

Реалізаційна модель одношарового перцептрона

¹Т.Б. Мартинюк, ²Л.М. Куперштейн, ¹А.Г. Буда

¹Вінницький національний технічний університет, Україна

²Вінницький фінансово-економічний університет, Україна

kuperok@mail.ru

Анотація

$$Y = f(S) = \text{sign}(S - \theta), \quad (2)$$

У статті розглядаються особливості реалізації моделі перцептрона з принципом обробки за методом різницевого зрізів. Запропоновано реалізацію моделі одношарового перцептрона на програмованих логічних інтегральних схемах фірми Xilinx у вигляді цифрового нейрочипу на базі конверсного процесора з обробкою за різницевими зрізами.

1. Вступ

За останні десятиліття інтерес до досліджень у напрямку нейронних технологій (штучних нейронних мереж (НМ)) значно зріс, що підтверджується великою кількістю публікацій. Це пояснюється насамперед тим, що застосування нейронної технології при її адекватній реалізації дозволяє істотно зменшити час розв'язання і, у деяких випадках, вирішити ті задачі, що дотепер вважалися нерозв'язуваними [1]. Серед основних класів задач, що розв'язуються за допомогою НМ, які лежать в основі усіх прикладних застосувань, виділяють такі: класифікація образів, кластеризація, прогнозування, оптимізація, асоціативна пам'ять, системи керування [2–4]. Найпершими, найпростішими, найпоширенішими НМ є перцептрони, які є базовими елементами для побудови практично усіх архітектур НМ [5, 6]. При цьому для забезпечення високої швидкодії, точності та надійності актуальним на сьогодні є саме їх апаратна реалізація, що підтверджується рядом розробок у цьому напрямку [7–11].

Метою даної роботи є аналіз та дослідження основних нейрооперацій одношарового перцептрону, а також його апаратна реалізація у вигляді цифрового нейрочипу.

2. Особливості одношарових перцептронів

Як відомо, один перцептрон може виконувати найпростіші процедури класифікації та розпізнавання [12, 6]. При цьому потужність нейрообчислень збільшується від з'єднання перцептронів в мережах. Найпростіша НМ – одношаровий перцептрон, який складається з групи перцептронів, що утворюють шар. Каскадуванням таких шарів отримують багатшарові мережі, в яких вихід одного шару є входом для наступного [6]. Саме таким НМ властиві великі обчислювальні можливості.

Проста модель перцептрона має трьохелементну структуру (помножувач, суматор, нелінійний перетворювач), в якій реалізується перемноження вектора вагових коефіцієнтів на вектор вхідних операндів з наступним підсумовуванням зважених входів та порівнянням отриманої суми із зовнішнім порогом обробки [5–7]. Математично ці операції можна записати так:

$$S = \sum_{i=1}^n w_i \cdot x_i, \quad (1)$$

де x_i – i -й вхідний сигнал; w_i – i -й ваговий коефіцієнт; S – суматорна функція; $f(S)$ – активаційна функція (нелінійне перетворення), Y – вихідний сигнал перцептрона, що реалізує найпростішу знакову функцію – функцію одиночного стрибка, θ – зовнішній поріг обробки. Така модель перцептрона реалізує простіший класифікатор, тобто є лінійним дискримінатором, оскільки вихідний сигнал Y формує значення так званої дискримінантної функції, яка залежить лише від лінійної комбінації входів [13]. Так, якщо вхідні вектори можуть належати одному з двох класів, перцептрон здатен розрізнити тип входу таким чином: $Y \geq 0$ – вхідний вектор належить першому класу, в протилежному випадку – другому. Простота такої моделі перцептрона не зменшує її цінності, оскільки на її основі ефективно розв'язуються задачі обробки двійкових даних, класифікації, розпізнавання, ідентифікації тощо [5, 6, 14–16].

Як видно із виразів (1) та (2) усі базові операції у класичній моделі перцептрона рознесені у часі і виконуються, зазвичай, різними блоками. При цьому, найбільш трудомістким і складним для апаратної реалізації з точки зору швидкодії є операції перемноження та підсумовування. Як відомо, обидві ці операції досить важко піддаються розпаралеленню. Крім того, не менш важливим при апаратній реалізації моделі перцептрона є аспект вибору елементної бази, оскільки вона також може вносити значні корективи у часові та апаратні витрати. Отже, для ефективної реалізації моделі перцептрона необхідно розглянути можливі перспективні варіанти реалізації операцій (1) і (2) та відображення їх на апаратній платформі.

3. Розпаралелене та суміщене виконання базових нейрооперацій

Серед способів розпаралелення операцій підсумовування уваги заслуговує метод паралельного підсумовування масиву чисел за різницевими зрізами (РЗ) [17]. Даний метод дозволяє реалізувати модель перцептрона, в якій виконується розпаралелена операція підсумовування з одночасним суміщенням її у часі з операцією порівняння із зовнішнім порогом обробки, на відміну від відомої моделі Розенблатта [6]. Це, в свою чергу, дозволяє значно зменшити час отримання вихідного значення перцептрона, оскільки перевищення суми над порогом можна отримати не формуючи кінцеву зважену суму. Детально спосіб порогової паралельної обробки за РЗ розглянуто і досліджено у ряді робіт [15, 18, 19]. Його суть полягає у тому, що в кожному циклі обробки виконується формування поточних часткових сум елементів векторного масиву, а також формування поточного порогу для наступного циклу обробки і формування результуючого сигналу. Такий підхід дозволяє прискорити процес обробки від 12% до 47% [19, 20]. Крім того, прискорення обробки за таким методом

також варіюється в залежності від кількості однакових вхідних даних перцептрона [15, 17]. Часові параметри процесу обробки за РЗ отримано в результаті комп'ютерного моделювання [20]. Звичайно, таке прискорення, що дає модель перцептрона на базі РЗ є досить важким аспектом при апаратному розв'язанні задач розпізнавання та класифікації, до яких зводяться більшість задач, що розв'язуються за допомогою штучних нейронних мереж.

Крім того, нейрооперації (1) та (2), що виконуються в перцептроні, доцільно розглядати у контексті використовуваної елементної бази, оскільки досить важко адаптувати різні паралельні алгоритми множення і підсумовування до певної технології з можливістю роботи всієї НМ у реальному часі.

4. Апаратна реалізація перцептрона

Описану вище модель перцептрона з обробкою РЗ реалізовано у вигляді конвеєрного процесора на базі систолічного масиву [21]. На рис. 1. наведено узагальнену структурну схему конвеєрного процесора, який виконує суміщене у часі підсумовування зважених вхідних операндів та порівняння із зовнішнім порогом обробки з формуванням вихідного сигналу виду (2).

Оброблення даних за РЗ дозволяє збільшити багатofункціональність конвеєрного процесора за

рахунок одночасного виконання разом із формуванням суми S елементів вхідного векторного масиву таких операцій: а) сортування за зростанням значень його елементів; б) визначення екстремальних елементів; в) порогове підсумовування елементів масиву з формуванням вихідного сигналу Y ; д) відновлення елементів вхідного масиву як зворотної операції [17, 22, 23]. Конвеєрний процесор (рис. 1) містить n комірок, причому i -а комірка містить арифметично логічний пристрій (АЛП) ALU1 для формування зрізів різниць, мініматор MIN для визначення мінімальної величини серед двох, які подаються на його входи. Мультиплексор MX1 і блок елементів I необхідні у кожній комірці для комутації входів, а регістри RGR, RGN – для тимчасового збереження інформації. Перший ярус містить АЛП ALU1, регістр RGR часткового результату і мультиплексор MX1. Остання n -а комірка апаратно значно складніша, оскільки саме у ній виконується нелінійне перетворення. Вона додатково містить регістри RGP, RGM, АЛП ALU2, мультиплексор MX2, елемент АБО OR, лічильники CT1, CT2, елемент I AND і блок NAND елементів I-НІ. Також конвеєрний процесор має вхід "Скид" встановлення в початковий стан процесора, n входів операндів, вхід "Адреса" керування пристроєм, вхід "Поріг θ " порогу обробки, вихід сигналу "Кінець" пристрою та вихід "Вихід Y " результату порогової обробки.

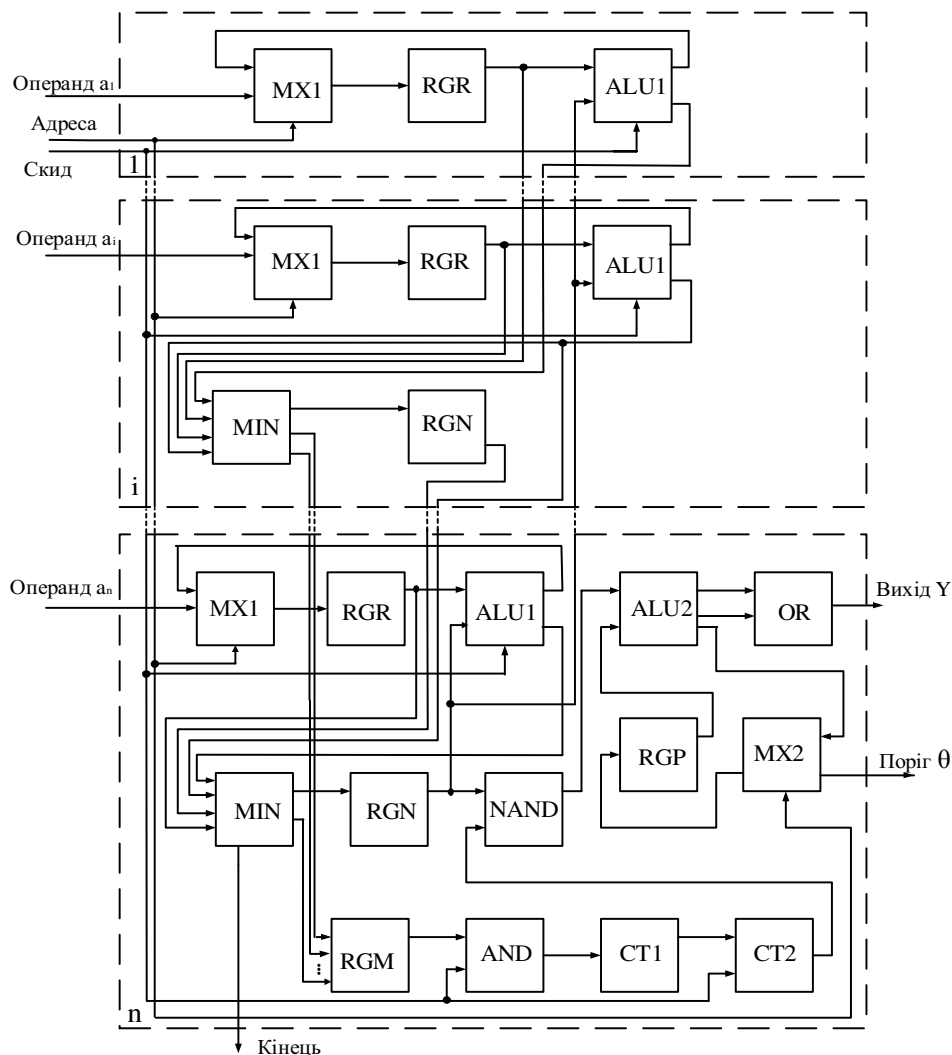


Рис. 1. Узагальнена структурна схема конвеєрного процесора

Завдяки „просторовій однорідності” операцій, які виконуються при обробленні векторних даних за РЗ [15, 24], конвеєрний процесор (рис. 1) використовує принцип розпаралелювання та конвеєризації обчислювального процесу.

Каскадування та об'єднання таких процесорів дає можливість отримання не тільки одношарового, але й багатшарового перцептрона, в якому відсутні функціональні обмеження, що накладаються на класичний перцептрон [6, 25].

Для апаратної реалізації моделі перцептрона найбільш доцільним та ефективним є формат нейрочипа у складі автономного нейрообчислювача. Аналіз засобів реалізації нейрочипів показав перспективність їх реалізації на базі програмованих логічних інтегральних схем (ПЛІС), що зумовлено високою інтеграцією, можливістю модифікації архітектури, значною кількістю зовнішніх виводів ПЛІС, розвинутою структурою швидкодіючих взаємозв'язків, підвищеною швидкістю обміну між кристалами. Все це у сукупності дозволяє створити на базі ПЛІС закінчений, досить регулярний фрагмент нейронної мережі, а потім, здійснюючи каскадування ПЛІС, будувати нейронні мережі довільної конфігурації і складності [7, 26]. За рахунок гнучкої реконфігурації ПЛІС можливе завантаження у кристали зовсім різних конфігурацій нейронних мереж, а, маючи розвинуту бібліотеку фрагментів мереж і універсальний апаратний комплекс [27] на одній програмно-апаратній платформі, можна швидко й ефективно вирішувати різноманітні задачі.

Запропоновану структуру конвеєрного процесора було спроектована й «розміщено» на ПЛІС XC95288XL-6-BG256 фірми Xilinx [19]. Результати моделювання показали, що при розміщенні процесора розмірності чотири 4-розрядних операнда використано 28% логічної ємності кристала ПЛІС, а саме 82 з 288 макрокомірок і 28 виводів з 192 можливих. Таким чином, на ПЛІС типу CPLD можна розмістити до чотирьох 4-вхідних нейронів. Використовуючи ПЛІС з більшою ємністю (наприклад, сімейства Spartan-3 або Virtex 4 [27]), можна отримати нейрочип, що містить не тільки адаптивний паралельний багатовхідний суматор, але й відповідну кількість помножувачів. Такі паралельні помножувачі з бібліотеки М-модулів використовуються при створенні універсальних і проблемно-орієнтованих нейроструктур на ПЛІС фірми Xilinx [28].

Разом з тим, реалізація операції множення залишається на сьогодні однією з найвимогливіших за апаратними та часовими витратами операцією. При цьому використання стандартних бібліотек є неоптимальним варіантом. Тому існує проблема реалізації ефективної закінченої нейроструктури на одному кристалі. Разом з тим, відомо результати експериментальної реалізації на VHDL методу швидкого множення на ПЛІС фірми Altera, яке при певній кількості та розрядності операндів дозволяє отримати вираш як за апаратними ресурсами ПЛІС, так і за швидкістю обчислень [29]. Щодо реалізації пам'яті вагових коефіцієнтів, то при їх сталому значенні (навчена нейромережа), їх також можна зберігати на ПЛІС, де розташовано й саму мережу, використовуючи внутрішню пам'ять кристала. У випадку реалізації нейронної мережі з навчанням у реальному часі або при значній кількості вагових коефіцієнтів потрібно використовувати зовнішній модуль оперативної пам'яті на базі ПЛІС [7].

Аналіз даних зі звітів імплементації нейрочипа на базі конвеєрного процесора на ПЛІС Xilinx в середовищі

Xilinx Foundation Series 2.1 показав, що час його спрацювання знаходиться у межах мікросекундного діапазону (0,23 мкс) [19]. Адаптувавши класичну методику оцінки продуктивності конвеєрних пристроїв [30, 31] до конвеєрної (систолічної) обробки за РЗ у запропонованому процесорі, визначено основні показники та коефіцієнти продуктивності реалізованого нейрочипа [32]. Для варіанту нейрочипу на 16 входів отримано розрахункове значення асимптотичної продуктивності рівне 2,15 млн.оп./с при такті роботи конвеєрного процесора в 466 нс, а коефіцієнт ефективності становить 0,056. При цьому продуктивність конвеєрного процесора з урахуванням кількості вхідних операндів (16) становить 1,9 млн.оп./с, а коефіцієнт завантаження конвеєра – 0,89. Також продуктивність конвеєрного процесора з урахуванням витрат часу запису у регістр, часу розповсюдження сигналу у міжз'єднаннях, часом зсуву синхроімпульсів оцінено у 1,04 млн.оп./с, а коефіцієнт використання конвеєрного процесора при цьому становить 0,485. Така обчислювальна потужність навіть при незначному розмірі вхідного вектора операндів може забезпечити виконання обробки нейрочипом на базі конвеєрного процесора у режимі близькому до РЧ. Для більш ефективної роботи конвеєрних процесорів на базі РЗ необхідно збільшити кількість оброблюваних операндів, що цілком можливо при використанні чипів ПЛІС більш великої логічної ємності, при цьому продуктивність обробки можна збільшити за рахунок більш швидкодіючих чипів ПЛІС.

Запропоновану розробку нейрочипа з одношаровим перцептроном можна використовувати як фрагмент багатшарової нейромережі, розміщеної на програмно-апаратній платформі з використанням інструментальних модулів цифрової обробки сигналів на ПЛІС XDSP-680 або XDPS-4М фірми Xilinx [7, 26].

5. Висновки

Досліджено особливості реалізації моделі перцептрона, принцип функціонування якого базується на методі РЗ. Використання цього методу дозволяє не тільки розпаралелити процес обробки, але й сумістити його з поетапним порівнянням із зовнішнім порогом, тобто апаратно реалізувати в одному функціональному блоці багатовхідний суматор та нелінійний перетворювач (з пороговою функцією активації).

Відображено модель перцептрона з обробкою за РЗ у вигляді конвеєрного процесора, який дає можливість сформувати вихідний сигнал не формуючи остаточної суми зважених вхідних операндів у разі перевищення її над зовнішнім порогом обробки.

Реалізація моделі перцептрона на ПЛІС фірми Xilinx на базі конвеєрного процесора з обробкою за РЗ дозволяє отримати нейрочип з досить регулярним фрагментом мережі і часом спрацювання у мікросекундному діапазоні, а отже і з можливістю каскадування ПЛІС для побудови нейронної мережі довільної конфігурації.

Теоретично оцінено продуктивність змодельованих нейрочипів на базі запропонованих конвеєрних процесорів при обробці 16 операндів, яка складає до 2 млн.оп./с. Коефіцієнт завантаженості конвеєра становить близько 50% для поліпшення чого необхідно збільшити кількість оброблюваних операндів, що цілком можливо при використанні чипів ПЛІС великої логічної ємності. Проте використання конвеєрних процесорів вже при такій обчислювальній потужності може забезпечити їх

практичне використання у різних прикладних нейросистемах.

6. Література

- [1] Бондарь Т.А., Логовский А.С. Нейрокомпьютеры в управлении // Зарубежная радиоэлектроника. – 1997. - №2. - С. 57-71.
- [2] От биологических нейронных сетей к искусственным нейронным сетям // Нейронные сети и нейрокомпьютеры. – Режим доступа: <http://21.net78.net/lesson.php?glava=1&print>
- [3] Комашинский В.И., Смирнов Д.А. Нейронные сети в системах управления и связи. – М.: Горячая линия – Телеком, 2003. – 94 с.
- [4] Галушкин А.И. О современных направлениях развития нейрокомпьютеров // Информационные технологии. - 1997. - № 5. – С. 2-5.
- [5] Каллан Р. Основные концепции нейронных сетей.: Пер. с англ. – М.: Издательский дом “Вильямс”, 2001. – 288 с.
- [6] Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. Пер. с англ.– М.: Мир, 1992. – 240 с.
- [7] Галушкин А.И. Нейрокомпьютеры. Кн. 3: Учеб. пособие для вузов / Общ. ред. А.И. Галушкина. – М.: ИПРЖР, 2000. – 528 с.
- [8] Сучасні напрямки розвитку нейрокомп'ютерних технологій // Організація інтелектуальних обчислень. Курс. Лекцій. – Режим доступу: <http://www.victoria.lviv.ua/html/oio/html/theme9.htm>.
- [9] Нейрочип впервые будет полностью управлять нервной клеткой // Nanonewsnetю. – Режим доступа: <http://www.nanonewsnet.ru/articles/2007/neirochip-vpervye-budet-polnostyu-upravlyat-nervnoi-kletkoi>
- [10] Куперштейн Л.М. До проблеми виявлення шахрайства у телекомунікаційних мережах на базі інтелектуальних технологій // Вісник Вінницького політехнічного інституту. – 2007. - №5. С. 78-84.
- [11] Васюра А.С., Куперштейн Л.М. Гібридна нейроподібна система на базі голографічного диска з різницево-зрізовою обробкою даних // Тези IV МНТК «PHOTONICS-ODS 2008». – Вінниця. – 2008. – С.98-99.
- [12] Розенблатт Ф., Принципы нейродинамики: Перцептроны и теория механизмов. – М.: Мир, 1965. – 480 с.
- [13] Сжов А.А., Шумский С.А. Нейрокомпьютинг и его применения в экономике и би знесе. – М.: Наука. – 1998. – 222 с.
- [14] Круглов В.В. Борисов В.В. Искусственные нейронные сети. Теория и практика. – М.: Горячая линия – Телеком, 2002. – 382 с.
- [15] Мартынюк Т.Б. Модель порогового нейрона на основе параллельной обработки по разностным срезам // Кибернетика и системный анализ. – 2005. - №4. – С. 78-89.
- [16] Архангельский В.И., Богаенко И.Н., Грабовский Г.Г., Рюмшин Н.А. Нейронные сети в системах автоматизации. - Киев: Техника,1999. - 364 с.
- [17] Мартынюк Т.Б. Рекурсивні алгоритми багатоперандної обробки інформації. Монографія. – Вінниця: УНІВЕРСУМ – Вінниця, 2000. – 216 с.
- [18] Мартынюк Т.Б., Куперштейн Л.М. Різницево-зрізова обробка як одна з альтернатив нейроподібної обробки даних // Праці 9 МНПК “Оброблення сигналів і зображень та розпізнавання образів (УкрОБРАЗ’2008)”. – Київ: Міжнародний науково-
- навчальний центр інформаційних технологій та систем, 2008. – С. 31-34.
- [19] Васюра А.С., Мартынюк Т.Б., Куперштейн Л.М. Методи та засоби нейроподібної обробки даних для систем керування. Монографія. – Вінниця: УНІВЕРСУМ – Вінниця, 2008. – 175 с.
- [20] Мартынюк Т.Б., Хом’юк В.В., Куперштейн Л.М., Матвеев С.С. Аналіз моделей паралельного підсумовування елементів числового масиву // Вісник Вінницького політехнічного інституту. - 2002. – № 6. – С. 51-56.
- [21] Патент № 67652А, МПК 7 G06G7/14. Конвейерний пристрій / А.С. Васюра, В.Г. Лисогор, Т.Б. Мартынюк, Л.М. Куперштейн, М.В. Дзись. - № 2003119839; Заявлено 03.11.2003; Опубл. 15.06.2004, Бюл. № 6. - 10с.
- [22] Мартынюк Т. Буда А., Хом’юк В., Куперштейн Л. Багатофункціональна регулярна структура для обробки векторних даних // Оброблення сигналів і зображень та розпізнавання образів (“УкрОБРАЗ’2004”). – Київ: Міжнародний науково-навчальний центр інформаційних технологій та систем, 2004. - С. 33-36.
- [23] Мартынюк Т.Б. Организация конвейерного процессора для порогового суммирования элементов векторного массива / Т.Б. Мартынюк, А.Г. Буда, Л.М. Куперштейн, Р.А. Расенко, Д.И. Климкина // “ПРИБОРОСТРОЕНИЕ-2004”: Междунар. науч.-техн. конф., 18-21 сент. 2004 г.: сб. тр. – Винница-Ялта, 2004. - С. 40-44.
- [24] Кожемяко В.П., Мартынюк Т.Б., Куперштейн Л.М. Модель „быстрого нейрона” с обработкой данных по принципу разностных срезов // Оптико-електронні інформаційно-енергетичні технології. – 2009, №2. С. 102-111.
- [25] Минский М., Пейперт С. Перцептроны.: Пер. с англ. – М.:Мир, 1971. – 365 с.
- [26] Логовский А. Технология ПЛИС и ее применение для создания нейрочипов // Открытые системы. – 2000. – №10. – Режим доступа: http://www.osp.ru/os/2000/10/019_print.htm.
- [27] Капитанов В.Д., Мистюков В.Г. Построение на ПЛИС фирмы Xilinx высокопроизводительных нейронных сетей // Scan Engineering Telecom. – 1999. – Режим доступа: http://www.scan.com/art_neur.pdf.
- [28] Грушвицкий Р.И., Мурсаев А.Х., Угрюмов Е.П. Проектирование систем на микросхемах программируемой логики. – СПб.: БХВ – Петербург, 2002. – 608 с.
- [29] Мо Джо Чо. Особенности реализации операции умножения на ПЛИС // Современные наукоемкие технологи. – 2008. – №4. – С. 114-116.
- [30] Вишенчук И.М., Черкасский Н.В. Алгоритмические операционные устройства и суперЭВМ. – К.: Техника, 1990. – 197 с.
- [31] Феселов В.Ф. Высокоскоростные вычисления: Архитектура, производительность, прикладные алгоритмы и программы для суперЭВМ. – М.: Радио и связь, 1988. – 432 с.
- [32] Мартынюк Т.Б., Куперштейн Л.М. Ефективність конвейерного процесора з різницево-зрізовим обробленням даних // Вісник Вінницького політехнічного університету. – 2008. – №5. – С. 69-77.