

ПРОБЛЕМНО-ОРІЄНТОВАНИЙ ПІДХІД ДО РЕАЛІЗАЦІЇ ШТУЧНИХ НЕЙРОННИХ СИСТЕМ

Вибрано принципи побудови, запропоновано методикку вибору елементної бази, розроблено структуру та проведено оцінювання основних параметрів комп'ютерних нейроорієнтованих систем.

Ключові слова: комп'ютерна система, проблемна орієнтація, нейроні, програмовані логічні інтегральні схеми, процесори цифрової обробки сигналів.

Постановка проблеми

Вимоги, що висуваються до комп'ютерних нейромережових засобів за продуктивністю та ефективністю використання обладнання не забезпечуються в рамках традиційних підходів до їх реалізації. Одним із основних шляхів підвищення ефективності використання обладнання та забезпечення виконання високої продуктивності є спеціалізація та проблемна орієнтація комп'ютерних засобів на реалізацію нейроалгоритмів. При спеціалізації та проблемній орієнтації здійснюється орієнтація структури, системи команд і принципів організації обчислень в комп'ютерних засобах на реалізацію нейроалгоритмів. В спеціалізованих і проблемно-орієнтованих комп'ютерних засобах в залежності від вимог за швидкістю нейроалгоритми можуть бути реалізовані програмним, мікропрограмним або апаратним шляхом. Кожен із перерахованих видів реалізації нейроалгоритмів має свої особливості, переваги і недоліки. В сучасних комп'ютерних засобах для реалізації нейроалгоритмів використовуються комбіновані підходи з перевагою одного з перерахованих видів. Вибір виду реалізацій ґрунтується на забезпеченні високої ефективності використання обладнання E та вимог конкретного застосування. Критерій ефективності використання обладнання E зв'язує продуктивність з витратами обладнання та дає оцінку комп'ютерним елементам за продуктивністю [1]. Кількісна величина ефективності використання обладнання для комп'ютерних засобів аналізу та синтезу зображень визначається наступним чином:

$$E = \frac{R}{t_{pz} W}$$

де R – складність нейромережових алгоритмів, яка визначається кількістю елементарних арифметичних операцій; t_{pz} – час розв’язання задачі; W – витрати обладнання на реалізацію комп’ютерних нейромережових засобів.

Розробка комп’ютерних нейромережових засобів з високою ефективністю використання обладнання вимагає широкого використання сучасної елементної бази (напівзамовних і замовних НВІС, однокристальних процесорів обробки сигналів), розробки нових методів, алгоритмів і спеціалізованих структур, орієнтованих на реалізацію базових операцій і нейромережових алгоритмів.

Тому актуальною проблемою є розробка комп’ютерних нейромережових засобів з високою ефективністю використання обладнання

Аналіз останніх досліджень та публікацій

Аналіз останніх досліджень і публікацій у галузі розробки комп’ютерних нейромережових засобів [1-9] показує, що такі засоби в основному реалізуються програмним шляхом на базі комп’ютерних систем з структурною організацією універсального типу. Така реалізація нейромереж має такі недоліки:

- апаратні засоби є надлишковими в функціональному та структурному відношенні;
- комп’ютерні системи не враховують вимоги конкретних застосувань щодо продуктивності, габаритів, споживаної потужності;
- невисока ефективність використання обладнання.

Апаратна реалізація нейромереж у вигляді спеціалізованої алгоритмічної системи, архітектура та організація обчислювального процесу в якій відображає структуру нейроалгоритму вимагає значних затрат обладнання та повного відпрацювання алгоритму.

Тому для реалізації штучних нейронних мереж (ШНМ) з високою ефективністю використання обладнання необхідно поєднувати універсальні і спеціалізовані підходи, програмні та апаратні засоби, тобто використовувати проблемно-орієнтований підхід.

Завдання і мета дослідження

Розробка комп’ютерних нейромережових систем з високою ефективністю використання обладнання вимагає проблемно-орієнтованого підходу з поєднанням універсальних і спеціалізованих засобів. При цьому розробка комп’ютерних нейроорієнтованих систем з заданими технічними параметрами повинна зводиться до доповнен-

ня обчислювального ядра додатковими спеціалізованими модулями. Розробка високопродуктивних спеціалізованих модулів вимагає широкого використання сучасної елементної бази, розпаралелювання і конвеєризації нейроалгоритмів.

Мета дослідження полягає в виборі принципів побудови, методики вибору елементної бази, розробленні структури та в оцінюванні основних характеристик комп'ютерних нейроорієнтованих систем.

Виклад основного матеріалу

Вибір принципів побудови комп'ютерних нейроорієнтованих систем. Розробку комп'ютерних нейроорієнтованих систем пропонується здійснювати на основі інтегрованого підходу, який охоплює [4]:

- сучасну елементну базу, апаратні та програмні комп'ютерні засоби;
- неромережеві методи та алгоритми;
- обчислювальні методи, алгоритми та НВІС-структури для реалізації базових операцій нейроалгоритмів.

В основу побудови комп'ютерних нейроорієнтованих систем необхідно покласти принципи, які дозволять зменшити вартість, терміни і розширити галузі їх застосування. Аналіз показує, що забезпечити дані вимоги можна при використанні таких принципів побудови [3,4]:

- змінного складу обладнання, що передбачає наявність ядра та змінних програмних і апаратних модулів, за допомогою яких ядро адаптується до вимог конкретного застосування;
- модульності, який передбачає розробку компонентів нейроорієнтованих систем у вигляді модулів, що мають вихід на стандартний інтерфейс;
- конвеєризації та просторового паралелізму обробки даних;
- відкритості програмного забезпечення, що передбачає можливість нарощування та його вдосконалення, максимального використання стандартних драйверів та програмних засобів;
- спеціалізації та адаптації апаратно-програмних засобів до структури нейроалгоритмів;
- програмованості архітектури шляхом використання репрограмованих логічних інтегральних мікросхем.

Вибір елементної бази для реалізації комп'ютерних нейроорієнтованих систем. Вибір елементної бази для реалізації комп'ютерних нейроорієнтованих систем доцільно здійснювати шляхом порівняння за параметром ефективності, який враховує продуктивність, габарити, потужність споживання та ціну. Даний параметр визначається так:

$$E_i = n_i t_i g_i k_i p_i s_i c_i \frac{\Pi_i}{\beta P_i \alpha S_i C_i},$$

де $i=1, \dots, m$; m – кількість варіантів; n_i – пороговий коефіцієнт, який враховує вимоги за продуктивністю; t_i – пороговий коефіцієнт, який враховує вимоги за температурою; g_i – пороговий коефіцієнт, який враховує спеціальні вимоги; p_i – пороговий коефіцієнт, який враховує вимоги за потужністю споживання; s_i – пороговий коефіцієнт, який враховує вимоги габаритів; c_i – пороговий коефіцієнт, який враховує вартість системи; Π_i – продуктивність i -о варіанта системи; P_i – потужність споживання i -о варіанта системи; S_i – габарити i -о варіанта системи, C_i – вартість i -о варіанта системи; β – коефіцієнт врахування потужності споживання; α – коефіцієнт врахування габаритів.

Аналіз [8,9] показав, що в даний час найчастіше для побудови комп'ютерних нейроорієнтованих систем використовується така елементна база:

- нейросигнальні процесори (нейрочіпи);
- систолічні процесори;
- програмовані логічні інтегральні схеми (ПЛІС);
- процесори загального призначення;
- процесори цифрової обробки сигналів (ПЦОС).

Ядро нейросигнальних процесорів являє собою типовий ПЦОС, а реалізована на кристалі додаткова логіка забезпечує виконання нейромережових операцій. Нейрочіпи спеціально розроблені для застосування в нейромережах і більшість з них орієнтовані на конкретні спеціалізовані системи.

Систолічні процесори - це матричні процесори, які працюють за конвеєрним принципом, виконуючи ритмічні обчислення і ритмічну передачу даних в системі і об'єднують в своєму складі деяке число процесорних елементів, а вся інша логіка реалізована на базі периферійних схем. При цьому може бути досягнута висока ступінь паралельності обробки даних. Недоліком цих систем може бути названа вузька спеціалізація обробних елементів, це веде до того, що систолічна матриця повинна бути оточена великою кількістю периферійних схем, що реалізують додаткову логіку, наприклад, модуль пам'яті ваг.

ПЛІС є однією із найперспективніших елементних баз для реалізації спеціалізованих компонентів комп'ютерних нейроорієнтованих систем. Такі мікросхеми складаються з множини однотипних елементів, що збільшує надійність їх роботи та спрощує проектування спеціалізованих компонентів.

Процесори загального призначення використовуються для реалізації програмних компонентів комп'ютерних нейроорієнтованих систем. Для цього розроблені бібліотеки для мов високого рівня, що реалізують різноманітні функції, які використовуються в комп'ютерних нейроорієнтованих системах, розроблені розширення об'єктно-орієнтованих мов, які дозволяють оперувати з нейромережами на рівні об'єктів та візуальні засоби розробки нейроорієнтованих систем.

ПЦОС найбільш часто застосовуються для реалізації процесорних ядер комп'ютерних нейроорієнтованих систем. Особливістю ПЦОС є високий ступінь спеціалізації та мала тривалість командного циклу, конвеєризація на рівні окремих мікроінструкцій та інструкцій, розміщення операндів більшості команд в регістрах, використання тіньових регістрів для збереження стану обчислень, поділ шин на шини команд і даних (гарвардська архітектура). Для ПЦОС характерним є наявність апаратного помножувача, що дозволяє виконувати множення чисел за один командний такт. Іншою особливістю ПЦОС є включення в систему команд таких операцій, як множення з накопиченням та різноманітних бітових операцій. В ПЦОС підтримується множинний доступ до пам'яті за один командний цикл. Основними фірмами, які випускають ПЦОС є Texas Instruments, Analog Devices та Motorola. При реалізації комп'ютерних нейроорієнтованих систем, розрахованих на побудову багатоканальних систем невеликої розмір-

ності - найкращим варіантом за критерієм якість/вартість є сигнальні процесори фірми Analog Devices (ціна аналогічного процесора Texas Instruments майже в два рази вища, при більш низьких показниках продуктивності).

Структура комп'ютерних нейроорієнтованих систем. Структуру комп'ютерної нейроорієнтованої системи можна представити у вигляді постійної частини F - процесорного ядра, яке реалізується на базі мікропроцесора ЦОС та змінної частини V - спеціалізованих модулів, які реалізують базові операції нейроалгоритмів. Структура комп'ютерної нейроорієнтованої системи наведена на рис.1.

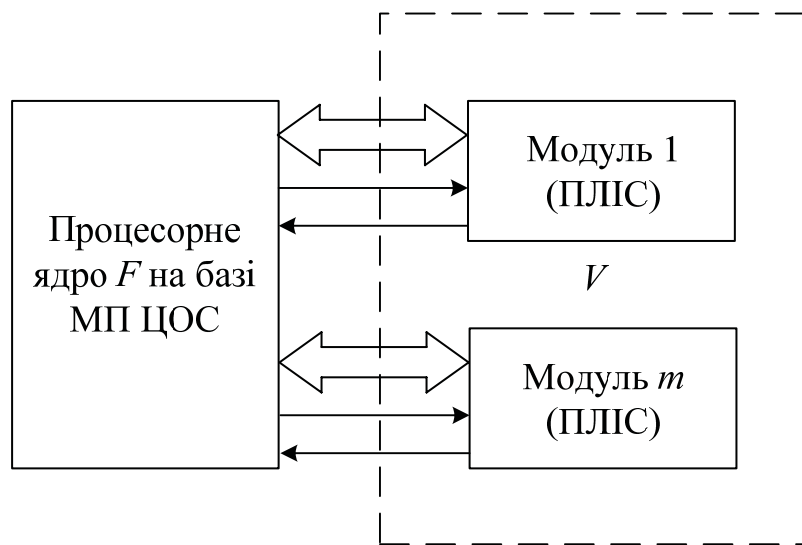


Рисунок 1 – Структура комп'ютерної нейроорієнтованої системи

Основою процесорного ядра комп'ютерної нейроорієнтованої системи є ПЦОС, який має набір команд добре пристосованих для виконання базових операцій нейроалгоритмів, а також повний набір команд загального призначення. Переважна більшість алгоритмів нейропарадигм зводиться до виконання обмеженого набору базових операцій типу “додавання – множення”. В таких ПЦОС за рахунок вдалих архітектурних і технологічних рішень вдалось поєднати високу швидкість виконання базових операцій нейроалгоритмів з ефективною реалізацією алгоритмів управління та прийняття рішень. Особливостями архітектури ПЦОС є:

- розвинута багатопинна організація, що дозволяє виконувати пересилання даних між різними функціональними вузлами;
- одночасне транспортування всіх операндів до операційного пристрою;

- використання апаратних засобів для виконання операції множення з підсумовуванням;
- виконання паралельно з основними операціями функцій адресації, переадресації, розгалуження програми;
- застосування апаратно вбудованих засобів організації циклічних обчислень, що переважають в нейроалгоритмах;
- використання модифікованої гарвардської архітектури;
- набір команд для виконання з підвищеною точністю арифметичних операцій;
- короткий командний цикл і виконання більшості команд за один цикл;
- здатність підтримувати конвеєрну обробку;
- зменшення довжини арифметичного конвеєра за рахунок використання багатопортової пам'яті;
- використання для обміну швидкодіючих каналів прямого доступу до пам'яті;
- механізм підтримки багатопроцесорності.

Необхідно відмітити, що основні характеристики процесорного ядра комп'ютерної нейроорієнтованої системи визначаються особливостями архітектури і технічними характеристиками ПЦОС. До числа таких характеристик відносяться: довжина інформаційного слова; число основних команд і час їх виконання; ємність пам'яті, що може адресуватися; ємності внутрішньокристалльної пам'яті даних і програм та кількість внутрішніх регістрів.

Структура комп'ютерних нейроорієнтованих систем залежить від конкретних вимог і множини нейроалгоритмів (N), які використовуються для розв'язання задач. При розв'язанні конкретної задачі здійснюється розподіл алгоритмів розв'язання задачі між обладнанням F і V

$$N = N_F + N_v$$

де N_F – множина алгоритмів, які виконуються на обладнанні F ; N_v – множина алгоритмів, які виконуються на обладнанні V .

В залежності від співвідношення N_F і N_v комп'ютерні нейроорієнтовані системи діляться на такі типи:

- з переважним використанням процесорного ядра (постійного обладнання F), коли $N_F \rightarrow N$, $N_V \rightarrow 0$, $N_F \gg N_V$;
- з переважним використанням спеціалізованих модулів (змінної частини V), коли $N_F \rightarrow 0$, $N_V \rightarrow N$, $N_F \ll N_V$;
- з рівномірним використанням постійного обладнання F і змінної частини V, коли $N_F \approx N_V$.

Перший тип комп'ютерних нейроорієнтованих систем характеризується тим, що основний обсяг обчислювальних потужностей зосереджений в процесорному ядрі.

В другому типі комп'ютерних нейроорієнтованих систем основні обчислювальні алгоритми реалізуються з допомогою спеціалізованих модулів, а процесорне ядро використовується для виконання допоміжних сервісних функцій.

Третій тип комп'ютерних нейроорієнтованих систем характеризується тим, що процесорне ядро забезпечує реалізацію алгоритмів управління, операцій введення-виведення та сервісних функцій, а спеціалізовані модулі – реалізують обчислювальні нейроалгоритми, які вимагають великого об'єму обчислень.

Аналіз оцінок основних параметрів комп'ютерних нейроорієнтованих систем. Однією з основних оцінок комп'ютерних нейроорієнтованих систем є швидкодія яка залежить від множини параметрів, основними з яких є [7]:

- тактова частота приймання вхідних даних та час запису в буферну пам'ять;
- тактова частота видачі результатів опрацювання та час циклу звертання до пам'яті;
- швидкодія ПЦОС та ПЛІС;

Для опрацювання неперервних потоків даних в режимі реального часу продуктивність комп'ютерних нейроорієнтованих систем повинна бути

$$P_{noc} \geq \frac{R}{t_{розв.зад}}$$

де R – складність нейроалгоритму; $t_{розв.зад}$ - час розв'язання задачі, який визначається так:

$$t_{розв.зад} = t_{вв} (1 - l_1) + t_{обр} + t_{вив} (1 - l_2)$$

де $t_{\text{вв}}$ – час введення інформації, l_1 – коефіцієнт врахування часового суміщення операцій введення з виконанням процесу обробки, $t_{\text{обр}}$ – час обробки, $t_{\text{вив}}$ – час виведення результатів розв'язання задачі, l_2 – коефіцієнт суміщення процесу виведення з процесом обробки.

В загальному ефективність комп'ютерних нейроорієнтованих систем визначається тим, як архітектура системи пристосована до розв'язання конкретної задачі та обчислюється так:

$$E_{\text{нс}} = \frac{V_p}{V_n},$$

де V_p – реальна швидкодія, V_n – номінальна швидкодія.

Витрати обладнання на реалізацію комп'ютерної нейроорієнтованої системи обчислюються за такою формулою:

$$W_{\text{нс}} = W_{\text{ПЯ}} + W_{\text{П}} + W_{\text{ПУ}} + \sum_{i=1}^n W_{\text{СМ}_i} m$$

де $W_{\text{ПЯ}}$ – витрати обладнання на процесорне ядро; $W_{\text{П}}$ – витрати обладнання на пам'ять, $W_{\text{ПУ}}$ – витрати обладнання на пристрій управління, $W_{\text{СМ}_i}$ – витрати обладнання на i -й спеціалізований модуль; m – кількість спеціалізованих модулів i -го типу; n – кількість типів спеціалізованих модулів.

Висновки

1. Структура комп'ютерних нейроорієнтованих систем представляється у вигляді постійної частини – процесорного ядра на базі ПЦОС та змінної частини, яка складається з спеціалізованих модулів, кількість яких залежить від вимог конкретного застосування.

2. Комп'ютерні нейроорієнтовані системи доцільно реалізовувати на основі інтегрованого підходу, який охоплює сучасну елементу бази, апаратні та програмні засоби, нейромережеві методи та алгоритми, обчислювальні методи, алгоритми та НВІС-структури та ґрунтується на таких принципах: змінного складу обладнання, модульності, конвеєризації та просторового паралелізму обробки даних, відкритості програмного забезпечення, спеціалізації та адаптації апаратно-програмних засобів до структури нейроалгоритмів.

3. Для вибору структури комп'ютерних нейроорієнтованих систем запропоновано використовувати критерій ефективності використання обладнання, який зв'язує продуктивність з витратами обладнання та дає оцінку елементам системи за продуктивністю..

ЛІТЕРАТУРА

1. Проблемно–ориентированные высокопроизводительные вычислительные системы: В.Ф. Гузик, В.Е. Золотовский: Учебное пособие. Таганрог:Изд-во ТРТУ, 1998. 236 с.
2. Уоссермен Ф.Нейрокомпьютерная техника. – М.: Мир,1992. – 259с.
3. Sorin Draghici. Neural Networks in Analog Hardware - Design and Implementation Issues.IEEE.February, 2000.
4. З.М. Любунь. Основи теорії нейромереж. Текст лекцій. Львів. Видавничий центр ЛНУ ім. Івана Франка. 2006, - 236с.
5. А.В. Палагин, В.Н. Опанасенко. Реконфигурируемые вычислительные системы. – К.: Просвіта, 2006.- 280с.
6. Николаев А.Б., Фоминых И.Б. Нейросетевые методы анализа и обработки данных. Учебное пособие. - М.: МАДИ (ГТУ), 2003, - 95с.
7. Цмоць І.Г. Інформаційні технології та спеціалізовані засоби обробки сигналів і зображень у реальному часі. – Львів: УАД, 2005.- 227с.
8. Грибачев В. П. Элементная база аппаратных реализаций нейронных сетей // Компоненты и технологии. 2006. № 8
9. Круг П.Г. Нейронные сети и нейрокомпьютеры: Учебное пособие по курсу «Микропроцессоры». – М.: Издательство МЭИ, 2002. – 176 с.