УДК 621.313

ПОЛТОРАК В.П.,
ВІТІЩЕНКО Н.С.

ВПЛИВ ОСНОВИ КОДУ НА ЕФЕКТИВНІСТЬ НАДЛИШКОВОГО КОДУВАННЯ ДАНИХ

Досліджений вплив основи надлишкових кодів на ефективність кодування у системах передавання даних. Обґрунтована верхня межа для основи коду, за якої досягається максимальна ефективність кодування. Обґрунтовано перспективи дослідження та використання кодів з короткою довжиною.

A code base impact on encoding efficiency was investigated for data transmission systems. Upper edge for code base is validated for encoding efficiency maximization. A short codes usage was grounded.

Відомості про навколишній світ передаються різноманітними каналами у формі дискретних символів – даних. Під час передавання у певному середовищі (металеве, оптичне, радіоефір, тощо) сигнали, що переносять дані, можуть бути спотворені завадами різної природи. Відповідно, дані на приймальному боці можуть не співпадати із тими, що були передані – виникають помилки. Виправлення помилок виконують за допомогою надлишкових (завадостійких) кодів.

Природна надлишковість є невід'ємною властивістю висхідних повідомлень (текстів, файлів, тощо). Будемо вважати, що джерело повідомлень містить лише безнадлишкові дані, як результат кодування висхідних повідомлень оптимальним нерівномірним кодом (ОНК), що до певної міри знімає їх природну надлишковість. Саме такі дані є незахищеними від спотворень у каналі передавання.

Захист їх надлишковим кодом передбачає цілеспрямоване, «дозоване» введення (додавання) надлишковості до повідомлень від джерела: наприклад, рівно на стільки, скільки потрібно для «виявлення до $\mathbf{t_в}$ можливих помилок», або ж – «виправлення до $\mathbf{t_{вп}}$ можливих помилок», тощо, що припадає на певну кількість **n** символів кодованих повідомлень. Звичайно, таке надлишкове кодування має сенс у разі, коли «дозована», обґрунтована кодом надлишковість менша за природну надлишковість висхідних повідомлень. У даній статті предмет дослідження обмежимо блоковими (**n, k**)-кодами, де **n** – довжина закодованого блоку даних, **k** – кількість інформаційних символів від джерела, які треба захистити від спотворень. Надлишковість зазвичай виступає у формі певної кількості **r** надлишкових символів коду, причому: $\mathbf{n = k + r}$.

Мірою втрат за умови надлишкового кодування (надалі – коду, кодування) є відносна надлишковість (надалі - надлишковість) $\mathbf{D = r/n}$. Її віддзеркаленням є відносна швидкість коду $\mathbf{R = k/n}$ (надалі – швидкість коду) – своєрідний технічний ефект від кодування. Вочевидь, що $\mathbf{D + R = 1}$, а самі **D** і **R** мають смисл питомих величин.

Мірою технічного ефекту, що досягається є, також, кількість помилок $\mathbf{t_в}$, або $\mathbf{t_{вп}}$. Доцільніше обрати за міру технічного ефекту питому величину, наприклад, $\mathbf{x = t_{вп}/n}$.

Поєднавши дві, визначені таким чином, характеристики, **R** і **x** оберемо відомий критерій оцінки ефективності кодів $\mathbf{R = f(x)}$ – межу для мінімальної кодової відстані, для оцінки ефективності кодів з різним значенням основи **q** , або – потужності алфавіту коду [1, 2]. Зазвичай, межу для мінімальної кодової відстані визначають при великих значеннях **n**, а сам критерій задається параметрично, через параметри кодів. Його зображують графічно на площині (**R;x**) і він є геометричним місцем точок **K**(**R;x**) з відповідними координатами. Вигляд **x** можна трохи модифікувати, врахувавши значення мінімальної кодової відстані **d** за Хемінгом: $\mathbf{d \geq 2\, t_{вп} + 1}$, до форми $\mathbf{x = d/2n}$, за умови, що **n** – велике. Різниця між дозволеними кодовими комбінаціями (надалі – КК) визначається мінімальною кодовою відстанню **d** коду, при тому, що вся множина КК розбита на дві підмножини – дозволених та заборонених комбінацій однієї довжини **n** [2].

Коди мають велику кількість параметрів, що їх характеризують, тому критерій ефективності кодів можна сформулювати одним із трьох способів [1]:

1. серед кодів з однаковими **n** і **d** кращим є код, який має більшу **k** (за однакових $\mathbf{x = d/2n}$);

2. серед кодів з однаковими **n** і **k** кращим є код, який має більшу **d (**за однакових **R);**

3. серед кодів з однаковими **d** і **k** кращим є код, який має меншу **n (**менше **r).**

Ці критерії органічно відповідають аналізу ефективності кодів на площині **R = f(x).**

В даній статті проведемо дослідження впливу основи коду **q** на ефективність надлишкового кодування даних, при $2 \le q < \infty$ за критерієм **(R;x).**

На рис.1 зображена класична структура системи передачі даних, що складається з джерела повідомлень, двох відносно самостійних пристроїв: Кодеку та Модему, каналу зв'язку та отримувача повідомлень. В рамках даної статті обговорюємо вхід Кодера та вихід Декодера.

За великого обсягу алфавіту сигналів, кожен символ може переносити набагато більше інформації ніж за **q = 2**. Зокрема, алфавіт, що містить **q** символів дозволяє передавати **I=log₂q** двійкових одиниць інформації на кожен символ.

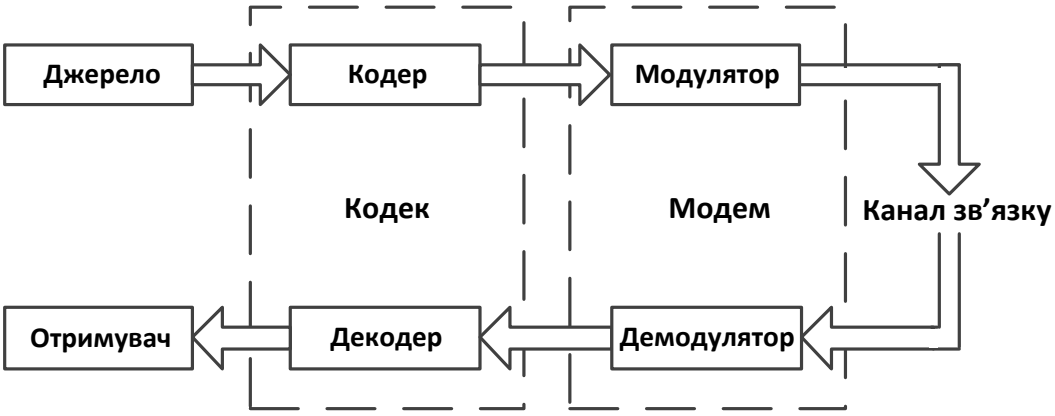

Рис. 1. Модель системи передачі даних

При фіксованих значеннях **n** та **k**, можна отримати верхню та нижню межі для найбільшої **d**. Розглянемо ці межі.

Відповідно до

$$\frac{tв}{n} = \frac{d-1}{2n} \approx \frac{d}{2n}, \qquad (1)$$

найкращі коди мають велику коректувальну здатність $\mathbf{t_в}$. Ці граничні залежності і є межами (верхніми або нижніми) для **d**.

Верхня межа Плоткіна є точною для низькошвидкісних кодів. У [1,2] вона записана у вигляді:

$$k \le n - \frac{qd-1}{q-1} + 1 - \log_q d. \qquad (2)$$

Поділивши (2) на n та спрямувавши n до $n \to \infty$, отримаємо нерівність [3]:

$$R \le 1 - \frac{2q}{q-1}\frac{d}{2n}, \qquad (3)$$

в якій відкинуто доданки вищих ступенів малості.

Верхні межі Хемінга і Елайеса, а також нижня межа Варшамова – Гілберта пов'язані з нерівністю Чернова [1], де використовується функція вигляду

$$\varphi(x) = x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x). \qquad (4)$$

Верхня межа Хемінга визначає максимальну можливу кількість дозволених кодових комбінацій $\mathbf{q^k}$ будь-якого коду при заданих значеннях **n** та **d.** В [2] межа Хемінга має наступний вигляд:

$$R \le 1 - \varphi\left(\frac{t_в}{n}\right). \qquad (5)$$

Врахувавши (1) та $n \to \infty$, дістанемо

$$R \le 1 - \varphi\left(\frac{d}{2n}\right) = 1 - \varphi(x), \qquad (6)$$

де $x = d/2n$.

Включивши до (6) вираз (4), матимемо рівняння межі Хемінга:

$$R = 1 - x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x). \qquad (7)$$

Область значень для **x** визначається виразом (7), з якого випливає межа Хемінга. Якщо збільшувати **q**, то лінія межі Хемінга піднімається вище над початковою межею, при **q = 2**, за рахунок збільшення **R**. При $q = \infty$, отримаємо асимптотичний вираз межі при великих значеннях основи **q**, тобто

$$R \le 1 - x. \qquad (8)$$

Верхня межа Елайеса, для середньо швидкісних кодів з **q** символами, визначається виразом [2]

$$\frac{d}{n} < \delta(R)\left[2 - \frac{\delta(R)q}{q-1}\right] + \varepsilon, \qquad (9)$$

де $\delta(R)$ – розв'язок рівняння $\varphi(x) = 1 - R$, утвореного з (5), а $\varphi(x)$ – функція (4); $\varepsilon > 0$ – як завгодно мала складова.

За допомогою межі Хемінга (7) та позначивши $x = \delta(R)$ дістаємо парні значення $x = \delta(R)$ та **R**. Після цього для кожного **R** за відомим $x = \delta(R)$ розраховуємо значення $d/2n$ межі Елайеса, користуючись виразом (9), перетвореним до вигляду

$$\frac{d}{2n} < \frac{1}{2}\delta(R)\left[2 - \frac{q}{q-1}\delta(R)\right] + \varepsilon . \quad (10)$$

Відповідно виразу (10), будуємо лінії для межі Елайеса при $2 < q < \infty$.

Нижня межа Варшамова-Гілберта – для великих значень **n** встановлює нижню межу для числа **r** перевірних розрядів, необхідного для забезпечення заданої **d**. Нижня межа оцінює число **r** при заданих **k** та **d**.

Межа Варшамова-Гілберта в [2] визначається виразом

$$q^{n-k} \le \sum_{i=0}^{d-2} C_n^i (q-1)^i , \quad (11)$$

де C_n^i – біномні коефіцієнти. Виключимо їх із розгляду, скориставшись асимптотичною оцінкою цих коефіцієнтів у вигляді нерівності Чернова

$$\sum_{i=0}^{t_в} C_n^i (q-1)^i \le q^{n\varphi(x)}, \quad (12)$$

де $\varphi(x)$ – функція виду (4).

З урахуванням (12), прологарифмуємо обидві частини нерівності (11) і знайдемо

$$n - k \le n\varphi(\frac{d-2}{n}) ,$$

звідки $R = \frac{k}{n} \ge 1 - \varphi(\frac{d}{n} - \frac{2}{n})$.

Спрямувавши **n** до $n \to \infty$, дістаємо асимптотичну форму нижньої межі Варшамова – Гілберта

$$R \ge 1 - \varphi(\frac{d}{n}) . \quad (13)$$

Порівнявши (13) з межею Хемінга (6), можна остаточно записати перетворений вираз (7) для межі Варшамова – Гілберта

$$R \ge 1 + \log_q \left[(\frac{2x}{q-1})^{2x} (1-2x)^{(1-2x)} \right] . \quad (14)$$

Завдяки цим межам, можна визначити області можливих значень параметрів надлишкових кодів та отримати конкретні числові значення для кожного з досліджуваних кодів при різних значеннях $\mathbf{t_в}$.

Оцінимо вплив **q** на ефективність ряду відомих надлишкових кодів: Хемінга, БЧХ та Ріда–Соломона, використавши отримані вирази критеріїв для мінімальної кодової відстані.

Код **Хемінга** має мінімальну кодову відстань **d** = 3 та здатен виправити поодинокі помилки в межах переданого блоку [1]. Мінімальне співвідношення коректувальних та інформаційних розрядів, нижче якого код не зберігає свої коректувальні властивості, визначається як $\mathbf{q^r - 1 = n}$ [1, 3].

Для порівняльного аналізу, з основних параметрів коду Хемінга були обрані: основа коду **q**, **r**, **n** та розраховані **k**, **x**, **R** та **D.**

Нижче наведено приклад одного з варіантів розрахунку параметрів для коду Хемінга з великим **n** при **q** = 4:

Оберемо **r** = 6, тоді:

- $n = \frac{q^r - 1}{q-1} = \frac{4^6 - 1}{4-1} = 1365$
- $k = n - r = 1365 - 6 = 1359$
- $x = \frac{tв}{n} = \frac{1}{1365} = 0.0007$
- $R = \frac{k}{n} = 0.996$
- $D = \frac{r}{n} = \frac{6}{1365} 100\ \% = 0.43\ \%$

У Таблиці 1 наведено параметри коду Хемінга, розраховані для різних **q**: 4, 8, 16, 32, 64, з метою побачити, як впливає зміна **q** на ефективність коду.

Табл.1. Параметри кодів Хемінга

$\mathbf{t_в = 1}$					
q	**4**	**8**	**16**	**32**	**64**
n	1365	4681	4369	1057	4161
x	0.0007	0.00021	0.00023	0.0009	0.00024
R	0.996	0.9989	0.9991	0.997	0.9992

Недвійкові коди **БЧХ** будуються за допомогою твірних поліномів P(x), які визначаються за заданими $\mathbf{d_{min}}$ та **n**. Найпоширенішим кодом БЧХ є код, для побудови якого застосовуються поле елементів GF(q) та розширене поле локаторів $GF(q^m)$. При побудові кодів БЧХ завжди забезпечується мінімально досяжна надмірність $r = 2mt_в$.

Для порівняльного аналізу, з основних параметрів коду БЧХ були обрані наступні: **q**, $\mathbf{t_в}$, **m**, $\mathbf{n|(q^m-1)}$, **x**, **r** , $k = n - r$, **D** та **R.**

Нижче наведено приклад одного з варіантів розрахунку параметрів для коду БЧХ, з виправленням 1 помилки при **q = 4**. Оберемо, наприклад, **m =** 5, тоді:

- $n : q^m - 1 = q^5 - 1 = 1023$
- $x = \frac{tв}{n} = \frac{1}{1023} = 0.0009$
- $r = 2mt_в = 10$
- $k = n - r = 1023 - 10 = 1013$
- $R = \frac{k}{n} = 0.99$
- $D = \frac{r}{n} = \frac{10}{1023} = 0.97\,\%$

У Таблиці 2 наведено параметри коду БЧХ, розраховані при різних **q**: 4, 8, 16, 32, 64 та кількості виправлених помилок $\mathbf{t_в}$= 1, 2 та 3, з метою побачити, як впливає зміна **q** на ефективність коду.

Табл.2. Параметри коду БЧХ

$\mathbf{t_в}$ = 1					
q	**4**	**8**	**16**	**32**	**64**
n	1023	4095	4095	1023	4095
x	0.0009	0.0002	0.0002	0.0009	0.00024
R	0.99	0.998	0.998	0.996	0.999
$\mathbf{t_в}$ = 2					
q	**4**	**8**	**16**	**32**	**64**
n	1023	4095	4095	1023	4095
x	0.0019	0.0005	0.0005	0.0019	0.0005
R	0.98	0.996	0.997	0.992	0.998
$\mathbf{t_в}$ = 3					
q	**4**	**8**	**16**	**32**	**64**
n	1023	4095	4095	1023	4095
x	0.0029	0.0007	0.0007	0.0029	0.0007
R	0.971	0.994	0.996	0.988	0.997

Коди **Ріда-Соломона** використовуються для передачі інформації по каналах з високою інтенсивністю завад, коли виникають помилки кратності два та більше, пачки помилок. Цей код є частковим випадком коду БЧХ і в ньому елементи та локатори знаходяться в одному скінченому полі Галуа.

Для порівняльного аналізу, з основних параметрів коду Ріда-Соломона були обрані наступні: **q**, $\mathbf{t_в}$, **n** = q - 1, $r = 2t_в$ та $k = n - r$, **x**, **R** та **D**.

Нижче наведено приклад одного з варіантів розрахунку параметрів для коду Ріда-Соломона, з виправленням 2 помилок при **q = 8**:

- $n = q - 1 = 7$
- $r = 2t_в = 4$
- $k = n - r = 3$
- $x = \frac{tв}{n} = \frac{2}{7} = 0.29$
- $R = \frac{k}{n} = 0.43$
- $D = \frac{r}{n} = \frac{4}{7} = 0{,}57\,\%$

У Таблиці 3 наведено параметри коду Ріда-Соломона, розраховані при різних **q**: 8, 16, 32, 64 та кількості виправлених помилок $\mathbf{t_в}$= 1, 2 та 3, з метою побачити вплив **q** на ефективність коду.

Табл.3. Параметри коду Ріда-Соломона

$\mathbf{t_в}$ = 1					
q	**4**	**8**	**16**	**32**	**64**
n	3	7	15	31	63
x	0.33	0.143	0.066	0.032	0.016
R	0.33	0.71	0.86	0.935	0.97
$\mathbf{t_в}$ = 2					
q	**4**	**8**	**16**	**32**	**64**
n		7	15	31	63
x		0.29	0.133	0.064	0.063
R		0.428	0.733	0.87	0.94
$\mathbf{t_в}$ = 3					
q	**4**	**8**	**16**	**32**	**64**
n		7	15	31	63
x		0.43	0.2	0.096	0.048
R		0.14	0.6	0.81	0.91

При подібних розрахунках, доводиться погодитися зі зміною довжини КК.

Отримані точки **K(R;x)** з відповідними координатами для кодів Хемінга, БЧХ та Ріда-Соломона з різною кількістю виправлених помилок нанесемо на графіки разом з теоретичними межами, отриманими вище.

На рис.2 зображені верхні межі Хемінга та нижні Варшамова-Гілберта при **q = 2** та **q = ∞.**

Межа Хемінга – одна з верхніх меж для **d**, зміст якої в тому, що не існують коди кращі за ті, що відображаються верхньою межею, тобто немає коду, координати якого перевищили б відповідні координати будь-якої точки верхньої межі.

Теоретична крива для межі Хемінга побудована за формулами (7) та (8). Теоретична крива для нижньої межі Варшамова–Гілберта на всіх наступних графіках побудована за виразом (14).

На графіку Рис. 2 видно, що для кодів, зі збільшенням відносної кількості помилок, що виправляються на довжину КК, швидкість коду значно зменшується.

Так, для коду Ріда-Соломона ($\mathbf{t_в}$ = 1)

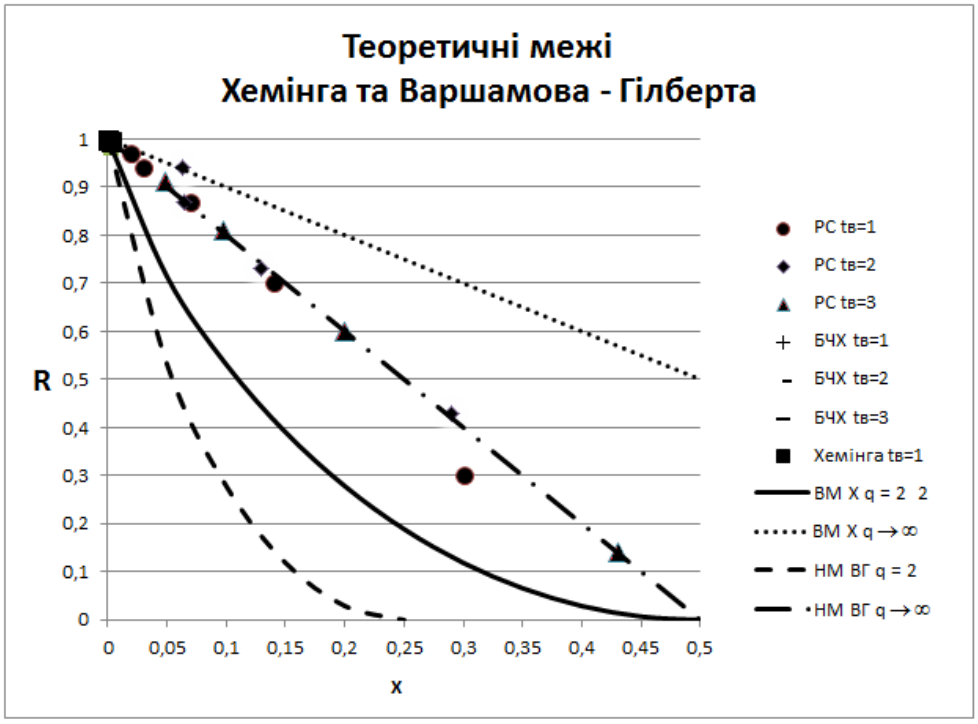

Рис. 2. Межі Хемінга та Варшамова-Гілберта з кодовими точками

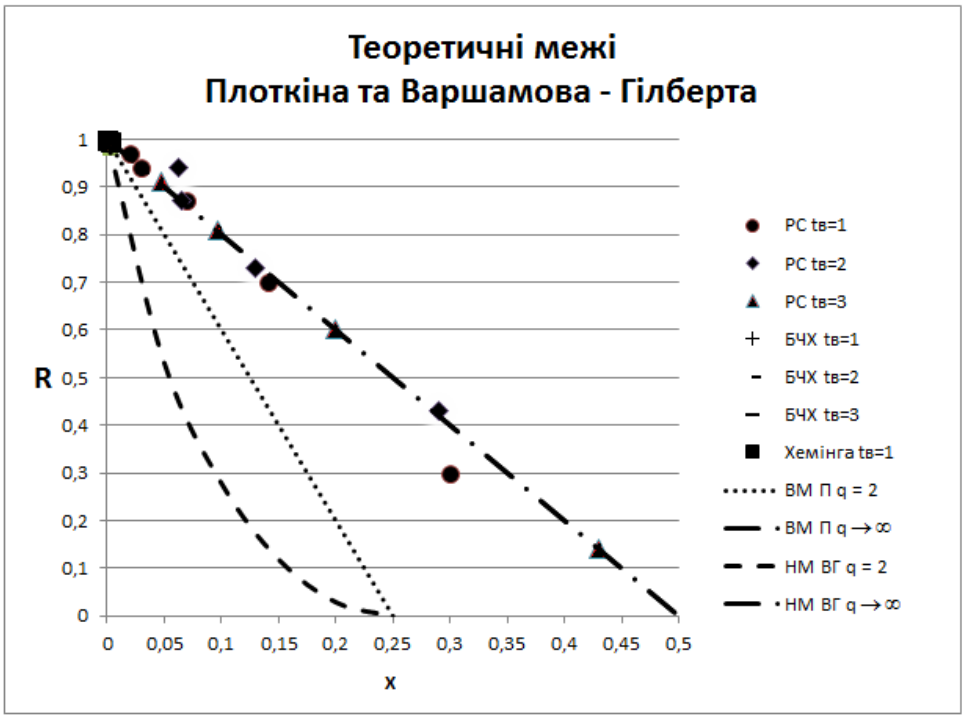

Рис. 3. Межі Плоткіна та Варшамова-Гілберта з кодовими точками

q = 4 **D** = r/n = 2/3 = 66,6%
q = 8 **D** = r/n = 2/7 = 28,5%
q = 16 **D** = r/n = 2/15 = 13,3%
q = 32 **D** = r/n = 2/31 = 6,45%
q = 64 **D** = r/n = 2/63 = 3,17%

В результаті бачимо, що при відносно постійній довжині кодової комбінації та зі збільшенням основи коду q зменшується надлишковість.

Збільшення основи коду **q** призводить до зміщення точок в сторону покращення характеристик ефективності коду.

На Рис. 3 зображені верхні межі Плоткіна та нижні Варшамова-Гілберта при **q = 2** та **q** = ∞**.**

Теоретична крива для верхньої межі Плоткіна побудована за формулою (3). З графіка випливає, що зі збільшенням **q** збільшується ефективність кодування. Так, при фіксованому

значенні **x**, збільшення **q** приводить до зростання швидкості **R** і зменшення надлишковості **D**.

Так, наприклад, для коду Хемінга

q = 4 **D** = r/n = 6/1365 = 0.43%

q = 8 **D** = r/n = 5/4681 = 0.12%

q = 16 **D** = r/n = 4/4369 = 0.09%

q = 32 **D** = r/n = 3/1057 = 1.1%

q = 64 **D** = r/n = 3/4161 = 0.072%

На рис.4 зображені верхні межі Елайєса та нижні Варшамова-Гілберта при **q = 2** та **q = ∞.**

Порівняння розташування точок для конкретних кодів із ходом кривих меж для мінімальної кодової відстані на всіх рисунках показує гарну відповідність точок межам при відповідних значеннях основи коду **q**.

Видно, що точки для конкретних кодів добре збігаються із розташуванням теоретичних кривих для однакових значень основи коду **q**.

Видно, що збільшення основи **q** призводить до суттєвого збільшення ефективності за **x** та **R**

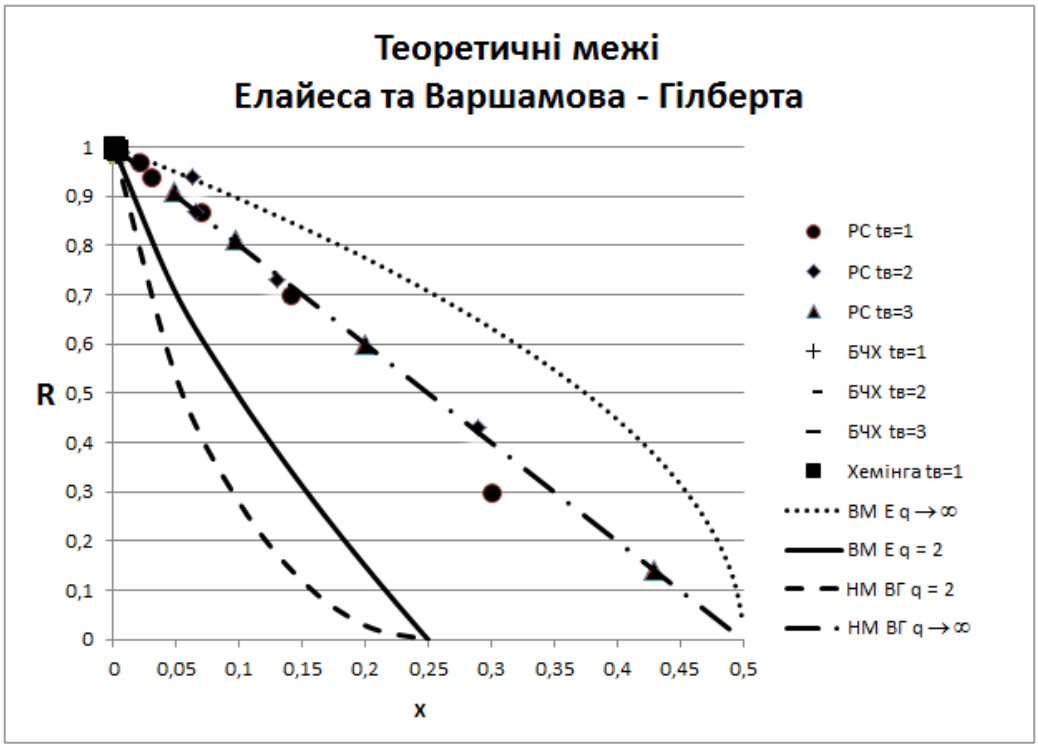

Рис. 4. Межі Елайеса та Варшамова-Гілберта з кодовими точками

Теоретична крива межі Елайеса побудована за виразами (7) та (10). Додатково на графіку нанесені точки, що відповідають ряду кодів із параметрами, розрахованими вище.

На рис.5 зображені верхні межі Хемінга та нижні Варшамова-Гілберта при **q = 2** та **q = ∞** в більшому масштабі, щоб покращити подання точок високошвидкісних кодів.

Так, наприклад, для коду БЧХ ($\mathbf{t_в}$ = 1)

q = 4 **D** = r/n = 10/1023 = 0.97%

q = 8 **D** = r/n = 8/4095 = 0.19%

q = 16 **D** = r/n = 6/4095 = 0.15%

q = 32 **D** = r/n = 4/1023 = 0.39%

q = 64 **D** = r/n = 4/4095 = 0.097%

Зі збільшенням основи коду спостерігається значне зменшення надлишковості для усіх трьох кодів. Але, звернувши увагу на відносну швидкість коду, бачимо, що для коду БЧХ – вона найбільша, і зі збільшенням **q** суттєво не змінюється.

надлишкових кодів незалежно від конкретного коду чи типу межі для мінімальної **d**.

Більш докладний аналіз ефективності кодів показує, що прирости ефективності при переході від **q** = 2 до більших значень **q** швидко скорочуються і практично стають нерозрізненними при **q** ≥ 32 від випадку **q** = ∞.

Задля збільшення ефективності надлишкових кодів, немає сенсу збільшувати основу **q** до нескінченності, оскільки максимальні значення ефективності, що характерні для випадку **q** = ∞, досягаються практично вже при **q** ≥ 32. Знайдене тут обмеження для **q** може слугувати певним «маяком» при конструюванні кодеків систем передавання даних, оскільки спрямування $q \to \infty$ означає необмежене збільшення складності апаратури та програмного забезпечення її реалізації, що не можна вважати за бажане. І,

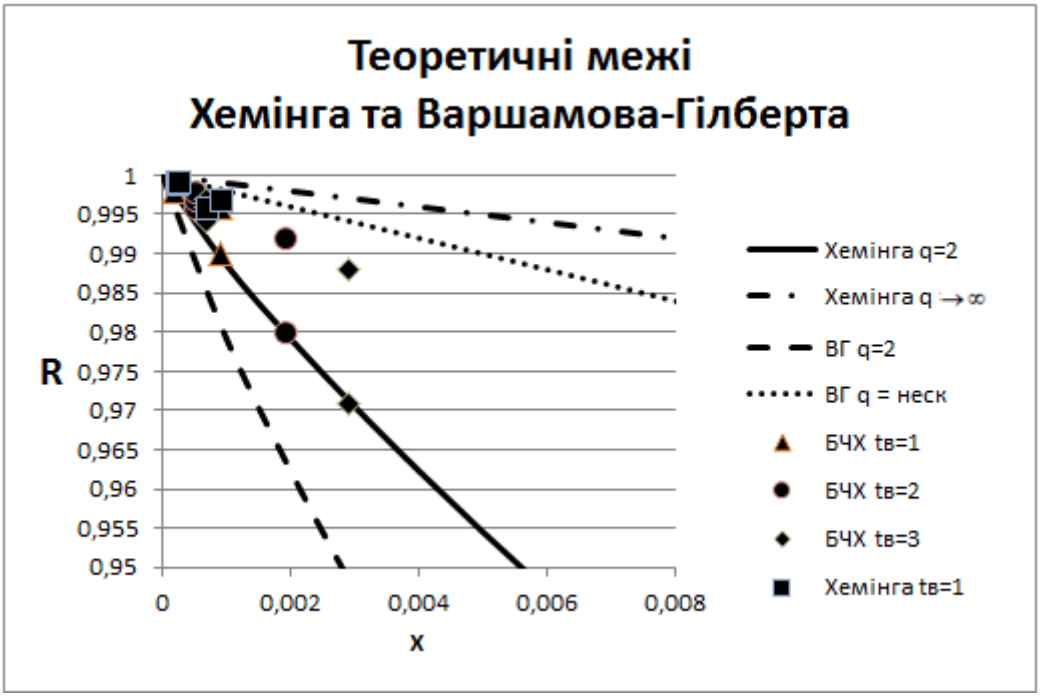

Рис. 5. Межі Хемінга та Варшамова-Гілберта з кодовими точками в збільшеному масштабі

навпаки, обґрунтоване обмеження основи коду **q** ≤ 32 дозволяє досягти максимальних значень ефективності кодів за умови менш витратних характеристик складності апаратури та програмного забезпечення кодеків систем передавання даних.

Детальний аналіз впливу **q** на характеристики надлишкових кодів дозволяє виявити ще одну неочевидну властивість. Мова йде про довжину коду **n**.

Дотепер ще існує відомий «штамп» у мисленні як математиків, так і інженерів «кодовиків» про те, що найбільш ефективними (за критерієм **D**) є «довгі» коди [1, 2] із великими **n**. Однак наше дослідження показує, що це є справедливим лише за умови фіксованого значення **q** = const. Так, до речі, і сформувалося свого часу наведене уявлення про «довгі» коди, як найбільш ефективні. Бо і до сього часу переважна більшість кодів, що використовуються на практиці є суто двійковими з **q** = 2.

Проведене в даній роботі дослідження показує, що високих показників ефективності надлишкового кодування можна досягти і при відносно невеликих довжинах **n** за рахунок збільшення основи коду **q**>2.

Це відкриває перспективи дослідження та використання відносно «коротких» кодів, для яких **n** є співставними з **q**.

Перелік літератури

1. Т. Касами, Н. Токура, Е. Ивадари, Я. Инагаки Теория кодирования. - М.: Мир, 1978. – 576с.
2. У. Питерсон , Э. Уэлдон Коды, исправляющие ошибкию - М.: Мир, 1976. – 590с.
3. Жураковський Ю.П., Полторак В.П. Теорія інформації та кодування: Підручник. – К.: Вища шк., 2001. – 255с.:іл.