності завмирань, і оптимальними будуть схеми, розглянуті в попередніх параграфах, з тією лише різницею, що в схемах мають вироблятися неперервні регулювання μ (регулювання посилення) і ϕ (регулювання фази) відповідно до очікуваних значень μ і ϕ . У зв'язку з труднощами підстроювання фази більш широке застосування знаходять некогерентні методи приймання, в яких відомості про фазу сигналу не використовуються. У каналах, де випадково змінюється тільки фаза сигналу, некогерентне приймання є оптимальним.

Формули, що визначають імовірність помилки елемента сигналу при повільних завмираннях для даного значення q_0 , залишаються тими ж, що й у каналі без завмирань. Але в процесі завмирань величина q_0 змінюється пропорційно μ^2 . Тому для визначення повної ймовірності помилки необхідно усереднити ймовірність $P_n(q_0)$ відповідно до розподілу $p(\mu)$, тобто

$$\overline{P_n} = \int_0^\infty p(\mu) P_n(q_0) d\mu = \int_0^\infty p(\mu) P_n\left(\frac{\mu\sqrt{q_0}}{\mu_0}\right) d\mu, \quad (1.74)$$

де враховано, що $q_0 = (\mu^2 / \mu_0^2) \overline{q_0}$.

Ймовірність помилки при когерентному прийманні бінарних сигналів в умовах повільних релєївських завмирань [17]

$$\overline{P}_n = \frac{1}{2} \left(1 - \sqrt{\frac{q_0}{\frac{2}{1-P_{12}} + q_0}} \right), \quad (1.75)$$

де P_{12} визначається за формулою (1.10).

Якщо $q_0 \gg 1$, маємо наближену формулу

$$\overline{P}_n \approx \frac{1}{2q_0(1-P_{12})}. \quad (1.76)$$

При некогерентному прийманні для систем з активною паузою ймовірність помилки у разі повільних релєївських завмирань визначається усередненням (1.72) за μ відповідно до формули

$$\overline{P}_n = \int_0^\infty 2\mu e^{-\frac{\mu^2}{\mu_0^2}} \frac{1}{2} e^{-\frac{\mu^2}{\mu_0^2} q_0} d\mu = \frac{1}{2+q_0}. \quad (1.77)$$

Графік залежності (1.77) наведений на рис. 1.5. З порівняння кривих видно, що наявність завмирань сигналу значно знижує завадостійкість системи зв'язку.

У випадку квазірелєївських завмирань щільність ймовірності $p(\mu)$ визначається узагальненим розподілом Релея (1.73). Для цього випадку повна ймовірність помилки при некогерентному прийманні бінарних сигналів визначається виразом [16]

$$\overline{P}_n = \frac{c^2 + 1}{q_0 + 2c^2 + 2} \exp\left(-\frac{c^2 \overline{q_0}}{q_0 + 2c^2 + 2}\right), \quad (1.78)$$

де $c^2 = \mu_p^2 / \mu_\Phi^2$.

Якщо $c = 0$, то цей вираз переходить у формулу (1.77) для релєївських завмирань.

Канал із квазірелєївськими завмираннями є проміжним випадком між каналом без завмирань і каналом з релєївськими завмираннями. Графіки ймовірності помилки для цього випадку на рис. 1.5 будуть лежати між кривими 2 і 3. При малих значеннях c графіки наближатимуться до кривої 3, а при великих – до кривої 2.

Природним методом усунення мультиплікативної завади, в тому числі й завмирання, є застосування *автоматичного регулювання посилення (АРП)*. За відсутності адитивної завади ідеальна система АРП дозволяє цілком усунути мультиплікативну заваду. Дія такої системи зводиться до множення сигналу x на $1/\mu$, тобто $z = (1/\mu)x = (1/\mu)\mu s = s$. За наявності адитивної завади $z = (1/\mu)x = (1/\mu)(\mu s + \omega) = s + \omega/\mu$. У цьому разі ми хоча і отримуємо сигнал постійної інтенсивності, але з флюктууючою за інтенсивністю адитивною завадою.

Ефективним заходом боротьби з завмираннями є *рознесення приймання сигналів*. Суть його полягає в тому, що передане повідомлення відтворюється не за одним прийнятим сигналом, а за двома чи більше сигналами, що несуть те саме повідомлення. Цими сигналами можуть бути сигнали кількох передавачів, які працюють на різних частотах (рознесення за частотою), або сигнали того самого передавача, прийняті на різні антени, рознесені за простором чи поляризацією. Можливо також і рознесення за часом (наприклад, повторення передачі).

Обробка кількох сигналів на прийомі зводиться до підсумовування за вагою. В окремих випадках це може бути просте додавання прийнятих сигналів чи вибір найбільшого з них.

РОЗДІЛ 2

ОПТИМАЛЬНЕ ПРИЙМАННЯ НЕПЕРЕРВНИХ ПОВІДОМЛЕНЬ ОПТИМАЛЬНА ФІЛЬТРАЦІЯ

2.1. Оптимальне приймання неперервних повідомлень

Нехай повідомлення перетворюються в деяке коливання $u(t)$, що може неперервно змінюватися в часі й набувати будь-якої форми. З такими повідомленнями ми зустрічаємося, наприклад, у телефонії, телеметрії, телебаченні. Визначимо умови оптимального приймання неперервних повідомлень.

Для простоти аналізу вважатимемо, що: 1) функція $u(t)$ набуває значень в межах від $+1$ до -1 ; 2) передане коливання є періодичним з періодом T (це завжди можна допустити, взявши значення T досить великим); 3) спектр коливання обмежений частотами від i_1/T до i_2/T . За цих умов функцію $u(t)$ можна подати у вигляді

$$u(t) = \sum_{i=i_1}^{i_2} \left(\lambda_{2i-1} \sqrt{2} \sin \frac{2\pi i}{T} t + \lambda_{2i} \sqrt{2} \cos \frac{2\pi i}{T} t \right) = \sum_{e=e_1}^{e_2} \lambda_e \varphi_e(t), \quad (2.1)$$

де λ_e – деякі параметри, що визначають передане коливання $u(t)$; $\varphi_e(t)$ – одиничні ортогональні функції; $e_1 = 2i_1 - 1$, $e_2 = 2i_2$.

Для передавання по каналу коливання $u(t)$ перетворюється в сигнал $s(u, t)$. Оскільки коливання (2.1) визначається $e_2 - e_1 + 1$ параметрами λ_e , то сигнал залежатиме від цих параметрів: $s(u, t) = s(\lambda_{e_1}, \dots, \lambda_{e_2}, t)$. Прийнятий сигнал (коливання) внаслідок накладання завади $\omega(t)$ дорівнює $x(t) = s(u, t) + \omega(t)$.

Задача полягає в тому, щоб за коливанням $x(t)$ відновити передане повідомлення $u(t)$ з якомога більшою точністю. Як ми вже встановили, найбільше, що може зробити приймач на підставі аналізу прийнятого сигналу $x(t)$, – це обчислити розподіл імовірностей $P(s/x)$ для всіх можливих реалізацій $u(t)$. При передаванні неперервних коливань функція $P(s/x)$ є щільністю розподілу ймовірності:

$$P(s_i/x) > P(s_j/x), \quad j \neq i.$$

Згідно з цим виразом оптимальний приймач обчислює апостеріорний розподіл імовірностей $P(s/x)$ в очевидному вигляді або видає на виході результат, еквівалентний до цієї функції, і реалізує на виході те повідомлення $u(t)$, при якому функція $P(s/x)$ є максимальною.

Відповідно до формули Бейеса вираз $P(s/x)$ для розглянутого випадку можна записати в наступному вигляді:

$$P(s/x) = kP(s)P(x/s), \quad (2.2)$$

де k – деяка стала, яка може бути обчислена з умови нормування $\int_s P(s/x) dx = 1$.

Припускаємо, що всі можливі повідомлення $u(t)$ та відповідні їм сигнали $s(u, t)$ рівноймовірні, тобто $P(s) = \text{const}$ для всіх реалізацій $u(t)$, що лежать в інтервалі $(-1, +1)$. При цьому відповідно до [9]

$$P(x/s_i) = \prod_{i=i_1}^{i_2} P(x_i/s_{i1}) = (\pi N_0)^{-\frac{n}{2}} \exp \left[-\frac{1}{N_0} \sum_{i=i_1}^{i_2} (x_i - s_{i1}) \right].$$

Із виразу (2.2) маємо:

$$P(s/x) = k_1 P(s) \exp \left\{ -\frac{1}{N_0} \int_0^T x(t) - s(u,t) \right. \left. \right\}^2 dt \left. \right\}, \quad (2.3)$$

де k_1 – параметр, який може бути обчислений для однієї реалізації випадкового процесу, що є моделлю повідомлення.

Звідси максимуму апостеріорної ймовірності $P(s/x)$ відповідає мінімальна за $u(t)$ величина

$$\Delta^2 = \int_0^T x(t) - s(u,t) \right. \left. \right\}^2 dt. \quad (2.4)$$

Отже, оптимальний приймач відтворює повідомлення $u(t)$, при якому середньоквадратичне відхилення Δ^2 має мінімальне значення. За відсутності завад такий приймач відтворює повідомлення без спотворень (без помилок): $x(t) = s(u,t)$, $v(t) = u(t)$ і $\Delta^2 = 0$, а за наявності завад помилка буде мінімальною.

З виразу (2.4) випливає, що операції оптимальної фільтрації та детектування дозволяють отримати максимальну інформацію з прийнятого сигналу $x(t)$ щодо переданого повідомлення $u(t)$. Структурна схема оптимального приймача наведена на рис. 2.1, а.

Запишемо вираз (2.3) в іншому вигляді:

$$P(s/x) = k_1 P(s) \exp \left[-\frac{1}{N_0} \int_0^T x^2(t) dt \right] \exp \left[-\frac{1}{N_0} \int_0^T s^2(t) dt \right] \exp \left[\frac{2}{N_0} \int_0^T x(t) s(u,t) dt \right]. \quad (2.5)$$

Перший експоненціальний множник, що не залежить від u , може бути включений у сталу k_1 . Другий множник взагалі не зображує операцію над $x(t)$; він може бути винесений у вигляді окремого множника, подібного до апіорної ймовірності. В багатьох випадках цей множник, рівний $\exp(-E/N_0)$, де E – енергія сигналу, також можна включити в сталу k_1 . Тоді вираз (2.5) можна записати так:

$$P(s/x) = k_1 P(s) \exp h(u), \quad (2.6)$$

де

$$h(u) = \frac{2}{N_0} \int_0^T x(t) s(u,t) dt. \quad (2.7)$$

Отже, в разі відомої апіорної ймовірності визначення апостеріорної ймовірності зводиться до обчислення функції $h(u)$, тобто взаємної кореляції між прийнятим сигналом $x(t)$ і переданим (очікуваним) сигналом $s(u, t)$. Іншими словами, *кореляційна обробка сигналу є оптимальною*, тобто вона забезпечує мінімальну ймовірність помилки.

Функція $h(u)$ легко обчислюється, коли всі параметри сигналу відомі. Досягається це за допомогою корелятора або погодженого фільтра, що звичайно має місце при передаванні *дискретних* повідомлень.

У разі передавання *неперервних* повідомлень ми маємо лише деяку апіорну інформацію про сигнал $s(u, t)$. Відомі, наприклад, несуча частота, вид модуляції, ширина спектра сигналу і т. ін. Відсутню частину інформації можна одержати внаслідок спостереження за попередньо

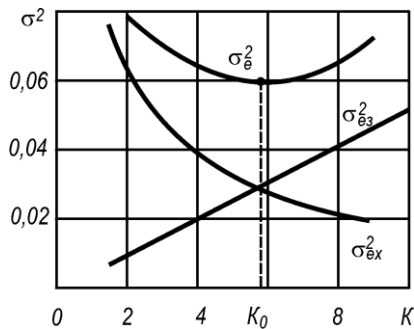


Рис. 2.2. До визначення оптимального коефіцієнта підсилення пристрою (системи) зв'язку

прийнятою реалізацією сигналу $x(t)$. Тому в результаті існує можливість на прийомі визначити оцінку сигналу $s(v, t)$ й обчислити функцію $h(v)$ для цієї оцінки:

$$h(v) = \frac{2}{N_0} \int_0^T x(t)s(v, t) dt. \quad (2.8)$$

Функцію $h(v)$ можна обчислити за допомогою слідувального фільтра СФ зі змінними параметрами (рис. 2.1, б) чи слідувального корелятора (рис. 2.1, в). Кожна з цих схем має основний інформаційний канал, на вихід якого надходить оціночне значення $v(t)$ переданого повідомлення $u(t)$, і канал зворотного зв'язку, за допомогою якого в схемі

на рис. 2.1, в формується опорний сигнал $s(v, t)$, а в схемі на рис. 2.1, б проводиться зміна параметрів фільтра. В схемі на рис. 2.1, б за допомогою керуючого елемента КЕ змінюють параметри фільтра так, щоб він був погоджений з неперервно змінюваним очікуваним сигналом $s(u, t)$. У схемі на рис. 2.1, в за допомогою керуючого елемента КЕ змінюється модульований параметр несучого коливання, формованого генератором Г. При частотній модуляції, наприклад, цим параметром є частота, при часово-імпульсній модуляції – часовий зсув імпульсів. Фільтр нижніх частот ФНЧ у цій схемі виконує роль інтегратора на інтервалі спостереження T , який пов'язаний з максимальною частотою F_m в спектрі

переданого повідомлення співвідношенням $T = 1 / 2F_m$.

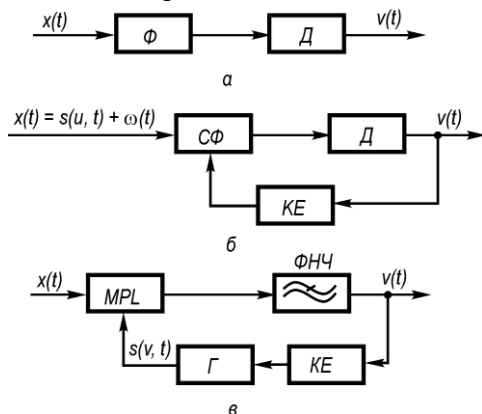


Рис. 2.1. Структурні схеми:
а – оптимального приймача неперервних сигналів; б – приймача зі слідувальним фільтром;
в – слідувального кореляційного приймача

Принцип слідувального приймання не залежить від видів модуляції. Вид модуляції визначає параметр, за яким має здійснюватися слідування. Схеми слідувального приймання дозволяють практично реалізувати завадостійкість, близьку до потенційної, і тим самим знизити поріг завадостійкості до граничного.

До цих пір ми припускали, що амплітуда і фаза несучого коливання сигналу незмінні в часі, а завада має сталу спектральну щільність N_0 . В каналах зі змінними параметрами амплітуда й фаза сигналу змінюються в часі. Тому схеми на рис. 2.1, б і рис. 2.1, в потрібно доповнити пристроєм автоматичного регулювання рівня сигналу й системного фазового автопідстроювання. Якщо значення N_0 невідоме або змінюється в часі, для здійснення оптимального приймання необхідний

спеціальний пристрій вимірювання й вирівнювання інтенсивності. Зокрема, для зосереджених за спектром чи за часом з різко вираженою інтенсивністю завад досить близькою до оптимальної є схема зі стиранням ділянок спектра чи тривалості сигналу, вражених завадою. Отже, в загальному випадку при оптимальному прийманні слід спостерігати і за формою сигналу, і за формою завади. Чим більше число параметрів при цьому врахується, тим вища завадостійкість. При цьому схема приймача має бути адаптивною.

Оптимальний приймач – це пристрій, який відтворює передане повідомлення $u(t)$ з найменшою помилкою. Оскільки корисний сигнал $s(u, t)$ залежить від $u(t)$ нелінійно, то оптимальний приймач також має бути нелінійним пристроєм (нелінійним фільтром). Прикладом такого пристрою (фільтра) є розглянутий вище слідувальний приймач. Таким

чином, теорію оптимального приймання можна розглядати як теорію оптимальної нелінійної фільтрації.

Загальна теорія нелінійної фільтрації охоплює різноманітні й дуже складні задачі. В даний час ця теорія розроблена лише для випадків, коли передане повідомлення $u(t)$ являє собою марковський чи нормальний випадковий процес.

2.2. Синтез пристроїв (систем) зв'язку за заданою структурною схемою

Задача синтезу пристрою зв'язку за заданою його структурною схемою полягає в тому, щоб за відомими характеристиками сигналу й завади визначити оптимальні значення параметрів пристрою зв'язку, при яких сумарна середня квадратична похибка мінімальна. Оскільки структурна схема пристрою відома, то сумарна середня квадратична похибка є функцією параметрів β_i пристрою:

$$\sigma_e^2 = f(\beta_1, \beta_2, \dots, \beta_n). \quad (2.9)$$

Для розрахунку оптимальних значень параметрів необхідно дослідити на мінімум функцію (2.9), розв'язавши наступну систему рівнянь:

$$\frac{\partial}{\partial \beta_i} \sigma_e^2 = 0, \quad i = \overline{1, n}. \quad (2.10)$$

В процесі синтезу пристрою зв'язку, як правило, треба обчислити оптимальне значення тільки настроюваних параметрів (коефіцієнт підсилення та одна чи дві сталі часу коректувального пристрою), кількість яких в пристроях невелика, що полегшує задачу оптимізації.

Приклад 2.1. Знайти оптимальне значення коефіцієнта підсилення пристрою зв'язку, передавальна функція якого в розімкненому стані

$$W_p(s) = \frac{K}{(Ts + 1)s}. \quad (2.11)$$

Розв'язання. На пристрій діють сигнал і завада, спектральні щільності яких $S_x(\omega) = N_x / \omega^2$; $S_3(\omega) = N_3$. Сумарна середня квадратична похибка системи

$$\sigma_e^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[|W_e(j\omega)W_\phi(j\omega)|^2 N_x + |W_3(j\omega)|^2 N_3 \right] d\omega = N_x \frac{1 + KT}{2K} + N_3 \frac{K}{2},$$

де $W_e(j\omega) = \frac{j\omega(1 + j\omega T)}{(j\omega)^2 T + j\omega + K}$ – частотна характеристика похибки системи; $W_\phi(j\omega) = 1/(j\omega)$ –

частотна характеристика формувального фільтра сигналу; $W_3(j\omega) = \frac{K}{(j\omega)^2 T + j\omega + K}$ – частотна характеристика замкненої системи.

Оптимальне значення коефіцієнта підсилення визначимо з умови (2.10), яка для даної задачі має вигляд

$$\frac{\partial}{\partial K} \sigma_e^2 = -\frac{N_x}{2K^2} + \frac{N_3}{2} = 0.$$

Тоді оптимальний коефіцієнт підсилення $K_{\text{опт}} = \sqrt{N_x/N_3}$.

На рис. 2.2 показані криві зміни складових середньої квадратичної похибки пристрою зв'язку в залежності від коефіцієнта підсилення для $N_x = 0,76 \cdot 10^{-4} \text{ рад}^2 \cdot \text{с}$; $N_3 = 0,305 \cdot 10^{-4} \text{ рад}^2 \cdot \text{с}$; $T = 0,1 \text{ с}$.

2.3. Фізичний зміст оптимальної смуги пропускання

Розглянемо пристрій зв'язку з частотною характеристикою (рис. 2.3, а)

$$W(j\omega) = \begin{cases} e^{j\omega} & \text{при } \omega \leq \omega_3; \\ 0 & \text{при } \omega > \omega_3, \end{cases} \quad (2.12)$$

де ω_3 – ширина смуги пропускання пристрою зв'язку.

Якщо сигнал і завада некорельовані, то дисперсія похибки системи з частотною характеристикою (2.12)

$$\sigma_e^2 = \frac{1}{\pi} \int_{\omega_3}^{\infty} S_x(\omega) d\omega + \frac{1}{\pi} \int_0^{\omega_3} S_3(\omega) d\omega = \sigma_{ex}^2 + \sigma_{e3}^2, \quad (2.13)$$

де σ_e – середня квадратична похибка; $S_x(\omega)$ – спектральна щільність сигналу; $S_3(\omega)$ – спектральна щільність завади; σ_{ex} – середня квадратична похибка відносно сигналу; σ_{e3} – середня квадратична похибка через дію завади.

Із виразу (2.13) випливає, що середня квадратична похибка пристрою зв'язку залежить від ширини смуги пропускання. Смугу пропускання пристрою зв'язку, при якій середня квадратична похибка набуває мінімального значення, називають *оптимальною*.

На рис. 2.3, б наведені графіки спектральних щільностей сигналу й завади. Оскільки вони не перекриваються, то оптимальна смуга пропускання системи дорівнює граничній частоті спектра сигналу. В цьому випадку середня квадратична похибка пристрою зв'язку дорівнює нулю, бо всі складові спектра сигналу відтворюються системою і жодна складова спектра завади не проходить на її вихід.

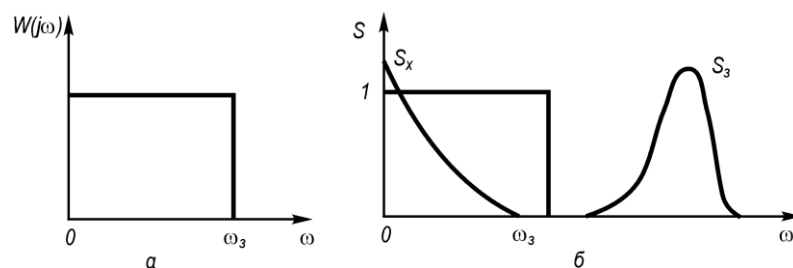


Рис. 2.3. До визначення оптимальної смуги пропускання для спектрів сигналу і завади, які не перекриваються (а), та ідеальна ЛЧХ (б)

Якщо графіки спектральних щільностей сигналу й завади перекриваються, то обидві складові середньої квадратичної похибки системи в (2.13) не дорівнюють нулю. Із рис. 2.4, а видно, що перша складова (середня квадратична похибка відтворення сигналу) визначається тією частиною спектральної щільності сигналу, яка розташована за смугою пропускання системи (площа 1 під графіком спектральної щільності сигналу). З розширенням смуги пропускання ця складова похибки зменшується (рис. 2.4, б). Друга складова середньої квадратичної похибки системи, зумовлена завадою, залежить від тієї частини спектральної щільності завади, яка збігається зі смугою пропускання системи (на рис. 2.4, а площа 2). З

розширенням смуги пропускання середньоквадратична похибка через дію завади зростає (рис. 2.4, б). Оптимальна смуга пропускання системи відповідає мінімальній середньоквадратичній похибці. Для її обчислення здиференціюємо (2.13) за смугою пропускання ω_3 і отриманий результат прирівняємо до нуля. В результаті отримаємо

$$-\frac{1}{\pi} S_x(\omega_{30}) + \frac{1}{\pi} S_n(\omega_{30}) = 0. \quad (2.14)$$

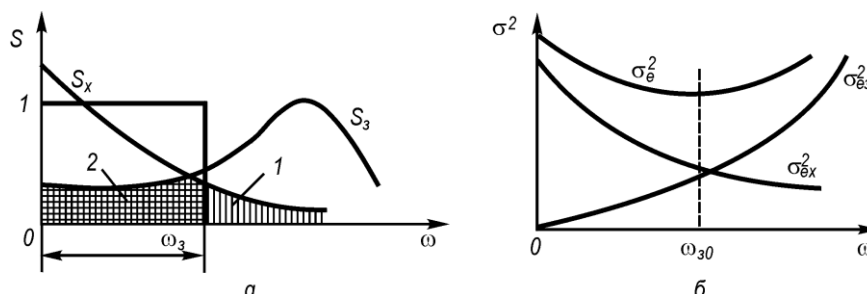


Рис. 2.4. До пояснення оптимальної смуги пропускання пристрою (системи) зв'язку (а) та залежність середньої квадратичної похибки пристрою зв'язку від частоти (б)

Із цього виразу випливає, що оптимальна смуга пропускання системи – це частота, на якій виконується рівність

$$S_x(\omega_{30}) = S_n(\omega_{30}). \quad (2.15)$$

Оскільки частотні характеристики систем зв'язку відрізняються від ідеальної характеристики (2.12), то оптимальна смуга пропускання, знайдена за (2.15), є наближеною. Як показує практика, похибки при цьому не перевищують 10–20 %.

Мінімум сумарної середньої квадратичної похибки системи визначається не тільки смугою пропускання, але й виглядом її частотної характеристики. Тому в загальному випадку синтез системи полягає в пошуку її оптимальних частотних характеристик з умови мінімуму середніх квадратичних похибок при заданих статистичних характеристиках сигналу і завади.

2.4. Синтез оптимальних пристроїв (систем) зв'язку

Сформулюємо постановку задачі синтезу оптимальних пристроїв (систем) зв'язку. На вхід пристрою, який проектується, діють стаціонарні сигнал і завада, автокореляційні функції яких відомі. Математичні сподівання сигналу й завади дорівнюють нулю. Бажаний вихідний сигнал пристрою зв'язку, що синтезується, визначається заданою частотною характеристикою. Необхідно знайти передавальну функцію пристрою, при якій сумарна середня квадратична похибка системи, що наведена на рис. 2.1, а, мінімальна:

$$\sigma_e = \sqrt{e^2(t)} = \min, \quad (2.16)$$

де $e(t) = y_6(t) - y(t)$ – похибка системи; $y_6(t)$ – бажаний вихідний сигнал системи; $y(t)$ – вихідний сигнал системи.

Розглянемо задачу оптимальної фільтрації. В цьому випадку

$$W_6(j\omega) = 1, \text{ тобто } y_6(t) = x(t). \quad (2.17)$$

Згідно з (2.16) дисперсія похибки

$$\sigma^2 = \left[\overline{x(t) - y(t)} \right]^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}^2. \quad (2.18)$$

Дисперсію вихідного сигналу пристрою зв'язку, що синтезується, знайдемо аналогічно дисперсії похибки системи [9]:

$$\sigma_y^2 = \int_{-\infty}^{\infty} \omega(\lambda) \int_{-\infty}^{\infty} \omega(\eta) R_f(\lambda - \eta) d\eta d\lambda, \quad (2.19)$$

де $\omega(\lambda)$ – імпульсна перехідна функція системи; $R_f(\lambda - \eta)$ – автокореляційна функція сумарного вхідного сигналу $f(t) = x(t) + n(t)$; $n(t)$ – завада.

Таким чином,

$$\sigma_{xy}^2 = \int_{-\infty}^{\infty} \omega(\lambda) R_{yf}(\lambda) d\lambda, \quad (2.20)$$

де $R_{yf}(\lambda)$ – взаємна кореляційна функція сигналу з сумарною дією.

Підставивши вирази (2.19) і (2.20) в (2.18), визначимо

$$\sigma_e^2 = \sigma_x^2 + \int_{-\infty}^{\infty} \omega(\lambda) \int_{-\infty}^{\infty} \omega(\eta) R_f(\lambda - \eta) d\eta d\lambda - 2 \int_{-\infty}^{\infty} \omega(\eta) R_{yf}(\lambda) d\lambda. \quad (2.21)$$

Синтез оптимальної системи зводиться до пошуку імпульсної перехідної функції з (2.21). Для розв'язання цієї задачі дамо варіацію імпульсної перехідної функції:

$$\omega(t) = \omega_0(t) + \delta\omega(t), \quad (2.22)$$

де $\omega_0(t)$ – шукана оптимальна імпульсна перехідна функція пристрою зв'язку, що проектується; δ – варіація імпульсної перехідної функції.

Імпульсну перехідну функцію, мінімізувавши дисперсію похибки (2.21), визначимо з умови

$$\left. \frac{\partial \sigma^2}{\partial \delta} \right|_{\delta=0} = 0. \quad (2.23)$$

Підставимо формулу (2.22) в (2.21). Тоді оптимальна імпульсна перехідна функція системи з урахуванням (2.23) має задовольняти рівнянню

$$2 \int_{-\infty}^{\infty} \omega(\lambda) \left[\int_{-\infty}^{\infty} \omega_0(\eta) R_f(\lambda - \eta) d\eta - R_{yf}(\lambda) \right] d\lambda = 0. \quad (2.24)$$

Оскільки неоптимальна імпульсна перехідна функція $\omega(t)$ – функція довільна, то рівняння (2.24) виконується тільки в тому випадку, коли

$$\int_{-\infty}^{\infty} \omega_0(\eta) R_f(t - \eta) d\eta - R_{yf}(t) = 0, \quad (2.25)$$

де змінна λ замінена на t .

Вираз (2.25) називають *рівнянням Вінера–Хопфа*.

Середня квадратична похибка виділення сигналу в усталеному режимі – стала величина, що визначається з виразу (2.21), в якому замість $\omega(\lambda)$ і $\omega(\eta)$ слід підставити початкові значення частот. Тоді з урахуванням (2.25)

$$\sigma_{e\min}^2 = \sigma_x^2 - \int_{-\infty}^{\infty} \omega_0(\lambda) \int_{-\infty}^{\infty} \omega_0(\eta) R_f(\lambda - \eta) d\eta d\lambda. \quad (2.26)$$

Розв'язання рівняння Вінера–Хопфа в часовій області є складним. Значно простіше розв'язати цю задачу в частотній області, тобто знайти оптимальну частотну характеристику пристрою зв'язку. З цією метою застосуємо до рівняння (2.25) перетворення Фур'є. В результаті отримаємо

$$W_0(j\omega)S(j\omega) = S_{xf}(\omega). \quad (2.27)$$

З цього рівняння знайдемо оптимальну частотну характеристику:

$$W_0(j\omega) = \frac{S_{xf}(\omega)}{S_f(\omega)} = \frac{S_{xf}(\omega)}{\psi(j\omega)\psi(-j\omega)}, \quad (2.28)$$

де $S_f(\omega)$ – спектральна щільність сумарного сигналу на вході проективної системи; $S_{xf}(\omega)$ – взаємна спектральна щільність сигналу з сумарним сигналом; $\psi(j\omega)$ – функція, всі полюси якої на площині комплексної змінної $p = j\omega$ розташовані в лівій півплощині; $\psi(-j\omega)$ – функція, всі полюси якої розташовані в правій півплощині.

У загальному випадку

$$\begin{aligned} S_{xf}(\omega) &= S_x(\omega) + S_{x3}(\omega); \\ S_f(\omega) &= S_x(\omega) + S_{x3}(\omega) + S_3(\omega) + S_{3x}(\omega). \end{aligned} \quad (2.29)$$

Якщо сигнал і завада некорельовані, то

$$S_{xf}(\omega) = S_x(\omega); \quad S_f(\omega) = S_x(\omega) + S_3(\omega). \quad (2.30)$$

Із виразів (2.28)–(2.30) випливає, що оптимальна частотна характеристика виділяє складові сигналу на частотах з достатньо великою спектральною щільністю сигналу і послаблює складові сигналу на частотах з максимальною спектральною щільністю завади.

З урахуванням оптимальної частотної характеристики мінімальна дисперсія похибки для некорельованих сигналу й завади

$$\sigma_{e\min}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_x(\omega)S_3(\omega)}{S_x(\omega) + S_3(\omega)} d\omega. \quad (2.31)$$

Таким чином, якщо спектри сигналу й завади не перекриваються, то середня квадратична похибка може дорівнювати нулю.

Спектральна щільність є парною функцією відносно частоти ω , тому полюси характеристики (2.28) розташовані на площині $p = j\omega$ як зліва, так і справа від уявної осі. Тому знайдена оптимальна частотна характеристика (2.28) відповідає фізично нереалізованій

нестійкій системі. Надалі синтез оптимальної системи зводиться до визначення *реалізовної оптимальної частотної характеристики*, найбільш близької до отриманої нереалізовної.

Реалізовна оптимальна частотна характеристика визначається так:

$$W_0(j\omega) = \frac{1}{\psi(j\omega)} \left[\frac{S_{xf}(\omega)}{\psi(-j\omega)} \right]^+, \quad (2.32)$$

де операція \dots^+ означає виділення доданків, полюси яких на площині комплексної змінної $p = j\omega$ розташовані зліва від уявної осі.

У випадку, що часто зустрічається в інженерній практиці, коли функція

$$\frac{S_{xf}(\omega)}{\psi(-j\omega)} \quad (2.33)$$

є дробово-раціональною функцією відносно частоти, виділення доданків з полюсами, розташованими зліва від уявної осі, здійснюється розкладанням (2.33) на прості дроби:

$$\frac{S_{xf}(\omega)}{\psi(-j\omega)} = \sum_{i=1}^n \frac{\alpha_i}{\omega - \omega_i}, \quad (2.34)$$

Де

$$\alpha_i = \left(\omega - \frac{1}{\lambda_i} \right) \frac{S_{xf}(\omega)}{\psi(-j\omega)}; \quad (2.35)$$

λ_i – полюси виразу (2.33).

Дисперсія похибки в системі з частотною характеристикою (2.32) буде більшою від значення, знайденого за виразом (2.31), і значення, знайденого без урахування фізичної реалізації оптимальної системи. Однак серед реалізовних пристроїв зв'язку система з характеристикою (2.32) забезпечує найменшу середньоквадратичну похибку виділення сигналу з його поєднання з завадою.

Згідно з (2.32) передавальна функція проективної оптимальної системи

$$W_0(s) = W_0(j\omega) \Big|_{j\omega=s}. \quad (2.36)$$

Тоді передавальна функція послідовного коректувального пристрою, що вмикається в ланку сигналу похибки,

$$W_k(s) = W_{p.o}(s) / W_{p.v}(s), \quad (2.37)$$

де $W_{p.o}(s)$ – передавальна функція розімкненої оптимальної системи; $W_{p.v}(s)$ – передавальна функція вихідної частини, складена з функціонально необхідних елементів пристрою (системи).

Приклад 2.2. Знайти передавальну функцію послідовного коректувального пристрою в оптимальній системі автопідстроювання частоти для випадку, коли спектральні щільності сигналу й завади визначаються виразами

$$S_x(\omega) = \frac{N_x}{T_x^2 \omega^2 + 1}; \quad S_3(\omega) = N_3. \quad (2.38)$$

Розв'язання. Інерційністю підсилювача і частотного дискримінатора системи знехтуємо. Тоді передавальна функція вихідної системи автопідстроювання частоти визначається виразом

$$W_{p.б}(s) = \frac{K_{p.б}}{T_r s + 1}, \quad (2.39)$$

де T_r – стала часу гетеродина; $K_{p.б}$ – коефіцієнт підсилення сигналу на початковій (вихідній) системі автопідстроювання частоти.

Спектральна щільність поєднання сигналу з завадою

$$S_f(\omega) = S_x(\omega) + S_3(\omega) = B \frac{1 + j\omega T_0}{1 + j\omega T_x} \cdot \frac{1 - j\omega T_0}{1 - j\omega T_x}, \quad (2.40)$$

де $B = N_x + N_3$; $T_0^2 = T_x^2 \frac{N_x}{N_x + N_3}$.

Отже,

$$\psi(j\omega) = \sqrt{B} \frac{1 + j\omega T_0}{1 + j\omega T_x}; \quad \psi(-j\omega) = \sqrt{B} \frac{1 - j\omega T_0}{1 - j\omega T_x}. \quad (2.41)$$

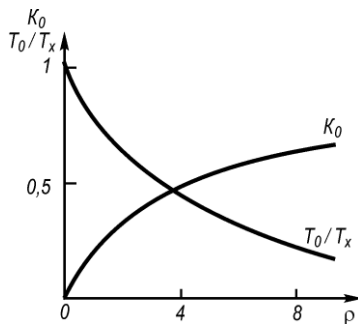


Рис. 2.5. Залежність оптимальних параметрів системи від відношення сигнал–шум

Відповідно до (2.32)

$$W_0(j\omega) = \frac{1}{\psi(j\omega)} \left[\frac{\alpha_1}{1 + j\omega T_x} + \frac{\alpha_2}{1 - j\omega T_0} \right]^+ \frac{N_x}{\sqrt{B}} = \frac{1}{\psi(j\omega)} \cdot \frac{\alpha_1}{1 + j\omega T_x} \cdot \frac{N_x}{\sqrt{B}}. \quad (2.42)$$

За методом невизначених коефіцієнтів

$$\alpha_1 = T_x / (T_x + T_0). \quad (2.43)$$

Згідно з (2.42) і (2.43) оптимальна частотна характеристика проективної системи автопідстроювання частоти

$$W_0(j\omega) = \frac{K_0}{T_j \omega + 1}, \quad (2.44)$$

де $K_0 = \alpha_1 \frac{N_x}{B}$.

Для очевидності виразимо параметри оптимальної частотної характеристики через коефіцієнт, рівний відношенню рівнів спектральної щільності сигналу й завади. В результаті отримаємо

$$K_0 = 1 - 1/\sqrt{1 + \rho}, \quad T_0 = T/\sqrt{1 + \rho}, \quad (2.45)$$

де $\rho = N_x / N_3$.

На рис. 2.5 показані залежності коефіцієнта підсилення K_0 і сталої змінної T_0/T_x від коефіцієнта ρ , з яких видно, що зі зростанням ρ стала часу оптимальної системи зменшується (її граничне значення дорівнює нулю), а коефіцієнт підсилення зростає (його граничне значення дорівнює одиниці). Таким чином, за відсутності завади оптимальна система є безінерційною ланкою (дисперсія похибки дорівнює нулю).

Дисперсію сумарної похибки в оптимальній системі автопідстроювання частоти визначимо за формулою [9]

$$\sigma_e^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[|1 - W_0(j\omega)|^2 S_x(\omega) + |W_0(j\omega)|^2 S_z(\omega) \right] d\omega = \frac{T_0 - (1 - K_0)^2 T_x}{2T_x(T_0 + T_x)} N_x + \frac{K_0^2}{2T_0} N_z.$$

Взявши до уваги введене значення ρ , матимемо

$$\sigma_e^2 = \frac{N_x}{2T_x \sqrt{1+\rho}} \left[1 + \frac{\rho}{(\sqrt{1+\rho})^2} \right]. \quad (2.46)$$

Обчислимо дисперсію похибки за формулою (2.31):

$$\sigma_{e\min}^2 = \frac{N_x}{2T_x \sqrt{1+\rho}}.$$

Отримане значення менше від обчисленого за оптимальною частотною характеристикою. Це пояснюється тим, що у формулі (2.31) не враховано фізичну реалізованість системи.

Передавальна функція розімкненої оптимальної системи автопідстроювання частоти

$$W_{p0}(s) = \frac{K_{p0}}{Ts + 1},$$

де $K_{p0} = K_0/(1 - K_0)$; $T = T_0/(1 - K_0)$.

Передавальна функція послідовності коректувального пристрою відповідно до виразу (2.37)

$$W_k(s) = K_k \frac{T_s + 1}{Ts + 1}.$$

Коректувальний пристрій в оптимальній системі автопідстроювання частоти вмикають між підсилувачем і дискримінатором, він може бути реалізований RC-ланкою з відставанням за фазою.

Підбиваючи підсумок, зазначимо, що теорія лінійної фільтрації Колмогорова–Вінера для стаціонарних процесів, викладена вище, має один істотний недолік: практична реалізація оптимальних фільтрів супроводжується великими труднощами. А якщо врахувати, що реальні модульовані сигнали не є стаціонарними, то побудова оптимального фільтра на базі лінійної теорії стає практично неможливою. Оскільки лінійний фільтр оптимальний для сигналу $s(u, t)$, а не для самого повідомлення $u(t)$, то при проектуванні нестационарних оптимальних систем фільтр Колмогорова–Вінера не застосовується. Розв'язання нестационарних задач оптимальних систем зв'язку базується на оптимальних фільтрах Калмана.

2.5. Синтез оптимальних фільтрів Калмана методом простору станів

Ефективним шляхом синтезу оптимальних нестационарних фільтрів є запропонований Калманом і Б'юсі метод простору станів. Відповідно до методики, викладеної в процес, що виділяється (повідомлення), зображується у вигляді компоненти чи лінійної комбінації компонент векторного процесу $x(t)$. Складові процесу $x(t)$ визначаються розв'язанням системи лінійних диференціальних рівнянь, збурюваннями в яких є білі шуми. Описаний векторний процес є окремим випадком багатовимірною марковського процесу. Тому підхід Калмана–Б'юсі до синтезу оптимальних лінійних фільтрів може бути названий *марковською теорією оптимальної лінійної фільтрації*. Будемо, однак, для зручності, а також зважаючи на розповсюджену термінологію, називати його *синтезом за методом простору станів*.

Зазначений підхід до синтезу оптимальних лінійних фільтрів тісно поєднується з загальною теорією оптимальної нелінійної фільтрації марковських процесів, розробленою Р. Л. Стратоновичем [35]. У [5] показано, як основні співвідношення для оптимальних фільтрів Калмана можуть бути отримані з результатів марковської теорії оптимальної нелінійної фільтрації. Оскільки задача оптимальної лінійної фільтрації простіша за задачу нелінійної фільтрації, то відповідний до неї математичний апарат розроблений на даний час більш детально.

Опис повідомлень. Розгляд синтезу оптимальних лінійних фільтрів методом простору станів почнемо з пояснення можливості зображення повідомлення у вигляді компоненти або лінійної комбінації компонент марковського процесу.

Нехай, наприклад, процес $\lambda(t)$, який підлягає виділенню за допомогою оптимального фільтра, є розв'язком лінійного стохастичного диференціального рівняння n -го порядку

$$A(p)\lambda(t) = k_1 B(p)n(t), \quad (2.47)$$

де $A(p) = p^n + a_{n-1}p^{n-1} + \dots + a_0$; $B(p) = b_m p^m + b_{m-1}p^{m-1} + \dots + b_1 p + 1$; n, t – білий шум; $m \leq n-1$. Процес $S(t)$ відповідно до (2.47) можна розглядати як результат впливу білого шуму $n(t)$ на формувальний фільтр з операторним коефіцієнтом передачі

$$K_f(p) = k_1 B(p) / A(p). \quad (2.48)$$

Зображення повідомлення $\lambda(t)$ у вигляді розв'язку рівняння (2.47) є достатньо універсальним. Так, якщо при синтезі оптимального лінійного фільтра задана спектральна щільність $S_\lambda(\omega)$ повідомлення $\lambda(t)$, яка описується дробово-раціональною функцією, то процес $S(t)$ можна розглядати як ефект на виході формувального фільтра (2.48) в усталеному режимі й відображати диференціальним рівнянням вигляду (2.47). Для цього достатньо прирівняти $S_\lambda(\omega) = |K_f(j\omega)|^2 = |k_1 B(j\omega) / A(j\omega)|^2$, визначити з цієї умови комплексний коефіцієнт передачі $K_f(j\omega)$ і відновити за $K_f(j\omega)$ диференціальне рівняння формувального фільтра.

Замінивши змінні в рівнянні (2.47), його можна подати системою лінійних диференціальних рівнянь першого порядку, до правих частин яких входить білий шум, але не його похідні. Якщо $B(p) = 1$, то вказаний перехід виконується за допомогою заміни змінних:

$$\left. \begin{aligned} dx_1(t)/dt &= x_2(t); \\ dx_2(t)/dt &= x_3(t); \\ \dots\dots\dots \\ dx_{n-1}(t)/dt &= x_n(t), \end{aligned} \right\} \quad (2.49)$$

де $x_1(t) = \lambda(t)$.

Рівняння (2.47) у нових змінних записується у вигляді

$$\frac{dx_n}{dt} = -a_{n-1}x_n - \dots - a_0x_1 + k_1n(t). \quad (2.50)$$

Сукупність компонент $x_1(t), x_2(t), \dots, x_n(t)$, що задовольняють рівняння (2.49), (2.50), утворить вектор-функцію, яка описує багатовимірний марковський процес $\mathbf{x}(t)$ в n -вимірному просторі (просторі станів). Систему рівнянь (2.49) можна записати у векторній формі:

$$d\mathbf{x}/dt = \mathbf{F}\mathbf{x}(t) + \mathbf{H}n(t), \quad (2.51)$$

де $\mathbf{x}(t)$ – n -вимірний вектор-стовпець із компонентами $x_1(t), x_2(t), \dots, x_n(t)$; \mathbf{F} і \mathbf{H} – відповідно матриця і вектор-стовпець вигляду

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & 1 \\ -a_0 & \dots & -a_1 & \dots & -a_{n-1} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ k_1 \end{bmatrix}. \quad (2.52)$$

Повідомлення $\lambda(t) = x_1(t)$ при такому підході є першою компонентою векторного процесу $\mathbf{x}(t)$ і пов'язане з ним співвідношенням

$$\lambda(t) = x_1(t) = \mathbf{c}_\lambda \mathbf{x}(t), \quad (2.53)$$

де \mathbf{c}_λ – n -вимірний вектор-рядок вигляду $\mathbf{c}_\lambda = 1 \ 0 \ 0 \ \dots \ 0$.

Якщо в рівнянні (2.47) коефіцієнти $B \ p = b_1, b_2, \dots, b_m$ відрізняються від нуля, то зображення його у вигляді системи рівнянь першого порядку з білими шумами в правих частинах ускладнюється. Дійсно, в цьому випадку із заміною змінних (2.49) в рівняннях з'являються похідні білого шуму. Існує декілька шляхів, які дозволяють позбутися похідних білого шуму на випадок, коли коефіцієнти b_1, b_2, \dots, b_m не дорівнюють нулю. Один із них полягає в наступному. Вводяться змінні:

$$\left. \begin{array}{l} \lambda(t) = x_1(t); \\ dx_1/dt = x_2(t) + c_1n(t); \\ \dots\dots\dots \\ dx_{n-1}/dt = x_n(t) + c_{n-1}n \ t, \end{array} \right\} \quad (2.54)$$

де c_1, c_2, \dots, c_{n-1} – сталі коефіцієнти, які вибираються надалі так, щоб усунути похідні білого шуму. Із (2.54) випливає, що

$$\frac{d^k x}{dt^k} = x_{k+1} + \sum_{i=0}^{k-1} c_{k-i} \frac{d^i n(t)}{dt^i} \quad \text{при } 1 \leq k \leq n-1,$$

$$\frac{d^n x}{dt^n} = \frac{dx_n}{dt} + \sum_{i=1}^{n-1} c_{n-i} \frac{d^i n(t)}{dt^i}. \quad (2.55)$$

З урахуванням (2.54) і (2.55) рівняння (2.47) записується у вигляді

$$\frac{dx_n}{dt} + \sum_{i=1}^{n-1} c_{n-i} \frac{d^i n(t)}{dt^i} + \sum_{k=0}^{n-1} a_k x_{k+1} + \sum_{i=0}^{n-2} \frac{d^i n(t)}{dt^i} \sum_{k=1}^{n-1-i} a_{k+i} c_k = \sum_{i=0}^m b_i \frac{d^i n(t)}{dt^i}. \quad (2.56)$$

Похідні білого шуму в рівнянні (2.56) усуваються в разі виконання наступних рівностей:

$$c_{n-i} + \sum_{k=1}^{n-1-i} a_{k+i} c_k = b_i, \quad i = 1, 2, \dots, n-1. \quad (2.57)$$

Якщо послідовно розв'язувати рівняння (2.57), починаючи з $i = n-1$, знайдемо шукані коефіцієнти c_1, c_2, \dots, c_{n-1} .

При виконанні рівностей (2.57) система рівнянь (2.54) і (2.56) набере вигляду

$$\left. \begin{aligned} dx_1 / dt &= x_2(t) + c_1 n(t), \\ \dots\dots\dots \\ dx_{n-1} / dt &= x_n(t) + c_{n-1} n(t), \\ dx_n / dt &= -\sum_{k=0}^{n-1} a_k x_{k+1} + \left(k_1 - \sum_{k=1}^{n-1} a_k c_k \right) n(t), \end{aligned} \right\} \quad (2.58)$$

де $x_1(t) = \lambda(t)$. Систему рівнянь (2.58) також можна записати у вигляді векторного рівняння (2.51), в якому матриця \mathbf{F} визначається виразом (2.52), а векторний стовпець

$$\mathbf{H} = \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_{n-1} \\ k_1 - \sum_{k=1}^{n-1} a_k c_k \end{bmatrix}.$$

Повідомлення $\lambda(t)$ в даному випадку також визначається через векторний процес $\mathbf{x}(t)$ рівністю (2.53).

Існує інший спосіб складання системи стохастичних диференціальних рівнянь першого порядку з завадами у вигляді білого шуму. В цьому випадку вводиться нова змінна $x_1(t)$, пов'язана з $\lambda(t)$ співвідношенням

$$\lambda(t) = B(p)x_1(t). \quad (2.59)$$

Після підстановки (2.59) в рівняння (2.47) останнє записується так:

$$A(p)x_1(t) = k_1 n(t). \quad (2.60)$$

Рівнянню (2.60) відповідає система диференціальних рівнянь першого порядку:

$$\left. \begin{aligned} dx_1 / dt &= x_2(t), \\ \dots\dots\dots \\ dx_{n-1} / dt &= x_n(t), \\ dx_n / dt &= -\sum_{k=0}^{n-1} a_k x_{k+1} + k_1 n(t). \end{aligned} \right\} \quad (2.61)$$

Ця система може бути записана у вигляді векторного рівняння

$$dx/dt = \mathbf{F}\mathbf{x}(t) + \mathbf{H}(t)n(t),$$

де $\mathbf{x}(t)$ – n -вимірний векторний процес; матриця \mathbf{F} і вектор-стовпець \mathbf{H} визначаються виразами (2.52). На відміну від попереднього випадку, повідомлення $\lambda(t)$ при даному способі заміни змінних виражається, як впливає з (2.59), (2.61), через кілька компонент векторного процесу $\mathbf{x}(t)$. Так, якщо $B(p) = 1 + T_1 p$, то

$$\lambda(t) = x_1(t) + T_1 x_2(t). \quad (2.62)$$

Зв'язок між повідомленнями $\lambda(t)$ і n -вимірним вектором $\mathbf{x}(t)$ встановлюється виразом

]

$$\lambda(t) = \mathbf{c}_\lambda \mathbf{x}(t),$$

де $\mathbf{c}_\lambda = [1 \quad T_1 \quad 0 \quad \dots \quad 0]$ – вектор-рядок, який містить n компонент.

Другий спосіб складання системи стохастичних диференціальних рівнянь дозволяє отримати у векторному рівнянні простіший порівняно з першим способом вираз для вектора-стовпця \mathbf{H} . Однак це досягається ціною ускладнення вектора-рядка \mathbf{c}_λ .

Проведений розгляд показує, як можна зобразити повідомлення у вигляді окремої компоненти чи лінійної комбінації кількох компонент векторного марковського процесу, якщо це повідомлення або описується лінійним диференціальним рівнянням n -го порядку достатньо загального вигляду (2.47), або задається дробово-раціональною спектральною щільністю.

На практиці виникають також задачі, в яких процес (повідомлення), що виділяється, описується диференціальним рівнянням чи системою диференціальних рівнянь зі змінними коефіцієнтами. При синтезі згладжуючих ланок складних пристроїв зв'язку на входи фільтра, який оптимізується, надходять разом із шумом кілька компонент процесу, який фільтрується, або фільтрівний процес в суміші з різними завадами. Щоб охопити ці випадки, а також ряд інших можливих випадків, загальна постановка задачі синтезу оптимального лінійного фільтра методом простору станів формується наступним чином.

Постановка задачі. Передбачається, що векторний процес $\mathbf{x}(t)$, який описує повідомлення $\lambda(t)$ в n -вимірному просторі станів, задовольняє векторному диференціальному рівнянню

$$dx/dt = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{H}(t)\mathbf{n}(t), \quad (2.63)$$

де $\mathbf{x}(t)$ – вектор-функція, що містить n компонент; $\mathbf{F}(t)$ – квадратна матриця розміром $n \times n$, яка залежить у загальному випадку від часу; $\mathbf{n}(t)$ – вектор білих шумів, який складається з l компонент; $\mathbf{H}(t)$ – матриця розміром $n \times l$.

Кореляційна матриця процесу, коли шум $\mathbf{n}(t)$ в загальному випадку передбачається нестационарним, має вигляд

$$\mathbf{R}_x(t, u) = M \mathbf{n}(t)\mathbf{n}^T(u) = \mathbf{Q}(t)\delta(t-u), \quad (2.64)$$

де $\mathbf{Q}(t)$ – квадратна матриця розміром $l \times l$.

Нагадаємо, що в найпростішому випадку повідомлення $\lambda(t)$ є однією з компонент векторного процесу $\mathbf{x}(t)$. Однак, як видно, наприклад, з формули (2.62), можливі й більш складні ситуації, коли повідомлення $\lambda(t)$ є комбінацією кількох компонент процесу $\mathbf{x}(t)$. У задачах одночасної фільтрації кількох повідомлень компоненти вектора $\mathbf{x}(t)$ можуть бути

пов'язані з різними повідомленнями. Обмежимося надалі обговоренням фільтрації одного повідомлення, який виражається через процес $\mathbf{x}(t)$ співвідношенням

$$\lambda(t) = \mathbf{c}_\lambda \mathbf{x}(t),$$

де \mathbf{c}_λ – вектор-рядок, що містить n елементів.

Фільтр, який оптимізується, в загальному випадку має декілька входів, на які можуть надходити в суміші з шумом повідомлення $\lambda(t)$ та його похідні. Сукупність процесів на m входах фільтра утворює спостережуваний векторний процес

$$\boldsymbol{\varphi}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{n}(t), \quad (2.65)$$

де $\boldsymbol{\varphi}(t)$ – вектор-функція, яка складається із m елементів; $\mathbf{C}(t)$ – прямокутна матриця розміром $m \times n$; $\mathbf{n}(t)$ – m -вимірний вектор білих шумів (шумів вимірювання) з кореляційною матрицею

$$\mathbf{R}(t, u) = \mathbf{R}(t)\delta(t - u); \quad (2.66)$$

тут $\mathbf{R}(t)$ – додатна визначена квадратна матриця розміром $m \times m$. В багатьох випадках процес $\boldsymbol{\varphi}(t)$ скалярний і являє собою адитивну суміш повідомлення $\lambda(t)$ і шуму. В цьому випадку $\mathbf{C}(t)$ – матриця розміром $1 \times n$, тобто вектор-рядок, що збігається з \mathbf{c}_λ .

Спостережуваний процес $\boldsymbol{\varphi}(t)$ надходить на вхід фільтра, починаючи з моменту $t = t_0$. Значення кореляційної матриці процесу $\mathbf{x}(t)$ у момент $t = t_0$ вважається відомим. Процес $\mathbf{x}(t)$ і завада $\mathbf{n}(t)$ вважаються незалежними. Фільтр має забезпечити виділення повідомлення $\lambda(t)$ з мінімальною середньоквадратичною похибкою в будь-який момент часу $t > t_0$.

Рівняння оптимального фільтра. Рівняння, які визначають структуру й характеристики оптимального фільтра, при розглянутій постановці задачі можуть бути отримані різними способами. Оскільки виведення їх досить громіздке, обмежимося тут обговоренням кінцевих результатів.

Структура оптимального фільтра описується векторним диференціальним рівнянням

$$d\mathbf{x}_0 / dt = \mathbf{G}_0(t, t) \boldsymbol{\varphi}(t) - \mathbf{C}(t)\mathbf{x}_0(t) + \mathbf{F}(t)\mathbf{x}_0(t), \quad (2.67)$$

де $\mathbf{x}_0(t)$ – векторний процес на виході оптимального фільтра, який відтворює з деякою помилкою процес $\mathbf{x}(t)$; $\mathbf{G}_0(t, t)$ – матриця змінних коефіцієнтів, яка утворюється з матриці $\mathbf{G}_0(t, u)$ при $t = u$; $\mathbf{G}_0(t, u)$ – матриця імпульсних перехідних функцій оптимального фільтра, яку також називають *перехідною матрицею*.

Початкове значення вектора $\mathbf{x}_0(t_0)$ в рівнянні (2.67) звичайно вважається нульовим.

Нагадаємо, що коли дія й реакція на виході фільтра описуються скалярними функціями, перехідна матриця містить єдину компоненту – звичайну імпульсну перехідну функцію фільтра. Якщо фільтр є багатовимірним, то компонента перехідної матриці, що знаходиться на перетині i -го рядка та j -го стовпця, – це імпульсна перехідна функція, яка пов'язує змінення i -ї компоненти вихідного процесу при дії у вигляді одиничного дельта-імпульсу з його j -ю компонентою. Останнє еквівалентне одиничному початковому значенню i -ї компоненти процесу на виході фільтра при нульових початкових значеннях решти компонент.

Матриця $\mathbf{G}_0(t, t)$, що входить у (2.67), визначається виразом

$$\mathbf{G}_0(t, t) = \mathbf{E}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t), \quad (2.68)$$

де $\mathbf{R}^{-1}(t)$ – матриця, обернена до матриці $\mathbf{R}(t)$; $\mathbf{E}(t)$ – матриця дисперсій помилок фільтрації:

$$\mathbf{E}(t) = M \mathbf{x}(t) - \mathbf{x}_0(t) \mathbf{x}(t) - \mathbf{x}_0(t)^T. \quad (2.69)$$

Матриця $\mathbf{E}(t)$ – квадратна симетрична розміром $n \times n$. Її елементами є дисперсії та взаємні дисперсії помилок фільтрації окремих складових вектора $\mathbf{x}(t)$. Іноді $\mathbf{E}(t)$ називають *матрицею коваріацій*.

Зміна помилок фільтрації в часі описується рівнянням

$$\begin{aligned} d\mathbf{E}/dt = & \mathbf{F}(t)\mathbf{E}(t) + \mathbf{E}(t)\mathbf{F}^T(t) - \\ & - \mathbf{E}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{E}(t) + \mathbf{H}(t)\mathbf{Q}\mathbf{H}^T(t). \end{aligned} \quad (2.70)$$

Рівняння (2.70) є матричним нелінійним диференціальним рівнянням Ріккати. Для його розв'язку необхідно задати початкове значення $\mathbf{E}(t_0)$ матриці дисперсій помилок. Якщо в момент $t = t_0$ процес $\mathbf{x}_0(t_0)$ на виході фільтра дорівнює нулю, то, як видно з (2.69), матриця $\mathbf{E}(t_0)$ дорівнює матриці дисперсій компонент процесу $\mathbf{x}(t)$, що фільтрується в момент $t = t_0$, тобто $\mathbf{E}(t_0) = M \mathbf{x}(t_0)\mathbf{x}^T(t_0)$.

Сукупність виразів (2.67), (2.68), (2.70) повністю визначає структуру й параметри оптимального фільтра. Структурна схема такого фільтра, побудована за рівнянням (2.67), зображена на рис. 2.6. Подвійними лініями на цьому рисунку позначені зв'язки між векторними процесами. При розв'язанні конкретних задач рівняння (2.67) доцільно подати у вигляді системи диференціальних рівнянь першого порядку. За цією системою рівнянь легко будуватися, як показано в п. 2.6, структура оптимального фільтра.

Найбільш складним етапом синтезу оптимальних фільтрів є розв'язання рівняння (2.70). Як правило, воно вимагає застосування ЕОМ.

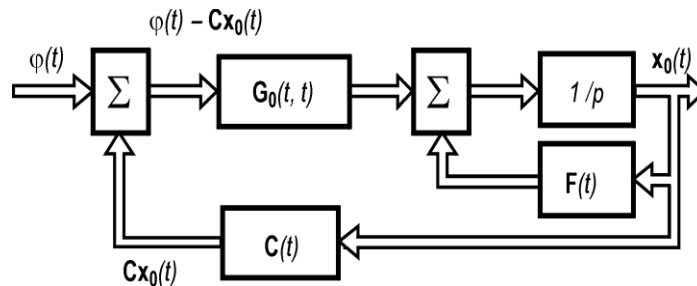


Рис. 2.6. Структурна схема фільтра, котрий побудований за рівнянням (2.67)

Одним із можливих шляхів розв'язання рівняння (2.70) є заміна нелінійного диференціального рівняння з матрицею розміром $n \times n$ системою лінійних диференціальних рівнянь, еквівалентних лінійному матричному диференціальному рівнянню з матрицею розміром $2n \times 2n$. Ця заміна виконується зображенням $\mathbf{E}(t)$ у вигляді

$$\mathbf{v}_1(t) = \mathbf{E}(t)\mathbf{v}_2(t), \quad (2.71)$$

де $\mathbf{v}_1(t)$ і $\mathbf{v}_2(t)$ – n -вимірні вектори, що задовольняють рівнянням

$$d\mathbf{v}_1/dt = \mathbf{F}\mathbf{v}_1 + \mathbf{H}\mathbf{Q}\mathbf{H}^T\mathbf{v}_2, \quad (2.72)$$

$$d\mathbf{v}_2 / dt = \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} \mathbf{v}_1 - \mathbf{F}^T \mathbf{v}_2. \quad (2.73)$$

Щоб переконатися в слушності виконуваного переходу, подамо (2.72), враховуючи (2.71), так:

$$\dot{\mathbf{E}}(t) \mathbf{v}_2(t) + \mathbf{E}(t) \dot{\mathbf{v}}_2(t) = \mathbf{F} \mathbf{E}(t) \mathbf{v}_2(t) + \mathbf{H} \mathbf{Q} \mathbf{H}^T \mathbf{v}_2(t). \quad (2.74)$$

Замінивши $\mathbf{v}_2(t)$ в (2.74) відповідно до (2.73), отримаємо

$$\left[\dot{\mathbf{E}}(t) - \mathbf{F} \mathbf{E}(t) - \mathbf{E}(t) \mathbf{F}^T + \mathbf{E}(t) \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} \mathbf{E}(t) - \mathbf{H} \mathbf{Q} \mathbf{H}^T \right] \mathbf{v}_2(t) = 0. \quad (2.75)$$

При виконанні рівняння Ріккати рівняння (2.75) задовольняється тотожно, що й доводить правомірність виконуваного переходу.

Виразимо матрицю дисперсій $\mathbf{E}(t)$ через перехідну матрицю $\mathbf{T}(t, t_0)$ системи рівнянь (2.72), (2.73). Матриця $\mathbf{T}(t, t_0)$ встановлює наступний зв'язок:

$$\begin{bmatrix} \mathbf{v}_1(t) \\ \mathbf{v}_2(t) \end{bmatrix} = \mathbf{T}(t, t_0) \begin{bmatrix} \mathbf{v}_1(t_0) \\ \mathbf{v}_2(t_0) \end{bmatrix}. \quad (2.76)$$

Зобразимо матрицю $\mathbf{T}(t, t_0)$ розміром $2n \times 2n$ у вигляді блочної матриці, елементи якої – матриці розміром $n \times n$:

$$\mathbf{T}(t, t_0) = \begin{bmatrix} \mathbf{T}_{11}(t, t_0) & \mathbf{T}_{12}(t, t_0) \\ \mathbf{T}_{21}(t, t_0) & \mathbf{T}_{22}(t, t_0) \end{bmatrix}.$$

На підставі (2.76)

$$\begin{aligned} \mathbf{v}_1(t) &= \mathbf{T}_{11}(t, t_0) \mathbf{v}_1(t_0) + \mathbf{T}_{12}(t, t_0) \mathbf{v}_2(t_0), \\ \mathbf{v}_2(t) &= \mathbf{T}_{21}(t, t_0) \mathbf{v}_1(t_0) + \mathbf{T}_{22}(t, t_0) \mathbf{v}_2(t_0). \end{aligned}$$

Звідси випливає, що

$$\begin{aligned} \mathbf{E}(t) [\mathbf{T}_{21}(t, t_0) \mathbf{v}_1(t_0) + \mathbf{T}_{22}(t, t_0) \mathbf{v}_2(t_0)] &= \\ = \mathbf{T}_{11}(t, t_0) \mathbf{v}_1(t_0) + \mathbf{T}_{12}(t, t_0) \mathbf{v}_2(t_0). \end{aligned} \quad (2.77)$$

Помноживши доданки, що входять до обох частин рівності (2.77), на $\mathbf{v}_2^{-1}(t_0)$, та врахувавши, що $\mathbf{v}_1(t_0) \mathbf{v}_2^{-1}(t_0) = \mathbf{E}(t_0)$, після перетворення отримаємо

$$\mathbf{E}(t) [\mathbf{T}_{21}(t, t_0) \mathbf{E}(t_0) + \mathbf{T}_{22}(t, t_0)] = \mathbf{T}_{11}(t, t_0) \mathbf{E}(t_0) + \mathbf{T}_{12}(t, t_0). \quad (2.78)$$

Для визначення матриці $\mathbf{E}(t)$ необхідно знати початкове значення $\mathbf{E}(t_0)$ і перехідну матрицю $\mathbf{T}(t, t_0)$. Щоб знайти матрицю $\mathbf{T}(t, t_0)$, запишемо систему рівнянь (2.72) і (2.73) у такому вигляді:

$$d\mathbf{v} / dt = \mathbf{A} \mathbf{v}(t), \quad (2.79)$$

де $\mathbf{v}(t)$ – складений вектор, компонентами якого є вектори $\mathbf{v}_1(t)$ і $\mathbf{v}_2(t)$; \mathbf{A} – блочна матриця:

$$\mathbf{A} = \begin{bmatrix} \mathbf{F}(t) & \mathbf{H}(t)\mathbf{Q}(t)\mathbf{H}^T(t) \\ \mathbf{C}^T(t)\mathbf{R}^{-1}\mathbf{C}(t) & -\mathbf{F}^T(t) \end{bmatrix}. \quad (2.80)$$

Існують різні методи обчислення перехідної матриці для системи, описуваної рівнянням (2.79). Якщо розмір матриці \mathbf{A} невеликий, а її складові – матриці \mathbf{F} , \mathbf{C} , \mathbf{H} , \mathbf{Q} – сталі в часі, то перехідну матрицю $\mathbf{T}(t, t_0)$ можна, наприклад, знайти за формулою

$$\mathbf{T}(t, t_0) = \mathbf{T}(t - t_0) = \mathbf{T}(\tau) = L^{-1} \mathbf{sI} - \mathbf{A}^{-1}, \quad (2.81)$$

де L^{-1} – операція оберненого перетворення Лапласа; \mathbf{I} – одинична матриця; $\mathbf{sI} - \mathbf{A}^{-1}$ – матриця, обернена до $\mathbf{sI} - \mathbf{A}$; s – оператор перетворення Лапласа.

В тих випадках, коли необхідно знайти значення коефіцієнтів $\mathbf{G}_0(t, t_0)$ в структурній схемі оптимального фільтра тільки в усталеному режимі, достатньо знайти усталене значення матриці $\mathbf{E}(t)$. Якщо, крім того, корисне повідомлення й завади є стаціонарними процесами, то в рівнянні (2.70) слід покласти $d\mathbf{E}/dt = 0$ і розв'язати отримане матричне алгебраїчне рівняння. Остання задача є значно простішою, ніж розв'язання повного рівняння (2.70).

2.6. Приклади синтезу оптимальних фільтрів

Розглянемо приклади синтезу лінійних фільтрів методом простору станів.

Приклад 2.3. Нехай повідомлення $\lambda(t)$ – стаціонарний випадковий процес, який описується диференціальним рівнянням першого порядку

$$\frac{d\lambda}{dt} = -\frac{1}{T}\lambda(t) + \frac{1}{T}n(t).$$

Такий процес відповідає білому шуму $n(t)$, пропущеному через інерційну ланку з операторним коефіцієнтом передачі $K(p) = 1/(1 + pT)$. Функція кореляції шуму $n(t)$ записується у вигляді

$$R_x(t, u) = S_s(0)\delta(t - u), \quad (2.82)$$

де $S_s(0)$ – спектральна щільність шуму $n(t)$ на нульовій частоті. Зазначимо, що при цьому спектральна щільність повідомлення $\lambda(t)$

$$S_\lambda(\omega) = \frac{2\mu\sigma_\lambda^2}{\omega^2 + \mu^2}, \quad (2.83)$$

де $\mu = 1/T$; $\sigma_\lambda^2 = S_n(0)/2T$ – дисперсія процесу $\lambda(t)$. Оскільки $\lambda(t)$ описується рівнянням першого порядку, то векторний процес $\mathbf{x}(t)$ в даній задачі одновимірний і складається з єдиної компоненти, яка збігається з $\lambda(t)$. Відповідно, процес $\mathbf{x}(t)$ описується рівнянням

$$\frac{dx}{dt} = -\frac{1}{T}x(t) + \frac{1}{T}n(t). \quad (2.84)$$

Припустимо, що суміш повідомлення й завади на вході фільтра має вигляд

$$\varphi(t) = x(t) + n(t), \quad (2.85)$$

де $n(t)$ – стаціонарний білий шум із спектральною щільністю $S_3(0)$ та функцією кореляції $R(\tau) = S_3(0)\delta(\tau)$.

Порівняння (2.82), (2.84), (2.85) із (2.64) і (2.65) показує, що в даній задачі матриці \mathbf{F} , \mathbf{H} , \mathbf{C} , \mathbf{Q} , \mathbf{R} є сталими скалярами і дорівнюють:

$$\mathbf{F} = -1/T, \quad \mathbf{H} = 1/T, \quad \mathbf{C} = 1, \quad \mathbf{Q} = S_3(0), \quad \mathbf{R} = S_n(0), \quad (2.86)$$

де S_3 і S_n – спектральна щільність завади і повідомлення відповідно.

Рівняння (2.67), що визначає структуру оптимального фільтра, в даному випадку конкретизується:

$$d\mathbf{x}_0/dt = \mathbf{G}_0(t, t) \varphi(t) - \mathbf{x}_0(t) - \mathbf{x}_0(t)/T, \quad (2.87)$$

де згідно з (2.68) і (2.86)

$$\mathbf{G}_0(t, t) = \mathbf{E}(t)/S_3(0). \quad (2.88)$$

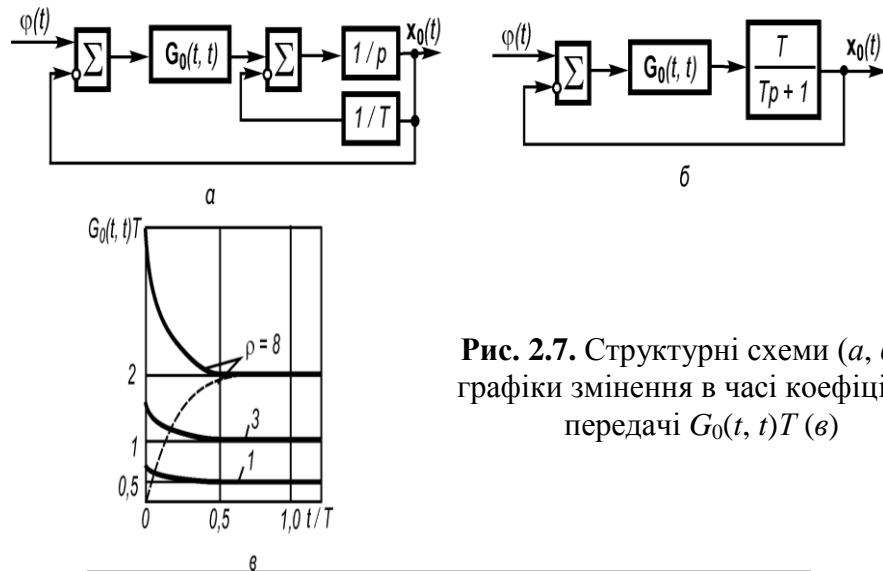


Рис. 2.7. Структурні схеми (а, б) та графіки змінення в часі коефіцієнта передачі $G_0(t, t)T$ (в)

Рівнянню (2.87) відповідає структурна схема, показана на рис. 2.7, а. Суматор, охоплений зворотним зв'язком, можна замінити інерційною ланкою. Тоді структурна схема набере вигляду, як на рис. 2.7, б.

Рівняння (2.70) для дисперсій похибок в даній задачі скалярне:

$$\frac{dE}{dt} = -\frac{2}{T}E(t) - \frac{1}{S_n(0)}E^2(t) + \frac{1}{T^2}S_3(0).$$

Замінімо його відповідно до (2.72), (2.73) системою лінійних диференціальних рівнянь:

$$\left. \begin{aligned} \frac{dv_1}{dt} &= -\frac{1}{T}v_1 + \frac{S_n v_2}{T^2}, \\ \frac{dv_2}{dt} &= \frac{v_1}{S_3(0)} + \frac{v_2}{T}. \end{aligned} \right\} \quad (2.89)$$

Із системи (2.89) випливає, що матриця \mathbf{A} має такий вигляд:

$$\mathbf{A} = \begin{bmatrix} -1/T & S_n(0)/T^2 \\ 1/S_3(0) & 1/T \end{bmatrix}. \quad (2.90)$$

Припустимо, що $t_0=0$, і використаємо для обчислення перехідної матриці $\mathbf{T}(t, 0)$ формулу (2.80). Матриця $s\mathbf{I} - \mathbf{A}$ згідно з (2.90) дорівнює

$$\begin{bmatrix} s+1/T & -S_n(0)/T^2 \\ -1/S_3(0) & s-1/T \end{bmatrix}.$$

Обернена до неї матриця

$$s\mathbf{I} - \mathbf{A}^{-1} = \frac{1}{s^2 - \gamma^2} \begin{bmatrix} s-1/T & S_n(0)/T^2 \\ 1/S_3(0) & s+1/T \end{bmatrix}, \quad (2.91)$$

де $\gamma = \sqrt{1 + S_n(0)/S_3(0)} = \sqrt{1 + \rho/T}$; $\rho = S_n(0)/S_3(0)$.

Виконавши обернене перетворення Лапласа для елементів матриці (2.91), дістанемо

$$\mathbf{T}(t, 0) = \begin{bmatrix} \text{ch } \gamma t - \frac{1}{\gamma T} \text{sh } \gamma t & \frac{S_n(0)}{T^2 \gamma} \text{sh } \gamma t \\ \frac{\text{sh } \gamma t}{S_3(0) \gamma} & \text{ch } \gamma t + \frac{\text{sh } \gamma t}{\gamma T} \end{bmatrix}; \quad (2.92)$$

отже,

$$\begin{aligned} \mathbf{T}_{11} &= \text{ch } \gamma t - \frac{\text{sh } \gamma t}{\gamma T}, & \mathbf{T}_{12} &= \frac{S_n(0) \text{sh } \gamma t}{T^2 \gamma}, \\ \mathbf{T}_{21} &= \frac{\text{sh } \gamma t}{S_3(0) \gamma}, & \mathbf{T}_{22} &= \text{ch } \gamma t + \frac{\text{sh } \gamma t}{\gamma T}. \end{aligned} \quad (2.93)$$

За формулою (2.78) з урахуванням (2.93) знаходимо, що дисперсія похибки

$$E(t) = \frac{E(0) \left[\text{ch } \gamma t - \frac{\text{sh } \gamma t}{\gamma T} \right] + \frac{S_n(0) \text{sh } \gamma t}{T^2 \gamma}}{E(0) \cdot \frac{\text{sh } \gamma t}{S_3(0) \gamma} + \text{ch } \gamma t + \frac{\text{sh } \gamma t}{\gamma T}}. \quad (2.94)$$

Значення $E(0)$ дорівнює дисперсії σ_λ^2 процесу $\lambda(t)$ і визначається співвідношенням (2.84), тобто $E(0) = \sigma_\lambda^2 = S_n(0)/2T$. При цьому вираз (2.94) для похибки набирає вигляду

$$E(t) = \sigma_\lambda^2 \frac{\sqrt{1+\rho} \cdot \text{ch } \gamma t + \text{sh } \gamma t}{\sqrt{1+\rho} \cdot \text{ch } \gamma t + (1+0,5\rho) \text{sh } \gamma t}. \quad (2.95)$$

Дисперсію похибки фільтрації в усталеному режимі знайдемо, взявши в (2.95) $t \rightarrow \infty$:

$$E_{\text{уст}} = \frac{S_n(0)/T}{1+\gamma T} = \frac{2\sigma_\lambda^2}{1+\sqrt{1+\rho}}.$$

Змінення в часі коефіцієнта передачі $G_0(t, t)$ за схемою оптимального фільтра описується, як впливає з (2.88) і (2.95), виразом

$$G_0(t, t) = \frac{1}{2T} \rho \frac{\sqrt{1+\rho} \cdot \text{ch } \gamma t + \text{sh } \gamma t}{\sqrt{1+\rho} \cdot \text{ch } \gamma t + (1+0,5\rho)\text{sh } \gamma t}. \quad (2.96)$$

На рис. 2.7, в суцільними лініями показано розраховане за формулою (2.96) для кількох значень ρ змінення нормованого коефіцієнта $G_0(t, t)T$ в часі. Як бачимо, коефіцієнт $G_0(t, t)$ максимальний в перший момент часу, а потім поступово зменшується, прямуючи до усталеного значення $G_{\text{уст}}$. Така поведінка коефіцієнта $G_0(t, t)$ оптимального фільтра пояснюється тим, що в перший момент часу повідомлення $\lambda(t)$ не проходить на вихід фільтра, й виникає значна помилка. Щоб швидко усунути її, необхідно розширити смугу пропускання слідкуючої системи, що й досягається збільшенням коефіцієнта $G_0(t, t)$. У процесі оброблення слідкуючою системою повідомлення $\lambda(t)$ вага помилки, спричиненої неточним відтворенням повідомлення $\lambda(t)$, зменшується. Одночасно збільшується флуктуаційна складова помилки. Щоб обмежити її, доцільно звужити смугу пропускання системи, зменшивши коефіцієнт $G_0(t, t)$. Початкове та усталене значення коефіцієнта $G_0(t, t)$, як випливає з (2.96), дорівнюють

$$G_0(0) = \frac{\rho}{2T}, \quad G_{\text{уст}} = \frac{\rho}{T(1+\sqrt{1+\rho})}. \quad (2.97)$$

Їх співвідношення $G_0(0)/G_{\text{уст}} = (1+\sqrt{1+\rho})/2$, яке характеризує ступінь змінення коефіцієнта $G_0(t, t)$, тим більше, чим більше відношення $\rho = S_{\text{н}}(0)/S_{\text{з}}(0)$ спектральних щільностей повідомлення й завади. Зі зростанням ρ зростає також значення коефіцієнта передачі $G_0(t, t)T$ в усталеному режимі.

Якщо повідомлення $\lambda(t)$ є нестационарним процесом, то поведінка коефіцієнта $G_0(t, t)$ оптимального фільтра змінюється. Розглянемо такий випадок. Нехай в умовах даної задачі вхідний сигнал $\lambda(t)$ утворюється вмиканням на вхід формувального фільтра білого шуму $n(t)$ в той же момент часу t_0 , коли суміш повідомлення й завади надходить для фільтрації на вхід оптимального фільтра.

Повідомлення $\lambda(t)$ в цьому випадку є процесом з наростаючою дисперсією $\sigma_{\lambda}^2(t) = \sigma_{\lambda_{\text{уст}}}^2 (1 - \exp(-2(t-t_0)/T))$, яка прямує до усталеного значення $\sigma_{\lambda_{\text{уст}}}^2 = S_{\text{н}}(0)/2T$. В момент $t=t_0$ повідомлення на виході формувального фільтра відсутнє і дисперсія $\sigma_{\lambda}^2(t_0) = 0$. Оскільки початкове значення $x_0(t_0)$ процесу на виході оптимального фільтра береться рівним нулю, то при $\sigma_{\lambda}^2(t_0) = 0$ дисперсія помилки $E(t_0)$ в момент $t=t_0$ також дорівнює нулю.

Зміну коефіцієнта $G_0(t, t)$ в схемі оптимального фільтра для цього випадку можна обчислити за формулами (2.86) і (2.94), поклавши в (2.94) $E(t_0) = 0$. Відповідна крива при $\rho = 8$ показана на рис. 2.7, в штриховою лінією. Як видно з рисунка, оптимальний коефіцієнт $G_0(t, t)$ в даному випадку збільшується зі зростанням дисперсії повідомлення, прямуючи до попереднього усталеного значення.

Комплексний коефіцієнт передачі знайденого оптимального фільтра, наведеного на рис. 2.7, б, в усталеному режимі дорівнює

$$\Phi(j\omega) = \frac{G_{\text{уст}}T/(1+j\omega T)}{1+G_{\text{уст}}T/(1+j\omega T)} = k_{\text{екв}} \frac{1}{1+j\omega T_{\text{екв}}}, \quad (2.98)$$

$$\text{де } T_{\text{екв}} = T/\sqrt{1+\rho}, \quad k_{\text{екв}} = \rho/(1+\rho+\sqrt{1+\rho}).$$

Оцінимо, який же вигравш досягається за рахунок змінності параметрів оптимального фільтра. Для цього порівняємо похибки фільтрації при виділенні стаціонарного повідомлення оптимальним фільтром зі змінними параметрами і оптимальним фільтром (2.98) зі сталими

параметрами. Якщо початкові умови на виході фільтра нульові, то в перший момент для обох фільтрів дисперсії похибки дорівнюють дисперсії повідомлення, тобто однакові. Однакові вони і в усталеному режимі. Виграш при використанні оптимального фільтра зі змінними параметрами проявляється в зменшенні помилки в перехідному режимі.

Виведемо співвідношення, яке описує зміну похибки фільтрації для оптимального фільтра (2.98). Дисперсія флукуаційної складової σ_1^2 , яка зумовлена дією завади $n(t)$, змінюється за законом

$$\sigma_1^2(t) = S_3(0) \int_0^t g_2(\tau) d\tau,$$

де $g_2(\tau)$ – імпульсна перехідна функція, яка відповідає комплексному коефіцієнту передачі (2.98):

$$g(t) = (k_{\text{екв}}/T_{\text{екв}}) e^{-t/T_{\text{екв}}}.$$

Після інтегрування отримаємо

$$\sigma_1^2(t) = S_3(0) \frac{k_{\text{екв}}^2}{2T} (1 - e^{-2t/T_{\text{екв}}}).$$

Комплексний коефіцієнт передачі, який пов'язує повідомлення $\lambda(t)$ і спричинену його зміненням складову похибки, дорівнює

$$K(j\omega) = 1 - \Phi(j\omega).$$

Йому відповідає імпульсна перехідна функція

$$g_2(t) = \delta(t) - \frac{k_{\text{екв}}}{T_{\text{екв}}} e^{-t/T_{\text{екв}}}.$$

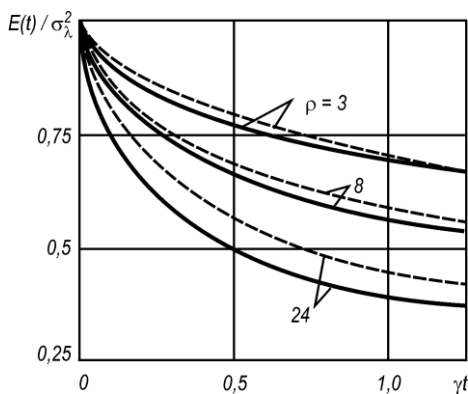


Рис. 2.8. Змінення нормованої похибки

Оскільки повідомлення $\lambda(t)$ не є білим шумом і має функцію кореляції $R_\lambda(\tau) = \sigma_\lambda^2 e^{-\mu|\tau|}$, то для обчислення дисперсії складової похибки, зумовленої неточним відтворенням повідомлення, застосуємо формулу [17]

$$\sigma_2^2(t) = 2 \int_0^t g_2(\tau) g_1(\tau) d\tau,$$

де $g_1(\tau)$ – функція, зображення якої за Лапласом дорівнює добуткові зображення $\sigma_\lambda^2/(s+\mu)$ кореляційної функції $R_\lambda(\tau)$ дії та передавальної функції $K(s) = 1 - \Phi(s)$, тобто

$$g_1 \tau = \sigma_\lambda^2 K(s)/(s+\mu).$$

Тоді, виконавши інтегрування, дістанемо

$$\sigma_2^2(t) = \sigma_\lambda^2 \left[1 - k_{\text{екв}} (1 - e^{-2t/T_{\text{екв}}}) \right].$$

Повідомлення й завада за умовою задачі незалежні. Тому змінення повної дисперсії похибки описується виразом

$$\sigma^2(t) = E(t) = \sigma_1^2(t) + \sigma_2^2(t) = \sigma_\lambda^2 \left[1 - (1 - e^{-2\gamma t}) \frac{\rho}{(1 + \sqrt{1 + \rho})^2} \right].$$

Змінення нормованої похибки $E(t)/\sigma_\lambda^2$, розраховане за наведеною формулою для кількох значень ρ , показане на рис. 2.8 штриховою лінією; суцільними лініями зображені обчислені за формулою (2.95) аналогічні криві для оптимального фільтра зі змінними параметрами. Порівняння кривих дозволяє зробити для розглянутого випадку фільтрації стаціонарного повідомлення наступні висновки:

- 1) виграш у зменшенні похибки фільтрації, який досягається за рахунок змінності параметрів, збільшується зі зростанням величини ρ (відношення повідомлення–завада);
- 2) інтервал часу, в межах якого проявляється перевага оптимального фільтра зі змінними параметрами, пропорційний до постійної часу T фільтра, що формує повідомлення, а отже, обернено пропорційний до ширини спектра повідомлення;
- 3) в цілому виграш за рахунок змінності параметрів порівняно невеликий, і для розглянутих співвідношень повідомлення–завада не перевищує 15 %.

Проте такі висновки не зменшують інтересу до оптимальних фільтрів зі змінними параметрами та методів їх синтезу. Якщо повідомлення чи завада описуються нестационарними випадковими процесами, то оптимальний фільтр навіть в усталеному режимі має змінні параметри, й заміна його фільтром зі сталими параметрами може призвести до істотного погіршення якості фільтрації.

Приклад 2.4. Припустимо, що повідомлення $\lambda(t)$ може бути зображене як процес, що формується на виході двох послідовно ввімкнених інтеграторів, і описане стохастичним диференціальним рівнянням другого порядку

$$d^2\lambda/dt^2 = k_{12}n(t). \quad (2.99)$$

Спектральна щільність процесу $\lambda(t)$ дорівнює $S_\lambda(\omega) = k_{12}^2 S_n(0)/\omega^4$. На вхід фільтра, що оптимізується, надходить суміш

$$\varphi(t) = \lambda(t) + n(t), \quad (2.100)$$

де $n(t)$ – білий шум із спектральною щільністю $S_n(0)$.

Необхідно знайти структуру і параметри оптимального фільтра, який мінімізує середньоквадратичну похибку відтворення процесу $\lambda(t)$ в будь-який момент часу t .

Для розв'язання поставленої задачі подамо рівняння (2.97) у вигляді системи рівнянь першого порядку:

$$dx_1/dt = x_2(t), \quad dx_2/dt = k_{12}n(t),$$

де $x_1(t) = \lambda(t)$. Отриману систему рівнянь запишемо в матричній формі:

$$\begin{bmatrix} dx_1/dt \\ dx_2/dt \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ k_{12} \end{bmatrix} n(t). \quad (2.101)$$

Із порівняння (2.101) і (2.63) випливає, що в даній задачі вектор $\mathbf{x}(t)$ містить дві компоненти: $x_1(t) = \lambda(t)$ і $x_2(t)$, а матриці \mathbf{F} , \mathbf{H} , \mathbf{Q} – сталі:

$$\mathbf{F} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 0 \\ k_{12} \end{bmatrix}, \quad \mathbf{Q} = S_n(0). \quad (2.102)$$

Рівняння (2.98) також запишемо в матричній формі, виразивши $\varphi(t)$ через вектор $\mathbf{x}(t)$:

$$\varphi(t) = 1 \quad 0 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + n(t). \quad (2.103)$$

Із порівняння (2.103) і (2.65) випливає, що в даному випадку матриці \mathbf{C} і \mathbf{R} конкретизуються у вигляді

$$\mathbf{C} = 1 \quad 0, \quad \mathbf{R} = S_3(0). \quad (2.104)$$

Оскільки вектор $\mathbf{x}(t)$ в даній задачі містить дві компоненти, то матриця дисперсій помилок $\mathbf{E}(t)$ має розмір 2×2 :

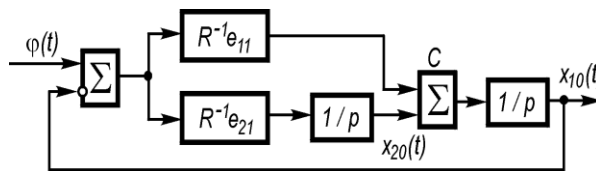
$$\mathbf{E}(t) = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix}. \quad (2.105)$$

Синтез фільтра зручно розпочати з визначення матриці коефіцієнтів $\mathbf{G}_0(t, t)$. Згідно з (2.68) $\mathbf{G}_0(t, t) = \mathbf{E}(t)\mathbf{C}^T\mathbf{R}^{-1}$, тому, враховуючи (2.103) і (2.102), отримуємо

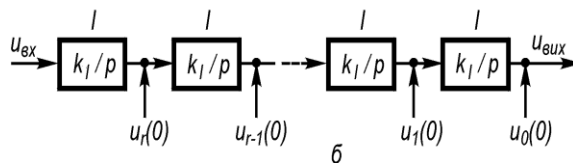
$$\mathbf{G}_0(t, t) = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mathbf{R}^{-1} = \begin{bmatrix} e_{11} \\ e_{21} \end{bmatrix} \mathbf{R}^{-1}. \quad (2.106)$$

Щоб знайти структуру оптимального фільтра для даної задачі, підставимо в рівняння (2.62) матриці \mathbf{F} , \mathbf{C} , $\mathbf{G}_0(t, t)$, які визначаються виразами (2.102), (2.104), (2.106), і запишемо його у вигляді системи рівнянь першого порядку:

$$\begin{aligned} dx_{10}/dt &= x_{20}(t) + R^{-1}e_{11} \varphi(t) - x_{10}(t), \\ dx_{20}/dt &= R^{-1}e_{21} \varphi(t) - x_{10}(t). \end{aligned} \quad (2.107)$$



а



б

Рис. 2.9. Змінення структурної схеми

Структурна схема оптимального фільтра, описуваного рівняннями (2.107), показана на рис. 2.9, а. Залишається знайти коефіцієнти, що входять до нього: $e_{11}(t)$ та $e_{21}(t)$. Для цього необхідно розв'язати рівняння (2.70). Обмежимося для простоти розглядом усталеного режиму, в якому $d\mathbf{E}(t)/dt = 0$. Обчислимо з урахуванням (2.72), (2.74) окремі доданки рівняння (2.70):

$$\mathbf{FE} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} = \begin{bmatrix} e_{21} & e_{22} \\ 0 & 0 \end{bmatrix}, \quad \mathbf{EF}^T = \begin{bmatrix} e_{21} & 0 \\ e_{22} & 0 \end{bmatrix},$$

$$\mathbf{EC}^T \mathbf{R}^{-1} \mathbf{CE} = \begin{bmatrix} e_{11}^2 & e_{11}e_{12} \\ e_{21}e_{11} & e_{21}e_{12} \end{bmatrix} \mathbf{R}^{-1}, \quad \mathbf{HQH}^T = \mathbf{Q} \begin{bmatrix} 0 & 0 \\ 0 & k_{12}^2 \end{bmatrix}.$$

Підсумувавши всі доданки (2.70) з урахуванням того, що матриця \mathbf{E} симетрична і $e_{12} = e_{21}$, отримаємо наступну систему алгебраїчних рівнянь:

$$2e_{21} - R^{-1}e_{11}^2 = 0, \quad e_{22} - e_{11}e_{21}R^{-1} = 0, \quad Qk_{12}^2 - R^{-1}e_{21}^2 = 0,$$

звідки

$$e_{21} = \sqrt{Qk_{12}^2 R} = \sqrt{k_{12}^2 S_n(0) S_3(0)},$$

$$e_{11} = \sqrt{2e_{21} R} = \sqrt[4]{4k_{12}^2 S_n(0) S_3^3(0)}. \quad (2.108)$$

В усталеному режимі комплексний коефіцієнт передачі розімкненої слідкуючої системи відповідно до рис. 2.9, б та рівностей (2.108)

$$K(j\omega) = \frac{K_a(1 + j\omega T_1)}{(j\omega)^2}, \quad (2.109)$$

$$\text{де } K_a = R^{-1}e_{21} = k_{12} \sqrt{S_n(0)/S_3(0)}; \quad T_1 = e_{11}/e_{21} = \sqrt[4]{4S_3(0)/k_{12}^2 S_n(0)}.$$

Отже, за умовами даної задачі оптимальний фільтр в усталеному режимі є слідкуючою системою з двома інтеграторами і демпфірувальним ланцюжком, параметри яких визначаються співвідношеннями (2.109).

2.7. Оптимальний фільтр для повідомлення, описуваного степеневим поліномом

Метод простору станів дозволяє достатньо просто знаходити структуру й параметри оптимального фільтра, який мінімізує похибку фільтрації в будь-який момент часу, та у випадку, коли повідомлення $\lambda(t)$ описується поліномом із випадковими коефіцієнтами:

$$\lambda(t) = a_0 + a_1 t + \dots + a_r t^r, \quad (2.110)$$

де a_0, a_1, \dots, a_r – випадкові величини з відомою кореляційною матрицею. Елементами цієї матриці є значення $M a_i a_j = \sigma_{ij}^2$. Застосування методу простору станів в даному випадку засновано на можливості зобразити повідомлення $\lambda(t)$ як процес на виході формувального фільтра, який складається із $r+1$ послідовно ввімкнених інтеграторів I (рис. 2.9, б). Дійсно, якщо зовнішня дія на вході ланцюжка інтеграторів відсутня, тобто $u_{\text{вх}}(t) = 0$, а початкові значення $u_i(0)$ на входах інтеграторів встановлені рівними

$$u_i(0) = \frac{a_i i!}{k_1^i},$$

то на виході ланцюжка утворюється процес (2.110). Формування повідомлення $\lambda(t)$ схемою, яка показана на рис. 2.7 в, описується диференціальним рівнянням $d^{r+1}\lambda/dt^{r+1} = 0$. Поклавши $\lambda(t) = x_1(t)$, це рівняння можна замінити системою $r+1$ рівнянь першого порядку:

Матрицю $\mathbf{E}(t)$ в даному випадку можна знайти, розв'язавши рівняння (2.115) шляхом переходу до системи лінійних рівнянь (2.72) та (2.73). Нагадаємо, що при цьому в процесі розв'язання виникає необхідність визначити перехідну матрицю $\mathbf{T}(t, t_0)$ розміром $2n \times 2n$, де $n=r+1$. Якщо для її пошуку використовується формула (2.81), то при обчисленні оберненої матриці $[s\mathbf{I} - \mathbf{A}]^{-1}$ зручно зобразити $s\mathbf{I} - \mathbf{A}$ у вигляді блочної матриці:

$$s\mathbf{I} - \mathbf{A} = \begin{bmatrix} \mathbf{m} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{bmatrix}. \quad (2.118)$$

Тоді відповідно до [11]

$$s\mathbf{I} - \mathbf{A} = \begin{bmatrix} \mathbf{M} & \mathbf{K} \\ \mathbf{B} & \mathbf{D} \end{bmatrix},$$

де

$$\begin{aligned} \mathbf{M} &= [\mathbf{m} - \mathbf{b}\mathbf{d}^{-1}\mathbf{c}]^{-1}, & \mathbf{B} &= -\mathbf{d}^{-1}\mathbf{c}\mathbf{M}, \\ \mathbf{D} &= [\mathbf{d} - \mathbf{c}\mathbf{m}^{-1}\mathbf{b}]^{-1}, & \mathbf{K} &= -\mathbf{m}^{-1}\mathbf{b}\mathbf{D}. \end{aligned} \quad (2.119)$$

Оскільки в задачі фільтрації повідомлення, описуваного поліномом, $\mathbf{Q} = 0$, то в матриці $s\mathbf{I} - \mathbf{A}$, як видно з (2.80), блок $\mathbf{b} = 0$. Рівності (2.119) при цьому набувають дуже простого вигляду:

$$\mathbf{M} = -\mathbf{m}^{-1}, \quad \mathbf{B} = -\mathbf{d}^{-1}\mathbf{c}\mathbf{M}, \quad \mathbf{D} = \mathbf{d}^{-1}, \quad \mathbf{K} = 0,$$

і процедура визначення матриці $\mathbf{E}(t)$ спрощується.

Кореляційну матрицю $\mathbf{E}(t)$ похибок фільтрації при $\mathbf{Q} = 0$ можна знайти також за формулою

$$\mathbf{E}(t) = \mathbf{\Omega}(t, 0)\mathbf{P}^{-1}(t)\mathbf{\Omega}^T(t, 0), \quad (2.120)$$

де $\mathbf{\Omega}(t, 0) = e^{\mathbf{F}t}$ – перехідна матриця формувального фільтра;

$$\mathbf{P}(t) = \mathbf{E}^{-1}(0) + \int_0^t \mathbf{\Omega}^T(\tau)\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}\mathbf{\Omega}(\tau)d\tau.$$

Рівність (2.120) є наведеним розв'язком рівняння (2.115), вираженим через легко обчислювану матрицю $\mathbf{\Omega}(t, 0)$. Розмір всіх матриць, які входять в (2.120), не перевищує $n \times n$.

Обсяг обчислень при визначенні матриці $\mathbf{E}(t)$ обома способами приблизно однаковий. Розглянемо побудову оптимального фільтра для повідомлення $\lambda(t)$, яке описується поліномом першого порядку.

Приклад 2.5. Нехай повідомлення

$$\lambda(t) = a_0 + a_1 t, \quad (2.121)$$

де a_0, a_1 – випадкові незалежні величини з нульовими математичними сподіваннями та дисперсіями σ_0^2 і σ_1^2 відповідно. На вхід фільтра надходить суміш

$$\varphi(t) = \lambda(t) + n(t),$$

де $n(t)$ – білий шум зі спектральною щільністю $S_3(0)$ і функцією кореляції $R(\tau) = S_3(0)\delta(\tau)$.

Фільтр має забезпечити мінімум середньоквадратичної похибки в будь-який момент часу t . Оскільки порядок полінома (2.121) $r=1$, то вектор $\mathbf{x}(t)$ і матриця \mathbf{F} в даному прикладі у відповідності до (2.13) набувають вигляду

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \quad (2.122)$$

Суміш $\varphi(t)$, виражена через векторний процес $\mathbf{x}(t)$, записується при цьому так:

$$\varphi(t) = 1 \quad 0 \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + n(t), \quad x_1(t) = \lambda(t).$$

Звідси після зіставлення з (2.65) випливає, що

$$\mathbf{C} = 1 \quad 0, \quad \mathbf{R} = S_3(0). \quad (2.123)$$

Вектор $\mathbf{x}(t)$ і матриці \mathbf{F} , \mathbf{C} , \mathbf{R} , як випливає із (2.122), (2.123), (2.102), (2.103), такі ж самі, як і в прикладі 2.4, тому знайдені там диференціальні рівняння (2.107) й структурна схема (рис. 2.7) оптимального фільтра залишаються правильними і для даного випадку. Однак коефіцієнти $e_{11}(t)$ і $e_{21}(t)$ ми отримуємо іншими. Для їх визначення необхідно розв'язати рівняння (2.115).

Застосуємо для обчислення матриці $\mathbf{E}(t)$ формулу (2.120). Знайдемо спочатку перехідну матрицю $\mathbf{\Omega}(t, 0) = e^{\mathbf{F}t}$. Згідно з (2.79), (2.81), (2.121)

$$\mathbf{\Omega}(t, 0) = L^{-1} \quad s\mathbf{I} - \mathbf{F}^{-1},$$

де

$$s\mathbf{I} - \mathbf{F} = \begin{bmatrix} s & -1 \\ 0 & s \end{bmatrix};$$

отже,

$$s\mathbf{I} - \mathbf{F}^{-1} = \begin{bmatrix} 1/s & 1/s^2 \\ 0 & 1/s \end{bmatrix}. \quad (2.124)$$

Визначивши обернене перетворення Лапласа для кожного елемента матриці (2.124), отримаємо:

$$\mathbf{\Omega}(t, 0) = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}. \quad (2.125)$$

Множення матриць $\mathbf{\Omega}^T$, \mathbf{C}^T , \mathbf{R}^{-1} , \mathbf{C} , $\mathbf{\Omega}$ і наступне інтегрування дає наступний результат:

$$\int_0^t \mathbf{\Omega}^T \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} \mathbf{\Omega} d\tau = \mathbf{R}^{-1} \begin{bmatrix} t & t^2/2 \\ t^2/2 & t^3/3 \end{bmatrix}. \quad (2.126)$$

Оскільки коефіцієнти a_0 і a_1 за умовою задачі – незалежні величини з дисперсіями σ_0^2 і σ_1^2 , то матриця $\mathbf{E}(0)$ відповідно до (2.115) і (2.116) має такий вигляд:

$$\mathbf{E}(0) = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}. \quad (2.127)$$

Із (2.120), (2.126) і (2.127) випливає, що

$$\mathbf{P}^{-1}(t) = \frac{1}{\Delta} \begin{bmatrix} 1/\sigma_1^2 + t^3/3R & -t^2/2R \\ -t^2/2R & 1/\sigma_0^2 + t/R \end{bmatrix}, \quad (2.128)$$

де $\Delta = (1/\sigma_0^2 + t/R)(1/\sigma_1^2 + t^3/3R) - t^4/4R^2$.

Підставивши (2.125) і (2.128) в формулу (2.120), знаходимо матрицю $\mathbf{E}(t)$. Її елементи $e_{11}(t)$ і $e_{21}(t)$, що визначають змінні коефіцієнти передачі в структурній схемі оптимального фільтра, дорівнюють:

$$e_{11}(t) = \frac{\sigma_0^2 + \sigma_1^2 t^2 + \frac{\sigma_0^2 \sigma_1^2 t^3}{3S_3(0)}}{1 + \frac{\sigma_0^2 t}{S_3(0)} + \frac{\sigma_1^2 t^3}{3S_3(0)} + \frac{\sigma_0^2 \sigma_1^2}{12S_3^2(0)} t^4},$$

$$e_{21}(t) = \frac{\sigma_1^2 t + \frac{\sigma_0^2 \sigma_1^2 t^2}{3S_3(0)}}{1 + \frac{\sigma_0^2 t}{S_3(0)} + \frac{\sigma_1^2 t^3}{3S_3(0)} + \frac{\sigma_0^2 \sigma_1^2}{12S_3^2(0)} t^4}.$$

Як видно з отриманих виразів, коефіцієнти кореляції та дисперсія помилки виділення повідомлення $\lambda(t)$ зі зростанням t прямують до нуля. Пояснюється це тим, що коефіцієнти a_0 і a_1 у виразі (2.121) стали в часі, й при необмеженому зростанні часу спостереження можуть бути виміряні з як завгодно високою точністю.

2.8. Поняття про оптимальну нелінійну фільтрацію

У п. 2.1 при розгляді оптимальної нелінійної фільтрації наведені схеми слідкувального приймання (рис. 2.1, б, в), які забезпечують мінімум середньоквадратичної похибки.

Синтез системи, яка здійснює нелінійне перетворення сигналу й забезпечує найбільшу точність слідкування за корисним параметром сигналу (повідомленням), котрий змінюється за час слідкування, може бути проведений методами оптимальної нелінійної фільтрації.

В теорії оптимальної нелінійної фільтрації поєдналися два основні підходи. Перший з них, запропонований Р.Л. Стратоновичем, заснований на описі параметра фільтрівного сигналу за допомогою марковського процесу. Цей підхід дозволив розв'язати ряд важливих і цікавих задач синтезу оптимальних нелінійних систем. Великий внесок у розвиток цього напрямку оптимальної нелінійної фільтрації та його застосування до синтезу оптимальних приймальних пристроїв при різних методах модуляції сигналу зробив В. І. Тихонов.

Підхід, розроблений І. А. Большаковим і В. Г. Рєпіним, заснований на припущеннях про нормальність процесу $\lambda(t)$ та високу точність його виділення, які дозволяють застосувати гауссову апроксимацію багатовимірного апостеріорного закону розподілу параметра $\lambda(t)$. У літературі такий підхід одержав назву квазілінійного методу. При проведенні синтезу системи оптимальної нелінійної фільтрації даним методом вхідна напруга записується у вигляді

$$u_{\text{вх}}(t) = u_c(t, \lambda, \lambda_n) + u_{\text{ш}}(t),$$

де $u_c(t, \lambda, \lambda_n)$ – напруга сигналу; $u_{\text{ш}}(t)$ – білий шум зі спектральною щільністю $S(\omega) = S(0)$; $\lambda(t)$ – корисний параметр сигналу, за яким ведеться слідкування; $\lambda_n(t)$ – вектор невідомих параметрів сигналу, які не несуть інформації (їх іноді називають заважаючими або супровідними).

При спрощуючому припущенні, що час кореляції заважаючих параметрів сигналу набагато менший за час кореляції корисного параметра (повідомлення), оптимальна система

нелінійної фільтрації є замкненою слідкуючою системою, функціональна схема якої зображена на рис. 2.10, *a*.

Параметр $\lambda(t)$, за яким ведеться спостереження (частота, фаза, тимчасове положення), нелінійно пов'язаний з радіосигналом $u_c(t, \lambda)$, що діє на вході радіотехнічної слідкуючої системи. Тому слідкуюча система має містити елементи, які здійснюють нелінійне перетворення радіосигналу. Задачу оптимізації структури всієї системи за критерієм мінімуму помилки слідкування слід розв'язувати при цьому на основі теорії оптимальної нелінійної фільтрації.

Нелінійну операцію виділення із сигналу інформації про величину неузгодженості виконує дискримінатор (рис. 2.10, *a*). Якщо дискримінатор вже вибраний або заданий, то задача синтезу звужується до оптимізації структури фільтра. Під фільтром тут розуміється вся інерційна частина системи, що слідкує за дискримінатором.

Припустимо, що помилка слідкування не виходить за межі лінійної ділянки характеристики дискримінатора, і подамо вихідну напругу дискримінатора у вигляді

$$u_{\text{вих}}(t) = S_D x(t) + \xi(t),$$

де S_D – крутість дискримінаційної характеристики; $x(t)$ – похибка слідкування (неузгодженість); $\xi(t)$ – флуктуаційна напруга на виході дискримінатора, яку вважатимемо незалежною від помилки слідкування.

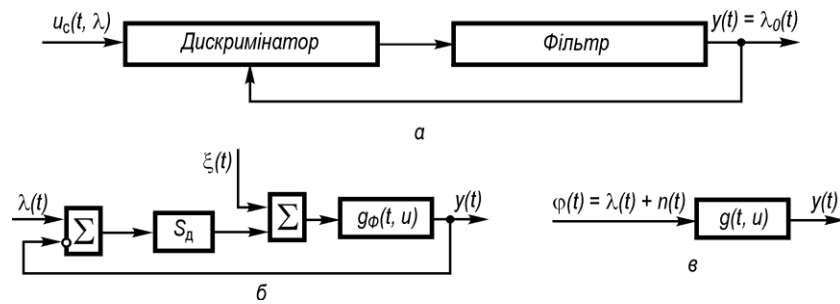


Рис. 2.10. Схеми слідкуючих систем:
a – функціональна; *б*, *в* – структурні

У загальному випадку оптимальний фільтр у контурі слідкуючої системи (рис. 2.10, *a*) може бути нелінійним. Введемо ще одне обмеження, і шукатимемо оптимальний фільтр у класі лінійних пристроїв. При цьому слідкуюча система, яка розглядається, виглядатиме так, як на рис. 2.9, *б*, де $g_\Phi(t, u)$ – імпульсна перехідна функція фільтра, що оптимізується (фільтр може мати змінні параметри).

З точки зору формування процесу $y(t)$ слідкуюча система, показана на рис. 2.9, *б*, може бути замінена еквівалентним фільтром (рис. 2.10, *в*) з імпульсною перехідною функцією $g(t, u)$, на вході якого діє суміш сигналу і завади:

$$\varphi(t) = \lambda(t) + \xi(t) / S_D = \lambda(t) + n(t). \quad (2.129)$$

Завада $n(t)$ відповідає флуктуаційному коливанию (шуму), перерахованому до входу дискримінатора. Імпульсна перехідна функція $g(t, u)$ еквівалентного фільтра пов'язана з аналогічною функцією $g_\Phi(t, u)$ фільтра в контурі слідкуючої системи інтегральним рівнянням [15]

$$S_D g_\Phi(t, u) = g(t, u) + S_D \int_u^t g(\theta, u) g_\Phi(t, \theta) d\theta. \quad (2.130)$$

Дискримінатор оптимальної системи виконує нелінійне перетворення вхідного сигналу, внаслідок чого формується напруга $z(t)$:

$$z(t) = \left. \frac{\partial Q_1(u_{\text{вх}}, \lambda, t)}{\partial \lambda} \right|_{\lambda=\lambda_0}, \quad (2.131)$$

де λ_0 – оцінка параметра $\lambda(t)$; $Q_1(u_{\text{вх}}, \lambda, t)$ – функція, що визначається співвідношенням

$$\int_{t-\Delta t}^t Q_1(u_{\text{вх}}, \lambda, t) dt = \ln P_\lambda(u_{\text{вх}}); \quad (2.132)$$

$P_\lambda(u_{\text{вх}})$ – функція правдоподібності, тобто залежність від λ багатовимірної щільності ймовірності реалізації вхідного процесу $u_{\text{вх}}(t)$, розглянутої на інтервалі часу Δt . Інтервал Δt вибирається значно більшим за час кореляції заважаючих параметрів, але значно меншим, ніж час кореляції корисного параметра сигналу. Функція правдоподібності $P_\lambda(u_{\text{вх}})$ обчислюється шляхом усереднення функції правдоподібності, записаної для сигналу, всі параметри якого, крім λ , відомі точно. Усереднення проводиться за всіма можливими значеннями заважаючих параметрів, з урахуванням їх апріорної щільності ймовірності.

Оскільки відповідно до (2.131) похідна функція Q_1 обчислюється в точці $\lambda = \lambda_0$, до дискримінатора (рис. 2.10, а) підводиться також оцінка $\lambda_0(t)$. Слідкуючий за дискримінатором фільтр в оптимальній системі є лінійним і, в загальному випадку, нестационарним. Його імпульсна перехідна функція $g_\Phi(t, u)$ збігається з імпульсною перехідною функцією оптимального фільтра в контурі лінійної слідкуючої системи, зображеної на рис. 2.10, б. При цьому в схемі на рис. 2.10, б, як показано в [15], доцільно взяти

$$S_D = K, \quad S_\xi(\omega) = S_\xi(0) = K, \quad (2.133)$$

де $S_\xi(0)$ – спектральна щільність процесу $\xi(t)$;

$$K = \overline{-\partial^2 Q_1 / \partial \lambda^2}. \quad (2.134)$$

Методика визначення імпульсної перехідної функції $g_\Phi(t, u)$ оптимального фільтра в контурі лінійної слідкуючої системи докладно викладена в [35]. За цією методикою для визначення оптимальної функції $g_\Phi(t, u)$ необхідно спочатку, розв'язавши інтегральне рівняння, знайти функцію $g(t, u)$ еквівалентного лінійного фільтра. Потім за знайденою функцією $g(t, u)$ визначається функція $g_\Phi(t, u)$ в результаті розв'язання інтегрального рівняння (2.130). Якщо повідомлення $\lambda(t)$ та завада $u_{\text{ш}}(t)$ стаціонарні і потрібно мінімізувати помилку фільтрації тільки в усталеному режимі, то оптимальний лінійний фільтр має сталі параметри, а його комплексний коефіцієнт передачі $K(j\omega)$ записується у вигляді

$$K(j\omega) = \frac{1}{k} \cdot \frac{\Phi(j\omega)}{1 - \Phi(j\omega)}, \quad (2.135)$$

де $\Phi(j\omega)$ – коефіцієнт передачі оптимального лінійного фільтра, який виділяє повідомлення $\lambda(t)$ із суміші з білим шумом, спектральна щільність якого дорівнює $1/k$.

Достоїнством методу, запропонованого І.А. Большаковим і В.Г. Репіним, є можливість роздільного синтезу дискримінатора й лінійного фільтра. Із зміненням спектра корисного

параметра $\lambda(t)$ дискримінатор оптимальної системи залишається незмінним, і знову доводиться шукати тільки оптимальний лінійний фільтр.

Марковський підхід до теорії оптимальної нелінійної фільтрації носить більш загальний характер, оскільки принципово його можна застосувати у випадку негауссової апостеріорної щільності ймовірності виділюваного параметра. Більшість практичних результатів при марковському підході отримані поки в рамках гауссового наближення, при якому апостеріорна щільність ймовірності процесу $\lambda(t)$ також вважається гауссовою. Однак і в цьому випадку можна вказати коло задач, розв'язання яких за марковським підходом має певні переваги. Трудомісткість синтезу на основі марковського підходу і квазілінійним методом приблизно однакова, якщо виконуються наступні умови:

- 1) повідомлення та завада є стаціонарними випадковими процесами, заданими своїми кореляційними функціями чи енергетичними спектрами;
- 2) мінімізація помилки фільтрації виконується в усталеному режимі;
- 3) всі параметри сигналу, крім корисного (тобто такого, що несе інформацію про повідомлення), відомі.

Деяку перевагу при виконанні цих умов, очевидно, можна надати квазілінійному методу. Якщо умови 1) і 2) не виконуються, то оптимальні згладжувальні ланцюги слідкуючої системи мають бути нестационарними. Синтез їх на основі марковського підходу простіший. Це аналогічно перевазі синтезу оптимальних лінійних нестационарних фільтрів методом простору станів у порівнянні з безпосереднім розв'язанням інтегральних рівнянь Вінера–Хопфа.

Якщо заважаючі параметри сигналу змінюються в часі, а період їх зміни чи час кореляції порівнянні з часом кореляції корисного параметра сигналу, то доцільно проводити синтез на основі марковської теорії.

Метод синтезу, заснований на марковській теорії, зручний також у тих випадках, коли потрібно з мінімальною похибкою відфільтрувати кілька параметрів (або кілька складових параметра) сигналу.

РОЗДІЛ 3

ВИКОРИСТАННЯ ЗАВАДОСТІЙКОГО КОДУВАННЯ В КАНАЛАХ СИСТЕМИ УПРАВЛІННЯ ІНФОКОМУНІКАЦІЙНОЮ МЕРЕЖЕЮ

3.1. Математична модель дискретного каналу

Питання передачі дискретної інформації в автоматизованих та інформаційних системах і мережах у разі, коли неможливо або невигідно з економічної точки зору використовувати для підвищення достовірності повідомлень зворотний канал зв'язку, викликають зацікавленість у широкого кола спеціалістів в галузі автоматики, проводового та радіозв'язку. Цій темі присвячена велика кількість статей та монографій. Однак, більшість публікацій присвячена висвітленню інформаційних каналів, в яких у якості завадостійких використовуються двійкові коректуючі коди. Це викликано, головним чином, простотою операцій у двійковій системі числення, які виконуються різними пристроями обчислювальної техніки при обробці, накопиченні та зберіганні інформації в автоматизованих та інформаційних системах та мережах.

Використання комбінованих видів модуляції в інформаційних каналах дозволяє набагато підвищити швидкість передачі даних у порівнянні з двійковими системами. І використання завадостійких кодів з алфавітом $q > 2$, які дозволяють виявляти та виправляти помилки, в таких системах є цілком логічним кроком. Такі коди одержали назву недвійкових (багатопозиційних, багатоосновних або q -ічних) [16, 22, 30, 33]. Не дивлячись на те, що деякі надмірні недвійкові коди, такі як узагальнений код Хеммінга, код Ріда-Соломона та інші, одержали досить широке застосування в інформаційних каналах [8, 15, 16, 29, 30, 33, 37], питання розвитку теорії недвійкового кодування залишаються досить актуальними.

Опис каналу складається із визначення обмежень на сигнали $S(t)$, передачу яких він забезпечує, та характеру перетворення $S(t) \rightarrow S^*(t)$, яке він виконує. Для дискретних каналів достатньо вказати дозволені сигнали (символи) або просто їх число m та дозволені моменти зміни сигналів. При $m = 2$ (такі канали називаються двійковими), дозволені моменти зміни сигналів (значущі моменти) повністю визначаються швидкістю модуляції. Перетворення $S(t) \rightarrow S^*(t)$ в ідеальному випадку зводиться до рівності $S(t) = S^*(t)$. Відмінність виникає із-за не удосконалення лінії зв'язку. Перетворення $S(t) \rightarrow S^*(t)$ має дві складових – детерміновану та випадкову. Детерміновані зміни сигналу в дискретному каналі полягає у затримці (зміщенні в часі). Вирішуючий пристрій, являється пороговим, видає сигнали фіксованої форми, тому змін сигналу не спостерігається. Випадкові зміни сигналу обумовлені завадами.

При перетворенні символів в дискретному каналі прийнята послідовність елементів повідомлення \vec{B}_i^* відрізняється від переданої \vec{B}_i . В загальному випадку ймовірнісні закономірності перетворення $\vec{B}_i \rightarrow \vec{B}_i^*$ дуже складні із-за складного характеру завади. Вплив завади можна представити як по елементне сумування послідовності \vec{B}_i з послідовністю помилок E_i . Тоді $\vec{B}_i^* = \vec{B}_i \oplus \vec{E}_i$.

В двійкових каналах можна розрізнити дві категорії вихідних елементів: неуражені та уражені. В простішому випадку початок уражених елементів, можна отримати, порівнюючи отримані послідовності \vec{B}_i^* із переданою \vec{B}_i . Це порівняння, як правило виконують не на безкінечному, а на кінцевому відрізку послідовностей (кодової комбінації, блоці, групі блоків).

Помилки можна розрізнити по типу (перехід $1 \rightarrow 0$ або $0 \rightarrow 1$) та кратності t . При рівності $p(1 \rightarrow 0) = p(0 \rightarrow 1)$ помилки вважаються симетричними. Кратність (число помилок) знаходиться числом помилок t , які приходяться на задане число одиничних елементів: помилки можуть бути однократними, двократними і т. д.

З точки зору впливу завад на вірність передачі канали поділяються на симетричні та несиметричні, канали з пам'яттю та без пам'яті, стаціонарні та нестаціонарні. В

симетричному каналі вірогідність передачі не залежить від статистики передаваної послідовності \vec{B}_i , а повністю визначається статистикою помилок \vec{E}_i . Стационарним каналом без пам'яті називають такий, в якому імовірність помилки в прийманні елемента не залежить від номера елемента в передаваній послідовності та не залежить від значень (нуль або одиниця) елементів, які йдуть перед ним або за ним.

Імовірностною моделлю дискретного каналу називають кінцеве описання каналу, яке дозволяє при відомих ймовірностях вхідного сигналу знаходити любі ймовірності характеристики дискретного каналу, не звертаючись до експериментальних даних або додатковим припущенням. Оскільки в симетричному каналі статистика \vec{B}_i^* повністю визначається статистикою помилок \vec{E}_i та не залежить від статистики \vec{B}_i , то для його опису достатньо потоку помилок \vec{E}_i .

При незалежних помилках достатньо знати єдиний параметр $p_{ном}$, щоб знайти розподілення випадкової величини. На основі теореми Бернуллі імовірність появи в n -елементній комбінації дорівнює t помилок $P(t, n)$ знаходиться біноміальним розподілом

$$P(t, n) = C_n^t p_{ном}^t (1 - p_{ном})^{n-t}, \quad (3.1)$$

при $0 \leq t \leq n$. З (3.1) видно, що ймовірність прийому неспотвореної комбінації ($t = 0$)

$$P(0, n) = (1 - p_{ном})^n, \quad (3.2)$$

Імовірність прийому помилкової комбінації, тобто комбінації, яка містить хоча б одну помилку

$$P(\geq 1, n) = 1 - P(0, n) = 1 - (1 - p_{ном})^n, \quad (3.3)$$

Імовірність появи m та більше помилок в комбінації довжиною n

$$P(\geq m, n) = \sum_{t=m}^{t=n} C_n^t p_{ном}^t (1 - p_{ном})^{n-t}, \quad (3.4)$$

Проста модель помилок дає лише приблизний стан до реальних каналів. Подальше дослідження потоку помилок в реальних каналах показали, що помилки в каналах зв'язку групуються. Ймовірність помилки в групі (пачці) помилок різко збільшується і складає значно більше $P_{ном}$.

Для опису ймовірностних закономірностей послідовності двійкових елементів при складних математичних моделях каналу зручно користуватися статистикою елементарних станів каналу, при якій канал описується декількома можливими станами. Число станів каналу та умовні ймовірності помилки ε в кожному із станів можуть бути різними. В самому першому приближенні канал може мати два стани – "гарний" та "поганий". Статистика станів \vec{D}_i має дві реалізації. Позначимо ймовірності знаходження каналу в "гарному" стані $P(D = 0) = P_0$, а в "поганому" $P(D = 1) = P_1$. Очевидно, що $P_0 + P_1 = 1$. Нехай в обох станах каналу можливі незалежні помилки з умовними ймовірностями ε_0 та ε_1 . Тоді ймовірності помилки

$$p_{ном} = P_0 \varepsilon_0 + P_1 \varepsilon_1. \quad (3.5)$$

При $\varepsilon_0 = 0$ та $\varepsilon_1 = 1$ статистика \vec{D}_i співпадає зі статистикою \vec{E}_i .

Математична модель каналу повинна забезпечувати інженерну можливість розрахунку основних характеристик, знання яких може бути необхідним при оцінці різних методів підвищення вірогідності передачі дискретної інформації. До таких характеристик в першу чергу відносяться: ймовірність помилки (неправильного прийому одиничного елемента) $P_{ном}$; розподілення інтервалів між помилками, а також розподілення довжин серій помилок, пакетів помилок; розподілення ймовірностей $P_n(t)$ – виникнення t помилок в блоці інформації довжиною n .

3.2. Аналіз моделей дискретного каналу

Випадкові дискретні процеси зручно описувати за допомогою ланцюгів Маркова.

Нехай система може знаходитися тільки в одному із K можливих несумісних станів. Позначимо ці стани $c = 0, 1, 2, \dots, K - 1$. Нехай ці стани можуть змінюватися одне іншим (або зберігатися) тільки заздалегідь фіксовані моменти часу $t_{-i}, \dots, t_{-1}, t_0, t_1, \dots, t_i$. В подальшому ці моменти часу будемо називати позиціями (кроками) та позначати їх номерами $-1, 0, 1, \dots, i$. Якщо ймовірність того чи іншого стану системи c_i на i -й позиції повністю знаходиться станом системи на n попередніх позиціях, то випадкова послідовність вказаних станів C_i називається n -зв'язним K -ічним ланцюгом Маркова. В простому випадку, коли c_i залежить тільки від стану на одній попередній позиції ($n = 1$), процес описується однозв'язним або простим ланцюгом Маркова.

Для опису простого ланцюга Маркова необхідно задати умовні ймовірності $p_{c_{i-1}c_i}$ того, що система на i -му кроці перейде в стан c_i при умові, що на $(i - 1)$ -му кроці вона знаходилася в стані c_{i-1} . Вказані умови ймовірності називають перехідними. Якщо початкові ймовірності $P_{c_{-1}c_0}$ вибрати співпадаючими з фінальними ймовірностями P_{c_0} , то ймовірності P_{c_j} при любых j також будуть співпадати з фінальними. Такий ланцюг називається стаціонарний.

Існують основні три моделі, основані на стаціонарних ланцюгах Маркова [4] (табл. 3.1).

Таблиця 3.1 - Моделі, основані на стаціонарних ланцюгах Маркова

Модель	1	2	3
Послідовність, передбачена ланцюгом Маркова	\bar{E}_i	$\bar{\Lambda}_i$	\bar{C}_i
Стан ланцюга	Шумові символи e_i	Інтервали між помилками λ_i	Стан каналу c_i
Можливі стани	0, 1	0, 1, 2, ..., $\lambda_m, > \lambda_m$	0, 1, 2,
Основа ланцюга	1	$\gg 2$	$\gg 2$
Зв'язність ланцюга	$\gg 1$	> 1	1

Всі три моделі дозволяють описати реально спостережамі \bar{e}_i (в межах від 1 до N) з любою потрібною точністю. Проте зручність використання моделей та їх складність, а також число параметрів, які описують модель, із підвищенням точності збільшуються не однаково. Як показали спеціальні дослідження, найбільш зручною з практичної сторони виявилася третя модель, яка описує послідовність стану каналу \bar{C}_i . По цій моделі дискретний канал при передачі кожного елемента \bar{b}_i із послідовності вхідних елементів \bar{B}_i знаходиться в одному із K можливих станів, які визначаються стаціонарним простим ланцюгом Маркова \bar{C}_i .

Послідовність помилок \vec{E}_i повністю визначається матрицею перехідних ймовірностей $p_{c_{i-1}c_0}$ порядку K та умовними ймовірностями помилок в кожному стані $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{h-1}$. Послідовність вихідних елементів $\vec{B}_i^* = \vec{B}_i \oplus \vec{E}_i$. Ймовірність помилки в дискретному каналі можна представити як

$$p_{\text{пом}} = \sum_{c=0}^{K-1} P_c \varepsilon_c, \quad (3.6)$$

де ε_c – умовна ймовірність помилки в стані c ($0, 1, 2, \dots, K-1$),

P_c – фінальні ймовірності станів.

Ймовірності $P_c = P_{c0}$, знаходяться по матриці з [1]:

$$P = \left\| p_{c_{-1}c_0} \right\| = \left\| \begin{array}{cccccc} P_{00}P_{01} & \cdot & \cdot & \cdot & \cdot & P_{0,K-1} \\ P_{10}P_{11} & \cdot & \cdot & \cdot & \cdot & P_{1,K-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{K-1,0}P_{K-1,1} & \cdot & \cdot & \cdot & \cdot & P_{K-1,K-1} \end{array} \right\|, \quad (3.7)$$

із системи рівнянь (з урахуванням того, що для стаціонарного процесу $P_{c_{-1}} = P_{c_0}$, якщо $c_{-1} = c_0$):

$$\left. \begin{array}{l} \sum_{c_0=0}^{K-1} P_{c_0} = 1; \\ \sum_{c_{-1}=0}^{K-1} P_{c_{-1}} p_{c_{-1}c_0} \end{array} \right\}, \quad (3.8)$$

при $c_0 = 0, 1, \dots, K-1$.

Фінальні (безумовні) ймовірності P_0 та P_1 знаходяться із співвідношень (3.8):

$$P_0 + P_1 = 1; P_0 = P_0 p_{00} + P_1 p_{10}; P_1 = P_0 p_{01} + P_1 p_{11}. \quad (3.9)$$

З урахуванням того, що сума ймовірностей в кожній строчці матриці дорівнює одиниці, фінальні ймовірності дорівнюють:

$$P_0 = p_{10}/(p_{01} + p_{10}); P_1 = p_{01}/(p_{01} + p_{10}). \quad (3.10)$$

Ймовірність помилки в прийомі одиничного елемента

$$p_{\text{пом}} = P_0 \varepsilon_0 + P_1 \varepsilon_1. \quad (3.11)$$

Ймовірність події, що в блоці довжиною n елементів міститься t помилок, дорівнює

$$P_n = P_\phi P_n(t) J, \quad (3.12)$$

$$J = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix}, \quad (3.13)$$

де J – одиничний вектор-стовбець, який складається із K одиниць,

$P_\phi = \|P_0, P_1, \dots, P_{K-1}\|$ – вектор-строчка фінальних ймовірностей ланцюга \vec{C}_i ,

$P_n(t) = \|P_n(t, c_{n+1}/c_1)\|$ – квадратна матриця $K \times K$ ймовірностей того, що в блоці довжиною n виникає t помилок при умові, що стан ланцюга в момент передачі першого символу було c_1 , а в момент передачі наступного найближчого буде c_{n+1} . З урахуванням вищезазначених формул вираз (3.12) прийме вигляд

$$P_n(t) = \begin{pmatrix} P_n(t, 0/0) & P_n(t, 1/0) & \dots & P_n(t, K-1/0) \\ \vdots & \vdots & & \vdots \\ P_n(t, 0/K-1) & P_n(t, 1/K-1) & \dots & P_n(t, K-1/K-1) \end{pmatrix}. \quad (3.14)$$

Описана модель дискретного каналу дозволяє досягнути достатньої точності при невеликому числі K станів c_i лише на відносно невеликих серіях елементів. При переході до характеристик великої довжини необхідно застосовувати каскадну ймовірностну модель дискретного каналу. В такому випадку дискретний канал в цілому також підлягає Марковській моделі. Використання каскадної моделі значно спрощує інженерне забезпечення при розрахунку параметрів дискретного каналу.

Історично першою моделлю дискретного каналу, яка враховує групування помилок, являється модель Гільберта. По цій моделі канал може знаходитися в одному із двох станів – "гарному", коли помилки неможливі, та в "поганому", коли виникають незалежні помилки з ймовірністю ε . Послідовність станів утворюють простий ланцюг Маркова. Модель Гільберта може бути отримана із загальної марковської моделі, якщо в ній прийняти $K = 2$, $\varepsilon_0 = 0$, $\varepsilon_1 = \varepsilon$.

Для відображення процесу пакетування помилок ймовірності змін станів каналу p_{01} та p_{10} повинні бути значно менше ймовірностей збереження стану каналу p_{00} та p_{11} .

Ймовірність помилки в каналі

$$p_{ном} = \varepsilon P_1 = \varepsilon p_{01} / (p_{01} + p_{10}). \quad (3.15)$$

Так як середня ймовірність помилки в дискретному каналі повинна бути (при пакетуванні помилок) значно менше умовної ймовірності помилки в пакеті, то $p_{ном} \ll \varepsilon$ та $p_{01} < p_{10}$.

Ймовірність виникнення пакету помилок (або проміжку між помилками) з даного одиничного елемента

$$p_n = p_{01} p_{10} / (p_{01} + p_{10}). \quad (3.16)$$

При групуванні помилок $p_n > p_{ном}$. Тому $p_{10} < \varepsilon$.

Якщо в моделі Гільберта прийняти $\varepsilon_0 = 0$, а $\varepsilon_1 = 1$, то вийде випадок моделі, в якій послідовність \vec{C}_i співпадає з послідовністю \vec{E}_i . Тут проста модель Маркова описує безпосередньо потік помилок.

Розрахунок ймовірності $P_n(t)$ виконується по такій же самій методиці, який викладений при розгляданні моделі Маркова.

По Моделі Сміта – Боуена – Джойса дискретний канал має три стани. Перший стан – "поганий", в якому виникають незалежні помилки з ймовірністю ε_0 , близькою до 0,5. Такий стан відповідає пакета помилок. Два інших стани – "гарні", один з яких відображає проміжки між пакетами помилок, а інший – проміжок між групами пакетів. Перехід із одного "гарного" стану в інший неможливий. Ймовірності незалежних помилок в "гарних" станах ε_1 та ε_2 значно менше 0,5, хоча і необов'язково однакові.

Послідовність трьох станів \vec{C}_i є простим ланцюгом Маркова. Очевидно, що модель Сміта – Боуена – Джойса може бути отримана із загальної моделі Маркова, якщо $K = 3$, $p_{01} = p_{10} = 0$.

Матриця перехідних ймовірностей

$$P = \begin{pmatrix} p_{00} & 0 & p_{02} \\ 0 & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix}. \quad (3.17)$$

Ймовірність помилки в дискретному каналі

$$p_{ном} = \varepsilon_0 P_0 + \varepsilon_1 P_1 + \varepsilon_2 P_2 \quad (3.18)$$

З урахуванням рівності $p_{01} = p_{10} = 0$ фінальні ймовірності дорівнюють:

$$\begin{aligned} P_0 &= \frac{P_{12}P_{20}}{P_{12}P_{20} + P_{02}P_{21} + P_{02}P_{12}}; \\ P_1 &= \frac{P_{02}P_{21}}{P_{12}P_{20} + P_{02}P_{21} + P_{02}P_{12}}; \\ P_2 &= \frac{P_{02}P_{21}}{P_{12}P_{20} + P_{02}P_{11} + P_{02}P_{12}}; \end{aligned} \quad (3.19)$$

Ймовірність $P_n(t)$ знаходиться таким же способом як і в (двох) попередніх випадку.

Як загальну математичну модель каналу управління найбільше використовують модель Л. П. Пуртова. Моделі за ланцюгів Маркова вимагають більш точного описування вихідних параметрів і більшого обсягу обчислень, тому їх застосовують для описування конкретних каналів. Модель Л. П. Пуртова описує ймовірність помилки кратності, яка дорівнює або більша, ніж t у блоці завдовжки n елементів залежно від імовірності помилки в елементі p і коефіцієнта групування помилок α :

$$P(Nk) = \left(\frac{Nk}{t} \right)^{1-\alpha}, \quad (3.20)$$

де Nk – довжина кодової комбінації.

Застосування моделі обмежене значеннями кратності помилки $t < \frac{Nk}{3}$, характерними

для більшості каналів.

Коефіцієнт групування помилок лежить у межах 0,5 ... 0,7 – для кабельних і 0,2 ... 0,4 – для радіоканалів. Граничні значення: $\alpha = 0$ – незалежні помилки і $\alpha = 1$ – всі помилки збираються в один пакет.

Дослідимо залежності $P(t)$ імовірності помилок кратності t у кодовій комбінації, використовуючи вищезгадану модель. Нехай кодова комбінація має $Nk = 128$ розрядів при ймовірності незалежної помилки $p = 10^{-6}$ і коефіцієнтах групування помилок $\alpha = 0,5; 0,6$ та $0,7$.

На рисунку 3.1 показані графіки цих залежностей.

З рисунку 3.1 видно, що, по-перше, при збільшенні коефіцієнта групування помилок за інших рівних умов зменшується ймовірність помилки кратності t , причому ця закономірність зберігається у всій області застосування моделі Л. П. Пуртова. В інших випадках імовірність помилки кратності t завжди більше ймовірності помилки кратності $t + 1$. І нарешті, зі збільшенням коефіцієнта групування помилок різниця між ймовірностями помилок кратності t і $t + 1$ зменшується.

Аналізуючи вищенаведений рисунок, можна зробити висновок, що в реальних каналах передавання керуючої інформації, помилки мають властивість до групування. В даному випадку простежується тенденція, що при даній кількості біт інформації ($Nk = 128$) найбільша ймовірність появи пачки помилок є тоді, коли довжина цієї пачки складає $t \rightarrow 5$. Отже, припустимо, що кратність помилки буде складати 4. Для виправлення цієї помилки, необхідно буде застосувати завадостійкий код, що буде мати велику коригувальну здатність, при малій надмірності.

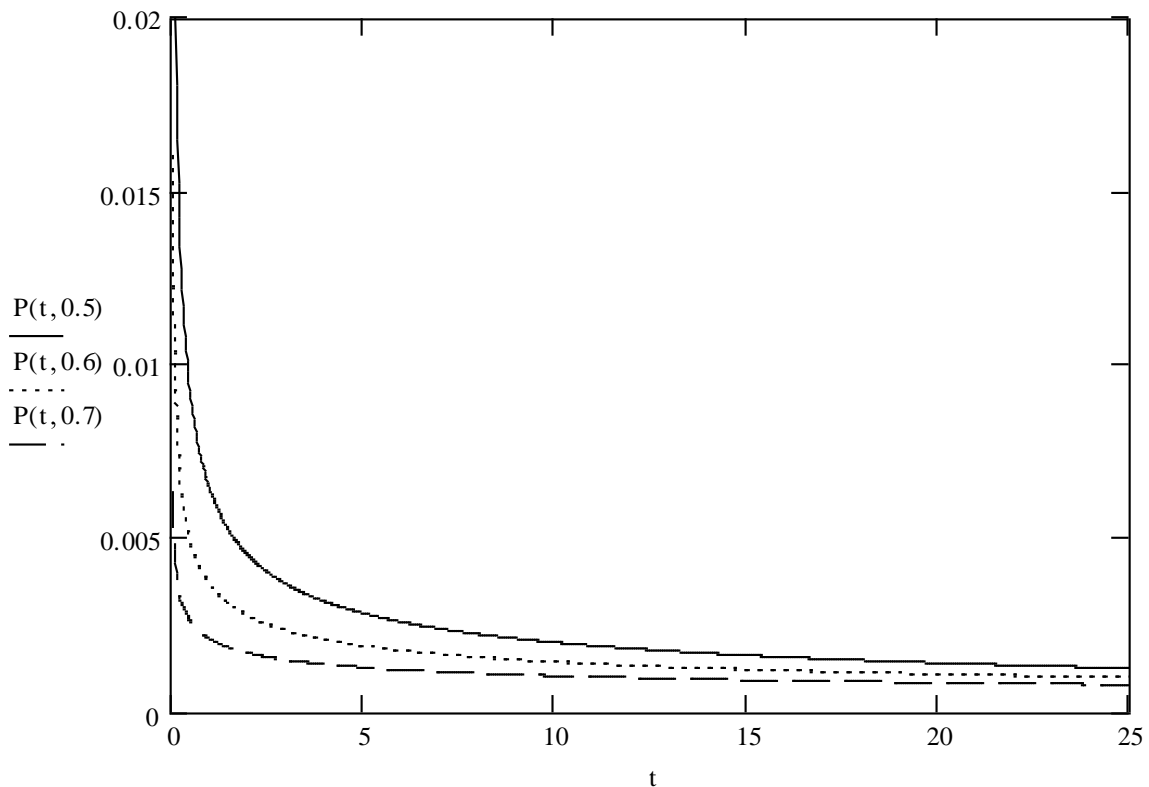


Рис. 3.1. Залежність $P(t)$ у каналах із групуванням помилок

3.3. Інтенсивність помилок та характер їх розподілу у часі

Існують такі способи кодування, які підвищують ефективність передачі інформації: оптимальне кодування, кодування з виявленням помилок і кодування з виправленням помилок [3].

Коди з виявленням та виправленням помилок підвищують завадостійкість передачі інформації.

Для вибору типу коректуючого коду, який використовується для вторинного кодування розглянемо інтенсивність помилок і характер їх розподілу у часі.

Ретельний розподіл статистичних характеристик послідовності помилок у реальних каналах зв'язку показав, що помилки є залежними і мають тенденцію до групування (пакування), тобто між ними є зв'язок – кореляція.

Більшу частину часу інформація на каналах зв'язку проходить без завад, а у деякі моменти часу виникають пакети помилок, в середині яких ймовірність помилки виявляється значно вищою середньої ймовірності помилок, яка знаходиться для досить великого відрізка часу передачі.

При груповому характері розподілу помилок один параметр – ймовірність помилки – не повністю характеризує канал, необхідні додаткові параметри, які відображають ступінь групування помилок у різного виду каналах.

Причини групового характеру помилок у каналах різного типу наступні: тропосферних радіоліній – завмирання та завади; радіорелейних ліній – короткочасні перерви та імпульсні завади; по кабельним лініям зв'язку – перевантаження групових трактів ВЧ апаратури, короткочасні переривання і різного роду станційні та експлуатаційні завади.

Для виділених телефонних кабельних каналів частота помилок лежить у межах від $(2...5) \cdot 10^{-4}$ до 10^{-5} . У комутуючих телефонних каналах $p=10^{-3}...10^{-4}$ у залежності від протяжності каналу, числа і завантаження АТС, які приймають участь у з'єднанні абонентів. Для радіорелейних і тропосферних каналів $p=10^{-3}...10^{-4}$ [8].

У всіх реальних каналах помилки розподілені у часі дуже нерівномірно. Особливо це характерно для кабельних каналів, у яких основна доля помилок зосереджена у визначені години доби, відповідаючи в основному збільшенню завантаження каналостворюючої апаратури і часу зміни чергувань обслуговуючого персоналу.

Оцінка достовірності обміну інформацією, при проектуванні систем передачі інформації, визначається допустимою ймовірністю завад повідомлень $P_{пом}$, при цьому вказується ймовірність завад двійкового символу переданого повідомлення P_3 , у випадку дії незалежних помилок у каналах зв'язку і $(1-P_3)$ – ймовірність відсутності завад. Тоді для двійкової послідовності, яка має n символів:

- ймовірність правильно прийнятої послідовності

$$P_{np}=(1-P_3)^n \quad (3.21)$$

- ймовірність помилки в прийнятій послідовності

$$P_{ном}=1-(1-P_3)^n \quad (3.22)$$

При використанні надлишкових кодів стає можливим знайти і виправити ту чи іншу кратність помилок у залежності від кодової відстані:

$$P_{ном}^i = C_n^i p_3^i (1-p_3)^{n-i} \quad (3.23)$$

де $i=1,2,3$ і т.д. - кратність помилок

$$C_n^i = \frac{n!}{i!(n-i)!} \quad (3.24)$$

Для кодів, які виправляють помилки кратності до t , ймовірність виправлення

$$p_{вип} = \sum_{i=1}^t p_{ном}^i \quad \text{чи} \quad p_{вип} = \sum_{i=1}^t C_n^i p_3^i (1-p_3)^{n-i} \quad (3.25)$$

Тоді можливо сформулювати чотири ситуації:

1) Кодове слово прийнято без помилок (ймовірність цієї події $P_{пр}$).

2) Кодове слово прийнято з помилкою (ймовірність $P_{пом.}$).

$$P_{пом} + P_{вип} = 1 \quad (3.26)$$

3) Кодове слово прийнято з помилкою, яка виправляється з ймовірністю $P_{вип.}$

4) Кодове слово прийняте з помилкою, яка не виправляється цим кодом. Ймовірність цієї події $P_{н.п.}$. Тоді

$$P_{пом} = P_{вип} + P_{н.п.} \quad (3.27)$$

З (3.27) виходить, що ймовірність появи не виправляємих помилок

$$P_{н.п.} = P_{пом} - P_{вип} \quad (3.28)$$

Підставимо в цей вираз формули (3.22) та (3.25) і отримаємо

$$P_{н.п.} = 1 - (1 - P_3)^n - \sum_{i=1}^i C_n^i p_3^i (1 - p_3)^{n-i} \quad (3.29)$$

Цей вираз дозволяє обчислити ймовірність не виправляємої помилки при передачі інформації за допомогою коду, який виправляє t -кратні помилки.

На практиці канали зв'язку характеризуються залежністю ймовірності завад наступного передаваного символу від спотворення попереднього, а також на проходження сигналів впливає сезонна та добова зміна погоди, присутність промислових завад, які змінюють інтенсивність у часі доби і днів тижня, взаємні завади і т.д. Все це приводить до пакетного характеру помилок у каналах передачі інформації.

Розрахунок ймовірностей помилок при пакетному розподілу проводять по формулі:

$$P_{пом} = \frac{p_e}{q} \cdot \sum_{b=1}^{b_{max}} \left(1 + \frac{n-1}{b}\right) \cdot \frac{b \cdot p_b}{\sum_{b=1}^{b_{max}} b \cdot p_b} \quad (3.30)$$

де b - довжина пакета помилок; p_b - умовна ймовірність появи пакета помилок довжиною b ; p_e - ймовірність спотворення двійкового символу; q - густина помилок у пакеті, яка рівна відношенню числа помилок у пакеті до довжини цього пакета b .

Вираз (3.30) визначає ймовірність помилки при передачі послідовності, яка має n символів.

Ймовірність $P_{н.п.}$ і $P_{в.п.}$ для кодів, які визначають пакети помилок, знаходять по формулам

$$P_{н.п.} = \frac{1}{2^r} \frac{p_e}{q} \sum_{b=l_k}^{b_{max}} \left[1 + \frac{n - \mathbf{1}_{l_k+1}}{b}\right] \frac{b p_b}{\sum_1^{b_{max}} b p_b}; \quad (3.31)$$

$$P_{в.п.} = \frac{p_e}{q} \left[\sum_1^{b_{max}} \left(1 + \frac{n+1}{b}\right) \frac{b p_b}{\sum_1^{b_{max}} b p_b} - \frac{1}{2^r} \sum_{b=l_k+1}^{b_{max}} \left[1 + \frac{n - \mathbf{1}_{l_k+1}}{b}\right] \frac{b p_b}{\sum_1^{b_{max}} b p_b} \right]; \quad (3.32)$$

де l_k – довжина знайденого пакету помилок.

Приблизні формули для знаходження $P_{н.п.}$ мають вигляд:

- для кодів, які виправляють помилки,

$$P_{н.п.} = \left(\frac{n}{t+1} \right)^{1-s}, \quad (3.33)$$

де t – кратність виправляємої помилки; s – показник групування помилок;

- для кодів, які знаходять помилки,

$$P_{н.п.} = \frac{P_e}{2^r} \left(\frac{n}{d} \right)^{1-s}, \quad (3.34)$$

де d – мінімальне кодова відстань; r – число перевірочних символів;

- для кодів які виявляють та виправляють помилки,

$$P_{н.п.} = \frac{\sum_{i=0}^t C_n^i \left(\frac{n}{d-t} \right)^{1-s}}{2^r} P_e, \quad (3.35)$$

Ці формули дають хороші результати, якщо число помилок e у комбінації з n символів які задовольняють вимогам $e < 0.3n$ [6].

При передачі інформації простим ненадлишковим кодом достовірність прийому залежить від типу каналу і вид завад у ньому. У більшості випадків достовірність, яка знаходиться недостатня. Її необхідно підвищити, щоб ймовірність помилкового прийому повідомлення споживачем була меншою, ніж ймовірність помилок у повідомленні без прийняття спеціальних заходів.

Застосування надлишкового коду є одним з шляхів підвищення достовірності.

Усі надлишкові коди можливо використовувати для:

- виявлення помилок;
- виправлення помилок;
- виявлення та виправлення помилок.

Для того, щоб збільшити достовірність за допомогою кодів, призначених для виявлення помилок, треба ввести зворотній канал зв'язку. Тоді кодова комбінація, прийнята по прямому каналу, аналізується для того, щоб визначити, чи належить вона до дозволеної комбінації. Дозволена комбінація поступає до споживача після того, як відкидаються перевірочні розряди. Якщо виявляється помилка, по зворотному каналу посилається сигнал запиту, по якому передавальний пристрій повторює передачу інформації. Тому передавальний пристрій повинен зберігати інформацію про відправлені сигнали у плині часу, достатнього для аналізу комбінації прийомним пристроєм і отримання можливого запиту про помилки.

Системи зі зворотнім каналом називають системами зі зворотнім зв'язком. Вони мають наступні переваги:

- виявляюча спроможність коду при однаковій надлишковості вище виправляючої;
- число логічних операцій, яке повинно виконуватися декодером для виявлення помилок, набагато менше, ніж число необхідних операцій для їх виправлення.

Єдиний недолік систем зі зворотнім зв'язком – це зменшення швидкості передачі інформації, але цей недолік виявляється тільки при поганому стані каналу зв'язку.

Виправлення помилок звичайно застосовується у тому випадку, коли у каналі зв'язку мають місце незалежні помилки чи короткі пачки помилок. Якщо вага помилок однакова з довжиною кодової комбінації, то виправлення пачок помилок приводять до невиправданих втрат обладнання на кодуючі та декодуючі пристрої.

Коди, які виправляють помилки, можуть виправляти помилки, вага яких чисельно не перевищує 20-25% довжини кодової комбінації. Найбільш вірогідні помилки з вагою, близькою до 50% довжини кодової комбінації. Тому при необхідності виправлення доречні ті способи, які дозволяють визначити перевірочні імпульси від інформаційних на час, який перевищує ймовірну довжину пачки помилок [8].

Таким чином, вибір способу підвищення достовірності передачі інформації залежить від багатьох факторів: необхідна достовірність прийому, допустима швидкість передачі, залежність від помилок у каналі зв'язку.

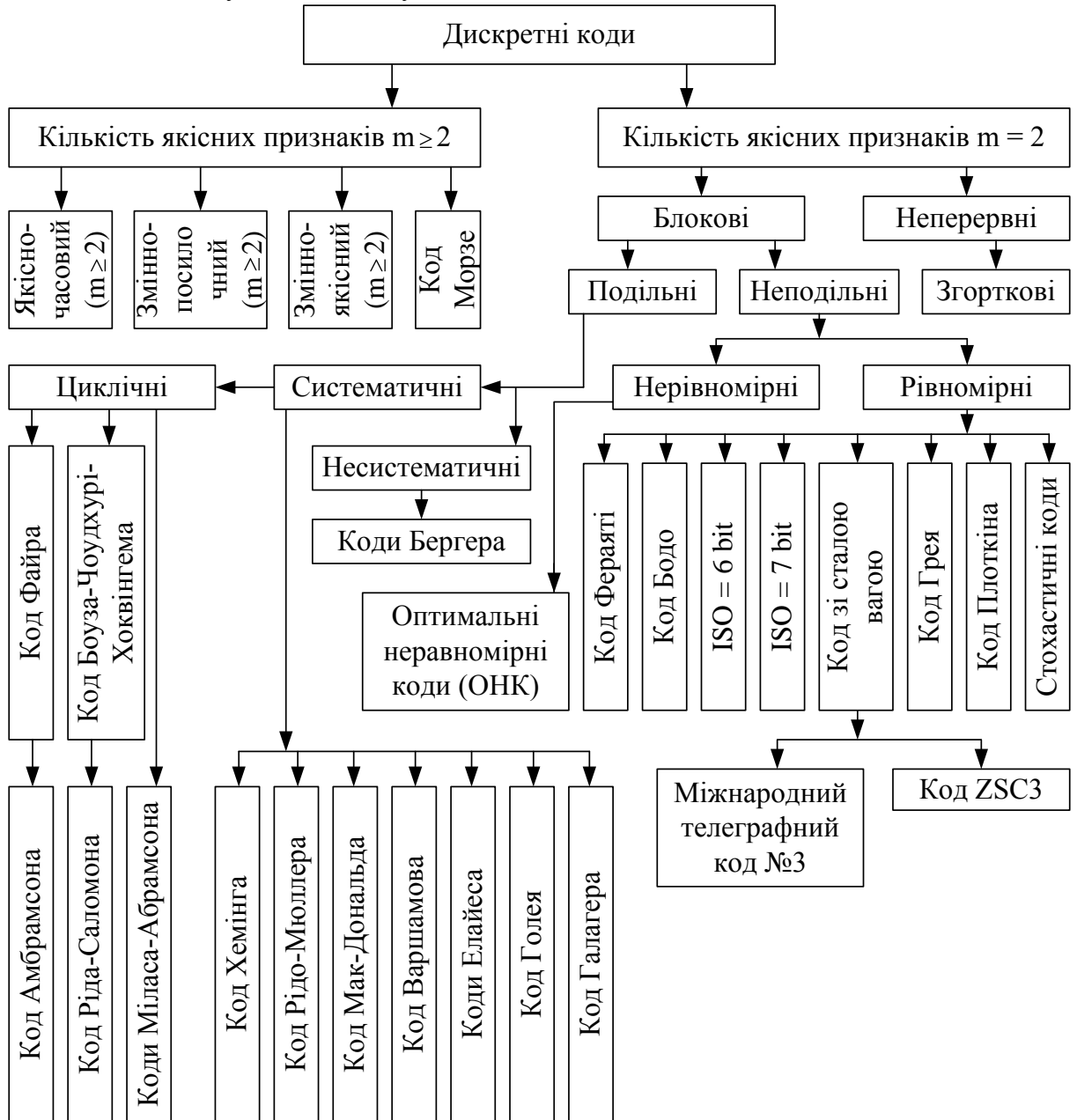


Рис. 3.2. Класифікація найбільш розповсюджених дискретних кодів

Таблиця 3.2. Коди та їх коректуючі здібності

Код	Наявність рекомендацій міжнародних організацій	Помилки	
		незалежні	пакети
		незалежні	пакети

		виявляємі	виправляємі	виявляємі	виправляємі
Кореляційний		+		+	
Інверсний	+	+		+	
Хемінга	+	+	+		
Голея		+	+		
Ріда-Малера		+	+		
Ітеративний	+	+	+	+	
БЧХ	+	+	+	+	+
Мажоритарний	+	+	+	+	
Файра				+	+
Абрамсона				+	+
Міласа-Абрамсона				+	+
Узагальнений код Хемінга	+	+	+		
Ріда-Соломона	+	+	+	+	+
Компаундний	+	+	+	+	+
Плоткіна		+	+		
Ланцюговий	+			+	+
Згортковий				+	+

На даний час розроблені десятки кодів, які теоретично можуть виявляти й виправляти довільну кількість помилок. При такому розмаїтті кодів спроба дати абсолютно точний їхній поділ на самостійні групи й підгрупи таким чином, щоб жодна з них не містила ознак інших груп, заздалегідь приречена на провал. Тому на рис. 3.2 наведено чисто умовний поділ кодів.

Одним з основних факторів, без якого неможливо зробити вірний вибір коду, є характер розподілу помилок у каналі зв'язку. У таблиці 3.2 приведені коди та їх коректуючі здібності.

З таблиці бачимо, що більшість надлишкових кодів виправляє тільки незалежні помилки. Лише невелика група кодів дозволяє виправляти пакети помилок.

Ступінь захисту інформації від помилок відповідним способом кодування залежить головним чином від мінімальної кодової відстані d_{min} даного коду.

Розрізняють три види кодової відстані: Хеммінга, Лі та матричну. Перша знайшла найбільше поширення в теорії кодування. Кодова відстань Хеммінга нероздільно пов'язана з поняттям ваги w кодової комбінації – кількістю її елементів, які не дорівнюють нулю. Кодова відстань Хеммінга d між двома комбінаціями однієї довжини n визначається як кількість однойменних розрядів (позицій), які мають неоднакові елементи. Так, для двійкових кодів, оскільки в двійковій арифметиці додавання однакових елементів дає 0, а неоднакових – 1, відстань Хеммінга між двома кодовими комбінаціями можна визначити порозрядним додаванням їх за модулем 2 та подальшим підрахунком кількості ненульових елементів, тобто визначенням ваги w такої суми.

Загальна кількість кодових комбінацій завдовжки n дорівнює 2^n , а кількість тих з них, які віддалені від заданої на відстань d , – кількості сполучень з n по d :

$$C_n^d = n!/[d!(n-d)!]. \quad (3.36)$$

Для виявлення всіх помилок кратністю v_e кодова відстань має становити $d \geq v_e + 1$, а виправлення помилок кратністю $v_{en} - d \geq 2v_{en} + 1$. Щоб виправити та виявити всі помилки, має виконуватися умова

$$d \geq v_{en} + v_e + 1. \quad (3.37)$$

Через те, що загалом кожний елемент (розряд) комбінації недвійкового (багатопозиційного) коду може мати на відміну від двійкового й понад однієї позиції ($m \geq 1$) з алфавіту q , кодова відстань при цьому визначається виразом

$$d = \sum_{i=1}^m d_i \quad (3.38)$$

де m – кількість позицій в кожному розряді (поодиноківому новому інтервалі, що відповідає тривалості одного елемента) кодової комбінації.

У метриці Хемінга кодова відстань, як і для двійкового коду, визначається кількістю однойменних розрядів з різними позиціями (символами):

$$d_i(x_k, x_l) = \begin{cases} 0, & x_k = x_l; \\ 1, & x_k \neq x_l. \end{cases} \quad (3.39)$$

У метриці Лі

$$d_i(x_k, x_l) = \min\{|x_k - x_l|, q - |x_k - x_l|\} = \min\{d_{j_{\text{mod}}}, q - d_{j_{\text{mod}}}\}, \quad (3.40)$$

де $d_{j_{\text{mod}}} = |x_k - x_l|$.

У модульній метриці $d_i(x_k, x_l) = |x_k - x_l|$, тобто слід виконувати віднімання за модулем q .

Найбільш раціональними треба вважати системи передачі інформації, у яких надлишкові коди використовуються тільки для виявлення помилок, тому що у реальних каналах часто спостерігаються пачки помилок довжиною у декілька десятків чи сотень символів і для їх виправлення був би потрібний код з довжиною кодової комбінації, яка вимірюється тисячами та десятками тисяч розрядів, що технічно важко здійснити.

Якщо відомий характер розподілення помилок у каналі, то це ще не говорить про те, що такі помилки присутні тільки в цьому каналі, а не модему, який використовується при дослідженні і можуть змінюватися тільки при зміні швидкості передачі, потужності сигналу і т.п. Характер помилок у значному ступені залежить від модему, що використовується. Тому завжди необхідно вибір модему і коду розглядати як єдину задачу та вибрати оптимальне рішення.

Після з'ясування характеристик каналу зв'язку подальший вибір коду характеризується вірогідністю помилки. По цьому параметру обираються коди у яких вірогідність виявлення помилки $P_{\text{н.п.}}$ менша за $P_{\text{пом.доп.}}$.

Для забезпечення швидкості передачі інформації, яка задана, потрібно вибрати код з мінімально необхідним числом перевірок розрядів, які забезпечують $P_{\text{н.п.}}$. При цьому не треба забувати про те, що виявляючі властивості визначаються не тільки числом перевірок розрядів, але і видом перевірок співвідношень, а для циклічних кодів - образуючим поліномом. Широке розповсюдження набули методи відбору кодів, які базуються на моделюванні реальних потоків за допомогою ЕОМ для визначення виду

перевірочних співвідношень чи виду образуючого поліному. Крім того, при виборі коду необхідно враховувати складність кодуєчих та декодуєчих пристроїв та їх надійність.

Отже, виходячи з таблиці 3.2 можемо сказати, що порівняно з іншими, найкращим кодом, що не тільки виявляє, а й виправляє незалежні (одиначні) помилки і має рекомендації міжнародних організацій та є порівняно простим в реалізації є код Хеммінга, як двійковий, так і узагальнений. Що ж стосується як пакетів, так і незалежних помилок, то найкращими по тим же параметрам є код БЧХ, Файра та код Ріда-Соломона.

Незважаючи на всі переваги завадостійкого кодування, все ж таки одним з його недоліків є збільшення інформаційного масиву. В таблиці 3.3 показано коефіцієнт збільшення інформаційного масиву для визначених вище оптимальних кодів для різних ймовірностей помилки одного елемента в лінії (каналі) зв'язку.

3.4 Двійкові групові коди

Лінійний систематичний груповий (блоковий) код

Застосувавши теорію лінійних просторів, можна дати таке визначення: лінійним систематичним груповим двійковим n, k -кодом називається код, у якого перевірні елементи b_j (де $j = l \dots r$) знаходяться як суми за модулем 2 обумовлених інформаційних елементів a_i (де $i = 1 \dots k$). У нелінійних двійкових n, k -кодах перевірні елементи знаходяться інакше.

Таким чином, у лінійному коді перевірні елементи визначаються як

$$\begin{aligned}
 b_1 &= \sum_{i=1}^k O \alpha_{ji} a_i; \\
 &\dots\dots\dots \\
 b_r &= \sum_{i=1}^k O \alpha_{ji} a_i,
 \end{aligned}
 \tag{3.41}$$

де $\sum O$ – знак суми за модулем 2; α_{ji} – коефіцієнти, значення яких дорівнюють 0 або 1.

Як випливає з (3.41), закон побудови лінійного коду визначається вибором kr коефіцієнтів α_{ji} .

Одна з властивостей лінійних кодів полягає в тому, що сума а модулем 2 будь-яких двох дозволених кодових комбінацій також є дозволеною комбінацією цього коду.

Побудова лінійного коду виконується на основі твірної (породжувальної) матриці. Ця матриця будується так, щоб:

- 1) кількість початкових кодових комбінацій дорівнювала k , тобто кількості інформаційних елементів первинного коду;
- 2) всі початкові кодові комбінації були різними;
- 3) нульова комбінація не входила до складу початкових;
- 4) всі початкові комбінації були лінійно незалежними;
- 5) кількість одиниць в кожній початковій комбінації була не меншою ніж d_{min} ;

Таблиця 3.3. Коефіцієнт збільшення інформаційного масиву

Ймовірність викривлення одного елемента в лінії (каналі зв'язку)	Кратність помилки	Код рекомендований для захисту інформації від помилок	Коефіцієнт збільшення інформаційного масиву (при $k=8$)
1	2	3	4
$1 \cdot 10^{-6}$	1	Хеммінга ($d=3$)	1,50
		Циклічний ($d=3$)	1,50
		Системат. ($d=3$)	1,50
$0,5 \cdot 10^{-5}$	1	Хеммінга ($d=3$)	1,50
		Циклічний ($d=3$)	1,50
		Системат. ($d=3$)	1,50
$1 \cdot 10^{-5}$	1	Хеммінга ($d=3$)	1,50
		Циклічний ($d=3$)	1,50
		Системат. ($d=3$)	1,50
$0,5 \cdot 10^{-4}$	1	Хеммінга ($d=3$) Циклічний ($d=3$) Системат. ($d=3$)	1,50 1,50 1,50
	2	Системат. ($d=5$) БЧХ ($d=5$)	2,25 2,58
$1 \cdot 10^{-4}$	1	Хеммінга ($d=3$) Циклічний ($d=3$) Системат. ($d=3$)	1,50 1,50 1,50
	2	Системат. ($d=5$) БЧХ ($d=5$)	2,25 2,58
$0,5 \cdot 10^{-3}$	1	Хеммінга ($d=3$) Циклічний ($d=3$) Системат. ($d=3$)	1,50 1,50 1,50
	2	Системат. ($d=5$) БЧХ ($d=5$) Компаундний ($d=5$)	2,00 1,84 2,30
$1 \cdot 10^{-3}$	1	Хеммінга ($d=3$) Циклічний ($d=3$) Системат. ($d=3$)	1,50 1,50 1,50
	2	Системат. ($d=5$) Компаундний ($d=5$) БЧХ ($d=5$)	2,25 2,35 2,58
$0,5 \cdot 10^{-2}$	2	Системат. ($d=5$) Файра ($d=5$) БЧХ ($d=5$) Компаундний ($d=5$)	2,25 1,39 2,58 3,52
	3	БЧХ ($d=7$) Файра ($d=7$) Компаундний ($d=7$) Узагальнений код Хеммінга Ріда-Соломона	3,87 1,59 3,69 1,87 2,37

б) кодова відстань між будь-якими парами початкових комбінацій також була не меншою ніж d_{min} .

де $\sum O$ – знак суми за модулем 2; α_{ji} – коефіцієнти, значення яких дорівнюють 0 або 1.

Як випливає з (3.41), закон побудови лінійного коду визначається вибором kr коефіцієнтів α_{ji} .

Одна з властивостей лінійних кодів полягає в тому, що сума а модулем 2 будь-яких двох дозволених кодових комбінацій також є дозволеною комбінацією цього коду.

Побудова лінійного коду виконується на основі твірної (породжувальної) матриці. Ця матриця будується так, щоб:

- 1) кількість початкових кодових комбінацій дорівнювала k , тобто кількості інформаційних елементів первинного коду;
- 2) всі початкові кодові комбінації були різними;
- 3) нульова комбінація не входила до складу початкових;
- 4) всі початкові комбінації були лінійно незалежними;
- 5) кількість одиниць в кожній початковій комбінації була не меншою ніж d_{min} ;
- 6) кодова відстань між будь-якими парами початкових комбінацій також була не меншою ніж d_{min} .

Підібрані за цим правилом початкові комбінації записуються у вигляді твірної матриці $G_{(n,k)}$, яка містить k рядків і n стовпців:

$$G_{(n,k)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} & b_{11} & b_{12} & \cdots & b_{1r} \\ a_{21} & a_{22} & \cdots & a_{2k} & b_{21} & b_{22} & \cdots & b_{2r} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{k1} & a_{k2} & \cdots & a_{kk} & b_{k1} & b_{k2} & \cdots & b_{kr} \end{bmatrix}. \quad (3.42)$$

Матрицю (3.42) подають також у вигляді двох під матриць: інформаційної E_k та перевірної $C_{(r,k)}$. Першу зручно записати в канонічній формі як одиничну під матрицю, що має k стовпців і стільки ж рядків:

$$E_k = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \quad (3.43)$$

Перевірною під матрицею $C_{(r,k)}$ будується підбором r -розрядних комбінацій з кількістю одиниць в рядку не меншою від $(d_{min} - 1)$. При цьому необхідно врахувати, що сума за модулем 2 будь-яких розрядів повинна мати більш як $(d_{min} - 2)$ одиниць.

Методи виправлення помилок у систематичному груповому коді. Розрізняють два основних таких методи: за допомогою кодового синдрому та за допомогою кодів-супутників.

Використання кодового синдрому. Перевірка кодової комбінації, що приймається, при цьому виконується зіставленням її перевірних розрядів з перевірними розрядами, які обчислюються на основі прийнятих інформаційних.

За відсутності помилки комбінація синдрому (пізнавач помилок) складається з одних нулів. Присутність хоча б одного ненульового елемента в комбінації синдрому вказує на спотворення елемента у прийнятій кодовій комбінації.

Між комбінацією синдрому та помилковою комбінацією, що спричинила створення синдрому, немає взаємної однозначної відповідності, тому за кожним синдромом

закріплюється та комбінація помилок, що виправляється, виникнення якої в каналі є найімовірнішим.

Побудова синдромів для виправлення подвійних, потрійних і помилок більшої кратності складна. Тому метод виправлення помилок за допомогою кодового синдрому використовується головним чином для виправлення поодиноких помилок, імовірність виникнення яких значно більша, ніж помилок більшої кратності.

Використання кодів-супутників. Цей метод передбачає побудову кодової таблиці з кодами-супутниками. Таким чином, для кожної кодової комбінації V_i лінійного систематичного коду дістають свою групу кодів-супутників, розташованих у відповідному стовпці, які зберігаються в пам'яті ЕОМ. У разі приймання комбінації, що збігається з одним із кодів-супутників, спотворена комбінація розшифровується як початкова робоча комбінація, до якої належить цей код-супутник.

Недоліками розглянутого методу є велика ємність пам'яті ЕОМ і значні затрати часу на перебір комбінацій кодів-супутників.

Коди Хемінга

Це одні з найпоширеніших систематичних кодів, які виправляють помилки. До кодів Хеммінга належать коди з мінімальною кодовою відстанню $d_{min} = 3$, що виправляють всі поодинокі помилки.

Формування r перевірних елементів у комбінаціях цих кодів виконують за k інформаційними елементами. Таким чином, довжина кодової комбінації $n = k + r$. Перевірними елементами є лінійні комбінації інформаційних елементів, тобто зважені суми інформаційних елементів з ваговими коефіцієнтами 1 та 0. Послідовність одиниць і нулів у кодовій комбінації називається ще кодовим вектором.

При передачі кодового вектора може бути спотворений будь-який елемент, кількість таких ситуацій $C_n^1 = n$. До цього слід додати ще одну ситуацію, коли помилка не виникає. Таким чином, загальна кількість 2^r комбінацій перевірних елементів має перевищувати кількість можливих помилкових ситуацій в коді з урахуванням відсутності помилок для правильного розрізнення їх і визначення місць помилки:

$$2^r \geq n + 1 \quad (3.44)$$

Оскільки $2^n = 2^{k+r} = 2^k \times 2^r$, можна записати

$$2^n \geq (n + 1) \times 2^k \quad (3.45)$$

2^n – повна кількість комбінацій коду.

Код будується таким чином, щоб в результаті $r = n - k$ перевірок отримати r -розрядне двійкове число, яке вказує на номер спотвореної кодової комбінації. Для цього перевірочні символи повинні мати знаходитись на номерах позицій, які виражаються степінню 2 ($2^0, 2^1, 2^2, \dots, 2^{r-1}$), так як кожний з них входить тільки в одне із перевірочних рівнянь. Таким чином, контрольні символи повинні знаходитись на першій, другій, четвертій і т. д. позиціях.

Перевірочна матриця коду повинна мати n стовпчиків та r рядків. Кожний стовпчик повинен складати двійкову комбінацію, яка вказує номер відповідної позиції коду.

Розширений код Хеммінга з кодовою відстанню $d_{min} = 4$ називається розширеним. Він забезпечує виправлення всіх однократних і виявлення всіх дво- та трикратних помилок. Для цього вводиться додатковий перевірний розряд b_0 який дописується до перевірної матриці Хеммінга з кодовою відстанню $d_{min} = 3$, завдяки чому остання збільшується до 4.

Код Хеммінга має істотний недолік: при виявленні будь-якого числа помилок він виправляє лише одиночні помилки. Надмірність семиелементного коду Хеммінга дорівнює

0,43. При збільшенні розрядності кодових комбінацій збільшується число перевірок, але зменшується надмірність коду. До того ж код Хеммінга не дозволяє знайти групові помилки, сконцентровані в пакетах. Довжина пакета помилок являє собою збільшену на одиницю різниця між іменами старшого і молодшого помилкових елементів.

Циклічні коди

Циклічні коди отримали досить широке застосування завдяки їх ефективності при виявленні та виправленні помилок. Назва кодів пішла від їх властивості, яка полягає у тому, що кожна кодова комбінація може бути отримана шляхом перестановки символів комбінації, яка належить цьому ж коду. Подання комбінацій в циклічних кодах виконують у вигляді поліномів формальної змінної x , що дає змогу звести дії над кодовими комбінаціями до дій над поліномами.

Лінійні систематичні n, k -коди, в яких циклічний зсув $a_{n-2}, a_{n-3}, a_{n-4}, \dots a_2, a_1, a_0, a_{n-1}$ дозволеної комбінації $a_{n-1}, a_{n-2}, a_{n-3}, \dots a_1, a_0$ також є дозволеною комбінацією, що належить цьому коду, називаються циклічними. Така циклічна перестановка елементів виникає після множення полінома на x .

Якщо прийняти умову, що поліном $P(x)$ є дільником двочлена $(x^n + 1)$, то базові комбінації, а разом з ними й всі дозволені комбінації коду здобувають властивість подільності на $P(x)$. Із цього слідує, що приналежність кодової комбінації до групи дозволених можна легко перевірити діленням її полінома на утворюючий поліном $P(x)$. Якщо залишок від ділення дорівнює нулю, то комбінація є дозволеною.

Ця властивість циклічного коду використовується для виявлення або виправлення помилок. Дійсно, якщо під впливом перешкод дозволена кодова комбінація трансформується в заборонену, то помилка може бути виявлена по наявності залишку при діленні комбінації на утворюючий поліном $P(x)$.

Таким чином, утворюючий поліном $P(x)$ повинен задовольняти вимозі – він повинен бути дільником двочлена $(x^n + 1)$. Вибір $P(x)$ однозначно визначає циклічний код і його коригувальні властивості.

Циклічні коди можуть бути двійковими та недвійковими.

Коди Боуза – Чоудхурі - Хоквінгема

Ці коди є різновидом циклічних кодів з кодовою відстанню $d_{min} > 5$. Вони дають змогу виявляти та виправляти будь-яку кількість помилок. При кодуванні задаються кількістю помилок, яку слід виправити, або мінімальною кодовою відстанню та загальною кількістю n елементів у кодовій комбінації. Кількість інформаційних k і перевірних r елементів визначають при побудові коду Боуза - Чоудхурі - Хоквінгема (БЧХ).

Декодування кодів БЧХ (виявлення та виправлення помилок) завдовжки $n = 15$ може виконуватися з використанням методики, яку застосовують для циклічних кодів з $d_{min} < 5$. При декодуванні кодів БЧХ з довжиною комбінацій $n > 15$ можуть виникнути деякі труднощі, пов'язані з великим обсягом обчислень для виявлення та виправлення помилок. У таких випадках при $k > n/2$, де k – кратність зсуву, рекомендується комбінацію, утворену після k – кратного зсуву і підсумовування з остачею, зсувати не праворуч, а ліворуч на n циклічних кроків.

Кількість помилок, які можуть виправляти коди БЧХ, не обмежена, але зі збільшенням кратності помилки значно зростає складність пристроїв декодування, що призводить до зменшення швидкості передачі інформації.

Код Файра

Двійковий код Файра призначений для виправлення поодиноких пачок помилок, для чого він потребує значно меншої кількості перевірних елементів порівняно з кодами БЧХ.

Під пачкою (пакетом) помилок розуміють не тільки групу помилок, розташованих поруч, а й групу або кілька спотворених і неспотворених елементів, які знаходяться між

двома спотвореними елементами. В останньому випадку до пачки помилок, крім двох крайніх спотворених елементів пачки, належать ще спотворені та неспотворені елементи, розташовані між ними, тобто в середині пачки.

Твірний поліном коду Файра визначається виразом

$$P_{\Phi}(x) = P(x)(x^c + 1) \quad (3.46)$$

де $P(x)$ – незвідний поліном степеня l , що належить h ; c – просте число, яке не повинно ділитися на h без остачі.

Поліном $P(x)$ має деякий степінь h , якщо h – найменше додатне число таке, що двочлен $x^h + 1$ ділиться на $P(x)$ без остачі. Для будь-якого l існує принаймні один незвідний поліном $P(x)$ степеня l , який належить числу

$$h = 2^l - 1. \quad (3.47)$$

Незвідний поліном $P(x)$ вибирається з таблиці так, щоб виконувалася умова (3.11), причому $l \geq b$, де $b = v_{en}$ – довжина пачки помилок.

Довжина коду Файра визначається виразом

$$n = \text{НСК}(c, h), \quad (3.48)$$

тобто є НСК чисел c та h , тому що тільки в цьому разі двочлен $x^h + 1$ буде ділитися на поліном $P_{\Phi}(x)$ без остачі.

Кількість перевірних елементів цього коду визначається так:

$$r = c + 1, \quad (3.49)$$

а інформаційних – так:

$$k = n - c - l \quad (3.50)$$

Код Файра виправляє будь-яку поодинокую пачку помилок завдовжки b або менше й одночасно виявляє будь-яку пачку помилок завдовжки $B \geq b$ або менше, якщо $c \geq b + B - 1$ і $l \geq b$.

Якщо користуватися цим кодом тільки для виявлення помилок, то можна виявити будь-яку комбінацію з двох пачок помилок, довжина найменшої з яких не перевищує l , а сума довжин обох пачок менша, ніж $c + 1$. Можна також виявити будь-яку поодинокую пачку помилок з довжиною, не більшою від $r = c + l$, де r – кількість перевірних елементів.

Код із багатократним повторенням

Код із багатократним повторенням (без інверсії) є подільним лінійним кодом. Він містить k інформаційних і $n_R k$ перевірних елементів, де $n_R \geq 2$ – кількість повторень початкової кодової комбінації. В цьому коді кожен k перевірних елементів є просто повтореними інформаційними елементами

$$b_j = b_{j+2k} = b_{j+3k} = \dots = b_{j+(n_R-1)k} = a_j, j = 1 \dots k. \quad (3.51)$$

Через те, що код має кодову відстань $d_{min} = n_R + 1$, він може використовуватися для виявлення та виправлення помилок. Процедура виявлення помилок у прийнятій кодовій комбінації полягає в порівнянні однойменних інформаційних і перевірних елементів. Незбіг їх свідчить про наявність помилок у прийнятій комбінації.

Ітеративні коди

Ітеративні коди характеризуються двома або більшою кількістю перевірок усередині кодової комбінації, а властивості цих кодів повністю визначаються їх параметрами.

Довжина n кодової комбінації, кількість інформаційних параметрів k та мінімальна кодова відстань d_{min} визначаються виразами.

$$n = \prod_{i=1}^S n_i; \quad k = \prod_{i=1}^S k_i; \quad d_{min} = \prod_{i=1}^S d_{min}, \quad (3.52)$$

де n_i, k_i, d_{min_i} – параметри ітерованих кодів; S – кратність ітерування; \prod – знак множення.

На практиці широко застосовуються двовимірні лінійні ітеративні коди з кодуванням за рядками та стовпцями з однією перевіркою на парність.

Суттєвим недоліком ітеративних кодів є порівняно висока надмірність їх, яка значно перевищує надмірність циклічних кодів, здатних виправляти ту саму кількість помилок за інших однакових умов. Однак їх використання в системах передачі даних зумовлює більш просте порівняно з циклічними кодами кодування та декодування за допомогою ЕОМ.

Каскадні коди

Збільшення мінімальної кодової відстані d_{min} і, як наслідок, здатності коду виправляти помилки можна досягти, якщо застосувати кілька ступенів кодування (каскадний принцип кодування). Такі коди дістали назву каскадних.

На практиці поширеними є каскадні коди, що складаються з двох кодів (два ступені кодування), які називаються внутрішнім і зовнішнім. При цьому зовнішній код використовується для кодування повідомлень, що надходять від джерела у вигляді первинного коду, а внутрішній – для кодування комбінацій зовнішнього коду перед передачею їх у канал зв'язку.

Як зовнішній код звичайно використовуються коди Ріда - Соломона або коди БЧХ. Вибір внутрішнього коду залежить від характеристик каналу зв'язку та інтенсивності виникнення помилок. Це може бути код БЧХ, код Хеммінга чи інший код. Взагалі задачею внутрішнього коду є забезпечення прийнятої ймовірності помилки, а зовнішнього – зниження результуючої ймовірності неправильного декодування до заданого значення. Каскадні коди знайшли широке застосування для передачі повідомлень по радіоканалах із великим рівнем завад, зокрема в супутникових лініях зв'язку.

3.5. Рекурентні коди

Рекурентними (неперервними) називаються коди, що подаються неперервною послідовністю кодових елементів без поділу на окремі комбінації.

Рекурентні коди дають суттєвий ефект при захисті інформації, яка передається по каналах, де можливе виникнення помилок великої кратності та початок помилок.

Від блокових рекурентні коди відрізняються тим, що дають змогу кодувати інформаційну послідовність неперервно, не поділяючи її на блоки фіксованої довжини n з k інформаційними та r перевірними елементами. Такі коди ще називаються ланцюговими.

Коректувальна здатність ланцюгового коду залежить від кроку додавання $t_{кр}$. Зміною довжини кроку $t_{кр}$ коректувальну здатність коду можна узгоджувати з характеристиками каналу зв'язку, зменшуючи чи збільшуючи допустиму частість помилок.

3.6. Недвійкові коди

Код із багатократним повторенням

Відмінність цього коду від q -коду з повторенням полягає в багатократному ($n_R \geq 3$) повторенні початкової кодової комбінації. Даний метод кодування застосовується при передачі інформації по каналах з високим рівнем завад для істотного підвищення вірогідності, коли немає можливості для цієї мети використати зворотний канал.

Якщо застосування двійкового коду з багатократним повторенням потребує відповідного аналогічного q -коду дає змогу або зберегти колишню швидкість передачі, або зменшити її несуттєво. Це пояснюється тим, що при використанні q -коду можна ввести аналог n_R -кратного повторення, збільшивши кількість позицій та ознак сигналу (алфавіт коду). В разі передачі інформації по радіоканалах, де, крім селективних завмирань, діють також завмирання сигналу в часі, аналог багатократного повторення можна ввести за кілька часових позицій (інтервалів).

Цей код є характерним представником класу низько швидкісних кодів, тому що його швидкість $R = k/n = 1/(n_R + 1)$.

Узагальнений код Хеммінга

Серед q -кодів найпростішими кодами, які мають алгебричну структуру й забезпечують нескладні процедури кодування та декодування, є лінійні блокові коди, що виправляють одну помилку. В класі двійкових кодів існує аналог їх – код Хеммінга. Хоча між цими кодами є суттєві відмінності, q -код часто називають узагальненим кодом Хеммінга, маючи на увазі узагальнення коду на недвійковий алфавіт $q > 2$.

На відміну від двійкового символами q -коду є елементи q -поля: його перевірна матриця H не є множиною всіх послідовних номерів позицій елементів коду в блоці завдовжки n . В q -коді, крім визначення місця помилки в блоці, необхідно ще знати її значення e .

q -коди прийнято поділяти на дві великі групи: коди з простою основою $q = p$, де $p \in \{3, 5, 7, 11, 13, \dots\}$; коди з основою q , що розкладається. Найбільший практичний інтерес викликає тільки окремий випадок цих кодів при $q = 2^l$, коли символи мають інформаційну ємність l бітів і можуть бути зіставлені з усіма l -розрядними двійковими числами.

Вибір q впливає на визначення операцій додавання, віднімання, множення і ділення під час виконання процедур кодування та декодування. Якщо основа q – просте число, то зручно використати апарат обчислень за модулем цього числа. Якщо ж $q = 2^l$, то необхідно звернутися до алгебричного апарата обчислень за модулем незвідного полінома. Символи коду при цьому ставляться у відповідність елементам скінченного поля порядку q .

Коди Ріда-Соломона

Ці коди використовуються для передачі інформації по каналах з високою інтенсивністю завад, коли виникають помилки кратності два і більше, пачки помилок, а також сполучення пачок і однократних помилок.

Довжина набору визначається вихідним полем. За вихідне поле береться, як правило, 2^l , де l визначає довжину набору (для двійкової підстави набори складаються із двох, чотирьох, восьми й т.д. символів). Практично використовуються тільки двійкові основи коду РС. Декодування цих кодів виконується на основі алгоритму Берлікемпа.

Коди РС відносяться до завершених кодів, тобто мають максимально можливі значення d_0 при заданих величинах n і R . Хоча зусиллями багатьох дослідників складність реалізації декодерів РС доведена при великих n до величин порядку $n^k \ln n$ (де k – невелике ціле число), абсолютна складність у числі операцій, виконуваних декодером, залишається ще дуже великою. Так, при реалізації декодера РС для коду (31, 15) необхідне виконання порядку 17 операцій на кожний біт декодуемого повідомлення, навіть якщо частина операцій проводити паралельно, а множення в недвійкових полях виконувати табличними методами.

Багатовимірні ітеративні коди

Для передачі інформації широко використовуються ітеративні коди, які відзначаються високою здатністю виявляти помилки. Двійковий ітеративний код, що реалізує перевірку на парність за рядками та стовпцями, виявляє всі одно-, дво- й трикратні помилки, деякі чотирикратні помилки і помилки більшої кратності, розташованих в кутах прямокутника.

Підвищення швидкості ітеративного коду можна досягти збільшенням основи q -коду. При $q > 1$ зростає обсяг інформації, що передається, оскільки кількість інформації, яка міститься в одному елементі кодової комбінації, визначається основою q -коду.

Слід відзначити, що кратність помилок, які виправляються q -кодом, вища порівняно з двійковим кодом, оскільки виправлення одного елемента q -коду відповідає виправленню $\log_2 q$ двійкових одиниць.

Недвійковий ланцюговий код

Рекурентні (неперервні) коди призначені в основному для виправлення пачок (пакетів) помилок. Для таких кодів характерним є те, що операції кодування та декодування виконуються над неперервною послідовністю елементів. Це є перевагою рекурентних кодів, оскільки дає більші можливості для використання надмірності, що вводиться.

Найпростіше рекурентні коди реалізуються при надмірності, що дорівнює 0,5, коли кількість перевірних множин (елементів кодових комбінацій) дорівнює кількості інформаційних. Такий код називається ланцюговим. З усіх рекурентних кодів цей код дістав найбільшого поширення.

У ланцюговому q -коді на передачі кожна перевірна множина формується за двома інформаційними, розташованими одна від одної на відстані $t_{кр} \geq 1$.

При передачі повідомлень по каналах, у яких помилки (спотворення елементів кодових комбінацій) групуються в пачки по $t_{кр}$ помилок у кожній, перевірні множини формуються так, щоб вони зв'язували інформаційні множини, що відстоять одна від одної на відстань $t_{кр}$, яка визначається станом каналу зв'язку.

Слід звернути увагу на те, що для можливості виявлення та виправлення спотворених крайніх (перших й останніх) інформаційних елементів у неперервну послідовність елементів на початку та в кінці треба вводити $t_{кр}$ додаткових елементів, які завчасно відомі на приймальному боці.

До недоліків ланцюгового q -коду треба віднести те, що він не виправляє послідовні помилки кратністю, більшою від $2t_{кр}$; надмірність цього коду дорівнює 0,5.

3.7. Завдостійке кодування в каналах передачі

Основні роботи К.Шеннона [1,2], у яких формулюються завдання завадостійкої передачі інформації з будь-якою наперед заданою точністю передачі інформації, пропонують як рішення цих завдань використовувати принцип випадковості використовуваних сигналів. Для завадостійкої передачі інформації пропонується використовувати випадкові (n,k) -коди, утворені шляхом випадкового вибору з 2^n можливих двійкових комбінацій довжиною n 2^k комбінацій, кожна з яких ототожнюється з однією з інформаційних комбінацій довжиною k . Використовуючи цю модель сигналів для передачі по каналу зв'язку К.Шеннон довів теорему про можливість передачі по каналу зв'язку інформації з імовірністю помилки, що залежить від параметрів n і k , і яка може бути зроблена як завгодно мала, шляхом вибору відповідних значень цих параметрів. Доказ цієї теореми мав фундаментальне значення для створення теорії завадостійкого кодування, хоча не давав конструктивних пропозицій про реалізацію такої можливості.

На практиці застосовується відносна невелика група алгебраїчних завадостійких кодів: коди Боуза-Чоудхурі-Хоквінгема (БЧХ), коди Ріда-Соломона (РС) і згорткові коди. Найбільш широко застосовуються циклічні коди з виявленням помилок, які є частковим випадком кодів БЧХ і застосовуються в стандартних протоколах X.25/2 (LAP-B, LAP-M). Знаходять

застосування коди РС із виправленням помилок у каналах радіозв'язку. У каналах супутникового зв'язку, що характеризуються незалежним характером помилок, широко застосовуються згорткові коди. Широкого застосування кодів з виправленням помилок не відбувається за причин складності реалізації виправлення помилок, великої залежності ймовірності помилки декодування від закону розподілу помилок.

У роботах з теорії інформації і завадостійкого кодування, написаних в 70-і роки, розглядалися коди з виправленням помилок. Спочатку коди на основі випадкових кодів К.Шеннона, потім алгебраїчні коди. Це пояснюється досягненням більш високих характеристик при передачі інформації кодами з виправленням помилок, у порівнянні із широко застосовуваними в цей час кодами з виявленням помилок. Перехід від кодів з виправленням до кодів з виявленням помилок можна пояснити декількома основними причинами:

- по-перше, більшою обчислювальною складністю реалізації кодека, що виправляє помилки;
- по-друге, необхідністю узгодження типу й параметрів коду, що виправляє помилки, з умовами передачі інформації, тобто інтенсивності й закону розподілу помилок у використовуваному каналі зв'язку;
- по-третє, застосуванням, як правило, каналів високої якості, високим ступенем відпрацьовування необхідних технічних рішень по реалізації циклічного коду в розроблених мікросхемах сполучення з каналам зв'язку, що випускаються цілим рядом фірм і стандартизацією протоколів канального рівня, у які вписана реалізація циклічного коду.

Тому для розгляду альтернативи застосування кодів з виправленням помилок варто шукати істотні причини для такого переходу. Сформулюємо властивості завадостійкого коду з виправленням помилок, які дозволяють говорити про таку альтернативу, а потім розглянемо можливий варіант побудови й застосування такого коду. Отже, такий код повинен мати властивості:

- Код має режими виявлення й виправлення помилок із забезпеченням в обох режимах гарантованої (наперед заданій) імовірності декодування з помилкою в довільному каналі зв'язку й надійною відмовою від декодування при неможливості виправлення помилки.
- Код має таку виправляючу здатність і дозволяє вибрати такі параметри n і k , що алгоритм передачі інформації, який їх використовує характеризується негіршими ймовірсно-тимчасовими характеристиками в порівнянні із застосуванням альтернативних кодів.
- Код забезпечує в режимі виправлення помилок виділення із заданою точністю частини правильно прийнятих символів навіть при кратності помилки, що перевищує виправляючу здатність коду.
- Код дозволяє декодувати кілька копій (однакових по змісту інформації кодових блоків) блоку з ефективністю, що перевищує ефективність декодування вихідного коду з виявленням чи виправленням помилок. Ця властивість може застосовуватися для роботи в паралельних каналах, при багаторазовій передачі повідомлення по одному каналу або в каналі зі зворотним зв'язком при обробці копій після прийому повтореного блоку.
- Процедури кодування й декодування коду містять тільки операції по модулю 2.
- Метод кодування повинен мати властивості випадковості сигналів на виході кодера, що забезпечують спільне рішення завдань забезпечення завадостійкості в постановці К.Шеннона.

Реалізація такої постановки завдання дозволить:

- Розширити спектр використовуваних каналів зв'язку по допустимому рівню якості каналів за рахунок використання каналів зниженої якості.

- Забезпечення гарантованої вірогідності заданого споживачем рівня (10^{-9} , 10^{-18} , 10^{-27}) при будь-якому виді перекручувань у каналі зв'язку;
- Зняти проблему точності (вірогідності) інформації при створенні глобальних гіперінформаційних просторів за умови передачі інформації практично по будь-яким каналам зв'язку;
- Забезпечити повернення до класичної постановки К.Шеннона в рішенні завдань завадостійкості, але в рамках єдиного алгоритму перетворення інформації.

3.7.1. Побудова і властивості стохастичних кодів з виправленням помилок

В 80-х роках були початі роботи зі створення нової конструкції кодів, що вписуються в структури існуючих мереж передачі даних, з метою підвищення техніко-економічний ефекту при передачі інформації по каналах зв'язку різної якості. В результаті проведених робіт були створені конструкції й алгоритми кодування (декодування) q-ічних стохастичних кодів з виправленням помилок на основі утворюючих двійкових кодів для каналів зв'язку різної якості [3].

Для цих кодів справедливі наступні оцінки, підтверджені теоретичними дослідженнями й статистикою випробувань практично реалізованих комплексів:

а) код забезпечує заздалегідь задану вірогідність (гарантовану ймовірність помилки декодування) як при виявленні, так і при виправленні помилок, пов'язану з обраною довжиною q-ічного символу й дозволеним у цьому випадку числа помилок, що виправляються, і відносно максимально можливого числа виправлених помилок t , пов'язаного з кодовою відстанню вихідного двійкового коду d ,

$$t = d - 2 \quad (3.53)$$

Дану властивість можна використовувати в дуплексних і симплексних каналах зв'язку.

б) У системі зі зворотним зв'язком (дуплексному каналі передачі даних) такий код з виправленням помилок забезпечує (див. наведені нижче таблиці):

- збільшення відносної (ефективної) швидкості передачі інформації, у порівнянні із застосуванням кодів, що виявляють помилки, у всьому інтервалі можливої якості каналу (тобто завжди);
- більш високу ймовірність успішного декодування кодового блоку у випадку виправлення помилок, щодо режиму виявлення помилок; при цьому канал передачі даних здобуває властивості каналу реального часу ("темпового" каналу), де інформація передається при значно меншому числі повторень, що максимально задовольняє вимогам по сполученню передачі даних і мови в одному каналі (цифрова передача мови критична до повторів).

в) сигнал з виходу кодера має характер "білого шуму", тому що використовується не один обраний випадково (n,k) -код, а ансамбль кодів, де на кожному черговому кодовому блоці відбувається зміна коду.

г) при наявності в приймачі двох і більше значень апіорі однакових до кодування на передавальній стороні кодових блоків (перша передача й повторення по запиту в системі зі зворотного зв'язку або багаторазова передача блоку в одноканальних й багатоканальних симплексних системах - "копій" блоків) існують алгоритми декодування копій, що дозволяють значно збільшити надійність доведення повідомлення в умовах інтенсивних перешкод у каналах зв'язку.

У результаті, стверджується, що розглянута конструкція кодів має сферу застосування, що збігається зі сферою застосування інформаційних систем і телекомунікаційної техніки взагалі.

Нижче наведені основні властивості стохастичних кодів, що виправляють помилки з гарантованою ймовірністю помилки декодування.

Основа коду обрана $q = 2^{32}$, тобто двійкова довжина q -ічного символу 32 біт, число таких символів у блоці n і k .

Імовірність помилки декодування стохастичних q -ічних кодів не залежить від виду й характеру перекручувань і зв'язано, в основному, з величиною q як у режимі виявлення помилок $(n, n-1)$ -коду (з одним надлишковим символом), так і в режимі виправлення помилок. При обраній основі q імовірність помилки після декодування при будь-якому характері перекручувань не перевищує

$$P_{\text{пом}} < q^{-1} = 2^{-32} < 10^{-9} \quad (3.54)$$

Число виправлених помилок t пов'язане з кодовою відстанню d вихідного двійкового коду співвідношенням $t = d - 2$ і приблизно відповідає числу виправлених помилок коду Ріда-Соломона з тими ж параметрами n і k (див. значення t у таблиці 3.3).

Примітка: Розглянуті коди виправляють помилки в імовірнісному сенсі, тобто помилка завжди виправляється тільки при кратності 1 (коли перекручений 1 q -чний символ), при кратності 2 і вище з контрольованою ймовірністю, що залежить від основи коду, тобто можлива помилка декодування (або відмова від декодування). Ці ймовірності при практичній реалізації можуть бути зроблені як завгодно малими.

Декодування (і кодування) використовує тільки двійкові операції з q -ічними символами. Так як число операцій декодування не залежить від q , то при збільшенні q знижується число операцій, що припадають на 1 біт переданої інформації. Число операцій декодування з виправленням помилок на один блок q -ічного коду може мати порядок величини βn двійкових операцій з операндами довжини

$$L = -\log q, (L = 32), \quad (3.55)$$

де коефіцієнт $\beta = 5 - 10$. У перерахуванні на 1 біт переданої інформації число операцій зменшується в L раз і має величину менш 1 оп/біт. У режимі виявлення помилок для $(n, n-1)$ -ічного коду число операцій кодування й декодування мінімально й становить n двійкових операцій з операндами довжиною L . Тобто для коду $(16, 15)$ при $q = 2^{32}$ число двійкових операцій кодування (декодування) становить 16 на блок довжиною $16 \times 32 = 512$ біт.

Значення ймовірності успішного декодування кодового блоку $R_{\text{пр}}(1)$ з першої передачі й ефективна швидкість, обчислена по формулі:

$$R_{\text{еф}} = \frac{k \times N_{\text{пр}}}{n \times N_{\text{пер}}} \quad (3.56)$$

де $N_{\text{пр}}$ і $N_{\text{пер}}$ - кількість прийнятих і переданих блоків відповідно, наведені нижче в таблиці 3.4.

Для стохастичних кодів існує алгоритм декодування "копій", коли для двох і більше значень кодового блоку стохастичного коду, у тому числі й для $(n, n-1)$ -коду з виявленням помилок, однакових при їхній передачі, можливо провести спільне декодування розширеного коду з виправленням помилок. При цьому, число помилок, що виправляються в блоці розширеного коду, значно перевищує число помилок, виправляються сумарно в кожному блоці, наприклад, якщо вихідний код виправляє $t=2$ помилки, то при 2-х кратному повторенні вихідного блоку в розширеному блоці виправляється не менш 6 перекручених q -ічних символів, при трикратному повторенні - не менш 10 символів і т.д. Гарантія вірогідності декодуваної інформації при цьому зберігається.

Режим декодування копій найбільш перспективний у симплексних радіоканалах, особливо при низькій якості каналу зв'язку, у тому числі при інтенсивній радіопротидії, а також у дуплексних каналах при спільному декодуванні не раніше декодованого й повтореного блоків.

Тимчасові ("темпові") характеристики залежать як від ефективної швидкості передачі $R_{эф}$, так і при рівній швидкості, від імовірності доведення блоку з першої (другий і т.д.) передачі. З таблиць 3.4, 3.5, 3.6 видно, що для кодів з виправленням помилок $P_{пр}(1)$ значно вище, ніж для кодів з виявленням помилок.

3.7.2 Порівняльна характеристика стохастичних кодів з виправленням помилок, і отримані результати випробувань програмно-апаратних засобів

У таблиці 3.4 наведені результати стендових перевірок стохастичних кодів, отримані з використанням програмного імітатора помилок каналу зв'язку. При проведенні перевірок задавалися значення ймовірності перекручування в каналі q-ічного символу P_q [перекручений випадковим образом у середньому кожний другий (1/2), кожний четвертий (1/4) і т.п. q-ічний символ]. Для спрощення проведення порівняльного аналізу проведено перерахування від заданого значення P_q до значення ймовірності перекручування двійкового символу P_0 для різного ступеня групування помилок оцінюється по моделі Пуртова з коефіцієнтом α (0 – незалежні помилки, 0,3 – слабе групування в провідному каналі, 0,5 – сильне групування в радіоканалі).

У таблиці 3.5 наведені результати порівняння розрахункових значень швидкості передачі при виявленні, виправленні однієї й двох помилок (перекручених q-ічних символів) стохастичним q-ічним кодом (довжина q-ічного символу $L = 8, 16$ або 32 біта) із вказівкою виграшу (с (+) або програшу в значенні швидкості (с (-) при виправленні помилки, щодо режиму виявлення помилки).

Таблиця 3.4. Результати стендових перевірок стохастичних кодів, отримані з використанням програмного імітатора помилок каналу зв'язку

Код	Якість каналу							
	P_q	1/2	1/4	1/8	1/16	1/32	1/64	1/128
	P_0	$\alpha=0$	$\alpha=0,3$	$\alpha=0,5$	$\alpha=0$	$\alpha=0,3$	$\alpha=0,5$	$\alpha=0$
		0,02142	0,00895	0,00416	0,002015	0,00099	0,000492	0,000245
		0,05942	0,02511	0,01173	0,00568	0,00280	0,001391	0,000693
		0,11532	0,04958	0,02332	0,011344	0,00559	0,002779	0,001385
Ефективна швидкість $R_{эф}$								
(4,3)		0,0319	0,1711	0,3339	0,4647	0,5009	0,5256	
(8,7)			0,0573	0,2524	0,4825	0,5817	0,5877	0,6852
(16,15)			0,0076	0,1336	0,3206	0,5073	0,6462	0,7012
(32,31)				0,0096	0,1772	0,3026	0,4455	0,7351
(8,4) (i.про t=2)		0,0635	0,2681	0,3815	0,3982	0,4062	0,4062	0,4062
(8,2) (i.o. t=2)		0,1051	0,1533	0,1559	0,1562	0,1562	0,1562	0,1562
(16,11) (i.o. t=2)			0,1157	0,4383	0,5627	0,6215	0,6405	0,6406

Продовження таблиці 3.4

(15,11) (i.o. t=2)		0,1139	0,2597	0,5483	0,6345	0,6802	0,6802
(16,7) (i.o. t=4)	0,0925	0,3539	0,3869	0,3906	0,3906	0,3906	0,3906
(32,26) (i.o. t=2)			0,1366	0,5330	0,7127	0,7687	0,7739
Імовірність прийому блоку $P_{пр}(1)$							
(4,3)	0,0175	0,3684	0,6315	0,8070	0,8594	0,9298	
(8,7)			0,3428	0,600	0,7142	0,8380	
(16,15)			0,2222	0,3777	0,5333	0,7333	
(32,31)			0,0096	0,1772	0,3026	0,6874	0,7351
(8,4) (i.про t=2)		0,695	0,935		0,9735		
(8,2) (i.o. t=2)	0,6816	0,9825	0,9961	0,9981			
(16,11) (i.o. t=2)		0,2308	0,6615	0,7692	0,9538		0,9724
(15,11) (i.o. t=2)			0,4769	0,8153	0,9076		
(16,7) (i.o. t=4)	0,2190	0,8571	0,9810	0,9905			
(32,26) (i.o. t=2)			0,1470	0,6862	0,8928	0,9608	0,9682

Таблиця 3.5. Результати порівняння розрахункових значень швидкості передачі при виявленні, виправленні однієї й двох помилок стохастичним q-ічним кодом

P_q	Код з виявленням помилок (d=2)			Код з виправлення однієї помилки (d=3)				Код з виправленням двох помилок (d=4)			
	n_{opt}	k_{opt}	R	n_{opt}	k_{opt}	R	$\Delta \%$	n_{opt}	k_{opt}	R	$\Delta \%$
0.500	2	1	0.008	7	4	0.002	-0.2	8	4	0.008	0
0.400	3	2	0.019	7	4	0.012	-0.7	8	4	0.040	+2.1
0.300	3	2	0.049	7	4	0.050	+0.1	8	4	0.124	+7.5
0.250	3	2	0.074	7	4	0.092	+1.8	8	4	0.187	+11.3
0.200	3	2	0.109	7	4	0.154	+4.5	8	4	0.258	+14.9
0.125	4	3	0.193	7	4	0.286	+9.3	8	4	0.353	+16.0

Продовження таблиці 3.5.

0.100	4	3	0.242	7	4	0.336	+9.4	16	11	0.399	+15.7
0.050	5	4	0.389	15	11	0.470	+8.1	16	11	0.587	+19.8
0.030	6	5	0.491	15	11	0.587	+9.6	32	26	0.682	+19.1
0.010	10	9	0.675	31	26	0.753	+7.8	64	57	0.833	+15.8
0.005	14	13	0.760	63	57	0.823	+6.2	128	120	0.882	+12.7
0.003	18	17	0.810	63	57	0.865	+5.5	128	120	0.919	+10.0
0.001	30	29	0.887	127	120	0.925	+3.8	128	120	0.931	+4.4

У таблиці 3.6 наведені результати порівняльного аналізу отриманих експериментально характеристик КПД для кодів з виявленням і виправленням помилок і кодів тільки з виявленням помилок за максимальним значенням $R_{эф}$, а також вигреш у порівнянні з кодом (32,31), що наближається по своїх властивостях до алгоритму виявлення помилок у каналній процедурі X.25/2 при довжині пакета 1024 біт і поліномі 32 ступеня.

Таблиця 3.6. Результати порівняльного аналізу отриманих експериментально характеристик КПД для кодів з виявленням і виправленням помилок і кодів тільки з виявленням помилок за максимальним значенням $R_{эф}$

№№ п/п	Якість каналу R_q	Код з виправленням помилки		Код з виявленням помилки		Вигреш, раз	Код (32, 31), раз	Оптимальний код і вигреш у порівнянні з кодом (32, 31), раз	
		Параметри коду	$R_{эф}$	Параметри коду	$R_{эф}$			$R_{эф}$	Оптимальний код
1	1/2	(8,2)	0,0793	-	-	абсол	-	(8,2)	абсол
2	1/4	(16,7)	0,3355	-	-	абсол	-	(16,7)	абсол
3	1/8	(16,11)	0,407	(4,3)	0,2407	1,69	-	(16,11)	абсол
4	1/16	(16,11)	0,5534	(8,7)	0,386	1,43	0,1468	(16,11)	3,93
5	1/32	(32,26)	0,6241	(8,7)	0,53	1,18	0,2457	(32,26)	2,24
6	1/64	(32,26)	0,7336	(8,7)	0,6075	1,2	0,4407	(32,26)	1,66
7	1/128	(32,26)	0,786	(8,7)	0,6835	1,15	0,6091	(32,26)	1,28
8	1/256	(32,26)	0,789	(16,15)	0,786	1,01	0,7277	(32,26)	1,08

3.8 Оцінка ефективності використання коду Файра

Оцінимо ефективність використання коду Файра для виправлення однократних помилок у системі передачі повідомлень. Умовно приймемо, що ймовірність спотворення одиничного елемента $P_e = 5 \cdot 10^{-3}$

При біномному характері розподілу помилок маємо:

$$P_{nn} = 1 - (1 - P_e)^n = 1 - (1 - 5 \cdot 10^{-3})^7 \approx 3,5 \cdot 10^{-2} \quad (3.57)$$

Інформаційна надмірність:

1) при виправленні однократної помилки:

$$g_{in1} = (1+r/k) = 1 + 3/4 = 1,75 \quad (3.58)$$

2) у разі виявлення помилок

$$g_{in2} = (1+r/k) \left(\frac{1}{1 - P'_{nn}} \right) = 1,75 \cdot \frac{1}{1 - 3,5 \cdot 10^{-2}} = 1,81 \quad (3.59)$$

Умовно візьмемо схемну надмірність для систем передачі повідомлень при виправленні помилок $q_{cx2} = 2,5$, тоді отримаємо:

$$g_1 = g_{in1} + g_{cx1} = 1,75 + 1,5 = 3,25 \quad (3.60)$$

$$g_2 = g_{in2} + g_{cx2} = 1,81 + 2,5 = 4,31 \quad (3.61)$$

Ймовірність появи помилки в повідомленні без надмірності:

$$P_{nk} = 1 - (1 - P_e)^k = 1 - (1 - 5 \cdot 10^{-3})^4 \approx 0,02 \quad (3.62)$$

а в повідомленні з використанням коду Файра, що виправляє одну помилку:

$$P_{nn}(1) = 1 - (1 - P_e)^n - n P_e (1 - P_e)^{n-1} = 1 - (1 - 5 \cdot 10^{-3})^7 - 7 \cdot 5 \cdot 10^{-3} (1 - 5 \cdot 10^{-3})^6 \approx 0,002$$

Ймовірність виникнення помилки при застосуванні цього самого коду для виявлення двократних помилок

$$P_{nn}(2) = 7 P_e^3 (1 - P_e)^4 + P_e^4 (1 - P_e)^3 = 7(5 \cdot 10^{-3})^3 (1 - 5 \cdot 10^{-3})^4 + 7(5 \cdot 10^{-3})^4 (1 - 5 \cdot 10^{-3})^3 + (5 \cdot 10^{-3})^7 \approx 0,86 \cdot 10^{-6}$$

Виграш у захисті від помилок

$$a_1 = \frac{P_{nk}}{P_{nn}(1)} = \frac{0,02}{0,002} = 10 \quad (3.63)$$

$$a_2 = \frac{P_{nk}}{P_{nn}(2)} = \frac{0,02}{0,86 \cdot 10^{-6}} \approx 23000 \quad (3.64)$$

Критерії ефективності

$$k_{1ef} = \log_2 \frac{a_1}{g_1} = \log_2 \frac{10}{3,25} \approx 1,6$$

$$k_{2ef} = \log_2 \frac{a_2}{g_2} = \log_2 \frac{23000}{4,31} \approx 12,3$$

Тобто ефективність використання коду Файра для виявлення та виправлення помилок відносно розрахунків становить 12,3.

Графік ефективності виявлення помилок

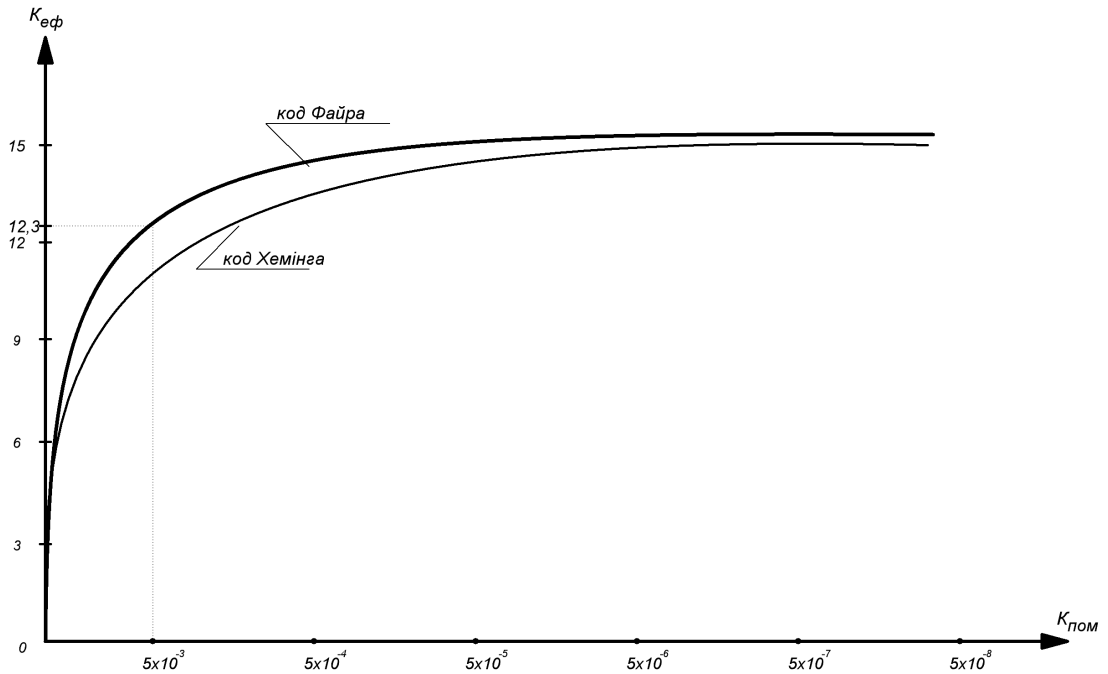


Рис. 3.4. Графік ефективності виявлення помилок

Графік ефективності виправлення помилок

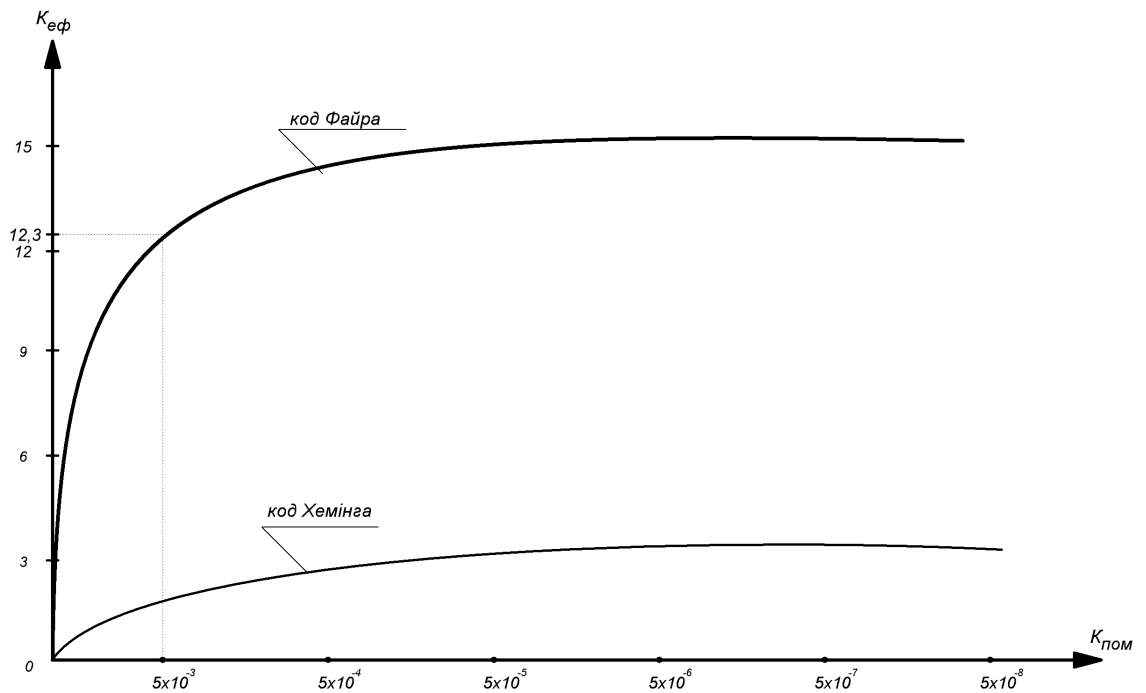


Рис. 3.5. Графік ефективності виправлення помилок

3.9. Розроблення блок-схем кодування та декодування кодом Файра

Так як код Файра є циклічним кодом, то кодування ним проходить так само, як і любым іншим кодом із цього класу (рис.3.6). При кодуванні вихідна комбінація з дописаними в молодші розряди r нулями ділиться на твірний поліном, а отриманий залишок дописується до вихідної комбінації. Отримана кодова комбінація передається в канал зв'язку.

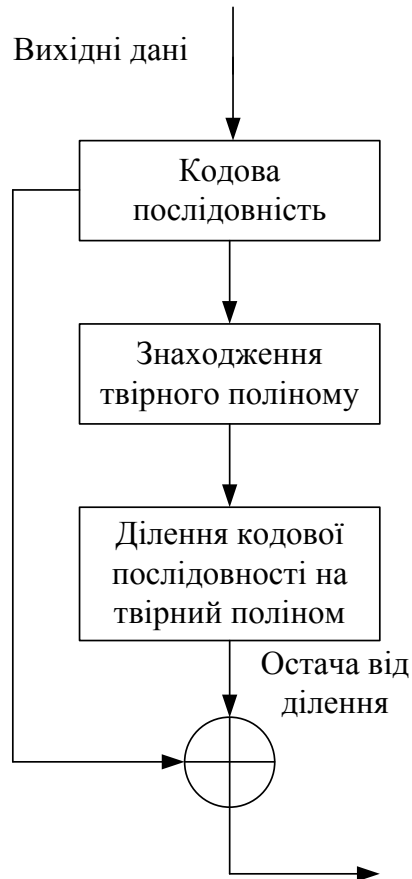


Рис. 3.6. Блок-схема кодування кодом Файра

При декодуванні виробляється роздільне ділення прийнятої кодової комбінації на поліном $P(x)$ ступеня l і на поліном $(x^c + 1)$.

У результаті такого ділення в зсувних логічних регістрах, відповідним багаточленам $P(x)$ та $(x^c + 1)$, виходять залишки $R_1(x)$ і $R_2(x)$, які будуть нульовими, якщо помилок не було. Якщо ж пройшла одиночна пачка помилок довжиною $b \leq l$, то залишки будуть відмінні від нуля й не рівні між собою.

Для виправлення даної пачки помилок продовжують ділення, здвигаючи кодові комбінації в логічних регістрах доти, поки не відбудеться збіг залишків у логічних регістрах. При цьому вхідна кодова комбінація здвигається й у буферному регістрі.

Збіг залишків від ділення в обох логічних регістрах свідчить про виявлення пакета помилок, а кількість додаткових зсувів кодових комбінацій без числа перевірочних розрядів вказує на місце, що займає помилковий пакет у декодуемій кодовій комбінації. Кодова комбінація остачі від ділення на незвідний полином є коригувальною комбінацією, що додається до спотвореної прийнятої комбінації в момент відкриття ключа, тим самим виправляючи виникаючі помилки (рис. 3.7).

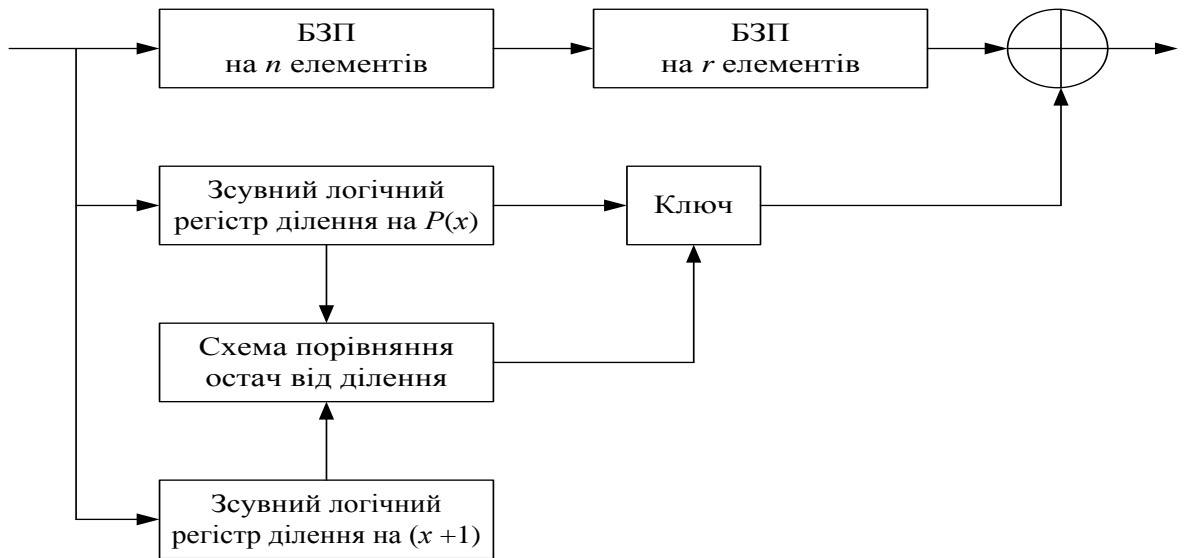


Рис. 3.7 – Блок-схема декодування кодом Файра

3.10. Програмна реалізація кодера Файра

Реалізації вище розробленої блок-схеми кодування кодом Файра буде здійснена на мові програмування Turbo Pascal. Це найбільш розповсюджена на даний час мова програмування.

Програма складається із п'яти модулів:

- основна програма;
- модуль вводу;
- модуль виводу;
- інтерфейс;
- модуль обробки помилок.

Модульна структура програми представлена на рисунку 3.8

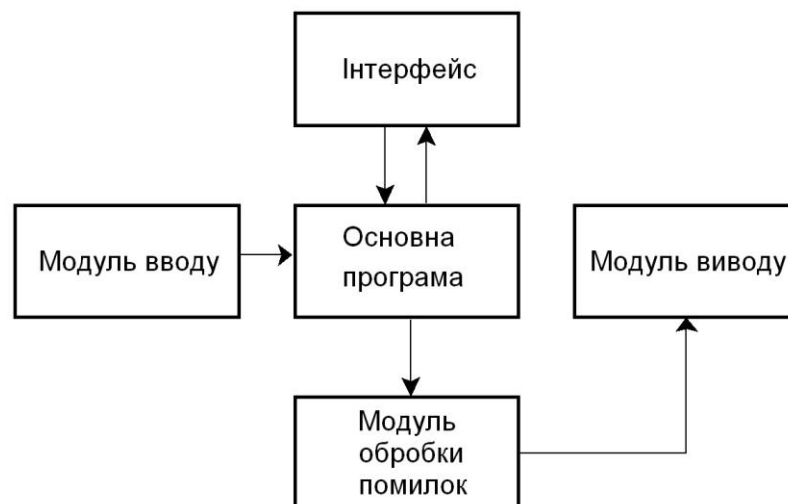


Рис. 3.8 – Модульна структура програми

Без тестування програм неможливо створити правильно працюючу програму, тому що завжди існують якісь граничні умови, за межами яких програма дає помилки. При цьому не слід плутати налагодження програм і тестування.

Налагодження програм – використовується при неправильній роботі програми.

Тестування – шукає помилки.

Існує багато методів тестування:

– статичне тестування є найбільш формалізованим і автоматизованим методом перевірки програм. Як еталони застосовуються правила структурної побудови програмних модулів і обробки даних. Перевірка ступеня цих правил проводиться без використання об'єктного коду програми шляхом формального аналізу тексту програми мовою програмування. Оператори й операнди тексту програм при цьому аналізуються в символному виді, тому такий метод називають символним тестуванням.

– детерміноване тестування є найбільш трудомістким і деталізованим. При детермінованому тестуванні контролюється кожна комбінація вихідних еталонних даних і відповідна їй комбінація результатів функціонування програми. Це дозволяє виявляти відхилення результатів від еталона й реалізуючих даних, при яких це відхилення відбулося.

– стохастичне тестування застосовується в тих випадках, коли неможливо в складних завданнях перебрати всі комбінації вихідних даних і проконтролювати результати функціонування програми на кожному з них. При цьому вихідні тестові дані задаються безліччю випадкових величин з відповідними розподілами й для порівняння отриманих результатів використовуються також розподіли випадкових величин. Стохастичне тестування застосовується в основному для виявлення помилок.

– тестування в реальному масштабі часу. У процесі такого тестування перевіряється виконання програм і обробка вихідних даних з урахуванням часу їхнього надходження, тривалості й пріоритетності обробки, динаміка використання пам'яті й т.д.

– при висхідному тестуванні насамперед перевіряються модулі нижніх ієрархічних рівнів, до яких поступово підключаються викликаючи їх модулі. При цьому забезпечується працездатність визиваємих компонентів і функції групи програм перевіряються в їхньому природному виконанні. Основні труднощі складаються в необхідності повного відновлення тестових наборів при підключенні кожної нової програми більш високого рівня.

– при спадному тестуванні перевірки починаються із програм керування й організації обчислювального процесу. Спочатку тестується керуюче ядро комплексу програм і програми рішення функціональних завдань, розміщених на вищих ієрархічних рівнях. До них поступово підключаються для тестування програми наступних більш низьких ієрархічних рівнів. Перевагою такого методу є можливість збереження й розвитку наборів тестових даних у міру підключення програм нижніх рівнів.

На практиці звичайно використовують три стратегії тестування:

– тестування програм як "Чорного ящика". Має на меті з'ясування обставин, у яких поведінка програм не відповідає специфікації. Тестові дані складаються й використовуються без обліку знань про внутрішню структуру програми.

– тестування програм як "Білого ящика". Тестові набори даних проектуються на основі внутрішньої логіки програми. Ціль тестування - перевірити кожну галузь, кожний шлях і кожний оператор. Специфікація програми при цьому не використовується.

– реальна стратегія звичайно сполучає обидва методи. За допомогою методу "чорного ящика" робиться:

- перевірка в нормальних умовах;
- аналіз граничних значень;
- перевірка у виняткових ситуаціях;
- припущення про помилку;

За допомогою методу "Білого ящика" робиться:

- покриття операторів;
- покриття рішень;
- покриття умов.

При тестуванні програми був використаний вихідний метод тестування. За допомогою цього методу спочатку були протестовані окремі модулі програми, а потім і вся програма.

Тестування системи містило в собі:

- тестування введення різних набору даних;
- одержання комбінацій для коду з будь-яким утворюючим багаточленом.

Тестування уведення різних наборів кодових комбінацій не показало ні однієї помилкової ситуації.

РОЗДІЛ 4

СТИСНЕННЯ ПЕРВИННИХ ПОВІДОМЛЕНЬ В ІНФОРМАЦІЙНИХ КАНАЛАХ СИСТЕМ УПРАВЛІННЯ

Список скорочень

SU TM – система управління телекомунікаційними мережами;
LZ – алгоритм Лемпеля – Зіва;
LZW - алгоритм Лемпеля – Зіва – Уелча;
LZH - алгоритм Лемпеля – Зіва – Хаффмена

4.1. Характеристики джерел повідомлень в системах управління телекомунікаційними мережами

Проблема скорочення обсягів переданих даних виникає в системах управління інфокомунікаційними мережами при передачі великих масивів вимірювальної інформації та інформації контролю за станом об'єктів. При реєстрації процесів, що протікають із високою швидкістю, або при необхідності передачі значних обсягів даних у системах із великою кількістю абонентів (промислових мережах) апаратура передачі даних іноді не в змозі забезпечити інформаційний обмін у необхідному темпі, при цьому інформація, що надходить від датчиків, як правило має високу надмірність.

Стиснення даних – процедура зменшення надмірності повідомлення з метою зменшення його обсягу. У процесі стиснення усувається природна надмірність, властива практично будь-якому джерелу інформації [22, 25, 32]. При розгортанні, перед видачею інформації споживачу, відбувається відновлення первинного повідомлення зі стиснутого. Усунення надмірності досягається за рахунок застосування ефективного кодування.

При розв'язанні питання про використання того або іншого алгоритму стиснення даних у конкретній системі передачі даних необхідно оцінити сильні і слабкі сторони алгоритму з урахуванням умов, у яких він буде працювати. Крім коефіцієнта стиснення важливе значення мають такі параметри, як швидкодія, необхідний обсяг пам'яті для роботи, здатність алгоритму адаптуватися до статистичних властивостей даних, що надходять, а при передачі даних у режимі реального часу – час затримки переданих даних. Тому алгоритм, що показує кращі результати в одних умовах, може працювати значно гірше або взагалі не підходити для іншої системи. В даному розділі проведено аналіз існуючих методів стиснення з метою визначення найбільш перспективних для застосування в СУ та розроблено алгоритми та методику експериментального дослідження обраних методів.

Кінцеві комбінаторні джерела

Комбінаторне джерело здатне породити тільки деякі повідомлення. Скажемо, допустимо вважати, що речення української мови вироблені деяким комбінаторним джерелом.

Кінцевим комбінаторним джерелом називається довільна кінцева множина S . В якості таких джерел, як правило, розглядаються множини слів деякого алфавіту A (словники).

Основною характеристикою джерела є його ентропія. Вона формалізує інтуїтивне поняття „кількість інформації, що переноситься одним елементом S ”. Ця кількість повинна, по-перше, залежати тільки від потужності $|S|$ множини S та, по-друге, зростати зі збільшенням $|S|$. Далі кількість інформації, що переноситься n елементами S , повинна бути в n раз більше кількості, що переноситься одним елементом, $n > 1$. Якщо ентропію джерела S позначити через $H(S)$, то можемо записати, що

$$H(S) = \varphi(|S|) \tag{4.1}$$

де $\varphi(|S|)$ – деяка монотонна функція, що задовольняє умові $\varphi(n|S|)=n\varphi(|S|)$.

Єдиною монотонною функцією, що задовольняє умові $\varphi(n \cdot k)=n\varphi(k)$ для будь-яких цілих n та k , є функція $\varphi(k)=C \log k$, де C – довільна позитивна константа.

Вважаючи $C=1$, приходять до двійникової ентропії. Одиниця вимірювання двійникової ентропії називається бітом. Таким чином, ентропія $H(S)$ кінцевої множини (джерела) S визначена як

$$H(S)=\log |S| \quad (4.2)$$

Часто в якості S розглядається підмножина множини E^n двійкових слів деякої довжини n . В цьому разі ентропією джерела на букву називається число

$$h(S)=\frac{1}{n} H(S) \quad (4.3)$$

Джерело, в якого ентропія на букву відмінна від 0 та 1, називається нетривіальним, так як воно виробляє деяку непусту та відмінну від E^n множини слів. Послідовність джерел називається асимптотичною нетривіальною, якщо її гранична ентропія на букву не рівна ні 0, ні 1. Це визначення, так само як і поняття комбінаторного джерела, дав Хартлі [1].

Особливої уваги заслуговують комбінаторні джерела з кінцевим числом станів (автоматні джерела). Таке джерело має деяке число k , $k \geq 1$, станів. При переході з одного стану в інший воно породжує деяку букву алфавіту A . Джерелу відповідає матриця S , елемент s_{ij} якої дорівнює 1, якщо з i -го стану джерело може перейти в j -е та 0 – в протилежному випадку, $1 \leq i, j \leq k$. Один зі станів називається початковим, деякі зі станів – кінцевими. Джерело породжує слово, якщо воно породжує послідовно всі його букви, проходячи від початкового стану до одного з кінцевих.

Імовірнісні джерела

Кінцевим імовірнісним джерелом називається довільна кінцева множина S з визначеним на ній розподілом імовірностей. Ця множина називається алфавітом, а її елементи – буквами.

При визначенні ентропії імовірнісного джерела зручно скористатися поняттям розбиття. Розбиттям A множини S (не обов'язково кінцевої) називається сукупність множин A_1, \dots, A_k , $k \geq 1$, що попарно не перетинаються і при об'єднанні дають S . Число k елементів розбиття A позначають $|A|$. Множини A_i називаються атомами розбиття A . На множині розбиттів множини S визначений частковий порядок: $B \geq A$, якщо кожний атом A є об'єднанням деяких атомів розбиття B . В цьому випадку кажуть, що B інформативніше A . Для будь-яких двох розбиттів A та B знайдеться розбиття, більш інформативніше, ніж кожне з них. Таким буде, наприклад, розбиття, атомами якого є всі можливі попарні перетинання атомів A та B . Воно позначається $A \vee B$, $A \vee B = \{A_i B_j\}$, де $A = \{A_1, \dots, A_{|A|}\}$, $B = \{B_1, \dots, B_{|B|}\}$ [3].

Найменш інформативним є розбиття, єдиним атомом якого служить сама множина S . Найбільш інформативне розбиття, атомами якого служать одноелементні множини.

Ентропією розбиття A простору S називається число

$$-\sum p(A_i) \log p(A_i) \quad (4.4)$$

де сума береться по всім атомам A_i розбиття A , $p(A_i)$ дорівнює імовірності A_i . Ентропією $H(S)$ кінцевого імовірнісного джерела називається ентропія його найбільш інформативного розбиття:

$$H(S) = -\sum p(A_i) \log p(A_i) \quad (4.5)$$

де A_i – елементи множини S ; $S = \{A_1, \dots, A_k\}$. Ці визначення належать Шеннону.

Стаціонарні джерела

Нехай A – деякий кінцевий алфавіт. Розглянемо множину A^∞ нескінчених послідовностей літер алфавіту A . На множині A^∞ визначена стаціонарна, тобто незалежна від зсуву в часі, імовірнісна міра. Множина A^∞ з визначеною на ній стаціонарною мірою називається стаціонарним джерелом. Тобто, до стаціонарних відносять ті джерела, у яких на множині X нескінчених послідовностей символів алфавіту x , визначена стаціонарна (незалежна від зсуву в часі) імовірнісна міра [1].

Якщо букви алфавіту $A = \{A_1, \dots, A_{|A|}\}$ виникають незалежно один від одного з деякими

імовірностями $p(A_1), \dots, p(A_{|A|})$, $\sum_{i=1}^{|A|} p(A_i) = 1$, то відповідне джерело називається джерелом

Бернуллі. Якщо імовірність появи чергової букви залежить лише від попередньої букви, то таке джерело називається марковським першого порядку. Воно визначається матрицею $P = (p_{ij})$, $1 \leq i, j \leq |A|$, де p_{ij} – імовірність появи букви A_j слідом за буквою A_i . По матриці P знаходиться стаціонарний вектор $p = (p_1, \dots, p_{|A|})$: $Pp = p$. Імовірність слова $A_{i_1} \dots A_{i_n}$ рівна $p_{i_1} p_{i_2 i_1} \dots p_{i_n i_{n-1}}$.

Аналогічно визначаються марковські джерела порядку t , коли імовірність появи чергової букви залежить від t попередніх, $t > 0$.

Іноколи буває доцільно розглядати декілька більш загальних джерел марковського типу, які, користуючись термінологією з теорії автоматів, назвемо джерелами Мура, та Мілі. Джерело Мура визначається кінцевим автоматом, що породжує в кожному зі станів одну з букв алфавіту A та переходить в черговий стан з деякою імовірністю. Робота джерела починається в одному зі станів в відповідності з деяким стаціонарним розподілом імовірностей. Джерело Мілі також визначається кінцевим автоматом. Однак букви породжуються ним в момент переходу з одного стану в інший.

Аналогічно стаціонарному імовірнісному джерелу визначається стаціонарне комбінаторне джерело. Розглянемо знову множину A^∞ нескінчених послідовностей літер алфавіту A . Підмножина $S \subseteq A^\infty$ називається стаціонарним комбінаторним джерелом, якщо для кожної послідовності $x \in S$ зсув x вліво знову належить S . Прикладом такого джерела може бути множина всіх можливих нескінчених послідовностей, що формуються кінцевим автоматом, при умові, що початковим може бути будь-яке з його станів.

Для бернуллівського джерела маємо

$$H(S) = - \sum_{i=1}^{|A|} p(A_i) \log p(A_i) \quad (4.6)$$

Для марковського джерела першого порядку обчислення ентропії дає нам

$$H(S) = - \sum_{i,j=1}^{|A|} p_{ji} \log p_{ji} = \sum_{j=1}^{|A|} p_j H(p_{j1}, \dots, p_{jk}) \quad (4.7)$$

Величина $H(p_{j1}, \dots, p_{jk})$ є умовною ентропією при умові, що з'явилась j -а буква. Ентропія джерела є зваженою сумою цих умовних ентропій з вагами, рівними стаціонарним імовірностям.

Марковські джерела є зручною математичною моделлю реальних джерел інформації та досить часто використовуються на практиці. У свій час засновник теорії марковських ланцюгів А.А. Марков, вивчаючи текст „Евгенія Онегіна”, замітив, що імовірність появи

голосної чи приголосної букви залежить від попередньої букви. Можна сказати, що він використав марковське джерело першого порядку з двома станами в якості моделі тексту пушкінського роману. Ідея А.А. Маркова знайшла продовження в основоположній роботі К.Шеннона, котрий приводить приклади англійських текстів, що породжені марковськими джерелами різних порядків. При цьому станами джерела можуть слугувати або букви, або слова.

Дискретні джерела

Розглянемо модель джерела: ансамбль A повідомлень, урахувавши розподіл імовірностей p_i , та розподіл проміжків часу τ_i , протягом яких джерело вибирає різні повідомлення a_i .

Продуктивність джерела щодо певного повідомлення a_i можна визначити як

$$V_{джі} = I_i / \tau_i \quad (4.8)$$

Її одиниця залежить від вибору одиниці кількості інформації I_i . Наприклад, це може бути біт за секунду.

Як правило, джерело вибирає досить велику кількість повідомлень протягом певного часу. Тому природно як загальну характеристику джерела прийняти середню за ансамблем його продуктивність, користуючись відомим методом статистичного усереднення:

$$V_{дж} = \sum_{i=1}^k p_i V_{джі} \quad (4.9)$$

Загалом, коли $\tau_i \neq \tau_j$ при $i \neq j$ вираз (1.8) після усереднення за часом можна перетворити до такого вигляду:

$$V_{джі} = I_i / \left(\sum_{i=1}^k p_i \tau_i \right) = I_i / \tau_{сер} \quad (4.10)$$

де $\tau_{сер}$ — середній час вибору джерелом одного повідомлення.

Вираз (4.10) дійсний також для більш поширеного випадку, коли джерело вибирає всі свої повідомлення за один і той самий проміжок часу $\tau = \tau_{сер}$.

З урахуванням (4.10) вираз (4.9) набуває вигляду:

$$V_{дж} = \frac{1}{\tau_{сер}} \sum_{i=1}^k p_i I_i = - \frac{1}{\tau_{сер}} \sum_{i=1}^k p_i \log p_i = \frac{H(A)}{\tau_{сер}} \quad (4.11)$$

Бачимо, що така загальна характеристика, як продуктивність дискретного джерела, визначається його середніми показниками: ентропією (середньою кількістю інформації в одному повідомленні) та часом утворення останнього [2].

Неперервні джерела

Стосовно неперервного джерела на відміну від дискретного можна говорити про нескінченний алфавіт повідомлень, кожне з яких відрізняється від сусідніх на нескінченно малу величину, та нескінченний ансамбль повідомлень.

Середня швидкість передачі інформації неперервним джерелом кількість інформації, що передається в одиницю часу, тобто

$$R_T(x, y) = \frac{I_T(x, y)}{T} = \frac{m}{T} [h(x) - h(x/y)] = F_o [h(x) - h(x/y)] \quad (4.12)$$

де $F_d = m/T$ – частота дискретизації повідомлення; $I_T(x, y)$ – кількість інформації в неперервному повідомленні тривалістю T ; $m = T/\Delta t$ – кількість відліків повідомлення тривалістю T , зроблених з інтервалом часу Δt ; $h(x)$ – диференціальна ентропія одного відліку повідомлення; $h(x, y)$ – умовна диференціальна ентропія одного відліку повідомлення [2].

4.2. Оцінки ефективності методів стиснення

Коефіцієнт стиснення даних. Для оцінки ефективності стиснення даних використовується декілька показників ступеня стиснення [22, 61]. Найбільш поширеним є коефіцієнт стиснення K_{cm} , який характеризує у скільки разів обсяг даних на виході компресора V_{cm} менший за обсяг даних на вході V_n :

$$K_{cm} = V_n / V_{cm} \quad (4.13)$$

Іноді застосовується зворотна величина, яку називають коефіцієнтом компресії K_k :

$$K_k = V_{cm} / V_n \quad (4.14)$$

Іноді для визначення ступеня стиснення використовується коефіцієнт стиснення даних K_{cd} :

$$K_{cd} = (1 - V_{cm} / V_n) \cdot 100\%, \quad (4.15)$$

який дорівнює обсягу даних, що були виключені з повідомлення у процесі його стиснення. При відсутності ефекту стиснення $K_{cd} = 0$, а у випадку максимального стиснення коефіцієнт K_{cd} наближається до 100%.

Швидкість роботи компресора. Для правильного вибору методу стиснення повідомлень при передачі даних недостатньо одного коефіцієнта стиснення. Необхідно враховувати також максимальну швидкість передачі, що забезпечується кожним методом. Максимальна швидкість стиснення визначається швидкістю роботи компресора, тобто кількістю символів N_c , які обробляються компресором за одиницю часу T :

$$R_{max} = N_c / T. \quad (4.16)$$

Щоб не відбувалося перевантаження компресора, продуктивність джерела повинна бути менше максимальної швидкості стиснення, тобто повинна виконуватися умова

$$H(A) < R_{max}. \quad (4.17)$$

Ця ж умова також повинна виконуватися для алгоритму декомпресії повідомлень. Однак у реальних алгоритмах стиснення даних швидкість декомпресора завжди перевищує швидкість компресора, тому головна увага при аналізі швидкості алгоритмів стиснення приділяється саме процедурі компресії.

Час затримки повідомлення в апаратурі стиснення. Цей параметр має велике значення при стисненні даних у реальному часі. Дозволений час затримки визначається ймовірністю аварійної ситуації, оскільки в цьому випадку вся накопичена в оперативній пам'яті інформація може бути втрачена.

Затримка повідомлення τ_z , у загальному випадку складається з часу накопичення даних в оперативній пам'яті кодера джерела (T) і часу обробки в кодері джерела ($\tau_{обр}$):

$$\tau_z = T + \tau_{обр}. \quad (4.18)$$

У деяких випадках процедура стиснення може здійснюватися паралельно з накопиченням даних в оперативній пам'яті. У цьому випадку час T в затримку не входить, і $\tau_3 = \tau_{обр}$.

Складність алгоритму стиснення. Складності залежить від складності алгоритму обробки, і, звичайно, чим вище необхідний коефіцієнт стиснення, тим складніший алгоритм обробки.

В літературі [40, 70, 74, 81, 83, 97] наводяться оцінки часової складності деяких алгоритмів стиснення. Так, для способу Віттера, який використовує хеширування, час роботи складає $O(1)$. Для матричного способу стиснення ця оцінка буде дорівнювати $O(n)$, а для способу стиснення Лемпеля-Зіва – $O(n^2)$. Стиснення із використанням алгоритму Хаффмена займає час $O(n \log(n))$.

Оскільки час роботи способів стиснення суттєво залежить від типу інформації, що підлягає стисненню, доцільно провести дослідження залежності часу роботи алгоритмів від розміру блоку, для різних способів стиснення та різних типів інформації, яка передається інформаційними каналами.

4.3. Класифікація методів стиснення

Стиснення інформації застосовується для прискорення та зниження витрат на її оброблення, зберігання й пошук, а також для зменшення ємності пам'яті, зайнятої в ЕОМ.

Під стисненням інформації розумітимемо операцію, внаслідок якої певному коду чи повідомленню ставиться у відповідність код або повідомлення меншої довжини.

Методи стиснення інформації поділяють за призначенням, характером і ступенем стиснення, швидкістю та ступенем відновлення початкового стану інформації (втратами).

За призначенням розрізняють дві великі групи способів стиснення: для передачі даних і для їх архівації. Різниця між ними полягає в тому, що перші оперують з незначними інформаційними масивами (до кількох десятків, сотень байтів), а другі зі значно більшим обсягом інформації (мегабайти).

За характером стиснення інформації розрізняють лінійні, матричні, комбіновані та каскадні способи.

До лінійних належать способи, за якими стиснення елементів інформаційного масиву виконується в одному з напрямків (горизонтальному або вертикальному). Залежно від цього лінійними способами можуть виконуватися поздовжнє (горизонтальне) та поперечне (вертикальне) стиснення інформації.

До матричних належать способи стиснення інформації, за якими елементи інформаційного масиву стискаються з використанням матричного принципу заміни повторюваних елементів.

Комбіновані способи поєднують одночасне використання для стиснення інформаційного масиву двох чи більше лінійних або/та матричних способів.

До каскадних належать способи стиснення інформації, за якими стиснення виконується послідовно різними способами.

За ступенем стиснення інформації розрізняють низькоефективні (з коефіцієнтом стиснення до 1,5), середноефективні (1,51...3) та високоефективні (понад 3) способи; за швидкістю стиснення/розпаковування — низько-, середньо- та високошвидкісні, при яких швидкість стиснення/розпаковування змінюється від кількох кілобайтів за секунду (низькошвидкісні) до кількох мегабайтів за секунду (високошвидкісні).

За ступенем відновлення початкового стану інформації (втратами) способи стиснення поділяють на без відновлення початкового стану інформації, з частковою її втратою та без втрати інформації (з повним її відновленням). Що стосується останнього поділу способів стиснення інформації, то до першої групи належать найпримітивніші, а до другої та третьої груп — складніші й ефективніші способи.

Так, до відомих способів стиснення інформації без відновлення її початкового стану можна віднести стиснення за допомогою поділу кодової комбінації на кілька частин і з порозрядним зсувом. Ці способи застосовуються дуже рідко, оскільки не гарантують повного відновлення стисненої інформації з точки зору неоднозначності утвореної при стисненні послідовності символів.

У той же час способи з частковою втратою інформації мають специфічне застосування, коли часткова її втрата майже не позначається на якості відновлюваної інформації [2].

Приведена класифікація зображена на рис. 4.1.

4.4. Способи стиснення даних при передачі

4.4.1. Лінійні способи

Стиснення даних із використанням замість повторень додаткових символів. Ці способи ґрунтуються на заміні повторюваних елементів деякими умовними символами. Вони є ефективними в тому разі, коли масиви інформації, які подаються у вигляді рядків або стовпців, розташованих у зростаючому порядку, мають однакові значення елементів в одних і тих самих розрядах, що характерно для техніко-економічної інформації. Таке стиснення даних дає змогу скоротити масив у кілька разів.

Так, якщо елементи повторюються на початку рядків (стовпців) відносно попередніх, то замість виключених розрядів у масив вводиться знак поділу p , який дає можливість відокремити елементи в згорнутому масиві. При розгортанні замість знака p поновлюють всі пропущені розряди, які були до елемента, що знаходився безпосередньо за p в стисненому тексті. Запис знаків, які знаходяться після p , виконується з кінця рядка (стовпця).

Як приклад розглянемо масив, у якому інформацію записано у вигляді рядків, що складаються з восьми десяткових знаків:

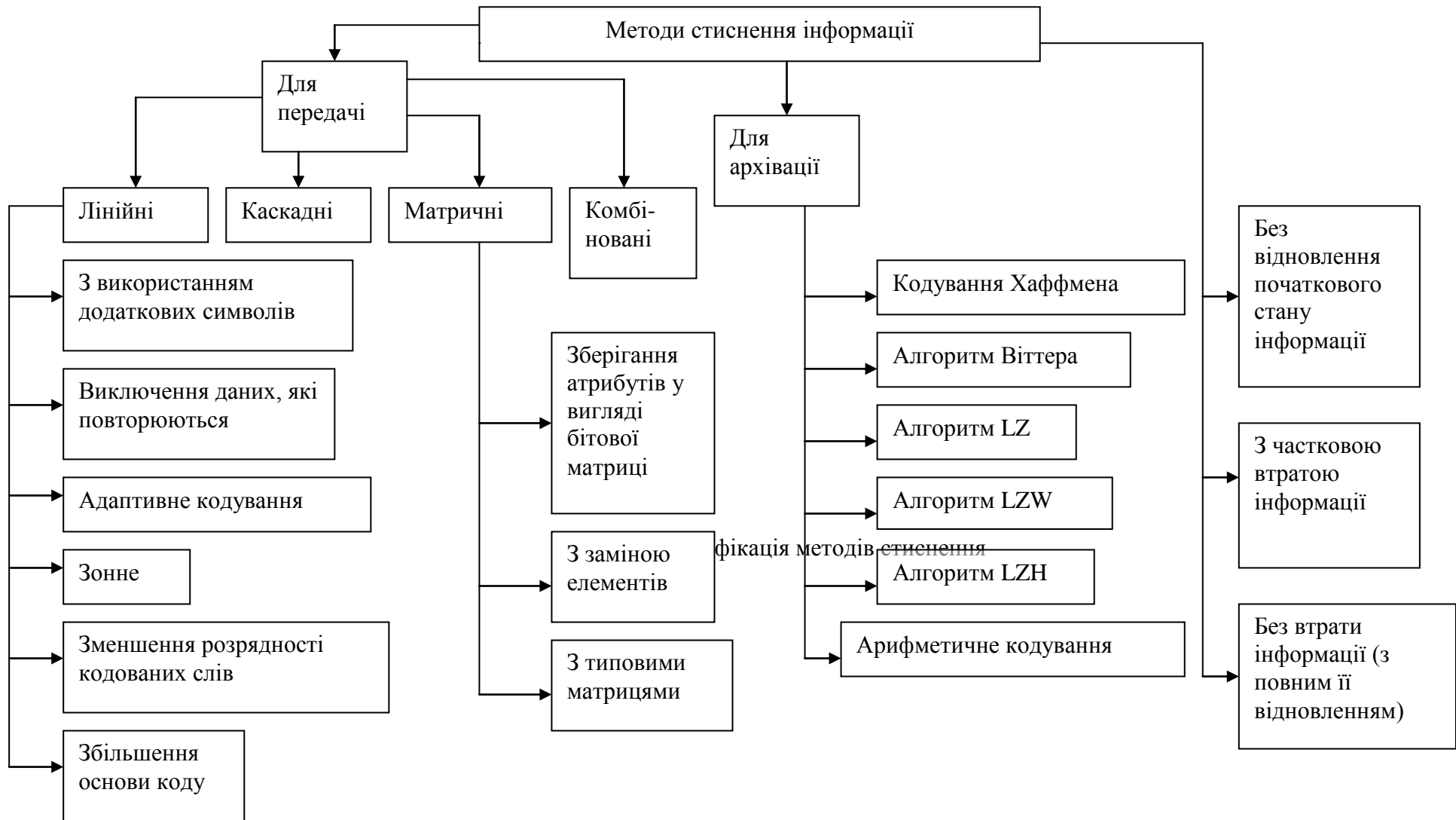
3 9 1 4 5 6 8 0	5 6 7 1 8 3 2 9
3 9 1 4 5 6 8 6	5 6 7 1 8 3 4 3
3 9 1 6 7 5 9 6	5 6 7 2 9 4 6 2
3 9 1 4 5 7 2 1	5 6 7 1 8 3 4 8
3 9 1 4 5 6 3 8	5 6 7 1 7 6 3 1

Згорнутий масив інформації матиме такий вигляд:

3 9 1 4 5 6 8 0	2 9 p 4 3 p 2 9
p 6 p 6 7 5 9 6	7 6 2 p 1 8 3 4
p 4 5 7 2 1 p 6	8 p 7 6 3 1
3 8 5 6 7 1 8 3	

Розгортку виконуємо з кінця масиву. При цьому на наступний рядок переходимо або після заповнення рядка, або при знаку поділу p :

3 9 1 4 5 6 8 0	5 6 7 1 8 3 2 9
..... 6 4 3
... 6 7 5 9 6	... 2 9 7 6 2
... 4 5 7 2 1	... 1 8 3 4 8
..... 6 3 8 7 6 3 1



Після заповнення пропущених цифр за аналогічними розрядами попереднього рядка дістаємо масив інформації, який був до стиснення.

Недоліком цього методу стиснення інформації є неможливість його застосування до впорядкованих масивів, у яких повторювані розряди зустрічаються не на початку рядків (стовпців).

Для згортання масивів, у яких в одному рядку (стовпці) є тільки одна повторювана ділянка, можна використати вищезгаданий спосіб з уведенням додаткового символу К (кінець рядка, стовпця). При цьому розгортання масиву ведеться від К до К. В разі фіксованої довжини рядка (стовпця) всі розряди, які знаходяться між К, разом із пропущеними розрядами мають утворювати повний рядок (стовпець). Так, початковий і згорнений по рядках масиви матимуть вигляд:

47356891	4 7 5 3 6 8 9 1
47536432	К р 4 3 2 К 3 5
35536690	р 6 9 0 К 6 9 р
69536241	2 4 1 К 7 р 1 0
79536210	К 8 р 8 0 2 К 1
89536802	6 р К 6 3 р К 4
16536802	6 9 р 5 К р 3 4
63536802	8 К
46936805	
46936348	

Поновлення масиву може виконуватися з початку або з кінця. За наявності в рядку (стовпці) кількох повторюваних ділянок замість р вводять спеціальні символи, що вказують необхідну кількість пропусків. При цьому необхідність у символі К, який визначає кінець рядка (стовпця), відпадає. Так, якщо для позначення заданої кількості пропусків у рядку ввести відповідно символи X - 2, Y - 3, Z - 5, то початковий і згорнений масиви десяткових знаків матимуть такий вигляд:

2 3 4 5 6 7 8 5	2 3 4 5 6 7 8 5
6 3 4 5 6 7 9 8	6 Z 9 8 X 7 X X
6 3 7 5 6 7 9 2	2 2 4 Y Y 8 9 3
2 4 7 5 6 7 9 2	X 2 3 1 1 Z X Z
8 9 3 5 6 2 3 1	4 8 9 3 8 Y 5 X
1 9 3 5 6 2 3 1	2 Y 7 Y 1 7 2 X
1 9 3 5 6 4 8 9	X 0
3 8 3 5 6 5 8 9	
2 8 3 5 7 5 8 9	
1 7 2 5 7 5 8 0	

Розгортання масивів при цьому методі стиснення виконується з їх початку або з кінця, а заповнення відповідної кількості пропусків замість додаткових символів X, Y, Z — перенесенням відповідної кількості символів, які знаходяться на однойменних розрядах попереднього рядка:

2 3 4 5 6 7 8 5
6 9 8
. . 7 2
2 4

8 9 3 . . 2 3 1
 1
 4 8 9
 3 8 . . . 5 . .
 2 . . . 7 . . .
 1 7 2 0

Стиснення інформації виключенням даних, які повторюються в різних файлах. За цим методом виконується така індексация (адресация) даних, яка дає змогу досягти логічної та фізичної незалежності їх. При цьому повторювані дані можуть відновлюватися багатократним зверненням до одного й того самого поля запису [5].

Стиснення інформації виключенням повторюваних символів. Виконується воно введенням двох додаткових символів, один з яких указує повторення, а інший — кількість їх (кількість літер, що повторюються в десятковій системі числення).

Наприклад, перелік виробів приладобудівного заводу:

1. Генератор сигналів низькочастотний ГЗ-109.
2. Генератор сигналів низькочастотний ГЗ-112/1.
3. Генератор сигналів спеціальної форми Г6-26.
4. Осцилограф СІ-81.
5. Осцилограф СІ-83.
6. Осцилограф СІ-65

у стисненому вигляді можна подати так:

1. Генератор сигналів низькочастотний ГЗ-109.
- 2.* 36—12/1.
3. * 17 спеціальної форми Г6-26.
4. Осцилограф СІ-81.
- 5.* 14—3.
- 6.* 13—65,

де * — знак повторення, а цифри після нього — кількість повторюваних знаків.

При відновленні даних розгортання виконується зверху вниз і справа наліво [4].

Зонне стиснення інформації. У цьому разі враховується те, що запис даних у пам'ять ЕОМ здійснюється за допомогою восьми восьмибітових слів, яким ставляться у відповідність літери та цифри повідомлення. Використання восьмибітового слова (байта) дає змогу закодувати $2^8 = 256$ знаків, тоді як реальні алфавіти з урахуванням цифр і деяких допоміжних символів містять до 50...60 знаків, тобто для кодування їх потрібні п'яти-шестибітові комбінації та аналогічні структури комірок пам'яті. Два-три біти, що залишаються, не вирішують проблему стиснення інформації, оскільки за їх допомогою можна записати тільки 4...8 знаків.

Однак, якщо використати півбайта для запису $2^4=16$ знаків деякого абстрактного алфавіту, а потім закодувати повідомлення в цьому $q=16$ -знаковому алфавіті по $m = 2$ знаки в кодовому слові, то одним байтом можна передавати ті самі $N=q^m = 16^2 = 256$ знаків.

Цей 16-знаковий алфавіт можна побудувати так, щоб 13 якісних ознак використовувалися як основні символи, а три — як допоміжні. Такий алфавіт матиме вигляд

0.	0000	4.	0100	8.	1000	C.	1100
1.	0001	5.	0101	9.	1001	D.	1101
2.	0010	6.	0110	A.	1010	E.	1110
3.	0011	7.	0111	B.	1011	F.	1111

Перші 12 символів умовно називатимемо ЦИФРА, а решту 4 — ЗОНА (співвідношення «цифр» і «зон» можна змінювати від 8:8 до 15: 1).

Умовимось, що в кодовому слові вторинного алфавіту перші чотири розряди завжди становитимуть зону, а чотири інші — цифри. Кількість можливих комбінацій вторинного алфавіту в даному разі буде $N = 4 \cdot 12 = 48$.

Для виконання зонного стиснення інформації потрібно знаки вторинного алфавіту розбити на зони. При цьому, якщо в тексті зустрічаються поруч знаки, які належать до однієї зони, номер її вказується тільки перед першим знаком, а запис наступних знаків обмежується записом цифрової частини їх.

Для того щоб знаки, які мають однакові зони, утворювали більш довгі послідовності при створенні кодових слів, у вторинному алфавіті необхідно враховувати статистичні характеристики алфавіту, з якого складаються тексти, які слід обробляти. Бажано також врахувати ймовірність різних сполучень деяких знаків.

Ефективність розбиття на зони встановлюють за допомогою потенціального коефіцієнта стиснення, який визначається виразом

$$K_{cm} = \frac{N_1}{N_2} \quad (4.19)$$

де N_1, N_2 — відповідно кількість байтів у первинному та стисненому текстах [5].

Стиснення інформації відбудеться тільки тоді, коли наступний знак у тексті належатиме до тієї самої зони. Тому ймовірність цієї події

$$p_1 = a^2 + b^2 + c^2 + d^2 \quad (4.20)$$

де a, b, c, d — ймовірності того, що наступний знак належить відповідно першій, другій, третій та четвертій зонам.

Ймовірність відсутності стиснення інформації

$$P_2 = 1 - P_1 \quad (4.21)$$

При стисненні на кожний знак відводиться півбайта, а при його відсутності — повний байт. Таким чином,

$$N_2 = N_1 P_1 / 2 + N_1 (1 - P_1) \quad (4.22)$$

$$\begin{aligned} K_{cm} &= \frac{N_1}{N_1 P_1 / 2 + N_1 (1 - P_1)} = \\ &= \frac{1}{(a^2 + b^2 + c^2 + d^2) / 2 + 1 - a^2 - b^2 - c^2 - d^2} = \\ &= \frac{2}{2 + a^2 + b^2 + c^2 + d^2 - 2a^2 - 2b^2 - 2c^2 - 2d^2} = \\ &= \frac{2}{2 - (a^2 + b^2 + c^2 + d^2)} \end{aligned} \quad (4.23)$$

Розподіл літер українського алфавіту та знаків на чотири зони наведено в табл. 4.1, де враховано частоту появи їх у тексті [2].

Таблиця 4.1 - Розподіл літер українського алфавіту та знаків на чотири зони.

Зона				Код знака в зоні
С	Д	Е	Ф	
Літера	Літера	Літера (знак)	Цифра (знак)	
Пробіл	М	Ф	-	0
О	П	Щ	1	1
Е	Є	Ш	2	2
А	З	Ц	3	3
Р	К	Ч	4	4

Л	Д	Ж	5	5
Т	Я	Х	6	6
Н	У	Ю	7	7
В	Ь	,	8	8
І	Б	.	9	9
И	Й	;	!	А
С	Г	:	?	В

Розглянемо стиснення інформації на конкретному прикладі. Закодуємо такий текст: «Постійний запам'ятовуючий пристрій є невід'ємною частиною електронної обчислювальної машини». У закодованій за допомогою табл. 4.1 формі він матиме вигляд

D1C1B69DAC7ADACOD3C3D1C3D0E8D6C618D7E74CADAC0D1C4AB649DAC0D2C0728
9D5E8D20C71E7C0E4C3B6A71E7C0252D4C641771901D9E4CAB5E7C835D8C7190D0C3E2CA7A
E9

Коефіцієнт стиснення тексту $K_{ст} = 1,3$.

Стиснення інформації зменшенням розрядності кодованих слів. За цим методом ефект стиснення інформації досягається завдяки поділу послідовності наперед упорядкованих чисел на кілька однакових відрізків, всередині яких відлік ведеться не за їх абсолютним значенням, а від межі попереднього відрізка. При цьому утворена розрядність чисел буде завжди менша від розрядності відповідних реальних чисел, для чого в пам'яті ЕОМ необхідно зберігати ці стиснені числа з розрядністю меншою, ніж розрядність реальних чисел.

Ємність пам'яті (двійкових розрядів) ЕОМ для розміщення N кодів визначається виразом

$$Q = N \log_2 M \quad (4.24)$$

де M — кількість кодів, що розміщуються в пам'яті ЕОМ.

Як впливає з (4.24), зі збільшенням M зростає довжина кодової комбінації. Тому для економії ємності пам'яті число $2^{\log_2 m}$ (де $\log_2 m$ округляється до найближчого цілого числа) розбивається на L однакових частин. Максимальне число в утвореному інтервалі чисел буде не більше $\log_2 \frac{M}{L}$, а ємність пам'яті для їх зберігання

$$Q_L = N \log_2 \frac{M}{L} \quad (4.25)$$

Якщо в пам'яті ЕОМ зберігати також адреси меж відрізків і порядкові номери чисел, що зберігаються, відлік яких ведеться від чергової межі, то розрядність чисел для запису номеру межі можна визначити як $\log_2 (N-1)$, а ємність пам'яті для зберігання номерів меж — як $(L-1) \log_2 (N-1)$, де $(L-1)$ — кількість меж між відрізками, причому $(L-1) \leq (N-1)$.

Загальна ємність пам'яті ЕОМ з урахуванням викладеного визначається виразом

$$Q' = N \log_2 \frac{M}{L} + (L-1) \log_2 (N-1) \quad (4.26)$$

Якщо продиференціювати (4.24) по L і прирівняти похідну

до нуля, то дістанемо значення $Q_{\text{тін}}$, яке буде при $L_{\text{опт}} = \frac{N}{\ln N}$.

Ємність пам'яті при оптимальній кількості зон, на які розбиваються числа, що зберігаються в пам'яті ЕОМ, знайдемо після підставлення в (4.24) значення $L_{\text{опт}}$:

$$Q'' = N \log_2 \frac{eM \ln N}{N} - \log_2 (N-1) \quad (4.27)$$

При $N > 100$ можна користуватися приблизною формулою

$$Q' = M \log_2 \frac{eM \ln N}{N} \quad (4.28)$$

Щоб знайти інформацію, записану в пам'ять ЕОМ, спочатку визначають $L_{\text{опт}}$, а потім — значення інтервалів між двома межами

$$C = \frac{2^{\lceil \log_2 M \rceil}}{L}, \quad (4.29)$$

де $\lceil \log_2 M \rceil$ — округлене до найближчого цілого числа значення $\log_2 M$. Після цього встановлюють інтервал, у якому знаходиться шукане число X :

$$K = X/C. \quad (4.30)$$

Адреса шуканого числа визначається як різниця між його абсолютним значенням і числом, що є граничним для цього інтервалу.

Виграш у ємності пам'яті визначається як

$$\Delta Q = Q - Q'' = N \log M - N \log_2 \frac{eM \ln N}{N} - \log_2(N-1). \quad (4.31)$$

Розглянемо конкретний приклад.

Якщо $M=800$, $X=300$ і $N=25$, то пошук числа 500 ведемо за такою послідовністю:

$$L=25/\ln 25=25/3.22=8; \quad C=1024/8=128; \quad K=X/C=300/128; \quad 3 > K > 2.$$

Отже, шукане число лежить у третьому інтервалі, де знаходяться числа від $128 \cdot 2 = 256$ до $383 = 256 + 127$ (додаємо 127, а не 128, оскільки в інтервалі знаходиться всього 128 чисел, з яких одне — нуль). Порядковий номер числа визначаємо як різницю між шуканим числом і числом, що є граничним для шуканого інтервалу:

$$300 - 256 = 44.$$

Знаходимо виграш у ємності пам'яті ΔQ :

$$Q = N \log_2 M = 25 \log_2 800 = 25 \cdot 10 = 250 \text{ двійкових розрядів};$$

$$Q'' = N \log_2 \left(\frac{eM \ln N}{N} \right) = 25 \log_2 \left(\frac{2.3 \cdot 800 \ln 25}{25} \right) = 200 \text{ двійкових розрядів};$$

$$\Delta Q = Q - Q'' = 250 - 200 = 50 \text{ двійкових розрядів.}$$

Стиснення інформації заміною деяких комбінацій літер одиничними символами. Цей метод ґрунтується на заміні деяких сполучень літер, які найчастіше зустрічаються в тексті, одиничними символами у вигляді двійкових кодових комбінацій, що не використовуються для подання знаків і символів при кодуванні як в ЕОМ, так і при обміні даними за допомогою ліній зв'язку.

Так, для подання алфавітно-цифрової та службової інформації в кодах КОІ-8 і ДКОІ з 256 можливих не застосовуються 167 символів, які не мають графічних еквівалентів. За цим способом стиснення інформації пропонується використати цей резерв для кодування біграм, що найчастіше зустрічаються в тексті. Ефективність стиснення науково-технічних текстів при цьому досягає 40% і більше. Даний метод можна поширити також на заміну в тексті сполучень з трьох і більше літер [4].

Стиснення інформації з використанням адаптивного кодування. Воно з успіхом застосовується при передачі техніко-економічної, статистичної та інших видів інформації, де, як правило, дуже висока вірогідність передачі потрібна не завжди.

Вимоги до вірогідності інформації можуть коливатися в широких межах залежно від характеру повідомлень. Так, при передачі статистичної інформації про кількість виробленої продукції помилка в молодших розрядах повідомлення (десятки та одиниці) менше впливатиме на правильність переданої інформації, ніж спотворення старших розрядів (мільйони та тисячі). При передачі текстових повідомлень вплив помилок на передачу ще менший.

Таким чином, доцільно передавати повідомлення з заданою вірогідністю у відповідних межах, використовуючи адаптивне кодування відносно джерела повідомлень, що дає змогу забезпечити вірогідність інформації, яка передається, залежно від вимог джерела повідомлень.

Для цього повідомлення необхідно поділити за категоріями з заданою вірогідністю передачі. На початку кожної категорії в інформацію вводяться додаткові службові комбінації, які при надходженні в кодер (під час передавання) та декодер (під час приймання) системи передачі перебудовують їх. Перебудова кодера може виконуватися також за сигналами від аналізатора інформації, якщо він установлений на вході системи передачі. Аналізатор обробляє інформацію, що надходить від джерела повідомлень, і поділяє її на категорії.

Ефективність адаптивного кодування залежить від співвідношення інформаційних об'ємів різних категорій в інформаційному потоці що стискується, тобто від ймовірності виникнення повідомлень визначеної категорії. Так, якщо загальна кількість повідомлень, що передаються N , множина категорій, які використовуються $X = \{x_1, x_2, \dots, x_k\}$, імовірності виникнення повідомлень визначених категорій $p(x_1), p(x_2), \dots, p(x_k)$, то задавшись визначеними довжинами n кодових комбінацій, котрими передаються повідомлення кожної з категорій множиною (n_1, n_2, \dots, n_k) , і тривалістю одного елемента масиву (символу) кодової комбінації $t_j, 1 \leq j \leq n_k$, можливо визначити основні характеристики даного методу стиснення.

Для цього скористаємось ентропією джерела інформаційного потоку

$$H(x) = -\sum_{i=1}^k p(x_i) \log_2 p(x_i) \quad (4.32)$$

де k - число категорій, що використовуються при стисненні інформації.

Середній час передачі одного повідомлення:

$$t_{ycp} = p(x_1) \sum_{j=1}^{n_1} t_j + p(x_2) \sum_{j=1}^{n_2} t_j + p(x_k) \sum_{j=1}^{n_k} t_j = \sum_{i=1}^k p(x_i) \sum_{j=1}^{n_i} t_j \quad (4.33)$$

Тоді, з урахуванням (4.32) і (4.33), швидкість передачі стиснутого інформаційного потоку буде визначатись за формулою:

$$R_{cm} = \frac{H(x)}{t_{ycp}} = \frac{-\sum_{i=1}^k p(x_i) \log_2 p(x_i)}{\sum_{i=1}^k p(x_i) \sum_{j=1}^{n_i} t_j} \quad (4.34)$$

Визначимо швидкість передачі інформаційного потоку до стиснення

$$R = \frac{-\sum_{i=1}^k p(x_i) \log_2 p(x_i)}{\sum_{j=1}^{nk} t_j} \quad (4.35)$$

де $\sum_{j=1}^{nk} t_j$ тривалість передачі кодової комбінації (одного повідомлення) до стиснення, котра

визначається часом передачі кодової комбінації максимальної тривалості, призначеної для передачі повідомлень категорій інформації самого більшого порядку.

Для передачі доцільно використовувати три категорії інформації з різною вірогідністю (з різною ймовірністю помилок): $10^{-3} \dots 10^{-4}$ — для передачі текстів; $10^{-5} \dots 10^{-6}$ - для передачі цифрових повідомлень від одиниць до тисячі; $10^{-6} \dots 10^{-9}$ - для передачі цифрових повідомлень більше тисячі та будь-яких дуже важливих цифрових і текстових повідомлень. Принципи поділу інформації на категорії можуть змінюватися залежно від конкретних умов і вимог.

Передачу інформації з різною вірогідністю можна забезпечити зміною способу кодування. Так, для передачі інформації по проводових каналах з вірогідністю $10^{-3} \dots 10^{-4}$ досить застосувати двійковий первинний код без уведення будь-якого захисту інформації; з вірогідністю $10^{-5} \dots 10^{-6}$ —

двійковий код, що виявляє помилки, а з вірогідністю $<10^{-6}$ — код, який виправляє відповідну кількість помилок. Тоді для передачі, наприклад, 32 кодових комбінацій їх довжина в першому випадку становитиме $n_1=5$, у другому — $n_2=6\dots7$, у третьому — $n_3=8\dots 12$ елементів.

Середня довжина кодової комбінації при використанні адаптивного кодування залежатиме від процентного співвідношення повідомлень різної категорії і загалом може бути визначена як

$$n_{сер} = \frac{n_1 N_1 + n_2 N_2 + n_3 N_3}{N} \quad (4.36)$$

де n_1, n_2, n_3 — довжини кодових комбінацій при передачі повідомлень першої, другої та третьої категорій; N_1, N_2, N_3 — відповідна кількість цих комбінацій; N — загальна кількість їх [6].

Коефіцієнт стиснення інформації можна визначити як відношення максимальної довжини n кодової комбінації до її середньої довжини $n_{сер}$:

$$K_{cm} = n/n_{сер} \quad (4.37)$$

Наприклад, при передачі $N=1000$ повідомлень, з яких $N_1=700, N_2=200$ і $N_3=100$, якщо $n_1=5, n_2=6, n_3=8$, матимемо

$$n_{сер} = (5 \cdot 700 + 6 \cdot 200 + 8 \cdot 100) / 1000 = 5,5 \text{ елемента}$$

і коефіцієнт стиснення

$$K_{ст} = 8 / 5,5 = 1,45.$$

Стиснення інформації збільшенням основи коду. Грунтується воно на перекодуванні кодованих послідовностей символів системи числення з меншою основою в систему з більшою основою. Так, якщо інформаційний масив, який складається з двійкових елементів, подати у вісімковій або шістнадцятковій системі числення (тобто перекодувати його), то це приведе до значного зменшення кількості елементів масиву (в кілька разів):

0000010100		111001011101
0010100111		001011101110
0100111001	^ ^	011101110000
0111001011	^ ^	101110000010
1001011101	^ ^	110000010100
1011101110	^ ^	000010100111
1101110000	^ ^	010100111001
1110000010	^ ^	100111001011
	10	

$N_1=192;$

01234567		053977
12345670		29CB B 8
23456701		4E5DC1
34567012		72EE0A
45670123		977053
56701234		BB829C
67012345		DCI4E5
70123456		E0A72E

$N_2=64;$

$N_3=48.$

При цьому коефіцієнти стиснення первинного масиву двійкових елементів у вісімкову і шістнадцяткову системи числення становитимуть

$$K_{ст8}=N_1/N_2=192/64=3;$$

$$K_{ст16}=N_1/N_3=192/48=4,$$

тобто залежать від кількості бітів, що несе один елемент інформаційного масиву (у вісімковій системі $8 = 2^3 \rightarrow 3$ біт/елемент, у шістнадцятковій — $16 = 2^4 \rightarrow 4$ біт/елемент).

4.4.2. Матричні методи стиснення

До ефективних матричних методів стиснення інформації при передачі даних належить **метод зі зберіганням атрибутів у вигляді бітової матриці**. При цьому методі скінченна кількість атрибутів (табл. 4.2) виноситься в першу частину («шапку») матриці, тілом якої є набір двійкових елементів, з яких 1 означає наявність, а 0 — відсутність атрибута (табл. 4.3). «Шапка» бітової матриці та її тіло можуть зберігатися на різних ділянках пам'яті ЕОМ [2].

Таблиця 4.2

Номер виробничої дільниці	Тип комплектувального виробу					
	0301	1017	2216	1100	1017	0301
0001	0301	1017	2216	1100		
0012	1008	0011	2001	2216	1017	
0015	1100	2001	2000	1008	0301	
0117	1100	0011	0301			
1206	2001	2216	0011	0301	1100	1017

Таблиця 4.3

Номер виробничої дільниці	Тип комплектувального виробу							
	0011	0301	1008	1017	1100	2002	2001	2216
0001	0	1	0	1	1	0	0	1
0012	1	0	1	1	0	0	1	1
0015	0	1	1	0	1	1	1	0
0117	1	1	0	0	1	0	0	0
1206	1	1	0	1	1	0	1	1

Значний ефект стиснення дає інший **матричний метод із заміною елементів**, що повторюються на деякій площі інформаційного масиву, одиничними елементами, які несуть ознаки цих ділянок масиву. Так, якщо ввести символи, що відбивають деякі обмежені площі інформаційного масиву, в яких елементи повторюються порівняно з однойменними розрядами попереднього рядка (наприклад, R-2x2, X-2x5, Y-3x2, Z-5x3), то інформаційний масив

1	2	3	4	5	6	7	8	9	0
2	2	3	4	8	6	7	5	9	0
3	2	3	4	7	6	7	3	9	0
3	2	3	4	7	6	7	3	9	0
3	2	3	4	7	6	7	3	9	0
3	2	3	4	7	8	9	1	9	0

після згортання матиме вигляд

```
1 2 3 4 5 6 7 8 9 0
2 Y 8 R 5 X 3 7 3 Z
Y 8 9 1
```

Розгортка виконується з початку масиву:

```
1 2 3 4 5 6 7 8 9 0
2 . . . 8 . . 5 . .
3 . . . 7 . . 3 . .
. . . . . . . . . .
. . . . . . . . . .
. . . . . 8 9 1 . .
```

Коефіцієнт стиснення $K_{ст} = 60/24 = 2,5$.

Стиснення інформації з типовими матрицями застосовується головним чином для відносно великих масивів. При цьому в пам'яті ЕОМ зберігаються матриці деяких найпоширеніших наборів символів, які зустрічаються в інформаційному масиві, що стискається. Для цього заздалегідь аналізують інформаційний масив, утворюють типові матриці і при стисненні замість наборів символів, які відповідають цій типовій матриці, в послідовність символів інформаційного масиву вводять адреси цих матриць [5].

Скориставшись розробленою програмою, котра дозволяє оцінити ефективність методів стиснення, що використовуються, покажемо порівняння трьох варіантів матричного стиснення:

- 1 – з апіорно заданими розмірами матриць;
- 2 – з використанням типових матриць;
- 3 – з попереднім аналізом вхідного потоку.

Стиснення з апіорно заданими розмірами матриць

Цей алгоритм матричного стиснення повністю збігається з класичним матричним стисненням, однак при цьому введені такі умови:

- геометрія масиву вибірки постійна і задана апіорно як у передавальному, та і у приймальному пристроях;
- розміри матриць задані апіорно як у передавальному, так і у приймальному пристроях;
- кожній матриці заздалегідь ставиться у відповідність деякий символ вторинного алфавіту.

Слід зазначити, що на коефіцієнт стиснення даного алгоритму буде впливати відповідність підібраних апіорно матриць (їх параметри) характеру вхідного інформаційного потоку. Розміри таких матриць, найменш чутливих до характеру інформаційного потоку, які, у той же час, дозволяють досягти найбільшого коефіцієнта стиснення, визначаються, як правило, за допомогою експертних оцінок.

Як основну перевагу даного алгоритму у порівнянні з іншими, можна відзначити досить високий коефіцієнт стиснення завдяки відсутності передачі у вихідному потоці будь-якої службової інформації про розміри матриць, які заздалегідь зафіксовані у передавальному і приймальному пристроях.

Стиснення з попереднім аналізом вхідного інформаційного масиву

Алгоритм такого стиснення аналогічний наведеному вище. Однак, при цьому розміри матриць не задаються апіорно, а визначаються у результаті аналізу масиву вибірки. Іншими

словами, даний алгоритм є багато прохідним: на першому проході визначаються всі можливі матриці даного масиву вибірки, далі виконується аналіз – визначення декількох (перших двох для алфавіту $q = 4$, перших шести при $q = 8$ тощо) матриць, які покривають найбільшу площу масиву вибірки; у другому проході виконується згортка масиву вибірки за допомогою звичайного алгоритму матричного стиснення. Далі масив вибірки передається у вихідний інформаційний потік, в кінці якого записується стоп символ, кількість і розміри визначених матриць. Розгортка повідомлень у приймальному пристрої виконується так само, як і у попередньому випадку, з тією різницею, що спочатку завантажується з вхідного потоку вся інформація до стоп-символу, далі зчитуються розміри матриць, і тільки після цього виконується дійсно розгортка. Всього може бути використано, як було вказано вище, тільки $(q - 3)$ матриць, тому що символи «0» і «1» - інформаційні, а один символ зарезервованій під стоп-символ.

Навіть без попереднього аналізу можна зазначити, що цей алгоритм буде трохи програвати попередньому, з огляду на багато прохідність, за часом, та, якщо врахувати необхідність передачі разом зі згорнутим масивом додаткової інформації про розміри матриць, за коефіцієнтом стиснення. Однак, стиснення самого інформаційного масиву виконується краще, тому що розміри матриць адаптуються до даного конкретного масиву вибірки.

4.4.3. Комбіновані методи

Комбіновані методи стиснення це новий клас способів стиснення інформації. З них як приклад можна навести **лінійно-матричне стиснення інформації**, при якому первинний інформаційний масив стискається за допомогою додаткових елементів, якими позначають повторення елементів у рядках порівняно з попереднім (наприклад, символами K-2, L-3, M-5 та в обмежених ділянках — матрицях символами X-2x2, Y-2x5, Z-3x5). До лінійно-матричних способів стиснення відносяться поздовжньо-матричне і поперечно-матричне стиснення, коли і поздовжнє (поперечне) і матричне стиснення використовуються одночасно. При цьому, на відміну від матричного стиснення, крім матриць розмірності $m \times n$ - число рядків і стовпців, введені матриці розмірності $1 \times n$ (або, при поперечно-матричному стисненні $m \times 1$). Тоді, якщо інформаційний масив має вигляд

1	2	3	4	5	6	7	8	9	0
2	2	3	4	5	6	7	5	9	0
2	2	3	4	5	6	4	6	9	0
2	2	3	4	5	6	4	4	9	0
3	2	3	4	5	6	4	2	9	0
3	2	3	4	5	7	4	2	9	0
3	2	3	4	5	7	4	2	9	0

то згорненим він матиме такий вигляд:

```

1 2 3 4 5 6 7 8 9 0
2 2 Z K 5 Y X 6 4 6
X 4 3 2 2 X 7 K L M

```

Розгортання виконується з початку масиву:

```

1 2 3 4 5 6 7 8 9 0
2 2 . . . . . 5 . .
. . . . . 6 4 6 . .
. . . . . . . 4 . .
3 2 . . . . . 2 . .
. . . . . 7 . . . .
. . . . . . . . . .

```

Коефіцієнт стиснення $K_{ст} = 70/30 = 2,33(3)$.

4.4.4. Каскадні методи

Застосовуються для збільшення коефіцієнта стиснення. При цьому, як правило, використовуються комбінації лінійних і матричних методів. Можна утворити такі каскадні способи стиснення, як кодо-зонний, кодо- та зонно-матричні тощо.

При використанні послідовного кодового і матричного способів стиснення отримують каскадний кодо-матричний метод.

В даному випадку вхідний інформаційний потік підлягає кодовому стисненню, а потім - матричному. При цьому коефіцієнт стиснення буде дорівнювати добутку:

$$K_{cm_{км}} = K_{cm_k} * K_{cm_m} = \frac{N_1 * \log_2 q_2}{(N_2^{j_0} + n_1 + n_2 + \dots + n_k) * \log_2 q_1}, \quad (4.50)$$

де N_1 - загальна кількість елементів (символів) вхідного інформаційного потоку; N_2 - число одиночних елементів (символів) в вхідному інформаційному потоці; n_1, n_2, \dots, n_k - число матриць різної розмірності; $k = \{1, 2, \dots, q\}$, тут $q = q_2 - q_1$; q_2 - алфавіт, що використовується для кодування вихідного (вторинного) інформаційного потоку; q_1 - частина алфавіту q_2 , що використовується у вторинному потоці для кодування одиночних символів (а також, алфавіт вхідного інформаційного потоку) [5].

4.4.5. Каскадно-комбіновані методи стиснення

Кодо-поздовжньо-матричне і кодо-поперечно-матричне стиснення відноситься до каскадно-комбінованих способів стиснення, так як такі способи передбачають на першому етапі кодове стиснення, а на другому комбіноване - поздовжньо-матричне (або поперечно-матричне) стиснення. Коефіцієнт стиснення визначається добутком коефіцієнтів стиснення і поздовжнього-матричного (поперечно-матричного), тобто:

$$K_{cm_{кпм}} = K_{cm_k} * K_{cm_{пм}}, \quad (4.51)$$

Або з урахуванням (4.51) отримаємо вираз:

$$K_{cm_{кпм}} = \frac{N_1 * \log_2 q_2}{(N_2^{j_0} + n_n + n_1 + n_2 + \dots + n_k) * \log_2 q_1}, \quad (4.52)$$

де N_1 - загальна кількість елементів (символів) вхідного інформаційного потоку; $N_2^{j_0}$ - число одиночних елементів (символів) в вхідному інформаційному потоці; $k = \{1, 2, \dots, q\}$, тут $q = q_2 - q_1$; q_2 - алфавіт, що використовується для кодування вихідного (вторинного) інформаційного потоку; q_1 - частина алфавіту q_2 , що використовується в вихідному інформаційному потоці для кодування одиночних символів; n_n - число матриць розмірності $1 \times m_2$ (при кодо-поздовжньому-матричному стисненні), або розмірності $m_1 \times 1$ (при кодо- поперечно-матричному стисненні).

4.4.6 Вибір методу стиснення при передачі

Для того, щоб вибрати найкращий спосіб стиснення скористаємося відомими залежностями коефіцієнта стиснення $K_{ст}$ від довжини інформаційного масиву при різних алфавітах q . Представимо ці залежності у вигляді таблиць та графіків при $q=4$, $q=8$ та $q=16$,

Таблиця 4.4. Залежність коефіцієнта стиснення $K_{ст}$ від довжини інформаційного масиву при $q=4$.

Метод стиснення	$K_{ст}$ при інформаційному масиві, кбіт				
	0.32	0.64	0.96	1.28	1.6
1. Лінійний з ρ	1.58	1.66	1.72	1.73	1.75
2. Лінійний з x, y, \dots, z	3.81	4.41	5.05	5.04	5.26
3. Кодовий	2	2	2	2	2
4. Матричний	4.10	4.42	5.0	4.92	5.23

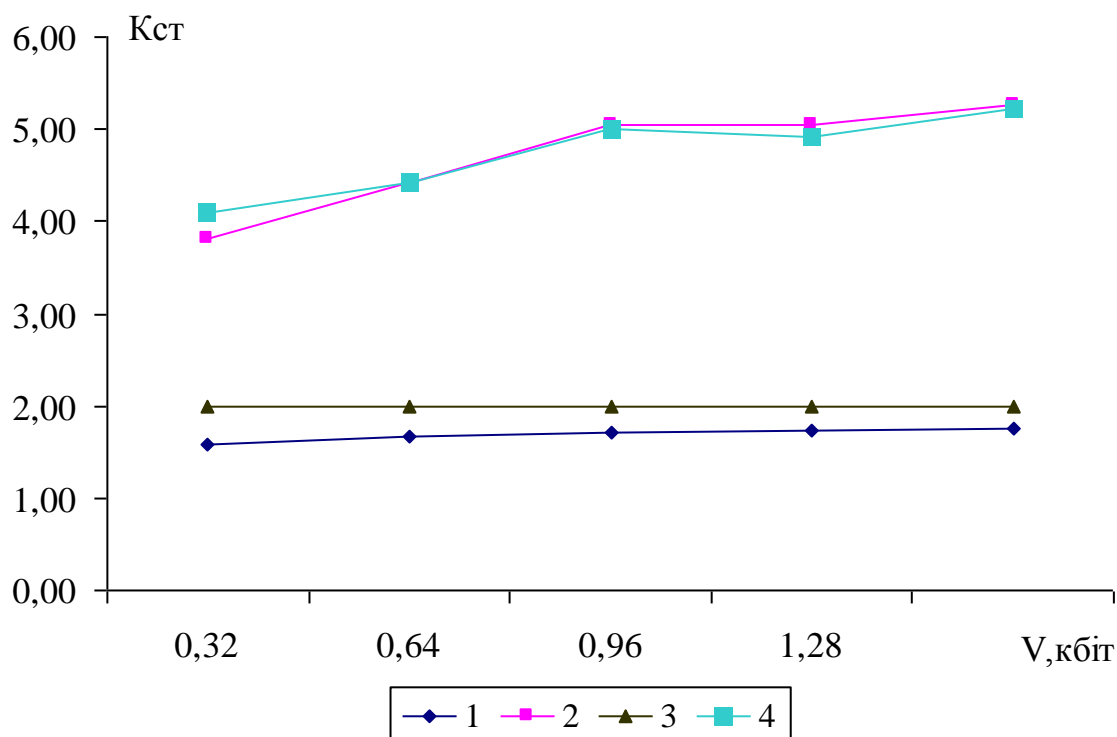


Рис.4.3. Графік залежності $K_{ст}$ від довжини інформаційного масиву при алфавіті $q=4$.

Таблиця 4.5. Залежність коефіцієнта стиснення $K_{ст}$ від довжини інформаційного масиву при $q=8$.

Метод стиснення	$K_{ст}$ при інформаційному масиві, кбіт				
	0.32	0.64	0.96	1.28	1.6
1. Лінійний з ρ	1.58	1.66	1.72	1.73	1.75
2. Лінійний з x, y, \dots, z	3.90	6.10	7.56	8.53	9.30
3. Кодовий	3.00	3.00	3.00	3.00	3.00
4. Матричний	4.10	6.21	8.07	9.62	10.53
5.Кодо-матричний	7.60	11.85	12.63	14.06	14.68
6.Лінійно-матричний	4.10	6.40	8.65	9.77	10.50
7.Кодо-лінійно-матричний	7.70	12.30	14.12	16.62	16.80

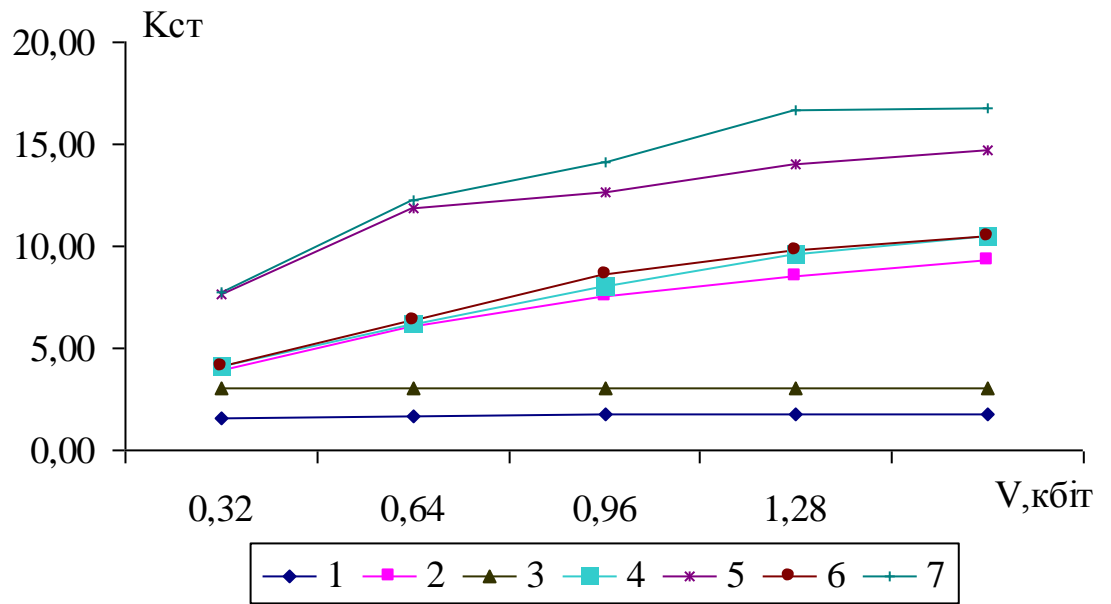


Рис. 4.4. Графік залежності $K_{ст}$ від довжини інформаційного масиву при алфавіті $q=8$.

Таблиця 4.6 - Залежність коефіцієнта стиснення $K_{ст}$ від довжини інформаційного масиву при $q=16$.

Метод стиснення	$K_{ст}$ при інформаційному масиві, кбіт				
	0.32	0.64	0.96	1.28	1.6
1. Лінійний з ρ	1.58	1.66	1.72	1.73	1.75
2. Лінійний з x, y, \dots, z	4.05	6.21	7.93	8.77	9.76
3. Кодовий	4.00	4.00	4.00	4.00	4.00
4. Матричний	4.10	6.40	8.20	9.70	10.81
5. Кодо-матричний	10.00	12.55	17.14	18.82	20.25
6. Лінійно-матричний	4.10	6.60	8.89	10.24	11.68
7. Кодо-лінійно-матричний	10.00	14.22	16.27	17.78	19.28

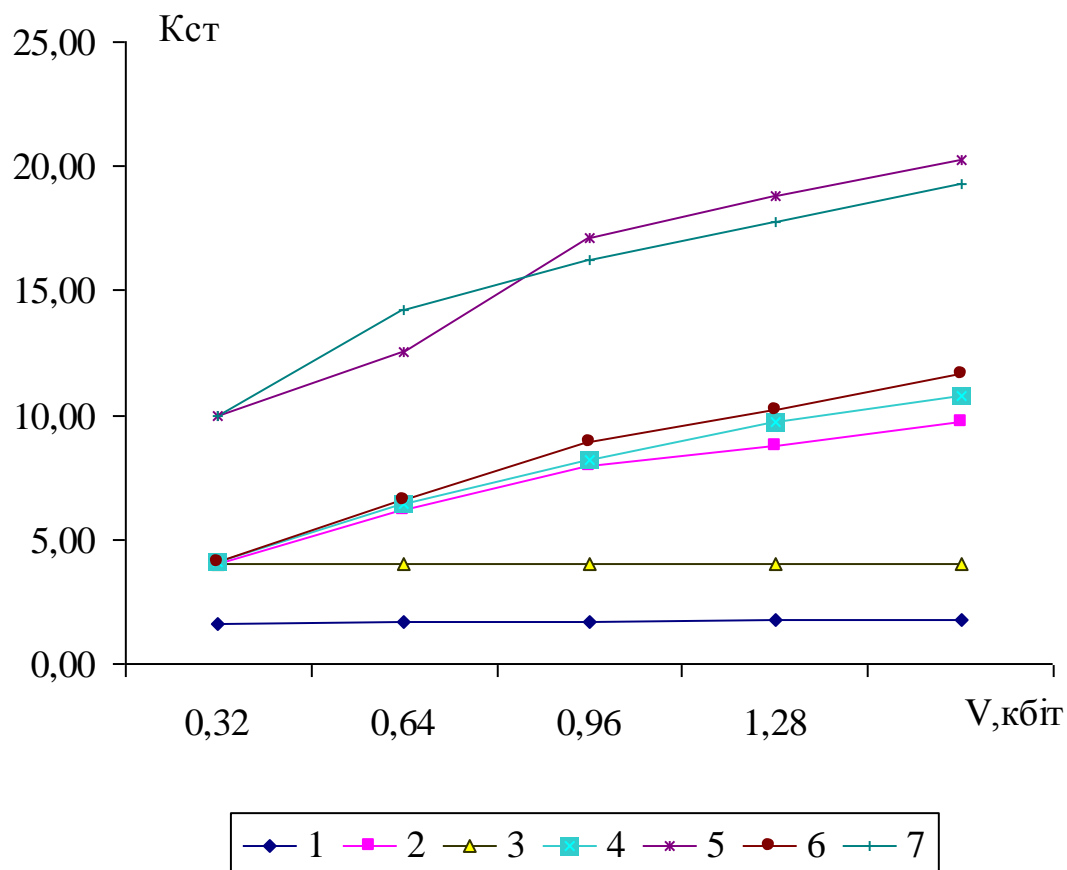


Рис.4.5. Графік залежності $K_{ст}$ від довжини інформаційного масиву при алфавіті $q=16$.

З таблиць та графіків можна зробити висновок, що для алфавіту $q=4$ найкращими способами стиснення є лінійний з x, y, \dots, z та матричний, а для алфавітів $q=8$ та $q=16$ - кодо-матричний та кодо-лінійно-матричний способи, тобто каскадні та комбіновані способи стиснення.

4.5. Класифікація алгоритмів стиснення при передачі та архівації

Будь-які алгоритми стиснення базуються на методах та способах які вказують основний напрямок стиснення. Деякі методи можуть мати досить багато різних алгоритмів реалізації, але основна ідея стиснення залишається. Існує два основних класи методів стиснення: оборотне і необоротне. При необоротному стисненні початкова інформація не може бути у відповідності відновлена із стисненої. Ці методи використовуються при передачі інформації, що належить для сприйняття людиною (аудіо дані, нерухомі і рухомі зображення), в тих випадках, коли нема необхідності відновити інформацію з стовідсотковою точністю. Недоліки цих методів очевидні - частина інформації втрачається при стисненні, перевага - висока ступінь стиснення, що звичайно, в декілька разів вища, чим при оборотному стисненні.

Оборотне стиснення дозволяє в точності відновити стиснену інформацію. Оборотне стиснення знаходить широке використання, забезпечуючи більш швидку передачу даних по каналам зв'язку і скорочення простору, необхідного для їх зберігання на носіях інформації.

За способом видалення надлишковості виділяють два основних різновиди алгоритмів оборотного стиснення: статистичні і словникові. Статистичні методи стиснення використовують кодування повідомлень, маючих різну імовірність появи, кодовими комбінаціями різної довжини. Символам, що більш часто зустрічаються привласнюються короткі кодові комбінації, тим що рідко - більш довгі.

Окремий клас статистичних методів стиснення складають методи, які використовують так зване кодування з передбаченням. Ці методи виконують збирання статистики про

взаємозалежності між символами по мірі надходження вхідних даних і потім намагаються передбачити черговий символ, що надходить. Якщо передбачене було вірним, воно не записується у вихідний потік. Якщо виникла помилка, алгоритм коректує статистику на основі отриманого символу. Чим краще виконується передбачення, тим менше невірних передбачених символів надійде в вихідний потік і вище коефіцієнт стиснення.

Іншим направленням в стисненні даних є словникові методи стиснення. Основна ідея їх полягає в тому, що в початковому тексті шукаються повторювані ланцюжки символів, котрі потім замінюються їх індексами в словнику. Словник - це список таких ланцюжків, котрі, як припускається, будуть часто використовуватись. Індеси формуються так, що в середньому займають менше місця, чим представлені ними ланцюжки, за рахунок чого і досягається стиснення.

Так само як і статистичне стиснення, словникове стиснення може бути статичним, напіваадаптивним та адаптивним. Статичні алгоритми працюють із заздалегідь заданим словником, що не змінюється в процесі кодування. Вони корисні в тих випадках, коли буде достатнім невисокий рівень стиснення, який досягається за рахунок невеликих витрат. Статичні методи ефективні також при стисненні такої інформації, як вихідні тексти програм на мовах програмування високого рівня.

Напіваадаптивне словникове стиснення є подальшим розвитком статичного підходу. При напіваадаптивному стисненні алгоритм працює в два проходи: на першому для даних, що кодуються, створюється словник, на другому – здійснюється власне стиснення даних за сформованим словником. При цьому досягається більш високий ступінь стиснення, однак, в умовах передачі інформації по каналах зв'язку, не завжди є можливість попереднього перегляду даних, що обмежує застосування цього методу.

При адаптивному способі словник будується на підставі даних, що обробляються. Цей спосіб швидко пристосовується до структури даних, однак вимагає великого обсягу пам'яті для збереження словника.

Крім поданої класифікації, всі способи стиснення повідомлень можна також розділити на однопрохідні і двопрохідні. Однопрохідний спосіб стиснення здійснює кодування символів у міру їхнього надходження, двопрохідний - здійснює попередній перегляд даних перед кодуванням. Статичні й адаптивні алгоритми звичайно є однопрохідними, напіваадаптивні - двопрохідними. Оскільки двопрохідні методи вимагають перед початком кодування наявності всіх даних, при використанні цих способів розмір блока даних сильно впливає на швидкість кодування і затримку даних. Це необхідно враховувати при організації стиснення даних у каналах зв'язку, які працюють у реальному часі.

За способами обробки даних способи стиснення можуть бути одно-, дво- та тривимірні. Одновимірний спосіб стиснення обробляє дані, що надходять, як один рядок. Більшість способів, які застосовуються для стиснення текстів і двійкових даних, є одновимірними.

Двовимірний спосіб представляє вхідні дані як послідовність рядків однакової або неоднакової довжини. При кодуванні такі способи використовують статистичну залежність не тільки між окремими символами, але і між сусідніми рядками даних. Двовимірні способи широко застосовуються для кодування зображень. Їх можна ефективно використовувати для стиснення даних, які мають чітку структуру, наприклад, записів бази даних або інформації, що надходить від датчиків. У цьому випадку кодер повинен мати інформацію про довжину запису, щоб перетворити одновимірний вхідний потік у двовимірний.

Тривимірні способи стиснення, крім розбиття вихідної інформації на рядки, здійснюють поділ груп рядків на окремі кадри і використовують статистичну залежність між даними сусідніх кадрах. Ці способи знаходять застосування при стисненні рухомих зображень.

Проілюструємо вищесказану класифікацію алгоритмів яка представлена на рисунку 4.6:

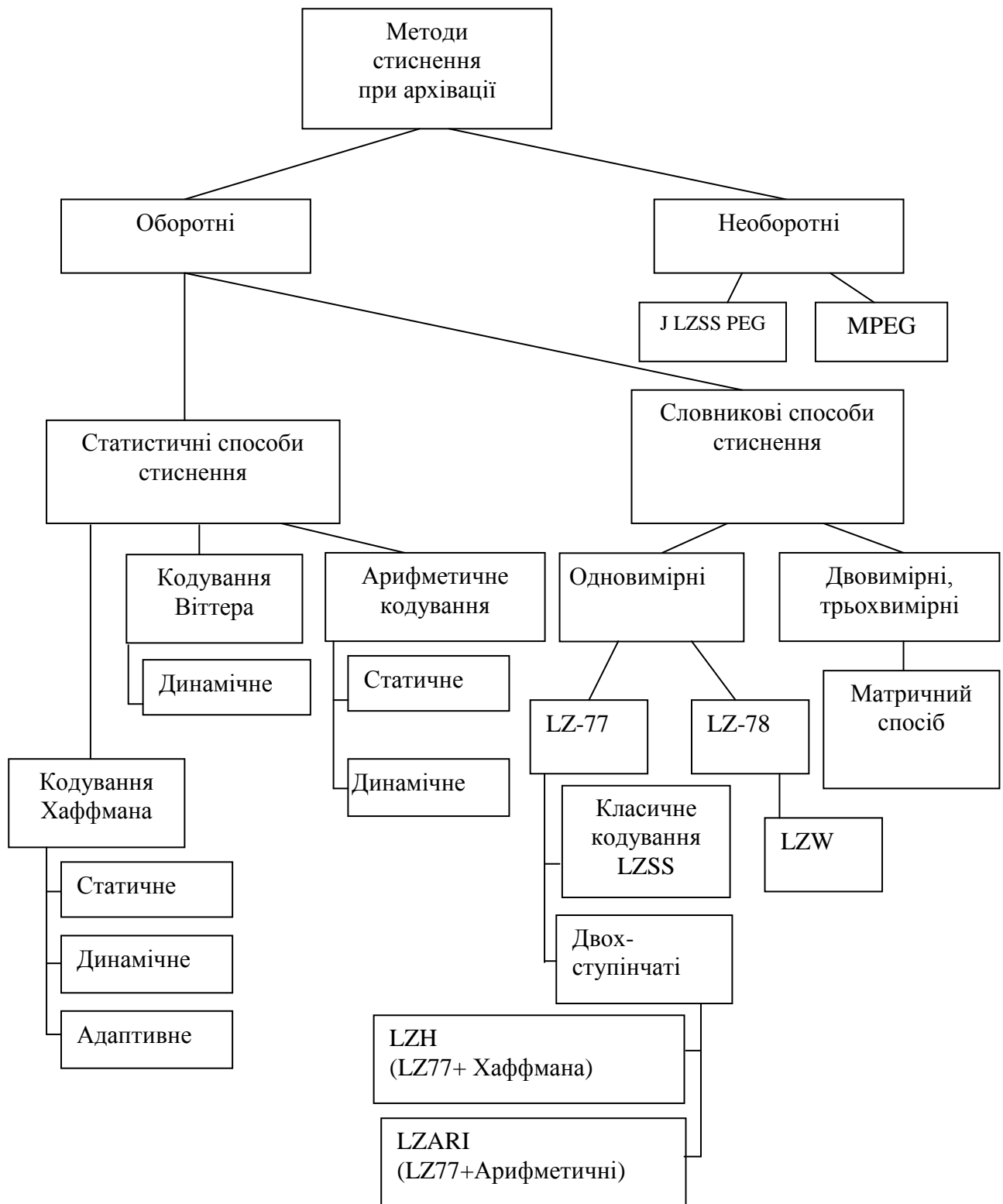


Рис. 4.6. Класифікація методів стиснення при архівації та передачі

4.5.1. Алгоритм стиснення Хаффмена

Щоб зрозуміти, навіщо потрібно адаптивне стиснення, потрібно спочатку з'ясувати, що собою представляє оригінальний алгоритм Хаффмана. В його основі лежить наступна ідея. Символам повідомлень ставиться у відповідність код змінної довжини, що складається з цілої кількості біт. Довжина коду визначається частотою появи символу в даних що підлягають стисненню, чим рідше зустрічається символ, тим довше його код. Більш того, процедура побудови кодів гарантує однозначність їх наступного декодування завдяки властивості унікальності префікса.

Яким же чином будується код символу? Для цього нам необхідно знати таблицю частот появи символів в вхідних даних, або імовірності появи символів, із алфавіту що має джерело повідомлень. На основі цієї таблиці будується дерево Хаффмана (H - дерево) за наступним алгоритмом:

- із символів, які зустрічаються в вхідних даних, створюють список вільних джерел. Кожен вузол характеризується вагою, під якою розуміють імовірність появи символу, чи кількість його появ в даних;
- вибираємо два вільних вузли, вага котрих мінімальна і створюємо для них батьківський вузол. При цьому вага батьківського вузла буде дорівнювати доданку ваги його дітей;
- додаємо батьківський вузол в список вільних вузлів, а його дітей, звідти, викреслюємо. Дугам, які з'єднують батьків з дітьми, ставимо у відповідність 0 і 1.
- два попередніх кроки повторюємо до тих пір, поки в списку вільних вузлів не зостанеться тільки один вузол, котрий називають коренем дерева.

Тепер, знаючи дерево Хаффмана, ми можемо легко визначити код будь-якого символу. Для цього необхідно пройти шлях від кореня дерева до вузла відповідного цьому символу. Отримана послідовність нулів та одиниць і буде, представляти собою код Хаффмана.

Розглянемо приклад, який виконує стиснення, користуючись вищеприведеним алгоритмом.

Нехай, задано кількість появ символів деякого алфавіту, ці дані зведено в таблицю 4.8:

Таблиця 4.8 - Кількість появ символів деякого алфавіту на виході джерела повідомлень.

Літера алфавіту	і	к	л	м	н	о	п	р
Кількість появ (частота)	7	8	3	4	3	10	3	5

На основі заданої таблиці побудуємо відповідне дерево Хаффмана та визначимо по ньому коди для кодові комбінації для кожного із символів.

Отримані кодові комбінації для літер алфавіту запишемо в таблицю 4.9.

Таблиця 4.9 - Коди символів алфавіту.

Літера алфавіту	і	к	л	м	н	о	п	р
Кодова послідовність	011	10	0100	0101	0001	11	0000	001

Проаналізувавши отримані дані, можливо сказати, що - ні один з кодів не є префіксом іншого коду, що гарантує однозначність декодування. Також видно, що самий короткий код - у найбільш часто зустрічаємих символів "к" і "о", а рідко зустрічаємі символи ("л", "м", "н" і "п") представлені самим довгим кодом.

Однак, оригінальний алгоритм досить простий, він має два значні недоліки. По-перше, для коректної роботи розпаковник повинен знати таблицю кодів, створену запаковником. Отже, її необхідно зберігати в архівному матеріалі, або передати декодеру іншими каналами, тим самим збільшуючи об'єм передаваної інформації. По-друге, так як нам необхідна статистика символів що містяться в початкових даних, то потрібно два проходи по ним - один для збору статистика, а другий вже для стиснення.

Позбавитись цих недоліків допомагає адаптивне стиснення. Завдячуючи йому число проходів скорочується до одного. Досягається це за рахунок того, що інформація про символи, що зустрічаються у початкових даних, збирається в процесі стиснення, а сам процес стиснення починається з "пустої" моделі.

Дійсно такий алгоритм починає працювати з "порожнього" дерева. Але "пусте" в даному випадку не означає "зовсім нічого не містить". В термінології адаптивного алгоритму вираз "пустий" означає, що до початку роботи у запаковника немає ніякої інформації про вхідні дані, тобто у випадку метода Хаффмана у нас немає статистики вмісту символів. Упаковник в дерево початково заносить два вузли, відповідних спеціальним символам. Перший з них - це признак закінчення вхідних даних (позначаємо його EOF - EndOfFile). Він потрібен для визначення моменту закінчення декомпресії. Другий кодовий елемент, позначає, що наступний символ був закодований без урахування дерева Хаффмана (позначимо через ESC). Його використовують в тому випадку, коли в даних що стискаються зустрічається символ, котрого ще немає в H - дереві (і відповідно, для нього неможливо визначити код Хаффмана). Кожному з цих спеціальних символів присвоюємо вагу, рівну одиниці. Так як, це службові символи і в вхідних даних вони будуть зустрічатися, то в процесі стиснення вони будуть переміщатися на самі довгі кодові комбінації.

Змінився зовнішній вигляд вузлів дерева. Крім полів, що вказують, що за символ зберігається і ваги вузла (w), додалося ще одне поле - номер вузла (n). Справа у тому, що у адаптивному алгоритмі використовується упорядковане дерево Хаффмана, тобто вузли дерева можуть бути перераховані в порядку зростання їх ваги. При цьому порядку при перерахуванні, кожний вузол буде знаходитися поряд з своїм "братом". Ця властивість буде використовуватись при оновленні дерева. Будемо послідовно зчитувати по одному символу із вхідних даних до тих пір, поки вони не закінчаться. Для кожного прочитаного символу виконуємо одну з наступних дій:

- якщо прочитаного символу немає в H - дереві, то в вихідну інформацію записується службовий ESC - код і сам прочитаний символ;
- інакше по H - дереву необхідно визначити код прочитаного символу і вивести цей код у вихідний потік;
- після виводу необхідно провести оновлення дерева Хаффмана прочитаним символом.

Після завершення цих операцій необхідно записати також признак завершення стиснення - код EOF.

Алгоритм роботи розпаковника виглядає наступним чином. Із запакованого пакета виділяються і зчитуються по одному коду символів. В залежності від коду виконується наступна дія:

- якщо це признак закінчення стиснення - закінчити розпаковування;
- інакше - якщо це ESC - код, то прочитати із пакета символ, в протилежному випадку визначити по H - дереву символ, відповідний прочитаному коду;
- вивести символ у вихідний пакет;
- провести оновлення дерева Хаффмана розкодованим символом.

Основну складність представляє процедура оновлення дерева, тому розглянемо її детальніше. При оновленні дерева виконується дві операції. Перша полягає в послідовному збільшенні ваги вузлів дерева. Спочатку збільшується вага вузла відповідного символу, потім вага його батька, далі вага батька батька і так до тих пір, поки не буде збільшена вага кореня. Друга операція буде виконуватись в тому випадку, коли збільшення ваги довільного із вузлів приведе до порушення властивості упорядкованості H - дерева (тобто нова вага вузла стане більше, чим вага вузла, наступного за ним в списку). В цьому випадку необхідна перестановка вузлів дерева. Нехай після зміни ваги вузла вона була рівною $w+1$. Тоді в списку вузлів ми шукаємо останній по номеру вузол з вагою w і переставляємо ці два вузли місцями (звичайно, замінивши для них батьків). На цьому перестановка завершена. Після перестановки продовжуємо операцію збільшення ваги вузлів - наступним в черзі збільшення ваги буде новий батько вузла, збільшення ваги якого і призвело до перестановки в дереві.

Далі в процесі стиснення в початкових даних зустрічається символ "А". При зміні ваги вузла, відповідно даному символу, виникає порушення властивості упорядкованості Н - дерева.

У відповідності до переставлених вузлів необхідно виконати корекцію ваги інших вузлів до кореня дерева Хаффмана. Структура дерева Хаффмана змінюється після того, як в пакеті що підлягає стисненню зустрічається в підряд ще символ "А" два рази.

В результаті даної обробки символ "А" із розділу рідких символів перейшов в групу досить часто зустрічаємих символів, і відповідно змінилась довжина його коду - з трьох до двох.

Необхідно також відмітити, відносно процедури оновлення дерева - в описаному вигляді вона використовується тільки в тому випадку, коли прочитаний із вхідного пакету символ вже існує в Н - дереві. В протилежному випадку (прочитаний символ зустрічається в вхідних даних вперше) дерево Хаффмана потрібно перебудувати.

Розглянемо особливості пов'язані з практичною реалізацією цього методу. Можливі проблеми пов'язані з тим що, при стисненні вхідного потоку іде поступове збільшення ваги вузлів Н - дерева. Неважко побачити, що максимальну вагу в дереві має корінь. Тому перше, що потрібно врахувати, це розрядність яка відведена під вагу кореня. Також важливо врахувати кількість розрядів відведених під двійковий код символу. Це правило особливо треба прийняти до уваги, коли частота появи символів у вхідних комбінаціях створюють послідовності Фібоначі (послідовність чисел натурального ряду, в котрій перші два елементи дорівнюють одиниці, а інші елементи, починаючи з третього, дорівнюють сумі двох попередніх). Якщо рахувати, що максимальна довжина коду 16 біт, то вага коріню, рівна 4181, може призвести до переповнення.

Вирішити проблему переповнення допомагає масштабування Н - дерева. При досягненні коренем своєї максимально допустимої ваги, необхідно "змасштабувати" вагу всіх вузлів дерева. Звичайно це реалізується шляхом ділення ваги вузлів, відповідних символів, котрі зустрічаються в вхідному пакеті, на деяке ціле число (досить часто виконується ділення на 2, так як така операція досить просто реалізується звичайним арифметичним здвигом) і подальшим перерахуванням ваги всіх інших вузлів. Але, як використовується цілочислене ділення, то в загальному випадку можлива втрата дробових частин частки від ділення, тому після процедури масштабування виконують перебудову дерева Хаффмана (в ряді випадків форма дерева може значно змінитися).

Масштабування Н - дерева допомагає не тільки уникнути переповнення, воно також дозволяє збільшити ступінь стиснення вхідних даних, якщо застосовується через деякі інтервали, так як дозволяє врахувати зміниння імовірності появи символів в залежності від положення в початковому пакеті. Але, нажаль, не існує правил вибору оптимального моменту масштабування для досягнення максимально можливого ступню стиснення.

4.5.2. Алгоритм стиснення Віттера

На початку 70 - х років були розроблені однопрохідні методи стиснення інформації, основані на класичній процедурі кодування Хаффмена. Всі ці методи незначно різнилися один від одного і сутність їх полягає в тому, що передавач будує дерево Хаффмана в темпі отримання даних від джерела тобто "на літу". В процесі кодування проходить навчання "навчання" кодера на статистичних характеристиках джерела повідомлень, в ході якого вираховуються оцінки початкових імовірностей повідомлення і виконується відповідна модифікація кодового дерева Хаффмена. У зв'язку з неперервним змінюванням кодового дерева цей процес отримав назву динамічного кодування Хаффмена. Очевидно, що для вірного відновлення стиснутих даних, декодер також повинен безперервно "навчатись" поряд з кодером, виконуючи синхронне змінювання кодової таблиці на прийомній стороні. Для забезпечення синхронності процесів кодування і декодування кодер видає символ в нестиснутому вигляді, якщо він вперше з'явився на виході джерела, і відмічає його на кодовому дереві. При повторній появі символу на виході кодера він передається нерівномірною кодовою комбінацією, що визначається позицією символу на поточному кодовому дереві. Кодер коректує дерево Хаффмена збільшенням частоти передачі символів, котрі вже введені в дерево, або нарощую дерево, добавляючи в нього нові вузли.

Однією з головних умов, котра повинна виконуватись при модифікації кодового дерева, є збереження властивостей хаффменовського дерева. При статистичному кодуванні символи розташовуються в списку в не зростаючому порядку ваги (імовірностей). Потім виконується об'єднання двох вузлів найменшої ваги W_i, W_j і заміна їх внутрішнім вузлом з вагою, рівною сумі початкової ваги W_i+W_j . Знову створений вузол розташовується в списку таким чином, щоб не руйнувався порядок розташування вузлів по вазі. Цей процес повторюється до тих пір, поки в списку не зостанеться один, так званий коріньовий вузол.

Вперше алгоритм синтезу динамічного коду Хаффмена був запропонований Н. Феллером в 1973 році, а потім модифікований Р. Галлагером і Д. Книтом. В зв'язку з цим він отримав назву "Алгоритм FGK".

Подальше удосконалення алгоритму динамічного кодування даних нерівномірними кодами FGK було запропоновано Д. Віттером. Цей алгоритм отримав назву алгоритму V. При пошуку шляхів оптимізації процедури кодування даних кодом Хаффмена автор виходив з того, що в новому алгоритмі число обмінів вузлів в процесі модифікації кодового дерева повинно обмежуватись деяким малим числом (в найкращому випадку одиницею), а динамічне хаффменовське дерево повинно будуватись таким чином, щоб мінімізувати не тільки сумарну довжину загального шляху $\sum W_j l_j$, але і величини $\sum l_j, \max \{l_j\}$. Мінімізація висоти дерева $h = \max \{l_j\}$ дозволяє зарадити створенню довгих кодових комбінацій при кодуванні чергового символу в повідомленні. Віттеру в значній ступені вдалося вирішити поставлену задачу. Розроблений ним алгоритм має в порівнянні з алгоритмом FGK дві переваги:

Кількість обмінів вузлами, при яких черговий вузол переміщається вгору по кодовому дереву в процесі його модифікації, обмежується одиницею. В алгоритмі FGK верхня границя кількості обмінів складає $l_j/2$, де l_j - довжина кодового слова для $Z_j(k+1)$ - го символу до початку процедури модифікації.

Алгоритм V мінімізує довжину зовнішнього шляху дерева l_j і гарантує дерево мінімальної висоти $h = \max \{l_j\}$ при умові мінімізації сумарної довжини зовнішнього шляху дерева $\sum W_j l_j$.

Сутність удосконалень алгоритму V полягає в введенні нової системи нумерації вузлів кодового дерева, яка отримала назву неявної нумерації (Implicit numbering). При неявній нумерації вузли хаффмановського дерева нумеруються в порядку збільшення за рівнями зліва на право, знизу вгору, тобто вузли більш низького рівня мають номери менші, чим вузли наступного рівня. Важливою особливістю неявної нумерації є дотримання необхідної умови побудови дерева, котра формулюється таким чином:

для кожної ваги

W всі зовнішні вузли (листя) дерева з вагою W повинні стояти попереду всіх внутрішніх вузлів ваги W .

Неважко бачити, що ця умова є одною з визначальних особливостей неявної нумерації порівняно з іншими алгоритмами.

Нумерація вузлів, що проводиться відповідно алгоритму FGK, не завжди відповідає неявній нумерації.

Однією з визначних особливостей алгоритму V є введення поняття блока еквівалентних вузлів. При цьому вузли x і y еквівалентні, якщо вони мають однакову вагу і обидва є внутрішніми або зовнішніми. Вузол блоку, маючий (при неявній нумерації) самий високий номер, має назву лідера блоку. Блоки упорядковуються по зростанню ваги, причому, блок листя вагою W повинен передувати блоку внутрішніх вузлів тієї ж ваги.

Особливістю алгоритму V також є спосіб модифікації дерева після отримання чергового символу. Головною операцією алгоритму по підтримці умови неявної нумерації є ковзання і прирощення (Slide And Increment). Суть цієї операції полягає в тому, що вузол, об'явлений черговим обмінюється з лідером свого блоку і потім ковзає в напрямку кореню дерева по сусідньому блоку, котрий безпосередньо прилягає до блоку чергового вузла. Ковзання продовжується до тих пір, поки черговий вузол не пройде весь блок і буде встановлено в голову цього блоку. Потім виконується прирощення ваги чергового вузла і новим черговим вузлом починається батько старого чергового вузла. Операція ковзання з прирощенням

продовжується до досягнення кореня дерева. При цьому вибір батьківського вузла залеже від того, чи є черговий вузол листом, або внутрішнім вузлом. Якщо черговий вузол був листом, тоновим черговим вузлом назначається його батьківський вузол, з котрим був пов'язаний черговий вузол до початку ковзання.

4.5.3. Алгоритм стиснення за допомогою арифметичного кодування

Арифметичне кодування є методом, який дозволяє стиснути послідовність вхідного алфавіту без втрат при умові, що відомий розподіл частот цих символів. Концепція методу була розроблена ще в роботах Еліаса в 60-х роках. В подальшому цей метод був істотно розвинутий і значно вдосконалений.

Арифметичне кодування є оптимальним, досягаючи теоретичної межі стиснення - ентропії вхідного потоку [4].

Текст, стиснутий арифметичним кодером, розглядається як деяка двійкова дріб із інтервалу $[0, 1)$. Результат стиснення можливо представити як послідовність двійкових цифр із цього дробу.

Ідея методу полягає в наступному: початковий текст розглядається як запис цього дробу, де кожний вхідний символ є цифрою з вагою, пропорційною імовірності його появи.

Нехай, алфавіт має два символи: a і b з імовірностями відповідно $3/4$ і $1/4$. Алгоритм стиснення Хаффмена не може зменшити об'єм слова в даному алфавіті.

Розглянемо інтервал $[0, 1)$. Розіб'єм його на частини, довжина котрих пропорційна імовірності елементів. В нашому випадку це $[0, 3/4)$ і $[3/4, 1)$. Суть алгоритму полягає в наступному: кожному символу в вхідному алгоритмі відповідає деякий підінтервал із $[0, 1)$. Пустому слову відповідає весь інтервал $[0, 1)$. Після отримання кожного наступного символу арифметичний кодер зменшує інтервал, вибираючи ту його частину, котра відповідає знову надійшовшому символу. Кодом ланцюжку є інтервал, виділений після всіх символів з ланцюжку, точніше кажучи, двійковий запис координати довільної точки із цього інтервалу.

Таким чином, довжина отриманого інтервалу пропорційна імовірності появи кодуемого ланцюжку.

Виконаємо наведений алгоритм для ланцюжку $aaba$, який представлено в таблиці 4.10

В якості коду можливо взяти довільне число із інтервалу, отриманого на кроці 4, наприклад 0.1.

Таблиця 4.10. Особливості роботи алгоритму арифметичного стиснення.

Крок	Ланцюжок що підлягає розгляду	Інтервал
0	""	$[0, 1)=[0, 1)$
1	a	$[0, 3/4)=(0, 0.11)$
2	aa	$[0, 9/16)=[0, 0.1001)$
3	aab	$[27/64, 36/64)=[0.011011, 0.100100)$
4	aaba	$[108/256, 135/256)=[0.01101100, 0.10000111)$

Арифметичний декодер працює синхронно з кодером: почав з інтервалу $[0, 1)$, він послідовно визначає символи вхідного ланцюжку. В особливості, в нашому випадку він спочатку розділить (пропорційно частотам символів) інтервал $[0, 1)$ на $[0, 0.11)$ і $[0.11, 1)$. Так як, число 0.1 (переданий кодером код ланцюжку $aaba$) знаходиться в першому з них, можливо отримати перший символ a . Потім ділимо перший підінтервал $[0, 0.11)$ на $[0, 0.1001)$ і $[0.1001, 0.1100)$ пропорційно частотам символів). Знову вибираємо перший, так як $0 < 0.1 < 0.1001$. Продовжуючи цей процес, ми однозначно декодуємо всі чотири символи. Для того, щоб декодер міг визначити

кінець ланцюжка, ми можемо передавати її довжину окремо, або додати до алфавіту додатковий символ "кінець ланцюжка".

При реалізації цього методу виникають дві проблеми: по-перше, необхідна арифметика з нескінченною точністю, і по-друге, результат кодування стає відомим лише при закінченні вхідного потоку.

4.5.4. Алгоритм стиснення LZ

Всі розглянуті вище методи і моделі кодування розглядали в якості вхідних даних ланцюжки символів (тексту) в деякому кінцевому алфавіті. При цьому остається відкритим питання про зв'язок цього вхідного алфавіту кодера з даними, що підлягають стисненню (представленими також у вигляді ланцюжків в (іншому) алфавіті, звичайно маючим 256 символів - біт).

В найпростішому випадку можливо використовувати в якості вхідного алфавіту кодеру саме ці символи (байти) вхідного потоку. Ступінь стиснення при цьому відносно невелика - порядку 50% для текстової інформації.

Значно більшого ступеня стиснення можливо досягти при виділенні із вхідного потоку ланцюжків що повторюються і кодування звернень на ці ланцюжки.

Метод, про який буде йти мова, належить Лемпелю і Зіву, звичайно, його називають LZ77 compression (по року публікації методу стиснення).

В його основі лежить наступна концепція: запаковник постійно зберігає деяку кількість останніх оброблених символів в деякому буфері (який також має назву словника що ковзає - sliding dictionary). Назва - словник що ковзає, обумовлена тим, що довжина його постійна; кожний раз, коли компресор кодує наступний ланцюжок, він дописує його в кінець словника і "обрізає" відповідну кількість символів на початку буфера. По мірі обробки вхідного потоку знову надійшовші символи потрапляють в кінець буфера, здвигаючи попередні символи і витискаючи самі старі. Розміри цього буфера можуть бути різні, вони змінюються в різних реалізаціях. Є методи які використовують буфер розміром 4 кбайт, а є моделі маючі значно більший буфер, наприклад 32 кбайт.

Алгоритм виділяє (шляхом пошуку у словнику) самий довгий початковий підрядок вхідного потоку, співпадаючий з одним з підрядків в словнику, і видає на вихід пару (length, distance), де length - довжина знайденого в словнику підрядку, distance - відстань від нього до вхідного підрядку (тобто, фактично індекс підрядку в буфері, віднятий з його розміру). В випадку, якщо такий підрядок не знайдено, в вхідний потік просто копіюється черговий символ вхідного потоку.

В першій версії алгоритму пропонувалося використовувати найпростіший пошук по всьому словнику. Час стиснення, при такій реалізації, був пропорційний добутку довжини вхідного потоку на розмір буфера, що не придатно для практичного використання. Однак, в подальшому було запропоновано використовувати двійкове дерево для швидкого пошуку в словнику, що дозволили на порядок збільшити швидкість роботи.

Таким чином, алгоритм Лемпеля - Зіва перетворює один потік початкових символів в два паралельних потоки length і distance. Очевидно, що ці потоки є потоками символів в нових алфавітах L і D, і до них можливо застосувати один з розглянутих вище методів (кодування Хаффмена, арифметичне кодування). Таким чином, ми підходимо до схеми двохступінчатого кодування, найбільш ефективного з практичного використовуємих в теперішній час. При реалізації цього методу необхідно досягнути узгодженого виводу обох потоків в один. Ця проблема звичайно розв'язується шляхом почергового запису кодів символів із обох потоків.

4.5.5. Алгоритм стиснення методом LZW

Незалежно від мови і характеру текстових повідомлень у різноманітних його частинах можна виявити групи символів, що неодноразово зустрічаються в тексті. Тому з метою скорочення надмірності повідомлення доцільно замість передачі груп символів, що зустрічаються повторно, робити посилання на аналогічні групи, передані в попередній частині повідомлення. Ця ідея лежить в основі методу компресії даних, запропонованого А. Лемпелем і Я. Зивом [7]. Метод Лемпеля-Зива відноситься до однопрохідних адаптивних методів стиснення послідовних даних, що не потребує апріорних знань статистичних характеристик джерела повідомлень. Алгоритм складається з правила для синтаксичного аналізу рядків повідомлення яке підлягає стисненню, що складає із символів кінцевого алфавіту, і схеми кодування. У процесі аналізу з рядків повідомлення виділяються підрядки перемінної довжини. При кодуванні цим підрядкам ставляться у відповідність кодові комбінації, що складаються з фіксованого числа елементів. Потім здійснюється пошук ідентичних підрядків у попередньому рядку. Після відшукування співпадаючих підрядків вибирається підрядок максимальної довжини і проводиться його заміна в поточному рядку повідомлення відповідною кодовою комбінацією.

Найбільше поширеним методом динамічного стиснення інформації є метод LZW. Цей метод компресії даних запропонований Т. Уелчем у 1984 р. Він одержав назву "метод LZW", тому що є подальшим розвитком методу LZ88. При кодуванні LZW -методом використовується таблиця рядків, що складає як з одиночних символів, так і з деяких буквсполучень. Причому кожному рядку відповідає своя двійкова комбінація (кодове слово).

Процедура динамічного кодування LZW -методом полягає в наступному. Відбувається початкове заповнення LZW - таблиці в неї вносяться рядки, що відображають статистику повідомлень на визначеній мові. При відсутності такої статистики таблиця містить тільки рядки, що складаються з одиночних символів. Потім, у міру надходження даних від джерела інформації, формуються рядки, що складаються з декількох символів. Тому що таблиця має обмежений розмір, то рядки, що зустрічаються в тексті рідко, виключаються, а на їхнє місце вносяться рядки, що мають велику частоту появи. Таким чином, у процесі накопичення статистики про повідомлення що стискається відбувається динамічна перебудова таблиці кодування й адаптація її до характеру переданих даних. Компресія даних починається з ініціалізації таблиці, при якій в неї вносяться рядки, що складаються з одиночних символів. Потім надходить перший вхідний символ, аналізований як префікс деякого рядка PREFIX. Після цього вводиться такий символ CHARACTER і утвориться розширений рядок шляхом об'єднання префікса й одиночного символу PREFIX + CHARACTER. Далі здійснюється зіставлення знову утвореного рядка з рядками, що існують у таблиці кодування. Якщо рядок PREFIX + CHARACTER є в таблиці, то вона стає новим префіксом, тобто $PREFIX = PREFIX + CHARACTER$ і вводиться такий символ CHARACTER і процедура зіставлення рядка в таблиці повторюється знову. У протилежному випадку, якщо послідовності PREFIX + CHARACTER у таблиці рядків немає, то на вихід кодера виводиться кодова комбінація, що відповідає рядку PREFIX, у таблицю вноситься додатковий рядок PREFIX + CHARACTER, а символ CHARACTER стає новим префіксом. Якщо вхідна послідовність не вичерпана, те процедура утворення рядків і їх співставлення повторюється.

Розробка блок-схем кодування та декодування кодом LZW

LZW-алгоритм розроблений Т. Велчем на основі алгоритму LZ78 Лемпеля та Зива. Він усуває в потоці даних надмірність, викликану появою в різних місцях потоку однакових ланцюгів символів. Ядром LZW-алгоритму є словник, що містить всі символи вхідного алфавіту й деякі рядки із цих символів. До початку роботи алгоритму в словник входять елементи в якості рядків, утримуючи кожний із символів вхідного алфавіту.

У ході роботи алгоритму ланцюги, що зустрічаються у вхідному потоці символів, заносяться в словник. При цьому у вихідний потік заноситься номер елемента словника, що містить рядок максимальної довжини, співпадаючу з рядком вхідного потоку. Словник

організований як сукупність двійкових дерев, і містить по одному двійковому дереву на кожен символ вхідного алфавіту.

Алгоритм LZW побудований навколо так званої таблиці фраз (словника) [3], що відображає рядки символів повідомлення, яке стискається в коди фіксованої довжини і дорівнює 12 біт. Таблиця має властивість попереджування, тобто для кожної фрази словника, що складається з деякої фрази w і символу K , фраза W теж міститься в словнику. Алгоритм роботи кодера LZW наведено на рис. 5.1 [3].

Описаний алгоритм кодера через свою простоту може бути ефективно використаний.

Декодер LZW має використовувати той же словник, що і кодер, будуючи його за аналогічними правилами при відновленні стиснутих даних. Кожен код, що зчитується, розбирається за допомогою словника на попередню фразу W і символ K . Потім рекурсія продовжується для попередньої фрази W доти, поки вона не виявиться кодом одного символу. При цьому завершується декомпресія цього коду. Відновлення словника відбувається для кожного коду, який декодується, крім першого. Після завершення декодування коду його останній символ, з'єднаний з попередньою фразою, додається в словник. Нова фраза одержує те ж значення коду (позицію в словнику), що привласнив їй кодер. Внаслідок такого процесу, крок за кроком декодер, алгоритм роботи якого наведено на рис. 4.7 [3], відновлює той словник, що побудував кодер. Звичайне декодування LZW набагато швидше, ніж процес кодування.

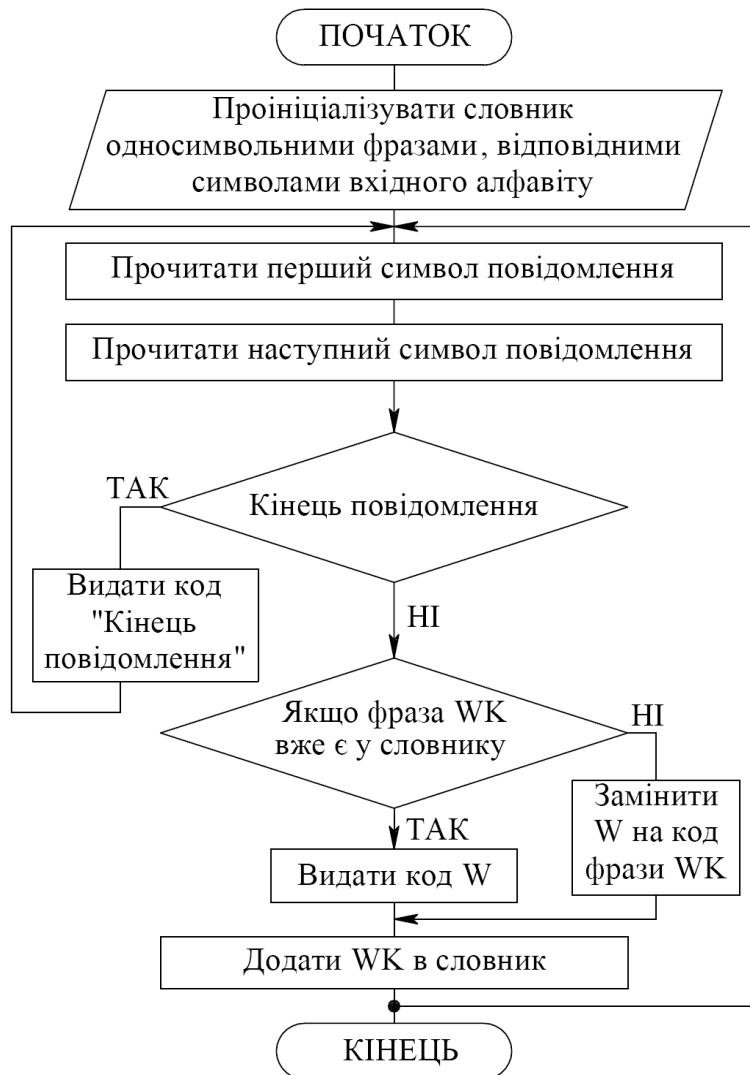


Рис. 4.7. Алгоритм роботи кодера LZW



Рис. 4.8. Алгоритм роботи декодера LZW

4.5.6. Алгоритм стиснення методом LZH

Метод стиснення LZH заснований на використанні двох методів стиснення даних: словниковий (LZ) і статистичний (Haffman) метод.

На початку виконання алгоритму стиснення відбувається ініціалізація словника. Словник логічно являє собою сукупність абстрактних структур даних, що містять набір деревоподібних структур, у котрих кожний корінь відповідає визначеному знаку алфавіту. При 8-розрядному форматі символу кількість таких дерев дорівнює 256. Деревоподібні структури являють собою набір відомих рядків, що починаються одним визначеним символом, а кожний вузол дерева надає один рядок із цього набору.

При виконанні компресії циклічно формується новий рядок шляхом додавання чергового одиночного символу до існуючого рядка, що знаходиться в таблиці статей словника, що виражається в додаванні нового вузла до кодового дерева.

Далі відбувається співставлення рядка, при якому послідовність символів із потоку даних співставлення зі статтею словника. Якщо рядок відповідає статті словника і не є створеним під час останнього виклику процедури зіставлення рядка, то вводиться наступний символ і добавляється до рядка. Потім цей крок повторюється. Якщо рядок не відповідає статті словника, або відповідає статті, створеної під час останнього виклику процедури зіставлення рядків, то останній символ убирається, а кодове слово відповідній статті словника видається на вихід кодера. Отриманий укорочений рядок представляє, таким чином, самий довгий зіставлений рядок, а останній

(відкинутий) символ є не зіставленим символом, із якого починається таке зіставлення рядків що надходять від джерела даних.

4.5.7. Матричний метод стиснення

З матричних методів стиснення найбільш придатним для використання в системах управління є адаптивно-матричний [22].

Для роботи алгоритму необхідно задати розмір блока даних. Для зручності і наочності вхідний потік розбивається на рядки символами “переведення рядка - повернення каретки”. У реальних умовах довжина рядка матриці визначається довжиною запису даних, що передаються.

Інформація зчитується блоками з вхідного файлу, одночасно робиться її аналіз і пошук елементів, що повторюються, у сусідніх рядках. Якщо в двох сусідніх рядках знайдені однакові ланцюжки символів, алгоритм здійснює перегляд наступних рядків, визначаючи кількість рядків матриці. Розміри матриці і номери рядків для кожного рядка матриці записуються в допоміжний масив. Якщо розміри матриці задовольняють визначеній умові, то вона позначається як та, що підлягає стисненню.

При записі інформації у вихідний потік для кожної виділеної матриці передається її перший рядок, перед яким ідуть ознака стиснення і розміри матриці. Наступні рядки матриці опускаються.

Даний метод не забезпечує найбільш оптимального виділення матриць у вхідному потоці (наприклад, якщо дві матриці перекриваються, то виділена буде перша що зустрілася, навіть якщо вона менша за розміром). Однак цей метод має високу швидкість стиснення, оскільки виділення матриць відбувається одночасно з прийомом даних.

4.5.8 Каскадний метод стиснення

Каскадний метод поєднує у собі стиснення інформації двома методами. Його відмінність від інших методів поданих у роботі в тому, що дані спочатку стискаються одним методом, а потім отриманий стиснутий масив обробляється іншим методом. Розгортання інформації робиться в зворотному порядку [61].

При реалізації каскадного методу стиснення (рис. 4.11) може виникнути проблема неоднозначності трактування стиснутого масиву даних. Це може відбутися в тому випадку, якщо алгоритм стиснення першого каскаду в результаті роботи згенерує масив даних, що містить керуючі символи алгоритму другого каскаду.

При розгортанні (декомпресії) такого масиву процедура другого каскаду інтерпретує ці символи як ознаки стиснення, і інформація буде перекручена.

Для вирішення цієї проблеми у роботі передбачається чіткий поділ способу вказівки ознаки стиснення для кожного каскаду, що не дозволяє використання однакових символів.

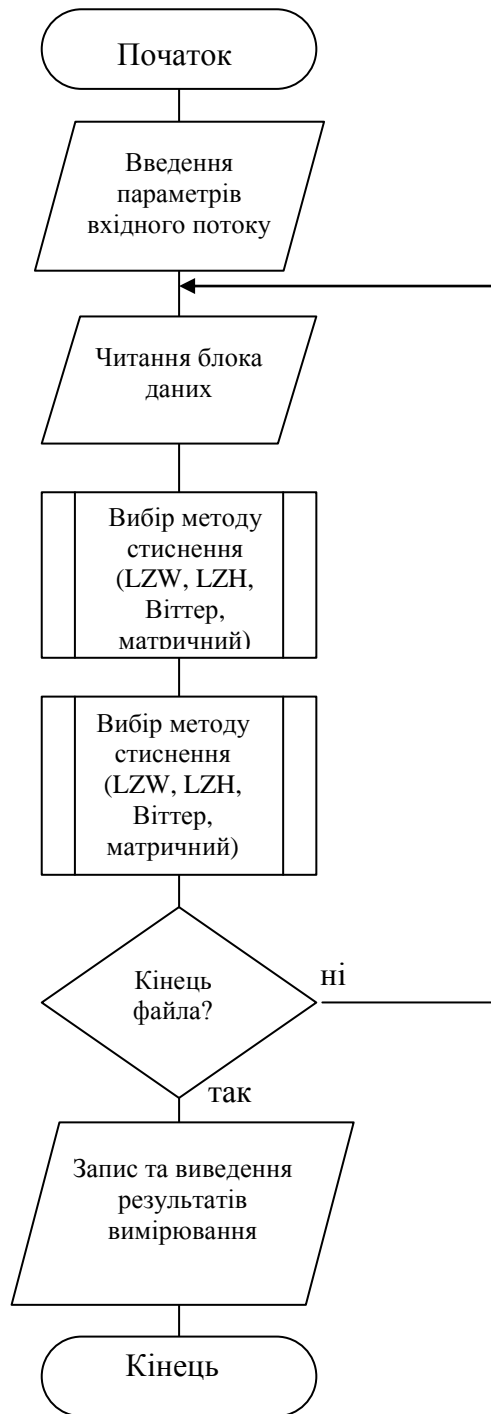


Рис. 4.9. Алгоритм каскадного методу стиснення

4.6. Кодо-матричне стиснення

На початку вихідна матриця переводиться в задану систему числення. Після цього здійснюється оптимізація розташування прямокутних областей, в яких повторюються рядки. Після виділення вибраних областей здійснюється стиснення по наступному алгоритму (рис. 4.10)

У заданій матриці шукаємо найбільшу під матрицю з тих, в яких повторюється один рядок декілька разів. Робимо це таким чином:

1. Спочатку задаємося вікном, верхній лівий кут якого співпадає з верхнім лівим кутом матриці і яке повністю займає перші два рядки матриці.

2. Якщо рядки вікна однакові, то займаємо поточне вікно як «краще вікно»;

3. Далі дивимося на рядок розташований нижче:

Якщо він ідентичний з попереднім, то збільшуємо вікно на одну позицію вниз, порівнюємо розмір вікна що вийшло з (ширина*(висота-1)) з розміром поточного «кращого вікна»;

Якщо отримане вікно більше, то його зберігають як «краще вікно». Таким чином продовжуємо нарощувати вікно вниз, доки черговий рядок не буде відрізнитись від попередніх, або доки не дійдемо кінця матриці.

4. Коли черговий рядок не співпадає з попереднім (або у п. 2 не співпали два рядки), то зменшуємо вікно, шляхом здвигу правої границі на одну позицію (колонку) вліво, до тих пір, поки не отримаємо вікно, в якому всі рядки однакові, або поки вікно не виродиться в стовпець.

5. Далі порівнюємо поточне «краще вікно» з отриманим, якщо останнє більше, то його зберігаємо як «краще».

6. Переходимо до пункту 3, якщо не досягнутий нижній рядок матриці

7. У результаті описаних дій отримаємо найбільшу під матрицю з тих, які своїм верхнім кутом виходять з лівого верхнього кута повної матриці і складаються з однакових рядків («краще вікно»). Збережемо цю підматрицю.

8. Далі робимо все вищезазначене (з п. 1) для наступної позиції: тобто тепер всі вікна будуть виходити з верхньої другої зліва позиції.

Отже для кожної позиції матриці ми знайдемо найбільшу під матрицю, верхнім лівим кутом якої є дана позиція, а рядки однакові.

9. З цього набору підматриць ми знайдемо найбільшу підматрицю, яку поміщаємо у список підматриць що стискаються.

Виконуємо всі пункти знову, але тепер кожне вікно перевіряємо: чи не накладається (не перехрещується) воно на під матриці зі списку матриць що стискаються. Якщо «так», то пропускаємо таке вікно. В результаті такого виконання дій всіх вищенаведених пунктів отримаємо наступну підматрицю, що не «накладається» на матриці зі списку матриць що стискають і яка дописується в цей список. Виконуємо до тих пір, поки список не перестане поповнюватись.

Переглядаємо повну(вихідну) матрицю зліва на право і згори до низу:

якщо поточна позиція не належить ні одній підматриці зі списку що стискаються, то її просто перепишемо;

якщо «натикаємося» на під матрицю зі списку, то записуємо символ «x», рядок підматриці, символ «u», число рядків у підматриці;

далі всі елементи матриці, що належать підматриці з розгляду знімаються.

І так до кінця вихідної матриці.

Розпаковка здійснюється у тій самій системі, в якій стискалась вихідна матриця. Після цього матриця переводиться в вихідну систему.

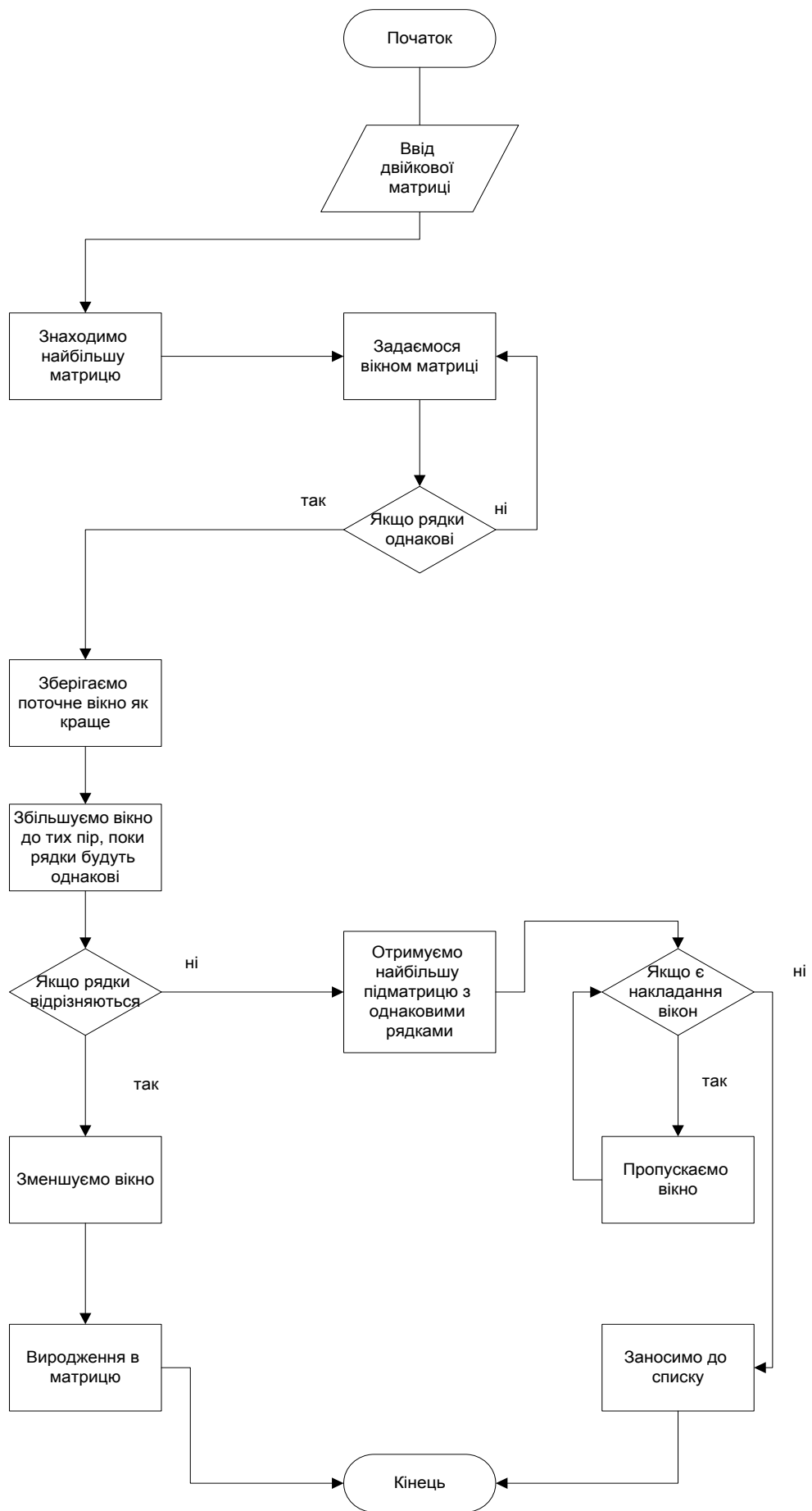


Рис. 4.10 Алгоритм кодо-матричного стиснення

4.7. Алгоритм повздовжнього та матричного стиснення

Алгоритм повздовжнього стиснення один із алгоритмів який виконує стиснення інформації завдяки заміні ланцюжків що повторюються, на більш короткі кодові комбінації.

В його основі лежить пошук в потоці інформації ланцюжків з однакових елементів і заміна їх на більш короткі, що знаходяться в алфавіті алгоритму.

Розглянемо вищенаведений алгоритм стиснення інформації не прикладі послідовності бінарних даних: 1000111000001100100001. Якщо проаналізувати таку послідовність, то можливо знайти ланцюжки символів що повторюються 1000111000001100100001. Нехай, два повтори підряд одного елементу буде еквівалентно символу - А, три повтори - В, чотири - С. Таким чином, з урахуванням вищенаведених означень, дані можливо представити у вигляді рядку виду: 10A1A0C11001B1.

З приведених означень видно, що ланцюжки, що мають два однакові елементи стиснути неможливо, бо стиснута таким чином послідовність не змінить об'єм інформації.

Розвертання виконується шляхом зчитування кожного символу послідовності і порівняння його з елементами словника. Якщо такий елемент маєтья, то елемент що знаходиться перед ним повторюється стільки разів, скільки це передбачено словником. (у нашому випадку словник містить таку інформацію: А -2, В -3, С - 4).

Схожим на метод повздовжнього стиснення є матричне стиснення інформації. В ньому виконується пошук комбінацій елементів що повторюються одна за одною, і заміна їх, як і в повздовжньому алгоритмі, на більш короткі комбінації які містить словник.

4.8. Комбінований метод підвищення ефективності передачі інформації

Як відомо, основною умовою при роботі існуючих систем управління (СУ) є відповідність їх якісних показників міжнародним стандартам, головним з яких є висока достовірність ($10^{-9} - 10^{-12}$) та забезпечення роботи усіх ланок СУ в режимі реального часу.

Висока достовірність передачі інформації може бути досягнута застосуванням завадостійкого кодування, тобто кодів, що мають високу здатність виявляти та виправляти помилки. При цьому рекомендується використовувати коди з достатньо великою збитковістю ($\geq 0,5$). Проте використання кодів з високою збитковістю приводить до збільшення інформаційного потоку в 1,5 та більше раз (див. табл. 3.3), що неминуче призводить до зниження швидкості передачі інформації.

Тому, опираючись на матеріали викладені вище, для підвищення ефективності передачі інформації по каналам зв'язку, з урахуванням джерел інформації, що використовуються, а також характеру та інтенсивності помилок в каналах зв'язку, пропонується метод комбінованого використання стиснення та завадостійкого кодування при передачі інформації. Даний метод дозволяє забезпечити високу достовірність передачі інформації без зниження її швидкості доставки, завдяки попередньому стисненню на передаючій стороні.

Якщо порівняти результати, що отримані при стисненні інформації з даними по збільшенню інформаційних масивів при застосуванні завадостійких кодів, що забезпечують потрібну достовірність передачі, то стане ясно, що комплексне використання стиснення та кодування інформації веде до підвищення ефективності використання каналів зв'язку.

Таким чином, одночасне використання способів стиснення, що пропонуються, та високоефективних способів кодування, дозволяє забезпечити вимоги, що пред'являються до каналів зв'язку, як по достовірності передачі по ним інформації, так і по швидкості її доставки, а також це дозволяє підвищити ефективність передачі інформації по каналам.

РОЗДІЛ 5

ВИКОРИСТАННЯ ШУМОПОДІБНИХ СИГНАЛІВ У СИСТЕМАХ ЗВ'ЯЗКУ

5.1. Основні поняття і визначення

Означення шумоподібних сигналів широкосмугових систем зв'язку. Шумоподібними сигналами (ШПС) називають такі сигнали, добуток ширини спектра F яких на тривалість T набагато більший за одиницю. Цей добуток називається *базою сигналу* і позначається B :

$$B = 2FT. \quad (5.1)$$

Системи зв'язку з шумоподібними сигналами (ШПС) відомі понад 40 років. За цей час їх переваги стали очевидними, а багато з недоліків усунуті. Процес розширення галузей використання систем зв'язку з ШПС незворотний, і в найближчому майбутньому увага до них посилюватиметься.

Основи теорії систем зв'язку з ШПС заклали В. О. Котельников [24] і К.Шеннон [49], а основи кодового поділу – Д. В. Агєєв. Системи зв'язку з ШПС посідають особливе місце серед інших систем зв'язку, що зумовлено їхніми властивостями: вони забезпечують високу заводозахищеність систем зв'язку; дозволяють організувати одночасну роботу багатьох абонентів у загальній смузі частот за асинхронно-адресним принципом роботи системи зв'язку, заснованим на кодовому поділі абонентів; дають змогу успішно долати багатоприменеве поширення радіохвиль шляхом поділу променів; забезпечують сумісність передавання інформації з вимірюванням параметрів руху об'єкта в системах рухомого зв'язку; забезпечують електромагнітну сумісність (ЕМС) широкосмугових систем з вузькосмуговими системами радіозв'язку і радіомовлення, системами телебачення, краще використання спектра частот на обмеженій території в порівнянні з вузькосмуговими системами зв'язку. Ці властивості систем зв'язку з ШПС відомі давно, але, оскільки потужності завод раніше були відносно невисокі, а елементна база не дозволяла реалізовувати пристрої формування й обробки в прийнятних габаритах, то довгий час системи зв'язку з ШПС широкого розвитку не набували. На цей час ситуація різко змінилася. Потужність завод на вході приймача може на кілька порядків перевищувати потужність корисного сигналу. Для забезпечення високої заводозахищеності необхідно використовувати системи з ШПС з надвеликими базами (десятки – сотні тисяч), а ансамблі (системи) сигналів мають складатися з десятків – сотень таких ШПС. Слід зазначити, що основи теорії ШПС з надвеликими базами сформувалися тільки останнім часом. У свою чергу реалізація пристроїв формування й обробки ШПС стає можливою в найближчому майбутньому завдяки бурхливому розвитку надвеликих інтегральних схем, спеціалізованих мікропроцесорів, приладів з поверхневими акустичними хвилями, приладів із зарядовим зв'язком.

Крім застосування в умовах впливу потужних завод, системи зв'язку з ШПС починають успішно конкурувати з широко розповсюдженими системами зв'язку з частотною модуляцією і частотним поділом каналів, застосовуваними для зв'язку з рухомими об'єктами. Створення малогабаритних пристроїв формування й обробки ШПС із базами $10^2 \dots 10^3$ дозволить широко впровадити ШПС у системи зв'язку і керування рухомими об'єктами у великих містах.

Шумоподібні сигнали, база яких $B \gg 1$, іноді називають *складними* на відміну від *простих* сигналів з базою $B \approx 1$. Оскільки спектр сигналів з обмеженою тривалістю має необмежену довжину, то для визначення ширини спектра використовують різні методи і прийоми. Для реальних ШПС, що складаються зі скінченного числа елементів, завжди можна однозначно визначити F і B .

У системах зв'язку з ШПС ширина спектра ШПС (F) завжди набагато більша за ширину спектра переданого повідомлення. В цифрових системах зв'язку, що передають інформацію у вигляді двійкових символів, тривалість ШПС і швидкість передавання інформації R пов'язані співвідношенням $T = 1/R$. Тому база ШПС

$$B = 2F / R \quad (5.2)$$

характеризує розширення спектра сигналу щодо спектра повідомлення. В аналогових системах зв'язку, верхня частота повідомлення яких дорівнює W , а частота відліку дорівнює $2W$,

$$B = F / (2W). \quad (5.3)$$

Якщо $B \gg 1$, то $F \gg R$ і $F \gg 2W$. Саме тому системи зв'язку з ШПС в іноземній літературі одержали назву *систем зв'язку з розширеним (або розподіленим) спектром*, а у вітчизняній літературі – *широкосмугових систем зв'язку*. Надалі термін “широкосмугові системи зв'язку” (ШСЗ) стосуватиметься тільки систем зв'язку з ШПС.

Завадозахищеність – це здатність системи зв'язку протидіяти впливові потужних завад. Завадозахищеність містить у собі скритність (захист від несанкціонованого доступу) системи зв'язку та її завадостійкість. Чим вищі скритність і завадостійкість, тим вища завадозахищеність системи зв'язку.

Завадостійкість ШСЗ визначається виразом, який пов'язує відношення сигнал–завада на виході приймача (на виході погодженого фільтра або корелятора) q^2 із відношенням сигнал–завада на вході приймача ρ^2 :

$$q^2 = 2B\rho^2, \quad (5.4)$$

де B – база ШПС;

$$\rho^2 = P_c / P_z, \quad (5.5)$$

де P_c, P_z – потужності ШПС і завади.

Часто значення енергії сигналу E та спектральної щільності потужності завади в смузі ШПС N_z є відомими. Тоді замість (2.4) для визначення q^2 користуються формулою

$$q^2 = 2E / N_z; \quad (5.6)$$

$$E = P_c T; \quad (5.7)$$

$$N_z = P_z F. \quad (5.8)$$

Відношення сигнал–завада на виході q^2 визначає робочі характеристики приймання ШПС, а відношення сигнал–завада на вході ρ^2 – енергетику сигналу і завади. Величина q^2 може бути отримана згідно з вимогами до системи (10...30 дБ), навіть якщо $\rho^2 \ll 1$. Для цього достатньо вибрати ШПС із базою B , що задовольняє (5.4). Як видно зі співвідношення (5.4), приймання ШПС погодженим фільтром або корелятором супроводжується посиленням сигналу (або придушенням завади) у $2B$ рази. Саме тому величину

$$K_{\text{ШПС}} = q^2 / \rho^2 \quad (5.9)$$

називають *коефіцієнтом підсилення ШПС при обробці* або просто *посиленням обробки*. Із (5.4), (5.9) випливає, що посилення обробки $K_{\text{ШПС}} = 2B$.

У ШСЗ приймання інформації характеризується відношенням сигнал–завада $h^2 = \frac{q^2}{2}$, або

$$h^2 = B + \rho^2. \quad (5.10)$$

На рис. 5.1, а подано залежності посилення обробки і бази ШПС від відношення сигнал–завада на вході ρ^2 , дБ, при значеннях q^2 (суцільні лінії) і h^2 (штрихові лінії), що дорівнюють 10, 20 і 30 дБ. Ці залежності

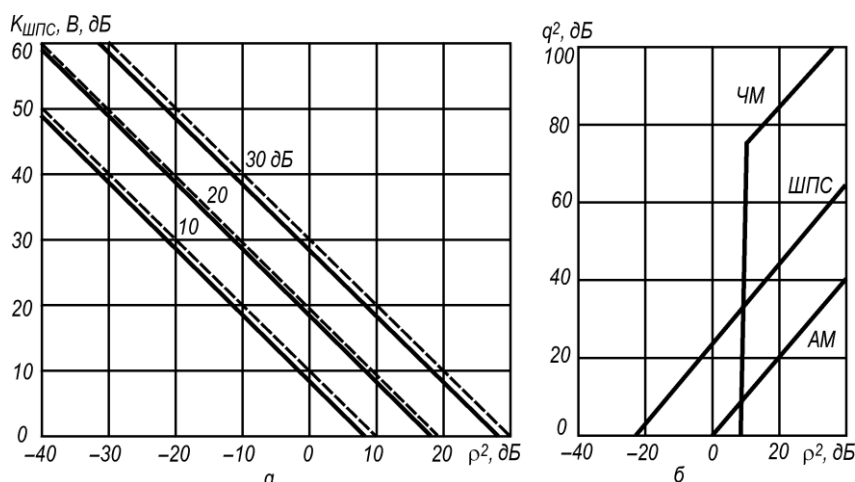


Рис. 5.5. Залежності посилення обробки і бази ШПС від відношення сигнал–завада на вході приймача (а) і завадостійкість систем зв’язку з ШПС (б)

побудовані згідно з формулами (5.4), (5.6). Наприклад, якщо необхідно отримати $h^2 = 20$ дБ, а на вході приймача $\rho^2 = -40$ дБ, то потрібна база має дорівнювати 60 дБ.

Співвідношення (5.4), (5.6) є фундаментальними в теорії систем зв’язку з ШПС. Вони отримані для завад у вигляді білого шуму з рівномірною спектральною щільністю потужності в межах смуги частот, ширина якої дорівнює ширині спектра ШПС. Разом з тим ці співвідношення слушні для широкого кола завад (вузькосмугових, імпульсних, структурних), що й визначає їхнє фундаментальне значення. У загальному випадку посилення обробки ШПС для довільних завад

$$K_{\text{ШПС}} \approx 2B, \quad (5.11)$$

де ступінь наближення залежить як від вигляду завад, так і від бази ШПС. У табл. 2.1 наведені значення посилення обробки для деяких систем зв’язку і навігації.

Наведені в перших чотирьох рядках таблиці параметри відповідають системам зв’язку 1960-х років, в п’ятому – сучасній системі GPS (Global Position System – багатосупутникова радіонавігаційна система).

* ФМ – фазоманіпульований сигнал;

ЧМ – частотно-маніпульований.

АМ). Для порівняння ЧМ і ШПС узяті однакові смуги частот, що відповідає $B = 100$. Завадостійкість системи зв’язку з ШПС розрахована згідно з (5.4), причому покладено, що інформація передається за допомогою широтно-імпульсної модуляції (ШІМ). Система з ЧМ має високу завадостійкість і забезпечує високу якість відтворення інформації за умови, що відношення сигнал–завада на вході вище за граничні значення $\rho^2 = 10 \dots 15$ дБ. Якщо ρ^2 нижче за граничні значення, завадостійкість системи зв’язку з ЧМ різко падає. Система з АМ і еквівалентною базою $B = 1$ працює лише при $\rho^2 > 0$, залежність q^2 від ρ^2 лінійна. Система зв’язку з ШПС забезпечує надійне приймання інформації і при $\rho^2 < 0$. Наприклад, якщо взяти $q^2 = 10$ дБ, то система зв’язку працюватиме при відношенні сигнал–завада на вході -13 дБ, тобто $\rho^2 = 0,05$. Таким чином, одним з основних призначень систем зв’язку з ШПС є забезпечення надійного приймання інформації при впливі потужних завад, коли відношення сигнал–завада на вході приймача ρ^2 може бути набагато меншим за одиницю.

Підкреслимо, що наведені співвідношення строго слушні для завади у вигляді гауссового випадкового процесу з рівномірною спектральною щільністю потужності (білий шум).

Скритність системи зв'язку – це здатність протистояти виявленню і вимірові параметрів. Скритність – поняття широке; оскільки виявлення ШПС і вимір параметрів можливі при різній первісній поінформованості (за апріорної невизначеності) про систему зв'язку, то можна вказати тільки основні співвідношення, що характеризують скритність. У випадку, коли відомо, що в даному діапазоні частот може працювати система зв'язку, але параметри її невідомі, то говорять про *енергетичну* скритність системи зв'язку, тому що її виявлення можливе за допомогою аналізу спектра (енергетичне виявлення). Характеристика виявлення (ймовірності помилкової тривоги і пропуску сигналу) цілком визначається відношенням сигнал–завада на вході приймача-аналізатора $\rho = P_c / P_z$, де завадою є власний шум приймача: $P_z = k_0 T_0 (N_{\text{ш}} - 1) F$; k_0 – стала Больцмана; T_0 – температура навколишнього середовища; $N_{\text{ш}}$ – коефіцієнт шуму приймача. Час виявлення ШПС, за умови $\rho^2 \ll 1$, приблизно визначається співвідношенням

$$T_{\text{в}} \approx aF, \quad (5.12)$$

де розмірна стала a залежить як від шумових властивостей приймача, потужності сигналу на вході, так і від необхідного відношення сигнал–завада на виході q^2 . Отже, чим ширший спектр ШПС, тим більший час виявлення, тим вища енергетична скритність системи зв'язку.

Якщо ШПС системи зв'язку відтворюються приймачем-аналізатором з високою завадозахищеністю, то час аналізу приблизно визначається співвідношенням, аналогічним за виглядом співвідношенню (5.12), але $a = Tb \left[k_0 T_0 \frac{N_{\text{ш}} - 1}{P_c} \right]$; тут T – тривалість сигналу; b – стала величина. Чим ширший спектр ШПС, тим більша база і час аналізу, тим вища параметрична скритність системи зв'язку. Отже, чим ширший спектр ШПС і чим більша його база, тим вища як енергетична, так і параметрична скритність.

Для боротьби з радіорозвідкою у завадозахищених системах зв'язку застосовують також зміну ШПС. Частота зміни ШПС, їхній вибір з деякого ансамблю (системи сигналів) обумовлюються багатьма вимогами до системи зв'язку і не можуть бути однозначними. Проте існує думка, що число сигналів у системі (або об'єм системи сигналів) має набагато перевищувати базу ШПС. Можна припустити, що для завадозахищених систем зв'язку об'єм системи сигналів L визначається статистичним законом

$$L \approx B^m, \quad (5.13)$$

де $m \geq 2$ – деяке число, яке визначає кількість ШПС.

Отже, використання ШПС підвищує завадостійкість і скритність системи зв'язку, тобто її завадозахищеність. ШПС використовують у супутникових і авіаційних системах зв'язку, радіорелейних лініях, у супутникових навігаційних системах. Очевидно, застосування ШПС у завадозахищених системах зв'язку буде розширюватися.

5.2. Кодовий поділ абонентів

Завадозахищені системи зв'язку є спеціальними, тому на ранньому етапі розвитку систем зв'язку з ШПС не передбачалось, що вони знайдуть широке застосування в комерційних системах зв'язку. Однак з розвитком асинхронних адресних систем зв'язку масове впровадження ШПС у системи радіозв'язку стало можливим завдяки кодовому поділу абонентів за допомогою ШПС, що відрізняються за формою.

При великих базах можна побудувати дуже багато різних ШПС. Наприклад, нехай ШПС являє собою фазоманіпульований сигнал, що складається з радіоімпульсів, фази яких 0 або π , а

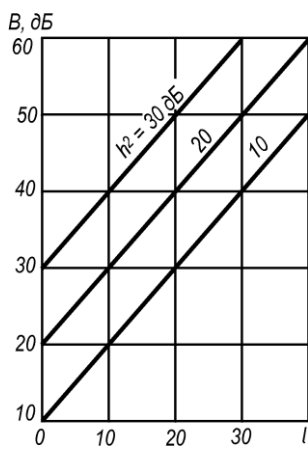


Рис. 5.2.
Завадостійкість ААСЗ

число їх дорівнює B . Можна побудувати множину сигналів (так званий повний код), число яких дорівнює 2^B , і вони між собою відрізнятимуться хоча б в одному імпульсі. Якщо взяти $B = 100$, то маємо $2^{100} \approx 10^{30}$ різних сигналів. З такої множини можна відібрати систему сигналів так,

щоб кожному абонентові в системі зв'язку відповідали свої власні сигнали. У такому разі всі абоненти можуть працювати в загальній смузі частот, а поділ їх можливий за рахунок розбіжностей форми ШПС. Такий поділ абонентів називається *кодовим*. При цьому ШПС є по суті адресою абонента; у цьому випадку принципово відпадає потреба в примусовій часовій синхронізації абонентів. Тому подібні системи зв'язку одержали назву *асинхронних адресних систем зв'язку* (ААСЗ). Вони

засновані на застосуванні ШПС і кодовому поділі абонентів.

В ААСЗ всі абоненти працюють у загальній смузі частот, тому при передаванні інформації ШПС різних абонентів перекриваються за часом і частотою, створюючи взаємні завади. Однак, використовуючи ШПС з великими базами, можна звести рівень взаємних завад до необхідного, щоб забезпечити якість приймання інформації. Якщо припустити, що на вході одного з приймачів системи зв'язку діє l ШПС з однаковими потужностями, то відношення сигнал–завада на виході приймача

$$h^2 = B/l. \quad (5.14)$$

Отже, збільшуючи базу ШПС, завжди можна досягти необхідної якості приймання інформації.

На рис. 5.2 подано залежності бази ШПС від числа активних абонентів, побудовані згідно з (5.14). За графіками можна визначити завадостійкість ААСЗ.

Ефективність ААСЗ. Як випливає з (5.14), підвищення завадостійкості ААСЗ при заданому числі активних абонентів можливе тільки за рахунок збільшення бази ШПС. При заданій швидкості передавання інформації збільшення бази потребує пропорційного розширення спектра ШПС відповідно до (5.2), а отже, виникає питання про ефективність використання радіоспектра в системах зв'язку з ШПС.

Ефективність використання радіоспектра характеризується *питомою щільністю активних абонентів* γ . Ця величина дорівнює числу активних абонентів, що припадає на 1 МГц смуги частот, тобто

$$\gamma = l/F, \quad (5.15)$$

де l – число активних абонентів, що одночасно працюють у смузі частот шириною F . Замінивши l у (5.15) згідно з (5.14), маємо

$$\gamma = T/h^2. \quad (5.16)$$

Часто питома щільність активних абонентів називається просто *ефективністю системи зв'язку*. З (5.16) випливає, що ефективність ААСЗ тим менша, чим більше відношення сигнал–завада на виході приймача. Отже, ААСЗ перспективніші тоді, коли висока якість передавання інформації не потрібна, що характерно для систем масового радіозв'язку. Наприклад, у радіотелефонній системі рухомого зв'язку з частотним поділом каналів максимальна ефективність $\gamma_{\max} = 250$ аб./МГц, тому що мінімальна ширина кожного каналу дорівнює 4 кГц, і на 1 МГц можна розмістити 250 частотних каналів, тобто активних абонентів. Однак для підвищення завадостійкості використовується частотна модуляція (ЧМ), і сусідні частотні канали рознесені на 25кГц. При цьому на 1 МГц можна розмістити 40 частотних каналів, тобто $\gamma_{\text{ЧМ}} = 40$ аб./МГц. Перші системи зв'язку з ШПС (“RACER” та ін.) мали низьку ефективність, з $\gamma_{\text{ШПС}} \approx 7 \dots 9$ аб./МГц. Із (5.16) випливає, що високу ефективність систем зв'язку з ШПС безпосередньо одержати важко. Наприклад, якщо взяти $T = 1/4$ кГц = 250 мс, а $h^2 = 10$, то $\gamma_{\text{ШПС}} = 25$ аб./МГц, тобто ефективність систем зв'язку з ШПС нижча за ефективність систем зв'язку з ЧМ.

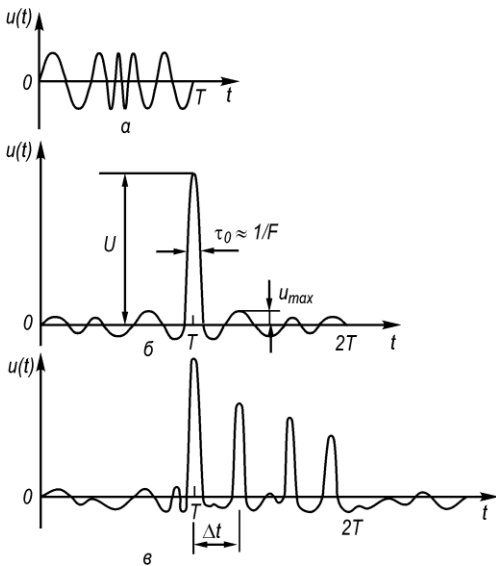


Рис. 5.3. Шумоподібний сигнал (а), автокореляційна функція (б) і поділ променів (в)

В останні роки запропоновано інший принцип побудови радіотелефонних систем рухомого зв'язку. Вся обслуговувана територія розбивається на велике число зон у вигляді стільників. У кожній зоні радіозв'язок ведеться на частотах, спеціально виділених цій зоні. Територіальне рознесення зон з однаковими частотними каналами уможливорює багаторазове використання тих самих частотних каналів. Ці системи зв'язку одержали назву *стільникових систем рухомого зв'язку* (ССРЗ). Приймання сигналів у таких системах принципово супроводжується взаємними завадами так само, як і в ААСЗ. Тому застосування ШПС у ССРЗ є перспективним, оскільки дозволяє успішно долати взаємні завади. Ефективність ССРЗ

$$\gamma_{ССРЗ} \approx 3,63 R_0 / D^2 / F_k,$$

де R_0 – радіус зони обслуговування; D – захисний інтервал; F_k – ширина частотного каналу. Якщо взяти $R_0 = 30$ км,

$D \approx 4,4$ км, а $F_k = 50$ кГц, то $\gamma_{ССРЗ} \approx 3333$ аб./МГц, тобто ефективність ССРЗ є набагато вищою за ефективність звичайних систем рухомого зв'язку. Якщо ширину частотного каналу збільшити до 100...200 кГц, то ефективність ССРЗ дорівнюватиме 1666...833 аб./МГц відповідно, проте все одно набагато перевищуватиме ефективність систем із ЧМ. При цьому можливе застосування ШПС із відносно невеликими базами (25...250), що дає змогу використовувати просту апаратуру формування й обробки ШПС із невисокою стабільністю частоти. У свою чергу, застосування ШПС вирішує проблему адресації великого числа абонентів. Стільникові системи рухомого зв'язку з ШПС дозволяють забезпечити зв'язком від 60 до 340 тисяч рухомих абонентів у великих містах. Крім того, в таких системах можна поєднати передавання телефонних повідомлень з визначенням місця розташування рухомих об'єктів і забезпеченням систем захисту.

Боротьба з багатопроменевістю. Застосування ШПС у системах зв'язку дозволяє боротися з багатопроменевістю поширення радіохвиль. Багатопроменевість виникає в тому випадку, коли радіохвилі приходять у точку прийому, відбившись від різних завад на шляху поширення (шари іоносфери, будинки, пагорби тощо). Через розбіжності в довжині шляху ці радіохвилі приходять з різним запізненням, внаслідок чого сигнали перекриваються в часі і між ними виникає інтерференція, що в свою чергу спричинює глибокі завмирання результуючого сигналу. Звичайно для компенсації завмирань передбачають збільшення потужності сигналу на 20 дБ. Інша річ – використання ШПС, оскільки при обробці ШПС погодженим фільтром відбувається стиск ШПС за часом. На рис. 5.3, а зображений ШПС із частотною модуляцією тривалістю T , на рис. 5.3, б зображена залежність між напругою на виході погодженого фільтра та відкликом фільтра на ШПС.

Цей відклик називається *автокореляційною функцією* (АКФ) ШПС. Хоча АКФ має тривалість $2T$, в ній можна виділити дві області, які різко відрізняються одна від одної. Перша область знаходиться в центрі АКФ; різкий викид у вигляді вузького імпульсу називають *центральною піком*. Його амплітуда дорівнює U , а тривалість

$$\tau_0 \approx 1 / F. \quad (5.18)$$

Чим ширший спектр ШПС, тим коротший центральний пік.

Другу область складають бічні піки з максимальним значенням u_{\max} .

Шумоподібні сигнали з великими базами мають властивості, що описуються двома співвідношеннями:

$$T / \tau_0 \approx B; \quad (5.19)$$

$$U / \gamma u_{\max} \approx \sqrt{\alpha B}, \quad (5.20)$$

де α – деяка стала, в загальному випадку залежна від бази B .

Співвідношення (5.19) визначає стиск ШПС – відношення тривалості ШПС T до тривалості центрального піка. Стиск ШПС дорівнює приблизно базі. Тому при $T = \text{const}$ із збільшенням F зменшується тривалість центрального піка τ_0 і зростає стиск.

Співвідношення (5.20) характеризує придушення бічних піків. Воно дорівнює відношенню амплітуди центрального піка U до амплітуди максимального бічного піка u_{max} . Чим більша база, тим більше придушення бічних піків. Границя АКФ ШПС із ростом бази прямує до вузького дельта-імпульсу. Таку АКФ має широкосмуговий шум.

На рис. 5.3, в зображений відклик погодженого фільтра на три ШПС, що надійшли по різних шляхах. Якщо затримка між променями Δt більша за тривалість центрального піка τ_0 , то промені розділяються; отже, центральні піки різних променів можна відділити один від одного, а потім об'єднати, усунувши затримку між ними. Цей принцип боротьби з багатопрореневістю був використаний в одній з перших систем зв'язку з ШПС "RAKE".

Отже, умова $\Delta t > \tau_0$ забезпечує поділ променів. Оскільки τ_0 і F пов'язані співвідношенням (5.18), то умова поділу променів записується так:

$$F\Delta t > 1. \quad (5.21)$$

Наприклад, якщо при поширенні радіохвиль існують два промені – прямий і відбитий від деякого об'єкта, – то затримка $\Delta t \approx 2d^2 / Rc$, де c – швидкість світла; R – відстань між передавачем і приймачем; d – відстань між відбиваючим об'єктом і прямим променем. У цьому випадку необхідно використовувати ШПС із шириною спектра

$$F \geq Rc / 2d^2. \quad (5.22)$$

Чим більше d , тим менше F , а тому при малих d можуть знадобитися ШПС із дуже широкими спектрами, що не завжди є реалізовним на практиці.

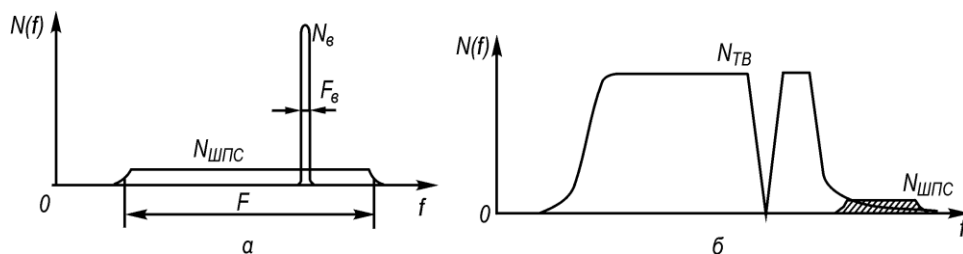


Рис. 5.4. Спектри широко- і вузькосмугової систем зв'язку (а), телевізійного сигналу і ШПС (б)

Вимір координат рухомих об'єктів і електромагнітна сумісність. Застосування ШПС дозволяє інтегрувати системи передавання інформації з системами траєкторних вимірів. Серед параметрів руху об'єкта найбільший інтерес становлять: 1) відстань між приймачем і передавачем; 2) їх відносна швидкість. Відстань вимірюється за затримкою сигналу в часі, а швидкість – за доплерівським зсувом частоти. Точність виміру і роздільна здатність за затримкою визначаються відношенням сигнал–завада q^2 (5.4) та шириною спектра сигналу і характеризуються похибкою

$$G t \approx 1 / qF. \quad (5.23)$$

Чим більші q і F , тим менша похибка виміру затримки, тим вищі точність виміру і роздільна здатність за відстанню. Точність виміру доплерівського зсуву частоти визначається відношенням сигнал–завада q^2 та тривалістю сигналу і характеризується похибкою

$$\sigma_f \approx 1/qT . \quad (5.24)$$

Чим більші q і T , тим менша похибка виміру доплерівського зсуву частоти, тим вищі точність виміру і роздільна здатність за швидкістю. Із (5.23), (5.24) випливає, що при спільному вимірюванні відстані та швидкості слід використовувати ШПС, тому що тільки для ШПС можна незалежно змінювати ширину спектра F і тривалість сигналу T . У системах зв'язку тривалість T звичайно визначається швидкістю передавання інформації. Тому підвищення точності виміру відстані можна досягти розширенням спектра F , тобто використовуючи ШПС.

Шумоподібні сигнали забезпечують добру електромагнітну сумісність (ЕМС) широкосмугових систем з вузькосмуговими системами радіозв'язку і мовлення. На рис. 5.4, а зображені спектри широко- і вузькосмугової систем зв'язку з ШПС.

Відповідно для ШПС спектральна щільність потужності $N_{\text{ШПС}} = P_{\text{ШПС}}/F$, а для вузькосмугового сигналу $N_{\text{в}} = P_{\text{в}}/F_{\text{в}}$. Завадостійкість системи зв'язку з ШПС визначається фундаментальним співвідношенням (2.4), в якому $\rho^2 = P_{\text{ШПС}}/P_{\text{в}}$. Посилення обробки дорівнює $2B$. Якщо вузькосмугова система зв'язку постійно займає визначений інтервал, то її спектр можна цілком придушити, використовуючи режекторний фільтр, настроєний на частоту вузькосмугової системи зв'язку. Отже, вплив вузькосмугової системи зв'язку на широкосмугову незначний. У свою чергу, широкосмугова система зв'язку також слабо впливає на вузькосмугову систему зв'язку. На виході приймача $N_{\text{ШПС}}F_{\text{в}} = P_{\text{ШПС}}F_{\text{в}}/F$. Тому відношення сигнал–завада на виході вузькосмугового приймача визначатиметься співвідношенням (5.4), в якому $\rho^2 = P_{\text{в}}/P_{\text{ШПС}}$, а $B = F/F_{\text{в}}$. Чим більше відношення $F/F_{\text{в}}$, тим краща фільтрація ШПС у вузькосмуговій системі зв'язку. Чим більша база ШПС, тим вища ЕМС широко- і вузькосмугової систем зв'язку.

Системи зв'язку з ШПС можна інтегрувати і з радіотелевізійними системами. На рис. 5.4, б зображений спектр телевізійного сигналу $N_{\text{ТВ}}$. Програми телебачення в одній територіальній зоні передаються по кількох каналах з великими захисними частотними інтервалами. Звичайно в цих частотних захисних інтервалах не допускається робота радіотехнічних систем, щоб не створювати завад телевізійним передачам. Однак в цих частотних інтервалах можна розмістити системи зв'язку з ШПС так, як це показано на рис. 5.4, а. Спектр ШПС розташований поблизу спектра телевізійного сигналу, там, де спектральна щільність останнього різко зменшується. При цьому взаємні завади обох систем будуть малими. Слід зазначити, що у разі використання сигналів з частотною модуляцією (замість ШПС) рівень взаємних завад зростає, тому що сигнали системи зв'язку і телебачення належать до одного класу і демодуються частотним детектором.

Отже, системи зв'язку з ШПС мають добру ЕМС із системами радіозв'язку, мовлення і телебачення. Раніше вже говорилося, що ШПС забезпечують високу ефективність використання радіоспектра в стільникових системах рухомого зв'язку. Якщо розглядати дію систем зв'язку в деякому замкненому просторі, то виявляється, що найкращу ЕМС в обмеженому діапазоні частот забезпечують ШПС, хоча самі по собі вони вимагають ширшої смуги, ніж традиційні вузькосмугові системи. Водночас загальна смуга частот при використанні ШПС буде меншою.

З розгляду основних властивостей ШПС випливає, що застосування ШПС у системах зв'язку дозволяє забезпечувати високу завадостійкість, скритність, адресність, працездатність у загальній смузі частот, подолання багатопроменевості, високі точності вимірів і роздільні здатності, добру ЕМС із багатьма радіотехнічними системами. Ці переваги надаються застосуванням ШПС із великими базами, що призводить до різкого ускладнення пристроїв формування й обробки, збільшення їхньої маси, об'єму, споживаної потужності. У більшості випадків перехід до ШПС із великими базами вимагає різкого розширення смуги частот, що спричинює певні труднощі зі створення широкосмугової елементної бази. Однак ці труднощі переборні. І тому ШПС уже зараз застосовуються в різних системах зв'язку і мають великі перспективи, особливо щодо систем масового радіозв'язку.

5.3. Основні структурні схеми широкосмугових систем зв'язку

Широкосмугові системи зв'язку з ШПС у залежності від призначення, технічних характеристик, бази ШПС, елементної бази можуть бути побудовані за різними схемами, перелічити які важко через численні варіанти їх. Для якісного уявлення про те, з яких основних пристроїв складаються ШСЗ, на рис. 5.5–5.8 наведені структурні схеми деяких систем зв'язку.

Структурна схема передавача і приймача цифрової системи зв'язку з фазоманіпульованим сигналом, призначена для передавання дискретних повідомлень. У передавачі (рис. 5.5, а) від джерела інформації *ДІ* послідовність двійкових одиниць 1 і 0 зі швидкістю $R = 1/T$ (рис. 5.5, в, I) надходить на перший вхід фазового модулятора *ФМ*. На другий вхід *ФМ* надходить фазоманіпульований сигнал (рис. 5.5, в, II) від генератора фазоманіпульованого сигналу *ГФМ*. Фазоманіпульований сигнал (ФМ-сигнал) має тривалість T і являє собою послідовність відеоімпульсів 1 і 0 тривалістю $\tau_0 = T/N$, де N – число імпульсів. На рис. 5.5, в, II $N = 13$. Звичайно вважають, що база ФМ-сигналу дорівнює числу імпульсів, тобто $B \approx N$. Ширина спектра ФМ-сигналу $F \approx 1/\tau_0$. Роботою *ГФМ* керує синхронізатор *С*, що формує необхідні сигнали керування і частоти. Послідовність ШПС у вигляді ФМ-сигналів, які переносять інформаційні символи (рис. 5.5, в, III), надходить у модулятор *Мод*, де здійснюється балансова модуляція коливання з несучою частотою ФМ-сигналом. Коливання з несучою частотою створюються генератором низької частоти *ГНЧ*. Підсилювач потужності *ПП* посилює ФМ-сигнал, який потім через антену випромінюється в простір.

У приймачі (рис. 5.5, б) сигнал проходить через змішувач *Зм*, переноситься за допомогою гетеродина *Г* на проміжну частоту, посилюється в підсилювачі проміжної частоти *ППЧ* і обробляється погодженим фільтром *ПФ*. Сигнал з виходу *ПФ* надходить на синхронізатор *С* і розв'язувальний пристрій *РП*. Синхронізатор здійснює пошук ФМ-сигналу за частотою і часом, накопичує сигнал для збільшення надійності синхронізації, керує режимом роботи розв'язувального пристрою.

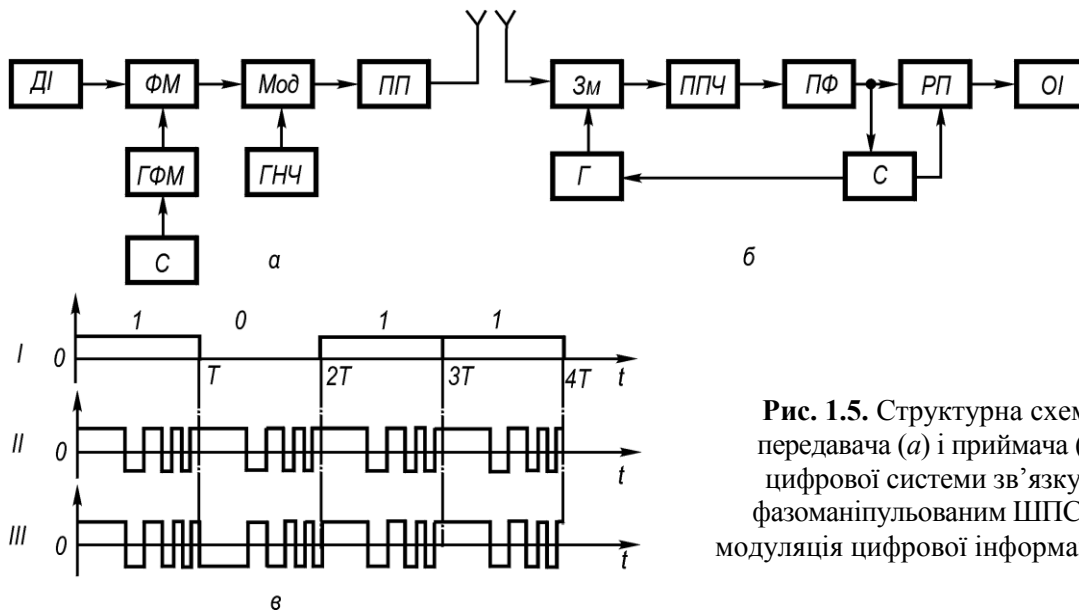


Рис. 1.5. Структурна схема передавача (а) і приймача (б) цифрової системи зв'язку з фазоманіпульованим ШПС та модуляція цифрової інформації (в)

Для пошуку ФМ-сигналу за частотою синхронізатором перенастроюється гетеродин. Після закінчення пошуку і набуття синхронізму на виході *РП* з'являється інформаційна послідовність у вигляді двійкових символів, що передається одержувачеві інформації *ОІ*. Приймач, зображений на рис. 5.5, б, є найпростішим. Разом з тим зазначимо, що погоджений фільтр і синхронізатор, який містить блоки пошуку і синхронізації, у разі великих баз ШПС є складними пристроями. Крім того, для пошуку ШПС і підтримки синхронізму приймач охоплений петлею зворотного зв'язку. Реальний приймач ШПС може містити кілька блоків пошуку і спостереження, в тому числі блок

пошуку ШПС за часом і часовою синхронізацією, блок фазового автопідстроювання частоти (ФАПЧ), що охоплені власними і взаємними зворотними зв'язками.

Структурна схема передавача і приймача радіотелефонної системи зв'язку з фазоманіпульованим ШПС. У передавачі (рис. 5.6, а) телефонне повідомлення (рис. 5.6, в, I) від джерела інформації *ДІ* надходить на вхід широтно-імпульсного модулятора *ШІМ*, з виходу якого ШІМ-сигнал (рис. 5.6, в, II) подається на вхід фазового модулятора *ФМ*. На другий вхід *ФМ* подається фазоманіпульований ШПС (рис. 5.6, в, III), який формується ГФМ. Фазоманіпульований сигнал з виходу *ФМ*, що містить інформацію (рис. 5.6, в, IV), надходить на вхід модулятора *Мод*, де здійснюється балансова модуляція коливання з несучою частотою від генератора низької частоти *ГНЧ*. Потім посилений в підсилювачі потужності *ПП* ФМ-сигнал через антену випромінюється в простір. Роботою широтно-імпульсного модулятора *ШІМ* і генератора ФМ-сигналу *ГФМ* керує синхронізатор *С*, що виробляє необхідні частоти і керуючі сигнали.

У приймачі (рис. 5.6, б) прийнятий сигнал у змішувачі *Зм* за допомогою гетеродина *Г* переноситься на проміжну частоту і після підсилювача проміжної частоти *ППЧ* надходить на корелятор *Кор*. Корелятор, як і погоджений фільтр, робить оптимальну обробку прийнятого сигналу. Хоча корелятор і погоджений фільтр відрізняються за принципом роботи, але забезпечують однакову завадостійкість прийому.

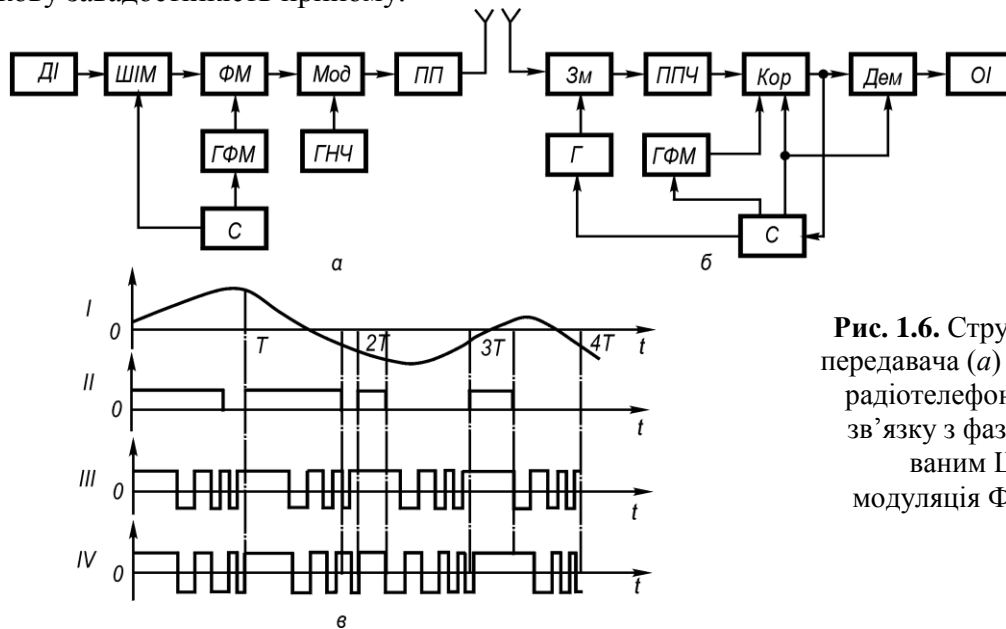


Рис. 1.6. Структурна схема передавача (а) і приймача (б) радіотелефонної системи зв'язку з фазоманіпульованим ШПС та модуляція ФМ ШПС (в)

Корелятор складається з перемножника й інтегратора. На другий вхід корелятора подається опорний сигнал у вигляді фазоманіпульованого ШПС (рис. 5.6, в, III). Напряга на виході корелятора містить телефонне повідомлення у вигляді ШІМ-сигналу, що подається на вхід демодулятора *Дем*, з виходу якого прийняте телефонне повідомлення передається одержувачеві інформації *ОІ*.

Робота приймача в цілому і його окремих блоків (*Г*, *ГФМ*, *Кор*, *Дем*) керується блоком пошуку ФМ ШПС за часом і частотою, а потім приймач підтримує синхронізм. Усе раніше сказане щодо синхронізатора приймача, зображеного на рис. 5.6, б, цілком стосується синхронізатора будь-якого приймача (рис. 5.7, б, 5.8, б).

Структурна схема передавача і приймача цифрової системи зв'язку з частотно-маніпульованим ШПС*. Відмінність передавача і приймача, зображених на рис. 5.7, від передавача і приймача на рис. 2.5 зводиться до наступного. У модуляторі *МодI* передавача (рис. 5.7, а) проводиться модуляція частотно-маніпульованого ШПС (ЧМ ШПС) дискретним повідомленням. ЧМ ШПС являє собою сигнал, що складається з *M* імпульсів, несучі частоти яких набувають одного з можливих значень: від f_0 до $f_0 + M - 1 / T$ з інтервалом між сусідніми значеннями $\Delta f = 1/T$. Усього використовується *M* частот, і жодна з них не застосовується двічі в

* Іноді такий ШПС називають сигналом зі стрибучою частотою.

одному ШПС. База такого сигналу $B \approx M^2$. ЧМ ШПС формується за допомогою частотного маніпулятора $ЧМ$, на перший вхід якого через шину (широка стрілка) подаються M частот від генератора сітки частот $ГСЧ$. На другий вхід подається кодова послідовність від генератора кодової послідовності ЧМ ШПС ($ГЧМ$), що визначає порядок зміни частот у ЧМ ШПС. У модуляторі $Мод2$ ЧМ ШПС переноситься на несучу частоту. Роботою $ГСЧ$, $ГЧМ$, $ГНЧ$ керує синхронізатор $С$.

У приймачі (рис. 5.7, б) ЧМ ШПС на проміжній частоті надходить на змішувач $Зм2$, де всі частоти сигналу переносяться на іншу проміжну частоту за допомогою опорного ЧМ ШПС, що надходить від частотного маніпулятора $ЧМ$. Призначення $ГСЧ$ і $ГЧМ$ таке ж саме, як $ГФМ$ та $ГНЧ$ у передавачі, що на рис. 5.5, а. З виходу ППЧ2 сигнал тривалістю T , що не має частотної маніпуляції, надходить на погоджений фільтр $ПФ$, а потім на розв'язувальний пристрій $РП$ і синхронізатор $С$. Останній робить пошук ЧМ ШПС за часом і частотою, потім підтримує синхронізм і керує роботою $Г$, $ГСЧ$, $ГЧМ$ і $РП$.

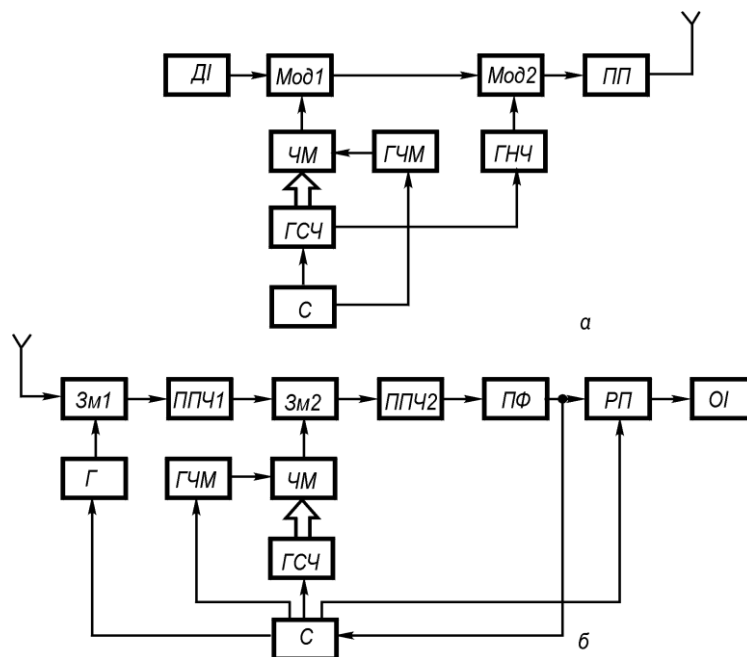


Рис. 5.7. Структурна схема передавача (а) і приймача (б) цифрової системи зв'язку з частотно-маніпульованим ШПС

Структурна схема передавача і приймача цифрової системи зв'язку з фазочастотно-маніпульованим ШПС. Фазочастотно-маніпульований ШПС (скорочено ФМ-ЧМ ШПС) є складним. При певній подвійній маніпуляції він складається з N імпульсів, групи яких передаються на M частотах. База такого сигналу $B \approx N \cdot M^2$. Оскільки ФМ-ЧМ ШПС є об'єднанням ФМ- і ЧМ-сигналів, то і схема передавача (рис. 5.8, а), і схема приймача (рис. 5.8, б) теж, у свою чергу, є об'єднанням схем передавачів, зображених на рис. 5.5, а і 5.7, а, та приймачів, зображених на рис. 5.5, б і 5.7, б. Для додаткового підвищення завадостійкості використовуються коректувальні коди, що формуються в передавачі (рис. 5.8, а) за допомогою кодера K і декодуються в приймачі (рис. 5.8, б) за допомогою декодера D .

У приймачі оптимальну фільтрацію здійснює корелятор $Кор$. Призначення інших блоків таке ж саме, як і в попередніх схемах.

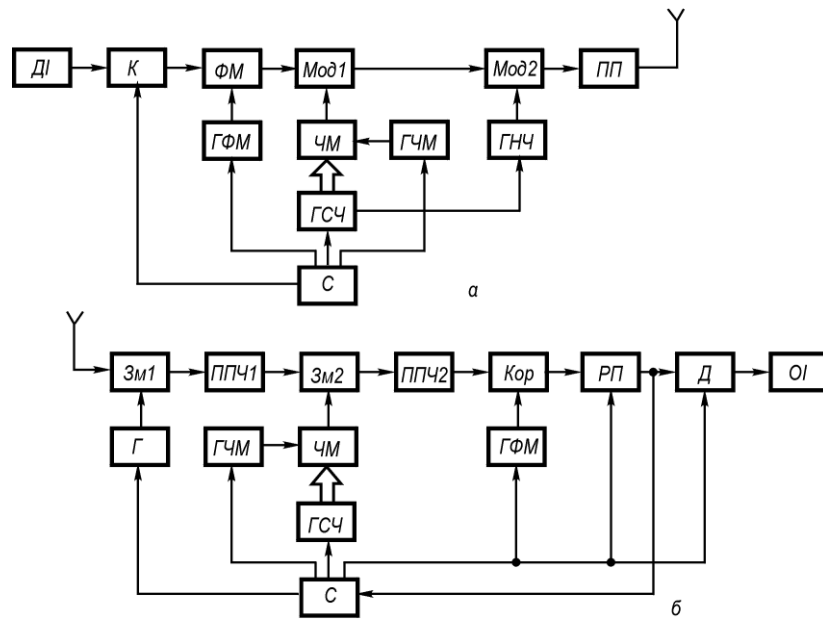


Рис. 5.8. Структурна схема передавача (а) і приймача (б) цифрової системи зв'язку з фазочастотно-маніпульованим ШПС і коректувальними кодами

Подані схеми не вичерпують усього різноманіття їх, але дозволяють виділити основні вузли широкосмугових систем зв'язку з ШПС. До таких вузлів належать генератори формування ШПС (або автомати формування ШПС з їх зміною), генератори сітки частот, погоджені фільтри, корелятори, блоки пошуку ШПС і синхронізації за часом і частотою. З викладеного матеріалу випливає, що розробник широкосмугової системи зв'язку має вміти вибрати тип ШПС і його базу, метод обробки, визначити час пошуку і синхронізації, знайти завадостійкість приймача ШПС у разі дії різного виду завад, визначити елементну базу, розробити для неї потрібні генератори ШПС, погоджені фільтри і корелятори, блоки пошуку і синхронізації. Крім цього, розробник повинен правильно спроектувати інші вузли передавача і приймача, знати, як проходить ШПС через вузли передавача і приймача та які втрати при цьому мають місце. На всі питання, що виникають у процесі проектування ШСЗ, у більшості випадків дати однозначні відповіді не можна. Тому проектування ШСЗ у даний час є інженерним мистецтвом, що ґрунтується на глибокому знанні теорії й техніки ШПС і на інтуїції розробника.

5.4. Шумоподібні сигнали та їхні спектри

Сигнал є змінною фізичною величиною, яка відображає повідомлення. Сигнал u , що є функцією часу t , записується у вигляді $u = u(t)$.

Множина сигналів $u_j(t)$, побудована за єдиним правилом, називається *системою сигналів*. Отже, система сигналів визначена, якщо є відомим правило побудови сигналів. Номер сигналу позначається індексом j . Якщо в системі кількість сигналів L , то можна пронумерувати сигнали натуральними числами від 1 до L і позначити $j=1, L$. Число L називається *об'ємом системи сигналів*.

Надалі розглядатимуться сигнали, які можна подати в такому вигляді:

$$u(t) = A(t) \cos[\omega_0 t + \theta(t)], \quad (5.25)$$

де $A(t)$ – обвідна сигналу; ω_0 – несуча частота; $\theta(t)$ – повільнозмінна частина фази сигналу.

Зображенню (5.25) відповідає радіочастотний сигнал. Оскільки розглядаються реальні сигнали (які можна сформувати й обробити), то всі функції часу і параметри правої частини (5.25) відомі.

Якщо сигнал заданий у загальному вигляді $u(t)$ і права частина (5.25) невідома, слід скористатися перетворенням Гільберта

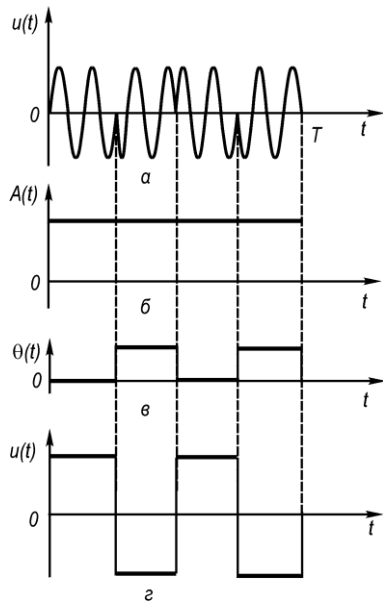


Рис. 5.9. Фазоманіпульований сигнал

$$u(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathcal{K}(x)}{x-1} dx, \quad \mathcal{K}(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(x)}{x-1} dx$$

і знайти спряжений сигнал $\mathcal{K} t$. У цьому випадку обвідна $A(t) = \sqrt{u^2(t) + \mathcal{K}^2(t)}$, фаза $\omega_0 t + \theta(t) = \arctg \mathcal{K}(t)/u(t)$.

Якщо функція θt неперервна і має неперервну першу похідну, то миттєва частота сигналу $\omega_0 t$ за означенням дорівнює першій похідній фази θt , тобто $\omega t = \omega_0 t + \dot{\theta} t$.

На рис. 5.9, а показаний фазоманіпульований сигнал (ФМ), що складається з чотирьох радіоімпульсів з однаковою несучою частотою, але з різними початковими фазами. На рис. 5.9, б, в зображено його обвідну $A t$ і нульову фазу θt . Обвідна є сталою величиною на інтервалі тривалістю T , а фаза дорівнює двом значенням: 0 або π .

Якщо несуча частота сигналу $\omega_0 t = 0$, то такий сигнал є *відеочастотним*. На рис. 5.9, г зображений відеочастотний сигнал $u t$ – послідовність позитивних і негативних

прямокутних імпульсів, – отриманий із ФМ-сигналу на рис. 5.9, а за умови, що $\omega_0 t = 0$. Оскільки знаки імпульсів відеочастотного сигналу визначаються початковими фазами імпульсів радіочастотного сигналу, то за аналогією з радіочастотним сигналом відеочастотний також називається *фазоманіпульованим* сигналом.

Спектр сигналу визначається перетворенням Фур'є

$$g(\omega) = \int_{-\infty}^{\infty} u(t) e^{-i\omega t} dt. \quad (5.26)$$

Спектр є функцією кутової частоти $\omega = 2\pi f$, де f – лінійна частота (надалі ω і f називатимуться просто частотою). Нескінченні межі інтегрування відповідають загальному випадкові. При визначенні спектра фінітного сигналу (із скінченною тривалістю) необхідно враховувати його розташування на осі часу t . Спектр може бути зображений у вигляді $g \omega = |g \omega| \exp i\phi \omega$, де $|g \omega|$ – амплітудний, а $\phi \omega$ – фазовий спектр сигналу.

Сигнал знаходиться за спектром оберненим перетворенням Фур'є

$$u(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\omega) e^{i\omega t} d\omega.$$

Ширина спектра. Спектр фінітних сигналів має нескінченну довжину, тому єдиного визначення ширини спектра не існує. Залежно від мети дослідження ширину спектра сигналу знаходять по-різному. Надалі ширина спектра визначатиметься так, щоб правильно відобразити суть розв'язуваної задачі. Цей підхід виправданий тим, що для сигналів, які входять в одну систему, будь-яке достатньо розумне визначення ширини спектра відбиватиме спектральні властивості кожного сигналу і системи сигналів у цілому. (Ширина спектра сигналу позначається F .)

Комплексна обвідна сигналу та її спектр. Радіосигнал (5.25) містить швидкозмінливий множник у вигляді косинусоїди, в аргумент якої входить несуча частота $\omega_0 = 2\pi f_0$. Відповідно спектр (5.26) цього сигналу складається з двох частотних смуг, зосереджених біля частот ω_0 і $-\omega_0$. При теоретичних дослідженнях доцільно для спрощення проміжних математичних операцій “звільнити” сигнал і його спектр від несучої частоти ω_0 . Це можна здійснити, ввівши комплексну обвідну сигналу.

Комплексна обвідна радіосигналу (5.25) визначається як

$$u(t) = |u(t)| \exp i\omega_0(t), \quad (5.27)$$

де модуль $|u(t)| = A t$ є обвідною сигналу $u(t)$. Перехід від комплексної обвідної до сигналу здійснюється за допомогою формули

$$u(t) = \operatorname{Re} [u(t) \exp i\omega_0(t)]. \quad (5.28)$$

На рис. 5.9, з зображена комплексна обвідна ФМ-сигналу, наведеного на рис. 5.9, а. Вона є послідовністю прямокутних відеоімпульсів і дійсною функцією часу. Це обумовлено тим, що початкові фази імпульсів ФМ-сигналу набувають одного з двох значень: 0 або π . У загальному випадку комплексна обвідна містить і дійсну, і уявну складові, але завжди є відеосигналом, що й пояснює перехід до неї від радіосигналу.

Спектр комплексної обвідної

$$G(\omega) = \int_{-\infty}^{\infty} u(t) e^{-i\omega t} dt. \quad (5.29)$$

Комплексна обвідна сигналу знаходиться відповідно до оберненого перетворення Фур'є

$$u(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{i\omega t} d\omega. \quad (5.30)$$

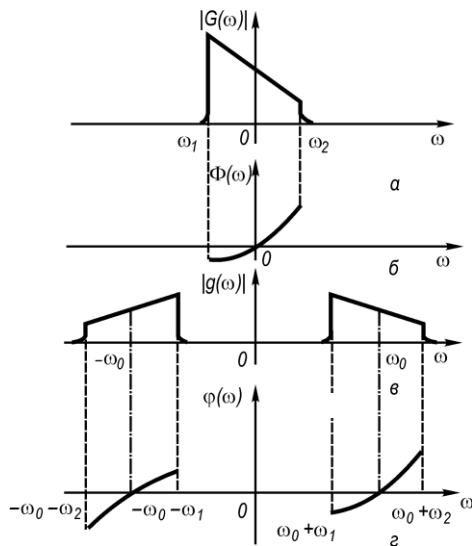


Рис. 5.10. Амплітудний спектр $|G(\omega)|$ комплексної обвідної сигналу (а), фазовий спектр $\Phi(\omega)$ (б), спектр сигналу (в)

Спектр комплексної обвідної можна зобразити у вигляді

$$G(\omega) = |G(\omega)| \exp i\Phi(\omega),$$

де $|G(\omega)|$ – амплітудний, а $\Phi(\omega)$ – фазовий спектри.

Спектр сигналу $g(\omega)$ і спектр його обвідної комплексної $G(\omega)$ пов'язані співвідношенням $g(\omega) = 0,5G(\omega - \omega_0) + 0,5G^*(\omega - \omega_0)$, де $*$ – знак комплексної спряженості.

Оскільки комплексна обвідна $u(t)$ – відеосигнал, то спектр $G(\omega)$ розташований в області відеочастот.

На рис. 5.10 зображений спектр $G(\omega)$ комплексної обвідної $u(t)$ довільного сигналу.

Енергія сигналу і частотно-часова площа. За означенням, енергія сигналу

$$E = \int_{-\infty}^{\infty} u^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |g(\omega)|^2 d\omega. \quad (5.31)$$

Для сигналів, у яких $|\theta'(t)|_{\max} \ll \omega_0$, енергія сигналу виражається через модулі комплексної обвідної та її спектра так:

$$E = \frac{1}{2} \int_{-\infty}^{\infty} |u(t)|^2 dt = \frac{1}{4\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 d\omega. \quad (5.32)$$

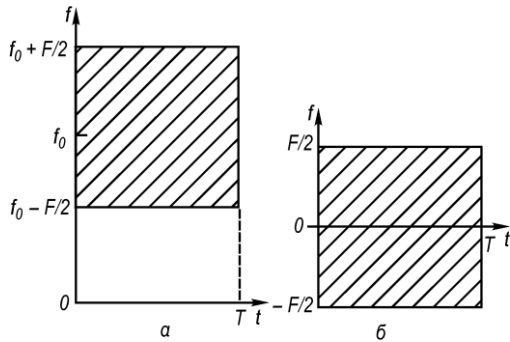


Рис. 5.15. Частотно-часова площина на радіочастоті (а) і на відеочастоті (б)

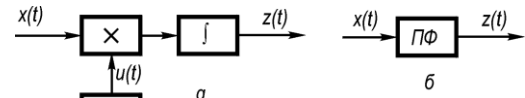


Рис. 5.12. Корелятор (а) і погоджений фільтр (б)

Як правило, більша частина енергії сигналу зосереджена в деякій смузі частот. Нехай F – ширина смуги частот, усередині якої зосереджена більша частина заданої енергії, а поза цією смугою – менша, якою можна знехтувати. Визначена в такий спосіб ширина смуги частот F вважається шириною спектра сигналу. В цьому випадку енергія сигналу зосереджена в частотно-часовому прямокутнику зі сторонами T (вздовж осі часу) і F (вздовж осі частот). Для передавання сигналу з допустимою точністю необхідно мати канал зі смугою частот, ширина котрої F , і час передавання T .

На рис. 5.11, а наведений приклад розподілу енергії сигналу на частотно-часовій площині (f, t). Спектр сигналу зосереджений біля несучої частоти f_0 і розташовується від $f_0 - F/2$ до $f_0 + F/2$. Рисунок має якісний характер, тому що для фінітних сигналів енергія розподілена в смузі $0 \leq t \leq T, |f| < \infty$. Оскільки комплексна обвідна є відеосигналом, то частотно-часовий прямокутник, по якому розподілена основна частина енергії сигналу, виглядатиме, як на рис. 5.11, б. Цей прямокутник виходить з базисного прямокутника, що на рис. 5.11, а, при зсуві останнього вниз за частотою на f_0 , що відповідає переходу від радіосигналу з несучою частотою f_0 до його комплексної обвідної.

5.5. Оптимальна обробка сигналів

Якщо на вході приймача сигнал $x(t)$ дорівнює сумі корисного сигналу $u(t)$ і завади $n(t)$ або тільки заваді, то оптимальний приймач за відомими параметрами сигналу обчислює так званий *кореляційний інтеграл*, а потім порівнює його величину з порогом z_0 . Якщо завада є гауссовим випадковим процесом, спектральна щільність якого рівномірна (білий шум), то кореляційний інтеграл має вигляд

$$z = \int_0^T x(t)u(t)dt . \quad (5.33)$$

Кореляційний інтеграл (5.33) визначається за допомогою корелятора (рис. 5.12, а) або погодженого фільтра (рис. 5.12, б). Основними елементами корелятора, як випливає з (5.33), є перемножник, генератор опорного сигналу *ГОС* й інтегратор. На перемножник надходять вхідний сигнал $x(t)$ і сигнал $u(t)$ від генератора опорного сигналу *ГОС*. Добуток $x(t)$ і $u(t)$ інтегрується з моменту приходу сигналу $t=0$ до моменту його закінчення $t=T$. Зазначимо, що корелятор є пристроєм зі змінними параметрами, тому що режим його роботи залежить від змінення $u(t)$ у часі. Оскільки операції множення й інтегрування лінійні, то корелятор є лінійним пристроєм. Зважаючи на те, що він відфільтровує сигнал від завад і є лінійним пристроєм зі змінними параметрами, його іноді називають *активним фільтром* на відміну від *пасивних* фільтрів, параметри яких є сталими в часі.

Погоджений фільтр є пасивним. Напряга на виході погодженого фільтра в момент закінчення сигналу $t=T$ з точністю до сталого множника a дорівнює напрузі на виході корелятора

$$u(T) = a \cdot z . \quad (5.34)$$

Імпульсна характеристика погодженого фільтра

$$h(t) = au(T - t) \quad (5.35)$$

за формою є дзеркально відбитим сигналом із запізненням T .

Спільним у кореляторі і погодженому фільтрі є рівність (з точністю до сталої) вихідних напруг у момент часу $t = T$. Це і визначає їхню взаємну еквівалентність щодо виявлення сигналу. А різниця між ними полягає в наступному. Оскільки корелятор є пристроєм зі змінними в часі параметрами, а погоджений фільтр – пристроєм зі сталими параметрами, то погоджений фільтр інваріантний до затримки сигналу і його початкової фази (наскільки ці величини зміняться в сигналі на вході фільтра, настільки вони зміняться й у сигналі на виході), а корелятор – не інваріантний.

Якщо сигнал має деякі невимірні або вимірні випадкові параметри, то структура оптимального приймача дещо змінюється, але його основна частина залишається такою самою, оскільки погоджений фільтр або корелятор завжди мають бути. Наприклад, при випадковій початковій фазі сигналу в приймачі з погодженим фільтром за останнім має йти детектор для виділення обвідної. В приймачеві кореляційного типу мають бути ще один, другий (квадратурний) канал і схема виділення обвідної. Саме тому в оптимальному приймачі завжди є погоджений фільтр або корелятор. Коефіцієнт передачі погодженого фільтра з імпульсною характеристикою (5.35)

$$k(\omega) = a^* g(\omega) \exp(-i\omega T), \quad (5.36)$$

де $g(\omega)$ – спектр сигналу; $*$ – знак комплексної спряженості. Із (5.36) випливають вирази для амплітудно-частотної характеристики (АЧХ)

$$|k(\omega)| = a |g(\omega)| \quad (5.37)$$

і фазочастотної характеристики (ФАХ) погодженого фільтра

$$\psi(\omega) = -\varphi(\omega) - \omega T. \quad (5.38)$$

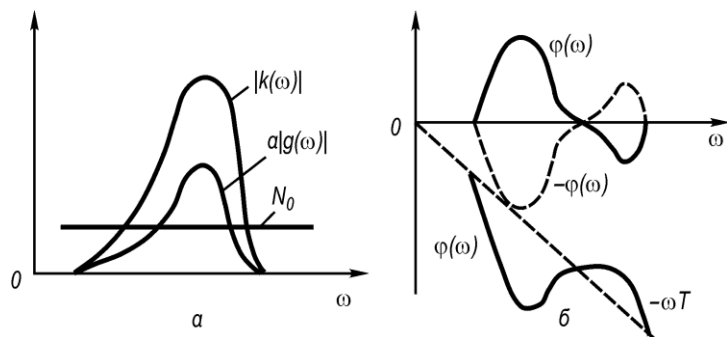


Рис. 1.13. Амплітудно-частотна (а) і фазочастотна (б) характеристики погодженого фільтра

Із (5.37) випливає, що АЧХ погодженого фільтра є тим вищою, чим більша спектральна щільність сигналу. На рис. 5.13, а зображені амплітудний спектр сигналу $a|g(\omega)|$, спектральна щільність завади N_0 і АЧХ фільтра $|k(\omega)|$, побудована відповідно до (5.37).

Фазовий спектр сигналу $\varphi(\omega)$ і ФЧХ фільтра $\psi(\omega)$, що описуються рівнянням (5.38), показані на рис. 5.13, б. Штриховою лінією зображені складові правої частини рівняння (5.38). ФЧХ фільтра компенсує фазовий спектр сигналу, внаслідок чого на виході фільтра в якийсь момент часу всі частотні складові відклику опиняються в фазі і, додаючись, дають максимум відклику.

Виняткова роль погодженого фільтра (або корелятора) в оптимальному приймачі пояснюється тим, що він максимізує відношення сигнал–завада на своєму виході. Це відношення у разі дії на вході фільтра білого шуму зі спектральною щільністю N_0 і сигналу з енергією E не залежить від форми сигналу:

$$q_0 = \sqrt{2E/N_0}. \quad (5.39)$$

При цьому максимальне значення сигналу на виході фільтра дорівнює aE , а середньоквадратичне значення завади – $a\sqrt{EN_0/2}$.

Результати, наведені в даному параграфі, строго слушні для завади у вигляді гауссового випадкового процесу з рівномірною спектральною щільністю потужності (білий шум). Разом з тим корелятор або погоджений фільтр (чи їхній набір, або модифікація) завжди входять в оптимальний приймач.

5.6. Кореляційні функції сигналів

Оптимальне приймання сигналів здійснюється за допомогою погоджених фільтрів або кореляторів. Нормований до певної напруги відклик погодженого фільтра, знайдений за допомогою інтеграла згортки:

$$r_{jk}(\tau) = \frac{1}{E} \int_{-\infty}^{\infty} u_j(t)u_k(t-\tau)dt, \quad (5.40)$$

де $u_j(t)$ – сигнал на вході фільтра, погодженого із сигналом $u_k(t)$; τ – зсув сигналу $u_j(t)$ щодо відклику $u_k(t)$. Енергії сигналів з номерами j і k дорівнюють E . При $j=k$ і $\tau=0$ з (5.40), відкидаючи індекси, маємо

$$r(0) = r_{\max} = \frac{1}{E} \int_{-\infty}^{\infty} u^2(\tau)dt = 1, \quad (5.41)$$

що й визначає нормованість відклику погодженого фільтра.

Вираз у правій частині (5.40) визначає інтегральний взаємозв'язок між сигналами $u_j(t)$ і $u_k(t)$ при деякому зсуві τ . Якщо τ – змінна величина, то $r_{jk}(\tau)$ – функціонал, що залежить як від функцій $u_j(t)$ і $u_k(t)$, так і від зсуву τ . Саме тому $r_{jk}(\tau)$ називається **кореляційною функцією (КФ)** сигналів $u_j(t)$ і $u_k(t)$. У залежності від того, погоджений або не погоджений сигнал з фільтром та чи є додатково доплерівський зсув несучої частоти сигналу, кореляційні функції мають різні зображення.

Взаємна функція невизначеності (ВФН) двох сигналів з номерами j і k , за означенням, виражається через комплексні обвідні сигналів і через їхні спектри так:

$$R_{jk}(\tau, \Omega) = \frac{1}{2E} \int_{-\infty}^{\infty} u_j(t)u_k^*(t-\tau)e^{i\Omega t} dt = \frac{1}{4\pi E} \int_{-\infty}^{\infty} G_j(\omega-\Omega)G_k^*(\omega)e^{i\omega\tau} d\omega, \quad (5.42)$$

де τ – зсув за часом між сигналами; Ω – доплерівський зсув частоти. З точністю до малих більш високого порядку нормований відклик погодженого фільтра пов'язаний із ВФН (5.42) співвідношенням

$$r_{jk}(\tau, \Omega) = \text{Re } R_{jk}(\tau, \omega) \exp(i\omega_0\tau). \quad (5.43)$$

Взаємкореляційна функція (ВКФ) є перетином ВФН при $\Omega=0$. Взявши $\Omega=0$, з (5.43) одержуємо

$$R_{jk}(\tau) = \frac{1}{2E} \int_{-\infty}^{\infty} u_j(t)u_k^*(t-\tau)dt = \frac{1}{4\pi E} \int_{-\infty}^{\infty} G_j(\omega)G_k^*(\omega)e^{i\omega\tau} d\omega. \quad (5.44)$$

Функція невизначеності (ФН). Якщо фільтр погоджений із сигналом, тобто $j=k$, то з (5.43), опустивши індекс j , одержуємо ФН:

$$R(\tau, \Omega) = \frac{1}{2E} \int_{-\infty}^{\infty} u(t)u^*(t-\tau)e^{i\Omega t} dt = \frac{1}{4\pi E} \int_{-\infty}^{\infty} G(\omega-\Omega)G^*(\omega)e^{i\omega\tau} d\omega. \quad (5.45)$$

Автокореляційна функція (АКФ) – перетин ФН при $\Omega = 0$. Взявши $\Omega = 0$, з (5.45) знаходимо АКФ:

$$R(\tau) = \frac{1}{2E} \int_{-\infty}^{\infty} u(t)u^*(t-\tau)dt = \frac{1}{4\pi E} \int_{-\infty}^{\infty} |G(\omega)|^2 e^{i\omega\tau} d\omega. \quad (5.46)$$

З рівності (5.46) видно, що АКФ є перетворенням Фур'є енергетичного спектра комплексної обвідної сигналу. Відповідно до оберненого перетворення Фур'є енергетичний спектр

$$|G(\omega)|^2 = 2E \int_{-\infty}^{\infty} R(\tau) e^{-i\omega\tau} d\tau. \quad (5.47)$$

Розглянемо приклад, що ілюструє властивості автокореляційної функції. На рис. 5.14, а, б зображені простий сигнал у вигляді прямокутного імпульсу і його автокореляційна функція. Максимум $R(\tau)$ припадає на момент закінчення сигналу ($t=T$). Це пояснюється тим, що $R(\tau)$ є нормованою напругою на виході фільтра, погодженого з вхідним сигналом, максимум якого збігається з моментом закінчення сигналу, тобто при $\tau=T$ максимум $R(\tau) = 1$ відповідно до (5.41).

Для автокореляційної функції, зображеної на рис. 5.14, б, енергетичний спектр (квадрат модуля амплітудного спектра) відповідно до (5.47) описується функцією

$$|G(\omega)|^2 = \sin^2(\omega T/2) / (\omega T/2)^2. \quad (5.48)$$

На рис. 5.15, а зображений фазоманіпульований шумоподібний сигнал (ФМ ШПС) тривалістю T , а на рис. 5.15, б – його АКФ. Елементарний імпульс має тривалість $\tau_0 = T/N$, де N – число імпульсів. Для ФМ ШПС, зображеного на рис. 5.15, а, $N=15$. Автокореляційна функція ФМ ШПС має центральний пік в інтервалі $(-\tau_0, \tau_0)$ з одиничною амплітудою та бічні піки, розподілені на інтервалах $(-T, -\tau_0)$ і (τ_0, T) .

Амплітуди бічних піків набувають різних значень, але в сигналах з “добрими” кореляційними властивостями вони малі, тобто істотно менші за амплітуду центрального піка, що дорівнює одиниці. Існують різні оцінки бічних піків як АКФ, так і ВКФ, ВФН,

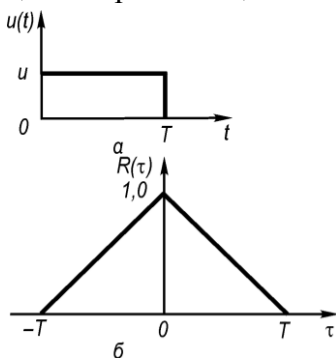


Рис. 5.14. Прямокутний імпульс (а) і його автокореляційна функція (б)

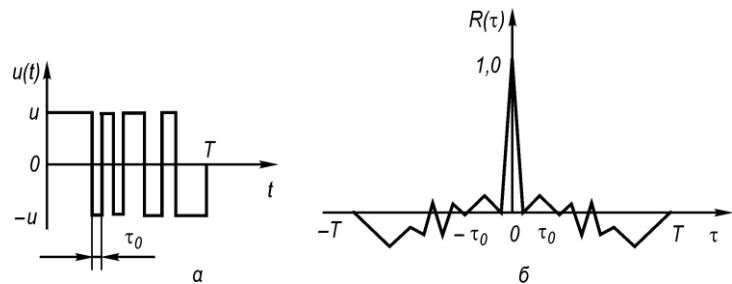


Рис. 5.15. Фазоманіпульований шумоподібний сигнал (а) і його автокореляційна функція (б)

ФН, проте всі вони описуються однаковими за формою відношеннями. Для ФМ ШПС оцінка бічних піків має вигляд

$$R = \alpha / \sqrt{N}, \quad (5.49)$$

де α – деяка величина, що залежить від вигляду оцінки, класу сигналу і, в загальному випадку, від N . Для довільних ШПС із базою B оцінка бічних піків

$$R = \beta / \sqrt{B}, \quad (5.50)$$

де β , як і α в (5.49), – деяка стала величина. Співвідношення (5.49), (5.50) визначають однакову залежність оцінок величини бічних піків від бази ШПС, оскільки величина N у ФМ ШПС є пропорційною базі B . Чим більша база, тим менші бічні піки. У граничному переході, коли $B \rightarrow \infty$, АКФ має вигляд трикутного імпульсу (рис. 5.16). Бічні піки на рис. 5.16 не ображені, оскільки при $B \rightarrow \infty$ вони наближаються до нуля відповідно до (5.49), (5.50). Тривалість центрального піка АКФ також прямує до нуля, оскільки $\tau_0 = T/N$, а з ростом бази B число імпульсів $N \rightarrow \infty$. АКФ, зображена на рис. 2.16, називається *ідеальною*, тому що вона не має бічних піків. Саме таку АКФ мають тривалі реалізації шуму.

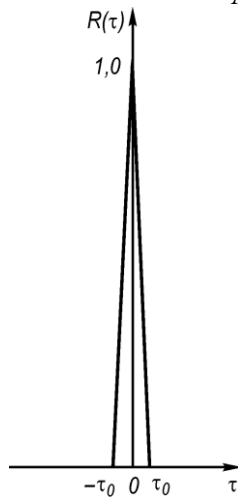


Рис. 5.16.
Ідеальна АКФ

Частотна кореляційна функція (ЧКФ) – перетин ФН при $\tau=0$. Взвзявши $\tau=0$, з (5.21) одержуємо

$$R(\Omega) = \frac{1}{2E} \int_{-\infty}^{\infty} |u(t)|^2 e^{i\Omega t} dt = \frac{1}{4\pi E} \int_{-\infty}^{\infty} G(\omega - \Omega) G^*(\omega) d\omega. \quad (5.51)$$

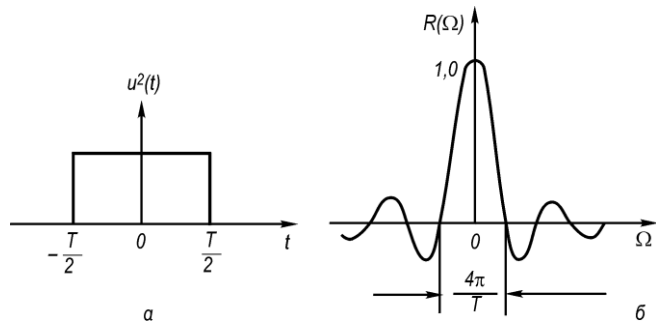


Рис. 5.17. Квадрат обвідної фазоманіпульованого сигналу (а) і його ЧКФ (б)

З першої рівності (5.51) випливає, що ЧКФ є перетворенням Фур'є квадрата обвідної сигналу. Вона не залежить від фазової структури сигналу, а визначається тільки квадратом модуля обвідної. Наприклад, для простого сигналу (рис. 5.14, а) і для фазоманіпульованого шумоподібного (рис. 5.15, а) квадрат обвідної дорівнює одиниці (рис. 5.17).

Тому ЧКФ сигналів, зображених на рис. 5.14, а і 2.15, а, однакові і записуються так:

$$R(\Omega) = (\sin \Omega T / 2) / (\Omega T / 2). \quad (5.52)$$

Нульові значення ЧКФ (рис. 117, б) ідуть з інтервалом $2\pi/T$.

Максимум і симетрія кореляційних функцій. Функції (5.40), (5.42), (5.46) називаються, як було зазначено раніше, *кореляційними функціями* (КФ). Відомо, що максимум КФ має місце лише при $j=k$ і $\tau=0$, $\Omega=0$, тобто тільки в центрі ФН (або АКФ і ЧКФ). Максимум

$$R_j(0,0) = R_k(0,0) = 1, \quad (5.53)$$

що аналогічно (5.41), а

$$|R_{jk}(\tau, \Omega)|_{j \neq k} < 1, \quad |R(\tau, \Omega)|_{\tau \neq 0, \Omega \neq 0} < 1.$$

Властивість симетрії КФ полягає в тому, що

$$R_{jk}(-\tau, -\Omega) = R_{jk}^*(\tau, \Omega) e^{i\Omega \tau}. \quad (5.54)$$

Із (5.54) випливає, що

$$|R_{jk}(-\tau, -\Omega)| = |R_j(-\tau, -\Omega)| = |R_j(\tau, \Omega)|, \quad (5.55)$$

$$R_j(-\tau) = R_j^*(\tau), \quad R_j(\Omega) = R_j^*(\Omega). \quad (5.56)$$

Об'єм і середньоквадратичні значення ВФН і ФН. Об'єм між поверхнею, описуваною квадратом модуля ВФН, і площиною невизначеності (або просто об'єм ВФН) дорівнює одиниці, тобто

$$\frac{1}{2\pi} \iint_{-\infty}^{\infty} |R_{jk}(\tau, \Omega)|^2 d\tau d\Omega = 1, \quad (5.57)$$

і не залежить від форми і номерів сигналів. Взявши $j=k$ і відкинувши індекси, маємо, що об'єм ФН також не залежить від форми сигналу і дорівнює одиниці, тобто

$$\frac{1}{2\pi} \iint_{-\infty}^{\infty} |R(\tau, \Omega)|^2 d\tau d\Omega = 1. \quad (5.58)$$

Формули (5.57) і (5.58) дозволяють знайти ефективні значення ВФН і ФН. Позначимо їх через $R_{jk \text{ еф}}$ і $R_{\text{еф}}$. Припустивши, що ВФН і ФН приблизно розподілені на прямокутнику зі сторонами $3T$ і $4\pi f$, згідно з (5.57), (5.58) можемо записати, що $R_{jk \text{ еф}}^2 \cdot 4FT = R_{\text{еф}}^2 \cdot 4FT = 1$. Звідси знаходимо

$$R_{\text{еф}} = R_{jk \text{ еф}} = R_{j \text{ еф}} = \frac{1}{2\sqrt{FT}} = \frac{1}{2\sqrt{B}}. \quad (5.59)$$

Із (5.59) видно, що чим більша база сигналу, тим менші ефективні значення ВФН і ФН. Формули (5.57)–(5.59) дуже важливі. Оцінка ефективного значення (5.59) збігається за формою з (5.50), але має визначений коефіцієнт, рівний 1/2. Як буде ясно з наступного матеріалу, оцінка (5.59) дає нижню границю, тобто найменше ефективне значення, оскільки отримана за умови рівномірного розподілу ВФН і ФН на частотно-часовій площині. Насправді ж розподіл цих функцій для реальних сигналів є нерівномірним, і тому ефективні значення ВФН і ФН перевищуватимуть знайдені за формулою (5.59).

Інтегральні рівності. Для визначення оцінок КФ широко використовують інтегральні рівності, що пов'язують між собою КФ різних сигналів. Однією з загальних інтегральних рівностей є

$$\int_{-\infty}^{\infty} R_{ml}(\tau_1, \Omega_1) R_{jk}^*(\tau_1, \Omega_1) e^{-iz\tau} d\tau = \int_{-\infty}^{\infty} R_{kl}(\tau, z) R_{jm}^*(\tau, z) e^{-i\Omega_1\tau} d\tau \quad (5.60)$$

(надалі індекс l буде опущений).

З формули (5.60) можна вивести окремі інтегральні рівності. Розглянемо їх.

А. Візьмемо $j=k=l=m$, $z=0$. Маємо рівність Бакулева

$$\int_{-\infty}^{\infty} |R(\tau, \Omega)|^2 d\tau = \int_{-\infty}^{\infty} |R(\tau)|^2 e^{-i\Omega\tau} d\tau. \quad (5.61)$$

Середня потужність сигналу ФН у перетині $\Omega = \text{const}$ є перетворенням Фур'є від квадрата АКФ.

Б. Візьмемо $j=m$, $k=l$, $z=0$. Маємо рівність Сталдера – Канна

$$\int_{-\infty}^{\infty} |R_{jk}(\tau, \Omega)|^2 d\tau = \int_{-\infty}^{\infty} R_j(\tau) R_k^*(\tau) e^{-i\Omega\tau} d\tau.$$

В. Візьмемо $j=m$, $k=l$, $z=\Omega=0$. Маємо

$$\int_{-\infty}^{\infty} |R_{jk}(\tau)|^2 d\tau = \int_{-\infty}^{\infty} |R_j(\tau) R_k^*(\tau)| d\tau. \quad (5.62)$$

З (5.62) випливає, що середнє значення квадрата модуля ВКФ сигналів з номерами j і k дорівнює середньому значенню добутку їх АКФ. Позначимо квадрати ефективних значень ВКФ через

$$R_{jk \text{ еф}}^2 = \frac{1}{2T} \int_{-\infty}^{\infty} |R_{jk}(\tau)|^2 d\tau, \quad R_{q \text{ еф}}^2 = \frac{1}{2T} \int_{-\infty}^{\infty} |R_q(\tau)|^2 d\tau, \quad (5.63)$$

де T – тривалість сигналу; $q=j$ або $q=k$. Використовуючи нерівність Буняковського–Шварца, з (5.63) дістаємо

$$R_{jk \text{ еф}}^2 \leq R_{j \text{ еф}} R_{k \text{ еф}}. \quad (5.64)$$

З (5.64) випливає, що для зменшення ефективного значення ВКФ необхідно зменшувати ефективні значення АКФ.

Використання наведених інтегральних рівностей для оцінки КФ буде проілюстровано далі.

5.7. Основні типи ШПС

Частотно-модульовані сигнали (ЧМ-сигнали) є неперервними, частота яких змінюється за заданим законом. На рис. 5.18, *а* зображений ЧМ-сигнал, частота якого змінюється за V-подібним законом від $f_0 - F/2$ до $f_0 + F/2$, де f_0 – центральна (несуча) частота сигналу; F – ширина спектра, яка дорівнює девіації частоти: $F = \Delta F$. Тривалість сигналу – T .

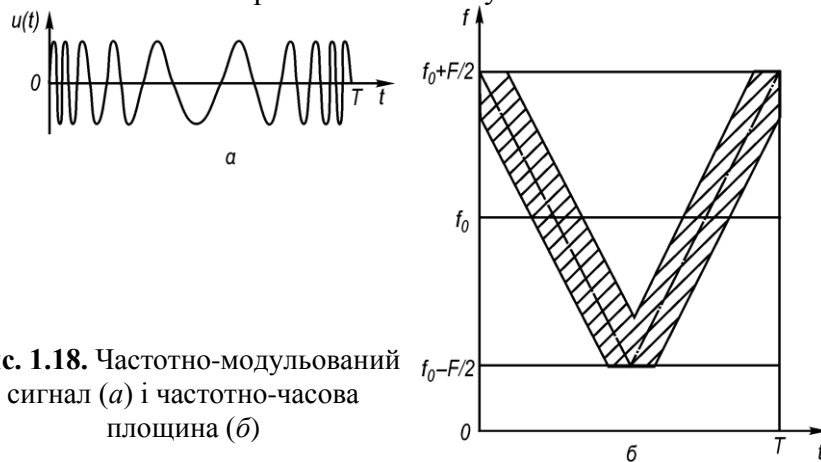


Рис. 1.18. Частотно-модульований сигнал (*а*) і частотно-часова площа (*б*)

На рис. 5.18, *б* зображена частотно-часова (f, t) площа, на якій заштриховано приблизний розподіл енергії ЧМ-сигналу за частотою і за часом. База ЧМ-сигналу

$$B = FT = \Delta f_d T, \quad (5.65)$$

де Δf_d – девіація частоти.

Частотно-модульовані сигнали знайшли широке застосування в радіолокаційних системах, оскільки для конкретного ЧМ-сигналу можна створити погоджений фільтр на приладах з поверхневими акустичними хвилями. Для систем зв'язку треба мати безліч сигналів. Необхідність швидкої зміни сигналів і переключення апаратури формування й обробки приводить до того, що реалізація зміни частоти стає дискретною. При цьому від ЧМ-сигналів переходять до БЧ-сигналів.

Багаточастотні сигнали (БЧ-сигнали) (рис. 5.19, *а*) є сумою N гармонік $u_1 t, \dots, u_N t$, амплітуди і фази яких визначаються відповідно до законів формування сигналів. На частотно-часовій площині (рис. 5.19, *б*) штрихуванням виділений розподіл енергії одного елемента (гармоніки) БЧ-сигналу на частоті f_k . Всі елементи (гармоніки) цілком перекривають виділений квадрат зі сторонами F і T . База сигналу B дорівнює площі квадрата. Ширина спектра елемента $F_0 \approx 1/T$. Тому база БЧ-сигналу

$$B = FT = F / F_0 = N, \quad (5.66)$$

тобто збігається з числом гармонік. БЧ-сигнали є неперервними; для їхнього формування й обробки важко пристосувати методи цифрової техніки. Крім цього, вони мають і такі недоліки, як поганий пік-фактор (нерізде вираження амплітуди сигналу, див. рис. 5.19, а) та потреба у великому числі частотних каналів N для одержання великої бази B .

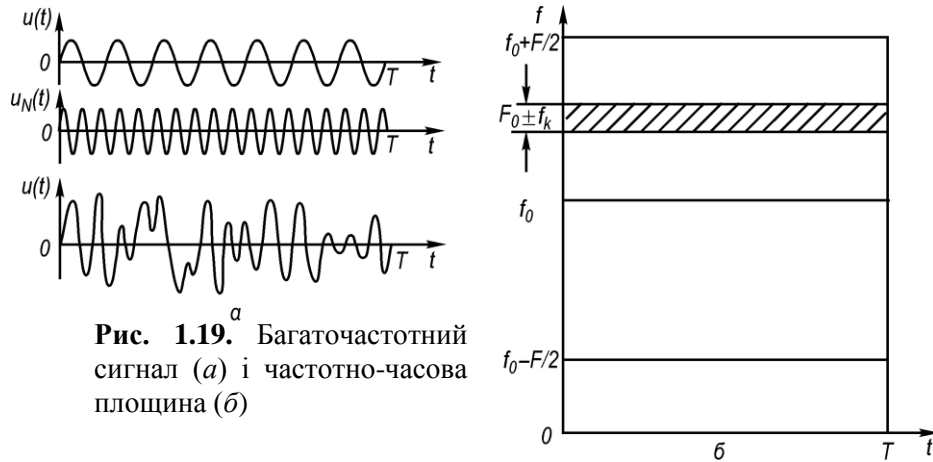


Рис. 1.19. Багаточастотний сигнал (а) і частотно-часова площа (б)

Фазоманіпульовані сигнали (ФМ-сигнали) є послідовністю радіоімпульсів, фази яких змінюються за заданим законом. Звичайно фаза набуває двох значень: 0 або π . При цьому радіочастотному ФМ-сигналові відповідає відео ФМ-сигнал (рис. 5.20, а), що складається з позитивних і негативних імпульсів. Якщо число імпульсів N , то тривалість одного імпульсу дорівнює $\tau_0 = T/N$, а ширина його спектра – приблизно ширині спектра сигналу: $F_0 = 1/\tau_0 = N/T$. На частотно-часовій площині (рис. 5.20, б) штрихуванням виділений розподіл енергії одного елемента (імпульсу) ФМ-сигналу. Всі елементи перекривають виділений квадрат зі сторонами F і T . База ФМ-сигналу

$$B = FT = T / \tau_0 = N, \quad (5.67)$$

тобто дорівнює числу імпульсів у сигналі.

Можливість застосування ФМ-сигналів у якості ШПС із базами $B = 10^4 \dots 10^6$ обмежена в основному апаратурою обробки. При використанні погоджених фільтрів у вигляді приладів на поверхнево-акустичних хвилях можливий оптимальний прийом ФМ-сигналів з максимальними базами $B_{\max} = 1000 \dots 2000$. ФМ-сигнали, оброблювані такими фільтрами, мають широкі спектри (порядку $10 \dots 20$ МГц) і відносно короткі тривалості ($50 \dots 100$ мкс). Обробка ФМ-сигналів за допомогою відеочастотних ліній затримки при перенесенні спектра сигналів в область відеочастот дозволяє одержувати бази $B = 100$ при $F \approx 1$ МГц, $T \approx 100$ мкс.

Досить перспективними є погоджені фільтри на приладах із зарядовим зв'язком (ПЗЗ). Згідно з опублікованими даними, за допомогою погоджених фільтрів ПЗЗ можна обробляти ФМ-сигнали з базами $10^2 \dots 10^3$ при тривалостях сигналів $B = 10^{-4} \dots 10^{-1}$ с. Цифровий корелятор на ПЗЗ

Але при використанні кореляторів необхідно в першу чергу вирішити питання про прискорене входження в синхронізм.

Оскільки ФМ-сигнали дозволяють широко використовувати цифрові методи і техніку формування й обробки та реалізовувати відносно великі бази їх, то вони є одним з перспективних видів ШПС.

Дискретні частотні сигнали (ДЧ-сигнали) є послідовністю радіоімпульсів (рис. 5.21, а), несучі частоти яких змінюються за заданим законом. Нехай число імпульсів у ДЧ-сигналі дорівнює M , тривалість імпульсу $T_0 = T/M$, його ширина спектра $F_0 = 1/T_0 = M/T$. Над кожним імпульсом (рис. 5.21, а) зазначена його несуча частота. На частотно-часовій площині (рис. 5.21, б) штрихуванням виділені квадрати, в яких розподілена енергія імпульсів ДЧ-сигналу. Як видно з рис. 5.21, б, енергія ДЧ-сигналу розподілена нерівномірно на частотно-часовій площині.

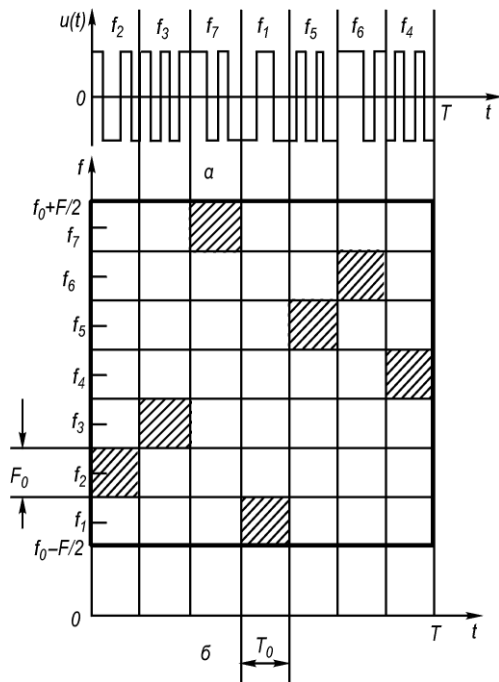


Рис. 5.22. ДСЧ-ФМ-сигнал (а) і частотно-часова площина (б)

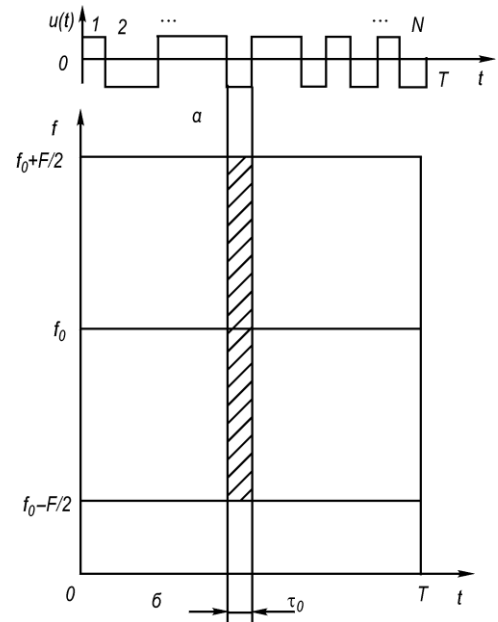


Рис. 5.20. Фазоманіпульований сигнал з кодовою фазовою модуляцією (а) і частотно-часова площина (б)

База ДЧ-сигналів здатний обробляти сигнали до бази $4 \cdot 10^4$.

Слід зазначити, що ФМ-сигнали з великими базами доцільно обробляти за допомогою кореляторів (на великих інтегральних схемах або на ПЗЗ). При цьому $B = 4 \cdot 10^4$ є граничною базою.

$$B = FT = MF_0 \cdot MT_0 = M^2 F_0 T_0 = M^2, \quad (5.68)$$

оскільки база імпульсу $F_0 T_0 = 1$. Із 5.68) випливає основне достоїнство ДЧ-сигналів: для

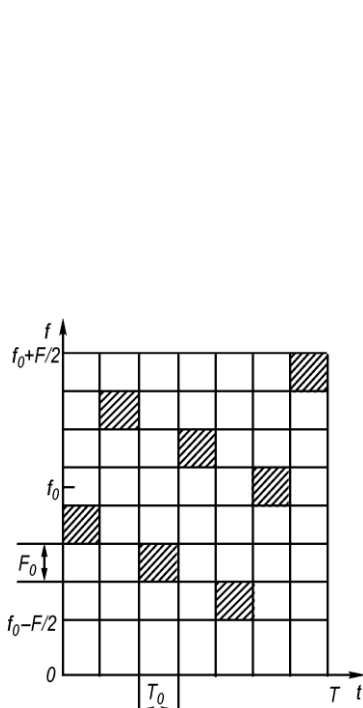


Рис. 5.23. Частотно-часова площина дискретного складеного сигналу з частотною маніпуляцією (ДСЧ-ЧМ-сигналу)

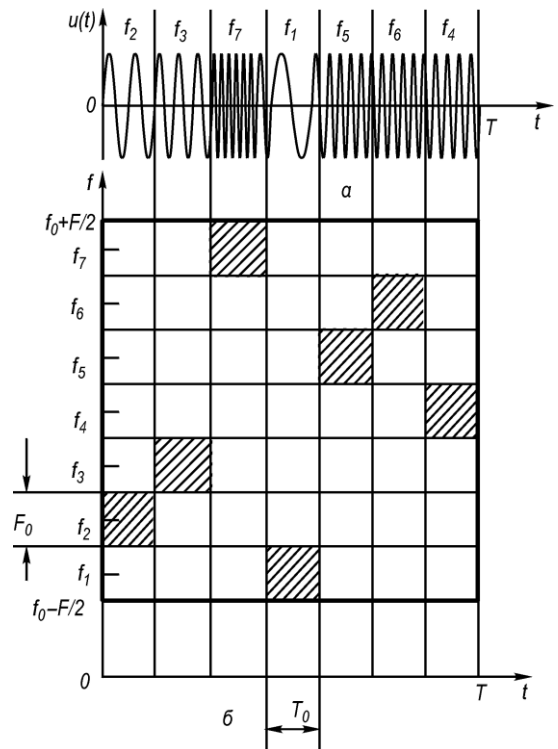


Рис. 5.25. Дискретний сигнал з кодовою частотною модуляцією (а) і частотно-часова площина (б)

одержання необхідної бази B число каналів $M = \sqrt{B}$, тобто значно менше, ніж для БЧ-сигналів. Саме це й обумовило застосування таких сигналів у системах зв'язку. Разом з тим для великих баз $B = 10^4 \dots 10^6$ використовувати тільки ДЧ-сигнали недоцільно, тому що число частотних каналів $M = 10^2 \dots 10^3$, тобто дуже велике.

Дискретні складені частотні сигнали (ДСЧ-сигнали) є ДЧ-сигналами, кожен імпульс яких замінений шумоподібним сигналом. На рис. 5.22, *а* зображений відеочастотний ФМ-сигнал, окремі частини якого передаються на різних несучих частотах. Номери частот зазначені над ФМ-сигналом. На рис. 5.22, *б* зображена частотно-часова площина, на якій штрихуванням виділено розподіл енергії ДСЧ-сигналу. Рис. 5.22, *б* за структурою не відрізняється від рис. 5.21, *б*, але на першому з них площа $F_0 T_0 = N_0$ дорівнює числу імпульсів ФМ-сигналу в одному частотному елементі ДСЧ-сигналу. База ДСЧ-сигналу

$$\begin{aligned} B &= FT = \\ &= M^2 F_0 T_0 = N_0 M^2. \end{aligned} \quad (5.69)$$

Число імпульсів повного ФМ-сигналу

$$N = N_0 M. \quad (5.70)$$

Зображений на рис. 5.22 ДСЧ-сигнал містить як елементи ФМ-сигнали. Тому такий сигнал скорочено називатимемо ДСЧ-ФМ-сигналом. За елементи ДСЧ-сигналу можна взяти ДЧ-сигнали. Розподіл енергії такого сигналу на частотно-часовій площині зображено на рис. 5.23. Якщо база елемента ДЧ-сигналу

$$B_0 = F_0 T_0 = M_0^2, \quad (5.71)$$

то база всього сигналу

$$B = M_0^2 M^2. \quad (5.73)$$

Такий сигнал можна скорочено позначати ДСЧ-ЧМ. Число частотних каналів у ДСЧ-ЧМ-сигналі дорівнює $M_0 M$. Якщо ДЧ-сигнал (рис. 5.21) і ДСЧ-ЧМ-сигнал мають рівні бази, то вони мають і однакове число частотних каналів. Тому особливих переваг ДСЧ-ЧМ-сигнал перед ДЧ-сигналом не має. Але принципи побудови ДСЧ-ЧМ-сигналу можуть виявитися корисними при побудові великих систем ДЧ-сигналів.

Таким чином, найбільш перспективними ШПС для систем зв'язку є ФМ-, ДЧ-, ДСЧ-ФМ-сигнали.

СПИСОК ЛІТЕРАТУРИ, РЕКОМЕНДОВАНОЇ ДЛЯ ПОГЛИБЛЕНОГО ВИВЧЕННЯ ДИСЦИПЛІНИ

1. Беркман Л.Н., Комарова Л.О., Чумак Н.С. Системи електрозв'язку та сигнали. Навчальний посібник, частина 1-4 підготовлено для самостійної роботи студентів вищих навчальних закладів – Київ: ДУТ ННІТІ, 2015.
2. Банкет В. Л., Дорофеев В. М. Цифровые методы в спутниковой связи. – М.: Радио и связь, 1988. – 448 с.
3. Барковський В. В., Барковська Н. В., Лопатін О. К. Математика для економістів: Теорія ймовірностей та математична статистика. – К.: Нац. акад. управління, 1997. – 225 с.
4. Баскаков С. И. Радиотехнические цепи и сигналы: Учеб. для вузов. – М.: Высш. шк., 1998. – 448 с.
5. Большаков И. А., Репин В. Г. Вопросы нелинейной фильтрации // Автоматика и телемеханика. – 1961. – № 4. – С. 53–57.
6. Варакин Л. Е. Системы связи с шумоподобными сигналами. – М.: Радио и связь, 1985. – 384 с.
7. Варакин Л. Е. Теория систем сигналов. – М.: Сов. радио, 1978. – 304 с.
8. Варакин Л. Е. Теория сложных сигналов. – М.: Сов. радио, 1970. – 375 с.
9. Вентцель Е. С. Теория вероятностей. – М.: Наука, 1969. – 576 с.
10. Витерби А. Д., Омура Д. К. Принципы цифровой связи и кодирования. – М.: Радио и связь, 1982. – 526 с.
11. Волочий Б.Ю. Передавання сигналів у інформаційних системах. Ч.1.: Нав. Посібник. – Львів: Видавництво Національного університету «Львівська політехніка», 2005. -196 с.
12. Джури Э. Импульсные системы автоматического регулирования / Пер. с англ.; Под ред. Я. З. Цыпкина. – М.: Мир, 1963. – 453 с.
13. Диксон Р. К. Широкополосные системы / Пер. с англ.; Под ред. В. И. Жуковского. – М.: Связь, 1979. – 303 с.
14. Зайцев Г. Ф., Стеклов В. К., Брицький О. І. Теорія автоматичного управління. – К.: Техніка, 2002. – 688 с.
16. Зюко А. Г. Помехоустойчивость и эффективность систем связи. – М.: Связь, 1973. – 359 с.
17. Зюко А. Г., Коробов Ю. Ф. Теория передачи сигналов: Учеб. для вузов. – М.: Связь, 1972. – 282 с.
18. Игнатов В. А. Теория информации и передачи сигналов: Учеб. для вузов. – М.: Сов. радио, 1979. – 280 с.
19. Кларк Дж., Кейн Дж. Кодирование и исправление ошибок в системах цифровой связи / Пер. с англ.; Под ред. Б. С. Цибакова. – М.: Радио и связь, 1987. – 392 с.
20. Кловский Д. Д. Теория передачи сигналов. – М.: Связь, 1972. – 282 с.
21. Кловский Д. Д. Теория передачи сигналов в задачах. – М.: Связь, 1978. – 252 с.
22. Кловский Д. Д., Шилкин В. А. Теория электрической связи: Сб. задач и упражнений. – М.: Радио и связь, 1990. – 280 с.
23. Коновалов Г. Ф. Радиоавтоматика: Учеб. для вузов. – М.: Высш. шк., 1990. – 335 с.

24. *Котельников В. А.* Теория потенциальной помехоустойчивости. – Л.: Госэнергоиздат, 1956. – 152 с.
25. *Маркюс Ж.* Дискретизация и квантование. – М.: Энергия, 1969. – 286 с.
26. *Назаров М. В., Кувшинов Б. И., Попов О. В.* Теория передачи сигналов. – М.: Связь, 1970. – 368 с.
27. *Панфілов І. П., Дирда В. Ю., Капацін А. В.* Теорія електричного зв'язку. – К.: Техніка, 1998. – 322 с.
28. *Первачев С. В.* Радиоавтоматика. – М.: Радио и связь, 1982. – 403 с.
29. *Передача дискретных сообщений: Учеб. для вузов / В. П. Шувалов, Н. В. Захарченко, В. О. Шварцман и др.; Под ред. В. П. Шувалова.* – М.: Радио и связь, 1990. – 464 с.
30. *Поляков П. Ф.* Широкополосные аналоговые системы связи со сложными сигналами. – М.: Радио и связь, 1981. – 153 с.
31. *Пугачев В. С.* Статистические методы в технической кибернетике. – М.: Сов. радио, 1971. – 192 с.
32. *Радиоавтоматика: Учеб. для вузов / Г. Ф. Зайцев, Г. Н. Арсеньев и др.* – М.: ООО “ДВК”, 2004. – Т. 1. – 523 с.
33. *Системы фазовой синхронизации / С. Н. Складенко, В. К. Стеклов, Р. В. Уваров, В. М. Чмилъ.* – К.: Техніка, 1994. – 160 с.
34. *Складенко С. М., Стеклов В. К., Беркман Л. Н.* Поштовий зв'язок. – К.: Техніка, 2004. – 904 с.
35. *Стратонович Р. Л.* К теории оптимальной нелинейной фильтрации случайных функций // Теория вероятностей и ее применение. – 1959. – № 2. – С. 44–49.
36. *Солодов А. В.* Теория информации и ее применение к задачам автоматического управления и контроля. – М.: Наука, 1967. – 436 с.
37. *Стилкер Дж.* Цифровая спутниковая связь / Пер. с англ. В. В. Маркова. – М.: Связь, 1979. – 592 с.
38. *Статистическая динамика радиотехнических следящих систем / С. В. Первачев, А. А. Валуев, В. М. Чиликин.* – М.: Сов. радио, 1973. – 488 с.
39. *Стеклов В. К., Беркман Л. Н.* Проектування телекомунікаційних мереж. – К.: Техніка, 2002. – 792 с.
40. *Стеклов В. К., Беркман Л. Н.* Теорія електричного зв'язку. . – К.: Техніка, 2006. – 552 с.
41. *Теория передачи сигналов: Учеб. для вузов / А. Г. Зюко, Д. Д. Кловский, М. В. Назаров, Л. М. Финк.* – М.: Связь, 1986. – 304 с.
42. *Теория электрической связи: Учеб. для вузов / А. Г. Зюко, Д. Д. Кловский, В. И. Коржик, М. В. Назаров.* – М.: Радио и связь, 1998. – 432 с.
43. *Тихонов В. И.* Нелинейная фильтрация и квазиэлектронный характер фазовой автоподстройки частоты // Изв. АН СССР. Техн. кибернетика. – 1965. – № 2. – С. 44–49.
44. *Туркин А. И.* Рекуррентный прием сложных сигналов (на основе метода вхождения и решения безостановочных экстремальных задач). – М.: Радио и связь, 1988. – 248 с.
45. *Финк Л. М.* Теория передачи дискретных сообщений. – М.: Сов. радио, 1970. – 727 с.
46. *Філіпський Ю. К.* Випадкові сигнали в радіотехніці. – К.: Вища шк., 1986. – 126 с.

47. Хлытчиев С. М., Тарасова Н. П., Лившиц В. М. Теоретические основы почтовой связи: Учеб. для вузов. – М.: Радио и связь, 1990. – 380 с.
48. Цыпкин Я. З. Основы теории автоматических систем. – М.: Наука, 1967. – 448 с.
49. Шеннон К. Работы по теории информации и кибернетике / Пер. с англ.; Под ред. Н. А. Железнова. – М.: Мир, 1963. – 829 с.