

УДК 004.032.26

Т. Б. Мартинюк, к. т. н., доц.;

А. В. Кожем'яко, к. т. н., доц. ;

Н. В. Фофанова, асп.

ОСОБЛИВОСТІ АПАРАТНОГО МОДЕЛЮВАННЯ ФУНКЦІЙ НЕЙРОНА

Розглянуто особливості використання оброблення векторних масивів даних за різницеви-ми зрізами, які можна розглядати як базис нейрооперацій формального нейрона. Проаналізовано можливості реалізації запропонованої структурної організації адаптивного багатовхідного суматора на ПЛІС.

Вступ

Розвиток сучасної елементної бази, зокрема, оптоелектронних ІС на базі смарт-пікселів [1] та програмованих логічних ІС (ПЛІС) [2, 3], дозволяє у практичному впровадженні нейротехнологій перейти від нейроемулаторів до нейроприскорювачів, тобто від програмного моделювання нейромереж (НМ) до їх апаратної реалізації у вигляді нейрочипів і нейрокомп'ютерів [4]. При цьому нейроприскорювачі виконуються як на базі апаратної реалізації типових нейрооперацій у нейромережному логічному базисі, так і на базі суміщеної апаратної реалізації зваженого підсумовування й програмної реалізації нелінійного перетворення [5]. Це пов'язано з тим, що на теперішній час існує велике розмаїття передатних функцій (функцій активації), які використовуються в якості базису нелінійного перетворення, тобто як одна з нейрооперацій формального нейрона (ФН) [4–6].

Метою роботи є розширення функціональних можливостей ФН з апаратною реалізацією його типових нейрооперацій на прикладі нейрона з пороговою функцією активації.

Постановка задачі

Серед відомих і широко застосовуваних функцій активації ФН можна відмітити такі: порогову або одиначного стрибка, лінійну з насиченням і сигмоїдну [6]. Сфера їх застосування охоплює НМ без навчання на базі ФН з пороговою та лінійною функціями активації та НМ з навчанням на базі ФН з переважно сигмоїдною функцією активації. На теперішній час найперспективнішими вважаються моделі так званих «швидких нейронів», які мають швидкісні характеристики за рахунок апаратної реалізації нейрооперацій, до яких відносяться операції зваженого підсумовування (1) і нелінійного перетворення (2):

$$s = \sum_{i=1}^n w_i \cdot x_i + b; \quad (1)$$

$$Y = f(s), \quad (2)$$

де w_i — вага i -го синапса (weight); b — зміщення (bias); s — стан нейрона (sum); x_i — i -й компонент вхідного вектора; Y — вихідний сигнал нейрона; f — нелінійне перетворення (функція активації).

Якщо для прискорення апаратного виконання нейрооперації зваженого підсумовування вигляду (1) можна застосувати, наприклад, організацію двовхідних суматорів у вигляді «дерева» [4], то для нелінійного перетворення запропоновано декілька варіантів. Наприклад, для апаратної реалізації логістичної або сигмоїдної функції можна застосувати розрядно-паралельні алгоритми Волдера, Пухова і Меджита, що призначені для обчислення елементарних функцій [7]. Пристрій для формування такої функції містить паралельні суматори, ПЗП та логічні елементи і може бути використаний для побудови швидкодіючих нейрочипів.

Іншими прикладами апаратної реалізації нейрооперацій нелінійного перетворення є застосування асоціативної нейромережі СМАС (Cerebellar Model Articulation Controller — мозочкова модель суглобного регулятора) [8]. СМАС призначена для запам'ятовування, відновлення та інтерп-

ретації функцій багатьох змінних. В цій нейромережі нелінійне перетворення аргументів функцій виконується умовно за допомогою алгоритму обчислення адреси комірок асоціативної пам'яті, в якій зберігаються числа, що визначають значення функції.

У роботі [9] наведено приклади навчання цієї мережі, які дозволяють їй формувати лінійну функцію і декілька варіантів ступінчастої функції. Алгоритм навчання СМАС являє собою ітеративну процедуру розв'язання системи лінійних алгебраїчних рівнянь. Недоліком такого підходу є значний час навчання НМ, значний об'єм пам'яті та наявність неточності відтворення необхідної функції.

Серед наведених функцій активації ФН інтерес до порогової функції обумовлений не в останню чергу її широким застосуванням в процесі моделювання біологічного нейрона [10]. Так, за теорією Хебба, кожний нейрон як нервова клітина мозку виконує просторово-часову сумачію вхідних сигналів, які надходять від збуджених нейронів, і, таким чином, визначає потенціал на своїй мембрані. У випадку, коли потенціал на мембрані перевищить порогове значення, нейрон збуджується [11].

Отже, базовим вузлом в апаратному моделюванні функції біологічного нейрона є просторово-часовий суматор, який формує сумарний сигнал множини вхідних сигналів, що подаються на його вхід одночасно. Таким чином, для апаратної реалізації пристрою для моделювання нейрона необхідно розв'язати такі задачі:

а) апаратно реалізовувати виконання двох важливих нейрооперацій: зваженого підсумовування і нелінійного перетворення;

б) забезпечити паралельне виконання операції підсумовування (згортки) зважених елементів вхідного вектора;

в) для прискорення процесу сумістити виконання двох нейрооперацій (1) і (2).

Одним з варіантів вирішення трьох наведених задач є використання нового підходу під час оброблення векторних масивів даних, а саме, оброблення на базі формування різницевих зрізів (РЗ) [12].

Оброблення за різницевиими зрізами як базис нейрооперацій

В основі оброблення векторних даних за РЗ лежить метод паралельного додавання тривалостей часових інтервалів [13], який замість традиційного підсумовування величин виконує накопичення кратних мінімальній складовій всіх доданків на кожному кроці обробки. Доведено універсальність цього методу: він може бути застосований для оброблення як аналогових, так і дискретних величин із поданням у цифровому коді [12]. Отже, замість фіксованого часу обробки, що характерно для традиційного підсумовування, де у найкращому випадку цей час дорівнює $O(\log_2 n)$, де n — кількість чисел у вхідному масиві, для алгоритму з використанням РЗ середній час можна визначити як випадкову величину з верхньою межею $O(n)$, яка залежить не тільки від розмірності n масиву чисел, але й від їх розподілу у масиві [12, 13]. Крім того, у роботі [12] показано, що застосування обробки за РЗ дозволяє сумістити виконання нейрооперацій (1) і (2) за умови порогової функції активації.

Послідовність операцій в процесі обробки за РЗ подано на рис. 1, де початковий векторний масив чисел розглядається як різницевий зріз A_0 , що містить ненульові елементи a_{i0} , тобто $A_0 \not\subset 0$ [15, 19]. Подальші РЗ A_j , ($j = 1, N$, де N — кількість циклів обробки) формуються за таким правилом:

$$A_j = \{a_{i,j-1} - q_j\}_{i=1}^n, \quad j = 1, N, \quad (3)$$

де q_j — внутрішній поріг обробки, який попередньо у кожному j -му циклі визначається таким чином:

$$q_j = \min \{a_{i,j-1}\}_{i=1}^n. \quad (4)$$

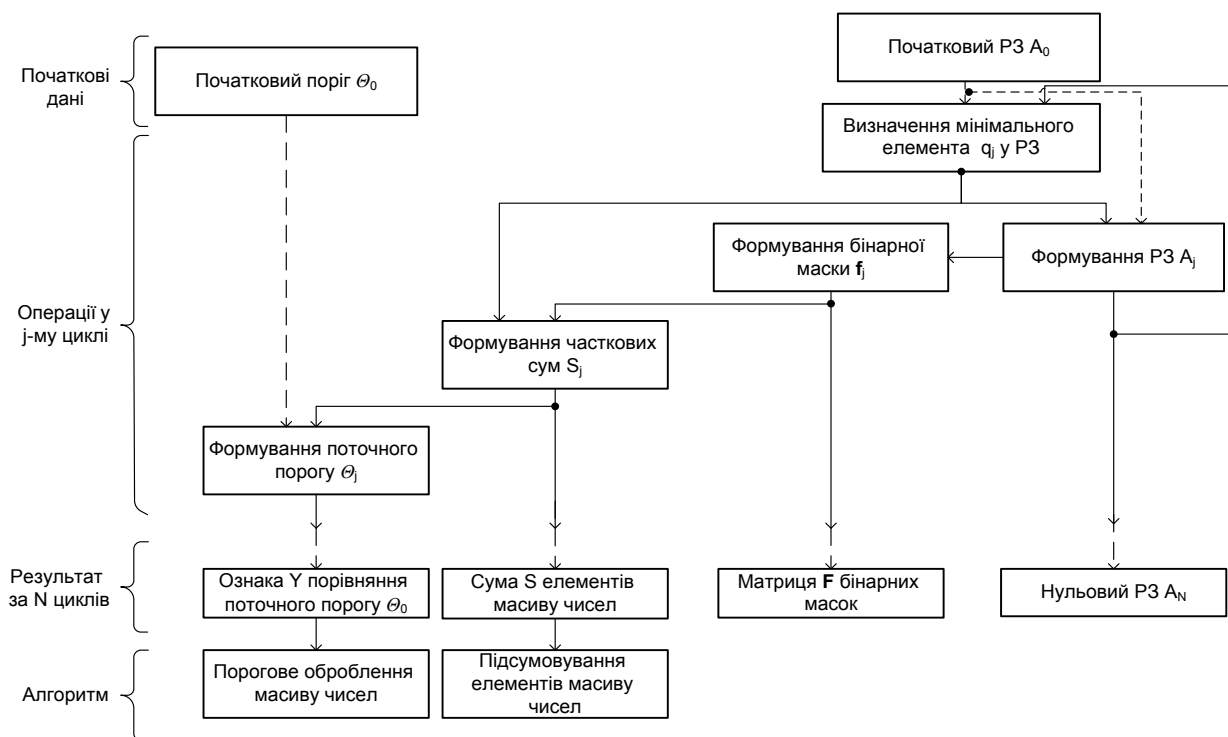


Рис. 1. Обробка за PЗ

Одночасно з обчисленням елементів наступного PЗ A_j у j -му циклі формуються відповідні елементи $f_{i,j}$ матриці F бінарних масок у вигляді

$$f_{i,j} = \begin{cases} 1, & \text{якщо } a_{i,j} \geq 0; \\ 0, & \text{якщо } a_{i,j} < 0. \end{cases} \quad (5)$$

Формування відповідних стовпців — бінарних масок f_i матриці F у кожному j -му циклі дозволяє визначити часткову суму S_j елементів PЗ A_{j-1} за такою формулою [12]:

$$S_j = q_j \sum_{i=1}^n f_{i,j}. \quad (6)$$

Накопичення S_j в подальшому забезпечує отримання остаточної суми за виразом [12, 13]

$$S = \sum_{i=1}^n a_{i,0} = \sum_{j=1}^N S_j = \sum_{j=1}^N q_j \left(\sum_{i=1}^n f_{i,j} \right). \quad (7)$$

У вигляді векторно-матричного перемноження співвідношення (7) можна записати так

$$S = (\mathbf{n}^T \cdot \mathbf{F}) \mathbf{q}, \quad (8)$$

де \mathbf{n} — одиничний вектор, T — символ транспонування; \mathbf{q} — вектор-стовпець вигляду

$$\mathbf{q} = \{q_j\}_{j=1}^N. \quad (9)$$

Виходом з процесу обробки є виконання умови

$$q_j = 0. \quad (10)$$

Таким чином, виконується інтегральна операція багатооперандного підсумовування або згортки елементів векторного масиву чисел [14]. Якщо розглядати елементи початкового PЗ A_0 як зважені вхідні сигнали, тобто за умови, що

$$a_{i,0} = w_i \cdot x_i, \quad i = \overline{1, n}, \quad (11)$$

то маємо спосіб реалізації зваженого підсумовування, тобто нейрооперації (1).

Особливістю обробки за РЗ є те, що сформовані часткові суми S_j (6) можна використати для порівняння у кожному j -му циклі з поточним зовнішнім порогом θ_{j-1} таким чином:

$$\theta_j = \theta_{j-1} - S_j, \quad j = \overline{1, N}, \quad (12)$$

де θ_0 — початкове значення зовнішнього порогу.

Вираз (12) є слухним за умови порогової функції активації (функції одиничного стрибка) [6], тобто для $\theta_0 = \text{const}$. В цьому випадку не обов'язково формувати остаточну суму S вигляду (7), оскільки процес порогової обробки завершується за виконання умови

$$\Delta_j \leq 0 \quad (13)$$

з формуванням одиничного значення вихідного сигналу Y ФН.

Таким чином, у реалізації порогового ФН вигляду

$$Y = \begin{cases} 1, & \text{якщо } S = \sum_{i=1}^n a_{i,0} \geq \theta_0; \\ 0 & \text{у протилежному випадку,} \end{cases} \quad (14)$$

можливе суміщення двох нейрооперацій (1) і (2), які виконуються на багатовхідному суматорі з використанням паралельної обробки за РЗ [12, 14]. При цьому час формування одиничного вихідного сигналу Y має випадковий характер, оскільки залежить не тільки від розмірності n вхідного масиву і розподілу його елементів, але й від величини зовнішнього порогу θ_0 . Максимальна кількість циклів обробки у самому несприятливому випадку становить $N = n$. Завдяки цьому у такому багатовхідному суматорі можна зберегти можливість формування остаточної суми S (7).

Паралельний адаптивний суматор

Отже, багатовхідний суматор з обробкою за РЗ має регулярну структуру і складається з n обчислювальних комірок, в яких паралельно обробляються елементи $a_{i,0}$ вхідного векторного масиву [15], а також має додаткову ($n + 1$) — у комірку для порівняння часткових сум S з поточним порогом θ_j і формування вихідного сигналу Y (рис. 2).

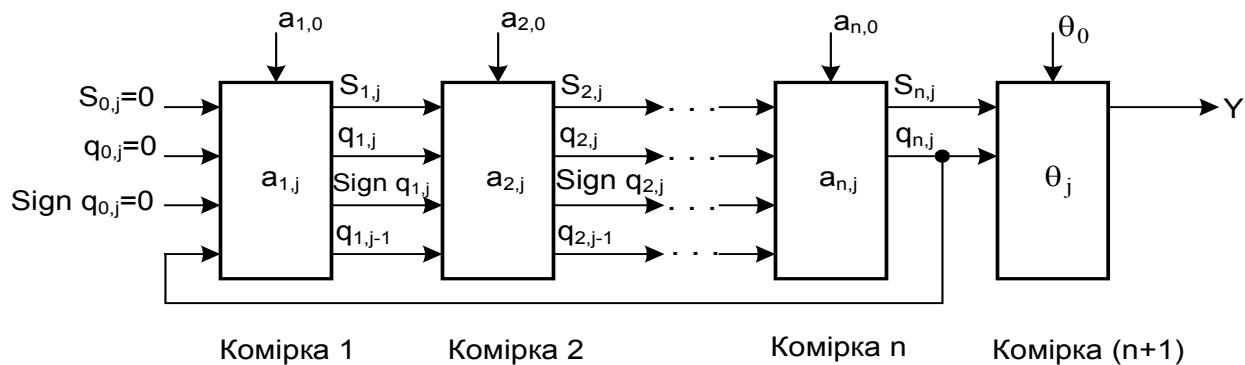


Рис. 2. Багатовхідний суматор

Між комірками передаються значення знаків часткових внутрішніх порогів $\text{sign } q_{ij}$, сформованих у попередніх комірках, а також внутрішній поріг q_{j-1} , сформований у попередньому ($j - 1$)-му циклі. Враховуючи рекурсивний характер базових співвідношень (4), (7), (12) для отримання детальної організації структури багатовхідного суматора в роботах [16] було обґрунтовано і реалізовано відображення регулярних ітераційних алгоритмів наведених базових операцій на одновимірний систолічний масив за відомою методикою Куна [17]. В результаті обробка векторного масиву чисел за РЗ являє собою конвеєрний процес [15].

Таким чином, запропонований багатовхідний суматор з архітектурою лінійного систолічного обчислювача і конвеєрним принципом обробки може бути використаний в якості цифрового просторово-часового суматора у складі пристроїв для моделювання нейронів з пороговою функцією активації. При цьому він може як адаптивний суматор працювати в трьох режимах:

- формального нейрона вигляду (14);
- формування скалярного добутку векторів вигляду

$$S = \sum_{i=1}^n a_{i,0} = \sum_{i=1}^n w_i \cdot x_i; \quad (15)$$

в) цифрового інтегратора вигляду

$$S = \sum_{i=1}^n x_i. \quad (16)$$

На рис. 3 показано схему пристрою для моделювання нейрона [18], що базується на адаптивному суматорі із запропонованою структурою, який працює в чотирьох режимах: крім трьох наведених вище ще й у режимі нейрона з лінійною функцією активації вигляду

$$Y = \begin{cases} S, & \text{якщо } \theta_0 - S \leq 0; \\ 0 & \text{у протилежному випадку.} \end{cases} \quad (17)$$

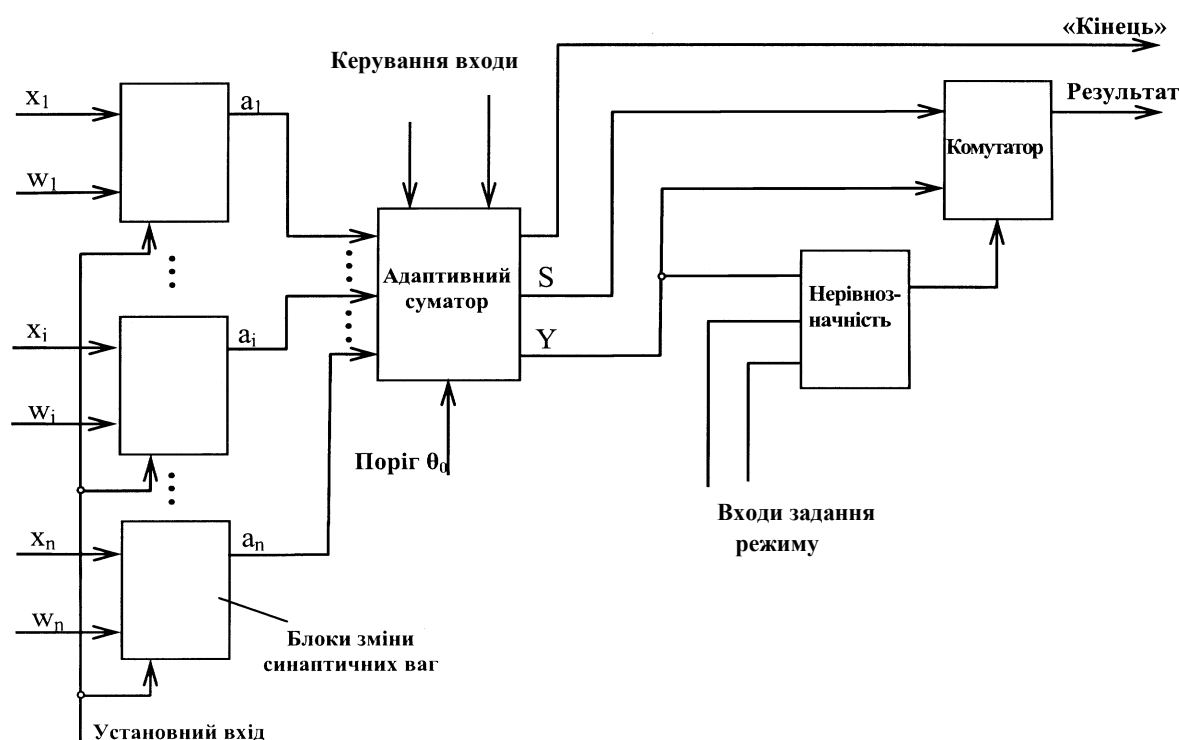


Рис. 3. Пристрій для моделювання нейрона

Отже, практично показано багатофункціональність адаптивного паралельного суматора з обробкою векторних даних за РЗ саме з апаратною реалізацією ФН як з пороговою, так і з лінійною вигляду (17) функціями активації. Такий суматор можна реалізувати у вигляді спеціалізованої прискорювальної плати, що використовується для моделювання нейронів та нейромереж.

Нейрообчислювач на ПЛІС

Серед апаратних нейроприскорювачів найперспективнішими вважаються ті, що реалізовані на базі ПЛІС, оскільки вони являють собою гнучкі нейрообчислювальні системи для науково-дослідних напрямів та мілкосерійного виробництва [4, 5]. Не в останню чергу це обумовлено можливістю розміщення на одному кристалі ПЛІС завершеного регулярного фрагмента нейромережі, на базі якого шляхом каскадування ПЛІС можна будувати нейромережі довільної конфігурації, складності структури і функціонального призначення [4].

Оскільки базовим будівельним елементом будь-якої нейромережі є саме нейрон, то доцільно зупинитися на прикладі реалізації на ПЛІС найпростішого нейрона, тобто нейрона з пороговою функцією активації, коли поріг є завантажувальною величиною.

В якості типової структури нейромережі з апаратною реалізацією нейрооперацій вигляду (1) і

(2) можна розглядати структуру паралельного програмованого обчислювача [4, 5], який використовується як апаратний прискорювач для керуючого РС (host-машини). Обчислювач містить (рис. 4): схему керування (СК), базові обчислювальні елементи (БОЕ1-БОЕ6), контролер зовнішньої шини (контролер E-bus) і контролер системної шини (контролер VME), два масиви статичної пам'яті (ОЗП0, ОЗП1) і блок швидкісних приймачів/передавачів (БШПП).

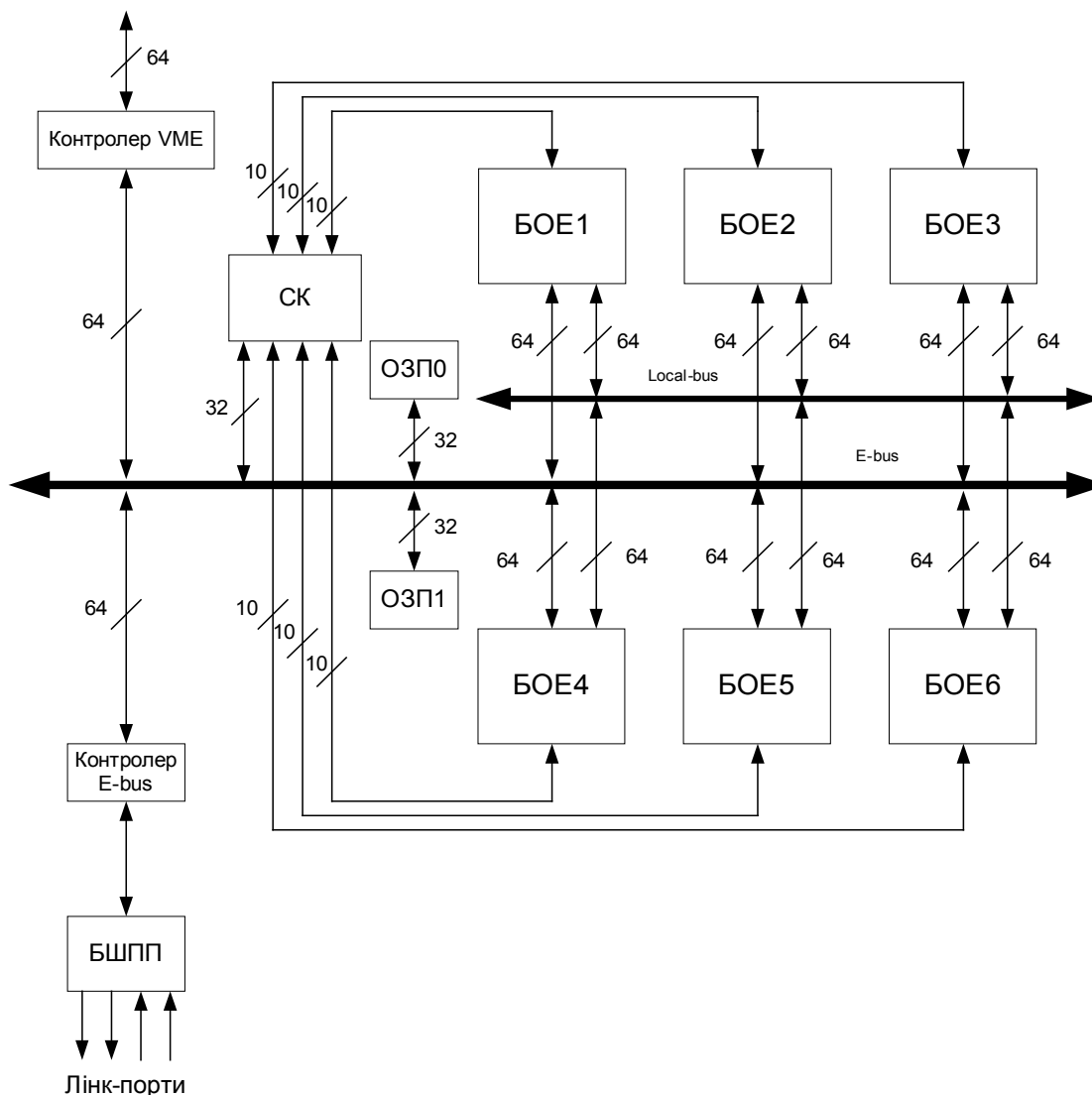


Рис. 4. Нейрообчислювач на ПЛІС

У відомій схемі обчислювачі БОЕ реалізуються на ПЛІС сімейства 10K фірми Altera і призначені для виконання простіших арифметичних операцій типу підсумовування, віднімання, множення та обчислення порогових функцій [4, 5]. Отже, запропонований адаптивний суматор, доповнений помножувачами для формування зважених вхідних сигналів і реалізований на одному кристалі ПЛІС дозволяє отримати нейрочіп, в якому реалізовано ФН з пороговою і лінійною функціями активації та з можливістю налаштування на один з чотирьох можливих режимів роботи [18].

Внутрішні режими БОЕ можуть бути використані для збереження значень вхідних коефіцієнтів, порогів та результатів обчислень, а пам'ять обчислювача призначена для збереження загальних коефіцієнтів, а також проміжних і остаточних результатів обчислень. Показану на рис. 4 структуру обчислювача можна розглядати як RISC-процесор (схема керування) і шість векторних процесів (БОЕ) з SIMD-архітектурою (одна команда для потоку даних) [4, 5].

Структура і набір команд RISC-процесора визначаються типом задач. Контролер шини VME є стандартним пристроєм, який реалізує інтерфейс з центральним процесором, зв'язок БОЕ між собою здійснюється за допомогою послідовного каналу приймачів/передавачів HOTlink фірми CYPRESS, а контролер зовнішньої шини керує передачею даних і містить набір чотирьох стандар-

тних FIFO та реєстрів керування і даних [4, 5]. Перевагами даного обчислювача є паралелізм архітектури, значна швидкодія через апаратну реалізацію арифметично-логічних операцій та можливість гнучкого переналаштування структури БОЕ, що розширює сферу його застосування.

Результати імплементації адаптивного суматора на базі ПЛІС фірми Altera показали, що в одній мікросхемі ПЛІС EP1K100FC256-3 можна розмістити до 60 комірок такого суматора. Крім того, час спрацювання адаптивного суматора з кількістю інформаційних входів до 60 не виходить за межі мілісекундного діапазону. З іншого боку, зі зменшенням кількості входів адаптивного суматора або за умови використання ПЛІС більшої ємності можна розмістити у мікросхемі ПЛІС не один ФН, а декілька, які можна розглядати як фрагмент шару нейромережі.

Висновки

1. Найбільша потреба у нейронах є найпростішою пороговою функцією активації як штучного аналога біологічного нейрона нервової системи людини виникає під час його апаратного моделювання в біомедичних дослідженнях.

2. Використання способу обробки даних за різницею зрізів як базису нейрооперацій дозволяє не тільки розпаралелити процес підсумовування векторних даних, але й сумістити у кожному циклі обробки виконання двох нейрооперацій: підсумовування зважених вхідних сигналів та нелінійне претворення у вигляді порівняння часткових сум з поточним зовнішнім порогом обробки. Це дозволяє значно розширити функціональні можливості ФН за рахунок його роботи у трьох режимах з формуванням відповідних вихідних сигналів.

3. Паралельний адаптивний суматор з можливістю роботи у чотирьох режимах має регулярну структуру завдяки рекурсивному характеру співвідношень для базових операцій обробки за РЗ, а це, у свою чергу, дозволяє розглядати його як лінійний систолічний масив з конвеєрною обробкою даних.

4. Використовуючи відому апаратно-програмну платформу для ПЛІС фірми Altera, можна побудувати нейрочіп для реалізації одного багатовхідного нейрона з пороговою і/або лінійною функціями активації або фрагмента шару нейромережі з декількох таких нейронів.

СПИСОК ЛІТЕРАТУРИ

1. Захаров С. М. Оптоэлектронные интегральные схемы с применением полупроводниковых вертикально излучающих лазеров / С. М. Захаров, В. Б. Федоров, В. В. Цветков // Квантовая электроника. — 1999. — Т. 28, № 3. — С. 189—206.
2. Грушвицкий Р. И. Проектирование систем на микросхемах программируемой логики / Р. П. Грушвицкий, А. Х. Мурсаев, Е. П. Угрюмов. — СПб. : БХВ — Петербург, 2002. — 608 с. — ISBN 5-94157-002-3.
3. Логовский А. Технология ПЛИС и ее применение для создания нейрочипов : [Электронный ресурс] // Открытые системы. — 2000. — № 10. — Режим доступа : <http://www.osp.ru/os/2000/10/178242>.
4. Галушкин А. И. Нейрокомпьютеры. Кн. 3 : учеб. пособие для вузов ; под. общ. ред. А. И. Галушкина. — М. : ИПРЖР, 2000. — 528 с. — ISBN 5-93108-007-4.
5. Шахнов В. А., Власов А. И., Кузнецов А. С., Поляков Ю. А. Нейрокомпьютеры: архитектура и реализация [Электронный ресурс]. — Режим доступа : http://www.citforum.ru/hardware/neurocomp/neurocomp_01.shtml.
6. Круглов В. В. Искусственные нейронные сети. Теория и практика. / В. В. Круглов, В. В. Борисов. — 2-е изд., стереотип. — М. : Горячая линия — Телеком, 2002. — 382 с. — ISBN 5-93517-031-0.
7. Бреус В. В. Модель «быстрого нейрона» на основе разрядно-параллельных алгоритмов [Электронный ресурс]. — Режим доступа : <http://www.ccssu.crimea.ua/eng/conf/oi2000/list.html>.
8. Аверьян Э. Д. Ассоциативная нейронная сеть СМАС. Ч. 1. Структура, объем памяти, обучение и базисные функции / Э. Д. Аверьян // Информационные технологии. — 1997. — № 5. — С. 6—14.
9. Аверьян Э. Д. Ассоциативная нейронная сеть СМАС. Ч. 2. Процессы обучения, ускоренное обучение, влияние помех, устранение влияния помех в двухслойной сети / Э. Д. Аверьян // Информационные технологии. — 1997. — № 6.
10. Патент 25525 України, МПК6 G06G7/60. Пристрій для моделювання нейрона / А. А. Смеров, Ю. М. Романишин, Ю. М. Сліпченко ; заявник і патентовласник Державний університет «Львівська політехніка». — № 97010322; заявл. 27.01.97; опубл. 25.12.98, Бюл. № 6.
11. Алгоритм работы циклической ансамблевой нейронной сети [Электронный ресурс] // Успехи современного естествознания. — 2005. — № 10. — Режим доступа : http://www.rae.ru/use/?section=content&op=show_article&article_id=7782622.
12. Мартынюк Т. Б. Модель порогового нейрона на основе параллельной обработке по разностным срезам / Т. Б. Мартынюк // Кибернетика и системный анализ. — 2005. — № 4. — С. 78—89.
13. Timchenko L. Parallel transformation / L. Timchenko, M. Grudin, T. Martyniuk, A. Kozhemiako // Управляющие системы и машины. — 1998. — № 5. — С. 93—95.
14. Патент 46877 України, МПК7 G06G7/14, G06G7/50. Конвеєрний підсумовуючий пристрій / Т. Б. Мартинюк, В. П.

Кожем'яко, А. В. Кожем'яко, І. А. Вербицький, С. А. Василецький; заявник і патентовласник ВНТУ. — № 99063405; заявл. 18.06.99; опубл. 17.06.02, Бюл. № 6.

15. Васюра А. С. Исследование процесса конвейерной обработки массива чисел / А. С. Васюра, Т. Б. Мартинюк, А. В. Кожемяко // Оптико-електронні інформаційно-енергетичні технології. — 2002. — № 1(3). — С. 85—94.

16. Мартинюк Т. Б. Особливості реалізації ітераційних алгоритмів багатооперандної обробки на систолічних масивах / Т. Б. Мартинюк, А. В. Кожем'яко, Л. М. Куперштейн // Оптико-електронні інформаційно-енергетичні технології. — 2002. — № 2(4). — С. 123—132.

17. Кун С. Матричные процессы на СБИС ; пер. с англ. / С. Кун. — М. : Мир, 1991. — 672 с. — ISBN 5-03-001857-3.

18. Патент 38491 України, МПК8 G06G7/00. Пристрій для моделювання нейрона / Т. Б. Мартинюк, Л. М. Куперштейн, І. В. Мороз, О. І. Чечельницький; заявник і патентовласник ВНТУ. — № 200810096; заявл. 04.08.08; опубл. 12.01.09, Бюл. № 1.

Рекомендована кафедрою лазерної та оптоелектронної техніки

Надійшла до редакції 15.12.09
Рекомендована до друку 28.09.09

Мартинюк Тетяна Борисівна — доцент, **Кожем'яко Андрій Вікторович** — доцент, **Фофанова Наталія Володимирівна** — аспірантка.

Кафедра лазерної та оптоелектронної техніки, Вінницький національний технічний університет