

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 681.3

О. П. ВОЙТОВИЧ, В. О. ВІТЮК, В. А. КАПЛУН

Вінницький національний технічний університет, м. Вінниця

ОСОБЛИВОСТІ ДОСЛІДЖЕННЯ ОЗНАК ШКІДЛИВОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ БЕЗ НАЯВНОСТІ ВИХІДНИХ КОДІВ

Анотація. В статті розглянуті особливості розслідування комп'ютерних інцидентів при знаходженні в атакованій системі шкідливого програмного забезпечення. Розглянуті відомі технології виявлення ознак шкідливого програмного забезпечення. Запропонована математична модель, що описує ознаки шкідливого програмного забезпечення. Удосконалено підхід для моніторингу шкідливих функцій, який реалізовано у вигляді програмного засобу.

Ключові слова: дослідження комп'ютерних інцидентів, шкідливе програмне забезпечення, HIPS, VIPS, пісочниця, узагальнена математична модель виявлення ознак шкідливого програмного забезпечення

Аннотация. В статье рассмотрены особенности расследования компьютерных инцидентов при нахождении в атакованной системе вредоносных программ. Рассмотрены известные технологии обнаружения признаков вредоносного программного обеспечения. Предложена математическая модель, описывающая признаки вредоносного программного обеспечения. Усовершенствован подход для мониторинга вредоносных функций, который реализован в виде программного средства.

Ключевые слова: исследования компьютерных инцидентов, вредоносное программное обеспечение, HIPS, VIPS, песочница, обобщенная математическая модель выявления признаков вредоносного программного обеспечения

Abstract. The peculiarities of computer incidents investigating in the system attacked by malware were considered. Known technologies of detecting malicious software are investigated. A mathematical model that describes the malware characteristics is described. Improved approach for detecting malware functions is implemented as software.

Key words: researching of computer incidents, malware, HIPS, VIPS, sandbox, mathematical model of malware detecting

Вступ

З перенесенням основних бізнес-процесів підприємств в інформаційно-комунікаційні системи все частіше трапляються випадки компрометації окремих систем. Причому, методи та шляхи атак стають все більш різноманітними, часто зловмисники впроваджують своє програмне забезпечення (ПЗ) для використання систем як «зомбі» для атак на інші системи та персональні комп'ютери [1].

Актуальність

Шкідливе ПЗ може проникнути до комп'ютерної системи через недостатнє дотримання політики безпеки з боку користувачів, наприклад через встановлення недовіреного ПЗ, або в результаті реалізації атаки зловмисником. При дослідженні комп'ютерних інцидентів недостатньо лише виявити, що певне ПЗ містить шкідливі (недокументовані) функції, необхідно знати, які саме процеси виконувались в операційній системі (ОС). Наприклад, яка інформація перехоплювалась, до яких файлів чи програм здобувався доступ, який мережевий порт відкривався для доступу тощо. Стандартні антивірусні системи зазвичай дають інформацію лише про назву та основні характеристики шкідливого ПЗ, не вказуючи необхідних для дослідника відомостей [2]. При цьому стоїть задача дослідження ПЗ без наявності вихідних кодів, що значно ускладнює роботу дослідника [1]. Основні ознаки наявності недокументованих функцій такі: запис у реєстр (особливо зміна гілок реєстру, що стосуються системних або інших програм), доступ на читання/запис до файлів та оперативної пам'яті, доступ до периферійних пристроїв, мережева активність, звертання до поштових систем, відправлення повідомлень, зміна налаштувань програм, зокрема інтернет-браузерів, звертання до тимчасових файлів інших програм, пошук за сигнатурами, перехоплення переривань та системних функцій тощо [3,4].

Мета

При дослідженні комп'ютерних інцидентів часто постає питання виявлення не тільки факту наявності шкідливого ПЗ, але й визначення, які саме функції виконувались. Метою даної статті є дослідження та розробка технологій, які дозволяють покращити виявлення ознак шкідливого ПЗ.

Аналіз проявів та ознак шкідливого ПЗ

Під терміном шкідливе ПЗ розуміють програмні засоби, що несанкціоновано впроваджуються у комп'ютерну систему, і, які здатні викликати порушення політики безпеки, завдавати шкоди інформаційним ресурсам, а в окремих випадках, й апаратним ресурсам комп'ютерної системи [3].

Найпопулярнішими видами шкідливого ПЗ є такі [5,6]: комп'ютерний вірус – різновид шкідливих програм, метою яких є проведення дій, що завдають шкоди власникові комп'ютерній системі. Відмінною особливістю вірусів є здатність до розмноження та впроваджуватись в тіло програм, завантажувального сектору, документа тощо; троянська програма – шкідлива програма, яка проникає на комп'ютер жертви під виглядом нешкідливої (наприклад, кодек, системне оновлення, заставки, драйвера тощо); мережевий черв'як – самостійна шкідлива програма, яка проникає на комп'ютер жертви, використовуючи

уразливості в ПЗ операційних систем; руткіт – програма, призначена для приховування слідів шкідливих дій зловмисника в системі.

Різні типи шкідливого ПЗ у комп'ютерній системі поводять себе по-різному, але можна виділити наступні ознаки, що найчастіше помічають користувачі комп'ютерних систем: виведення на екран неочікуваних повідомлень або зображень; подача непередбачених звукових сигналів; довільний запуск на комп'ютерних програм без виклику користувача; поява попереджень від міжмережевого екрану про спробу програми вийти в Інтернет; відправлення листів через електронну пошту на адреси, які збережені в контактах; у поштової скриньці знаходиться велика кількість повідомлень без зворотної адреси і заголовка.

Якщо користувач помічає, що з комп'ютером відбувається подібне, то з великою мірою вірогідності можна передбачити, що комп'ютер заражений шкідливим ПЗ.

Крім даних ознак існують непрямі ознаки зараження комп'ютера: часті зависання і збої в роботі комп'ютера; повільна робота комп'ютера при запуску програм; неможливість завантаження операційної системи; зникнення файлів і каталогів або спотворення їх вмісту; часте звернення до жорсткого диска (часто блимає лампочка на системному блоці); інтернет-браузер «зависає» або поводить несподіваним чином (наприклад, вікно програми неможливо закрити).

За цими ознаками, які може помітити користувач відбуваються операції, дані про які можна отримати лише спеціалізованими засобами. А саме дані про:

- читання/запис даних у файлової системі – створення, видалення, редагування файлів, каталогів, дописування інформації в файл;
- модифікації пам'яті – створення чи завершення процесів, створення прихованих процесів;
- зміни реєстру – створення нових записів в реєстрі, редагування або видалення існуючих;
- зовнішня мережева активність – отримання чи відсилання інформації через мережу;
- внутрішня мережева активність - отримання чи відсилання інформації через localhost;
- перехоплення хуків клавіатури;
- відкриття портів;
- запуск файлів в операційній системі;
- встановлення чи заміна драйверів.

Отже, для виявлення ознак шкідливого ПЗ, його потрібно досліджувати за наведеними вище параметрами.

Аналіз існуючих технологій виявлення ознак шкідливого ПЗ

На сучасному етапі розвитку систем захисту ПЗ серед програм, які включають компоненти виявлення ознак шкідливого ПЗ, найпопулярнішими є такі: ThreatExpert [7], Process Monitor[8], Defense Wall HIPS[9], SafenSoft SysWatch Deluxe [10] та деякі інші. Вони використовують один з трьох основних методів контролю активності ПЗ: HIPS, VIPS та Пісочниця (sandbox).

Технологія HIPS – це технологія контролю активності, заснована на перехопленні звернень до ядра ОС і блокуванні виконання потенційно небезпечних дій ПЗ, яке працює в режимі користувача, виконуваних без відома користувача [5]. За допомогою власного драйвера перехоплює всі звернення ПЗ до ядра ОС. У разі спроби виконання потенційно небезпечної дії з боку ПЗ, HIPS-система блокує виконання даної дії і видає запит користувачеві, який вирішує дозволити або заборонити виконання даної дії. Схема взаємодії процесів і ядра ОС в методі HIPS показано на рис. 1.

Переваги систем, побудованих на методі HIPS:

- низьке споживання системних ресурсів;
- не вимогливі до апаратного забезпечення ПК (можуть працювати на різних платформах);
- можливість визначення загроз нульового дня;
- можливість визначення руткітів, які працюють в режимі користувача.

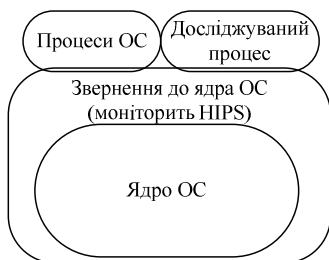


Рисунок 1 - Схема взаємодії процесів і ядра ОС в технології HIPS

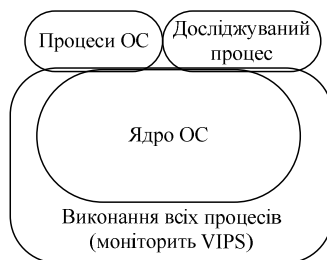


Рисунок 2 - Схема взаємодії процесів і ядра ОС в технології VIPS

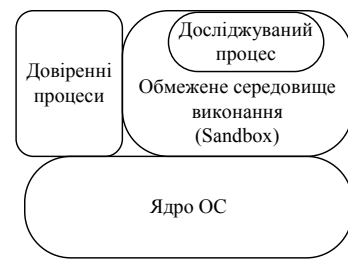


Рисунок 3 - Схема взаємодії процесів і ядра ОС в технології пісочниця

УДК 004.415.532

О. А. РЕМІННИЙ

Вінницький національний технічний університет, м. Вінниця

ПАТЕРНИ АВТОМАТИЗОВАНОГО ФУНКЦІОНАЛЬНОГО ТЕСТУВАННЯ КОРИСТУВАЦЬКИХ ІНТЕРФЕЙСІВ

Анотація. Проведено класифікацію дизайн патернів функціонального автоматизованого тестування. Вперше виділено групи патернів функціонального автоматизованого тестування, до яких приводить багатoshарова архітектура системи тестування. Зроблена класифікація патернів за їх призначенням. Виділено архітектурний архетип системи автоматизованого тестування у вигляді багатoshарового композитного рішення з високою ізоляваністю шарів та мета фреймворком.

Ключові слова: класифікація патернів, користувацький інтерфейс, тестування.

Аннотация. Проведена классификация дизайн паттернов функционального автоматизированного тестирования. Впервые выделены группы паттернов функционального автоматизированного тестирования, к которым приводит многослойная архитектура системы тестирования. Произведена классификация паттернов по их назначению. Выделено архитектурный архетип системы автоматизированного тестирования в виде многослойного композитного решения с высокой изолированностью слоев и мета фреймворком.

Ключевые слова: классификация паттернов, пользовательский интерфейс, тестирование.

Abstract. Made classification of functional automated test design patterns. Patterns are grouped into categories caused by a multilayer architecture of the test system. Made classification of patterns according to their purpose. Allocated architectural archetype of automated testing as a multilayered composite solution with high layers isolation and reusable meta framework block.

Keyword: classification of patterns, user interface, testing.

Вступ

Зараз існує досить широка кількість підходів до розробки програмного забезпечення: Об'єктно орієнтоване програмування, функціональне програмування, багато їх піднапрямків (Domain Driven Design, Test Driven Design, Behavior Driven Design та ін.) Дані підходи мають в основі певні декларативні концепції та теореми, які спрощують процеси визначення початкової архітектури системи, розуміння системи, передачу знань між розробниками і т.п.

Процес розробки рішення комплексу автоматизованих функціональних тестів певної програми не сильно відрізняється від процесу створення самої програми. Автоматизоване тестування є досить молодим напрямком і на даний момент швидко розвивається, стандартизується та удосконалюється. Створюються нові утиліти для взаємодії з тестованою системою (SUT – System Under Test).

На даний момент досить критичною частиною розробки автоматизованих тестів став етап підготовки та визначення базових практик та підходів, за допомогою яких буде розроблятися система автоматизованого тестування. На рис 1 можна побачити як взаємодіють між собою система автоматизованого тестування та тестер:

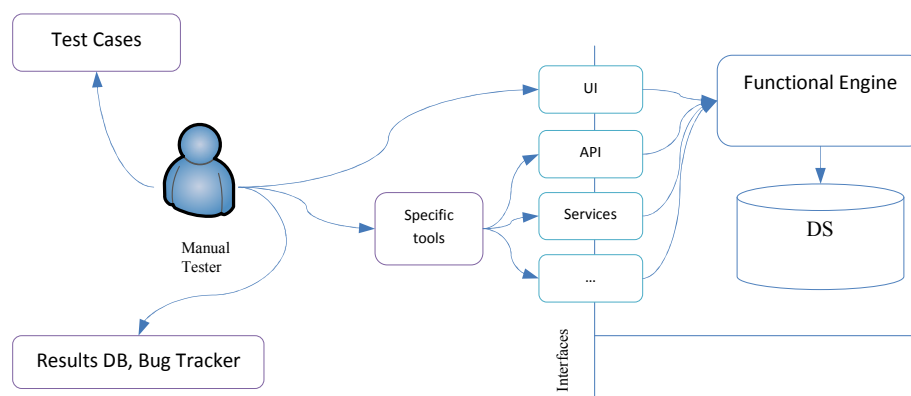


Рисунок 1 - Взаємодія тестера з тестованою системою

Центральною частиною такої системи є особа тестер (Manual Tester). Використовуючи ручну взаємодію та візуальний аналіз аплікації, а також спеціальні утиліти доступу (Specific tools) до невізуальних інтерфейсів тестованої системи, тестер відтворює викладені в тест кейсах (Test Cases) сценарії. У випадку збою чи непрогнозованої поведінки системи, тестер додає інформацію про некоректну поведінку в систему відстеження помилок системи (Bug tracker).

Основною ціллю автоматизованого тестування є видалення людської взаємодії з тестованою системою. З огляду літературних джерел видно, що на ринку присутня досить велика кількість систем автоматизованого тестування. Як виявив літературний пошук [1], на даний момент існує більше 20 комерційних та безкоштовних продуктів які дозволяють доступатись до інтерфейсів системи (interfaces), і які декларують певні конкретні вимоги та рекомендації по роботі конкретно з їхнім продуктом. Однак немає чітко відокремлених практик, які були б незалежними від постачальника продукту автоматизованого тестування.

Також постачальники програмних продуктів досить часто вдаються до маркетингових хитрощів і описують переваги їхніх систем при малій кількості функціональних тестів, а як показує практика найдорожчою в процесі роботи з системою виявляється підтримка існуючих тестів. А для того щоб правильно розробити фреймворк автоматизованого тестування, потрібно керуватись ні від чого незалежним набором правил та практик.

Мета

Основною задачею даної статті є виділення набору базових патернів автоматизованого функціонального тестування графічних інтерфейсів програмних додатків та їх групування за призначенням.

Постановка задачі

Перед тим як перейти до визначення набору патернів, звернемось до самого поняття патернів. Термін патерн дизайну вперше був визначений архітектором Крістофером Олександром.

"Кожен патерн описує проблему, яка виникає знову і знову в нашому середовищі, а потім описує основні рішення цієї проблеми, таким чином, що ви можете використовувати це рішення мільйон разів, ніколи не вирішуючи ті ж самі задачі".

Одним з основоположників патернів дизайну в програмуванні можна вважати групу співавторів GOF [2]. Відповідно їх визначення:

"Кожен патерн дизайну систематично окреслює, пояснює і оцінює важливі повторювані дизайни в об'єктно-орієнтованій системі".

Далі перейдемо до визначення самих патернів по категоріях. Як приклад для застосування розглянемо наступну задачу автоматизації веб сайту (рис 2). Додаток має в собі набір сторінок, серед яких є логін сторінка. Кожен тест має пройти через логін сторінку для виконання подальшого функціоналу.

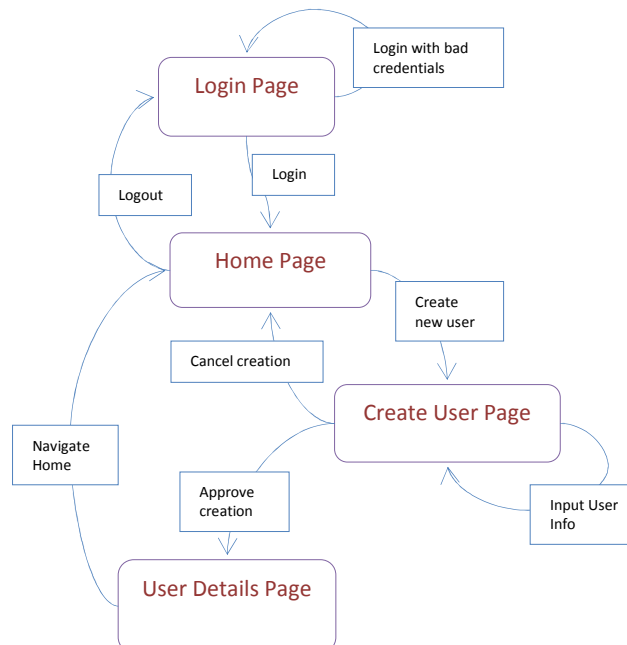


Рисунок 2 – Приклад простої веб-аплікації з мінімальним набором сторінок та функціоналу

Патерни автоматизованого функціонального тестування

На рисунку 3 наведена загальна класифікаційна схема патернів автоматизованого тестування. Вони розбиті на групи відповідно до тих цільових задач. Далі наведено більш детальний опис кожної категорії та самих шаблонів.

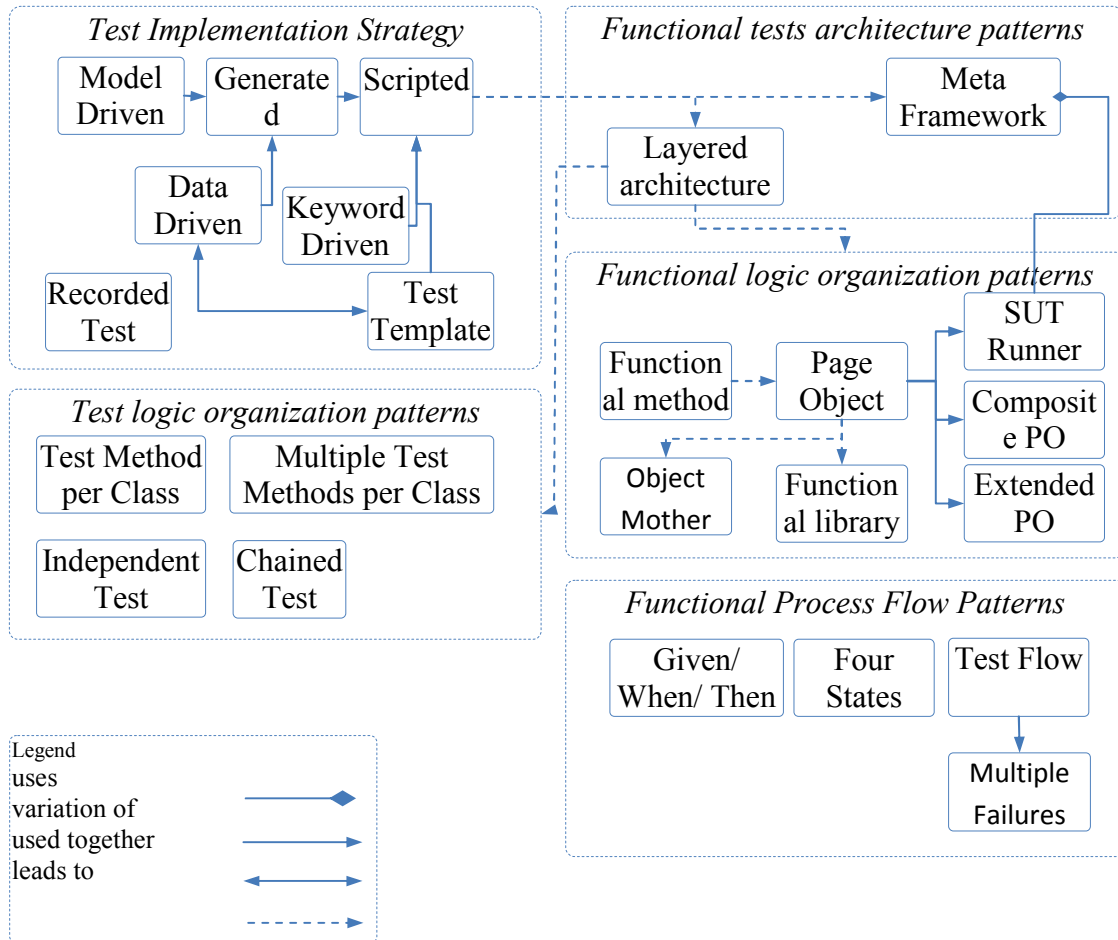


Рисунок 3 - Класифікаційна схема шаблонів автоматизованого функціонального тестування

Патерни імплементації тестів

Дана категорія об'єднує набір патернів які дозволяють обрати стратегію розробки рішення автоматизованого тестування.

Записані тести (Recorded). Імплементація тесту виконується за рахунок утиліти автоматизованого тестування, яка вміє виконувати запис та відтворення. В деякій мірі вважається поганою практикою, так як дуже дорогі в підтримці.

Запрограмовані (Scripted). Імплементація тесту виконується програмістом.

Імплементація базового шаблону (Test Template). Імплементація базового класу шаблону тесту. Варіації тесту – за рахунок наслідування та розширення функціоналу шаблону.

Імплементація за даними (Data Driven). Імплементація базового тесту визначається тест кейсом. Варіації тесту – за рахунок набору різних вхідних комбінацій даних.

Імплементація за Ключовими словами (Keyword Driven). Імплементація тесту за допомогою ключових слів (як Натиснути, Ввести і т.п.).

На даний момент існує декілька програмних утиліт, які дозволяють імплементувати тести за допомогою ключових слів. Кроки тесту мають вигляд набору з ключового слова, назви контролю на екрані та вхідних параметрів.

Імплементація з моделі (Model Driven). Будь яка аплікація в конкретний момент часу з конкретними вхідними даними може знаходитись тільки в одному конкретному коректному стані. Відповідно до такого визначення можна стверджувати що програма є кінцевою машиною станів (кінцевим автоматом). Враховуючи цей факт та наявність моделі станів та переходів (прикладом є рисунок X) можна визначити певні набори переходів між сторінками які б максимально покривали функціонал програми.

Архітектурні патерни

Багатошарова структура тестової системи (Multi Layered Test Solution). Розділяє логіку тестової системи на окремі логічні шари.

Це добре розповсюджена ідея - розподілену систему архітектурно розділяти на окремі шари. Перший рівень інкапсулює логіку представлення, другий - рівень бізнес-логіки, а третій шар відповідає

за зберігання даних. Використання цієї парадигми дозволяє знизити вартість обслуговування програмних додатків, так як компоненти всередині кожного рівня можуть бути змінені без впливу на інші рівні. Той же принцип може бути застосований до системи тестування.

Код тестів можна розділити на три шари: шар інструменту доступу до користувацьких інтерфейсів програмного додатку, шар функціональної логіки та шар тест кейсу. Кожен шар має певну відповідальність із загальною метою скорочення витрат на підтримку тестів і полегшення створення нових тестів.

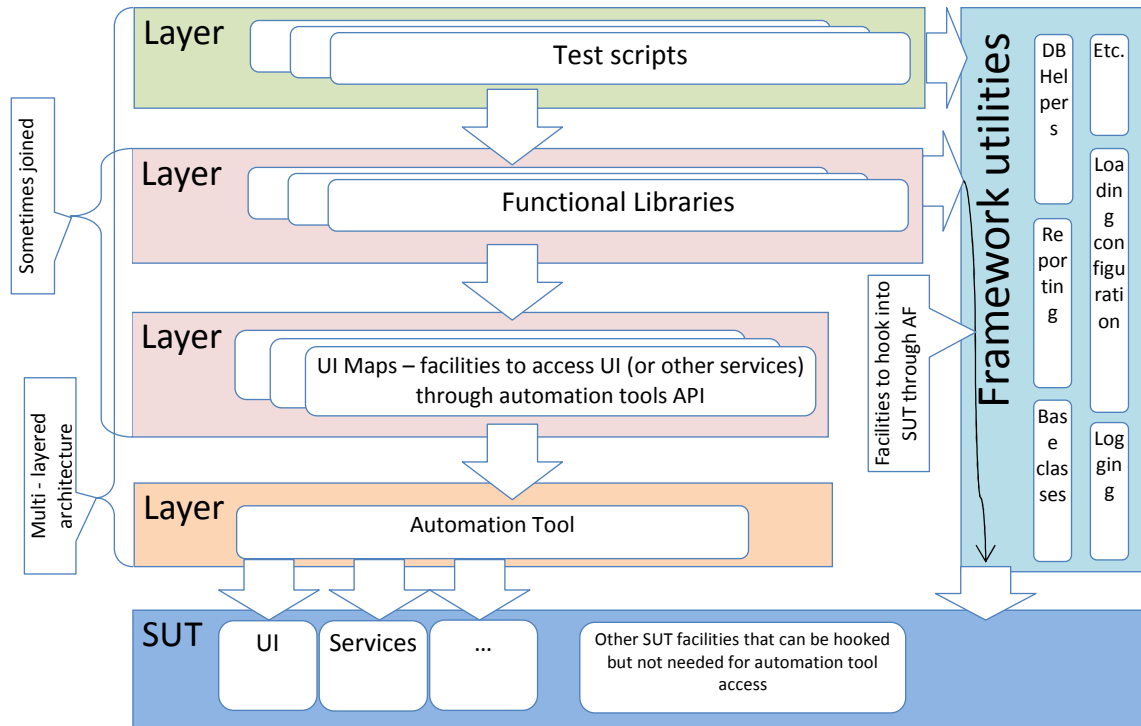


Рисунок 4 – Архітектурний архетип - багатошарова архітектура тестової системи

Мета Фреймворк (Meta Framework). Виділяє набір базових незалежних утилітарних класів, які можуть використовуватись незалежно від обраних інструментів тестування та перевикористовуватись між проектами.

Такі рішення зазвичай потрібні у випадку коли тестуються різні проекти всередині однієї корпорації і корпоративний стандарт вимагає уніфікованого інтерфейсу результатів. Також Мета фреймворк покращує показники перевикористання коду між проектами, оскільки кращі утилітні методи можуть включатись до його складу. Базові класи як для функціональних так і тестових об'єктів спрощують передачу знань між різними проектами. На рисунку X Мета фреймворк виділено в правій частині.

Функціональні композиційні патерни

Дана група шаблонів описує як можна виділяти частини коду в універсальні одиночні компоненти, призначені для багаторазового використання.

Функціональний Метод (Functional method). Абстрагує конкретну функцію аплікації від її реалізації на UI, API чи іншому рівні.

Багато утиліт для автоматизованого тестування дозволяють створювати так звані «записані сценарії» - коли розробник тесту робить певні дії з конкретною аплікацією і вони автоматично створюють скрипт тесту, який потім можна буде повторно програти і перевірити чи він виконався коректно після змін в програмі.

Зміна вигляду логін сторінки призведе до того, що потрібно буде провести зміни в усіх запропонованих методах. Якщо ж абстрагувати логін метод у вигляді Application.Login(username, password), і у всіх тестах використовувати саме цей метод, то у випадку зміни логін сторінки нам потрібно буде модифікувати лише один функціональний метод, і з зміни автоматично розповсюдяться на всі тести де цей метод використовується.

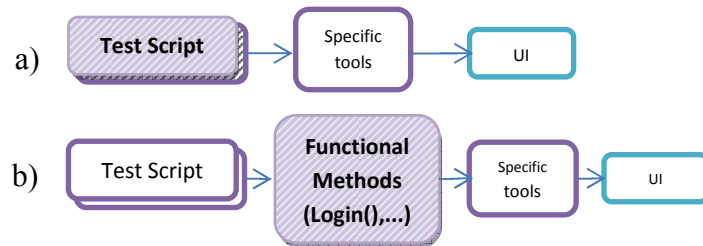


Рисунок 5 - Взаємодія тестового скрипту з користувацьким інтерфесом без проміжного шару функціональних методів (а) та з шаром функціональних методів (б)

Об'єкт сторінки (Page Object). Групує функціональні методи певної сторінки.

Функціональні методи для описаної на рисунку X аплікації можна винести в єдиний клас, так як їх є небагато. Однак для покращення підтримуваності коду буде краще згрупувати методи відповідно до сторінок, які представляють ці методи: сторінка PageLogin, методи: Login(); сторінка PageHome: методи: Logout(), CreateUser().

Функціональна бібліотека (Functional library) Групує функціональні об'єкти або (і) функціональні методи певної конкретної аплікації в один модуль, придатний для багаторазового використання.

Об'єкт запуску та закриття тестованої системи (SUT Runner). Дозволяє початковий запуск системи що підлягає тестуванню, її початкову ініціалізацію. Після закінчення тесту – звільнює ресурси пов'язані з системою.

Серед функціональних методів можна виділити набір таких, які не мають відношення до тестування функціоналу, як наприклад старт веб-браузера та навігація на логін сторінку SUT. Після закінчення тесту веб-браузер потрібно закрити. SUT Runner відповідає за такі загальні активності.

Джерело об'єктів (Object Mother). Створює об'єкти у вигляді, ініціалізованому і потрібному для виконання тесту.

Перевізник (Transporter). Централізує управління навігацією в тестованій системі в залежності від вимог тестів.

Даний об'єкт інкапсулює в собі всю логіку, пов'язану зі здійсненням навігації всередині тестованої системи. Таким чином задачі бізнес логіки не перетинається з навігацією всередині системи.

Для випадку наведеного на рисунку X, матимемо клас Transporter з методами NavigateToLogin(), NavigateToHomePage(), NavigateToCreateuser() і т.д. Як варіант, кожен окремий об'єкт сторінки може мати визначені власні методи транспортування, тоді - сам виконуватиме роль перевізника.

Композитний об'єкт сторінки (Composite Page Object). Агрегує менші об'єкти сторінки у одному зовнішньому об'єкті.

Даний патерн дозволяє структурувати об'єкти сторінки більш об'єктно орієнтовано, виділяючи під об'єкти, що можуть перевикористовуватись на різних сторінках, і включати їх в батьківський об'єкт.

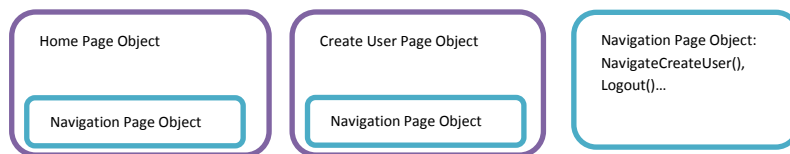


Рисунок 6 – Використання об'єкта сторінки Navigation через агрегацію в об'єктах сторінки Home та Create User

Розширений об'єкт сторінки (Extended Page Object). Розширяє базовий об'єкт сторінки через наслідування. Альтернатива композитному об'єкту сторінки.

Процесуальні патерни

Дана група патернів описує, за яким алгоритмом можна будувати функціональний тест, komponуючи його з послідовності викликів функціональних композиційних блоків.

Дано/Коли/Тоді (Given/When/Then). Ділить процес виконання тесту на три етапи:

- дано (визначення попередніх умов);
- коли (задає конкретні операції по роботі з контекстом);
- тоді (перевірка результатів).

Чотирьох етапний тест (4 stages test). Ділить процес виконання тесту на чотири етапи:

- встановлення початкових умов;
- виклик бізнес функцій;
- перевірка результатів;

- очищення системи.

Тестовий потік (Flow Test). Дозволяє на протязі одного тесту виконувати як бізнес операції так і перевірки. Вони можуть чередуватись для досягнення кінцевих цілей тесту.

Декілька помилок (Multiple Failures). Створює механізм який дозволяє продовжувати роботу тесту після некритичної помилки.

Патерни групування тестів

Дана група описує шаблони впорядкування користувацького коду в проекті автоматизованих тестів. Незалежний Тест (Standalone Test). Повертає тестовану систему до стану того ж в якому вона була до виконання тесту.

Пов'язаний Тест (Chained Test). Попередній тест виставляє стан тестованої системи, який потрібен буде наступним тестам.

Тестовий метод в класі (Test Method per Test Class). Окремий тестовий метод розташовується в окремому тестовому класі.

Груповані тестові методи в класі (Grouped Test Methods in Test Class). Окремий тестовий метод розташовується в окремому тестовому класі.

Висновки

В даній роботі проведено класифікацію дизайн патернів функціонального автоматизованого тестування. Вперше виділено такі патерни як Функціональний метод, Функціональна бібліотека, Об'єкт запуску та закриття тестованої системи, Композитний об'єкт сторінки, Розширений об'єкт сторінки. Визначено, що рушійною силою побудови тестувального рішення є вибір конкретного патерну імплементації тестів. Вперше виділено групи патернів функціонального автоматизованого тестування, до яких приводить багат шарова архітектура системи тестування. Зроблена класифікація патернів за їх призначенням. Вперше виділено архетип системи автоматизованого тестування у вигляді багат шарового композитного рішення з високою ізоляваністю шарів та перевикористовуванним блоком – мета фреймворком.

Серед подальших активностей можна виділити наступні – порівняння патернів які є альтернативами один одному.

Список літератури

- 1) List of GUI testing tools Testing [Електронний ресурс] // Режим доступу до файлу: http://en.wikipedia.org/wiki/List_of_GUI_testing_tools
- 2) Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software/ Addison-Wesley Professional – 1994. - 416p.
- 3) Gerard Meszaros, xUnit Test Patterns: Refactoring Test Code / Gerard Meszaros – 2007. – 833p.
- 4) Misha Rybalov, Design Patterns for Customer Testing [Електронний ресурс] // Режим доступу до файлу: <http://www.autotestguy.com/archives/Design%20Patterns%20for%20Customer%20Testing.pdf>
- 5) Ryan Gerard, Amit Mathur, Meta-Framework: A New Pattern for Test Automation [Електронний ресурс]/ Режим доступу до файлу: http://www.associationforsoftwaretesting.org/?dl_name=Meta-Framework.pdf

Відомості про авторів

Ремінний Олександр Андрійович – аспірант кафедри АІВТ. Вінницький національний технічний університет, м. Вінниця, вул. Хмельницьке шосе 95, (0432) 598243, Oleksandr.Reminnyi@gmail.com

УДК 638.322

В. П. ТАРАСЕНКО, О. К. ТЕСЛЕНКО, О. Ю. ЯНОВСЬКА

Національний технічний університет України «Київський політехнічний інститут», м. Київ

РЕАЛІЗАЦІЯ ОБЕРНЕНИХ ПІДСТАНОВОК НА ПРОСТОМУ ДВОМОДУЛЬНОМУ КАСКАДІ КОНСТРУКТИВНИХ МОДУЛІВ

Анотація. Визначено умови реалізації на простому двомодульному каскаді підстановок, обернених до підстановок, які також реалізуються на простому двомодульному каскаді. Визначена залежність структур (типів) конструктивних модулів простого двомодульного каскаду, який реалізує обернену підстановку, від структур (типів) конструктивних модулів простого двомодульного каскаду, який реалізує пряму підстановку. Теоретичні результати проілюстровані на прикладах. Важливість одержаних результатів полягає в забезпеченні теоретичної та практичної бази для реалізації прямих та обернених підстановок довільної розрядності на логічних структурах лінійної складності.

Ключові слова. функціональні перетворення на комбінаційних пристроях, двомодульний каскад, прямі та обернені підстановки, структури конструктивних модулів каскаду.

Аннотация: определены условия реализации на простом двомодульном каскаде подстановок, обратных к подстановкам, которые также реализуются на простом двомодульном каскаде. Определена зависимость структур (типов) конструктивных модулей простого двухмодульного каскада, который реализует обратную подстановку от структур (типов) конструктивных модулей, которые реализуют прямую подстановку. Теоретические результаты проиллюстрированы на конкретных примерах. Важность полученных результатов состоит в обеспечении теоретической и практической базы для реализации прямых и обратных подстановок произвольной разрядности на логических структурах линейной сложности.

Ключевые слова: функциональные преобразования на комбинационных устройствах, прямые и обратные подстановки, структуры конструктивных модулей каскада

Abstract. there are defined conditions of realization of permutation, which are inverse to ones, realized by two-module cascade. It is defined the dependence of the structure (type) of constructive modules of two-module cascade, realizing the inverse permutation on the structure (type) of constructive modules, realizing the direct one. Theoretical results were illustrated with examples. The importance of results obtained lies in provision of theoretical and practice basis for realization of direct and inverse permutations with arbitrary digit capacity using simple logical structures with linear complexity.

Key words: functional transformations using combinational devices, direct and inverse permutations, structure of constructive modules.

Вступ

Розвиток технології ПЛІС (Programmable Logic Devices – PLD) [1,2] призвів до появи можливості реалізації на мікросхемах доволі складних проектів протягом порівняно короткого часу. Це спонукає до пошуку нових методів реалізації засобів для спеціалізованих обчислень, інженерна розробка яких раніше була економічно недоцільною. Подібні обчислення добре відомі, наприклад, з практичного застосування результатів таких розділів математики, як теорія груп та підстановок [3]. Звичайно підстановки визначаються на множині $\mathbf{A} = \{0, 1, \dots, k-1\}$ із k елементів будь якої природи. Якщо $p(x)$ – підстановка на множині \mathbf{A} , то обернена підстановка $p^{-1}(x)$ визначається як $p^{-1}(p(x)) = p(p^{-1}(x)) = x$. За визначенням пряма і обернена підстановки пов'язані властивостями ізоморфізму. Однак ізоморфним підстановкам можуть відповідати їх програмні чи апаратні реалізації, що суттєво відрізняються за своїми основними характеристиками. Тому велике значення має систематизоване дослідження підстановок і, в першу чергу, їх апаратних реалізацій. Результати таких досліджень потенційно можуть мати і теоретичний інтерес, як операційна база відповідних перетворень.

Зауважимо, що ця робота в змістовному відношенні по суті є продовженням роботи [4] і значною мірою використовує науково-поняттєвий доробок, термінологію і позначення попередньої роботи.

Постановка задачі

Реалізація функціональних перетворень за допомогою комбінаційних пристроїв забезпечує максимальну їх продуктивність, але складність реалізації в загальному випадку зростає по експоненті із ростом кількості входів. Оскільки логічні функції часто мають досить велику кількість аргументів, то їх безпосередня реалізація на ПЛІС не завжди можлива. Тому актуальною є задача пошуку часткових рішень, коли складність реалізації підстановок комбінаційними пристроями буде мати поліноміальний характер. Одним із таких рішень є реалізація підстановок за допомогою простих одновимірних каскадів конструктивних модулів (ОККМ) [3]. Такі каскади є комбінаційними пристроями і мають структуру, показану на рис. 1, де X - вектор вхідних первинних змінних, $F(x)$ - вектор значень первинної вихідної функції, $r_1..r_n$, $h_1..h_n$ - змінні, що подаються відповідно на правий та лівий боковий вхід відповідного модуля, F_r , F_h - ліва та права бокові функції відповідно.

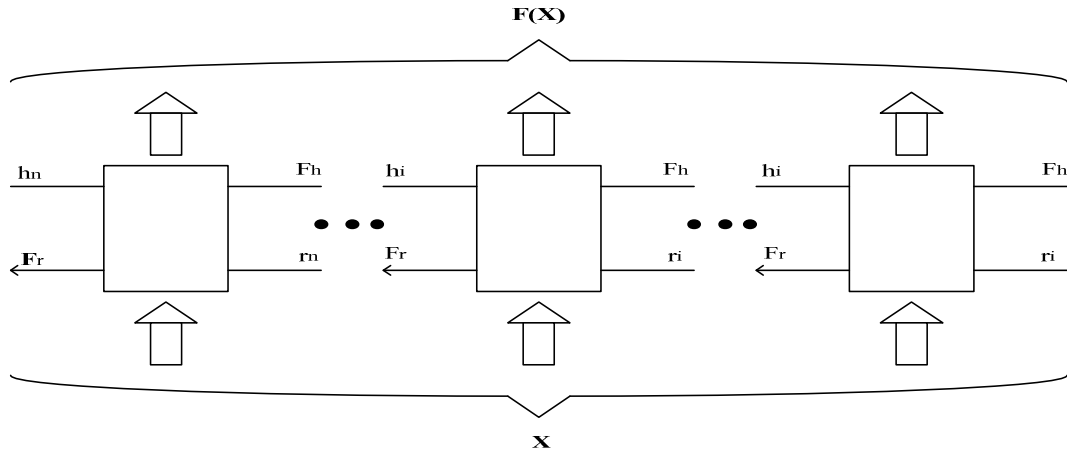


Рисунок 1 – Загальна структура ОККМ

Очевидно, що коли пряма та обернена підстановки реалізуються безпосередньо на ОККМ або, принаймні, на нескладній матричній структурі, то такі підстановки доцільно використовувати в ізоморфних перетвореннях у відповідних алгоритмах для досягнення оптимальних характеристик по певних критеріях. Якщо ж одна із таких підстановок має просту реалізацію, а інша – складну, то вони можуть знайти застосування в криптографічних перетвореннях (односторонні підстановки, односторонні функції). Таким чином, виникає задача визначення, по-перше, можливостей і складності реалізації підстановок, обернених до підстановок, які безпосередньо реалізуються на ОККМ, та, по-друге, структур конструктивних модулів (КМ) для реалізації обернених підстановок на основі відомих структур КМ, які реалізують пряму підстановку [4].

Двомодульний каскад

Структура двомодульного каскаду показана на рис.2. В роботі [4] розглядаються властивості двомодульного каскаду за умови, що каскад в цілому реалізує повну підстановку.

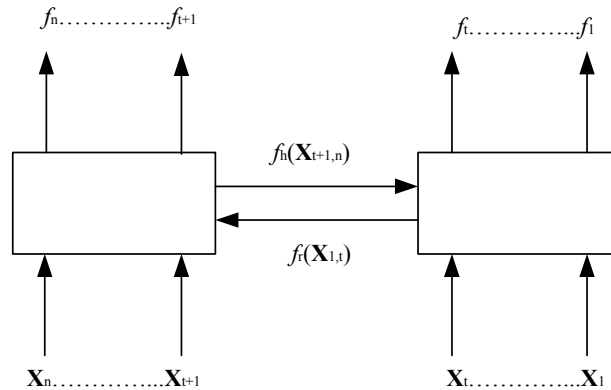


Рисунок 2 – Двомодульний каскад

Позначимо $\mathbf{X}_{1,t}$ кортеж (впорядковану послідовність) аргументів $\langle x_1, x_2, \dots, x_t \rangle$, що надходять на первинні входи правого модуля ОККМ (молодші розряди аргументу), аналогічно $\mathbf{X}_{t+1,n} \rightarrow \langle x_{t+1}, x_{t+2}, \dots, x_n \rangle$ - кортеж аргументів, що надходять на первинні входи лівого модуля (старші розряди аргументу), $\mathbf{F}_{1,t}$ - кортеж функцій $\langle f_1, f_2, \dots, f_t \rangle$, що реалізуються на первинних виходах правого модуля ОККМ, аналогічно $\mathbf{F}_{t+1,n} \rightarrow \langle f_{t+1}, f_{t+2}, \dots, f_n \rangle$ - кортеж функцій, що реалізуються на первинних виходах лівого модуля ОККМ, $\mathbf{S}_{1,t}$ - множина кодів значень аргументів із кортежа $\mathbf{X}_{1,t}$, $\mathbf{S}_{t+1,n}$ - множина кодів значень аргументів із $\mathbf{X}_{t+1,n}$.

Значення логічної функції $f(\mathbf{X}_{1,t})$ на боковому виході правого модуля поділяє множину $\mathbf{S}_{1,t}$ на дві підмножини \mathbf{S}_a і \mathbf{S}_b так, що для будь-яких значень аргументів $a_1, a_2 \in \mathbf{S}_a$ ($a_1 \neq a_2$) значення функції $f_r(a_1) = f_r(a_2)$, відповідно $\mathbf{F}_{t+1,n}(a_1, \mathbf{X}_{t+1,n}) = \mathbf{F}_{t+1,n}(a_2, \mathbf{X}_{t+1,n})$, для будь-яких $b_1, b_2 \in \mathbf{S}_b$ ($b_1 \neq b_2$) $f_r(b_1) = f_r(b_2)$, відповідно $\mathbf{F}_{t+1,n}(b_1, \mathbf{X}_{t+1,n}) = \mathbf{F}_{t+1,n}(b_2, \mathbf{X}_{t+1,n})$, а $\mathbf{F}_{t+1,n}(a_1, \mathbf{X}_{t+1,n}) \neq \mathbf{F}_{t+1,n}(b_1, \mathbf{X}_{t+1,n})$ ($f_r(a_1) \neq f_r(b_1)$).

Побудуємо наступні множини кодів аргументів із $\mathbf{X} - \mathbf{S}_{ca} = \mathbf{S}_c \times \mathbf{S}_a$, $\mathbf{S}_{da} = \mathbf{S}_d \times \mathbf{S}_a$, $\mathbf{S}_{cb} = \mathbf{S}_c \times \mathbf{S}_b$ та $\mathbf{S}_{db} = \mathbf{S}_d \times \mathbf{S}_b$ як декартові добутки [5] відповідних підмножин. Утворені множини попарно не мають спільних елементів, а їхнє об'єднання дорівнює $\mathbf{S}_{1,n}$ - множині всіх кодів аргументів із \mathbf{X} [4].

Парою підстановки будемо називати впорядковану пару кодів – коду аргументу із $\mathbf{S}_{1,n}$ (або $\mathbf{S}_{t+1,n}$) та коду значень функцій (далі – коду функцій) із кортежу $\mathbf{F}_{1,n}$ (або $\mathbf{F}_{t+1,n}$). Нехай \mathbf{P}_{ca}^\wedge - множина пар підстановки з кодами функцій із $\mathbf{F}_{1,t}(\mathbf{X}_{1,t}, c)$, коли аргументи із $\mathbf{X}_{1,t}$ приймають всі можливі коди із множини \mathbf{S}_a , а \mathbf{T}_{ca}^\wedge - множина пар підстановки з кодами функцій із $\mathbf{F}_{t+1,n}(a, \mathbf{X}_{t+1,n})$, коли аргументи із $\mathbf{X}_{t+1,n}$ приймають всі можливі коди із множини \mathbf{S}_c . Аналогічно визначимо множини пар підстановки \mathbf{P}_{da}^\wedge , \mathbf{T}_{ad}^\wedge , \mathbf{P}_{cb}^\wedge , \mathbf{T}_{bc}^\wedge та \mathbf{P}_{db}^\wedge , \mathbf{T}_{bd}^\wedge . Указані множини пар підстановки однозначно визначають повну підстановку та кортеж функцій $\mathbf{F}(\mathbf{X})$, які реалізує двомодульний каскад. Крім того, множини пар підстановок із точністю до інверсії функцій $f_r(\mathbf{X}_{1,t})$ та $f_h(\mathbf{X}_{t+1,n})$ визначають логічні функції, які реалізуються модулями каскаду.

Відома [4] класифікація модулів двомодульного каскаду за умови, що каскад в цілому реалізує повну підстановку. Модулі можна розділити на 6 типів, досліджуючи властивості множин \mathbf{P}_{da}^\wedge , \mathbf{T}_{ad}^\wedge , \mathbf{P}_{cb}^\wedge , \mathbf{T}_{bc}^\wedge та \mathbf{P}_{db}^\wedge , \mathbf{T}_{bd}^\wedge . Множини \mathbf{P}^\wedge та \mathbf{T}^\wedge з однаковим комплектом буквених індексів будемо називати спорідненими. Можливі три варіанти розбиття чотирьох множин кодів функцій на дві пари для забезпечення балансності (врівноваженості) [6]. Перший варіант (а) – $(\mathbf{P}_{ca}, \mathbf{P}_{cb})$, $(\mathbf{P}_{da}, \mathbf{P}_{db})$. Другий варіант (б) – $(\mathbf{P}_{ca}, \mathbf{P}_{da})$, $(\mathbf{P}_{cb}, \mathbf{P}_{db})$. Третій варіант (в) – $(\mathbf{P}_{ca}, \mathbf{P}_{db})$, $(\mathbf{P}_{da}, \mathbf{P}_{cb})$. У кожному із цих варіантів виділимо два випадки наявності спільних елементів, що важливо для забезпечення ортогональності [4]. У першому з них серед чотирьох множин $\mathbf{P}_{ca}, \mathbf{P}_{cb}, \mathbf{P}_{da}, \mathbf{P}_{db}$ кодів функцій існують три множини, які попарно мають спільні елементи (тобто, обов'язково існує множина, яка має спільні елементи із двома іншими). Такі типи конструктивних модулів позначаються цифрою 3. В другому випадку серед множин $\mathbf{P}_{ca}, \mathbf{P}_{cb}, \mathbf{P}_{da}, \mathbf{P}_{db}$ попарно спільні елементи можуть мати тільки дві множини (тобто будь-яка із множин може мати спільні елементи лише з однією з інших). Такі типи конструктивних позначаються цифрою 2. В роботі [4] наведені характеристики конструктивних модулів типів 3а, 3б, 3в, 2а, 2б, 2в. також показано, що не кожна пара типів конструктивних модулів є сумісною (тобто реалізує повну підстановку), а тільки : (3а-2а), (3б-2в), (3в-2б), (2а-2а), (2а-2б), (2б-3в), (2б-2в), (2в-2б), (2в-2в). При цьому сумісність має симетричний характер, тобто неважливо, який із модулів буде зліва (старші розряди), а який справа (молодші розряди).

Реалізація обернених підстановок

Розглянемо тепер реалізацію обернених підстановок за допомогою двомодульного ОККМ. Кортеж $\mathbf{G}(\mathbf{X})$ реалізує обернену підстановку, якщо $\mathbf{G}(\mathbf{F}(\mathbf{X})) = \mathbf{X}$. Оскільки аргументи множини \mathbf{X} та значення кортежів \mathbf{G} та \mathbf{F} функцій приймають всі значення із множини $\mathbf{S}_{1,n}$, то формування кортежу \mathbf{G} можна однозначно виконати наступним чином. Нехай для будь якого $\mathbf{S} \in \mathbf{S}_{1,n}$ $r = \mathbf{F}(s)$, тоді $\mathbf{G}(r) = s$.

Згідно з [4], реалізація прямої підстановки на простому двомодульному каскаді характеризується наступними множинами пар підстановки: $\mathbf{M}_{ca}^{\wedge} = \mathbf{T}_{ac}^{\wedge} \times \mathbf{P}_{ca}^{\wedge}$, $\mathbf{M}_{da}^{\wedge} = \mathbf{T}_{ad}^{\wedge} \times \mathbf{P}_{da}^{\wedge}$, $\mathbf{M}_{cb}^{\wedge} = \mathbf{T}_{bc}^{\wedge} \times \mathbf{P}_{cb}^{\wedge}$, $\mathbf{M}_{db}^{\wedge} = \mathbf{T}_{bd}^{\wedge} \times \mathbf{P}_{db}^{\wedge}$. В свою чергу множини пар підстановки T^{\wedge} та P^{\wedge} можна подати наступним чином:

$$\begin{aligned} \mathbf{T}_{ac}^{\wedge} &= (\langle \mathbf{S}_c \rangle, \langle \mathbf{T}_{ac} \rangle), \mathbf{T}_{ad}^{\wedge} = (\langle \mathbf{S}_d \rangle, \langle \mathbf{T}_{ad} \rangle), \\ \mathbf{T}_{bc}^{\wedge} &= (\langle \mathbf{S}_c \rangle, \langle \mathbf{T}_{bc} \rangle), \mathbf{T}_{bd}^{\wedge} = (\langle \mathbf{S}_d \rangle, \langle \mathbf{T}_{bd} \rangle), \\ \mathbf{P}_{ca}^{\wedge} &= (\langle \mathbf{S}_a \rangle, \langle \mathbf{P}_{ca} \rangle), \mathbf{P}_{da}^{\wedge} = (\langle \mathbf{S}_a \rangle, \langle \mathbf{P}_{da} \rangle), \\ \mathbf{P}_{db}^{\wedge} &= (\langle \mathbf{S}_b \rangle, \langle \mathbf{P}_{db} \rangle), \mathbf{P}_{cb}^{\wedge} = (\langle \mathbf{S}_b \rangle, \langle \mathbf{P}_{cb} \rangle), \end{aligned} \quad (1)$$

де використання кутових дужок вказує на взаємну впорядкованість елементів множин \mathbf{T} і \mathbf{S} та \mathbf{P} і \mathbf{S} у відповідності до $\mathbf{F}(\mathbf{X})$.

Очевидно, що обернена підстановка буде характеризуватись наступними множинами пар підстановки

де коди аргументів та коди значень функцій міняються місцями:

$$\begin{aligned} \mathbf{M}_{ca}^0 &= \mathbf{T}_{ac}^0 \times \mathbf{P}_{ca}^0 = (\langle \mathbf{T}_{ac} \rangle, \langle \mathbf{S}_c \rangle) \times (\langle \mathbf{P}_{ca} \rangle, \langle \mathbf{S}_a \rangle), \\ \mathbf{M}_{da}^0 &= \mathbf{T}_{ad}^0 \times \mathbf{P}_{da}^0 = (\langle \mathbf{T}_{ad} \rangle, \langle \mathbf{S}_d \rangle) \times (\langle \mathbf{P}_{da} \rangle, \langle \mathbf{S}_a \rangle), \\ \mathbf{M}_{cb}^0 &= \mathbf{T}_{bc}^0 \times \mathbf{P}_{cb}^0 = (\langle \mathbf{T}_{bc} \rangle, \langle \mathbf{S}_c \rangle) \times (\langle \mathbf{P}_{cb} \rangle, \langle \mathbf{S}_b \rangle), \\ \mathbf{M}_{db}^0 &= \mathbf{T}_{bd}^0 \times \mathbf{P}_{db}^0 = (\langle \mathbf{T}_{bd} \rangle, \langle \mathbf{S}_d \rangle) \times (\langle \mathbf{P}_{db} \rangle, \langle \mathbf{S}_b \rangle). \end{aligned} \quad (2)$$

Вияснимо факт існування наступних спільних декомпозицій кортежів функцій $\mathbf{G}_{t+1,n}(\mathbf{X})$ та $\mathbf{G}_{1,t}(\mathbf{X})$:

$$\mathbf{G}_{t+1,n}(\mathbf{X}) = \mathbf{G}_{t+1,n}(\mathbf{X}_{t+1,n}, f_r'(\mathbf{X}_{1,t})) \quad (3)$$

$$\mathbf{G}_{1,t}(\mathbf{X}) = \mathbf{G}_{1,t}(f_h'(\mathbf{X}_{t+1,n}), \mathbf{X}_{1,t}) \quad (4)$$

Розглянемо спільну декомпозицію (3). Коди аргументів в (2) згідно з визначенням – це коди значень функцій в (1). В свою чергу, згідно з [4], множини кодів функцій \mathbf{P}_{ca} , \mathbf{P}_{da} , \mathbf{P}_{cb} та \mathbf{P}_{db} розбиваються на дві пари таким чином, щоб множини, які належать одній парі не мали спільних елементів і були доповненням одна одній до множини $\mathbf{S}_{1,t}$. Утворимо попарні перетини цих множин, тобто $\mathbf{A}_1 = \mathbf{P}_{ca} \cap \mathbf{P}_{da}$, $\mathbf{A}_2 = \mathbf{P}_{ca} \cap \mathbf{P}_{cb}$, $\mathbf{A}_3 = \mathbf{P}_{ca} \cap \mathbf{P}_{db}$, $\mathbf{A}_4 = \mathbf{P}_{da} \cap \mathbf{P}_{cb}$, $\mathbf{A}_5 = \mathbf{P}_{da} \cap \mathbf{P}_{db}$, $\mathbf{A}_6 = \mathbf{P}_{cb} \cap \mathbf{P}_{db}$. Очевидно, що $\mathbf{A}_i \cap \mathbf{A}_j = \emptyset$, $j = 1, 2, \dots, 6, i \neq j$, $\mathbf{A}_1 \cup \mathbf{A}_2 \cup \mathbf{A}_3 \cup \mathbf{A}_4 \cup \mathbf{A}_5 \cup \mathbf{A}_6 = \mathbf{S}_1$. Крім того, дві із шести множин \mathbf{A}_i обов’язково будуть пустими, тобто в загальному випадку непустих множин \mathbf{A}_i не більше 4. Враховуючи, що в ОККМ типу 2 існують пари співпадаючих множин, то в цьому випадку кількість непустих множин \mathbf{A}_i лише 2.

Твердження 1. Нехай $r_1, r_2 \in \mathbf{A}_i (r_1 \neq r_2, \mathbf{A}_i \neq \emptyset)$. Тоді

$$\mathbf{G}_{t+1,n}(\mathbf{X}_{t+1,n}, r_1) = \mathbf{G}_{t+1,n}(\mathbf{X}_{t+1,n}, r_2) \quad (5)$$

Згідно з визначенням множин пар підстановки рівність (5) справедлива для будь яких $r_1 \neq r_2$, які містяться в одній із множин \mathbf{P} , за умови, що змінні із $\mathbf{X}_{t+1,n}$ є значенням кодів функцій із спорідненої

множини \mathbf{T} . Якщо множина \mathbf{A}_i не пуста, то це означає, що відповідні множини \mathbf{P} мають спільні елементи, але тоді, згідно з [4], споріднені множини \mathbf{T} не мають спільних елементів і доповняють одна одну до множини $\mathbf{S}_{t+1,n}$. Це означає, що рівність (5) справедлива на всіх значеннях змінних із $\mathbf{X}_{t+1,n}$ для будь яких $r_1, r_2 \in \mathbf{A}_i (r_1 \neq r_2)$. Твердження доведено. Аналогічно для кортежу $\mathbf{G}_{1,t}(\mathbf{X})$ (4).

Із твердження 1 випливає, що коефіцієнти спільних декомпозицій кортежів $\mathbf{G}_{t+1,n}(\mathbf{X})$ та $\mathbf{G}_{1,t}(\mathbf{X})$ принаймні не перевищують кількості не пустих множин \mathbf{A}_i . В свою чергу, в ОККМ, які належать до типу 3 (3а, 3б, 3в), кількість не пустих множин \mathbf{A}_i дорівнює 3 або 4, а в ОККМ типу 2 (2а, 2б, 2в) – дорівнює 2.

Отже в загальному випадку на простому двомодульному каскаді можуть бути реалізовані підстановки, обернені до підстановок, які реалізуються лише на КМ типу 2.

Розглянемо визначення структур КМ для реалізації обернених підстановок на основі відомих структур КМ, які реалізують пряму підстановку.

Для визначеності будемо розглядати КМ, які використовуються в молодших розрядах двомодульного каскаду, де будь який із них подається наступними множинами пар підстановки $\mathbf{P}_{ca}^{\wedge}, \mathbf{P}_{da}^{\wedge}, \mathbf{P}_{db}^{\wedge}, \mathbf{P}_{cb}^{\wedge}$. Згідно з попереднім маємо:

$$\begin{aligned} \mathbf{P}_{ca}^{\wedge} &= (\langle P_{ca} \rangle, \langle S_a \rangle), \mathbf{P}_{da}^{\wedge} = (\langle P_{da} \rangle, \langle S_a \rangle), \\ \mathbf{P}_{cb}^{\wedge} &= (\langle P_{cb} \rangle, \langle S_b \rangle), \mathbf{P}_{db}^{\wedge} = (\langle P_b \rangle, \langle S_b \rangle). \end{aligned} \quad (6)$$

Враховуючи симетричність при під'єднанні модулів КМ різних типів в двомодульний каскад, одержані результати для КМ в молодших розрядах легко узагальнюються і для КМ в старших розрядах.

Розглянемо КМ типу 2а, який характеризується наступною структурою:

$$\begin{aligned} \mathbf{P}_{ca} \cap \mathbf{P}_{cb} &= \emptyset, \mathbf{P}_{ca} \cup \mathbf{P}_{cb} = \mathbf{S}_{1,t}; \\ \mathbf{P}_{da} \cap \mathbf{P}_{db} &= \emptyset, \mathbf{P}_{da} \cup \mathbf{P}_{db} = \mathbf{S}_{1,t}; \\ \mathbf{P}_{ca} &= \mathbf{P}_a \text{ і } \mathbf{P}_{cb} = \mathbf{P}_{db} \\ f_r(\mathbf{S}_a) &= a, f_r(\mathbf{S}_b) = b, a, b \in \{0, 1\}, a \neq b \end{aligned}$$

Згідно (6)

$$\mathbf{P}_{ca}^0 = \mathbf{P}_{da}^0 = \mathbf{S}_a, \mathbf{S}_a^0 = \mathbf{P}_{ca}^0 = \mathbf{P}_{da}^0, \mathbf{P}_{cb}^0 = \mathbf{P}_{db}^0 = \mathbf{S}_b, \mathbf{S}_b^0 = \mathbf{P}_{cb}^0 = \mathbf{P}_{db}^0.$$

Маємо:

$$\begin{aligned} \mathbf{P}_{ca}^0 \cap \mathbf{P}_{cb}^0 &= \emptyset, \mathbf{P}_{ca}^0 \cup \mathbf{P}_{cb}^0 = \mathbf{S}_{1,t}; \\ \mathbf{P}_{da}^0 \cap \mathbf{P}_{db}^0 &= \emptyset, \mathbf{P}_{da}^0 \cup \mathbf{P}_{db}^0 = \mathbf{S}_{1,t}; \\ \mathbf{P}_{ca}^0 &= \mathbf{P}_{da}^0 \text{ і } \mathbf{P}_{cb}^0 = \mathbf{P}_{db}^0; \\ f_r^0(\mathbf{S}_a^0) &= a, f_r^0(\mathbf{S}_b^0) = b, a, b \notin \{0, 1\}, a \neq b \end{aligned}$$

Отже для реалізації оберненої підстановки КМ типу 2а перетворюються в КМ також типу 2а. КМ для реалізації обернених підстановок будемо в подальшому називати оберненим КМ.

Розглянемо приклад. Нехай $n=5, t=3$. В табл. 1 та в табл. 2 задано відповідні КМ, а в табл. 3 – підстановка, яка реалізується двомодульним каскадом.

Таблиця 1 – КМ для молодших розрядів

f_r	1	1	1	1	1	0	0	1
$\mathbf{F}_{1,3}$ при $f_h=0$	4	7	5	6	3	1	0	2
$\mathbf{F}_{1,3}$ при $f_h=1$	2	4	3	5	7	0	1	6
\mathbf{X}	0	1	2	3	4	5	6	7

$$\begin{aligned}
 \mathbf{S}_{a(a=0)} &= \{5, 6\}, \quad \mathbf{S}_{b(b=0)} = \{0, 1, 3, 4, 7\}, \\
 \mathbf{P}_{ca(c=0, a=0)} &= \mathbf{P}_{da(d=1, a=0)} = \{1, 0\}, \\
 \mathbf{P}_{cb(c=0, b=1)} &= \mathbf{P}_{ab(d=1, b=1)} = \{2, 3, 4, 5, 6, 7\}, \\
 \mathbf{P}_{ca}^{\wedge} &= \langle \{5, 6\}, \{1, 0\} \rangle, \mathbf{P}_{da}^{\wedge} = \langle \{5, 6\}, \{0, 1\} \rangle, \mathbf{P}_{cb}^{\wedge} = \langle \{0, 1, 2, 3, 4, 7\}, \{4, 7, 5, 6, 3, 2\} \rangle, \\
 \mathbf{P}_{db}^{\wedge} &= \langle \{0, 1, 2, 3, 4, 7\}, \{2, 4, 3, 5, 7, 6\} \rangle
 \end{aligned}$$

Таблиця 2 – КМ для старших розрядів

f_h	1	0	1	1
$\mathbf{F}_{4,5}$ при $f_r=0$	3	0	1	2
$\mathbf{F}_{4,5}$ при $f_r=1$	2	0	3	1
\mathbf{X}	0	1	2	3

$$\begin{aligned}
 \mathbf{S}_{c(c=0)} &= \{1\}, \quad \mathbf{S}_{d(d=1)} = \{0, 2, 3\}, \\
 \mathbf{T}_{ac(c=0, a=0)} &= \mathbf{T}_{bc(d=1, a=0)} = \{0\}, \\
 \mathbf{T}_{ad(c=0, b=1)} &= \mathbf{T}_{db(d=1, b=1)} = \{3, 1, 2\}, \\
 \mathbf{T}_{ac}^{\wedge} &= \mathbf{T}_{bc}^{\wedge} \langle \{1\}, \{0\} \rangle, \mathbf{T}_{ad}^{\wedge} = \langle \{0, 2, 3\}, \{3, 1, 2\} \rangle, \mathbf{T}_{bd}^{\wedge} = \langle \{0, 2, 3\}, \{2, 3, 1\} \rangle,
 \end{aligned}$$

Таблиця 3 – Пряма підстановка

$\mathbf{F}(\mathbf{X})=$	22	24	23	25	27	30	31	26	04	07	05	06	03	01	00	02
$\mathbf{X}=$	00	01	02	03	04	05	06	07	10	11	12	13	14	15	16	17
$\mathbf{F}(\mathbf{X})=$	32	34	33	35	37	10	11	36	12	14	13	15	17	20	21	16
$\mathbf{X}=$	20	21	22	23	24	25	26	27	30	31	32	33	34	35	36	37

Розглянемо формування КМ для оберненої підстановки. Обернений КМ для молодших розрядів.

$$\begin{aligned}
 \mathbf{P}_{ca}^0 &= \mathbf{P}_{da}^0 = \mathbf{S}_a = \{5, 6\}, \\
 \mathbf{S}_a^0 &= \mathbf{P}_{ca} = \mathbf{P}_{da} = \{1, 0\}, \\
 \mathbf{P}_{cb}^0 &= \mathbf{P}_{ab}^0 = \mathbf{S}_b = \{0, 1, 2, 3, 4, 7\}, \\
 \mathbf{S}_b^0 &= \mathbf{P}_{cb} = \mathbf{P}_{cb} = \{2, 3, 4, 5, 6, 7\}, \\
 \mathbf{P}_{ca}^{\wedge 0} &= \langle \{1, 0\}, \{5, 6\} \rangle, \mathbf{P}_{da}^{\wedge 0} = \langle \{0, 1\}, \{5, 6\} \rangle, \mathbf{P}_{cb}^{\wedge 0} = \langle \{4, 7, 5, 6, 3, 2\}, \{0, 1, 2, 3, 4, 7\} \rangle.
 \end{aligned}$$

Звідси впливає наступна структура (табл. 4).

Таблиця 4 – Структура оберненого КМ для молодших розрядів

f_r^0	0	0	1	1	1	1	1	1
$\mathbf{G}_{1,3}$ при $f_h^0 = 0$	6	5	7	4	0	2	3	1
$\mathbf{G}_{1,3}$ при $f_h^0 = 1$	5	6	0	2	1	3	7	4
\mathbf{X}	0	1	2	3	4	5	6	7

Обернений КМ для старших розрядів:

$$\begin{aligned}
 \mathbf{S}_{c(c=0)}^0 &= \{0\}, \mathbf{S}_{d(d=1)}^0 = \{1, 2, 3\}, \\
 \mathbf{T}_{ac(c=0, b=1)}^0 &= \mathbf{T}_{bc(d=1, a=1)}^0 = \{1\},
 \end{aligned}$$

$$\mathbf{T}_{ad(c=0,b=1)}^0 = \mathbf{T}_{bd(d=1,b=1)}^0 = \{0, 2, 3\},$$

$$\mathbf{T}_{ac}^0 = \mathbf{T}_{bc}^0 = \langle \{0\}, \{1\} \rangle, \mathbf{T}_{ad}^0 = \langle \{3, 1, 2\}, \{0, 2, 3\} \rangle, \mathbf{T}_{bd}^0 = \langle \{2, 3, 1\}, \{0, 2, 3\} \rangle.$$

Звідси впливає наступна структура (табл. 5)

Таблиця 5 – Структура оберненого КМ для старших розрядів

f_h^0	0	1	1	1
$G_{4,5}$ при $f_r^0 = 0$	1	2	3	0
$G_{4,5}$ при $f_r^0 = 1$	1	3	0	2
\mathbf{X}	0	1	2	3

Таблиця 6 – Обернена підстановка

$\mathbf{G}(\mathbf{X}) =$	16	15	17	14	10	12	13	11	25	26	30	32	31	33	37	34
$\mathbf{X} =$	00	01	02	03	04	05	06	07	10	11	12	13	14	15	16	17
$\mathbf{G}(\mathbf{X}) =$	35	36	00	02	01	03	07	04	05	06	20	22	21	23	27	24
$\mathbf{X} =$	20	21	22	23	24	25	26	27	30	31	32	33	34	35	36	37

Аналогічно попередньому, шляхом простого перейменування кодів функцій та кодів аргументів в множині пар підстановок, КМ, обернений до КМ типу 2б може мати тип 2б або 2в, а КМ, обернений до КМ типу 2в може мати тип 2в або 2б. Конкретний тип оберненого КМ, в цьому випадку, залежить від типу другого КМ двомодульного каскаду, який реалізує пряму підстановку. Так, для типів (2а-2б) прямої підстановки, обернена підстановка відповідає типам (2а-2б), для типів (2б-2в) – типи (2в-2б), для типів (2в-2в) – типи (2в-2в).

Розглянемо окремі випадки для КМ типу 3. Згідно з попереднім, для КМ типу 3 існують 3 або 4 непусті множини $\mathbf{A}_i (i = 1, 2, \dots, 6)$. Розглянемо, при яких умовах можливе об'єднання множин $\mathbf{A}_i, \mathbf{A}_j (i \neq j)$ в один клас по спільній декомпозиції виду (3). Нехай $r_1 \in \mathbf{A}_i, r_2 \in \mathbf{A}_j, (r_1 \neq r_2, i \neq j)$. Тоді таке об'єднання не можливе, якщо

$$\mathbf{G}_{t+1,n}(\mathbf{X}_{t+1,n}, r_1) \neq \mathbf{G}_{t+1,n}(\mathbf{X}_{t+1,n}, r_2). \quad (7)$$

Вказана нерівність має місце при існуванні принаймні одного такого значення $t \in \mathbf{S}_{t+1,n}$, де $\mathbf{G}_{t+1,n}(t, r_1) \neq \mathbf{G}_{t+1,n}(t, r_2)$. Для цього серед елементів множин пар підстановки \mathbf{T}^\wedge , споріднених до відповідних множин \mathbf{P}^\wedge (тобто тих, де множини кодів функцій \mathbf{P} використовувались при створенні \mathbf{A}_i та \mathbf{A}_j) повинна бути принаймні одна пара, де коди функцій однакові, а коди аргументів – різні.

Очевидно, що ця умова виконується для будь яких множин \mathbf{T}^\wedge , які мають спільні елементи множин кодів функцій та різні множини кодів аргументів. Якщо дві множини \mathbf{T}^\wedge мають однакові множини кодів аргументів та мають спільні елементи множин кодів функцій, то для невиконання (7) необхідне повне співпадання пар підстановок для спільних кодів функцій. З КМ типу 3 сумісні КМ лише типу 2, де множини кодів функцій співпадають. При цьому лише для КМ типу 2а можливе співпадання множин кодів функцій при співпаданні множин кодів аргументів. Звідси впливає, що коли в КМ типу 2а виконується умова:

$$\mathbf{T}_{ac}^\wedge = \mathbf{T}_{bc}^\wedge \text{ або } \mathbf{T}_{ad}^\wedge = \mathbf{T}_{bd}^\wedge, \quad (8)$$

то з будь яким КМ типу 3а утворюється підстановка, обернена до якої також реалізується на простому двомодульному каскаді. Зауважимо, що при виконанні обох умов (8), функції КМ не будуть залежати від значень молодших розрядів

Розглянемо наступний приклад. Нехай КМ в старших розрядах як і раніше задається табл. 2. Відзначимо, що $\mathbf{T}_{ac}^\wedge = \mathbf{T}_{bc}^\wedge = \langle \{1\}, \{0\} \rangle$.

Нехай КМ для молодших розрядів задається наступною табл.7.

Таблиця 7

f_r	0	1	0	0	0	0	1	0
$F_{1,3}$ при $f_h = 0$	4	7	5	6	1	3	0	2
$F_{1,3}$ при $f_h = 1$	2	7	3	5	4	0	1	6
X	0	1	2	3	4	5	6	7

Для даного КМ маємо:

$$\begin{aligned}
 S_{a(a=0)} &= \{0, 2, 3, 4, 5, 7\}, \quad S_{b(b=1)} = \{1, 6\}, \\
 P_{ca(c=0,a=0)} &= \{4, 5, 6, 1, 3, 2\}, \quad P_{da(d=1,a=0)} = \{2, 3, 5, 4, 0, 6\}, \\
 P_{cb(c=0,b=0)} &= \{7, 0\}, \quad P_{db(d=1,b=1)} = \{7, 1\}, \\
 P_{ca}^{\wedge} &= \langle \{0, 2, 3, 4, 5, 7\}, \{4, 5, 6, 1, 3, 2\} \rangle, \quad P_{da}^{\wedge} = \langle \{0, 2, 3, 4, 5, 7\}, \{2, 3, 4, 0, 6\} \rangle, \\
 P_{cb}^{\wedge} &= \langle \{1, 6\}, \{7, 0\} \rangle, \quad P_{db}^{\wedge} = \langle \{1, 6\}, \{7, 1\} \rangle, \\
 A_1 &= P_{ca} \cap P_{da} = \{2, 3, 4, 5, 6\}, \quad A_2 = P_{ca} \cap P_{cb} = \emptyset, \\
 A_3 &= P_{ca} \cap P_{db} = \{1\}, \quad A_4 = P_{da} \cap P_{cb} = \{0\}, \\
 A_5 &= P_{da} \cap P_{db} = \emptyset, \quad A_6 = P_{cb} \cap P_{db} = \{7\}, \\
 M_{ca}^{\wedge} &= T_{ac}^{\wedge} \times P_{ca}^{\wedge}, \quad M_{da}^{\wedge} = T_{ad}^{\wedge} \times P_{da}^{\wedge}, \quad M_{cb}^{\wedge} = T_{bc}^{\wedge} \times P_{cb}^{\wedge}, \quad M_{db}^{\wedge} = T_{bd}^{\wedge} \times P_{db}^{\wedge}, \\
 T_{ac}^{\wedge} &= T_{bc}^{\wedge} \langle \{1\}, \{0\} \rangle, \quad T_{ac}^{\wedge} = \langle \{0, 2, 3\}, \{3, 1, 2\} \rangle, \quad T_{bd}^{\wedge} = T_{bc}^{\wedge} \langle \{0, 2, 3\}, \{3, 2, 1\} \rangle.
 \end{aligned}$$

Враховуючи (7), сформуємо множини M^{\wedge} :

$$\begin{aligned}
 M_{ca}^{\wedge} &= T_{ac}^{\wedge} \times P_{ca}^{\wedge} = \langle \{1\}, \{0\} \rangle \times \langle \{0, 2, 3, 4, 5, 7\}, \{4, 5, 6, 1, 3, 2\} \rangle, \\
 M_{da}^{\wedge} &= T_{ad}^{\wedge} \times P_{da}^{\wedge} = \langle \{0, 2, 3\}, \{3, 1, 2\} \rangle \times \langle \{0, 2, 3, 4, 5, 7\}, \{4, 3, 5, 4, 0, 6\} \rangle, \\
 M_{cb}^{\wedge} &= T_{bc}^{\wedge} \times P_{cb}^{\wedge} = \langle \{1\}, \{0\} \rangle \times \langle \{1, 6\}, \{7, 0\} \rangle, \\
 M_{db}^{\wedge} &= T_{bd}^{\wedge} \times P_{db}^{\wedge} = \langle \{0, 2, 3\}, \{2, 3, 1\} \rangle \times \langle \{1, 6\}, \{7, 1\} \rangle.
 \end{aligned}$$

Таблиця 8 – Пряма підстановка

F(X)	32	27	33	35	34	30	21	36	04	07	05	06	01	03	00	02
X	00	01	02	03	04	05	06	07	10	11	12	13	14	15	16	17
F(X)	12	37	13	15	14	10	31	16	22	17	23	25	24	20	11	26
X	20	21	22	23	24	25	26	27	30	31	32	33	34	35	36	37

Для формування оберненої підстановки використовуємо обернений КМ для старших розрядів, заданий в табл. 5.

Розглянемо побудову оберненого КМ для молодших розрядів. Маємо:

$$\begin{aligned}
 M_{ca}^{\wedge 0} &= T_{ac}^{\wedge 0} \times P_{ca}^{\wedge 0} = \langle \{0\}, \{1\} \rangle \times \langle \{4, 5, 6, 1, 3, 2\}, \{0, 2, 3, 4, 5, 7\} \rangle, \\
 M_{da}^{\wedge 0} &= T_{ad}^{\wedge 0} \times P_{da}^{\wedge 0} = \langle \{3, 1, 2\}, \{0, 2, 3\} \rangle \times \langle \{2, 3, 5, 4, 0, 6\}, \{0, 2, 3, 4, 5, 7\} \rangle, \\
 M_{cb}^{\wedge 0} &= T_{bc}^{\wedge 0} \times P_{cb}^{\wedge 0} = \langle \{0\}, \{1\} \rangle \times \langle \{7, 0\}, \{1, 6\} \rangle, \\
 M_{db}^{\wedge 0} &= T_{bd}^{\wedge 0} \times P_{db}^{\wedge 0} = \langle \{2, 3, 1\}, \{0, 2, 3\} \rangle \times \langle \{7, 1\}, \{1, 6\} \rangle,
 \end{aligned}$$

Сформуємо множини **A**:

$$\begin{aligned} A_1 &= P_{ca} \cap P_{da} = \{2, 3, 4, 5, 6\}, \quad A_2 = P_{ca} \cap P_{cb} = \emptyset, \\ A_3 &= P_{ca} \cap P_{db} = \{1\}, \quad A_4 = P_{da} \cap P_{cb} = \{0\}, \\ A_5 &= P_{da} \cap P_{db} = \emptyset, \quad A_6 = P_{cb} \cap P_{db} = \{7\}. \end{aligned}$$

Оскільки елементи множини A_1 є кодами аргументів для множин $M_{ca}^{\wedge 0}$ та $M_{da}^{\wedge 0}$ і елементи множини A_4 є кодами аргументів для множин $M_{da}^{\wedge 0}$ та $M_{cb}^{\wedge 0}$, а $T_{ac}^{\wedge 0} = T_{bc}^{\wedge 0}$, то елементи множин A_1 та A_4 входять в один і той же клас еквівалентності по декомпозиції (3). Дійсно, легко бачити, що для будь якого елемента r цих множин функція $G_{4,5}(X_{4,5}, r)$ визначається рядком при $f_r^0 = 0$ або рядком при $f_r^0 = 1$ в табл.5. Звідси впливає наступна структура оберненого КМ для молодших розрядів.

Таблиця 9 – Обернений КМ типу 3а для молодших розрядів.

f_r^0	0	1	0	0	0	0	0	1
$G_{1,3}$ при $f_h^0 = 0$	6	4	7	5	0	2	3	1
$G_{1,3}$ при $f_h^0 = 1$	5	6	0	2	4	3	7	1
X	0	1	2	3	4	5	6	7

В результаті буде реалізована підстановка, яка задана в табл. 10. Легко бачити, що ця підстановка є оберненою до підстановки, яка задана в табл. 8

Зауважимо, що всі отримані результати мають місце, якщо доведення проводить базуючись на множинах T , а не P .

Таблиця 10 – Обернена підстановка

16	14	17	15	10	12	13	11	25	36	20	22	24	23	27	31
00	01	02	03	04	05	06	07	10	11	12	13	14	15	16	17
35	06	30	32	34	33	37	01	05	26	00	02	04	03	07	21
20	21	22	23	24	25	26	27	30	31	32	33	34	35	36	37

Висновки

Таким чином встановлено:

- 1) Якщо підстановки реалізуються простим двомодульним каскадом з використанням КМ лише типу 2 (2а, 2б, 2в), то обернена підстанова також реалізується на простому двох модульному каскаді.
- 2) При використанні КМ типів 3б та 3в обернена підстанова не може бути реалізована на простому каскаді (з одним боковим виходом).
- 3) При використанні КМ типу 3а реалізація оберненої підстановки на простому двох модульному каскаді можлива за умови, що в КМ типу 2а при фіксованому значенні на боковому виході (0 або 1) функції КМ на первинних виходах не залежать від значення на боковому вході.
- 4) Для КМ типу 2а або 3а обернений КМ має відповідно тип 2а або 3а. Для КМ типу 2б обернений КМ має тип 2б, якщо пряма підстанова реалізована парою 2а-2б і тип 2в, якщо пряма підстанова реалізована парою 2в-2б. Для КМ типу 2в обернений КМ має тип 2б, якщо пряма підстанова реалізована парою 2в-2б і тип 2в, якщо пряма підстанова реалізована парою 2в-2в.

Можливим напрямком подальших досліджень є поширення одержаних результатів для більш складних КМ та реалізації прямих та обернених підстановок на багатомодульних каскадах.

Література

- 1) Опанасенко В.Н., Сахарин В.Г. ПЛИС типа FPGA фирмы Xilinx: возможности, проектирование и применение// Электронные системы и компоненты. – 2003, № 4, с.7-11
- 2) Кузелин М.О., Кнышев Д.А., Зотов В.Ю. Современные семейства ПЛИС фирмы Xilinx. – М: «Горячая линия-Телеком», 2004, 440 с.
- 3) Тарасенко В.П., Тесленко О.К., Яновська О.Ю. Проблеми апаратної реалізації підстановок. Наукові записки УНДІЗ, №2, 2007, с 52-58
- 4) В.П. Тарасенко, О.К. Тесленко, О.Ю. Яновська. Реалізація повних підстановок на простому двомодульному каскаді конструктивних модулів. Інформаційні технології та комп'ютерна інженерія. №1(11), 2008, с.88-97

5) Кантор Г. Труды по теории множеств.- Москва. Наука, 1985.

6) Логачев О.А., Сальников А.А, Яценко В.В. Булевы функции в теории кодирования и криптологии. – Москва. Издательство МЦНМО, 2004.

Відомості про авторів

Тарасенко Володимир Петрович, д.т.н., проф. Завідуючий кафедрою системного програмування і спеціалізованих комп’ютерних систем НТУУ «КПІ».

Тесленко Олександр Кирилович, к.т.н., ст.н.с. Доцент кафедри системного програмування і спеціалізованих комп’ютерних систем НТУУ «КПІ».

Яновська Олена Юрївна, Асистент кафедри звукотехніки і реєстрації інформації НТУУ «КПІ».

УДК 004.27; 004.31; 004.382.2.

Ю. С. ЯКОВЛЕВ

Институт кибернетики имени В.М. Глушкова НАН Украины, г. Киев

ОСОБЕННОСТИ АРХИТЕКТУРЫ И СТРУКТУРЫ РЕКОНФИГУРИРУЕМЫХ ГРАФИЧЕСКИХ УСКОРИТЕЛЕЙ

Анотация. розглянуто особливості побудови графічних прискорювачів з реконфігурацією шляхом використання: комутаційних схем для вибору оптимальних ресурсів прискорювача; масштабування системи за рахунок використання модифікованої кільцевої шини; застосування ПЛИС для конфігурації прискорювача під тип вирішуваної задачі; архітектури типу "процесор-в-пам'яті" із застосуванням запропонованого методу розподілу графічної задачі по процесорах системи. При цьому запропоновані архітектурно-структурні рішення захищені патентами України.

Ключові слова: реконфігурація, графічний прискорювач, "процесор-в-пам'яті", програмована логічна інтегральна схема (ПЛИС).

Аннотация. рассмотрены особенности построения графических ускорителей с реконфигурацией путем использования: коммутационных схем для выбора оптимальных ресурсов ускорителя; масштабирования системы за счет использования модифицированной кольцевой шины; применения ПЛИС для конфигурации ускорителя под тип решаемой задачи; архитектуры типа "процессор-в-памяти" с применением предложенного метода распределения графической задачи по процессорам системы. При этом предложенные архитектурно-структурные решения защищены патентами Украины.

Ключевые слова: реконфигурация, графический ускоритель, "процессор-в-памяти", программируемая логическая интегральная схема (ПЛИС)

Abstract. features of construction of accelerators with reconfiguration by usage are observed: diagrammes of connections for sampling of optimal resources of the accelerator; scalings of system at the expense of usage of the updated ring bus; applications the PLIS for an accelerator pattern under type of the solved task; architectures of type "processor-in-memory" with application of the offered method of allocation of the graphics task on system processors. Thus offered is architectural-structural solutions are protected by patents of Ukraine.

Keywords: reconfiguration, the accelerator, the "processor-in-memory", the programmed logical chip (PLC)

Введение

Обработка графических изображений используется в различных областях, в том числе в играх для отображений различных игровых сцен, в медицине, например, при обработке гистограмм, снимков и т.д., в военном деле, в системах автоматизированного проектирования (САПР) и т.д. При этом на каждом этапе развития средств вычислительной техники (ВТ) требования к качеству и скорости обработки изображений постоянно возрастают, что в принципе может быть достигнуто за счет расширения полосы пропускания информации по каналу процессор-память, совершенствования системы памяти, обеспечивающей малое время доступа к данным, применением более мощных процессоров и множества других факторов, которые в совокупности, как правило, представлены в существующем множестве графических ускорителей основных фирм – производителей АТІ (AMD) и NVIDIA [1, 2]. Практически во всех указанных случаях их применения заметный положительный эффект может быть достигнут, если алгоритмы подлежат хорошему распараллеливанию и содержат большое количество простых однотипных вычислений. С этой точки зрения, обработка изображения как раз и является той областью, где распараллеленная часть алгоритма, как правило, имеет большой процент по сравнению с другими целевыми прикладными программами.

Помимо графических ускорителей для повышения качества и скорости обработки изображений используются также различные средства, выполненные на микросхемах ASIC (например, устройства цифровой обработки сигналов – ЦОС), на микросхемах памяти ("Processor-in-memory" или PIM-системы), обширный класс которых относят к категории CIMA (Computing-in-Memory Architecture), а также средства, обладающие свойствами реконфигурации, в том числе – использующие программируемые логические интегральные схемы (ПЛИС)

Актуальность

Актуальность данной работы продиктована, с одной стороны, необходимостью улучшения качества отображения графической информации за счет совершенствования средств и способов обработки графики и, с другой стороны, – отсутствием систематизированного подхода к анализу и построению перспективных архитектурно-структурных решений средств обработки графики, обеспечивающих по сравнению с известными более высокое качество и скорость обработки изображений при меньших затратах ресурсов системы и улучшенных параметрах пользовательских характеристик (например, энергопотребления и тепловыделения).

Цель

В соответствие с этим, целью данной работы является анализ и отображение существующих и перспективных архитектурно - структурных решений, обладающих свойствами реконфигурации, выполненных на современной элементной базе и ориентированных на обработку графики, включая PIM-системы, а также программируемые логические интегральные схемы – ПЛИС и однокристалльные

графические процессоры. При этом не рассматриваются, так называемые, графические ускорители фирм ATI (AMD) и NVIDIA, анализу и применению которых посвящены многочисленные публикации.

Задачи

Основными задачами являются:

- 1) Анализ особенностей архитектурно-структурной организации специализированных графических процессоров, выполненных на современной элементной базе и ориентированных на обработку графики (например, совершенствованные технологии VLSI, технологии цифровой обработки сигналов ЦОС и др.)
- 2) Анализ особенностей архитектурно-структурной организации и применения средств для обработки графических изображений, в том числе PIM-систем, обладающих свойствами реконфигурации.
- 3) Сравнение основных параметров реконфигурируемых графических систем, выполненных с применением различных архитектурно-структурных решений.

Решение задач

Все возрастающие требования к качеству и скорости обработки изображения приводит к выбору высокопроизводительного микропроцессора при построении соответствующих систем, для которых обычно предъявляются весьма строгие требования к таким параметрам, как габариты, стоимость, потребляемая мощность и время вывода нового изделия на рынок. Эти требования часто не могут быть удовлетворены только применением более мощного микропроцессора. Поэтому в промышленных изделиях функции обработки изображения осуществляются как с применением графических ускорителей фирм ATI (AMD) и NVIDIA, так и с применением специализированных процессоров таких, как процессоры цифровой обработки сигналов (ЦОС) или – Digital Signal Processors (DSPs), а также с помощью специфического программного обеспечения Application Specific Standard Products (ASSPs). Однако, при увеличении сложности обработки изображения для ЦОС необходимо использовать большое количество параллельных модулей. Такие скоростные ЦОС становятся дорогими, и их производительность имеет тенденцию отставать от соответствующих требований при обработке изображения. С другой стороны, ASSPs являются негибкими, дорогими и трудоемкими, отнимающими много времени при модификации, несмотря на то, что врожденный параллелизм при обработке изображения предлагает приложению высокую эффективность технологии вычислений (High Performance Computing - HPC).

Аппаратное ускорение – эффективный способ увеличить производительность при использовании специализированной аппаратной архитектуры, которая выполняет параллельную обработку. С появлением программируемой вентильной матрицы (Field Programmable Gate Array – FPGA), специализированная аппаратная архитектура может быть реализована с более низкой стоимостью, при этом мощность потребления может быть уменьшена, поскольку схема оптимизируется под приложение. Реконфигурируемое вычисление может также уменьшить размер канала “память-процессор” (и, следовательно, стоимость) и обеспечить дополнительную гибкость за счет динамической реконфигурации во время выполнения приложения. В результате производительность обработки изображения может быть улучшена по сравнению с мощным микропроцессором на несколько порядков, что соответствует требованиям для многих приложений.

Операции обработки изображения включают большое количество операции умножения-накопления (MAC). При этом микропроцессоры выполняют несколько последовательных шагов (выборка машинной команды, декодирование, выполнение и получение результата на запрос) для каждой операции MAC. Так как эти шаги выполняются последовательно, то трудно ускорить вычисление без архитектурных изменений. FPGAs может обеспечить высокоэффективное MAC – вычисление, используя специализированную архитектуру, которая в состоянии выполнить несколько операций MAC одновременно

Аппаратные сопроцессоры – ускорители базовых операций обработки графики. Ниже рассматриваются две разновидности архитектур, которые вычисляют базовые операции алгоритма обработки изображений: взаимную корреляцию в пространственных и спектральных областях. Другие операции, такие как фильтрация и свертка можно рассматривать как специфические случаи корреляционных [3]. Обе архитектуры могут быть реконфигурированы, чтобы соответствовать различным приложениям и размерам FPGA. На рис. 1 представлена упрощенная архитектура аппаратного средства для обработки графики, которая отнесена к категории пространственной. Основой этой архитектуры является систолический массив модулей MAC, подключенных к памяти, где каждый модуль MAC выполняет операции и посылает результат к следующему модулю в той же самой строке. Последний MAC каждой строки выдает результат вычисления всей строки, который подается на вход MAC следующей строки и т.д. (рис.1). Результат вычисления матрицы получают путем суммирования результатов последовательных строк N при накоплении результата предыдущей строки к следующей.

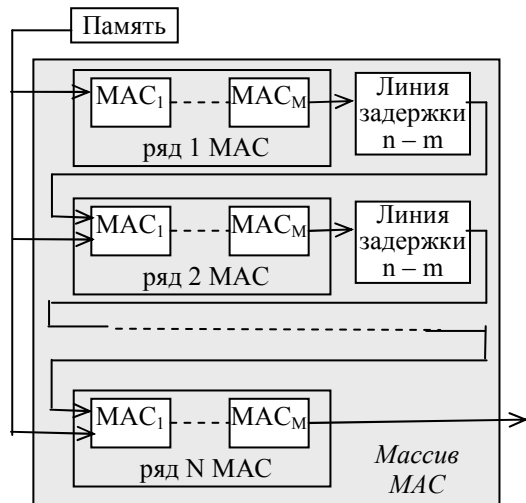


Рисунок 1 – Фрагмент пространственной архитектуры реконфигурируемого графического ускорителя

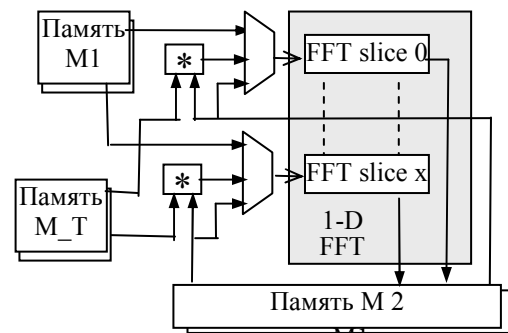


Рисунок 2 – Фрагмент архитектура для вычисления взаимной корреляция в спектральной области

Такая архитектура ориентирована на выполнение операций линейной фильтрации, свертки и методов корреляции, которые обычно используются во многих приложениях обработки изображения.

Строка задержки $N - M$ циклически повторяется и вставляется между выводом последнего MAC модуля в каждой i -й строке и вводом результата в первый модуль MAC в следующей $(i + 1)$ строке. При этом схемы задержки могут быть реализованы как сдвиговые регистры или как память типа FIFOs в зависимости от выбранной технологии и типа устройства. Для этого итерационного процесса требуется промежуточная память для хранения результата каждого уровня, а также мультиплексор между массивом MAC и памятью, чтобы выбрать соответствующие входные данные при каждой итерации.

Недостатком этой архитектуры является необходимость применения линий задержки и малая емкость памяти, размещенная на кристалле FPGA. Поэтому при организации на FPGA памяти приходится увеличивать её емкость, одновременно уменьшая тактовую частоту в два раза. При этом увеличение ширины шины памяти в два раза обеспечивают требуемую массивом MAC скорость за счет передачи данных одновременно для двух последовательных пикселей.

На рис. 2 представлена архитектура для вычисления взаимной корреляция (CC) в спектральной области [3]. В этой архитектуре 1-D FFT массив используется вместо массива MAC, приведенного на рис.1. Для вычисления взаимной корреляция в спектральной области изображение разбивается на строки или столбцы, чтобы выполнить параллельную обработку. Модули FFT (Fast Fourier Transform) вычисляют строки/столбцы FFTs и называются секторами FFT.

Максимальное число распределенных секторов FFT зависит от размера изображения и выбранного устройства для реализации. В этом процессе используются три различных блока памяти: блок памяти M1 (рис. 2), предназначенный для запоминания входного изображения и результата, блок памяти M_T для запоминания FFT(T) и блок памяти M2 для запоминания промежуточных результатов. У всех этих блоков памяти – одинаковая емкость, часть которой предназначена для хранения размера входного изображения, а другая – для хранения результатов вычисления FFT. Эти блоки памяти могут быть выполнены, используя внутренние блоки микросхемы FPGA. Блоки памяти M1 и M2 являются независимыми, чтобы обеспечить достаточное количество портов памяти для каждого сектора FFT. Чтобы избежать конфликтов при доступе к памяти, обращения к блокам памяти осуществляют по диагонали при сдвиге во времени секторов FFT друг относительно друга. Описанная архитектура полезна также для свертки и фильтрации спектральной области, и требуется только незначительные изменения: в частности при реализации свертки удаляют сопряженную операцию при выводе информации из блока памяти M_T.

Анализ экспериментальных результатов, приведенных в [3], показал, что при применении FPGA можно достичь ускорения больше, чем на два порядка величины относительно программных реализаций. Например, пространственная аппаратная реализация архитектуры на FPGA для шаблона 16×16 обеспечивает в 688 раз быстрее, чем эквивалентная программная реализация и в 244 раза быстрее, чем спектральная реализация на современном персональном компьютере без потери точности. При этом можно отметить, что у пространственной архитектуры – умеренное потребление ресурсов FPGA, которое увеличивается с увеличением размеров шаблона. Основное ограничение для пространственной

архитектуры - число секторов DSP, доступных в выбранном устройстве FPGA. С другой стороны, спектральная архитектура потребляет большинство ресурсов FPGA, включая логику, внедренные блоки памяти и DSP сектора, и только немного быстрее, чем пространственная реализация. Однако, спектральная архитектура может быть выгодной для больших размеров шаблона, поскольку продолжительность обработки не зависит от размера шаблона.

Реконфигурируемые однокристалльные графические ускорители. Рассмотренные выше архитектуры ориентированы на выполнение отдельных (базовых) операций алгоритма обработки изображений. Чтобы управлять потоком данных и выполнить остальную часть алгоритма, необходима интеграция с другими элементами (микропроцессором или PC).

Эта проблема решается в пределах сценария “Система на Программируемом Чипе”(“System on Programmable Chip” – SoPC). Микропроцессор, размещенный на FPGA, подключен к аппаратному сопроцессору, который ускоряет самые сложные операции. Поскольку все компоненты SoPC располагаются в единственном чипе, то соответственно уменьшается размер, стоимость и потребляемая мощность. Для исследования реальной системы была выбрана микросхема Virtex 5 FPGA, XC5V50T фирмы Xilinx. Система предназначена для высокоэффективной обработки изображения и содержит внедренный микропроцессор, блоки памяти и коммуникационные шины между элементами, а также аппаратный сопроцессор, который ускоряет критические операции. Упрощенная структурная схема такой системы приведена на рис.3. При этом сопроцессор обладает свойствами реконфигурации и спроектирован с использованием пространственной архитектуры из-за её многосторонности (многофункциональности) и меньшим потреблением ресурсов. Архитектура сопроцессора показана на рис. 4.

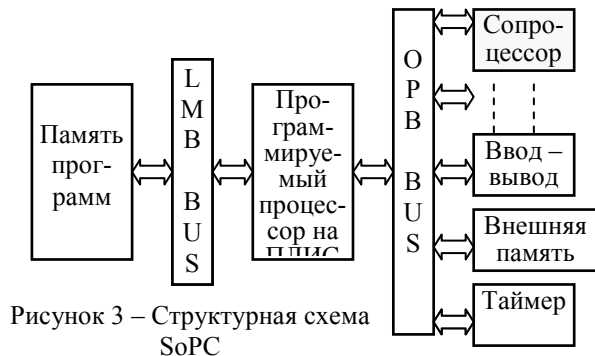


Рисунок 3 – Структурная схема SoPC

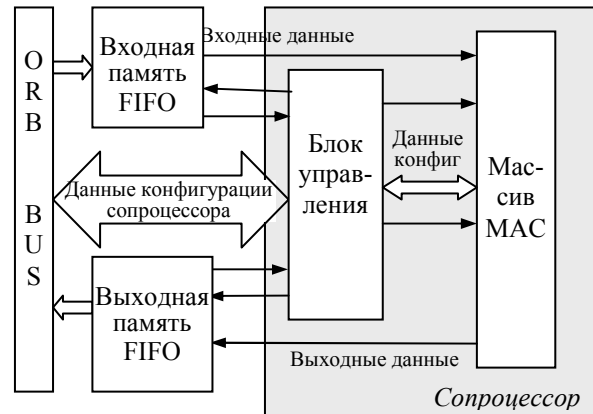


Рисунок 4 – Упрощенная архитектура сопроцессора (рис.3)

Параметры конфигурации упаковываются в 4-х 32-разрядных регистрах так, чтобы полная конфигурация сопроцессора могла быть осуществлена для 4-х транзакций OPB. Как только сопроцессор запрограммирован, обработка данных начинается немедленно, пока данные доступны при вводе в порядке поступления. Сопроцессор работает параллельно с порядком поступления, и выходные данные памяти выводятся также в порядке поступления. Данные ввода-вывода передаются из/в оперативную память по каналу прямого доступа в память без вмешательства микропроцессора.

Как только результаты готовы, они сохраняются в порядке поступления и посылаются в микропроцессор через OPB. Реконфигурируемая производительность сопроцессора была измерена для предложенного алгоритма. Результаты были получены с и без сопроцессора для сравнения. Программные результаты были получены с Pentium IV PC, 3.2 ГГц и 2 Гбайта оперативной памяти. Потребление ресурса сопроцессора главным образом определяется размером массива MAC и длиной строк задержки. Потребление ресурса FPGA умеренно. Половина FPGA остается пустой, поэтому больше компонентов может быть добавлено в систему SoPC. Результаты SoPC были проанализированы без обнаружения потерь точности в алгоритмах. Операционная частота SoPC составляла 100 МГц, она является и частотой OPB. Производительность была измерена для предложенных алгоритмов. Эти измерения также взяты для SoPC без сопроцессора и PC. SoPC работает намного быстрее (до 49 раз быстрее) чем программируемый процессор на ПЛИС (Microblaze) и немного быстрее, чем PC. Для больших размеров изображений SoPC в 157 раз работает быстрее, чем Microblaze и приблизительно в 3 раза быстрее чем PC.

Таким образом, экспериментальные результаты демонстрируют, что все представленные решения обеспечивают ускорение до 3 порядков величины по отношению к эквивалентной программной реализации. Эта архитектура может быть реконфигурируема для многих приложений и к потребностям

многих систем обработки изображений. Архитектура является масштабируемой для любого семейства FPGA и приспособляемой к любому FPGA. Чтобы получить аппаратное ускорение при обработке изображения, реконфигурируемые аппаратные средства и сопроцессор интегрируются в SoPC. Эта система может обеспечить гибкость, микропроцессор, который размещен на FPGA, обеспечивая тем самым высокую производительность при выполнении сложных (критических) операций.

Реконфигурируемые графические ускорители с архитектурой "Processor-in-memory" (процессор-в-памяти). Большинство PIM – архитектур лучше всего функционирует как сопроцессоры компьютерных систем, выполняя функции, для которых они оптимизированы, в то время как другие прикладные программы, в которых основной является последовательная часть, реализуются главным (хост) процессором. Следует отметить, что PIM-система хорошо приспособлена под обработку изображения [4], поскольку обеспечивает возможность параллельной обработки с помощью множества процессоров (процессорных ядер – ПЯ) большого количества фрагментов алгоритма (несколько сотен и более), выполняя одновременные соответствующие вычисления для сотен и тысяч пикселей изображения. При этом, благодаря особенностям архитектурно-структурной организации PIM-систем, обеспечивается широкая полоса пропускания по каналу процессор-память, что существенно сказывается на повышении производительности графической системы в целом. Из-за малых размеров данных, связанных с большинством операций обработки графики (8 – 16 бит), вычислительная мощность современных 32-или 64-разрядных процессоров с возможностями обработки с плавающей запятой, не является необходимой. Вместо этого, матрица или вектор из более простых 8-разрядных элементарных процессоров (PEs), специализированных к определенным потребностям применений, может быть весьма эффективной

Начало обработки изображений с помощью средств типа PIM-систем положило простейшее устройство с архитектурой SIMD – Single-Instruction stream and Multiple-Data stream (один поток команд – поток множества данных), содержащее на одном кристалле множество одноразрядных средств обработки со схемами сдвига влево и вправо, каждый из которых подключен к одному столбцу данных, размещенных в памяти [5]. Это один из первых нового поколения CIMA (Computing-in-Memory Architecture) проектов, который предложен и построен как CRAM проект, разработанный Университетом Торонто в 1992. Благодаря распараллеливанию алгоритма обработки изображений и в итоге широкой эквивалентной полосе пропускания по каналу множество процессоров – множество столбцов данных, размещенных в памяти, это устройство, используемое как приставка-ускоритель к основному процессору, обеспечивает производительность порядка Тера (Тера) команд в секунду. Устройство подобного типа также предложено в [6], где часть вычислений выполняется в узле PIM, однако основные вычисления выполняются на суперскалярном процессоре MP. Тем самым PIM-система используется в качестве ускорителя МП. При этом, для увеличения производительности системы, увеличивают количество узлов памяти типа PIM. Подобный простой проект PIM - устройства Terasys, был объявлен в 1995. Этот проект содержит большую матрицу простых элементов вычисления (типично более чем 1.000) которые сформированы в матрице DRAM. Элементарные процессорные элементы объединены по выходному сигналу усилителей считывания и управляются единственным устройством управления. Поэтому эта архитектура работает как SIMD архитектура - в любом цикле обращения к памяти типа DRAM, каждый процессор в матрице выполняет одну и ту же команду над её собственными локальными данными. Несмотря на то, что с архитектурной точки зрения, этот проект – самый простой, тем не менее, он может теоретически обеспечить высокую производительность при реализации алгоритмов с хорошим распараллеливанием. Однако, эти типы массово – параллельных, однопроцессорных элементов SIMD-приборов в составе PIM-системы могут быть эффективны только на наборе прикладных программ, которые могут быть легко распараллелены. Для прикладных программ с существенным количеством последовательных вычислений, увеличение скорости за счет применения этой архитектуры – ограничено.

В целом можно отметить, что в настоящее время диапазон проектируемых устройств, которые подпадают под архитектурный стиль "Processor-in-memory", богат и различен [3, 4], и большинство из них могут успешно применяться для обработки графических изображений.

Дальнейшее развитие PIM-систем, ориентированных на обработку графики, шло одновременно с развитием интегральной технологии по пути совершенствования процессорных элементов, коммутационной среды и обеспечения возможностей реконфигурации с целью настройки архитектуры на особенности приложения и возможностей реконфигурируемой среды (микросхемы ПЛИС). С этой точки зрения особый интерес представляют архитектура PIM-систем типа VIRAM (vector-intelligent-random access memory) и DIVA (Data IntensiVe Architecture – интенсивная архитектура данных) [7]. При этом VIRAM, используя новую технологию PIM, содержит встроенную в кристалл динамическую оперативную память (DRAM) с векторным сопроцессором, обеспечивая тем самым эффективное использование полосы пропускания по каналу память-процессор. Особенностью архитектура чипа

VIRAM является то, что это – законченная система на чипе, содержащая элементарные процессоры и стандартную динамическую оперативную память (DRAM).

PIM-система типа DIVA содержит множество чипов PIM в качестве интеллектуальных сопроцессоров и использует широкие информационные каналы памяти большой емкости, применяя при этом мелко модульную параллельность. Каждый чип DIVA PIM-системы включает ЗУ с аппаратными средствами вычисления и коммуникации и содержит два информационных канала, механизмы которых координируются единственным блоком управления: 32-разрядный скалярный информационный канал и 256-разрядный широкий информационный канал. Скалярный информационный канал – стандартная архитектура RISC-процессора с расширенными функциями DIVA-specific для того, чтобы координироваться с широким информационным каналом. Широкий информационный канал воздействует на составные объекты (суперслова) 256 бит, выполняя параллельные работы SIMD на установленных по размеру полях объекта (8, 16, и 32-разрядных полях). В дополнение к условным арифметическим действиям и логическим операциям, широкое арифметико-логическое устройство также поддерживает большой набор операций для того, чтобы манипулировать данными, включая перегруппировку данных в пределах широкого слова-операнда, передачи между широкими и скалярными регистрами, собирая в структуру и распаковывая операции. Кроме того, широкое арифметико-логическое устройство поддерживает выборочное выполнение команд на базе информационного канала в зависимости от состояния кодов условия.

Графический ускоритель, построенный на принципах потоковой обработки. Другой подход к повышению производительности системы за счет эффективного использования ширины полосы пропускания памяти основан на принципах потоковой обработки. Например, программируемый потоковый микропроцессор Imagine [7], ориентирован на обработку интенсивных приложений, отличающихся высоким параллелизмом данных с небольшим глобальным многократным их использованием. Imagine – это 32-разрядная архитектура с поддержкой выполнения операций для 16-и 8-разрядных данных, обеспечивая в результате в два и в четыре раза повышение пиковой производительности соответственно.

В [8] предлагается архитектура PIM-системы, ориентированная на обработку компьютерной графики с массовым обращением к памяти. Чтобы достигнуть максимальной производительности, рабочая нагрузка должна быть равномерно распределена среди процессоров, а размещение данных и распределение задачи пользователя должны уменьшить коммуникацию между процессорами,

Предложения по повышению эффективности графических ускорителей, выполненных на PIM-системах.

Реконфигурируемая PIM-система с выбором оптимальных ресурсов средствами коммутации

Совершенствование PIM-системы, ориентированной на обработку графики, предлагается путем расширения её функциональных возможностей за счет динамической настройки архитектуры системы на класс решаемых задач, а также за счет оптимальной загрузки всех процессорных элементов полезной (вычислительной) работой. При этом за основу принимаются: максимальный учет особенностей построения и применения существующих PIM-систем и методов их организации и ориентация PIM-систем на обработку хорошо распараллеливаемых алгоритмов при массовом обращении к памяти за данными, а также возможность реализации PIM-чипа на современных и перспективных ПЛИС. Показано, что основные свойства ПЛИС как элементной базы соответствуют особенностям структурной организации PIM-систем, на основании чего определены факторы, определяющие пригодность PIM-системы для реализации на ПЛИС [9].

Для расширения функциональных возможностей на кристалле PIM-системы размещены (на рис. 5 закрашены серым цветом) [10]:

- набор (вместо одного) проблемно-ориентированных (или специализированных) ведущих процессоров (ВП) вместе с соответствующими блоками памяти типа КЭШ;
- блок анализа входного управляющего пакета (БОУП1) для выработки всей совокупности управляющих сигналов, необходимых для функционирования БИС;
- селектор С1 для выбора и активации ведущих процессоров;
- набор селекторов С2 и С3 для выбора процессорных ядер (ПЯ) и соответствующих банков памяти;
- устройство для загрузки программ и данных в массивы памяти (УЗг);
- устройство интерфейса диагностики и отладки (Ин.Д/О)

Селекторы (коммутаторы) С1 – С3 управляются блоком обработки управляющего пакета и формирования управляющих сигналов (БОУП1), необходимых для работы всей системы.

Управляющий пакет содержит поля и признаки, отражающие последовательность действий PIM-системы, обеспечивающих: настройку архитектурно-структурного образа системы перед запуском её на решение конкретной задачи, перестройку системы в процессе работы, а также различные режимы работы системы памяти, в том числе: в качестве обычной памяти с использованием емкости памяти,

размещенной только на кристалле (чипе), а также на кристалле и дополнительной внешней памяти, подключенной к чипу через интерфейс внешней памяти; в качестве "процессора-в-памяти" с использованием ресурсов обработки информации только собственного кристалла, а также с использованием собственных ресурсов обработки информации и дополнительных ресурсов за счет других чипов, подключенных через интерфейс ввода/вывода и интерфейс внешней памяти.

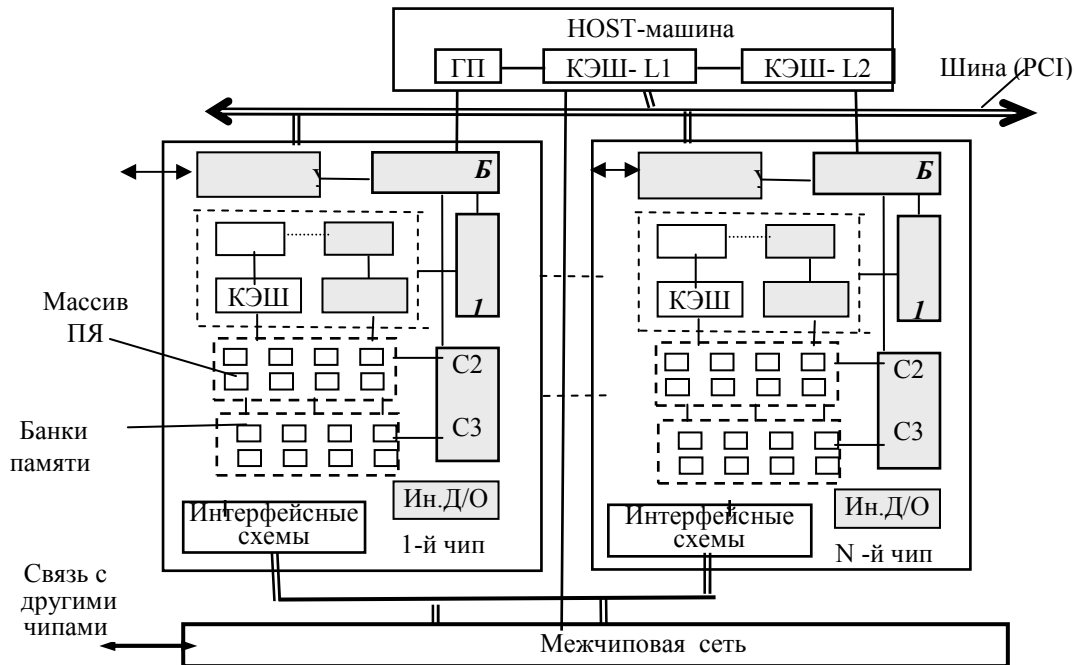


Рисунок 5 – Увеличенная структурная схема реконфигурируемой PIM-системы

Кроме того, управляющий пакет обеспечивает формирование управляющих сигналов и сигналов синхронизации для работы всех узлов системы памяти: управление работой КЭШ-памяти с учетом реализации механизмов когерентности; реализация режимов "Чтение/Запись" при обращении к памяти внутри чипа и за его пределами через интерфейс внешней памяти и др.

Преимущества предлагаемой PIM-системы по сравнению с известными PIM-системами прежде всего состоит в том, что в ней реализована возможность повышения производительности системы (от 2,5 до 10 раз) за счет исходной настройки с помощью селекторов C1 – C3 архитектуры на класс решаемых задач, а также её перестройки в процессе работы (в динамике). Это достигается путем использования оригинальных схмотехнических решений и соответствующих режимов работы БИС памяти, обеспечивающих: выбор оптимального ведущего процессора, выбор необходимого массива памяти для решения конкретной задачи с возможностью подключения дополнительной памяти, находящейся за пределами чипа; выбор оптимальной разрядности данных и количества обрабатываемых слов в пределах всей строки массива памяти, ориентируясь на соответствующие методы обработки и имеющиеся ресурсы для распараллеливания алгоритма решаемой задачи.

Кроме того, существенно снижена мощность, потребляемая одним чипом за счет исключения "холостых пробегов" процессорных ядер массива памяти, так как они загружены только полезной вычислительной работой. При этом сокращается количество одновременно передаваемых бит, так как передачи по линиям связи не всегда реализуются полной N-разрядной строкой (как в известных структурах подобного типа), а при необходимости произвольными кратными R-разрядными группами (например, частями строки разрядностью R, 2R, ..., αR ; $\alpha R \leq N$). Это обеспечивает возможность более глубокого распараллеливания процесса обработки информации при тех же исходных ресурсах за счет повышения эффективности использования процессоров. Увеличивается также и серийность PIM-чипа и соответственно уменьшается стоимость, поскольку он из специализированного преобразован в чип широкого назначения за счет настройки его архитектуры и необходимых ресурсов на различные типы решаемых задач, в том числе – в части обработки графики.

Реконфигурируемая PIM-система с кольцевой модифицированной шиной. Для дальнейшего повышения эффективности применения PIM-систем в качестве графических ускорителей необходимо было решить следующие задачи: упростить коммутационную среду между компонентами системы,

сохранив при этом возможности распараллеливания алгоритма обработки графики и исключив ограничения по её масштабированию. Разместить на том же кристалле блок служебных функций по управлению памятью, сократив при этом время реализации наиболее важных его функций, в частности – время распределения реализуемого алгоритма по процессорам системы.

На основе анализа существующих РІМ-систем и выявления их недостатков предложена оригинальная архитектурно-структурная организация системы такого класса с модифицированной кольцевой шиной, которая содержит на одном кристалле вместе со средствами обработки и хранения данных, набор блоков, которые реализуют служебные функции по управлению памятью и распределения программ обработки графики по процессорам системы. Блок-схема такой РІМ-системы приведена на рис.6, которая защищена патентом Украины на изобретение [11].

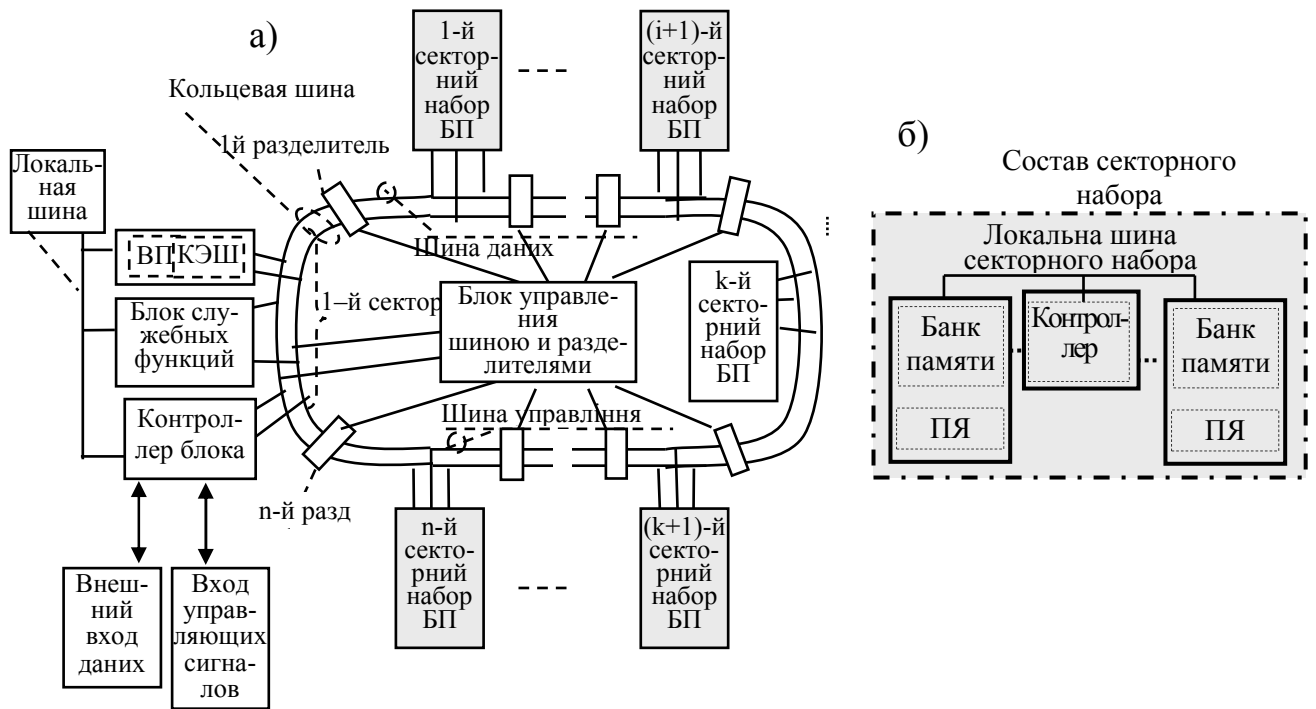


Рисунок 6 – РІМ-система со средствами поддержки вычислительного процесса: а) общая структура РІМ - системы з кольцевою шиною; б) состав секторного набора

Особенностью такой структуры является то, что в кольцевую шину введены секторные разделители, которые управляются специальным блоком. Два соседних секторных разделителя выделяют сектор кольцевой шины, длина которого меньше длины всей кольцевой шины в n раз, где n – количество секторов, при этом емкостная нагрузка на секторную часть кольцевой шины также уменьшается в n раз, и каждый банк памяти (БП) секторного набора (рис. 6,б) может общаться с секторными шинами быстрее. При широкополосной передаче сигналов по кольцевой шине емкостная нагрузка всей кольцевой шины также существенно уменьшена, поскольку каждый секторный разделитель при сквозном прохождении сигналов через него усиливает сигнал, возобновляя форму его импульса. Кроме того, возможность блокировки с помощью секторных разделителей каждого сектора кольцевой шины от влияния других секторов, позволяет реализовать параллельные вычисления в секторных наборах заблокированных секторов, что также существенно увеличивает производительность системы в целом и снимает ограничения по наращиванию параллельно работающих секторных наборов.

Блок служебных функций реализует предложенный модифицированный алгоритм обработки изображений, основанный на равномерном (с точки зрения загрузки процессоров) распределении распараллеленных фрагментов алгоритма по процессорам РІМ-системы [12] с аппаратной его поддержкой, структурная схема которого защищена патентом Украины [13] и содержит: блок памяти микропрограмм, блок распределения фрагментов приложения, блок управления, блок памяти таблиц распределения, а также логический блок и блок буферной памяти вместе с интерфейсом с рабочей РІМ-системой. Предложенная система уменьшает общее время распределения по процессорам алгоритма обработки изображения, а также расширяет сферу ее приложения для рабочих систем, выполненных на разных аппаратно-программных платформах.

Оценка эффективности применения реконфигурируемых графических ускорителей. Реконфигурируемые процессоры могут адаптироваться к изменению потребностей обработки графических изображений (например, задач обработки media-фильтров) для легко распараллеливаемых алгоритмов. Интегрирование памяти большой емкости на том же самом кристалле, что и логические схемы, обеспечивает возможность получения более высокой ширины полосы памяти, чем была бы практически достигнута на традиционной архитектуре. Например, работая с 256-разрядной шиной от DRAM-банка (сегмента), можно достичь пропускную способность для каждого сегмента памяти до 4.8GB/s., размещенного на чипе. С архитектурной точки зрения, это устройство является весьма простым, однако теоретически может обеспечить высокую производительность (табл. 1) [14].

Таблица 1. Параметры производительности СИМА- архитектур для различных задач

Тип архитектуры	Время обработки, мс		
	Гистограмм	Фильтра	Пороговой величины
Традиционная RISC-архитектура	19	67	17
Процессор-в-памяти	71	0,6	0,1
Векторная DRAM-память (333МГц)	28	50	0,9
Одиночный мультипроцессорный чип с 4-мя процессорами (333МГц)	8	25	–
Merged CFD без быстрого переноса (16K LE's)	37	6	6,5
Merged CFD с быстрым переносом (16K LE's)	12	1,2	1,0

Для СИМА-системы с четырьмя банками памяти, общая ширина полосы обеспечивает пропускную способность до 19.2GB/s.. Это – больше, чем на порядок по сравнению с традиционными системами. Для сравнения, система оперативной памяти AlphaServer 8200 обеспечивает максимальную пропускную способность только 1.2GB/s. Таким образом, наиболее очевидное преимущество СИМА-архитектур по сравнению с другими типами архитектур при обработке изображения - увеличенная ширина полосы пропускания памяти и более низкая задержка сигнала при обращении к памяти, размещенной на чипе. Например, если на DRAM- чипе можно передавать одновременно 128 бит каждые 6 наносекунд, то данные полного буфера изображения емкостью 1МБ можно передавать через память меньше, чем за 1/2000 с. Конечно, это требует достаточной скорости процессора, чтобы не отставать от этой огромной ширины полосы. Оптимальная архитектура СИМА должна обеспечить высокий уровень вычислительного параллелизма, чтобы эксплуатировать полную ширину полосы пропускания памяти [7].

Технология реализации PIM – систем, как и других систем класса СИМА, позволяет основную оперативную память выполнить в непосредственной близости к элементарным процессорам, обеспечивая тем самым более низкую латентность памяти и значительно более широкий её интерфейс, чем в обычных микропроцессорах. Так разработанная за рубежом PIM-система VIRAM-1 имеет пиковую производительность 1.6 GFlop/s для данных разрядностью 32 бита и является чипом малой мощности, потребляя всего 2 ватта энергии.

В таблице 2 показаны различия в параметрах между PIM-системами типа VIRAM, DIVA, потоковым процессором Image, а также классической системой типа Power3, которая представляет собою 64-битовую реализацию процессора PowerPC с поддержкой выполнения восьми команд в цикле.

Таблица 2 – Различия в параметрах VIRAM, Image, DIVA и архитектуры Power3 [7].

Параметры	VIRAM	Imagine memory	DIVA (1 PIM)	Power 3
Пропускная способность по каналу процессор-память, ГВ/с	6,4	2,7	1,77	1,6
Пиковая производительность, Гфлоп/с	1,6	20	1,3	1,5
Пиковая производительность, Флоп/слово	1	30	1	3,75
Тактовая частота, МГц	200	600	166	375
Область (размеры) чипа	15x18 мм (270 мм ²)	12x12мм (144мм ²)	9,8x9,8 мм (96мм ²)	270мм ²
Ширина поддерживаемых данных, бит	64/32/16	42/16/8	64/32/16/8	64
Количество транзисторов, млн. шт.	130	21	55	16
Потребляемая мощность, Вт	2	4	1,6	33

Как большинство обычных суперскалярных архитектур, Power3 многократно использует КЭШ, чтобы уменьшить задержку основной памяти.

Согласно табл.2, Imagine имеет на порядок более высокую пиковую производительность, в то время как VIRAM – в два раза больше полосу пропускания памяти и потребляет половину мощности. При этом у VIRAM и DIVA имеется достаточная полоса пропускания, чтобы реализовать одну операцию при доступе к памяти, в то время как Image, требует 30 операций

Выводы

Таким образом, несмотря на совершенствования в технологии VLSI, которая обеспечила более высокие вычислительные возможности и большие блоки памяти, классическая архитектура, не в состоянии достаточно эффективно использовать эти достижения для эффективных параллельных вычислений из-за узких мест канала “память – процессор”, так как графическая обработка требует интенсивной коммуникации памяти процессора для передач данных, текстурирования, и буферизации. Архитектура PIM с малым временем доступа к памяти может увеличить производительность системы.

Применение реконфигурируемых архитектур существенно увеличивает производительность в десятки и сотни раз по сравнению с программной реализацией обработки графики и высокопроизводительным процессором за счет настройки архитектуры и структуры системы на размеры графических изображений и их сложность, а также за счет широкой шины (полосы пропускания) памяти. С этой точки зрения архитектурно-структурная организация КС с модифицированной кольцевой шиной, а также реконфигурируемой PIM-системы имеют обоснованную перспективу использования и дальнейшего развития вследствие особенностей их архитектурно-структурной организации. При этом в настоящее время все новые решения по созданию систем такого типа защищены патентами Украины.

Список литературы

1. Романюк О.Н. Класифікація графічних відеоадаптерів./ Романюк О.Н., Довгалоук Р.Ю., Олійник С.В.//Наукові праці ДонНТУ випуск 14(188). Серія “Інформатика, кібернетика та обчислювальна техніка”. С. 2011. – 215.
2. Додонов А.Г. Сравнительный анализ многопроцессорных систем формирования графических изображений / Додонов А.Г., Мельников А.Н, Резник Я.В.//The Sixth International Conference “INTERNET –EDUCATION - SCIENCE” Vinnytsia, Ukraine, October 7 –11, 2008. С. 369 – 373
3. Almudena Lindoso . Hardware Architectures for Image Processing Acceleration/ Almudena Lindoso and Luis Entrena// Электронный ресурс. Режим доступа: <http://www.intechopen.com/books/image-processing/hardware-architectures-for-image-processing-acceleration>. – Дата доступа: 17.10.2013.
4. Яковлев Ю.С. Однокристалльные компьютерные системы высокой производительности. Особенности архитектурно- структурной организации и внутренних процессов: Монография / Ю.С. Яковлев. – Винница: ВНТУ, 2009. – 294 с.
5. Thinh M. Le. SIMD Processor Arrays for Image and Video Processing: A Review/ Thinh M. Le, Snelgrove W. M., and Panchanathan S.//Электронный ресурс. Режим доступа: <http://www.thinhmle.com/imagesnus/SIMDProcessor ArraysforImageandVideoProcessingAREview.pdf>. – Дата доступа: 17.10.2013.
6. Tashev T. Design of Advanced Computer Architectures, Based on PIM - Processors in Memory/ Tashev T., Tashev S., Tasheva N.// Information technologies and control. – 2010. – №3. – С. 19 – 22
7. Olikier Leonid. Evaluation of Architectural Paradigms for Addressing the Processor-Memory Gap/ Olikier Leonid, Husbands Parry, Jacqueline Gorden Griem// Электронный ресурс. Режим доступа: <http://www.escholarship.org/uc/item/5s59f02m#page-3>. htm. – Время доступа: 18.10.2013
8. Jae Chul Cha. A PIM (Processor-In-Memory) for Computer Graphics: Data Partitioning and Placement Schemes/ Jae Chul Cha and Sandeep K. Gupta //World Academy of Science, Engineering and Technology 14 - 2008. С. –123 – 131. – Электронный ресурс. Режим доступа: <http://www.waset.org/journals/waset/v14/v14-22.pdf>.– Время доступа: 13.10.2013
9. Палагин А.В. Особенности подхода к выбору ПЛИС для проектирования PIM- систем / Палагин А.В. , Яковлев Ю.С. , Елисеєва Е.В. // Математичні машини і системи. – 2012. – № 3. – С. 19–28.
10. Патент № 6259 Украина, МПК G06F13/00, G06F12/00. Система пам’яті з інтеграцією функцій зберегання та обробки інформації на одному кристалі /Сергієнко І. В., Кривонос Ю. Г., Палагін О. В., Коваль В. М., Яковлев Ю. С., Тихонов Б. М.; Інститут кібернетики Імені В.М.Глушкова НАН України; 15.04.2005. Бюл. № 4. – 24с.
11. Пат. на винахід 99164 Україна, МПК G06F 15/16, G06F 13/42. Інтелектуальна розподілена система пам’яті з кільцевою шиною/ Палагін О.В., Яковлев Ю. С., Тихонов Б. М., Єлісеєва О. В.; Інститут кібернетики ім. В.М. Глушкова НАН України; заявл. 16.07.2010; опубл. 25.07.2012, Бюл. № 14.– 21с.

12. Пат. 71719 Україна, МПК G06F 9/44, G06F 9/45. Спосіб розподілу програми користувача для комп'ютерної системи / Сергієнко І. В. Палагін О. В., Боюн В. П., Яковлев Ю. С., Єлісеєва О. В.; Інститут кібернетики ім. В.М. Глушкова НАН України; заявл. 03.01.2012; опубл. 25.07.2012, Бюл. № 14.– 15с.

13 Пат. 73424 Україна, МПК G06F 9/44 (2006.01), G06F 9/45(2006.01). Система для розподілу програми користувача / Сергієнко І. В. Палагін О. В., Боюн В. П., Яковлев Ю. С., Єлісеєва О. В.; Інститут кібернетики ім. В.М. Глушкова НАН України; заявл. 27.02.2012; опубл. 25.09.2012, Бюл. № 18.

14. Landis David. Evaluation of Computing in Memory Architectures for Digital Image Processing Applications / Landis David, Hulina Paul and Deno Scott, Roth Luke and Coraog Lee// Електронний ресурс. Режим доступа: http://design.ecs.psu.edu/SmartDIMM/papers/ICCD_CIMA.pdf. – Дата доступа: 16.10.2013

Сведения об авторе

Яковлев Юрий Сергеевич – докт. техн. наук, зав. отделом, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины.

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ОБЧИСЛЮВАЛЬНІ МЕТОДИ

УДК 004.023

Ю. М. ЗОРІН, С. В. ПОДОЛЬСКИЙ

Національний технічний університет України "Київський політехнічний інститут", м. Київ

РОЗВ'ЯЗАННЯ КВАДРАТИЧНОЇ ЗАДАЧІ ПРО ПРИЗНАЧЕННЯ МЕТОДОМ ЛОКАЛЬНИХ ОПТИМУМІВ

Анотація. У роботі запропоновано евристичний алгоритм розв'язання квадратичної задачі про призначення. На відміну від існуючих евристичних алгоритмів, які потребують визначення декількох вхідних параметрів, запропонований алгоритм використовує лише один параметр і базується на ідеї використання комбінації локальних оптимумів в околі перестановок двох елементів. При цьому він дозволяє отримати більш якісні розв'язки за дещо менший час в порівнянні з відомими алгоритмами.

Ключові слова: квадратична задача про призначення, евристичний алгоритм, комбінаторна оптимізація, локальний мінімум.

Аннотация. В работе предложен эвристический алгоритм решения квадратичной задачи о назначениях. В отличие от известных эвристических алгоритмов, требующих задания нескольких входных параметров, предлагаемый алгоритм использует только один параметр, и базируется на идее использования комбинации различных локальных оптимумов в окрестности перестановок двух элементов. При этом он позволяет получить более качественные решения за несколько меньшее время по сравнению с известными алгоритмами.

Ключевые слова: квадратичная задача о назначениях, эвристический алгоритм, комбинаторная оптимизация, локальный минимум.

Abstract. The paper presents a new heuristic algorithm for the quadratic assignment problem. Unlike existing heuristic algorithms that require a few input parameters setting, the proposed algorithm uses only one parameter and is based on the idea of exploiting a combination of different local optima in the neighborhood of the two elements permutations. At the same time it allows to obtain solutions of better quality in slightly lesser time compared to the known algorithms.

Keywords: quadratic assignment problem, heuristic algorithm, combinatorial optimization, local minimum.

Вступ

Квадратична задача про призначення [1] по праву вважається однією з найскладніших задач комбінаторної оптимізації. Проблема формулюється таким чином. Є N об'єктів і N їх місць призначення, задані відстані між місцями призначення та інтенсивності потоків ресурсів між об'єктами. Задача полягає в розміщенні об'єктів у місцях призначення таким чином, щоб мінімізувати суму добутків відстаней на інтенсивність відповідних потоків. Інакше кажучи, задана цільова функція

$$\min \sum_{i=1}^N \sum_{j=1}^N d_{ij} f_{\pi(i)\pi(j)},$$

де $\pi(x)$ - номер об'єкта, призначеного місцю з номером x , d_{ij} - відстань між місцями i і j ,

$f_{\pi(i)\pi(j)}$ - інтенсивність потоків між об'єктами $\pi(i)$ і $\pi(j)$. Оскільки задача відноситься до класу

NP - hard, не існує "точних" алгоритмів, що дозволяють розв'язати її за поліноміальний час. До речі, задача про комівояжера може розглядатися як спеціальний випадок квадратичної задачі про призначення, в якому всі об'єкти об'єднані за допомогою інтенсивностей потоків з константним значенням в кільце, а всі інші потоки дорівнюють нулю. Таким чином, квадратична задача про призначення є більш складною ніж задача про комівояжера.

Окрім значного теоретичного інтересу, який викликає квадратична задача про призначення, існує широкий спектр практичних її застосувань, включаючи такі області як планування, паралельні та розподілені обчислення, статистичний аналіз даних, балансування запуску турбін і виробництво комп'ютерів [2]. Серед найбільш відомих евристичних алгоритмів розв'язання квадратичної задачі про призначення можна згадати Fast Ant System [3] і Simulated Annealing [4].

Мета дослідження

Метою роботи є розробка евристичного алгоритму розв'язання квадратичної задачі про призначення, який не поступається відомим в швидкодії й якості розв'язків, але не потребує налаштування численних параметрів, яке зазвичай здійснюється експериментальним шляхом і вимагає значних зусиль.

Алгоритм розв'язання квадратичної задачі про призначення.

В основі роботи алгоритму лежить використання локальних оптимумів, що не повторюються. Пошук локального оптимуму ґрунтується на тому, що за поліноміальний час можна перевірити доцільність

всіх $N(N-3)/2$ можливих перестановок пар елементів в деякому векторі, що є кандидатом на розв'язок задачі розміром N [5].

Кожна перестановка пари елементів i і j розв'язку π призводить до зміни значення цільової функції на деяке значення $\Delta(\pi, i, j)$, яке може бути обчислено за формулою

$$\begin{aligned} \Delta(\pi, i, j) = & (d_{ii} - d_{jj})(w(\pi_j, \pi_j) - w(\pi_i, \pi_i)) + (d_{ij} - d_{ji})(w(\pi_j, \pi_j) - w(\pi_i, \pi_i)) + \\ & + \sum_{k \neq i, j} ((d_{ki} - d_{kj})(w(\pi_k, \pi_j) - w(\pi_k, \pi_i)) + (d_{ik} - d_{jk})(w(\pi_j, \pi_k) - w(\pi_i, \pi_k))) \end{aligned}$$

Обчислювальна складність такої операції становить $O(n)$.

Здається, що при перевірці, або здійсненні, деякої перестановки розглядати її повторно недоцільно. Однак в подальшому, ця перестановка може виявитися корисною, якщо після її застосування хоча б одна інша перестановка виявиться ефективною і буде виконана. При цьому можна зробити евристичне припущення, що перестановки, які були здійснені раніше, є оптимізуючими з більшою ймовірністю, ніж ті, що були здійснені нещодавно. Це припущення базується на тому факті, що чим більше перестановок було здійснено після деякої конкретної перестановки, тим більше добутків відстаней на відповідні потоки пари об'єктів цієї перестановки могло потенційно зменшитись за абсолютним значенням. Виходячи з цього розроблено такий алгоритм знаходження локального оптимуму.

1. Створити список допустимих перестановок та заповнити його випадковою послідовністю всіх можливих перестановок пар індексів масиву розміру N .
2. Створити порожній список заборонених перестановок.
3. Якщо список допустимих перестановок порожній, закінчити пошук.
4. З ймовірністю, пропорційною номеру перестановки, нумеруючи з кінця списку допустимих, вибрати з нього випадкову перестановку пар індексів.
5. Якщо вибрана перестановка є оптимізуючою, виконати її й додати список заборонених перестановок в кінець списку допустимих. Список заборонених перестановок після цього встановлюється порожнім.
6. Додати вибрану на 4-му кроці перестановку в кінець списку заборонених.

У пункті 4 при ймовірнісному виборі індексу наступної перестановки замість використання колеса рулетки, алгоритмічна складність якого становить $O(N)$, пропонується вибір індексу перестановки за допомогою відображення $R \rightarrow i$, де R – генератор псевдовипадкових чисел з рівномірним розподілом, i – індекс наступної перестановки у списку перестановок, яка буде використана. Зазначене відображення виконується таким чином. Нехай кількість доступних перестановок дорівнює S . Необхідно обрати перестановку з ймовірністю, пропорційною індексу i цієї перестановки в списку. Розглянемо суми виду

$$\begin{aligned} 1 + 2 + 3 + \dots + i &= \frac{i(i+1)}{2} \\ 1 + 2 + 3 + \dots + i + (i+1) &= \frac{i(i+1)(i+2)}{2} \end{aligned}$$

Тоді при попаданні величини R в діапазон від $\frac{i(i+1)}{2}$ до $\frac{i(i+1)(i+2)}{2} - 1$ включно можна знайти індекс i , що задовольняє зазначеним вище умовам. Для натурального R , розв'язавши квадратне рівняння

$$R = \frac{i(i+1)}{2}, R \in \left[1, \frac{S(S+1)}{2}\right],$$

отримаємо індекс

$$i = \frac{-1 + \sqrt{1 + 8R}}{2}$$

Складність такої операції становить $O(1)$. Експериментально встановлено [5], що запропонована техніка вибору індексу перестановки дає збільшення швидкодії приблизно в 1,5 рази у порівнянні з підходом, коли всі перестановки обираються з однаковою ймовірністю.

В запропонованому алгоритмі введені поняття «інтенсивність» та матриця інтенсивностей M , елементи m_{ij} якої відповідають бажаності призначення j -го об'єкта i -му місцеположенню. Використовуючи запропоновану матрицю інтенсивностей, на кожній ітерації алгоритму виконується конструювання розв'язків, які потім і приводяться до локальних оптимумів за зазначеним вище алгоритмом. У випадку, якщо внаслідок конструювання отримуємо локальний оптимум, очевидним є факт передчасної збіжності алгоритму, що потребує виконання диверсифікації.

Конструювання нового розв'язку виконується за таким алгоритмом.

1. Із списку місцеположень, яким ще не призначені об'єкти, обрати випадковим чином з рівномірною ймовірністю довільне місцеположення i .
2. За допомогою колеса рулетки серед ще непризначених об'єктів вибрати для даного місцеположення об'єкт j з ймовірністю, що є пропорційною відповідному значенню елемента m_{ij} з матриці інтенсивностей M .
3. Якщо залишилися непризначені об'єкти (або місцеположення), перейти до кроку 1, інакше – завершити конструювання розв'язку.

Якщо зробити припущення, що кількість можливих локальних оптимумів як мінімум на порядок менша, ніж кількість всіх можливих розв'язків $N!$, то для запобігання передчасній збіжності алгоритму слід уникати повторного впливу одного й того ж локального оптимуму на матрицю інтенсивностей. Тобто доцільною є результуюча матриця, елементи якої формуються як комбінація або перетин різних локальних оптимумів. У зв'язку з цим необхідно вести список локальних оптимумів і виконувати в ньому пошук при спробі додавання нового. З метою підвищення швидкодії доцільним є зберігання не самих векторів, а лише їх хеш-кодів.

Остаточню узагальнений алгоритм передбачає таку послідовність дій.

1. Ініціалізувати матрицю інтенсивностей довільним додатним числом та створити порожній список хеш-кодів потенційних розв'язків задачі, а також ініціалізувати значення сумарної вартості розв'язків $C_H = 0$.
2. Сконструювати новий розв'язок задачі за допомогою колеса рулетки.
3. Якщо хеш-код сконструйованого розв'язку міститься в списку хеш-кодів, виконати диверсифікацію, скинувши в початкові відповідні значення матриці інтенсивностей для даного розв'язку, і перейти до кроку 2.
4. Оптимізувати сконструйований розв'язок до локального оптимуму.
5. Якщо хеш-код отриманого на 4-му кроці локального оптимуму вже міститься в списку хеш-кодів, перейти до кроку 2.
6. Додати хеш-код отриманого локального оптимуму до списку хеш-кодів та додати до сумарної вартості розв'язків значення цільової функції поточного розв'язку

$$C_H = C_H + C.$$

7. Помножити відповідні значення елементів m_{ij} матриці інтенсивностей M для призначень $i \rightarrow j$ отриманого оптимуму на значення коефіцієнту

$$k = \left(\frac{C_H}{C \times H} \right)^\alpha,$$

де C_H – сумарна вартість розв'язків, C – значення цільової функції поточного розв'язку, H – довжина списку хеш-кодів, α – коефіцієнт інтенсифікації.

8. Якщо значення цільової функції даного локального оптимуму є меншою за найкращу на даний момент, запам'ятати поточний розв'язок.
9. Якщо виконується умова критерію зупинки, завершити алгоритм і повернути відомий оптимальний розв'язок, інакше – перейти до кроку 2.

В запропонованому алгоритмі значення інтенсивностей m_{ij} матриці M , що відповідають бажаності призначень, фактично змінюються в залежності від того, на скільки кращим або гіршим від середньостатистичного є отриманий розв'язок. Вартість середньостатистичного розв'язку отримується як середня вартість всіх відомих розв'язків C_H / H . Тоді значення $C_H / (C \times H)$ фактично означає, у скільки разів значення поточного розв'язку менше, ніж значення середньостатистичного розв'язку. Оскільки зазвичай після деякої кількості ітерацій вартість кожного наступного розв'язку стає досить близькою до вартості оптимального, то дане відношення буде близьким до одиниці. Для того, щоб підвищити «контрастність» між вартістю поточного розв'язку та середньостатистичного, дане відношення підноситься до степеня $\alpha > 1$. Таким чином, ми маємо автоматичну корекцію бажаності призначень розв'язків як кращих, так і гірших, ніж середньостатистичні.

З викладеного випливає, що алгоритм має лише один вхідний параметр - коефіцієнт інтенсифікації α . За замовчанням можна використовувати значення $\alpha = 1$. Очевидно, що початкові значення елементів матриці інтенсивностей при такому підході не мають значення, а лише мають бути однаковими.

Результати тестування

Показники швидкодії й якості отриманого розв'язку розробленого алгоритму наведені в порівнянні з результатами авторських реалізацій алгоритмів Fast Ant System (FANT) і Simulated Annealing (SA). В якості вхідних даних бралися матриці розмірністю 20, заповнені випадковими числами у діапазоні від 1 до 99 включно. Кількість ітерацій для кожного з алгоритмів була підібрана таким чином, щоб забезпечити приблизно однаковий середній час виконання всіх алгоритмів. У табл. 1 наведені усереднені за 1000 запусків результати роботи алгоритмів.

Таблиця 1 – Порівняння результатів роботи FANT, SA та запропонованого алгоритму

	Час виконання	Кількість ітерацій	Значення цільової функції
SA	585мс	1350	864112
FANT	581 мс	2071	863995
Запропонований алгоритм	568 мс	1000	863226

В результаті проведених експериментів для згенерованих вхідних наборів було визначено оптимальне значення коефіцієнта інтенсифікації $\alpha = 2$.

Висновки

Результати числових експериментів доводять, що запропонований алгоритм у порівнянні з алгоритмами FANT та SA дозволяє отримати більш якісні розв'язки за дещо менший час при довільних вхідних даних. Перспективним є впровадження схеми, за якою всі значення приросту вартості розв'язку внаслідок парних перестановок обчислюються заздалегідь перед знаходженням локального оптимуму, а потім коректуються після кожної перестановки пари елементів [6].

Запропонований алгоритм потребує визначення лише одного вхідного параметру α , який може бути заданий апіорі.

Список літератури

1. Assignment problems and the location of economic activities [Text] / T. C. Koopmans, M. J. Beckman // *Econometrica*. — 1957. — № 25. — P. 53-76.
2. The quadratic assignment problem [Text] / E. L. Lawler // *Management Science*. — 1963. — № 9. — P. 586-599.
3. FANT : Fast Ant System [Technical Report] / E. D. Taillard // *Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale*. — 1998.
4. An improved annealing scheme for the QAP / D. T. Connolly // *European Journal of Operational Research*. — 1990. — № 46. — P. 93-100.
5. Зорін Ю.М., Подольський С.В. Алгоритм перетину локальних оптимумів для оптимізації задачі про квадратичні призначення // III наук.-техн. конф. «Прикладна математика та комп'ютинг». Тези доповідей. — К.: НТУУ «КПІ», 13-15 квітня 2011. — С. 353 – 356.

6. Зорін Ю.М., Подольський С.В. Табу-пошук для квадратичної задачі про призначення // IV наук.-техн. конф. «Прикладна математика та комп'ютинг». Тези доповідей. — К.: НТУУ «КПІ», 12 квітня 2012. — С. 273 – 276.

Відомості про авторів

Зорін Юрій Михайлович – доцент каф. СПіСКС, факультет прикладної математики, Національний технічний університет України «КПІ».

Подольський Сергій Валентинович – EPAM Systems, Software Engineer.

УДК 004.725

М. В. НАЗРУК, В. В. ПАСІЧНИК

Національний університет «Львівська політехніка», м. Львів

МОДЕЛЮВАННЯ МІСЬКОГО ОСВІТНЬОГО СЕРЕДОВИЩА ЯК ПРОФІЛЬНОЇ СОЦІАЛЬНОЇ МЕРЕЖІ

Анотація. Проаналізовано процеси моделювання соціальних мереж в контексті розвитку сфери освіти крупного міста. Запропонована формальна модель соціальної освітньої мережі міста (СОММ), як проблемно-орієнтованої складної мережі крупного соціополісу. З використанням апарату теорії складних мереж подано множину основних параметрів та інформаційних характеристик відповідної моделі СОММ.

Ключові слова: соціальна мережа, мережевий граф, агенти.

Аннотация. Проанализированы процессы моделирования социальных сетей в контексте развития сферы образования крупного города. Предложенная формальная модель социальной образовательной сети города (СОММ), как проблемно-ориентированной сложной сети крупного социополиса. С использованием аппарата теории сложных сетей подано множество основных параметров и информационных характеристик соответствующей модели СОММ.

Ключевые слова: социальная сеть, сетевой граф, агенты.

Abstract. The processes of modeling of social networks in the context of education large cities were analyzed. A formal model of social education network of the city (SENC) as a large sociopolis was offered. Using the theory of complex networks the set of basic parameters and informational characteristics of a corresponding model (SENC) was given.

Keywords: social network, network graph, agents.

Постановка проблеми у загальному вигляді

Сучасний етап розвитку світової цивілізації диктує та обумовлює процеси формування інформаційного суспільства, що передбачає широке проникнення інформаційно-технологічних механізмів та інформаційних систем у повсякденне життя соціуму, зумовлюючи активне використання інформаційно-комунікаційних технологій (ІКТ) в усіх сферах соціального буття. Закономірним наслідком, означених вище явищ, є інформатизація освітніх процесів, яка передбачає використання комп'ютерних та телекомунікаційних технологічних розроблень, інфокомунікаційних соціальних сервісів; пришвидшення процесу здобуття нових і закріплення раніше набутих знань, процесів обміну інформацією; просторове наближення та соціопсихологічна адаптація інформативних та пізнавальних освітніх матеріалів до кінцевого споживача, який використовує їх для власного самонавчання та професійного становлення в процесі неперервного навчання впродовж життя.

Сучасні соціальні комп'ютерні мережі є одним з найбільш універсальних інструментів комунікування і популярним інформаційно-технологічним сервісом, який сьогодні утримує активну увагу значної частини масової Інтернет-аудиторії. В сфері освіти цей інформаційний феномен сприяє формуванню та розвитку технологій електронного навчання, пропонуючи нові інноваційні, організаційні та методичні рішення. Водночас використання в реальних освітніх процесах популярних нині соціальних мереж, таких як «ВКонтакте», Facebook, Twitter та ін. не є масовим через відсутність інструментарію, профільно розробленого та адаптованого до навчальних цілей, обмеженість швидкостей доступу в мережах до Інтернет ресурсів з навчальних аудиторій, перенасиченість цієї категорії мереж розважальним контентом.

Це в свою чергу обумовлює необхідність розроблення моделі соціальної освітньої мережі міста (СОММ), що уможливить проведення комплексного системного дослідження соціально-освітніх зв'язків, які притаманні таким освітнім структурам, як: дитячі садки, школи, технікуми, коледжі, університети – з однієї сторони та бізнесом і суспільною громадою міста – з іншої, з метою реалізації виваженої спільної освітньої, наукової та культурної діяльності, встановлення найхарактерніших закономірностей поширення потоків інформації в соціумі та подання їх опису у термінах теорії складних мереж.

Аналіз сучасних досліджень та публікацій

Використання феномену соціальних комп'ютерних мереж, як одного з варіантів реалізації освітнього соціокомунікаційного середовища складної системи взаємодії учасників навчального процесу, цікавить сьогодні багатьох дослідників. Зокрема у [1, 2] аналізується досвід використання соціальних мереж в освітніх цілях, розглядаються перспективи розвитку технологій та систем електронного навчання. Пропонується широкий спектр різноманітних математичних моделей поширення інформації між учасниками соціальних мереж: моделі з порогами [3], моделі незалежних каскадів [4], моделі на основі ланцюгів Маркова, моделі Ізінга [3] і т. ін.

Відомо, що соціальна мережа може бути подана як граф з скінченною множиною вершин (агентів моделі, з'єднаних ребрами, які відображають взаємодію агентів) [5, 6].

З використанням графових формалізмів можливе моделювання і таких нетривіальних та дещо, на перший погляд, неklasичних, а саме квантових і квантоподібних (КП) властивостей соціальних мереж.

Причому КП властивості соцмереж мають два профілюючих джерела – топологія власне мережі [7, 8] та КП властивості психіки та поведінкової психології людей, що взаємодіють в середовищі цієї мережі.

В роботі [9] розглянуто клас оптимізаційних задач управління соціальними мережами та подані методи їх вирішення. Запропоновані відповідні математичні моделі, означені їх особливості та сфери їх практичного застосування.

В деяких роботах виділяються класи моделей інформаційного впливу, управління і протиборства [7], однак, ряд властивостей соціальних мереж ще потребує дослідження та розроблення адекватних підходів та формалізмів їх коректного моделювання [8].

Мета даної статті – проаналізувати множину загальних параметрів моделі соціальної освітньої мережі міста (СОММ) з використанням апарату теорії складних мереж.

Основний матеріал

У зв'язку з поширеним різночитанням та трактуванням поняття освітнє середовище подамо сутнісну структуру цього терміну, використовуючи термінологічні статті популярного інформаційного ресурсу, яким є Wikipedia – (<http://www.wikipedia.org/>).

Освітнє середовище – сукупність об'єктивних зовнішніх умов, факторів, соціальних об'єктів, необхідних для успішного функціонування галузі освіти. Це система впливів (педагогічних, виховних, освітніх, професійних, економічних, соціокультурних) і умов формування особистості, а також можливостей її розвитку, які містяться в соціальному і просторово-предметному оточенні.

Освітнє середовище міста ми визначаємо як багатогранне та поліфункціональне утворення, що існує й розвивається в реальному часі й територіальному просторі, підлягає коригуванню, управлінню, координації й через який здійснюється педагогічний вплив зовнішніх об'єктивних і суб'єктивних факторів на суб'єкти навчального процесу.

Міська освітня мережа – це об'єднання локальних комп'ютерних мереж та інформаційних ресурсів освітніх установ всіх видів і рівнів освіти міста в єдину інформаційно-комунікаційну проблемно-орієнтовану складну мережу, яка активно взаємодіє із зовнішніми глобальними мережами та їх ресурсами.

Соціальна мережа – соціальна структура, утворена індивідами або організаціями, яка відображає різноманітні зв'язки між ними через різнорівневі соціальні взаємовідносини.

Модель предметної області

Якісно під соціальною мережею розумітимемо соціальну структуру, яка складається з множини агентів (суб'єктів – індивідуальних чи колективних) і визначеної на ній множини відношень (сукупності зв'язків між агентами, наприклад дружба, співпраця, комунікація і т.п.).

Формально модель соціальної освітньої мережі міста подамо кортежем, який містить чотири елементи:

$$M = \langle A, V, R, F \rangle,$$

де $A = \{a_1, a_2, \dots, a_n\}$ – множина суб'єктів (агентів) соціальної освітньої мережі міста. В ролі агентів СОММ виступатимуть вихованці, учні, студенти та інші особи, які навчаються, а також науково-педагогічні працівники, батьки та представники підприємств, установ, кооперативних, громадських організацій крупного міста, які беруть участь у навчально-виховній роботі.

$V = \{v_1, v_2, \dots, v_m\}$ – множина властивостей агентів, тобто особисті дані, національні особливості, вік, рівень освіти, інтереси, соціальний статус та ін.

$R = \{r_1, r_2, \dots, r_k\}$ – множина відношень між агентами, наприклад, дружба, співпраця, комунікація, навчання та інші форми взаємодії, які притаманні суб'єктам освітньої мережі міста.

$F = \{f_1, f_2, \dots, f_v\}$ – множина дій, які допускаються з агентами через зміну їх властивостей і відношень між ними.

Характеристика параметрів моделі СОММ

Освітня мережа крупного міста має просторово-предметні умови, систему міжособистісних відношень між суб'єктами навчально-виховного процесу та простір різноманітних видів діяльності, необхідних для соціалізації осіб, що навчаються у відповідності з віковими та індивідуальними соціокультурними особливостями розвитку. В контексті крупного міста як складного соціополісу зазвичай вона складається із навчальних закладів, наукових, науково-методичних і методичних установ,

науково-виробничих підприємств, державних і місцевих органів управління освітою та самоврядування в галузі освіти, тобто має властивості ієрархічної структури.

Таким чином, соціальну освітню мережу міста (СОММ) представимо у вигляді мережевого графа $G = (A, R)$ (рис.1.), де A – не порожня скінченна множина вузлів (агентів), R – множина невпорядкованих пар різних елементів з A , які називатимемо ребрами (відношеннями між агентами) мережі.

Порядком мережі називатимемо загальну кількість її вузлів n , $n \in A$, розміром мережі – число її відношень m , $m \in R$.

Структурні властивості мережі виражатимемо за допомогою матриці суміжності – булевої матриці A , розмірності $n \times n$ з елементами a_{ij} ($i, j = 1, \dots, n$), де $a_{ij} = 1$, якщо вузли i та j з'єднанні між собою (дві вершини з'єднані, якщо відповідні люди знайомі, чи дружать між собою, навчаються в одному закладі, перебувають у шлюбі, пов'язані діловими контактами тощо), в протилежному випадку $a_{ij} = 0$ [10].

Кожен вузол характеризується ступенем, тобто кількістю зв'язків, які входять в нього, таким чином ступінь вузла i визначається:

$$d_i = \sum_j a_{ij}.$$

Для прикладу, ступінь вузла в СОММ показуватиме рівень комунікабельності особи, тобто кількість її соціальних контактів.

На рис. 1. множина вершин A графу G включає n підмножин – блоків (рівнів) $A_i \subset A$ ($i = 1, \dots, n$), які відповідають структурним освітянським підрозділам різних рівнів (дошкільна освіта, загальна середня, професійно-технічна, вища та ін.). В межах кожного блоку містяться вузли (агенти) a_{ij} .

Множина ребер R графу G являє собою набір пар вузлів не лише в межах одного рівня, а і має такі пари вузлів, як $r_{jl}^i = \{a_{ji}, a_{l_{i+1}}\}$. Ребро графу $r_{jl}^i \in R$, таким чином з'єднує j -тий вузол рівня A_i з l -тим вузлом рівня A_{i+1} .

Відзначимо, що у межах кожного блоку агенти об'єднані за спільними властивостями (наприклад, вік, рівень освіти, навчання в одному закладі), тобто утворюють, так звані, кліки (clique) V_{a_i} , $i = 1, \dots, n$ – підграфи чи класи, для яких зв'язки між вузлами всередині є міцніші та чисельніші, ніж між вузлами інших підграфів. Таким чином, наша соціальна освітня мережа матиме властивості асортативного змішування.

Тенденцію до утворення груп взаємозв'язаних вершин (кліків) характеризуватиме коефіцієнт кластерності – ймовірність того, що два найближчих сусіди деякого вузла самі є найближчими сусідами [10]. Іншими словами, якщо вузол j має q_j найближчих сусідів з числом t_j зв'язків між ними, то локальний коефіцієнт кластерності рівний:

$$C_j(q_j) = \frac{t_j}{q_j(q_j - 1)/2}.$$

На рис. 1. a_{k_i} ($i = 1, \dots, m$) – керуючі агенти, вузли з великою кількістю зв'язків (хаби), які визначатимуть поведінку мережі.

Шлях між вершинами визначатимемо кількістю кроків, які необхідно здійснити, щоб дістатися вздовж існуючих ребер від однієї вершини до іншої. Природно, вершини можуть бути з'єднані прямо або опосередковано. Для всієї мережі можна ввести поняття середнього шляху:

$$l = \frac{2}{n(n+1)} \sum_{i \geq j} S_{ij},$$

де n – кількість вузлів, S_{ij} – найкоротша відстань між вузлами i та j .

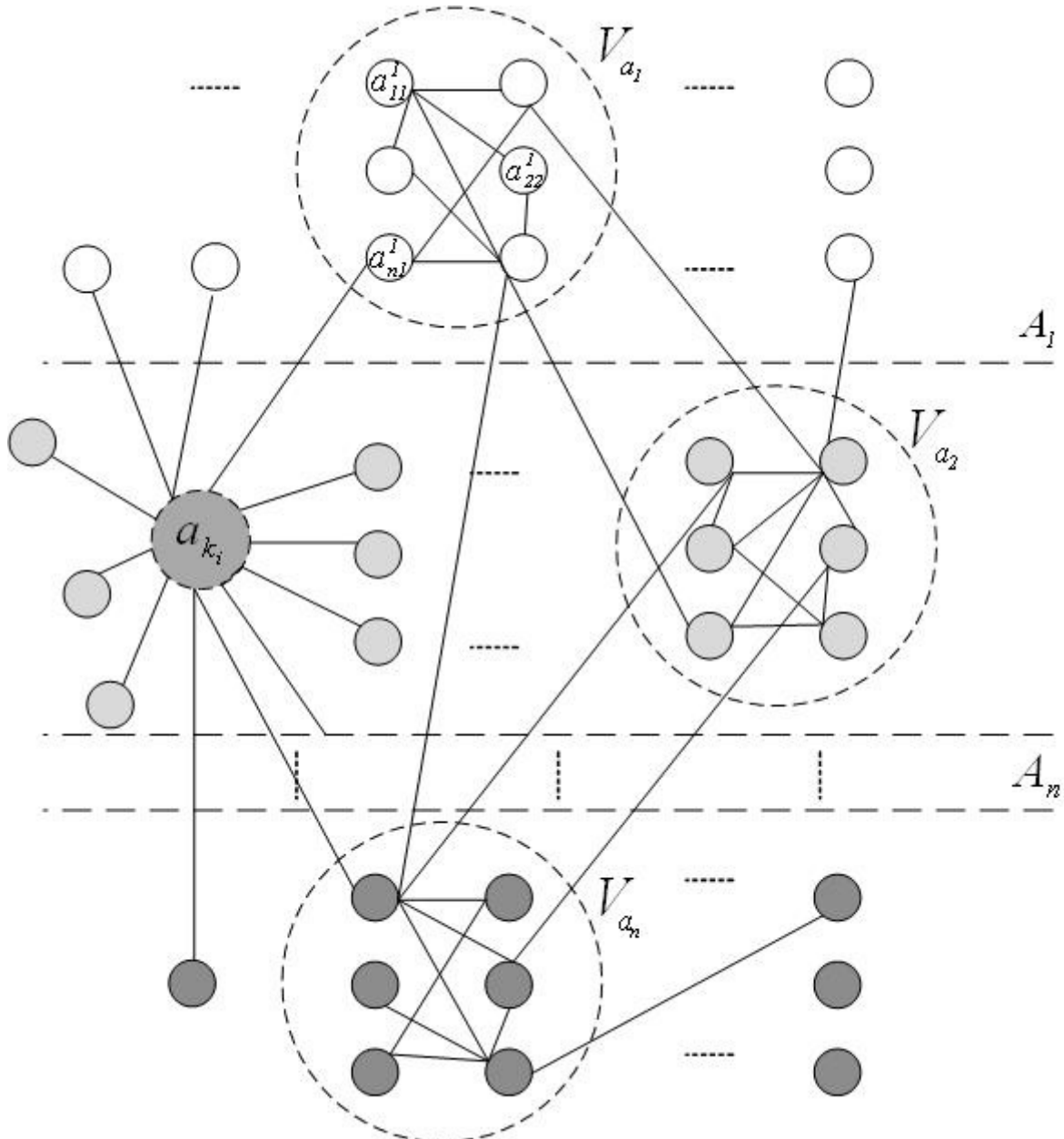


Рисунок 1 – Взаємодія агентів в моделі COMM

Кількість найкоротших шляхів, що проходять через вершину показуватиме такий параметр, як посередництво [10]. Вузли з найбільшим посередництвом відіграватимуть головну роль у встановленні зв'язків між іншими вузлами в мережі. Посередництво b_m вузла m визначається за формулою:

$$b_m = \sum_{i \neq j} \frac{B(i, m, j)}{B(i, j)},$$

тут $B(i, j)$ – загальна кількість найкоротших шляхів між вузлами i та j , $B(i, m, j)$ – кількість найкоротших шляхів між вузлами i та j , що проходять через вершину m .

Користувачі мережі тісно взаємодітимуть між собою, тому можна виділити ряд факторів, які будуть впливати на поведінку агентів, їх активність, вибір того чи іншого навчального контенту (рис.2.).

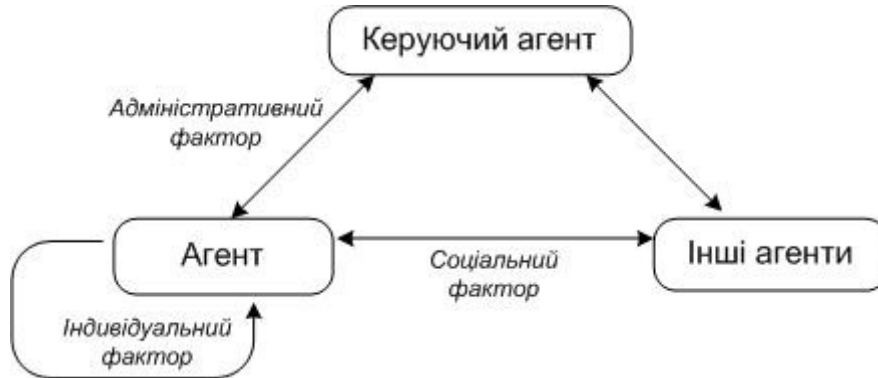


Рисунок 2 – Фактори, що впливатимуть на поведінку агентів в освітній соціальній мережі

Індивідуальний фактор – це власні переконання та інтереси агента без зовнішнього впливу; під соціальним фактором позначимо взаємозв'язок з іншими агентами освітньої мережі; адміністративний фактор – це вплив з боку керуючого агента a_k [3].

Оскільки агенти в мережі впливають один на одного, то степінь впливу задамо матрицею $B = \left\| b_{ji} \right\|$ розмірності $n \times n$, де $b_{ji} \geq 0$ – рівень довіри j -го агента i -му.

Відзначимо, що мережева взаємодія в моделі соціальної освітньої мережі міста може бути представлена, як поєднання чотирьох компонентів: простір (S), час (T), інформація (I), енергія (E):

$$N = S \cup T \cup I \cup E$$

В нашому випадку, простір представляє собою сукупність умов, а саме: інформаційно-комунікаційних, навчально-методичних, психологічних, соціальних та ін., що визначають рівень взаємодії осіб (агентів) в СОММ, багатоманітність їх внутрішніх і зовнішніх зв'язків.

Час – показник динаміки росту мережевої взаємодії користувачів СОММ.

Інформація (навчального, особистого, суспільного, професійного та ін. характеру), яка поширюється між особами в СОММ визначає зміст мережевої взаємодії.

Енергія представляє різні форми, способи існування і життєдіяльності соціальної освітньої мережі міста.

Висновки

Розроблення моделі соціальної освітньої мережі міста (СОММ) дозволяє формувати повні і адекватні описи та подання інтеграційних процесів в системі навчальних закладів різних видів і рівнів освітньої галузі міста та єдиного загального освітнього простору, в якому будуть реалізовані гармонійні організаційно-управлінські механізми діяльності муніципальної системи освіти, запроваджуватимуться технології сучасних методів і форм навчання, забезпечуватиметься функціонування розгорнутих експериментальних майданчиків з різних інноваційних напрямків у сфері освіти, буде організована ефективна взаємодія з соціальною сферою: закладами культури, охорони здоров'я, спорту, дозвілля, батьківськими спільнотами і т. п.

Авторами було проаналізовано специфіку побудови моделі освітньої мережі міста, описано множину загальних параметрів та інформаційних характеристик моделі СОММ з використанням апарату теорії складних систем. Запропоновано подавати мережеву взаємодію в моделі соціальної освітньої мережі міста як поєднання чотирьох базових компонентів, а саме: простору, часу, інформації та енергії.

Подальші дослідження спрямовані на детальне розроблення комплексної математичної моделі освітньої мережі міста та її формальний опис з метою імітації і коректного прогнозування поведінки такого роду мереж при зміні їх структурних властивостей. Дослідження, базуючись на фізичних

аналогіях, процесів росту мереж та мережевого поширення інформаційних потоків в соціумі, зокрема з використанням таких аналогій щодо явищ дифузії, масопереносу та ін.

Список використаної літератури

- 1) Фещенко А. В. Социальные сети в образовании: анализ опыта и перспективы развития / А. В. Фещенко // Сборник статей «Гуманитарная информатика». – 2011. – Вып. 6. – С. 124-134.
- 2) Клименко О. А. Социальные сети как средство обучения и взаимодействия участников образовательного процесса / О. А. Клименко // Теория и практика образования в современном мире: материалы междунар. заоч. науч. конф. февраль 2012 г., Санкт-Петербург. – СПб.: Реноме, 2012. – С. 405-407.
- 3) Губанов Д.А. Социальные сети: модели информационного влияния, управления и противоборства / Д. А. Губанов, Д. А. Новиков, А. Г. Чхартишвили. – М.: Физматлит, 2010. – 228 с.
- 4) Rolfe M. Social Networks and Threshold Models of Collective Behavior / M. Rolfe. – Chicago: University of Chicago, 2004.
- 5) Karimi F. Threshold Model of Cascades in Temporal Networks / F. Karimi, P.Holme – 2012. – ArXiv: 1207.1206v1.
- 6) Santoro N. et al. Time-Varying Graphs and Social Network Analysis: Temporal Indicators and Metrics / Santoro N.et al – 2011. – ArXiv: 1102.0629.
- 7) Perseguers S. Quantum Random Networks / S. Perseguers, M. Lewenstein, A. Acín, J. Cirac // Nature Physics, 2010 – DOI 10.1038/NPHYS1665.
- 8) Takaguchi T. et al. Importance of Individual Events in Temporal Networks / T. Takaguchi – 2012. – ArXiv: 1205.4808v1.
- 9) Зуев А. С. Модели управления мнениями агентов в социальных сетях / А. С. Зуев, Д.Н. Федянин // Проблемы управления. – 2011. – № 2. – С. 37-45.
- 10) Евин И.А. Введение с теорию сложных сетей / И.А. Евин // Компьютерные исследования и моделирование. – 2010. – Том 2. – С. 121-141.

Відомості про авторів

Назарук Марія Володимирівна – аспірант, Національний університет «Львівська політехніка».

Пасічник Володимир Володимирович – д.т.н., професор, Національний університет «Львівська політехніка».

БІОЛОГІЧНІ ТА МЕДИЧНІ ПРИЛАДИ І СИСТЕМИ

УДК 681.586.773

Й. І. СТЕНЦЕЛЬ, С. В. ПАВЛОВ, С. М. ЗЛЕПКО

Східноукраїнський національний університет імені Володимира Даля, м. Луганськ
Вінницький національний технічний університет, м. Вінниця

ПРИНЦИПИ ФІЗИЧНОГО ТА МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ НЕІНВАЗИВНИХ МЕТОДІВ ДІАГНОСТИКИ СТАНУ ЗДОРОВ'Я ЛЮДИНИ

Анотація. У роботі наведено результати аналізу теоретичних та експериментальних досліджень роботи біологічного організму, принципи діагностування стану людського організму. Показано, що діагностика стану здоров'я людини базується на основі прямого перенесення енергії, маси та кількості руху без врахування принципів їх створення, реологічних переходів, перетворень і стоку того чи іншого явища перенесення.

Ключові слова: біологічний організм, діагностика, стан, явище перенесення, реологічний перехід, стік, похибка, контроль, час, джерело, випромінювач.

Аннотация. В работе приведены результаты анализа теоретических и экспериментальных исследований работы биологического организма, принципы диагностирования состояния человеческого организма. Показано, что диагностика состояния здоровья человека базируется на основе прямого переноса энергии, массы и количества движения без учета принципов их создания, реологических переходов и стока того или иного явления переноса.

Ключевые слова: биологический организм, диагностика, состояние, явление переноса, реологический переход, с ток, погрешность, контроль, время, источник, излучатель.

Abstract. The paper presents the results of the theoretical and experimental studies of the biological organism, the principles of diagnostics of the human body. It is shown that the diagnosis of human health is based on the direct transfer of energy, mass and momentum excluding principles of their creation, rheological transitions, transformations and flow of a transfer phenomena.

Keywords: biological organism, diagnosis, condition, transport phenomena, flow switch, flow, accuracy, control, time, source, radiator.

Вступ

Вивчення процесів перенесення енергії, маси та кількості руху привело до подальшого, більш чіткого розуміння багатьох фундаментальних аспектів явищ, котрі протікають в біологічній структурі і зокрема в організмі людини. За минулий час вченими отримано детальне пояснення багатьох механізмів руху матеріальних, теплових та енергетичних потоків і їх перенесення в організмі людини як у необмеженому, спокійному її стані, так і в замкнених об'ємах тіла. Велика увага приділена також найбільш складним видам перенесення, який об'єднує властивості як зовнішніх, так і внутрішніх процесів.

Явища перенесення в людському організмі, котрі виникають у звичайних умовах під дією впливових сил, надзвичайно різноманітні. Вони існують в об'ємах одно- та багатозначних біологічних структурах. Матеріальний чи тепловий об'єм такої структури будь-якого масштабу може піддаватися дії впливової сили, яка виникає однократно чи багатократно від багатьох різноманітних видів і суміщення фізико-хімічних процесів. Впливова сила виникає через різницю тиску, депресію температур чи маси, механічних напружень електричної, магнітної та інших сил.

За останні роки в біомедичній кінетиці появився новий науковий напрямок, суть якого полягає в комплексному вивченні біохімічних процесів у сукупності з фізичними процесами перенесення кількості теплоти, маси та енергії. Процеси, котрі рахувалися в класичній біохімічній кінетиці збуреннями та викривляли хід реакції біоструктури на механічні, теплові та енергетичні подразники, набули зацікавленість якраз в комбінації з біохімічними процесами. Таке суміщення біохімічної кінетики з теорією дифузії [1-6], теплопередачі [7-11], гідродинамікою [12-16] дозволило отримати низку теоретично цінних результатів, розробити нові методи вивчення стану здоров'я людини та підвести науковий фундамент під теорією таких важливих біомедичних процесів, як створення пухлин, розчинення біомедичних препаратів, основні процеси біохімічних технологій при діагностуванні та лікуванні тощо.

Класична біохімічна кінетика [1, 3, 5] вивчає протікання біохімічної реакції в ідеалізованих умовах: за сталою (як в часі, так і в просторі) температурою тіла та сталих у біоструктурі концентраціях речовин. Задачею макроскопічної біохімічної кінетики є вивчення біохімічних перетворень у реальних умовах її макроскопічного протікання в організмі людини, тобто з врахуванням сторонніх фізико-хімічних процесів, котрі накладаються на основний. Найважливішими з фізичних процесів є дифузія початкових речовин, котрі уводяться в організм людини, і продуктів біохімічних реакцій, а також виділення та розповсюдження їх результату за рахунок перенесення теплових потоків, енергетичних полів та кількості руху, наприклад, потоком крові, лімфориди, води тощо. На такі процеси сильно впливає характер руху таких потоків і полів, котрі приводять до конвекційного та кондуктивного перенесення кількості маси, енергії (наприклад, теплової) та руху [10-13].

Вивчення макроскопічної біокінетики дозволяє розробляти методи діагностики істинної біохімічної кінетики та механізму біохімічних процесів, безпосереднє дослідження яких ускладнюється факторами перенесення маси, енергії та кількості руху [1, 5, 6,10,15,16]. Конвекційне перенесення речовини в організмі людини пов'язане з турбулентним рухом крові та лімфатичних рідин у відповідних біологічних каналах та рідинно-твердих речовин у шлунково-кишковому тракті. Таким чином, макроскопічна біохімічна кінетика являє собою результат синтезу двох наукових напрямів: біохімічної кінетики, з однієї сторони, і теорії процесів дифузії речовин перетворення (спалювання) і теплопередачі біохімічної енергії — з іншої.

Мета статті

Виконати аналіз біологічного організму як об'єкта вимірювального контролю та діагностики та розробити принципи фізичного та математичного моделювання неінвазивних методів діагностики стану здоров'я людини на основі теорії реологічних переходів та перетворень.

Постановка задачі

Майже всі явища перенесення маси, енергії та кількості руху мають загальні риси але відрізняються від процесів перенесення, які обумовлені звичайними видами прикладання сили, типу вимушеного руху [12-18]. Характерною відмінністю є те, що заздалегідь дуже мало відомо про результуючий напрямок руху в організмі людини, який виникає під дією тої чи іншої сили. Поля напрямку руху маси, енергії чи кількості руху завжди тісно пов'язані поміж собою тому їх необхідно розглядати сумісно, Сам напрямок руху може бути порівняно незначним. Це значить, що швидкості перенесення в цьому напрямку достатньо малі, а інерційні течії прилягають до поверхні, котрі рівномірно розподіляють потік маси, енергії чи кількості руху. Існує багато різних сил, як різниця густин при перенесенні кількості маси речовини, різниця температур при перенесенні кількості тепла та маси, різниця потенціалів при перенесенні кількості електромагнітної чи іншої енергії, різниця зусиль при перенесенні механічної енергії, котрі викликають рух відповідних потоків [2-7]. Рух матеріального чи енергетичного потоку є фіксованим у просторі через наявність стійкої поверхні розділу фаз, наприклад поверхні розділу між біологічною структурою та рідиною. Як відомо [23, 24], у хімічній технології всі тепло-масообмінні та хімічні процеси супроводжуються відповідними реологічними переходами та перетвореннями, які в значній мірі суттєво впливають на їх результат. Так як у біологічному організмі мають місце аналогічні процеси, то актуальною є задача вивчити біологічні процеси на основі фізичних переходів та математичних перетворень, враховуючи, що кожний фізичний перехід має зону реологічного переходу, а математичне перетворення описує таку зону інтегральною імпульсною дельта-функцією Дірака.

Структура біохімічних переходів та перетворень у біологічному організмі

Так як в основі роботи кожного біологічного організму лежить біохімічне перетворення, то його біологічний процес у першому наближенні можна представити схемою, яка складається з основного біохімічного реактора, первинного ректифікаційного апарату, вторинного ректифікаційного апарату, сепараторів, холодильників, фільтрів і трубопроводів (рис. 1). У біологічному організмі основний біохімічний реактор поєднується з первинним ректифікаційним апаратом і називається шлунком. Вхідним матеріальним потоком такого реакторно-ректифікаційного апарату (РРА) є роздіблена та змочена харчова речовина. Під дією соляної кислоти й різноманітних добавок (ферментів) у шлунку проходять біологічні хімічні реакції, в результаті котрих харчова речовина розділяється на відповідні фракції, котрі виводяться зі шлунку. Тобто в шлунку проходить первинна переробка харчової речовини (ППХР) на легку та важку частини. Легка частина ППХР є найбільш цінною для біологічного організму, виводиться через стінку шлунку і направляється для живлення кровеносної системи, підшлункових органів, почок, печінки та інших важливих органів. Важка частина поступає в кишковий тракт, де в результаті вторинних біохімічних реакцій виводяться допоміжні фракції, котрі створюють відповідні зворотні зв'язки та забезпечують стабілізацію вихідних координат основних біологічних органів, включаючи й мозок.

Легка частина ППХР складається з багатьох фракцій, котрі відводяться зі шлунку відповідними елементами (за аналогією з тарілками ректифікаційних колон) і за заданою біопрограмою направляються відповідними мікротрубними каналами до того чи іншого органу біологічного організму, котрий у свою чергу є вторинним біохімічним реактором. Відходи переробки легкої частини ППХР виводяться з організму через потові залози, а також через інші системи. Кишковий тракт важкої частини ППХР теж являє собою аналог ректифікаційного апарату, в якому на відповідних ділянках виводяться ті чи інші її фрагменти. Останні за відповідною біопрограмою поступають у різні елементи біологічного організму. Між первинним і вторинним ректифікаційним апаратами розташовуються різні біосистеми, серед яких основними можуть бути наступні: мозкова система, до якої входить блок головного та спинного мозку; кровеносна система, до якої можна віднести серце, (рушійно-силовий блок), печінку (блок фільтрації

крові) та периферійний судинний блок; моченосна система, до якої входять: почки, блок накопичення та виведення мочі з біологічного організму; лімфеносна система, котра складається з блоку створення та накопичення лімфатичної рідини та блоку перенесення цієї рідини по лімфосудинах; м'язосилова система, яка складається з шлунково-кишечного м'язового блоку, блоку формування рухо-поступальних і поворотних дій, блоку формування роботи судинної та лімфатичної систем; кістково-формуєчої системи, до складу якої входить блок накопичення твердої біологічної структури, блок накопичення кальцію та інших інгредієнтів, а також блоку виробництва мастильної системи для суглобів. Окрім систем, котрі забезпечують основні функції біологічного організму, до складу останнього входить система підготовки біореагентів, котрі забезпечують роботу шлунка як біореактора. До цієї системи можна віднести, наприклад, блок жовчного міхура, підшлункову залозу, селезінку тощо.

Таким чином біологічний організм є складною біокомп'ютерною багатопараметричною системою управління, інтерфейсом зв'язку якої є сам біологічний організм з його здоровою, нервово-подразнюючою, термодинамічною, біохімічною, електромагнітною та іншими малими системами, котрі являють собою давачами різноманітної інформації вимірювального контролю. За аналогією з комп'ютерно-інтегрованими системами контролю та управління технологічними процесами (KICKY-TP) в біологічному організмі можна виділити три основні блоки (контролери) приймання, обробки та видання вимірювальної інформації: первинний «контролер» - головний мозок, у котрому записана біопрорама функціонування біологічного організму, вторинний «контролер» - спинний мозок, у котрому записана програма рухомих дій і третинний «контролер» - серцева система з її джерелами енергоживлення всіх елементів біосистеми, забезпечуючи захист від різноманітних електромагнітних впливів, а також відповідний їх ритм роботи [19-22]. На рис. 1 лініями зі стрілками показані основні функціональні зв'язки первинного реакторно-ректифікаційного апарату (шлунку) і вторинного RECTИФІКАЦІЙНОГО апарату (кишечника) з основними життєзабезпечуючими системами.

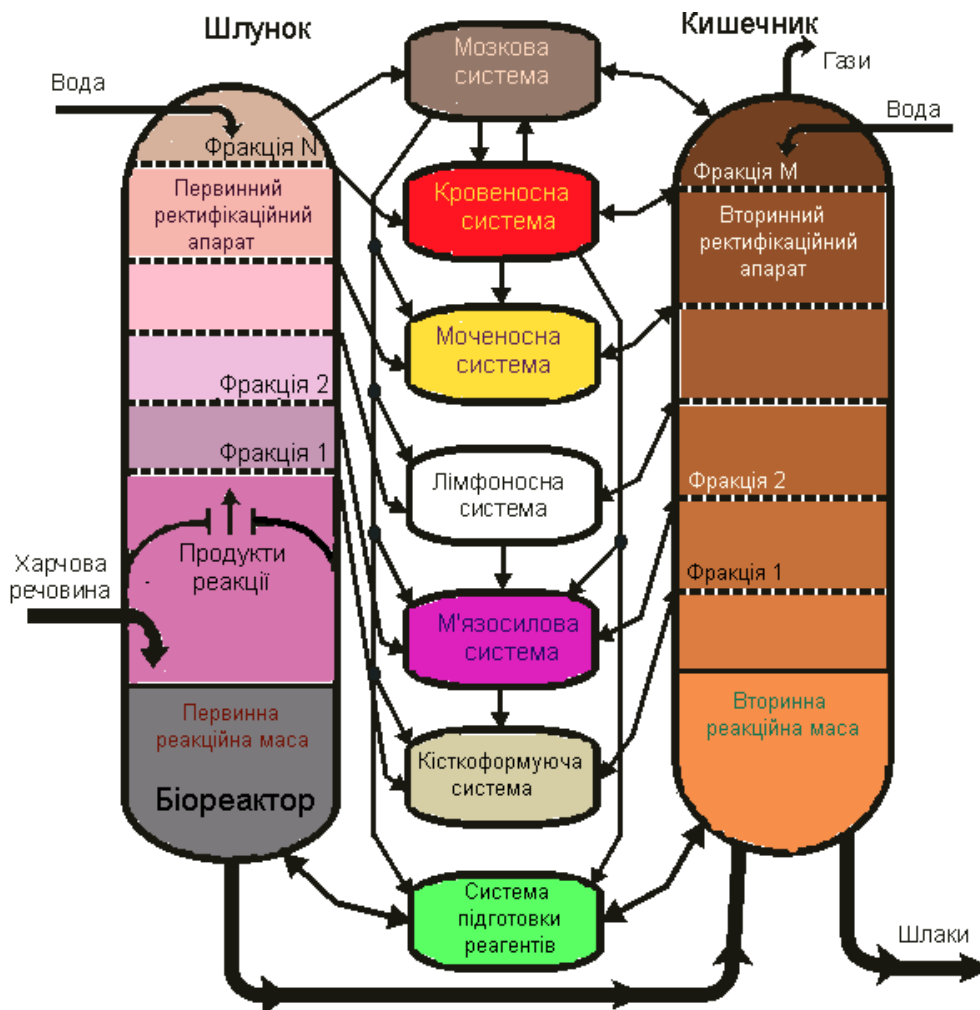


Рисунок 1 – Спрощена структурна схема біологічного організму

Між ректифікаційними апаратами та іншими системами існують як прямі так і зворотні зв'язки. Окрім того між самими системами теж існують відповідні зв'язки. Зворотні зв'язки в біологічному організмі за аналогією з системами автоматичного регулювання виконують стабілізуючу роль, тобто підтримують основні життєзабезпечуючі параметри (ОЖП) на заданому рівні, притаманному здоровому організму, наприклад, стабілізація температури біоорганізму при зміні температури навколишнього середовища за рахунок потовиділення, стабілізації кислотно-сольового складу реакційної маси шлунку за рахунок додаткового споживання води, стабілізація величини світлового потоку за рахунок звуження чи розширення зіниць ока та багато інших. За аналогією з КІСКУ-ТП ОЖП мають відповідні допустимі межі сигналізації та блокувань. Так, наприклад, при порожньому шлунку поступає сигнал з головного мозку про необхідність його наповнення, при повному шлунку – втрачається апетит до їжі, при високому або низькому кислотно-сольовому складі реакційної маси шлунку появляються в ньому больові симптоми.

Біологічний організм як об'єкт контролю, управління та реологічних переходів

Кожен апарат чи система біологічного організму є об'єктом контролю та управління, який має вхідні x_i , вихідні y_j і збурюючі z_n координати. У кожному апараті чи системі проходять відповідні переходи енергії, маси чи кількості руху, в результаті чого мають місце реологічні перетворення маси біоречовин, їх теплової енергії та кількості руху [23]. Так як такі процеси відносяться до основних явищ перенесення, котрі в теорії масо-теплоперенесення описуються законами Фур'є, Фіка та Ньютона, то з їх допомогою можна описувати аналогічні процеси в біологічному організмі. Як і в інших тепло-масообмінних процесах, при реологічних переходах біологічних організмів структурно мають місце наступні їх складові: область джерела живлення, тепла і кількості руху, зону реологічного переходу (перетворення маси, енергії і кількості руху) та область стоку створюваної енергії, маси та кількості руху (рис.2). У найпростішому вигляді кожний реологічний перехід складається з трьох блоків: 1 – блок є джерелом накопичення маси, енергії та кількості руху; 2 – блок є зоною реологічного перетворення маси, енергії та кількості руху за рахунок фізико-хімічного перетворення; 3 – блок стоку (виведення) нового продукту, енергії та кількості руху. Реологічний процес проходить як в часі t , так і за просторовими координатами x, y, z .

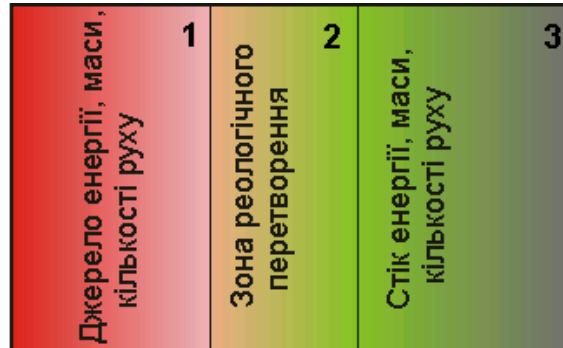


Рисунок 2 – Структурна схема реологічного переходу

Принцип роботи реологічного переходу полягає в тому, що з блоку 1 маса (енергія чи кількість руху) неперервно поступає в блок 2, в якому за рахунок фізико-хімічних дій перетворюється в іншу форму (наприклад, процеси розкладання, хімічні реакції, кількість руху), а нова речовина (енергія чи кількість руху) з блоку 2 виводиться (стікає) в інше середовище, яким може бути газ, рідина чи тверда речовина [24]. З цього випливає важливий для подальших досліджень висновок, що в блоках 1 і 3 мають місце процеси явищ перенесення як за часом, так і за просторовими координатами, а в блоці 2 – тільки за часом. Якщо прийняти, що джерело та стік є безмежними, то процес реологічного переходу можна зобразити схемою, приведеною на рис.3.

Як показують теоретичні та експериментальні дослідження, зона реологічного перетворення являє собою об'єм $V_{рп} = x_{рп} y_{рп} z_{рп} = const$ для відповідного реологічного переходу і є функцією тільки часу перетворення.

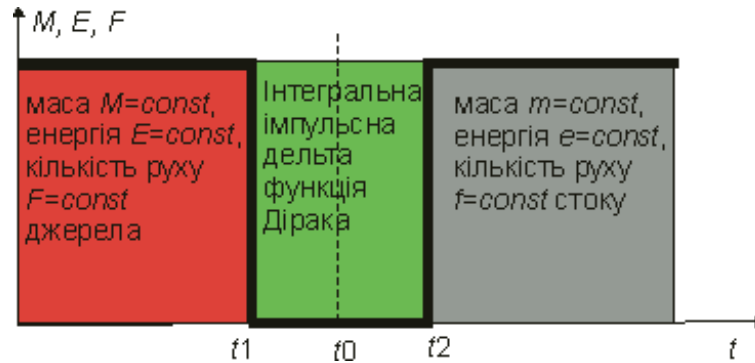


Рисунок 3 – Структурна схема реологічного переходу

Як показують теоретичні та експериментальні дослідження, зона реологічного перетворення являє собою об'єм $V_{RP} = x_{RP}y_{RP}z_{RP} = const$ для відповідного реологічного переходу і є функцією тільки часу перетворення.

Принципи фізичного та математичного моделювання реологічних перетворень у біологічному організмі

Якщо параметри джерела позначити функцією $\xi_D(x, y, z, t)$, параметри стоку функцією $\xi_C(x, y, z, t)$, а параметри інтегральної імпульсної δ -функції Дірака $\xi_I(V_{RP}, t)$, то процес реологічного переходу, показаний на рис. 3, можна описати наступними рівняннями для:

$$\text{- джерела перенесення } \xi_D(x, y, z, t) = \begin{cases} 0 & \text{при } t = 0 \\ 1 & \text{при } t = \infty \end{cases};$$

$$\text{- стоку перенесення } \xi_C(x, y, z, t) = \begin{cases} 0 & \text{при } t = 0 \\ 1 & \text{при } t = \infty \end{cases};$$

- зони реологічного перетворення

$$\xi_I(V_{RP}, t) = \int_{t_1}^{t_2} f(\zeta) \delta(\zeta - t_0) d\zeta = \begin{cases} 0 & t_0 < t_1, \quad t_0 \geq t_2 \\ f(t_0 + 0) & t_1 < t_0 < t_2 \end{cases},$$

де ζ - деяка змінна; $\delta(\zeta - t_0)$ - ядро лінійного інтегрального перетворення; t_0 - середнє значення часу фазового переходу; $\delta(\zeta) = \frac{d}{d\zeta} f(\zeta)$ - одинична ступінчаста імпульсна δ -функція; $f(\zeta)$ - функція розподілення маси, енергії та кількості руху в зоні реологічного перетворення.

Для структурної схеми рис. 3 функція $f(\zeta) = 1$. У реальних умовах перенесення маси, енергії та кількості руху їх функції розподілення $\xi_D(x, y, z, t)$, $\xi_C(x, y, z, t)$ і $\xi_I(V_{RP}, t)$ не дорівнюють одиниці, а саме реологічне перетворення може бути складним, наприклад, багатастадійним. Окрім того процес стоку, котрий є накопичувальним, може бути як аперіодичним, так і коливальним та описуватися диференціальним рівнянням високого порядку. Такі реологічні переходи мають місце в кожному блоці чи системі біологічного організму, котрий характеризується великою кількістю внутрішніх контролюючих параметрів. Багато з цих параметрів володіють сигнальними функціями, за рахунок чого біологічний організм відчуває біль в тому чи іншому елементі (блоці, системі), відсутність апетиту, рухових ефектів, запаморочення тощо. У таких випадках біологічний організм не може швидко (або зовсім не може) справитися з такою ситуацією і вимагає зовнішнього втручання в процес діагностування небажаної ситуації та зовнішнього впливу на її зменшення чи усунення. Враховуючи багатопараметричність біологічного організму, його діагностування є складною задачею. Серед різноманітних методів діагностики такого організму є неінвазивні, котрі будуються, як правило, на принципах перенесення маси, енергії (теплової, ультразвукової, оптичної, електромагнітної тощо) та кількості руху.

Висновки

В основу роботи біологічного організму покладений принцип перенесення маси, енергії та кількості руху. Такі явища перенесення характеризуються наявністю джерела, звідки маса, енергія та кількість руху передаються в елементи біологічного організму, терплять відповідне перетворення (наприклад, маса в масу, маса в енергію, маса в кількість руху, теплова енергія в кількість руху, інфрачервона енергія в теплову тощо). Створена в зоні реологічного переходу нова маса, енергія чи кількість руху переносяться шляхом тепло-масоперенесення в біологічному організмі, створюючи таким чином стік маси, теплової енергії та кількості руху. Перетворення маси, енергії та кількості руху в біологічному організмі проходить за відповідними фізико-хімічними законами: теплопровідності Фур'є, масоперенесення Фіка, перенесення кількості руху Ньютона, до яких можна додати ще й закон кінетики хімічних реакцій Арреніуса. Указані закони побудовані для ідеальних випадків, коли джерела маси, енергії та кількості руху є безмежними в часі та просторі й не враховують принципів накопичення створеної маси, енергії та кількості руху. При перенесенні маси, енергії та кількості руху створюється зона реологічного перетворення (аналогічно пограничного шару в гідродинаміці), у котрій проходять відповідні фізико-хімічні перетворення. У результаті таких перетворень створюються нові параметри, як концентрації речовин, теплові ефекти, термо- і масодинамічні процеси, котрі якраз і характеризують роботу того чи іншого біологічного органу. Існуючі неінвазивні методи діагностики побудовані на визначенні вторинних показників, які є характерними для одноступеневого процесу перенесення: *джерело* → *стік*. Найбільш повну інформацію про стан здоров'я біологічного організму можна отримати за первинними показниками, якими характеризуються зони реологічних перетворень.

Література

- 1) Франк-Каменецкий Д.А. Диффузия и теплопередача в химической кинетике. – М.: Наука, 1987. – 502.
- 2) Рамм В.М. Абсорбционные процессы в химической промышленности. – М.: Госхимиздат, 1951. – 386 с.
- 3) Берд Р., Стюарт В., Лайтфут Е. Явления переноса. – М.: Химия, 1974. – 688 с.
- 4) Баррер Р. Диффузия в твердых телах. – М.: Издательство, 1948. – 504 с.
- 5) Астарида Дж. Массоперенос с химической реакцией. – М.: Химия, 1971. – 493 с.
- 6) Темпел Б. Хемосорбция. – М.: ИЛ, 1958. – 382 С.
- 7) Кутателадзе С.С. Основы теории теплообмена. – М.: Машгиз, 1962. – 368 с.
- 8) Лыков А.В. Теория теплопроводности. – М.: Высш. шк., 1967. – 599 с.
- 9) Лыков А.В. и Михайлов Ю.А. Теория тепло-и массопереноса. – М.: Госэнергоиздат, 1963. – 389с.
- 10) Эккерт Э.Р., Дрейк Р.М. Теория тепло-и массообмена. – М.: Госэнергоиздат, 1962. – 562 с.
- 11) Де Гроот С., Мазур П. Неравновесная термодинамика. Пер. С франц. – М.: Мир, 1964, – 456 с.
- 12) Вайнберг А.М. Математическое моделирование процессов переноса. Решение нелинейных краевых задач. – Москва-Иерусалим, 2009. – 210 с.
- 13) Бояджиев Х, Бешков В. Массоперенос в движущихся пленках жидкости /Пер. с англ.- М. : Механика. – Вып. 43. – 1988. – 136 с.
- 14) Кафаров В.В. Основы массопередачи. – М.: Высш.шк., 1962. – 416 с.
- 15) Гиршфельдер Дж., Кертис Ч., Берд Р. Молекулярная теория газов и жидкостей. Пер. С англ.- М.: Издат инлит, 1961. – 929 с.
- 16) Шлихтинг Г. Теория пограничного слоя. Пер. С нем. – М.: Наука, 1969. – 742 с.
- 17) Маделунг Э. Математический аппарат физики. Пер. С англ.. – М.: Физматгиз, 1968. – 618 с.
- 18) Самарский А.А., Михайлов В.П. Математическое моделирование. Нелинейные дифференциальные уравнения математической физики. - М.: «Наука», 1987. - 280 с.
- 19) Павлов С.В. Неінвазивні оптико-електронні прилади та системи діагностики мікроциркуляції периферійного кровообігу. Дис. на здоб. наук. ступеня докт. техн. наук. – Вінниця, 2008.- 349 с.
- 20) Павлов С.В, Кожем'яко В.П., Петрук В.Г., Колісник П.Ф. Фотоплетизмографічні технології контролю серцево-судинної системи: (Монографія) – Вінниця: УНІВЕРСУМ-Вінниця, 2007. – 254 с.
- 21) Кожем'яко В.П., Павлов С.В., Станчук К.І. Оптико-електронні методи і засоби для обробки та аналізу біомедичних зображень: (Монографія) /Вінниця: УНІВЕРСУМ-Вінниця, 2006. – 203 с.
- 22) Шиян А.А., Злепко С.М., Павлов С.В., Хаїмзон І.І. Інформаційні технології для управління діяльністю людини. Монографія. – Вінниця: ВНТУ, 2012. – 316 с.
- 23) Стенцель Й.І. Математичне моделювання хімічних процесів на основі теорії реологічних переходів. Вісник Східноук.нац університету. Науковий збірник. №5 (111), Ч.2.- 2007. – с.91-97.
- 24) Стенцель Й.І. Фотоколориметричні газоаналізатори. Монографія. – К.: ІСДО. 1995. – 126 с.

Відомості про авторів

Стенцель Йосип Іванович – докт. техн. наук, професор, завідувач кафедри комп'ютерно-інтегрованих систем управління Технологічного інституту Східноукраїнського національного університету імені Володимира Даля.

Павлов Сергій Володимирович - докт. техн. наук, професор, завідувач кафедри загальної фізики та фотоніки Вінницького національного технічного університету. 21021, м. Вінниця, Вінницький національний технічний університет, вул. Хмельницьке шосе 95.

Злепко Сергій Макарович - докт. техн. наук, професор, завідувач кафедри проектування медико-біологічної апаратури. 21021, м. Вінниця, Вінницький національний технічний університет, вул. Хмельницьке шосе 95.

УДК 681.586.773

Й. І. СТЕНЦЕЛЬ, С. В. ПАВЛОВ, О. Д. АЗАРОВ, Л. І. ПЕТРОСЯН

Східноукраїнський національний університет імені Володимира Даля, м. Луганськ
Вінницький національний технічний університет, м. Вінниця**ФІЗИЧНІ ТА МАТЕМАТИЧНІ МОДЕЛІ РЕОЛОГІЧНИХ ПЕРЕХОДІВ У БІОЛОГІЧНОМУ ОРГАНІЗМІ**

Анотація. У роботі наведено результати досліджень роботи біологічного організму. Показано, що в біологічному організмі проходять процеси, котрі супроводжуються явищами перенесення енергії, маси та кількості руху. Такі процеси обумовлені біохімічними реакціями, біомагнітними процесами, а також наявністю границь переходу, в яких створюються нові речовини, види енергії та витрати біоречовин. Розглядаються фізичні моделі реологічних перетворень, а також, а також узагальнена математична модель біологічного організму.

Ключові слова: біологічний організм, метод, фізична модель, зона перенесення, маса, енергія, кількість руху, обмеження, математична модель.

Аннотация. В работе приведены результаты исследований работы биологического организма. Показано, что в биологическом организме протекают процессы, которые сопровождаются явлениями переноса энергии, массы и количества движения. Такие процессы обусловлены биохимическими реакциями, биомагнитными процессами, а также наличием границ перехода, в которых создаются новые вещества, виды энергии и расходы биовеществ. Рассматриваются физические модели реологических преобразований, а также обобщенная математическая модель биологического организма.

Ключевые слова: биологический организм, метод, физическая модель, зона переноса, масса, энергия, количество движения, ограничения, математическая модель.

Abstract. The paper presents the results of studies of the biological organism. It is shown that in biological organisms are processes which are accompanied by energy transfer phenomena, mass and momentum. Such processes are caused by biochemical reactions biomagnetic processes, as well as the presence of transition boundary, which creates new substances, forms of energy and cost biorechovyn. We consider the physical model rheological changes as well, as well as a generalized mathematical model of a biological organism.

Keywords: biological organism, method, physical model, a zone transfer, mass, energy, momentum, limitations, mathematical model.

Вступ

Біологічні процеси за своєю структурою та характером у багатьох випадках аналогічні процесам, котрі мають місце в хімічній промисловості. У всіх випадках як біологічні, так і хімічні процеси супроводжуються явищами перенесення маси, енергії та кількості руху. Як у перших, так і в других процесах проходять перетворення маси, енергії та кількості руху, у результаті чого створюються нові речовини, види енергії та руху, котрі відповідним чином накопичуються та переносяться в інші матеріальні середовища, створюючи таким чином їх стік. У науковій літературі велика увага явищам перенесення приділялася технологічним процесам: хімічним, біохімічним, фармацевтичним, нафтопереробним і надзвичайно мало процесам перенесення маси, енергії та кількості руху, які пов'язані з життєдіяльністю біологічного організму у тому числі й людського. Як вказується в [1-9], явища перенесення маси, енергії та кількості руху відносяться до нелінійних, розподілених за часом і просторовими координатами, у зв'язку з чим виникають труднощі їх описання в аналітичній, доступній для подальшого використання формі. Тому дослідники, як правило, суттєво спрощували фізичні моделі процесів явищ перенесення, обмежували їх граничними та крайовими умовами таким чином, щоби нелінійне диференціальне рівняння того чи іншого явища перенесення привести до лінійного чи квазілінійного.

Відомо [1], що гравітаційна сила $F = g(\rho_r - \rho)$ є рухомою силою, яка призводить до виникнення перенесення кількості маси чи тепла в середовищі, яким може бути й біологічний організм. В аналітичному описанні вона входить у загальне векторне рівняння балансу сил і кількості руху. Іншими балансовими рівняннями є рівняння нерозривності (баланс маси) і рівняння балансу, які описують будь-який процес перенесення, котрий викликає зміну густини (щільності). Таким чином, завжди мають місце по крайній мірі три рівняння, котрі визначають параметри перенесення. До таких параметрів відносяться швидкість, тиск, енергія й концентрація. Крім того, необхідні рівняння, котрі зв'язують параметри стану процесу перенесення. Потрібно також знати коефіцієнти молекулярного перенесення, тобто в'язкість для ньютонівської рідини, коефіцієнт енергопровідності, коефіцієнт дифузії компонентів у середовищі та деякі інші коефіцієнти.

При дослідженні явищ перенесення конкретизуються задачі, котрі накладаються на механізми, які викликають рух, у вигляді геометричної форми, умов на характер поверхні, навколишнє середовище тощо. Величинами, котрі потрібно визначити аналітично або експериментально при будь-якому явищі перенесення, є потік перенесення та поле швидкостей. Наприклад, коефіцієнт енерговіддачі α від поверхні до середовища визначається через енергетичний потік Q і задані умови за енергією E_0 і E_∞ , де E_0 і E_∞ — енергія з випромінюючої поверхні тіла площею S та середовища необмеженої товщини відповідно [2]. Відповідно рівняння статки для передачі енергетичного потоку Q записується в такому вигляді

$$Q = \alpha S(E_0 - E_\infty), \quad (1)$$

Загальними рівняннями перенесення для рухомих потоків служать звичайні рівняння гідромеханіки. Рухомим механізмом у вимушених потоках є градієнт густини, енергії та швидкості потоку. Рівняння руху, викликаного перенесенням енергії (аналогічно маси), мають вигляд [3-5]:

$$\frac{D\rho}{D\theta} = -\rho\nabla v \text{ або } \frac{\partial\rho}{\partial\theta} = -(\rho\nabla v + v \cdot \nabla\rho); \quad (2)$$

$$\rho \frac{Dv}{D\theta} = \rho g - \nabla p + \mu \nabla^2 v + \frac{1}{3} \mu \nabla(\nabla v); \quad (3)$$

$$\rho c_p \frac{DE}{D\theta} = \nabla k \nabla E + \beta E \frac{Dp}{D\theta} + \mu \Phi + q''', \quad (4)$$

де: ρ - густина (або щільність); θ - час перенесення маси, енергії та кількості руху від джерела до середовища; v - швидкість перенесення; g - прискорення земного тяжіння; $\bar{\rho}$ - визначальний параметр процесу перенесення (наприклад, густина потоку); ∇v - градієнт зміни швидкості; $\nabla\rho$ - градієнт зміни визначального параметра; ∇p - градієнт статичного тиску; μ - динамічна в'язкість речовини; c_p - питома теплоємність; β - коефіцієнт об'ємного розширення; Φ - дисипативна функція; ∇E - градієнт зміни енергії; k - коефіцієнт енергопровідності; q''' - питома потужність об'ємних джерел енергії.

Перше рівняння (рівняння нерозривності) виражає умову зберігання імпульсу маси. Це скалярне рівняння зв'язує миттєву швидкість зміни щільності субстанції в деякій точці поля, виражену через повну похідну $D/D\theta$, з місцевою швидкістю розширення або стискування ∇V , яке обумовлене полем швидкості. Друге рівняння (векторне) виражає рівність сили, обумовленої місцевим прискоренням, сумі місцевої об'ємної сили, сили, обумовленої градієнтом тиску, та сил в'язкості. Третє рівняння (скалярне) виражає закон зберігання енергії. Як вказується в [9-13], основна трудність при розв'язку рівнянь має місце через можливі зміни параметрів перенесення μ , k і щільності (густини) ρ . Параметри μ і k залежать головним чином від температури і в деяких випадках можна приймати, що вони є сталими. Але, щоби отримати рух маси чи енергії, потрібно завжди враховувати зміну щільності енергії чи густини речовини. Градієнти концентрації хімічних компонентів у середовищі, яке знаходиться в полі гравітаційної сили, можуть створювати реологічні переміщення, котрі викликані градієнтними силами. У багатьох випадках зміна концентрації речовини є єдиною рушійною силою перенесення маси. Тоді параметр E входить у рівняння (2) у вигляді енергії, якою може бути й температура, та обумовлений перенесенням теплового, ультразвукового, світлового чи іншого потоку. Щоби зв'язати конвекційне та дифузійне перенесення хімічних компонентів у біологічному організмі, потрібні додаткові рівняння зберігання, аналогічні рівнянню (3) для енергії. Якщо одночасно проходить дифузія декількох різних хімічних компонентів, то потрібно стільки ж таких рівнянь. Рівняння балансу для конвекційного перенесення з врахуванням дифузії та будь-яких хімічних реакцій приймає вигляд

$$\frac{DC}{D\theta} = \nabla \cdot D_C \nabla C + v''', \quad (5)$$

де D_C - коефіцієнт дифузії компонента; v''' — швидкість створення компонента C в одиниці об'єму.

Розглянуті повні рівняння, які описують енергоперенесення в динамічних потоках мають складний вигляд. Такими рівняннями рекомендується користуватися у випадках, де потрібно враховувати різні фізичні процеси та механізми [12-14]. Перший детальний розрахунок перенесення, викликаного виштовхуючою силою, зроблено Лоренцом [15]. Розглянуті найпростіші умови: стаціонарне ламінарне перенесення маси в шарі, який примикає до зануреної в спокійний навколишній газ вертикальної поверхні; температура газу є сталою та рівною температурі T_∞ безмежного середовища, температура випромінюючої тепло поверхні також підтримується сталою та рівною T_0 . Якщо $T_0 > T_\infty$, то (якщо густина рідини (газу) зменшується з ростом температури, що звичайно має місце) витискуюча сила $B(x, y) = g(\rho_\infty - \rho)$ є позитивною, то рух потоку направлений знизу вверх і навпаки. Якщо поверхня має в напрямку z велику протяжність, то процес перенесення описується в системі декартових координат x, y повною системою рівнянь (1) - (3). Найпростішим видом перенесення є стаціонарний ламінарний

рухомий потік. Тоді член ∇V у рівняннях (2) і (3) можна прийняти рівним нулю. Енергетичні ефекти, які описуються в рівнянні (4) членом, котрий враховує поле тиску, розподіленими джерелами q''' і дисипативним членом $\mu\Phi$, іноді дуже малі. Таким чином, різниця $(g\rho - \nabla p)$ між об'ємною силою та силою тиску в рівнянні (3) з деяким наближенням можна прийняти в багатьох випадках рівною $g(\rho - \rho_\infty)$. Таким чином, для ламінарного потоку рівняння (2) - (4) приймають вигляд:

$$\nabla \cdot V = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0; \quad (6)$$

$$\rho \left(u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right) = \mu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + g(\rho_\infty - \rho); \quad (7)$$

$$\rho \left(u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \right) = \mu \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right); \quad (8)$$

$$\rho c_p \left(u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} \right) = k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right). \quad (9)$$

При рішенні рівнянь (6) - (9), як правило, приймають, що розподілення швидкості $v(x, y) = 0$. Тоді з рівняння (6) випливає, що $u(x, y) = u(y)$. При цій умові рівняння (6) і (8) вилучаються, а рівняння (7) і (9) спрощуються. У рівнянні (7) конвекційним членом $u \partial u / \partial x$ (перенесення кількості руху) нехтують, так як $u = u(y)$. Крім того, нехтують ефектами в'язкості й енергопровідності. Накінець, залежність густини (щільності) від зміни енергії $\rho(E)$ приймається лінійною

$$\rho(E) = \rho(E_\infty) [1 - \beta(E - E_\infty)], \quad (10)$$

де β - термічний коефіцієнт об'ємного розширення середовища.

Рівняння, які після таких спрощень залишилися, для функцій $u(y)$ і $E(x, y)$ мають вигляд:

$$\mu \frac{\partial^2 u}{\partial y^2} + g\rho\beta(E - E_\infty) = 0, \quad \text{і} \quad k \frac{\partial^2 E}{\partial y^2} - \rho c_p u \frac{\partial E}{\partial x} = 0.$$

Система рівнянь складається з нелінійних рівнянь, так як до них входить член, який містить похідну $\partial E / \partial x$. Тому для подальшого спрощення автори приймають, що оскільки в області перенесення тепла енергія змінюється від E_∞ при $x = 0$ до $E(x, y)$ при $x = L$, то величину $\partial E / \partial x$ можна наближено подати як $(E - E_\infty) / L$. Тоді залежність $E(x)$ вилучається і $E(x, y) = E(y)$. Уводячи замість $E(y)$ нову змінну $\phi = [E(y) - E_\infty] / [E_0 - E_\infty]$, отримують звичайну систему диференціальних рівнянь за координатою y у вигляді:

$$\frac{d^2 u}{dy^2} + \frac{g\rho\beta(E_0 - E_\infty)}{\mu} \phi = 0, \quad (13)$$

$$\frac{d^2 \phi}{dy^2} - \frac{\rho c_p u}{kL} \phi = 0. \quad (14)$$

З цього випливає, що процеси перенесення маси, енергії та кількості руху розглядалися односторонньо за схемою «джерело \Rightarrow стік», причому у всіх випадках джерело на границі перенесення має безмежну концентраційну та енергетичну потужність. Рівняння (2) - (4) описують процес перенесення енергії, маси та кількості руху тільки від джерела з максимальним потенціалом φ_0 до іншого середовища з

меншим потенціалом φ . Різниця цих потенціалів $\Delta\varphi = \varphi_0 - \varphi$ є рушійною силою процесу перенесення. Для масоперенесення за таку рушійну силу приймається різниця концентрацій $\Delta C = C_0 - C$ речовин або різниця їх густини $\Delta\rho = \rho_0 - \rho$ (виштовхуюча сила); для енергоперенесення рушійною силою може бути різниця температур $\Delta T = T_0 - T$ або різниця енергетичних потенціалів $\Delta E = E_0 - E$, де C_0, ρ_0, T_0, E_0 - відповідно концентрація, густина, температура та енергія на границі їх переходу в інше середовище. З вищеведеного випливає, що явища перенесення маси, енергії та кількості руху в класичній теорії розглядалися як одностадійний процес – від джерела до деякого іншого середовища, про що свідчать рівняння (2) - (4), в яких є тільки одна швидкість перенесення маси $D\rho/Dt$, об'єму DV/Dt та енергії DE/Dt . Але ці швидкості характеризують лише перенесення від джерела безмежної потужності до іншого середовища, а фактично від джерела до зони реологічних перетворень і не враховують яким чином проходить перетворення маси, енергії та кількості руху в іншому середовищі та де накопичується новостворена маса, енергія чи кількість руху. Права частина рівнянь (2) - (4) описує процеси перенесення маси, енергії та кількості руху за геометричними координатами x, y, z , а також вплив зміни енергії та кількості руху на процес перенесення маси, маси та енергії на процес перенесення маси тощо. Як правило, перенесення маси, енергії та кількості руху здійснювалося в середовищі безмежних розмірів, про що свідчать граничні умови, котрі накладувалися на методи аналітичних рішень цих рівнянь. Рівняння Лоренца описують процеси перенесення маси, енергії та кількості руху при умові безмежної потужності джерел, сталих швидкостях $\frac{D\rho}{Dt} = const, \frac{DV}{Dt} = const, \frac{DE}{Dt} = const$ і часу $t = \infty$. Практично у всіх випадках приймалося, що новостворена маса, енергія чи кількість руху не накопичується, або переноситься в безмежність, або дорівнює кількості маси, енергії та кількості руху, яка передана джерелом. Окрім того приймалося, що час перенесення маси, енергії та кількості руху і час перенесення новоствореної маси, енергії та кількості руху є одним і тим же часом, що призводить до парадоксу про безмежну швидкість перенесення [15].

Мета статті

Розробити фізичні та математичні моделі перенесення кількості руху, енергії та маси в біологічному організмі людини на основі теорії реологічних переходів та перетворень.

Постановка задачі

У біологічному організмі, у тому числі й людському, протікають тепломасо-об'ємні процеси, котрі супроводжуються біохімічними процесами, в результаті чого створюються матеріальні та теплові потоки, швидкість яких залежить від кількості руху, обумовленої роботою судинно-серцевої системи. Задача полягає в тому, що на основі теорії реологічних переходів і перетворень виконати аналіз таких біологічних процесів і розробити фізичні та математичні моделі.

Фізичні моделі реологічних перетворень маси, енергії та кількості руху в біологічному організмі

Аналіз процесів перенесення кількості руху, маси та енергії показує, що при дослідженні процесів масо-енергоперенесення приймалося, що джерела маси та енергії (концентраційний та енергетичний напір) мають безмежну потужність і ніколи не враховувалося, що між джерелом і досліджуваним тілом завжди є деяка межа розділу (реальна чи уявна), через яку здійснюється перенесення імпульсу кількості руху, маси та енергії [14 - 15]. Перехід імпульсу кількості руху, маси та енергії через межу розділу називається реологічним переходом, а процес, який протікає при цьому переході – реологічним перетворенням. Фізична модель найпростішого реологічного переходу показана на рис. 1.

Джерело з безмежною потужністю (кількості руху $Q_L = Q_{L0}$, маси $Q_M = Q_{M0}$, енергії $Q_E = Q_{E0}$) на межі розділу «джерело-тіло» (далі під терміном «тіло» називатимемо біологічний організм) зліва характеризується деякими своїми сталими параметрами, наприклад, температурою нагріву $T = T_{G0} = const$, концентрацією $C = C_{G0} = const$, енергією $E = E_{G0} = const$. У початковий момент часу $t = 0$ тіло характеризується своїми параметрами: $q_l = q_{l0}$; $q_m = q_{m0}$ і $q_e = q_{e0}$ або температурою $T = T_{\Omega 0} = const$, концентрацією $C = C_{\Omega 0} = const$, енергією $E = E_{\Omega 0} = const$. Причому $q_l < Q_L < q_l$, $q_m < Q_M < q_m$ і $q_e < Q_E < q_e$.

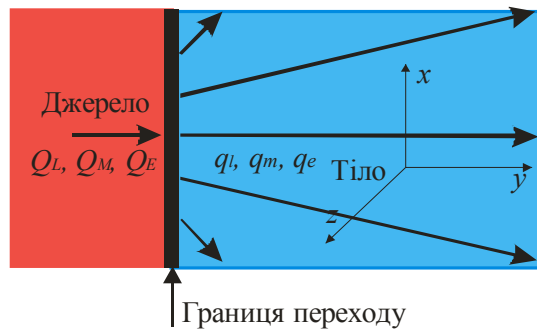


Рисунок 1 – Фізична модель реологічного переходу

Якщо джерело й тіло миттєво з’єднуються між собою, то кількість руху, маси та енергії від джерела передається тілу за рахунок того чи іншого перетворення. Якщо передається імпульс кількості теплоти, то такий перехід підпорядковується закону Фур’є, якщо передається імпульс маси, то закону Фіка, якщо передається імпульс кількості руху, то закону Ньютона. При умові безмежної потужності джерела на границі розділу фаз (на границі між джерелом і тілом) створюється пригранична зона як показано на рис. 2.

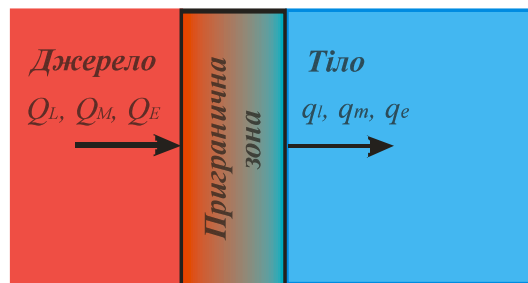


Рисунок 2 – Процес перенесення з одним реологічним переходом

Кількість руху Q_L , маси Q_M чи енергії Q_E від джерела з безмежною потужністю передається в приграничну зону, в якій відповідним чином перетворюється і вже в перетвореному вигляді розповсюджується (стікає) в тілі. Пригранична зона може бути дуже вузькою, наприклад, ламінарний приграничний шар на вертикальній стінці, між незмішуваними рідинними середовищами, між двома тілами, які труться поміж собою, при передачі високочастотної енергії від одного тіла до іншого тощо, а також достатньо широкою зоною, як наприклад, при наявності гнійних процесів, адсорбційно-десорбційних процесів, хімічних перетворень у шлунково-кишковому тракті тощо. Реологічні перетворення, які мають один реологічний перехід, відносяться до найпростіших (наприклад, нагрівання тіла до температури нижчої від критичної), процеси розчинення речовин в тілі, при яких не створюються другорядні речовини, наприклад, розчинення цукру, змішування кислот чи спиртів тощо. Одномірні технологічні процеси, як правило, описуються найпростішими лінійними диференціальними рівняннями, такими як рівнянням Фур’є для теплопровідності $q_T = -\lambda \nabla T$, рівняннями Фіка для масопровідності $q_m = -D_m \nabla C$, рівнянням Ньютона для перенесення кількості руху, де λ - коефіцієнт теплопровідності; T - температура; D_m - коефіцієнт дифузії; C - концентрація перетворюваних речовин. Хімічні процеси в тілі, як правило, супроводжуються одночасним перенесенням кількості руху, маси та енергії. Такі процеси (багатомірні) мають дві й більше зони реологічних переходів, а відповідно, перетворень. За принципом дії зони розділяються на теплообмінні, масообмінні, енерго-масообмінні та реакційні. Реакційні розділяються на одностадійні, двостадійні та багатостадійні [15]. На рис. 3 показана фізична модель двостадійного реологічного перетворення.

До таких процесів можна віднести енерго-масообмінні процеси, наприклад, процеси абсорбції, десорбції, ректифікації тощо. Багатостадійні реологічні процеси, як правило, мають місце в шлунково-кишковому тракті та серцево-судинній системі тіла при наявності хімічних перетворень. На рис. 4 показана чотиристадійна фізична модель при наявності одностадійної хімічної реакції. Першим реологічним переходом може бути процес перемішування харчових компонентів у шлунку, другим – процес нагрівання реакційних компонентів джерелом тепла за рахунок екзотермічного розкладання реакційних компонентів і кровоносна система, третім – процес масоперенесення концентрації реагуючих компонентів у

шлунку й четвертим – процес хімічного перетворення, який може бути як екзотермічним, так і ендотермічним.

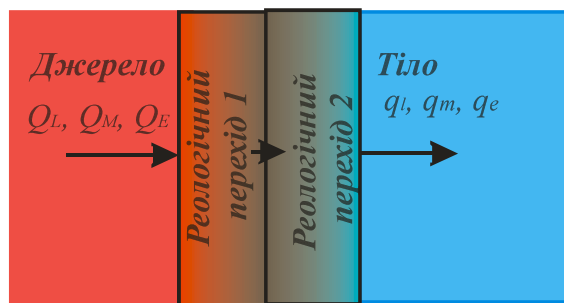


Рисунок 3 – Процес перенесення з двома реологічними переходами

При багатостадійних реологічних хімічних перетвореннях проміжні продукти створюють нові реологічні перетворення. Для таких процесів характерним є те, що стоками можуть бути як проміжні продукти, так і кінцевий продукт багатостадійного хімічного перетворення.

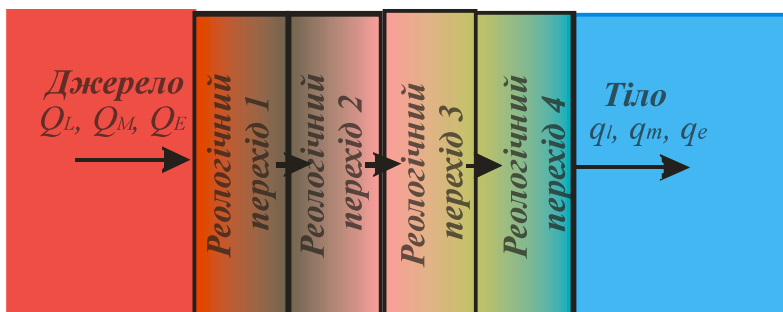


Рисунок 4 – Процес перенесення з чотирма реологічними переходами

Узагальнена математична модель явищ перенесення

На кожному реологічному переході імпульс вхідної величини зменшується, а вихідної – збільшується. У загальному випадку кожний реологічний перехід являє собою деяку інтегральну імпульсну дельта-функцію Дірака, як показано на рис. 5, де позначено: Q, Q_0 - поточна та початкова концентрації; q, q_0 - поточна та кінцева кількість імпульсу; $Q = f(z, t)$ - функція зміни концентрації; $q = f(z, t)$ - функція зміни концентрації; z - впливовий фактор або сукупність впливових факторів; t - час реологічного переходу; z_1, t_1 - параметри початку реологічного переходу; z_2, t_2 - параметри кінця реологічного переходу.

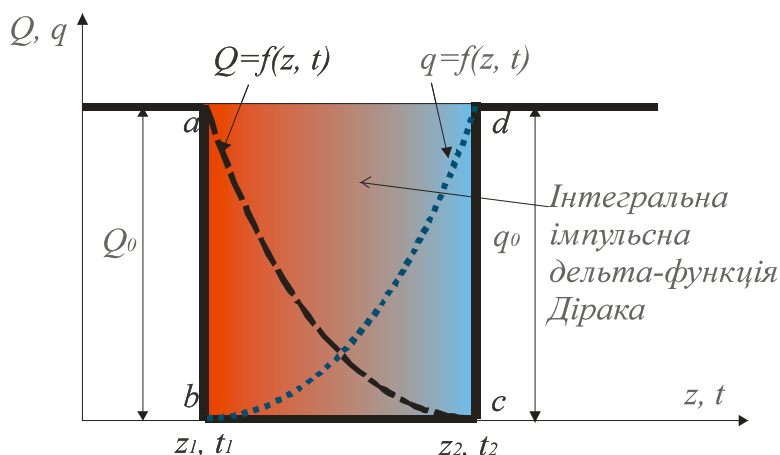


Рисунок 5 – Графічне зображення зони реологічного переходу

У загальному випадку $q \neq Q$, а час реологічного переходу $\theta = t_2 - t_1$. Площина, яка обмежена точками a, b, c, d , є інтегральною імпульсною дельта-функцією Дірака з ядром у форму деякої функції, яка описує процеси, котрі проходять в зоні реологічного переходу. Так як всередині інтегральної імпульсної дельта-функції Дірака, як правило, знаходяться дві криві реологічного переходу – крива зміни того чи іншого явища перенесення (кількості руху, маси та енергії) $Q = f(z, t)$ і крива стоку процесу перетворення $q = f(z, t)$, то завжди виконуватиметься баланс кількості руху, маси та енергії, тобто

$$\bar{Q}(\bar{z}, \theta) = \bar{\varphi}(\bar{z}, \theta) \cdot \bar{v}(\bar{z}, \theta) + \bar{q}(t), \quad (15)$$

де $\bar{\varphi}(\bar{z}, \theta)$ - вектор перенесення потенціалу; \bar{z} - вектор напрямку руху перенесення; $\bar{v}(\bar{z}, \theta)$ - вектор швидкості перенесення.

Так як процеси явищ перенесення кількості руху, маси та енергії незворотні, то їх можна показати у формі реологічних переходів, як показано на рис. 6, а. На цьому рисунку показаний найпростіший варіант одномірного реологічного переходу, коли імпульс джерела $Q = Q_0 = const$. Такий процес може мати місце при нагріванні тіла, розчиненні твердих речовин, змішуванні рідин з різними концентраціями тощо. На рис. 6, б) приведений графік НРП зміни вхідного параметра $Q(z, \theta)$ за деякою координатою зони z реологічного переходу. При $z = z_1$ значення параметра $Q(z_1, \theta) = Q_0$, а при $z = z_2$ - $Q(z_2, \theta) = Q_k$, де Q_0, Q_k - початкове та кінцеве значення вхідного визначального параметра. Так як у зоні реологічного переходу (ЗРП) визначальний параметр $Q(z, \theta)$ змінюється від Q_0 до Q_k за відповідним законом (Фур'є, Фіка чи Ньютона), то згідно з [3, 4] такий процес перенесення визначального параметра можна описати рівнянням вигляду

$$\frac{Q(z, \theta) - Q_k}{Q_0 - Q_k} = \operatorname{erf}\left(\frac{z}{2\sqrt{\xi \cdot \theta}}\right), \quad (16)$$

де ξ - коефіцієнт перенесення визначального параметра (для перенесення теплоти – коефіцієнт температуропровідності, для перенесення маси – коефіцієнт дифузії).

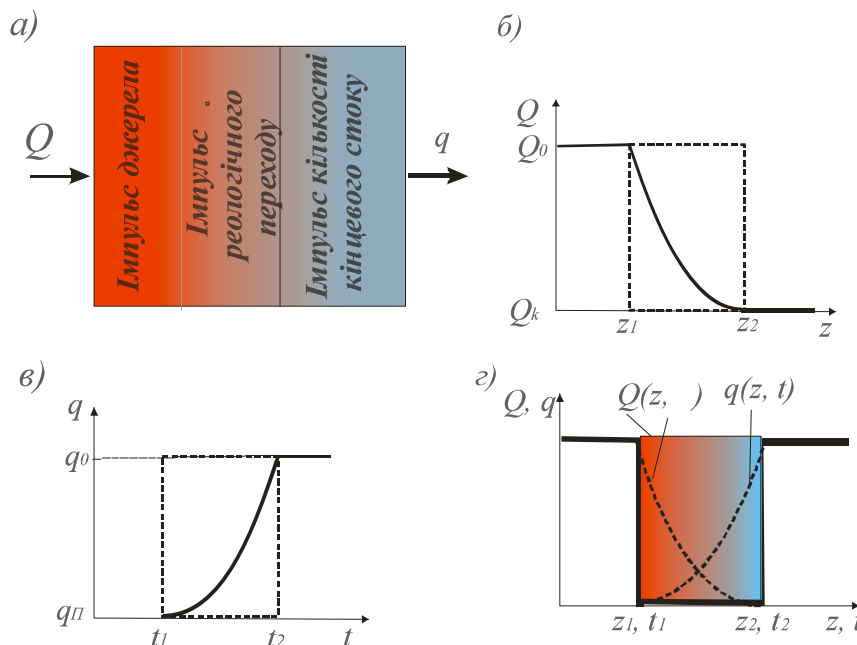


Рисунок 6 – Фізична модель (а) і графіки незворотних реологічних переходів (НРП): б) – графік НРП імпульсу кількості притоку від джерела; в) – графік незворотного імпульсу кінцевого стоку; г) – графік інтегральної імпульсної δ – функції Дірака

У зоні реологічного переходу (ЗРП) вихідний визначальний параметр змінюється від деякого початкового значення $q(z, t) = q_{II}$ до максимального $q(z, t) = q_0$ за законом, який визначається процесом реологічного перетворення, залежить від потенціалу перенесення, швидкості, просторової координати та інших параметрів перенесення як тиск, в'язкість, густина тощо (рис. 6, в). Крива $q(z, t) = f(z, t)$ формує стік вихідного визначального параметра, який може змінюватися від $t = t_1 \geq 0$ до $t = t_2 \leq \infty$. Наприклад, такий процес може бути описаний таким диференціальним рівнянням інтегруючої динамічної ланки другого порядку

$$\tau_2^2 \frac{d^3 q(z, t)}{dt^3} + \tau_1 \frac{d^2 q(z, t)}{dt^2} + \frac{dq}{dt} = kQ(z, \theta), \quad (17)$$

де τ_1, τ_2 - деякі сталі часу реологічного перетворення; k - коефіцієнт перетворення.

Для ламінарних потоків руху рідин можна прийняти, що прискорення руху незначне. Тоді $\rho Dv / D\theta = 0$ і система рівнянь (2) - (4) з врахуванням (17) приводиться до наступної системи нелінійних диференціальних рівнянь:

$$\left\{ \begin{array}{l} \frac{d\rho}{d\theta} + (\rho \nabla v + v \nabla \rho) = \sum_{i=0}^n \tau_i^i \frac{d^{i+1} \rho}{dt^{i+1}}; \\ \rho c_p \frac{dE}{d\theta} + \nabla k \nabla E + \beta E \frac{dp}{d\theta} + \mu \Phi + q^m = \sum_{i=0}^n \tau_i^i \frac{d^{i+1} E}{dt^{i+1}}. \end{array} \right. \quad (18)$$

$$\left\{ \begin{array}{l} \frac{d\rho}{d\theta} + (\rho \nabla v + v \nabla \rho) = \sum_{i=0}^n \tau_i^i \frac{d^{i+1} \rho}{dt^{i+1}}; \\ \rho c_p \frac{dE}{d\theta} + \nabla k \nabla E + \beta E \frac{dp}{d\theta} + \mu \Phi + q^m = \sum_{i=0}^n \tau_i^i \frac{d^{i+1} E}{dt^{i+1}}. \end{array} \right. \quad (19)$$

де n - кількість фазових реологічних переходів; τ_i^i - сталі часу стоку результату реологічного перетворення.

Висновки

Система рівнянь (18) - (19) являє собою математичну модель енерго-масоперенесення в біологічному організмі, яка є функцію ядра інтегральної імпульсної дельта-функції Дірака. Ліва частина цих рівнянь описує процес конвекційно-кондуктивного перенесення з часом θ , а права - процес стоку результату реологічного перетворення з часом t . Час стоку t може дорівнювати або бути більшим часу θ , але ніколи не меншим. Рівняння (18) і (19) виключають парадокс про безмежність швидкості перенесення, що є важливим при дослідженні реальних процесів енерго-масоперенесення. Так як ці рівняння описують процеси енерго-масоперенесення в області інтегральної імпульсної дельта-функції Дірака, то на границях цієї функції (див. рис. 5, границі a-b і c-d) похідні за визначальним параметром перенесення дорівнюють нулю. Для цих границь $\frac{d\rho}{d\theta} = \frac{d\rho}{dt} = 0$ і $\frac{dE}{d\theta} = \frac{dE}{dt} = 0$, що може бути однією з умов аналітичного розв'язку нелінійних диференціальних рівнянь (18) і (19).

Література

- 1) Берд Р., Стьюарт В., Лайтфут Е. Явления переноса. - М.: Хмиия, 1974. - 688 с.
- 2) Лыков А.В. Теория теплопроводности. - М.: Высш. шк., 1967.-599 с.
- 3) Эккерт Э.Р., Дрейк Р.М. Теория тепло-и массообмена. - М.: Госэнергоиздат, 1962. - 562 с.
- 4) Вайнберг А.М. Математическое моделирование процессов переноса. Решение нелинейных краевых задач. - Москва-Иерусалим, 2009. - 210 с.
- 5) Франк-Каменецкий Д.А. Диффузия и теплопередача в химической кинетики. - М.: Наука, 1987. - 502.
- 6) Стенцель Й.І. Фотоколориметричні газоаналізатори. Монографія. - К.: ІСДО. 1995. - 126 с.
- 7) Стенцель Й.І. Математичне моделювання хімічних процесів на основі теорії реологічних переходів. Вісник Східноук.нац університету. Науковий збірник. №5 (111), Ч.2.- 2007. - с.91-97.
- 8) химической кинетики. - М.: Наука, 1987. -502.
- 9) Кафаров В.В. Основы массопередачи. - М.: Высш.шк., 1962. - 416 с.
- 10) Рамм В.М. Абсорбционные процессы в химической промышленности. - М.: Госхимиздат, 1951. - 386 с.

- 11) Хоблер Т. Теплопередача и теплообмінники. – Л.: Гос.наук.-техническое узд-во хим. лит, 1961. – 820 с.
- 12) Гиршфельдер Дж., Кертис Ч., Берд Р. Молекулярная теория газов и жидкостей. Пер. С англ.- М.: Издат инлит, 1961. – 929 с.
- 13) Шлихтинг Г. Теория пограничного слоя. Пер. С нем. – М.: Наука, 1969. – 742 с.
- 14) Маделунг Э. Математический аппарат физики. Пер. С англ.. – М.: Физматгиз, 1968. – 618 с.
- 15) Вайнберг А.М. Математическое моделирование процессов переноса. Решение нелинейных краевых задач. – Москва-Иерусалим, 2009. – 210 с.

Відомості про авторів

Стенцель Йосип Іванович – докт. техн. наук, професор, завідувач кафедри комп'ютерно-інтегрованих систем управління Технологічного інституту Східноукраїнського національного університету імені Володимира Даля.

Павлов Сергій Володимирович - докт. техн. наук, професор, завідувач кафедри загальної фізики та фотоніки Вінницького національного технічного університету. 21021, м. Вінниця, Вінницький національний технічний університет, вул. Хмельницьке шосе 95.

Азаров Олексій Дмитрович – докт. техн. наук, професор, завідувач кафедри обчислювальної техніки, 21021, м. Вінниця, Вінницький національний технічний університет, вул. Хмельницьке шосе 95

Петросян Лілія Іванівна – науковий співробітник кафедри комп'ютерно-інтегрованих систем управління Технологічного інституту Східноукраїнського національного університету імені Володимира Даля.

КОМП'ЮТЕРНІ СИСТЕМИ ТА КОМПОНЕНТИ

УДК 621.3

Л. В. КРУПЕЛЬНИЦЬКИЙ, В. Я. СТЕЙСКАЛ, С. В. БОГОМОЛОВ

Вінницький національний технічний університет, м. Вінниця

РОЗРОБКА ТА МЕТРОЛОГІЧНА АТЕСТАЦІЯ КАНАЛІВ ТОЧНОГО ЧАСУ В АВТОМАТИЗОВАНІЙ СИСТЕМІ ДЛЯ ОФІЦІЙНОГО ТЕЛЕРАДІОМОНІТОРИНГУ

Анотація. Проаналізовано функціональне призначення системи офіційного телерадіомоніторингу. Запропоновано оригінальну методику з використанням відеокамери для одночасного знімання зображень індикаторів часу взірцевого приладу і того, який досліджується. Розглянуто підхід, щодо визначення границі допустимих абсолютних похибок у вимірювальних каналах точного часу.

Ключові слова: канали точного часу, вимірювальні канали, телерадіомоніторинг, автоматична система, метрологічна атестація.

Аннотация. Проанализировано функциональное назначение системы официального телерадиомониторинга. Предложена оригинальная методика с использованием видеокамеры для одновременной съемки изображений индикаторов времени образцового прибора и того, который исследуется. Рассмотрен подход относительно определения границы допустимых абсолютных погрешностей в измерительных каналах точного времени.

Ключевые слова: каналы точного времени, телерадиомониторинг, автоматическая система, метрологическая аттестация.

Abstract. The functional purpose of the official TV radio monitoring were analyzed. The original technique of using video cameras for simultaneous image capture time indicator exemplary device and what is studied. The approach of designating the permissible absolute error in measuring the exact time channels.

Keyword: channels of precise time, measurement channels, TV and radio monitoring, automatic system, metrological certification.

Вступ

Система офіційного моніторингу теле-, радіопрограм призначена для багатоканального приймання, вибіркового перегляду, запису і зберігання ефірних, кабельних та супутникових програм телебачення і радіо з метою планового і оперативного контролю відповідності змісту телерадіопрограм Законам України про телерадіомовлення, рекламу та іншим чинним документам, які стосуються діяльності Національної ради України з питань телебачення і радіомовлення.

В основу розробки моніторингової системи покладено теоретичні та практичні результати досліджень у двох суміжних напрямках інформаційних технологій:

- побудови високопродуктивних багаторозрядних аналого-цифрових перетворювачів з використанням систем числення з ваговою надлишковістю (в тому числі, на основі кодів Фібоначчі та «золотої пропорції») [1,2];
- створення спеціалізованих студійних, вимірювальних і моніторингових аналого-цифрових комп'ютерних систем для телерадіомовлення.

В частині спеціалізованих апаратних компонентів ефективність системи ґрунтується на методах надлишковості, автокалібрування і самоконтролю обладнання, а в частині програмного забезпечення – на методах адаптивного ущільнення, кореляційного та спектрального аналізу відео- та аудіосигналів.

Мета

Проведення аналізу розробки каналів точного часу в автоматизованій системі для офіційного телерадіомоніторингу та здійснення їх метрологічної атестації.

Постановка задач

- 1) Проаналізувати функціональне призначення системи офіційного телерадіомоніторингу та виділити їх основні метрологічні характеристики.
- 2) Проаналізувати запропоновану методику визначення похибок вимірювання часу.
- 3) Розглянути підхід, щодо визначення границі допустимих абсолютних похибок у вимірювальних каналах точного часу.

Розв'язання задач

Однією з основних функціональних характеристик автоматизованої системи офіційного телерадіомоніторингу є можливість визначення часу мовлення та тривалості телерадіопередач і їх фрагментів. Метрологічній атестації [3] підлягають вимірювальні канали «Синхро» в режимах запису відеоаудіофайлів передач (режим запису) та при визначенні часу мовлення та тривалості телерадіопередач при відтворенні записаних файлів (режим відтворення).

В режимі запису забезпечується вставка відміток точного часу в відеоаудіосигнали під час їх запису в комп'ютерні файли. Типовий вимірювальний канал має структуру, наведену на рисунку 1. Супутниковий приймач сигналів точного часу «Синхро-01СТ» приймає сигнали точного часу з системи глобального позиціонування GPS, декодує їх в секундні імпульси та в адресно-часовий код стандарту SMPTE, відеоінсертер титрів точного часу «Синхро-TB2» декодує вказаний код та наносить на кожен

кадр зовнішнього відеосигналу титри логотипу «RADA, годин, хвилин, секунд, номеру кадру, дня, місяця та року» як ілюструє рисунок 2 [4]. Відеоаудіосервер записує відео та аудіосигнали у вигляді файлів. Сynchronізація системного часу комп'ютера відбувається по мережі Ethernet від тайм-сервера, вмонтованого в «Синхро-01СТ». Аудіосигнали на початку кожного файлу помічаються звуковими відмітками точного часу стандарту DTMF. Назва кожного відео та аудіофайлу містить час початку запису, сформований системним годинником комп'ютера.

У режимі відтворення записаних відеоаудіофайлів оператор робочої станції за допомогою програми EditorARM аналізує зміст телерадіопередач, а також визначає час мовлення та тривалість передач і їх фрагментів. Для незалежного визначення часу мовлення радіопрограм при її відтворенні крім програми EditorARM може застосовуватись декодер-індикатор DTMF-відміток точного часу аудіозаписів «Синхро-DTMF» [4]. Схема типового вимірювального каналу в режимі відтворення і аналізу наведена на рисунку 3.

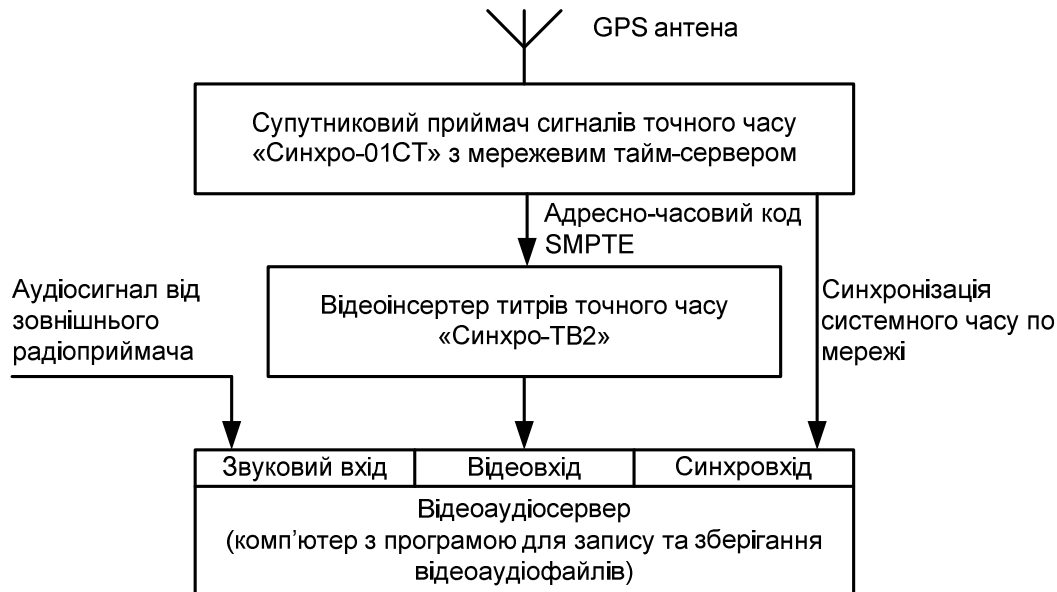


Рисунок 1 – Структурна схема вимірювального каналу «Синхро» в режимі запису

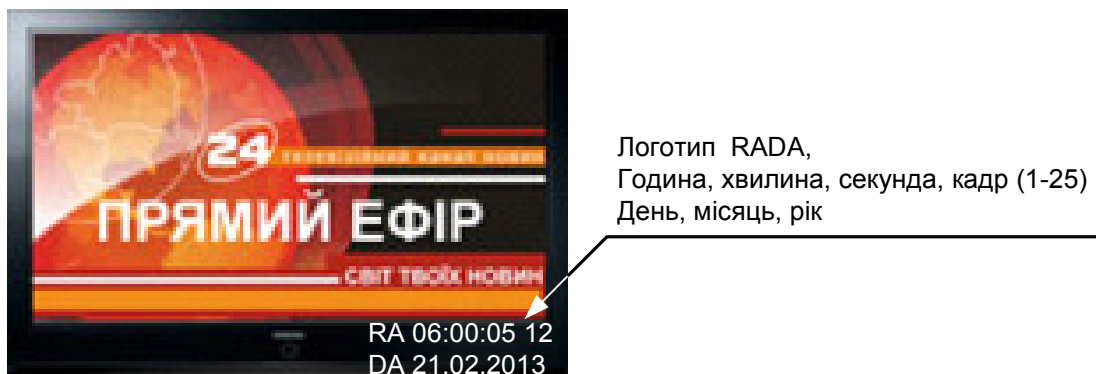


Рисунок 2 – Формат титрів точного часу на кожному кадрі відео файлів

Отже, до складу вимірювальних каналів (ВК) «Синхро» входять як апаратні, так і програмні компоненти. Значення поточного часу представлені титрами точного часу на відеокадрах, звуковими позначками на початку аудіофайлів, а також наявні в іменах відеоаудіофайлів та у системних годинниках комп'ютерів мережі. Аналіз сукупності вказаних факторів дозволив сформуванати набір метрологічних характеристик і якісних параметрів моніторингової системи при визначенні часу мовлення і тривалості телерадіопередач, які подано в таблиці 1.

Визначення границі допустимої абсолютної похибки секундного імпульсу на виході супутникового приймача «Синхро-01СТ» є досить тривіальною метрологічною задачею і проводиться шляхом співставлення з наявним взірцевим датчиком точного часу. Проте, дослідження інших характеристик носить досить специфічний характер, оскільки кінцеві результати вимірювань часу представляються не в

електричному вигляді, а у вигляді зображень на відеокдрах, на цифрових індикаторах «Синхро-01СТ» і «Синхро-DTMF», у вікнах часу операційної системи та операторської програми комп'ютерів.

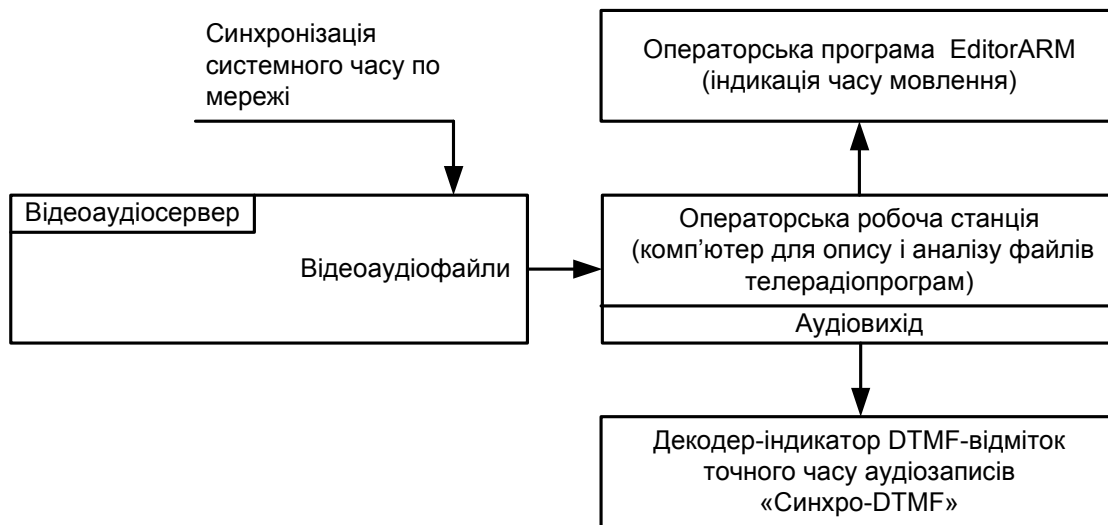


Рисунок 3 – Структурна схема вимірювального каналу «Синхро» в режимі відтворення

Таблиця 1 – Метрологічні характеристики і параметри якості вимірювальних каналів

Назва режиму роботи	Метрологічні характеристики вимірювальних каналів	Характеристики параметрів якості і технічного рівня
Режим запису	Границі допустимої абсолютної похибки генеральної сукупності ВК формування часу секундного імпульсу на виході супутникових приймачів «Синхро-01СТ»	Границі допустимої абсолютної похибки генеральної сукупності ВК формування титрів часу відеоінserterами «Синхро-ТВ2». Границі допустимої абсолютної похибки генеральної сукупності ВК відтворення часу системними годинниками відеоаудіосерверів.
Режим відтворення	Границі допустимої абсолютної похибки генеральної сукупності ВК визначення часу мовлення телепередач в операторській програмі. Границі допустимої абсолютної похибки генеральної сукупності ВК визначення часу мовлення радіопередач декодером-індикатором «Синхро-DTMF».	Границі допустимої абсолютної похибки генеральної сукупності ВК визначення тривалості часу телерадіопередач в операторській програмі.

На думку авторів, їм вдалося запропонувати оригінальну методику визначення похибок вимірювань часу, значення якого представлені у вигляді цифрових зображень на індикаторах приладів або на екрані комп'ютера. В запропонованій методиці використовується відеокамера, яка одночасно проводить знімання візрцевого індикатора часу і індикатора часу пристрою, що досліджується. Звичайна сучасна цифрова miniDV-відеокамера формує відеозапис з 25-ти відеокадрів в секунду (стандарт PAL), тобто роздільна здатність вимірювань складає 0.040 с, що цілком достатньо для даного застосування. Аналіз відеозапису проводиться безпосередньо на відеокамері або на комп'ютері в режимі покадрового перегляду. Різницю значень часу між показами візрцевого пристрою та пристрою, що досліджується визначають прямим порівнянням в межах одного чи кількох кадрів (якщо індикація містить дробові значення часу – секунди, долі секунд або номери кадрів) або шляхом підрахунку кількості кадрів між відповідними моментами зміни значень секунд чи хвилин на обох пристроях (якщо дробові значення часу відсутні). За цією методикою, до речі, можна визначити до долей секунди похибку годинників з цифровими чи аналоговими шкалами, що не мають індикації секундних значень (фіксуються моменти переходу між хвилинами).

Нижче подамо скорочені схеми виконання окремих досліджень з визначення границі допустимих абсолютних похибок часу в вимірювальних каналах «Синхро», які ілюструють викладену вище ідею.

Оцінку границі допустимої абсолютної похибки формування титрів часу відеоінserterами «Синхро-ТВ2» проводять для кожного з відібраних вимірювальних каналів відповідно до схеми вимірювань, яку показано на рисунку 4, методом порівняння цифрових значень часу на індикаторі «Синхро-01СТ» і на титрах відеокадрів, згенерованих inserтером «Синхро-ТВ2».

Вимірювання проводять фіксуючи на відеокамеру значення точного часу на індикаторі «Синхро-01СТ». В реальному часі відеосигнал з лінійного виходу відеокамери поступає на вхід inserтера «Синхро-ТВ2», який також додає до кожного кадру титри часу. Синхронізація «Синхро-ТВ2» здійснюється від «Синхро-01СТ» адресно-часовим кодом SMPTE. Вихідний сигнал «Синхро-ТВ2», записується в відеофайл відеоаудіосервера штатною програмою відеоархіватора і для контролю виводиться на монітор, як проілюстровано прикладом на рисунку 4.

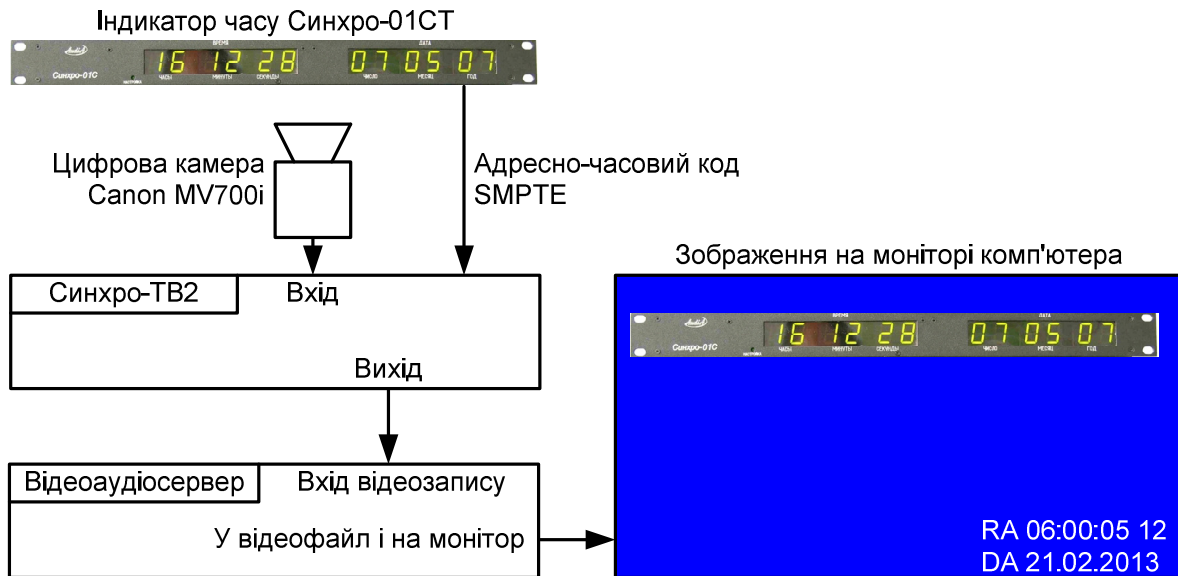


Рисунок 4 – Схема вимірювання абсолютної похибки формування титрів часу відеоінserterом «Синхро-ТВ2» з використанням відеокамери

Аналіз записаного відеофайла здійснюють шляхом покадрового перегляду зображення в програмі Media Player Classic, яка входить до складу операційної системи Microsoft Windows XP відеоаудіосервера. Використовуючи номери кадрів, наявні в титрах часу визначають кількість кадрів, на які різняться моменти зміни секунд на зображенні індикатора «Синхро-01СТ» і в титрах часу, згенерованих «Синхро-ТВ2». Оскільки тривалість одного кадру становить 1/25 с або 40 мс, то кількість кадрів перераховують в значення абсолютної похибки в мілісекундах.

Оцінку границі допустимої абсолютної похибки визначення часу мовлення радіопередач декодером-індикатором «Синхро-DTMF» проводять відповідно до схеми вимірювань на рисунку 5 методом порівняння цифрових значень часу на індикаторі «Синхро-01СТ» і на індикаторі «Синхро-DTMF». Для цього запуск відтворення записаного звукового файлу з відеоаудіосервера здійснюють в момент часу, який співпадає за хвилинами і секундами з моментом початку запису цього звукового файлу.



Рисунок 5 – Схема вимірювання абсолютної похибки визначення часу мовлення радіопередач декодером-індикатором „Синхро-DTMF”

Для автоматичного запуску відтворення звукового файлу в момент часу, який точно співпадає з моментом початку запису цього файлу використовують функцію операційної системи Windows XP «Призначені завдання» з панелі керування. Наприклад, запис звукового файлу розпочато в 16 год. 00 хв. 00 с, про що свідчить назва файлу, а тривалість файлу складає 15 хв. Розміщуємо файл в «Призначених завданнях» на 16 год. 00 хв. 00 с наступної доби. Відтворення файлу автоматично розпочнеться саме в цей час програмою, яка налаштована операційною системою на відтворення звукових файлів. На початку відтворення файлу декодер-індикатор «Синхро-DTMF» засинхронізується і почне відлік часу запису файлу. Вимірювання проводять одночасно, фіксуючи на відеокамеру значення точного часу на індикаторі «Синхро-01СТ» і на індикаторі «Синхро-DTMF». Аналіз записаного відеофайла здійснюють шляхом покадрового перегляду зображення безпосередньо на відеокамері. Визначають кількість кадрів, на які різняться моменти зміни секунд на зображенні індикатора «Синхро-01СТ» і на індикаторі «Синхро-DTMF».

Дослідження границі допустимої абсолютної похибки визначення часу мовлення телепередач в операторській програмі для кожного з відібраних каналів проводять методом порівняння цифрових значень часу в титрах записаного відеофайлу і в вікні часу операторської програми. Аналіз відтворюваного відеофайлу здійснюють шляхом покадрового перегляду зображення в операторській програмі EditorARM, яка входить до складу спеціалізованого програмного забезпечення відеоаудіосервера. Використовуючи номери кадрів, наявні в титрах часу, визначають кількість кадрів, на які різняться моменти зміни секунд на відеозображенні і в вікні часу операторської програми. Кількість кадрів різниці перераховують в значення абсолютної похибки в мілісекундах.

Висновки

1) Проаналізовано функціональне призначення системи офіційного телерадіомоніторингу, що дало можливість виділити основні метрологічні характеристики і параметри якості системи при визначенні часу мовлення та тривалості телерадіопередач.

2) Запропоновано оригінальну методику з використанням відеокамери для одночасного знімання зображень індикаторів часу вірцевого приладу і того, який досліджується, що дозволило здійснити метрологічну атестацію вимірювальних каналів «Синхро».

3) Розглянуто підхід, щодо визначення границі допустимих абсолютних похибок у вимірювальних каналах точного часу. Визначення абсолютної похибки здійснюється шляхом покадрового перегляду відеозапису на комп'ютері або безпосередньо на відеокамері. Вказана методика була практично успішно апробована [5] і дозволила оцінити значення похибок часу до десятих долей секунди.

Список літератури

1) Крупельницький Л. В. Аналого-цифрові пристрої систем, що самокоригуються, для вимірювань і оброблення низькочастотних сигналів : монографія / Л. В. Крупельницький, О. Д. Азаров. – Вінниця : УНІВЕРСУМ – Вінниця, 2005. – 167 с. – ISBN 966-641-126-1

2) Азаров О. Д. Основи теорії аналого-цифрового перетворення на основі надлишкових позиційних систем числення : монографія / Азаров О.Д. – Вінниця : УНІВЕРСУМ – Вінниця, 2004. – 257 с.

3) Метрологія. Терміни та визначення: ДСТУ 2681-94. – [Чинний від 01.01.1995]. – К. : Держстандарт України, 1995. – 72 с.

4) Специализированное и измерительное оборудование собственной разработки и производства: каталог разработок НТЦ «Аналого-цифровые системы» ВНТУ/ [авт.-упоряд. Стейскал В.Я., Крупельницький Л.В. и др.]. — Винница : 2013. — 40 с.

5) Система офіційного телерадіомоніторингу. Вимірювальні канали «Синхро» для визначення часу мовлення та тривалості телерадіопередач : УО 4728690 / 8.144-2008 ДМА. – Офіц. вид. – Львів : ДП НДІ «Система», 2008. – 40 с. – (Програма та методика метрологічної атестації)

Відомості про авторів

Крупельницький Леонід Віталійович - к.т.н., доцент кафедри обчислювальної техніки, Вінницький національний технічний університет, Хмельницьке шосе, 95, м. Вінниця.

Стейскал Віктор Ярославович – к.т.н., завідувач науково-технічного центру «Аналого-цифрові системи», Вінницький національний технічний університет, Хмельницьке шосе, 95, м. Вінниця, 21021.

Богомолов Сергій Віталійович – к.т.н., старший викладач кафедри обчислювальної техніки, Вінницький національний технічний університет, Хмельницьке шосе, 95, м. Вінниця.

УДК 519.254: 621.317

П. Ф. ЩАПОВ, Р. П. МИГУЩЕНКО

НТУ «Харьковский политехнический институт», г. Харьков

СИНТЕЗ ДВУМЕРНЫХ ДИАГНОСТИЧЕСКИХ ПАРАМЕТРОВ ПРИ КОВАРИАЦИОННОМ АНАЛИЗЕ ТРЕХМЕРНЫХ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ ВИБРОСИГНАЛОВ

Анотація. Представлені результати ковариційного аналізу неперервного вейвлет-перетворення спектрально-нестационарних вібросигналів. Запропонована модель кусочно-лінійної апроксимації любого з перетину трьохмірного вейвлет-зображення послідовностями регресій з випадковими коефіцієнтами, дисперсійний аналіз яких дозволяє виділити чотири критеріальні двомірні статистики Фішера, які залежать або від зсуву, або від масштабу. Відображено, що максимально інформативними для завдань вібраційної діагностики є F-статистики, залежні від масштабу. Отримані F-статистики, будучи випадковими величинами з відомим законом розподілу, дозволяють перейти від якісної оцінки вимірювальних характеристик до кількісної і оцінити ймовірності помилок діагностики на базі стандартних параметричних моделей.

Аннотация. Представлены результаты ковариационного анализа непрерывного вейвлет-преобразования спектрально-нестационарных вибросигналов. Предложена модель кусочно-линейной аппроксимации любого из сечений трехмерного вейвлет-изображения последовательностями регрессий со случайными коэффициентами, дисперсионный анализ которых позволяет выделить четыре критеріальні двумерные статистики Фишера, которые зависят либо от сдвига, либо от масштаба. Отражено, что максимально информативными для задач вибрационной диагностики являются F-статистики, зависящие от масштаба. Полученные F-статистики, являясь случайными величинами с известным законом распределения, позволяют перейти от качественной оценки измерительных характеристик к количественной и оценить вероятности ошибок диагностики на базе стандартных параметрических моделей.

Abstract. Results of the kovariatsionny analysis of continuous veyvlet-transformation of spectral and non-stationary vibrosignals are presented. or from scale. The model of piecewise and linear approximation of any of sections of the three-dimensional veyvlet-image by sequences of regressions with the casual coefficients which dispersive analysis allows to allocate Fischer's four criteria two-dimensional statisticians who depend either on shift, or on scale is offered. Reflected that as informative for problems of vibration diagnostics are F-statistics that depend on the scale. The resulting F-statistic, as random variables with known distribution law, allow you to go from a qualitative assessment to quantitative measurement characteristics and evaluate the probability of error diagnosis based on standard parametric models.

Введение

Вейвлет-анализ случайных сигналов, содержащих диагностическую информацию об изменениях в состоянии сложного физического объекта, наиболее эффективен тогда, когда сигналы отличаются локализацией спектральной нестационарности. Визуальное различие трехмерных графических моделей вейвлет-преобразования сигнала для разных состояний объекта диагностики – яркий пример информационной избыточности трехмерного вейвлет-изображения. Однако, выделение в такой трехмерной модели информативных по видам состояний объекта, параметров, меняющихся, как свою максимальную величину, так и ее положение на плоскости «сдвиг-масштаб» – задача сложная. Главная проблема обеспечения статистически приемлемых, для параметрического тестирования вероятностных свойств информационных параметров, как случайных функций только одной (либо масштаб, либо сдвиг) переменной. Замена трехмерного вейвлет-изображения амплитуд спектральных составляющих случайного сигнала двумерными изображениями нескольких (от двух до четырех) информативных параметров упростит дискриминацию состояний объекта диагностики. Более того, появляется возможность автоматизации процедур не визуального, а количественного распознавания визуально-неразличимых трехмерных вейвлет-изображений, относящихся к различным состояниям объекта. Такое распознавание возможно, если информативные двумерные параметры содержат информацию о соотношениях между систематической и случайной составляющими амплитуды для любой точки трехмерного вейвлет-изображения. Это позволяет совместно контролировать не только огибающую его поверхности, но и спектральную плотность сигнала в локальных зонах наблюдения последнего.

Применение вейвлет-анализа существенно расширило возможности в получении новой, фактически дополнительной информации, недоступной при других подходах в задачах распознавания состояний сложных объектов различной физической природы. Графическая интерпретация информационной составляющей вейвлет-изображений различной мерности (вплоть до дисперсии вейвлет-коэффициентов) широко используется в практических приложениях исследовательских научных разработок в странах СНГ [1, 2] и дальнего зарубежья [3, 4, 5].

Достаточно эффективными подходами к сжатию трехмерных вейвлет-изображений являются использование двумерных контурных детекторов (метод Канни) [2, 5], мультимасштабные методы выделения контурных двумерных моделей [6, 7]. Однако такие подходы обеспечивают получение информации в качественном, а не количественном выражении. Если же рассматривать методы получения количественно выраженной информации, то информативные параметры – одномерны и информационно ограничены. В основном, это дисперсии вейвлет-коэффициентов [8] или классические методы сжатия данных для задач архивизации [9].

Цель статьи

Описание статистического подхода к синтезу информативных параметров вейвлет-изображения, как статистик дисперсионного отношения Фишера, позволяющих планировать контроль локализованных спектральных изменений случайных вибрационных измерительных сигналов.

Ковариационная модель вейвлет-изображения

Пусть $X(t)$ – случайный сигнал, который содержит измерительную информацию в форме спектральных изменений, включая изменения своей дисперсии во времени. Прямое непрерывное вейвлет-преобразование осуществляет перевод исходного сигнала в набор вейвлет-коэффициентов в координатах масштаба и смещения. Такое преобразование выполняется согласно выражения:

$$W_{\Psi}(a, b)X = \frac{1}{\sqrt{C_{\Psi}}} \int \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right) X(t) dt, \quad (1)$$

где $W_{\Psi}(a, b)X$ – вейвлет-коэффициенты, $\Psi(t)$ – материнский вейвлет, a и b – параметры, определяющие соответственно масштаб (частота) и смещение (время) функции $\Psi(t)$, C_{Ψ} – нормировочный множитель.

Для восстановления исходного сигнала по известным вейвлет-коэффициентам, используется обратное непрерывное вейвлет-преобразование:

$$X(t) = \frac{1}{\sqrt{C_{\Psi}}} \iint \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right) [W_{\Psi}(a, b)X] \frac{dad b}{a^2}. \quad (2)$$

Для непосредственного вычисления вейвлет-коэффициентов, интегралы (1, 2), описывающие прямое и обратное вейвлет-преобразования, могут быть представлены в виде суммы, взятой по дискретному набору масштабов. Несмотря на дискретность набора, подобное представление обеспечивает достаточно точное приближение вейвлет-коэффициентов.

Обозначим как $x(t)$ – реализацию сигнала $X(t)$, полученную в ходе измерений в моменты $t_1, \dots, t_i, \dots, t_N$. Осуществим непрерывное вейвлет-преобразование реализации $x(t)$ в соответствии с математической моделью [10]:

$$W_{g_k}(a, b)x = \frac{a}{\sqrt{C_{g_k}|a|}} \sum_{i=1}^N x_i \left[g_{k-1}\left(\frac{t_{i-1}-b}{a}\right) - g_{k-1}\left(\frac{t_i-b}{a}\right) \right], \quad (3)$$

где $W_{g_k}(a, b)x$ – вейвлет-коэффициенты, которые вычислены при помощи гаусовского вейвлета k -го порядка, $g_{k-1}(t)$ – гаусовский материнский вейвлет $(k-1)$ -го порядка, C_{g_k} – нормировочный множитель, x_i – отсчеты реализации $x(t)$ в моменты времени t_i , $i = \overline{1, N}$.

Математическая модель (3) является моделью трехмерных наблюдений характеристик $W_{g_k}(a, b)x = Y$, a и b . Их взаимосвязь, а также взаимная межгрупповая изменчивость определяются особенностями спектральной нестационарности сигнала $x(t)$ на всем интервале $[t_1, t_N]$ его наблюдения.

Трехмерную модель (3) можно заменить конечным множеством двумерных моделей

$$\{Y_{jl} = F_a[a_j | b_l = const] + \delta_{jl}\}_{l=1}^L, \quad (4)$$

$$\{Y_{jl} = F_b[b_l | a_j = const] + \varepsilon_{jl}\}_{j=1}^J. \quad (5)$$

Функции F_a и F_b это неслучайные функции (тренды), а δ_{jl} и ε_{jl} – случайные величины (случайные остатки). Функции F_a и F_b можно оценить, представив их последовательностями линейных регрессий со случайными коэффициентами. Дисперсионный анализ таких последовательных случайных регрессий (групп двумерных наблюдений) позволяет выявить статистически значимые аддитивные и мультипликативные изменения трендов на фоне неустраняемых шумов (случайных остатков).

Критериальные F -статистики такого анализа, построенные на отношениях дисперсий систематического и случайного (остаточного) изменений вейвлет-коэффициентов, Y_{jl} несут информацию об усредненных, на интервалах двумерных наблюдений геометрических параметрах всего трехмерного вейвлет-изображения в координатах Y , a и b . Теория дисперсионного анализа случайных регрессий (ковариационный анализ [11]) выделяет четыре независимые F -статистики, позволяющие тестировать на статистическую значимость независимые составляющие дисперсионного разложения вейвлет-коэффициентов Y_{jl} по сечениям a и b трехмерного вейвлет-изображения модели (3).

Плотности распределения вероятностей F -статистик известны. Это либо центральные, либо нецентральные F -распределения, число степеней которых определяется планом ковариационного анализа (количеством регрессий, общим числом отсчетов по осям a и b , количествами отсчетов по группам двумерных наблюдений, уровнем значимости α принятия решений, видом модели односторонней классификации групп наблюдений).

Рассмотрим общий вид любой из двумерных моделей множеств (4) или (5) в форме кусочно-линейной, по математическому ожиданию, последовательности k -групп двумерных наблюдений

$$Y_{si} = A_s + B_s z_{si} + \Delta_{si}, \quad (6)$$

где s – номер группы наблюдений, для которой оцениваются коэффициенты A_s и B_s частной (групповой) регрессии; $s = \overline{1, K}$; i – номер наблюдения в s -той группе; $i = \overline{1, n_s}$ (причем, $\sum_{s=1}^K n_{s_k} = J$ для модели (4), в которой $l = \text{const}$ и $\sum_{s=1}^K n_{s_k} = L$ для модели (5), в которой $j = \text{const}$); n_s – число наблюдений в s -той группе; Δ_{si} – случайный остаток, соответствующий либо остатку δ_{jl} (модель (4)), либо остатку ε_{jl} (модель (5)).

Требования к случайным остаткам Δ_{si} стандартны [11]: 1) нулевые матожидания; 2) постоянство остаточной дисперсии; 3) отсутствие взаимной корреляции; 4) нормальность закона распределения вероятностей. Выполнение первого требования обеспечивается использованием метода наименьших квадратов при оценивании коэффициентов A_s и B_s , а из оставшихся требований главным является второе. Требования 3) и 4) для ковариационного анализа большой роли не играют [11].

Дисперсионное разложение общей суммы квадратов отклонений Y_{si} в модели (6)

$$Q = \sum_{s=1}^K \sum_{i=1}^{n_s} (Y_{si} - \bar{Y})^2$$

от общего среднего \bar{Y} включает пять составляющих

$$Q = Q_0 + Q_1 + Q_2 + Q_3 + Q_\Delta, \quad (7)$$

из которых Q_Δ – остаточная сумма, отвечающая за неустраняемый случайный шум трехмерного вейвлет-изображения. За локальные внутригрупповые геометрические особенности изображения отвечают суммы Q_1 , Q_2 (частные аддитивные смещения) и Q_3 (мультипликативные изменения). Сумма Q_0 характеризует общее линейное изменение средних значений Y_{si} при монотонных

изменениях j (модель (4)) и l (модель (5)). F -статистики, величина которых отражает статистическую значимость систематических (общего и локальных) изменений рельефа вейвлет-изображения по сечениям a и b (масштаб и смещение), определяются дисперсионными отношениями [11]

$$F_0 = \frac{Q_0}{Q_\Delta} (N - 2k), \quad (8)$$

$$F_1 = \frac{Q_1}{Q_\Delta} (N - 2k), \quad (9)$$

$$F_2 = \frac{Q_2}{Q_\Delta} \left(\frac{N - 2k}{k - 2} \right), \quad (10)$$

$$F_3 = \frac{Q_3}{Q_\Delta} \left(\frac{N - 2k}{k - 1} \right), \quad (11)$$

Поскольку суммы дисперсионного разложения (7) $Q_0 \div Q_3$, Q_Δ зависят от реализаций Y_{si} вейвлет-коэффициентов Y_{jl} -двумерных моделей (4) и (5), то F -статистики (8) – (11) – это функции либо аргумента a (масштаб), либо b (смещение).

Результаты экспериментальных исследований

Для проверки эффективности математических моделей ковариационного анализа случайных зависимостей вейвлет-коэффициентов от значения параметров a и b (3) были проанализированы два класса вибросигналов (по 64 реализации в каждом), несущих информацию о состоянии топливной системы дизельных агрегатов Д80 тепловозов ТГМ4.

На рис. 1 представлены выходные вибросигналы пьезоэлектрического акселерометра АП-18, установленного на трубку топливного насоса высокого давления дизельного агрегата для двух состояний форсунки а) – рабочей (в нормальном состоянии) и б) – с неисправностью («ляющая» форсунка).

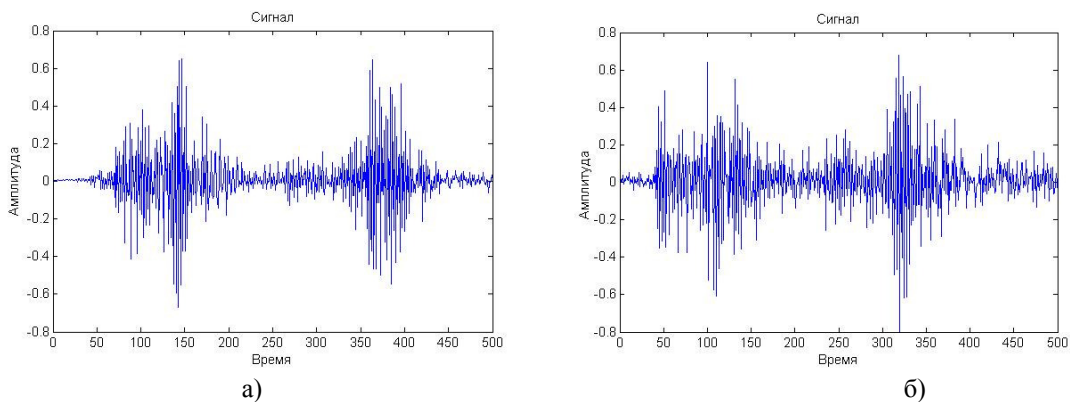


Рисунок 1 – Вибросигналы, соответствующие рабочему и неисправному состоянию форсунки дизельных агрегатов

Непрерывные вейвлет-преобразования этих сигналов, осуществленные в соответствии с математической моделью (1), представлены на рис. 2. Вариант а) соответствует рабочему состоянию, б) – с функциональным нарушением в работе форсунки. На рис. 3 представлены соответствующие варианты а) и б), которые являются двумерными проекционными моделями трехмерных вейвлет-изображений из рис. 2 а), б).

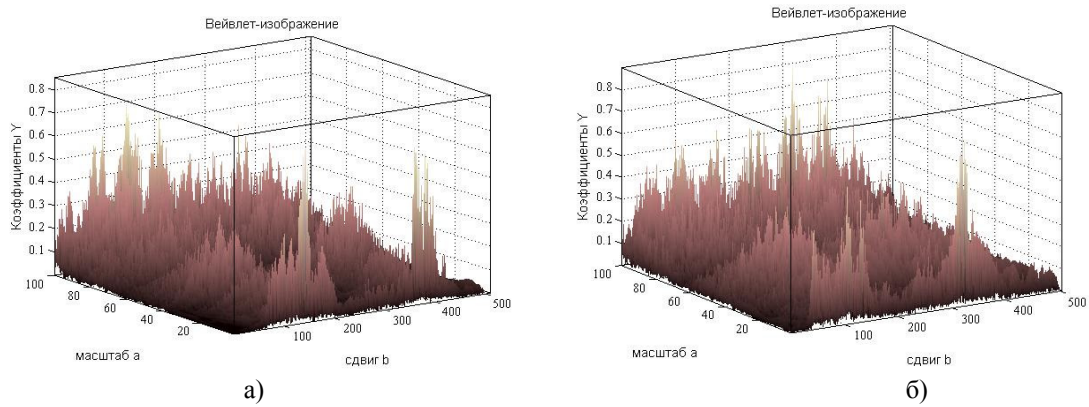


Рисунок 2 – Вейвлет-ізображення, відповідні робочому і несправному стану форсунки дизельних агрегатів

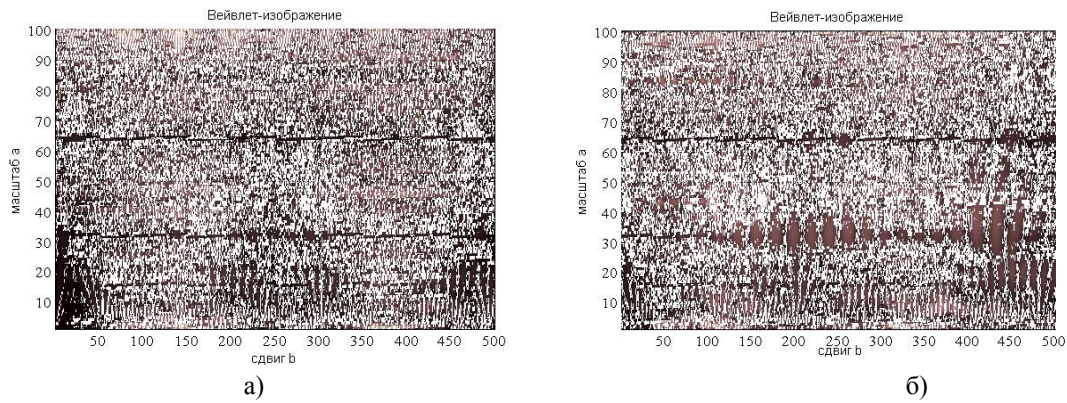


Рисунок 3 – Двумірні проєкційні моделі вейвлет-ізображень, відповідні робочому і несправному стану форсунки

На рис. 4–7 (варіанти а), б)) представлені F -статистики коваріаційного аналізу для реалізацій на рис. 2, розраховані за рівняннями (8) – (11), які несуть кількісну інформацію про складові дисперсійного розкладу (7). Ці рисунки наочно ілюструють можливість кількісного отримання вимірної інформації про стан функціонального стану об'єкта вібраційної діагностики. Варіанти а) і б) рис. 4–7 відрізняються не тільки максимальними значеннями F -статистик (в 2 – 4 рази), але й конфігурацією функціональних залежностей, відповідних F -статистик від значень масштабу a .

Аналогічні залежності отримані для F -статистик, залежних від значень зсуву b .

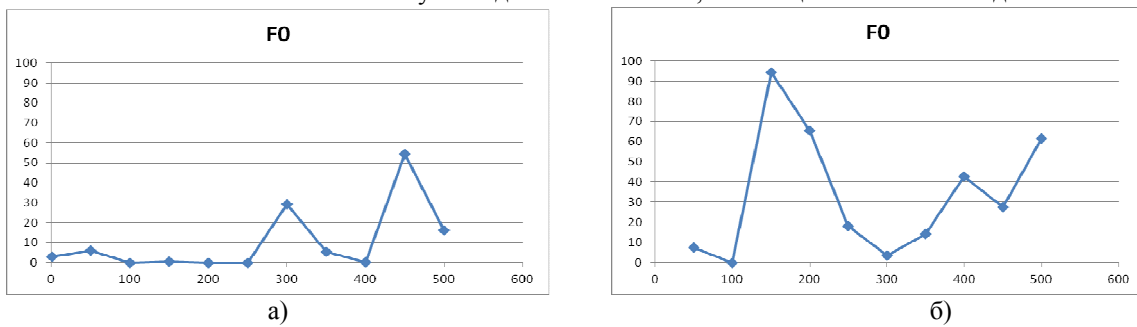


Рисунок 4 – Реалізація статистики F_0 для робочого і несправного стану форсунки дизельних агрегатів

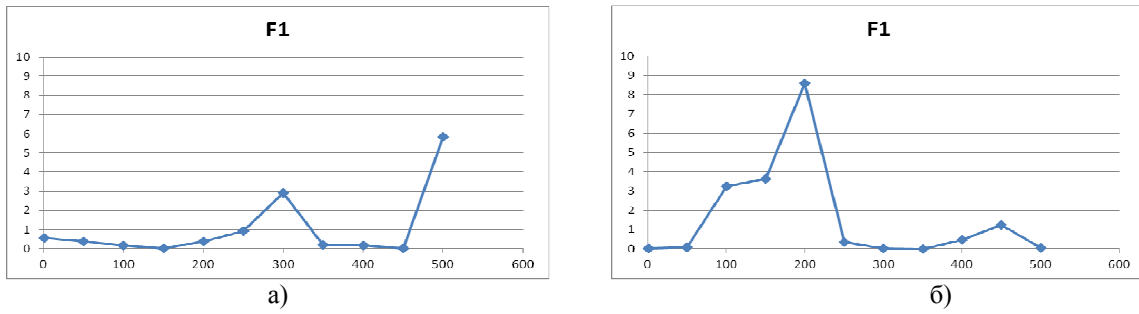


Рисунок 5 – Реализация статистики F_1 для рабочего и неисправного состояния форсунки дизельных агрегатов

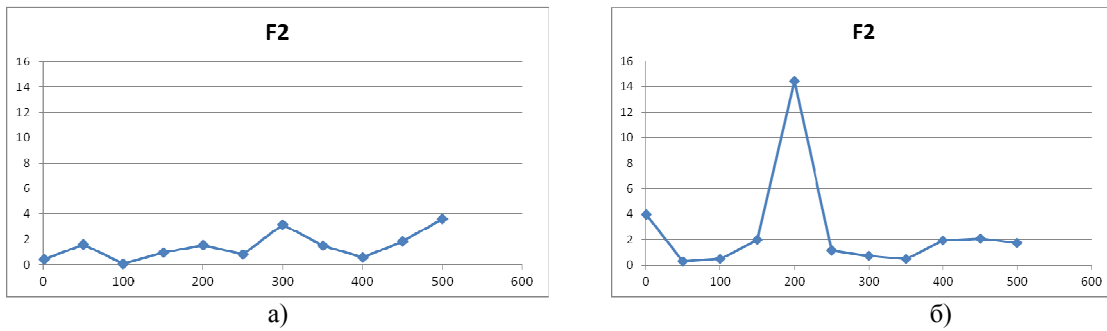


Рисунок 6 – Реализация статистики F_2 для рабочего и неисправного состояния форсунки дизельных агрегатов

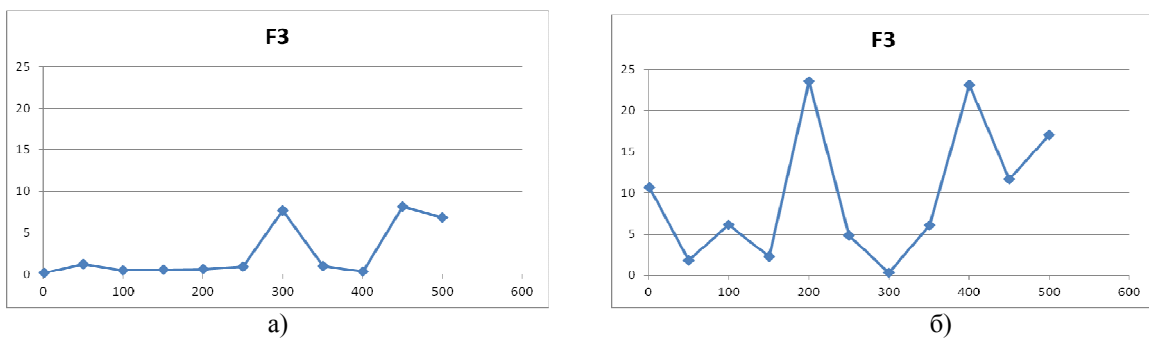


Рисунок 7 – Реализация статистики F_3 для рабочего и неисправного состояния форсунки дизельных агрегатов

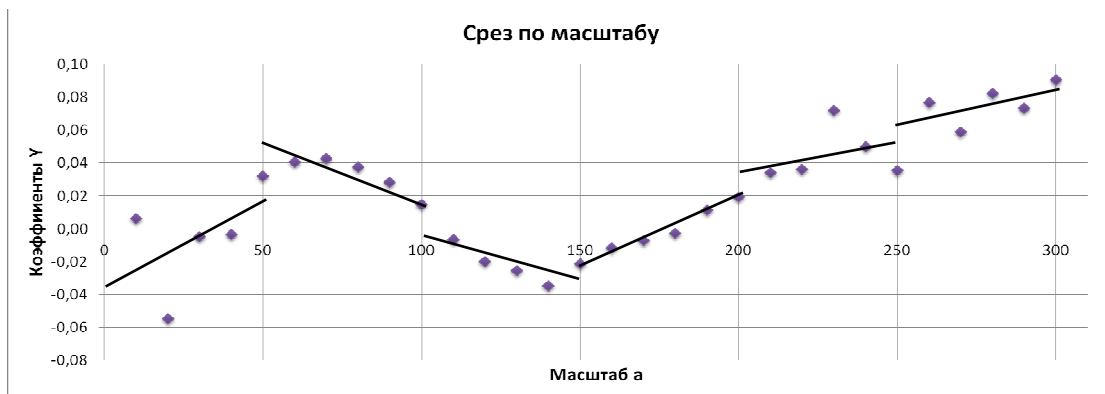


Рисунок 8 – Кусочно-линейная аппроксимация последовательности вейвлет-коэффициентов регрессии со случайными параметрами

На рис. 8 показан пример кусочно-линейной по математическому ожиданию аппроксимации последовательности вейвлет-коэффициентов регрессиями со случайными параметрами. Такие аппроксимации характерны для любого из сечений трехмерного вейвлет-изображения при ковариационном анализе последовательности вейвлет-коэффициентов.

Выводы

1. Доказана эффективность статистического подхода для анализа трехмерной модели вейвлет-преобразования спектрально-нестационарных вибрационных сигналов. Такой анализ позволяет уменьшить размерность вейвлет-изображения, выделив при этом четыре информативных критериальных F -статистик дисперсионных отношений, имеющих значимое функциональное различие по видам состояния объекта вибрационной диагностики.

2. Полученные F -статистики, являясь случайными величинами с известным законом распределения, позволяют оценить вероятности ошибок диагностики на базе стандартных параметрических моделей дискриминантного анализа.

3. Ковариационный анализ вейвлет-преобразований позволяет планировать измерения для локализованных во времени спектрально-нестационарных вибрационных сигналов.

Список литературы

- 1) Дремин И.М. Вейвлеты и их использование / И.М. Дремин, О.В. Иванов, В.А. Нечитайло // УФН. – 2001. – Т.171. – №5. – С.465-501.
- 2) Кветний Р.Н. Метод виділення контурів на основі вейвлет-перетворення з використанням двовимірних фільтрів // Інформаційні технології та комп'ютерна інженерія. – 2010. – №3. – С.26-34.
- 3) Meyer Y Wavelets: Algorithms and Applications (Philadelphia: SIAM, 1993).
- 4) Wavelets: An Elementary Treatment of Theory and Applications (Ed T Koornwinder) (Singapore: World Scientific, 1993).
- 5) Jun Li. A Wavelet Approach to Edge Detection: a thesis to The Department of Mathematics and Statistics in partial fulfillment of the requirements for the degree of Master of Science in the subject of Mathematics / Jun Li. – Huntsville, Texas, 2003. – 80 с.
- 6) Антошук С.Г. Иерархический подход при контурном анализе изображений / С.Г. Антошук, О.В. Бабилунга, А.А. Николенко, Е.В. Ткаченко // Штучний інтелект. – 2005. – №3. – С.603-609.
- 7) Jianjia Pan. Edge Detection Combining Wavelet Transform and Canny Operator Based on Fusion Rules / Jianjia Pan // International Conference on Wavelet Analysis and Pattern Recognition, Baoding. – 2009. – С.324-328.
- 8) Thurner S, Feurstein M C, Teich M C Phys. Rev. Lett. 80 1544 (1998).
- 9) Antonini M, Barlaud M, Mathieu P, Daubechies I IEEE Trans. On Image Process. 1 205 (1992).
- 10) G. Ososkov, A. Shitov. Gaussian Wavelet Features and their Applications for Analysis of Discretized Signals. — Computer Physics Communications, vol. 126 (2000), pp. 149–157.
- 11) Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке: Методы планирования эксперимента / Н. Джонсон, Ф.Лион; пер. с англ. под ред. Э.К. Лецкого. – М.: Мир, 1981. – 520 с.

Відомості про авторів

Щапов Павел Федорович, д.т.н., профессор кафедры ИИТС, НТУ «Харьковский политехнический институт», Харьков.

Мигущенко Руслан Павлович, к.т.н., доцент кафедры ИИТС, НТУ «Харьковский политехнический институт», Харьков.