Національний університет “Львівська політехніка”

СЕГІН Андрій Ігорович

УДК 62.001.57:681.518

# МОДЕЛІ ДЖЕРЕЛ ІНФОРМАЦІЇ ТА МЕТОДИ ЇХ ФОРМУВАННЯ В СИСТЕМАХ РЕАЛЬНОГО ЧАСУ

01.05.02 – математичне моделювання та обчислювальні методи

Автореферат дисертації на здобуття
наукового ступеня кандидата технічних наук

Львів – 2001

Дисертацією є рукопис
Робота виконана в Івано-Франківському державному технічному університеті нафти і газу, Міністерство освіти і науки України

Науковий керівник – доктор технічних наук, професор
Николайчук Ярослав Миколайович,
Тернопільська академія народного господарства, завідувач кафедри "Спеціалізовані комп'ютерні системи"

Офіційні опоненти: – доктор фізико-математичних наук, професор, Яворський Ігор Миколайович
Фізико-механічний інститут ім. Г.В. Карпенка
НАН України, завідувач відділом відбору й обробки стохастичних сигналів.

– кандидат технічних наук, старший науковий співробітник, Алішов Надір Ісмаіл-Огли,
Інститут кібернетики ім. В.М.Глушкова
НАН України, ведучий науковий співробітник відділу управляючі машини і системи.

Провідна установа – Державний науково-дослідний інститут інформаційної інфраструктури Державного комітету зв'язку та інформації і НАН України, відділ інформаційних технологій і систем, м. Львів.

Захист відбудеться "29" червня 2001р. о $16^{\underline{00}}$ годині на засіданні спеціалізованої вченої ради Д 35.052.05 у Національному університеті "Львівська політехніка" (79013, Львів-13, вул. С. Бандери, 12, ауд. 226 головного корпусу)

З дисертацією можна ознайомитися у бібліотеці Національного університету "Львівська політехніка" (м. Львів-13, вул. Професорська, 1)

Автореферат розісланий "28" травня 2001 року.

Вчений секре т ар
спеціалізованої вченої ради Ткаченко С.П.

# ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність. Розробка моделей джерел інформації (ДІ), які обслуговуються розподіленими автоматизованими системами є актуальною задачею. Широкому використанню таких моделей сприяє швидкий розвиток комп'ютерної техніки та технічних засобів формування, обробки, передавання та відображення даних. Разом з цим, в проблемах моделювання ставляться нові задачі, які полягають у застосуванні більш складних комплексів моделей для різнотипних об'єктів керування (ОК), з метою підвищення якості та швидкості обробки даних, отримання повнішої інформації, покращення ергономічних характеристик та багато інших.

На сьогодні є розроблено ряд моделей джерел інформації, введених зарубіжними вченими, серед яких найбільш видатні: Шенон, Хартлі, Фрітч і вченими країн СНГ: Колмогоровим, Кузьміном, Боюном, Николайчуком, які відображають статистичні, кореляційні, спектральні, ентропійні та інші властивості і системні характеристики ОК. В той же час, практично, відсутня узагальнена система моделей джерел інформації, яка б базувалася на перспективних принципах глибокого розпаралелення процесів побудови моделей, формалізованих конвеєрних процедур їх формування. Дана задача особливо актуальна при побудові моделей ДІ розподілених систем керування реального часу.

Розробка та реалізація узагальненої системи моделей ДІ дозволяє суттєво підвищити функціональні можливості програмно-інформаційного забезпечення комп'ютерних систем збору, обробки, передавання даних та керування складними об'єктами, що в свою чергу знижує вартість комп'ютерних систем, собівартість організації руху даних, підвищує рівень достовірності ідентифікації станів об'єктів керування, а, в загальному, підвищує рівень надійності розподілених систем контролю та керування.

Таким чином створення узагальненої системи інформаційних моделей об'єктів керування, їх систематизація, дослідження властивостей і способів формування моделей ДІ в різних дискретних базисах та розробка нових моделей, які забезпечують розширення функціональних можливостей програмно-інформаційного забезпечення розподілених систем є актуальною науковою задачею.

Зв'язок роботи з науковими програмами, планами, темами. Розробка інформаційних моделей, дослідження їх властивостей і методів побудови, а також методів формування та цифрової обробки даних в дискретних базисах здійснювались в межах науково-дослідних робіт, що проводились Івано-Франківським державним технічним університетом нафти і газу (ІФДТУНГ):

— НДР "Розробка методологічних основ (теорії, моделей, алгоритмів, процедур і технічних засобів) діагностування і автоматизованого управління об'єктами нафтогазового комплексу України". (Отримано формули що ви-

ражають оцінку ентропії через кореляційні функції).

— Держбюджетна тема Д-6-Ф № держреєстрації 01340001278 “Розробка теоретичних основ вертикальної інформаційної технології та вдосконалення методів автоматизованого контролю розподілених об’єктів нафтогазовидобувних підприємств” ТЗ ИГАЦ–537–98. (Розроблено методи кодування знакозмінних сигналів в інтегральній інформаційній технології, методи приведення процесів зі спадаючими ділянками до процесів, які не мають від’ємних приростів та побудовано інформаційну модель багатоканального джерела інформації в інтегральній технології.)

Мета і задачі дослідження. Мета роботи полягає в теоретичному обґрунтуванні та створенні системи моделей джерел інформації, методів їх побудови, орієнтованих на формування і відображення в системах реального часу; розроблення алгоритмів розпаралелення побудови кореляційних моделей та пакету програм діагностування станів об’єктів керування на основі теорії інформації, статистичного і спектрального аналізу.

Основні задачі дослідження:

— аналіз відомих моделей джерел інформації, їх властивостей та методів побудови;
— порівняння ефективності методів формування даних в різних базисах та цифрової обробки інформації про стани об’єктів керування, які описуються моделями джерел інформації;
— розроблення теоретичних основ формування даних для побудови моделей об’єктів керування;
— розроблення структурних схем розпаралелених алгоритмів побудови моделей джерел інформації;
— розроблення програмного забезпечення обробки даних в розподілених системах, що реалізують розроблені та досліджені моделі ДІ .

Об’єкт дослідження: об’єкти керування як джерела інформації.

Предмет дослідження: моделі джерел інформації.

Методи дослідження. При виконанні поставлених задач використовувалися методи та алгоритми цифрової обробки сигналів, статистичного і спектрального аналізу, апарат дискретної математики, теорія інформації, теорії множин та графів, математичне моделювання з використанням елементів теорії полів Галуа, комп’ютерне моделювання.

Наукова новизна одержаних результатів полягає в наступному:

— одержали подальший розвиток кореляційні моделі джерел інформації, шляхом узагальнення їх властивостей та синтезом нових аналітичних виразів кореляційних оцінок;
— вперше введено модель глобальної дисперсії, побудованої на основі інформативних елементів матриці коефіцієнтів кореляції багатомірних ДІ, та, розробленої на її основі, п’ятої логіко-статистичної інформаційної моделі,

які дозволяють ідентифікувати відхилення складних багатоканальних ДІ на основі визначення середньої дисперсії кореляційних зв'язків між процесами в каналах ДІ та характеризуються мінімальною надлишковістю даних в класі розроблених моделей;

— теоретично обґрунтовано ентропійні моделі джерел інформації, які базуються на кореляційній, структурній, модульній, полярній, знакові та еквівалентній функціях кореляції, що дозволяє враховувати статистичні взаємозв'язки між станами ДІ;

— розроблена ідентифікаційна модель двовимірних зображень на основі власних функцій, яка дозволяє спростити алгоритми розпізнавання графічних зображень, за рахунок відсутності операцій масштабування.

Практичне значення одержаних результатів полягає в тому, що за результатами проведених автором теоретичних досліджень розроблено алгоритми та програмне забезпечення обробки даних в розподілених комп'ютерних системах, які дозволяють контролювати стани складних об'єктів керування за багатьма параметрами в реальному масштабі часу. Практичну цінність мають такі результати:

— для підвищення швидкодії формування моделей, що особливо важливо в систем реального часу, розроблені розпаралелені алгоритми побудови моделей, які є основою для створення програмного забезпечення багатопроцесорних обчислювальних середовищ або процесорів паралельної дії.

— розроблено структурні схеми формувачів логіко-статистичних інформаційних моделей, апаратна реалізація яких на низовому рівні дозволяє суттєво зменшити надлишковість даних (у 2 рази і більше разів), що значно розвантажує лінії передавання, приймально-передавальні модулі та обчислювальні центри на вищих рівнях розподілених комп'ютерних систем

— розроблено алгоритмічне та програмне забезпечення обробки інформації на основі досліджених в роботі моделей;

— удосконалено структуру багатоканального АЦП Галуа, який здійснює перетворення аналогових сигналів в цифрові коди в базисі Галуа. Удосконалення полягає у зменшенні в 2 рази кількості компараторів у порівнянні з відомою структурою АЦП такого типу.

Теоретичні і методологічні положення побудови інформаційних моделей, що досліджувались в дисертаційній роботі використані при дослідженні та аналізі аварійних та передаварійних станів об'єктів електроенергетики, викладанні курсу "Автоматизовані системи управління", і курсовому проектуванні на кафедрі автоматизованого управління в технічних та організаційних системах Івано-Франківського державного технічного університету нафти і газу та при викладенні курсу "Низові комп'ютерні системи" на кафедрі спеціалізованих комп'ютерних систем Тернопільської академії народного господарства..

Особистий внесок здобувача полягає в наступному:

— розроблено модель глобальної дисперсії, на основі якої створено п’яту логіко-статистичну модель, та проаналізовано їх властивості [1];
— створено систему класифікації моделей джерел інформації, яка найбільш повно охоплює відомі класи інформаційних моделей, і розроблено рекомендації щодо їх використання на різних рівнях ієрархії систем управління [2, 6];
— досліджено властивості кореляційних моделей і розроблені графи розпаралелених алгоритмів їх формування [3-4];
— проаналізовано час затримки аналого-цифрового перетворення різними класами багатоканальних АЦП і розроблено вдосконалену структурну схему АЦП Галуа [4];
— введено ентропійні моделі, які базуються на структурній, модульній та еквівалентній функціях кореляції [5];
— обґрунтовано моделі розпізнавання образів на основі кореляційних оцінок та власних функцій [7-8].

Апробація результатів дисертації. Основні наукові результати і теоретичні положення дисертаційної роботи доповідались та отримали схвалення на VI міжнародній науково-практичній конференції “Проблеми створення, інтеграції і використання науково-технічної інформації на сучасному етапі” – Київ, 1999; XIV міжнародній міжвузівській школі-семінарі “Методи і засоби технічної діагностики” – Івано-Франківськ, 1999; 3-ій міжнародній конференції комп’ютерні технології друкарства: алгоритми, сигнали, системи “Друкотехн –2000” – Львів, 2000; науково-технічній конференції професорсько-викладацького складу Івано-Франківського державного технічного університету нафти і газу – Івано-Франківськ, 1999.

Публікації по роботі. Результати проведених в роботі досліджень відображені в 8 публікаціях: 5 статтях у фахових журналах і тез 3 доповідей; 4 праці опубліковані одноосібно.

Структура і об’єм дисертації. Дисертація складається з вступу, чотирьох розділів, списку використаних джерел та додатків. Загальний об’єм дисертації 186 сторінок, 68 рисунки на 56 сторінках, 3 таблиці на 7 сторінках, 4 додатки на 38 сторінках. Список використаних джерел містить 86 найменувань.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтована актуальність теми дисертації, сформульовані мета та задачі досліджень, наукова новизна і практична значимість отриманих результатів. Наведені дані про реалізацію та впровадження результатів роботи, її апробацію та публікації.

У першому розділі здійснено огляд існуючих інформаційних моделей, які розроблені на базі теорії ймовірності, теорії інформації, теорії цифрової обробки сигналів, теорії розпізнавання образів та інших теоретичних основах. В результаті проведеного аналізу здійснена класифікація інформаційних мо-

делей, в якій найбільш повно охоплено інформаційні моделі різних класів (рис. 1).

Рис. 1. Класифікація моделей джерел інформації.

Формалізація інформаційних моделей, досліджених в дисертації, здійснена на основі виразу, який в загальній формі має вигляд:

$$y_i = Op\{z_{l-n}, K, z_{l-1}, x_{i1}, x_{i2}, K, x_{im}\};$$

де $\overline{O}$ – деяка послідовність внутрішніх станів системи, від яких залежить образ $y_i$; $l$ – номер поточного стану; $n$ – максимально віддалений від поточного стан системи, який впливає на результат перетворення.

Проаналізовано також інформаційні технології, які відіграють важливу роль при організації автоматизованих систем та способі формування дискрет-

них даних та їх представлення, на основі яких будуються інформаційні моделі. В інформаційній технології, в якій передаються не самі значення, а різниця між поточним і попереднім дискретним відліком запропоновано здійснювати кодування приростів бітами Галуа, що забезпечує підвищену завадостійкість в порівнянні з бінарною, тернарною та показниковою системами кодування. При умові, що прирости можуть бути або додатними або рівними нулю, кодування відбувається наступним чином: $G \to +1,\ \overline{G} \to 0,$ де $\overline{G}$ – інвертований біт Галуа. Кодування приростів в базисі Галуа довільних вхідних аналогових сигналів потребує інших способів.

В бінарній системі позитивний приріст кодується бітом Галуа $G$, а від'ємний – інвертованим бітом Галуа $\overline{G}$; відсутність приросту (нульовий приріст) – послідовністю прямих та інвертованих бітів Галуа $G\,\overline{G}$. При тернарній Галуа системі прирости кодуються двобітовою реверсивно-кодовою послідовністю: $G\,G \to +1;\ \overline{G}\,\overline{G} \to 0;\ \overline{G}\,G \to -1;$ або змішаним способом: $G1 \to +1;\ \overline{G}0 \to 0;\ G0 \to -1.$

<u>У другому розділі</u> досліджено такі типи кореляційних функцій: $R_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} \overset{0}{x}_i \cdot \overset{0}{x}_{i+j}$ – кореляційна; $K_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot x_{i+j}$ – коваріаційна; $\rho_{xx}(j) = \frac{R_{xx}(j)}{D_x}$ – нормована кореляційна; $H_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{sgn}[\overset{0}{x}_i] \cdot \mathrm{sgn}[\overset{0}{x}_{i+j}]$ – знакова; $P_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \mathrm{sgn}[\overset{0}{x}_{i+j}]$ – полярна; $C_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} (x_i - x_{i+j})^2$ – структурна; $G_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} \left| x_i - x_{i+j} \right|$ – модульна; $\overset{\vee}{F}_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} \overset{\vee}{z}_i[x_i, x_{i+j}]$ – еквівалентності; $\overset{\vee 2}{F}_{xx}(j) = \frac{1}{n}\sum_{i=1}^{n} \overset{\vee 2}{z}_i(x_i, x_{i+j})$ – квадратична еквівалентності, де $\overset{\circ}{x}_i = x_i - M_x$ – центроване значення; $\mathrm{sgn}[x_i] = \begin{cases} 1, x_i > 0; \\ 0, x_i = 0 \\ -1, x_i < 0 \end{cases}$ – знакова функція; $\hat{z}_i[x_i, x_{i+j}] = \hat{z}_i = \begin{cases} x_i, \text{ якщо } x_i \ge x_{i+j}, \\ x_{i+j}, \text{ якщо } x_i < x_{i+j}, \end{cases}$ – функція менше з двох", та введено нову кореляційну функцію: $Tm_{xx}(j) = \frac{1}{n}\sum_{i}^{n} \overset{\vee}{z}_i \cdot \left| x_i - x_{i+j} \right|$, яка розширює функціональні можливості кореляційної моделі еквівалентності. Отримано аналітичний зв'язок цієї функції з кореляційною, коваріаційною, структурною та модульною. Аналіз кореляційних моделей показує, що найбільшу чутливість

до сигналів з шумами має нововведена модель (рис. 2). Це дозволяє використовувати її для виявлення появи шумів, які порушують нормальний стан ОК.

Рис. 2. Графіки зміни дисперсії кореляційних моделей при збільшенні дисперсії шуму з нормальним законом розподілу.

Проаналізовані спектральні моделі в базисах Фур'є, Радемахера та Крестенсона і показано, що для сигналів з широким спектром в базисі Фур'є, спектральний аналіз зручніше проводити в базисі Крестенсона, оскільки одна базисна функція Крестенсона включає всі кратні гармоніки базису Фур'є.

Розроблені ентропійні моделі, які базуються на кореляційних функціях:

$\overline{H(x_i, x_{i+j}\, x_i)} = \frac{1}{2T}\sum_{j=1}^{T}\log_2(2\pi e)^2\left[C_{xx}(j)\left(D_{0_x} - \frac{C_{xx}(j)}{4}\right)\right]$ – на основі структурної функції;

$\overline{H(x_i, x_{i+j})} = \frac{1}{T}\sum_{j=1}^{T}\log_2\left[\frac{\pi e\sqrt{\pi}}{2}G_{xx}(j)\cdot\sqrt{8-\pi g_{xx}^2(j)}\right]$ – на основі модульної функції, де $g_{xx}(j) = \frac{G_{xx}(j)}{\sigma_{0_x}}$ – нормована модульна функція кореляції;

$\overline{H(x_i, x_{i+j})} = \frac{1}{T}\sum_{j=1}^{T}\log_2\left[\pi^2 e\frac{M_x - F_{xx}(j)}{\sigma_{0_x}}\sqrt{\frac{8\sigma_{0_x}}{\pi} - [M_x - F_{xx}(j)]^2}\right]$ – на основі функції еквівалентності;

$\overline{H(x_i, x_{i+j})} = \frac{1}{T}\sum_{j=1}^{T}\log_2\left[2\pi e\sigma_{0_x}\cdot\sqrt{1-\sin^2\left(\frac{\pi}{2}H_{xx}(j)\right)}\right]$ – на основі знакової функції;

$\overline{H(x_i, x_{i+j})} = \frac{1}{T}\sum_{j=1}^{T}\log_2\left[\frac{2\pi e}{\sigma_{0_x}}\cdot\sqrt{2-\pi\cdot P_{xx}^2(j)}\right]$ – на основі полярної функції.

Даний клас моделей, по відношенню до кореляційних, здійснюють інтегральну оцінку імовірності переходів між станами і завдяки представленню в логарифмічному просторі забезпечують менші об'єми даних. Ентропійна модель, виражена через $\overset{\vee}{F}_{xx}(j)$, обчислюється з більшою швидкодією, завдяки простішій реалізації функції "менше з двох" ніж операцій множення, віднімання, піднесення до квадрату. Ентропійну модель на основі $H_{xx}(j)$ доцільно використовувати для обробки даних ДІ, які формують цифрові сигнали із значеннями +1 і –1.

В порівнянні з ентропією Хартлі і $\varepsilon$-ентропією моделі, виражені через кореляційні функції, враховують статистичні залежності між станами і є більш адекватними. Крім того, їх розрахунок простіший ніж ентропії Шенона за рахунок того, що непотрібно визначати закони розподілу ймовірності.

Для спрощення подання та відображення поведінки складного ДІ запропоновані кластерні моделі, які крім кількісних показників враховують і семантичні значення станів об'єкта $S_i$ та переходів між ними $S_i \to S_j$. Якщо за кількісні показники прийняти імовірність переходів між станами то для ОК з $n$ станами, критичним кількісним коефіцієнтом $\alpha$ і критичним семантичним коефіцієнтом $\beta$, кластеризація здійснюється на основі матриць ймовірностей переходу між станами та семантичних коефіцієнтів:

$$P = \begin{pmatrix} p_{11} & p_{12} & K & p_{1n} \\ p_{21} & p_{22} & K & p_{2n} \\ M & M & M & M \\ p_{n1} & p_{n2} & K & p_{nn} \end{pmatrix}, \qquad Q = \begin{pmatrix} q_{11} & q_{12} & K & q_{1n} \\ q_{21} & q_{22} & K & q_{2n} \\ M & M & M & M \\ q_{n1} & q_{n2} & K & q_{nn} \end{pmatrix}.$$

Таким чином, кластерні моделі можна подати в 4 варіантах: відображення $S_i \to S_j$, для яких $p_{ij} \ge \alpha$; відображення $S_i \to S_j$, для яких $p_{ij} \ge \alpha$ і позначенням семантичних ваг переходів; відображення $S_i \to S_j$, для яких $q_{ij} \ge \beta$ і відображення $S_i \to S_j$, для яких $p_{ij} \ge \alpha$ і $q_{ij} \ge \beta$.

Застосування кластерних моделей дозволяє покращити ергономічне сприйняття великих об'ємів інформації оператором і підвищити швидкість його реагування на відхилення ОК від норми.

Для узагальненого контролю та аналізу багатоканального ОК, в якого аналітичні співвідношення між процесами в каналах не існують або складно знаходяться, але є взаємопов'язаними, розроблена модель глобальної дисперсії. Глобальна дисперсія будується на основі матриці коефіцієнтів взаємокореляції між сигналами в каналах ДІ. Кореляційна матриця для $m$-канального джерела має розмірність $m \times m$ і симетрична відносно головної діагоналі, оскільки $r_{ij} = \begin{cases} r_{ji}, & \text{при } i \ne j, \\ 1, & \text{при } i = j; \end{cases}$

де $r_{ij}$ – коефіцієнт кореляції між процесами в каналах i та j; $i=\overline{1,m}$, $j=\overline{1,m}$. Інформативним елементам, які знаходяться під/над головною діагоналлю присвоюються порядкові номера $s=1,K,N$, де $N=0{,}5m(m-1)$. Тоді глобальна дисперсія обчислюється за виразом: $D_G=\frac{1}{N}\sum_{s=1}^{N} r_s{}^2$.

Для більш повної характеристики стану багатоканального ДІ введено коефіцієнт $p_s$, який враховує різну семантичну значимість степені кореляції між процесами в різних каналах складного ДІ, і, відповідно, семантична глобальна дисперсія визначатиметься за виразом: $D_S=\frac{1}{N}\sum_{s=1}^{N} p_s\cdot r_s{}^2$.

Із введенням семантичних коефіцієнтів ефективність використання значно зростає, особливо на об’єктах з великою різницею семантичної ваги коефіцієнтів кореляції.

На основі розробленої глобальної дисперсії запропоновано п’яту логіко-статистичну інформаційну модель (ЛСІМ 5), яка дозволяє фіксувати в дворівневому просторі недопустимі відхилення глобальної дисперсії за межі встановленої апертури $E5$. Математично модель описується наступним чином:

$$g=\begin{cases}0, & \text{при } D_G\in E5,\\ 1, & \text{при } D_G\notin E5.\end{cases}$$

Проведений аналіз моделей розпізнавання образів дозволив створити нові моделі цього класу, які базуються на оцінці кореляції між еталонними та досліджуваними об’єктами. В загальному вигляді модель описується виразом:

$$X=\frac{1}{l}\sum_{l=1}^{N-1} w_i * (e_l * x_l)^*,$$

де $*$ – ознака операцій множення, віднімання, модулю, піднесення до квадрату, “менше (більше) з двох”; $w_i$ – ваговий коефіцієнт; $e_i$ та $x_i$ – відповідно координати еталонного та досліджуваного об’єкта. Прикладом може служити модель, яка виражається через структурну і модульну функції

$$XCG=\frac{1}{N-1}\sum_{l=0}^{N-1}\left(w_l-|e_l-x_l|\right)^2$$

і є втіленням тільки одного з можливих варіантів. Потрібно відмітити, що образ об’єкту $x_i$ ідентифікується з еталонним образом $e_j$ з групи еталонів $E$, при $X=\max$, якщо використані оцінки $K_{xe}, R_{xe}, \check{F}_{xe}, \check{F}^2_{xe}$ і при $X=\min$, якщо – оцінки $C_{xe}, G_{xe}, \hat{F}_{xe}, \hat{F}^2_{xe}$.

Розроблено метод розпізнавання графічних образів на основі власних функцій, згідно якого вибираються власні функції з довільними фазами, які

утворюють деяку множину $V$. У випадку розпізнавання літер українського алфавіту достатньо п’ять функцій (рис 3, а). На рис. 3, б наведено приклад розкладу літери $H$ за власними функціями та подання її у вигляді вектора:

$$v_H \{V2\ V2\ V2\ V2\ V5\ V2\ V2\ V2\ V2\}.$$

Кожен еталонний образ $e_k$ із множини еталонів $E$ (еталонних зображень літер алфавіту) має свій визначаючий вектор $v_{e_k}$. В результаті сканування деякого графічного об’єкту формується вектор розкладу цього об’єкту по власних функціях. Відповідні координати цього вектора порівнюються з координатами векторів, що описують еталонні образи. При порівнянні значень кожної координати вектора графічного об’єкту з відповідними значеннями координат векторів еталонів, еталонам присвоюється ознаку співпадання $s_i^k$

$$s_i^k = \begin{cases} 1, \text{ якщо } a_i^x = a_i^{e_k}, \\ 0, \text{ якщо } a_i^x \neq a_i^{e_k}; \end{cases}$$

де $k = 1, K$, $m$ – номер еталону; $m$ – загальна кількість еталонів в множині $E$.

Звичайно, ознакам $s_i^k$ можна надати нерівноцінні значення, які б враховували вагу кожної координати.

а)

б)

Рис. 3 а) – Форми дискретних власних функцій; б) – Розклад літери $H$ по власних функціях.

Таким чином, графічний об’єкт ідентифікується з тим еталоном, з яким він має найвищий коефіцієнт співпадання, при цьому немає потреби масштабувати графічний об’єкт. Особливістю методу є те, що він дозволяє ідентифікувати об’єкт з нахилом та нечітко вираженою геометрією.

У третьому розділі розроблені структури алгоритмів формування кореляційних моделей, та узагальненого алгоритму побудови ентропійних моделей та глобальної дисперсії (рис. 4). Запропоновані графи строго визначають структури алгоритмів, на основі яких реалізуються можливості прискорення розрахунку моделей із застосуванням методів розпаралелення, колективного та багаторазового використання проміжних результатів.

Рис. 4. Узагальнений граф розрахунку глобальної дисперсії і ентропії.

Подано методику побудови комплексу інформаційних моделей руху даних, які дозволяють вирішити проблеми узгодження та синхронізації роботи елементів системи і передавання даних між ними в інформаційному плані, тобто дають можливість узгодити всі операції в системі, пов'язані з формуванням, обробкою та передаванням даних. Для визначення економічних затрат на організацію руху даних в АСУ запропонована вартісна модель, яка

формується на основі питомої вартості операцій: $V = V_1 + V_2 + V_3 = \sum_{j=1}^{3} V_j$, де $V_j$ – вартість операцій: 1 – формування, 2 – обробки, 3 – приймання даних; $V_j = \sum_{i=1}^{k_j} e_i \cdot t_i$, де $k_j$ – кількість операцій $j$-го типу, $e_i$ – питома вартість операції за одиницю часу; $t_i$ – час виконання операції.

Розроблені методи формування і представлення знакозмінних сигналів в інтегральній інформаційній технології (ІІТ) в базисі Галуа, особливістю якої є інкрементна зміна відліків (в класичному варіанті) і біт орієнтоване подання інформації кодами Галуа.

Перший метод полягає у використанні двохбітової реверсивно-кодової шкали Галуа. В цьому методі кожен дискретний відлік кодується двома бітами, де перший біт приймає значення прямого біта Галуа $G$ при зростанні інтегрального значення на один квант, при цьому другому біту присвоюється інвертоване значення $\overline{G}$. Якщо інтегральне значення зменшується на один квант, то другий біт рівний $G$, а перший встановлюється в $\overline{G}$. Таким чином, перші біти складають пряму послідовність, яка відображає тільки зростання інтегрального значення $S^+$, а другі біти утворюють реверсивну послідовність, яка характеризує тільки спадання інтеграла параметра $S^-$. Тоді миттєве значення при такому способі кодування визначається різницею між інтегральними значеннями за виразом: $x_i = S_i^+ - S_{i-1}^+ + S_{i-1}^- - S_i^-$

Другий метод полягає у використанні подвійного інтеграла параметра процесу. Проте цей спосіб придатний для процесів, які змінюються за синусоїдальною функцією або деякими іншими функціями, в яких перший інтеграл приймає невід'ємні значення.

Третій метод полягає у приведенні відліків до діапазону невід'ємних значень, шляхом додавання константи $C$, що вибирається рівною модулю найменшого від'ємного значення, яке може приймати сигнал та інтегрування результуючого сигналу.

Найбільш ефективний спосіб формування даних у вигляді кодів Галуа – асинхронний, який полягає в тому, що біти Галуа генеруються тільки в моменти часу зміни інтегрального сигналу на один квант. При такому способі генерується ланцюг кодів Галуа (рис. 5), який забезпечує завадозахищеність даних та зменшення їх об'єму.

Для визначення інтегральних значень сигналу в дискретний момент $i$, в кожному каналі зчитуються останні $n$ біт, де $n$ – вибрана розрядність коду.

Рис. 5. Кодування даних в багатоканальних ОК.
Д – давач; ПФС – перетворювач форми сигналу.

У четвертому розділі запропонована оцінка впливу старіння інформації на адекватність кореляційних моделей при використанні різних типів аналого-цифрових перетворювачів. Для нормованої кореляційної моделі така оцінка буде мати вигляд: $\gamma_c = \frac{1}{M}\sum_{k=1}^{M}\left(r_{xy_k} - r^*_{xy_k}\right)^2$, яка по суті, виражає дисперсію відхилення. Здійснено аналіз різних типів багатоканальних аналого-цифрових перетворювачів та запропоновано структурну схему модифікованого багатоканального АЦП Галуа.

Для зменшення надлишковості даних, які передаються для обробки та аналізу на вищі рівні автоматичних та автоматизованих систем, запропоновані структурні схеми формувачів ЛСІМ 1-4, які можна реалізувати апаратними засобами і використовувати на низовому рівні обчислювальних мереж. На рис. 6 показано характеристики стиснення даних досліджених моделей і їх залежність від кількості каналів та розрядності відліків, з яких видно, що модель глобальної дисперсії та ЛСІМ 5 характеризуються найбільшими коефіцієнтами стиснення даних.

З метою підвищення швидкості формування статистичних моделей розроблено методи, застосування яких показано на прикладі побудови кореляційної оцінки. Перший метод полягає в тому, що кореляційну оцінку обчислюють з різним кроком дискретних зсувів $j = 0, 1, K, m$, $j = 0, 2, 4, K, 2m$, $j = 0, 4, 8, K, 4m$ і т. д., поки дозволяє об'єм $N$ масиву даних. В загальному вигляді таку процедуру можна записати виразами: $R_{xx}(j) = \frac{1}{n_1}\sum_{i=1}^{n_1} \overset{0}{x}_i \overset{0}{x}_{i+j}$, де $i = \overline{1, n_1}$; $j = (0, 1, K, m)\cdot 2^k$, $k = \overset{\vee}{E}\left[\log_2 \frac{N - n_1}{m}\right]$ – кількість повторень процедури. Таким чином отримаємо $k$ оцінок, в результаті усереднення яких отримаємо кінцеву оцінку.

Другий метод побудови ковзних статистичних моделей, який забезпечує вищу швидкодію обчислення, відрізняється від попереднього тим, що $i = (0, 1, K, n_1)\cdot 2^k$ та $k = \overset{\vee}{E}\left[\log_2 \frac{N}{n_1 + m}\right]$.

а)

б)

Рис. 6. Графіки залежності коефіцієнта стиснення даних від розрядності відліків – а) та кількості каналів складних ДІ – б) в різних класах моделей джерел інформації.

Приведені методи забезпечують прискорення обчислення ковзних статистичних моделей при їх побудові для значних зсувів в часі.

Здійснено аналіз способів відображення сигнальних моделей та їх формалізація: $CM[N, Cl, m, Sp, D]$, де $N$ – кількість рівнів квантування ($N = 2$ – дворівнева модель, $N = n$ – багаторівнева); $Cl$ – спектр кольорів ($Cl = bw$ – чорно-білий, $Cl = col$ – кольоровий); $m$ – вимірність представлення інформації ($m = 1$ – в одномірному просторі, $m = 2$ – в двомірному просторі, $m = 3$ – тримірному просторі); $Sp$ – спосіб представлення інформації ($Sp = s$ – структури, графи і схеми, $Sp = d$ – діаграми і гістограми, $Sp = g$ – графіки, $Sp = t$ – таблиці, $Sp = bs$ – блок-схеми, $Sp = iw$ – інформаційні вікна); $D$ – динамічність ($D = dyn$ – динамічні, $D = st$ – статичні). Таке формальне представлення дає можливість емпірично оцінити ефективність різних способів відображення даних моделями ДІ.

На прикладі резервуару, як джерела інформації, подано методики побудови комплексу квазістаціонарних моделей, які включають часову модель, впоряд-

ковані часові моделі по станах та по змінах, вибіркової часової моделі по станах. На основі запропонованих моделей розраховуються статистичні характеристики математичне сподівання $M_x$, дисперсія $D_x$, кореляційна функція $R_{xx}(j)$, спектральна характеристика $S(\omega)$ кожного стану, взаємокореляційні зв'язки між ними, імовірності його знаходження в кожному стані: $p_i = \frac{T_i}{T}$, $T_i$ – час перебування ОК в стані $S_i$; $T$ – загальний час спостереження. Квазістаціонарні моделі є також основою для побудови добових рапортів технологічних станів ОК, матриці ймовірностей переходу між ними та кластерних моделей.

Комплекс розроблених та досліджених моделей ДІ, ліг в основу розробки програмного-інформаційного забезпечення систем реального часу. Кореляційні, ентропійні та статистичні моделі використано для ідентифікації та аналізу аварійних станів об'єктів електроенергетики.

В додатках приведені приклади побудови розроблених моделей, в тому числі і для реальних об'єктів, способи відображення моделей, текст програми обчислення моделей та акти впровадження результатів роботи.

## ВИСНОВКИ ПО РОБОТІ

1. Здійснена формалізація з єдиних методологічних позицій комплексу інформаційних моделей, який включає сигнальні, статистичні, кореляційні, ентропійні, логіко-статистичні, кластерні, квазістаціонарні моделі, моделі руху даних та розпізнавання образів. Використання поданого узагальненого комплексу моделей джерел інформації можливе для широкого класу ОК, дозволяє отримати різносторонні характеристики об'єкту та ідентифікувати відхилення його станів від норми за багатьма ознаками.

2. Проаналізовано існуючі та розроблені нові типи кореляційних моделей, які характеризуються різною швидкодією, точністю та складністю реалізацією алгоритмів, що дозволило створити рекомендації щодо їх ефективнішого використання. На основі досліджених кореляційних оцінках розроблено ентропійні моделі, які враховують статистичну залежність між станами ДІ, але, на відміну від кореляційних, відображають усереднене значення цієї залежності в логарифмічному просторі, що дозволяє значно зменшити об'єм даних. Досліджено відомі та запропоновано нові типи кластерних моделей, які забезпечують ефективніше сприйняття інформації оператором за рахунок відображення найбільш важливих станів ОК за кількісними та семантичними показниками.

3. Вперше запропонована модель глобальної дисперсії та п'ятої ЛСІМ, побудованих на основі матриці коефіцієнтів кореляції, які забезпечують інтегроване відображення станів складного джерела інформації і мінімальний об'єм даних серед досліджених моделей. Дані моделі здатні здійснити стиснення даних в тисячі або десятки тисяч разів.

4. Запропоновано методи представлення знакозмінних сигналів в інтегральній технології з використанням кодів поля Галуа, які дозволяють змен-

шити об’єм і підвищити завадозахищеність даних, які є основою для побудови моделей ДІ. Відповідно, достовірність вихідних даних забезпечує більш адекватне відображення моделями станів ОК.

5. Розроблено структури алгоритмів розрахунку кореляційних моделей з глибоким розпаралеленням операцій, що дозволяє підвищити швидкодію їх виконання.

6. За результатами досліджень проведених в дисертації розроблено алгоритми та програмне забезпечення, яке практично використано для аналізу об’єктів електроенергетики в промислових умовах. Створене програмне забезпечення може бути використані при незначній адаптації для дослідження об’єктів інших галузей народного господарства.

## ПУБЛІКАЦІЇ ЗА ТЕМОЮ ДИСЕРТАЦІЙНОЇ РОБОТИ

1. Сегін А. І. Подання інформаційних моделей та методика їх побудови // Розвідка і розробка нафтових і газових родовищ. Івано-Франківськ: ІФДТУНГ, Серія: технічна кібернетика та електрифікація об’єктів паливно-енергетичного комплексу. – 1997. – Т. 6, № 34. – С. 23-34.

2. Николайчук Я. М., Сегін А. І. Моделі джерел інформації та методи їх представлення // Методи та прилади контролю якості. ІФДТУНГ, 1998, № 2. – С. 80–84.

3. Сегін А. І. Побудова алгоритмів формування моделей джерел інформації для автоматизованих систем реального часу // Розвідка і розробка нафтових і газових родовищ. Івано-Франківськ: ІФДТУНГ. Серія: методи і засоби технічної діагностики. –1999. – Т. 8, № 36. – С. 84–93.

4. Сегін А. І. Оцінка впливу старіння інформації на кореляційні моделі багатоканальних об’єктів управління // Розвідка і розробка нафтових і газових родовищ. Івано-Франківськ: ІФДТУНГ Серія: технічна кібернетика та електрифікація об’єктів паливно-енергетичного комплексу. – Т. 6, № 36, – 1999. – С. 80–88.

5. Сегін А. І. Теоретичні основи ентропійних моделей об’єктів управління // Матеріали науково-технічної конференції професорсько-викладацького складу Івано-Франківського державного технічного університету нафти і газу. – Івано-Франківськ, 1999. – С. 89–90.

6. Николайчук Я. М., Сегін А. І. Інформаційні моделі об’єктів управління в багаторівневих мережах автоматизованих систем // Матеріали VI міжнародної науково-практичної конференції “Проблеми створення, інтеграції і використання науково-технічної інформації на сучасному етапі” –

7. Николайчук Я., Сегін А., Сабадаш І. Метод розпізнавання графічних зображень на основі власних функцій Карунена-Лоєва. // Комп’ютерні технології друкарства. Збірник наукових праць. – Львів, – № 5, – 2000. – С. 344–347.

8. Сегін А. Ідентифікація двовимірних зображень на основі інформаційної моделі хемінгового простору // Комп’ютерні технології друкарства. Збірник наукових праць. – Львів, – № 4, – 2000. – С. 344–347.

## АНОТАЦІЯ

Сегін А.І. Моделі джерел інформації та методи їх формування в системах реального часу.

Дисертація у вигляді рукопису на здобуття наукового ступеню кандидата технічних наук зі спеціальності 01.05.02 – математичне моделювання та обчислювальні методи, Національний університет “Львівська політехніка”, Львів. 2001.

В дисертаційній роботі розроблені та досліджені моделі джерел інформації, які належать до класів статистичних, кореляційних, ентропійних, спектральних, логіко-статистичних, кластерних, розпізнавання образів, квазістаціонарних моделей та моделей руху даних. Створено інформаційні моделі та методи кодування даних знакозмінних сигналів в інтегральній технології.

Здійснено аналіз впливу завад з нормальним законом розподілу ймовірностей та запізнення аналого-цифрового перетворення на точність побудови кореляційних моделей. Досліджено залежність коефіцієнтів стиснення даних представлених моделей від розрядності відліків та кількості каналів джерела інформації.

Ключові слова: математична модель, кореляційна функція, ентропія, модель джерела інформації, інформаційна технологія.

## ABSTRACT

Segin A. I. Models of information sources and methods of them forming at real-time systems.

Thesis is represented as manuscript to acquire for degree of candidate of technical sciences in specialty 01.05.02 – Mathematical modeling and computation methods. National University “Lvivska Politechnica”, Lviv, 2001.

Information sources models belonging to classes of statistical, correlation, entropy, logic-statistical, clasteral, pattern recognition and quasi-stationary models were worked out and researched in thesis. Information models and methods of data coding of bipolar signals in integral technologies are created.

Influences analyze of interference with normal law of distributive probability and analog-digital transformation delay on formation precision of correlation function was implemented. Dependence of data compression coefficient of represented models from data rank and channels number of information source was investigated.

Key word: mathematical model, correlation function, entropy, information source model, information technology.

# АННОТАЦИЯ

Сегин А.И. Модели источников информации и методы их формирования в системах реального времени.

Диссертация в виде рукописи на соискание учёной степени кандидата технических наук по специальности 01.05.02 – математическое моделирование и численные методы. Национальный университет “Львивська политэхника” г. Львов –2001 г.

В диссертационной работе разработаны и исследованы модели источников информации, которые принадлежат к классам статистических, корреляционных, энтропийных, спектральных, логико-статистических, кластерних, распознавания образов, квазистационарных моделей и моделей движения данных. Созданы информационные модели и методы кодирования данных сигналов с изменяющимся знаком в интегральной технологии.

Осуществлен анализ влияния помех с нормальным законом распределения и опоздания аналого-цифрового преобразования на точность построения корреляционных моделей. Исследована зависимость коэффициентов сжатия представленных моделей от разрядности представления данных и количества каналов источника информации.

Построены графы алгоритмов вычисления корреляционных моделей различных типов с глубоким распараллеливанием выполнения операций, на основе которых увеличено быстродействие их расчёта, а также построен обобщённый граф вычисления энтропийной модели и глобальной дисперсии, базирующихся на нормированной корреляционной модели.

Разработаны структурные схемы формировщиков ЛСИМ 1 – 4, которые позволят реализовать модели аппаратными средствами на низовых уровнях вычислительных систем, что обеспечит значительное уменьшение объёма данных, передаваемых на высшие уровни системы для последующей обработки и анализа.

Предложены методы кодирования биполярных данных в интегральной технологии в базисе Галуа, а также методы приведения сигналов с интервалами уменьшения амплитуды к сигналам не имеющих отрицательных приращений амплитуды, которые наиболее удобны при использовании интегральной информационной технологии.

Разработаны программные средства построения и отображения статистических, корреляционных, спектральных, логико-статистических информационных моделей.

Ключевые слова: математическая модель, корреляционная функция, энтропия, модель источника информации, информационная технология.