

І.Г. Цмоць, д.т.н., О.В. Скорохода, к.т.н., В.Б. Красовський,  
НУ «Львівська політехніка», м. Львів

## МОДЕЛІ ТА НВІС-СТРУКТУРИ ФОРМАЛЬНОГО НЕЙРОНА ПАРАЛЕЛЬНО-ВЕРТИКАЛЬНОГО ТИПУ З СУМІЩЕННЯМ ПРОЦЕСІВ НАДХОДЖЕННЯ ТА ОПРАЦЮВАННЯ ДАНИХ

*Анотація.* Розроблено моделі та методи опрацювання даних починаючи як з молодших, так і зі старших розрядів для формального нейрона паралельно-вертикального типу з суміщенням процесів надходження та опрацювання даних, синтезовано НВІС-структури та проведено оцінювання витрат обладнання на їхню реалізацію.

*Аннотация.* Разработаны модели и методы обработки данных, начиная как с младших, так и со старших разрядов для формального нейрона параллельно-вертикального типа с совмещением процессов поступления и обработки данных, синтезированы СВИС-структуры и проведена оценка затрат оборудования на их реализацию.

*Abstract.* The model and methods of data processing from high- and low-order bit for the parallel-vertical formal neuron with a simultaneous bitwise receipt and processing of input data and weight coefficients have been designed. VLSI-structure for its implementation has been synthesized, evaluation of the costs of equipment has been carried out.

*Ключові слова:* формальний нейрон, неймережа, паралельно-вертикальне опрацювання, НВІС-реалізація.

*Ключевые слова:* формальный нейрон, нейросеть, параллельно-вертикальная проработка, СВИС-реализация.

*Keywords:* formal neuron, neural networks, parallel vertical processing, VLSI-implementation.

**Вступ.** Широке застосування штучних нейронних мереж (ШНМ) у різних областях науки, техніки і виробництва вимагає від них високих технічних характеристик. Однією з найпоширеніших вимог, які ставляться до засобів реалізації ШНМ, є забезпечення високої продуктивності. Ця проблема виникає, як правило, при використанні ШНМ для вирішення проблем у режі мі реального часу, що накладає певні обмеження на процес опрацювання інформації [1-2].

Для забезпечення високої продуктивності НВІС-реалізації і з зменшення кількості виводів інтерфейсу пропонується опрацювання даних у нейронних елементах здійснювати паралельно розрядними зрізами (вертикально) на основі багатооперандного підходу. Таке опрацювання даних у нейронних елементах будемо називати паралельно-вертикальним опрацюванням [3-8].

У [3] показано, що відомі моделі нейронів не орієнтовані на НВІС-реалізацію, оскільки не ґрунтуються на елементарних арифметичних

операціях і вимагають значної кількості виводів інтерфейсу. У цій статті запропоновано модель та структуру нейрона паралельно-вертикального типу, яка, як і модель, розроблена у [3], забезпечує орієнтацію нейрона на НВІС-реалізацію та його адаптацію до вимог конкретного застосування.

**Основна частина.**

**Модель формального нейрона паралельно-вертикального типу з одночасним порозрядним надходженням, опрацюванням вхідних даних і вагових коефіцієнтів.** У моделі формального нейрона паралельно-вертикального типу з одночасним порозрядним надходженням та опрацюванням коефіцієнти та вхідні дані можуть надходити як старшими, так і молодшими розрядами вперед. Модель формального нейрона паралельно-вертикального типу з одночасним порозрядним надходженням і опрацюванням вхідних даних і вагових коефіцієнтів старшими розрядами вперед аналітично записується так:

$$y = f_{(p-s)} \left( f_a \left( f_z \left( f_{P_M} \left( f_{P_j} \left( f_{(s-p)} (w_{ji}), f_{(s-p)} (x_{ji}) \right) \right) \right) \right) \right). \quad (1)$$

Особливістю даної моделі формального нейрона є суміщення у часі процесів надходження розрядних зрізів вхідних сигналів і вагових коефіцієнтів, їх послідовно-паралельних перетворень і формування часткових результатів  $P_{ji}$ .

Структуру моделі формального нейрона паралельно-вертикального типу з одночасним порозрядним надходженням, опрацюванням вхідних даних і вагових коефіцієнтів старшими розрядами вперед, яка реалізує вираз (1), подано на рис. 1.

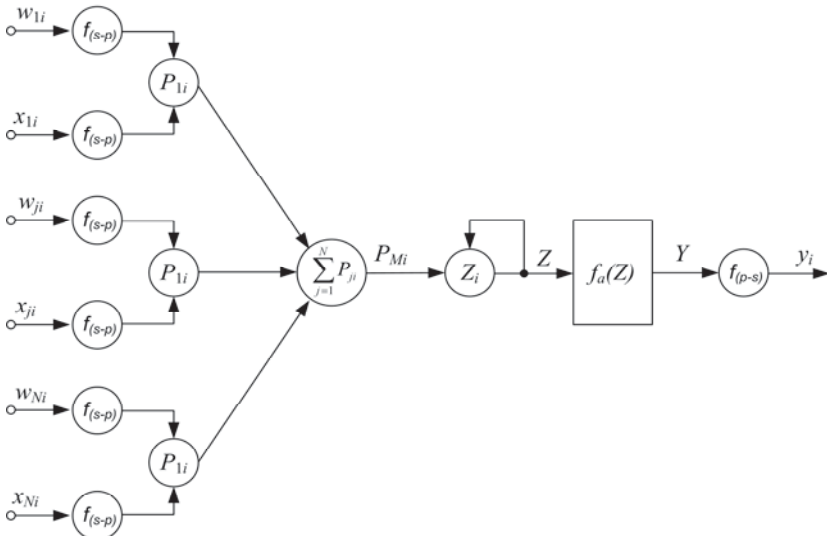


Рис. 1. Модель формального нейрона паралельно-вертикального типу з надходженням і опрацюванням даних старшими розрядами вперед

Формування часткових результатів при надходженні операндів старшими розрядами вперед у цій моделі відбувається так:

$$P_{ji} = \begin{cases} 0, & \text{коли } w_{ji} = x_{ji} = 0 \\ x_{j0}x_{j1}\dots x_{j(i-1)}, & \text{коли } w_{ji} = 1, x_{ji} = 0 \\ w_{j0}w_{j1}\dots w_{j(i-1)}, & \text{коли } w_{ji} = 0, x_{ji} = 1 \\ x_{j0}x_{j1}\dots x_{j(i-1)} + w_{j0}w_{j1}\dots w_{j(i-1)}, & \text{коли } w_{ji} = x_{ji} = 1 \end{cases} \quad (2)$$

При надходженні вагових коефіцієнтів, вхідних даних і формуванні часткових результатів  $P_{ji}$  починаючи з старших розрядів ускладнюється реалізація послідовно-паралельного перетворення та формування часткових результатів, а також збільшується розрядність підсумовувача макрочасткових результатів. Формування скалярного добуту  $\sum_{j=1}^N W_j X_j$  в даній моделі

здійснюється послідовним підсумовуванням макрочасткових результатів за формулою  $Z_i = 2Z_{i-1} + P_{Mi}$ , де  $Z_0=0$ .

Модель формального нейрона паралельно-вертикального типу з одночасним порозрядним надходженням і опрацюванням вхідних даних і вагових коефіцієнтів молодшими розрядами вперед аналітично записується так:

$$y = f_{(p-s)} \left( f_a \left( f_z \left( f_{P_i} \left( f_w \left( w_{ji} \right), f_x \left( x_{ji} \right) \right) \right) \right) \right) \right) \quad (3)$$

де  $f_w : R^N \rightarrow \left( \underbrace{R^N \rightarrow \dots \rightarrow R^N}_n \right)$  – оператор одночасного послідовно-паралельного перетворення  $N$  вагових коефіцієнтів;  $f_x : R^N \rightarrow R^N \times R^n$  – оператор одночасного послідовно-паралельного перетворення  $N$  вхідних даних

Структуру моделі формального нейрона паралельно-вертикального типу, яка реалізує вираз (3), подано на рис. 2, де Рг – регістр.

Надходженні вагових коефіцієнтів і вхідних даних починаючи з молодших розрядів передбачає реалізацію  $N$  паралельно-послідовних перетворень, формування в кожному такті часткових результатів і шляхом їх підсумовування відповідних частин макрочасткових результатів. Формування частин макрочасткових результатів зводиться до інтегральної однорозрядної макрооперації групового підсумовування  $N$  розрядів часткових результатів і їх підсумовування з допомогою накопичуючого суматора. Кількість макрочасткових результатів залежить від розрядності множників  $X_j$  та кількості розрядів множника, які аналізуються для отримання часткового результату. Визначається кількість макрочасткових результатів за формулою:

$$r = \left\lceil \frac{n+1}{g} \right\rceil, \quad (4)$$

де  $r$  – кількість груп, на які розбиваються множники  $X_j$ ;  $n$  – розрядність множників  $X_j$ ;  $g$  – кількість розрядів у групі, які аналізуються для отримання часткових результатів.

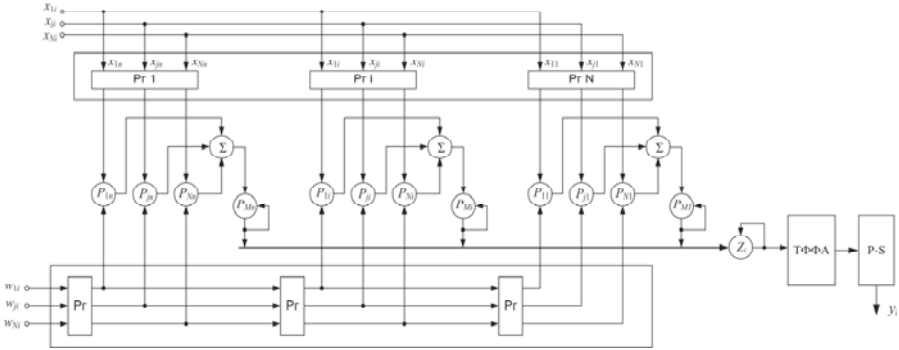


Рис. 2. Модель формального нейрона паралельно-вертикального типу з надходженням і опрацюванням даних молодшими розрядами вперед

Формування  $i$ -го макрочасткового результату з аналізом одного розряду множника здійснюється за формулою:

$$P_{Mi} = \sum_{j=1}^n 2^{-i} \sum_{j=1}^N w_{ji} x_{ji}. \quad (5)$$

Формування скалярного добуту у цій моделі здійснюється послідовним підсумовуванням макрочасткових результатів за формулою  $Z_i = 2^{-i} Z_{i-1} + P_{Mi}$ , де  $Z_0=0$ . Підсумовуванням макрочасткових результатів із зсувом вправо накопичених результатів забезпечує зменшення розрядності накопичуючого суматора.

У загальному випадку у даній моделі формального нейрона кількість макрочасткових результатів залежить як від розрядності операндів, так і від алгоритму формування часткових результатів. Необхідно також зауважити, що час обчислення результату виходу нейрона, в основному, визначається розрядністю операндів, а не їхньою кількістю.

**Нейроелемент паралельно-вертикального типу з одночасним порозрядним надходженням, опрацюванням вхідних даних і вагових коефіцієнтів.**

На основі моделі формального нейрона (1) розроблено базову структуру нейроелемента паралельно-вертикального типу із суміщенням процесів надходження і формування часткових результатів для розрядних зрізів вхідних даних і вагових коефіцієнтів (рис. 3).

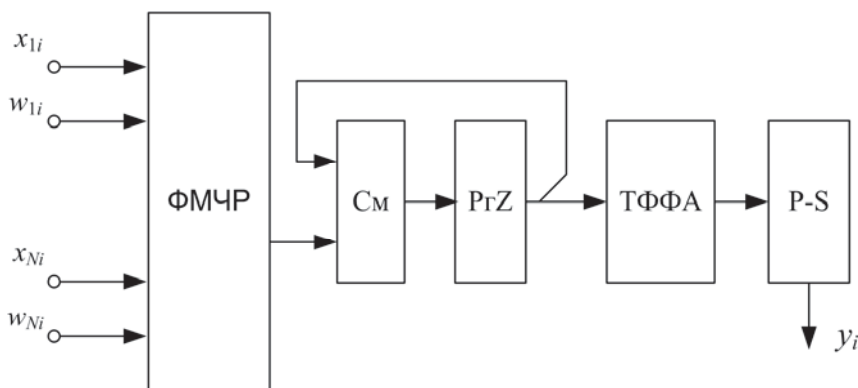


Рис. 3. Базова структура нейроелемента паралельно-вертикального типу із суміщенням процесів надходження і опрацювання даних

Надходження вхідних даних  $x_{ji}$  і вагових коефіцієнтів  $w_{ji}$  у такий нейроелемент може здійснюватися як старшими, так і молодшими розрядами вперед. Якщо вхідні дані та вагові коефіцієнти надходять старшими розрядами вперед, формування макрочасткових результатів  $P_{Mi}$  здійснюється так:

$$P_{Mi} = \sum_{j=1}^N [(x_{j0}x_{j1} \dots x_{ji})w_{ji} + (w_{j0}w_{j1} \dots w_{ji-1})x_{ji}]. \quad (6)$$

Структуру пристрою, який реалізує алгоритм обчислення скалярного добутку з надходженням операндів послідовним кодом старшими розрядами вперед, подано на рис. 4, де ФЧР – формувач часткових результатів, ПІ – тактові імпульси.

У цьому пристрої надходження вхідних значень  $X_j$  і вагових коефіцієнтів  $W_j$  здійснюється одночасно послідовним кодом старшими розрядами вперед. Розряди чисел  $x_{j1}, \dots, x_{jm}$  і  $w_{j1}, \dots, w_{jm}$  записуються у регістрах Рг  $X_1$ , Рг  $W_1, \dots, \text{Рг } X_N$ , Рг  $W_N$  і надходять у ФЧР, на виходах якого формується макрочастковий результат обчислення  $P_{Mi}$  скалярного добутку. Формування макрочасткових результатів обчислення  $P_{Mi}$  здійснюється за допомогою логічних елементів І та багатовходового суматора БСм.

Макрочасткові результати  $P_{Mi}$  формуються за допомогою логічних елементів І та багатовходового суматора БСм. Скалярний добуток  $Z$  обчислюють, підсумовуючи зсунуту вліво суму часткових результатів  $Z_{i-1}$  при нерухомому макрочастковому результаті  $P_{Mi}$  відповідно до формули:

$$Z_i = 2Z_{i-1} + P_{Mi}, \quad Z_0 = 0. \quad (7)$$

Обчислення скалярного добутку в даному пристрої виконується за  $n$  тактів, тривалість яких визначається так:

$$t = t_{Pz} + t_{2N-C_M},$$

де  $t_{2N-C_M}$  – час додавання  $2N$   $n$ -розрядних чисел.

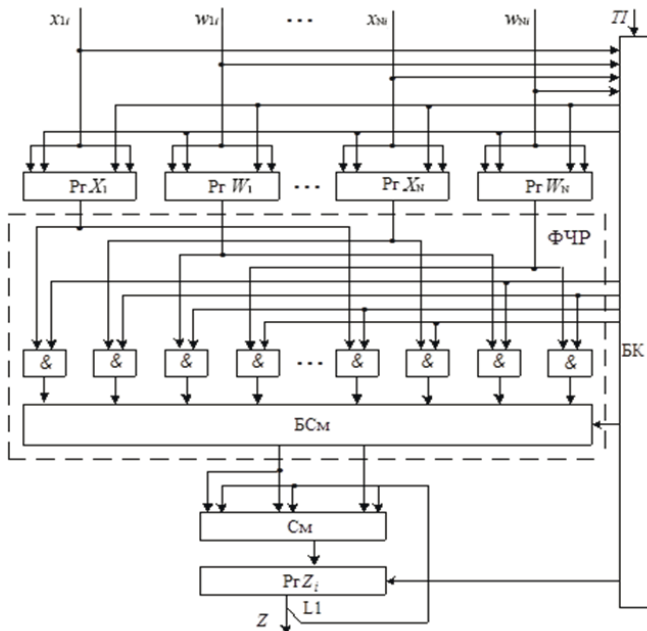


Рис. 4. Структура пристрою обчислення скалярного добутку з надходженням операндів послідовним кодом старшими розрядами вперед

Апаратні затрати на реалізацію пристрою обчислення скалярного добутку з надходженням операндів послідовним кодом старшими розрядами вперед визначаються так:

$$W = 2(N + 1)W_{P_2} + 2NnW_i + W_{2N-C_M} + W_{C_M 2n} + W_{BK},$$

де  $W_b$ ,  $W_{2N-C_M}$ ,  $W_{C_M 2n}$ ,  $W_{BK}$  – апаратні затрати на реалізацію відповідно елемента І,  $2N$ -входового суматора,  $2n$ -розрядного суматора і блоку керування.

Якщо вхідні дані  $x_{ji}$  і вагові коефіцієнти  $w_{ji}$  надходять молодшими розрядами вперед, формування частин макрочасткових результатів зводиться до групуового підсумовування  $N$  однорозрядних чисел. Кількість макрочасткових результатів залежить як від розрядності множників, так і від кількості розрядів, які аналізують для їх формування. Кількість макрочасткових результатів  $r$  визначають за формулою (4).

Формування макрочасткових результатів може здійснюватися прямим шляхом ( $g=1$  і  $g=2$ ) або на основі попередніх обчислень ( $g>2$ ). Формування  $h$ -го макрочасткового результату при аналізі одного розряду множників ( $g=1$ ) здійснюється так:

$$P_{Mh} = \sum_{i=1}^n 2^{-i} \sum_{j=1}^N w_{ji} x_{jh}, \quad (8)$$

Синтезовано та досліджено нейроелемент паралельно-вертикального типу із суміщенням процесів надходження і формування часткових результатів для випадку надходження операндів молодшими розрядами вперед. Розроблений нейроелемент забезпечує однотипність і простоту реалізації, зменшення кількості зовнішніх виводів і досягнення високої продуктивності за рахунок глибокого розпаралелювання до бітового рівня та використання конвеєризації. Час обчислення скалярного добутку в такому нейроелементі визначається, передусім, розрядністю операндів, а не їх кількістю. Визначено, що підвищити ефективність нейроелемента можна роздільним або комплексним використанням методів, які дадуть змогу зменшити кількість макрочасткових результатів, час їхнього формування та підсумовування.

Структура пристрою обчислення скалярного добутку з послідовним формуванням макрочасткових добутків подана на рис. 5, де Рг – регістри; ФЧР – формувач часткових результатів; БСм – однорозрядний  $N$ -входовий суматор; ФМЧР – формувач макрочасткових результатів; См – суматор; ПЕ – процесорний елемент.

У кожному такті роботи на входи пристрою надходить розрядний зріз операндів  $x_{ji}$  і  $w_{ji}$ , які записуються в регістри відповідно РгХ і РгW. Дані з виходів регістрів РгW надходять на входи всіх ПЕ, а з виходів РгХ – тільки на вхід ПЕ<sub>1</sub>. У ПЕ<sub>l</sub> за допомогою ФЧР формуються розряди часткових результатів пар операндів шляхом множення операндів  $X_j$  на коефіцієнти  $K_{jl}$ , які визначається при аналізі  $l$ -х груп із  $g$  розрядів вагових коефіцієнтів  $W_j$ . На  $j$ -у виході ФЧР отримуємо  $h$ -і розряди часткового результату для  $j$ -ї пари операндів, які обчислюються так:

$$P_{jl} = A_j K_{jl},$$

де  $K_{jl}$  – коефіцієнт, який визначається при аналізі  $l$ -х груп із  $g$  розрядів вагового коефіцієнту  $W_j$ .

Паралельним додаванням  $h$ -х розрядів часткових результатів на БСм отримуємо  $h$ -у частину макрочасткового результату  $P_{lh}$ , яка надходить на входи ФМЧР, де виконується підсумовування за такою формулою:

$$P_{lk} = 2^{-1} P_{l(k-1)} + P_{lh},$$

де  $k=1, \dots, s$ ,  $P_{l0} = 0$ .

Отримані на виходах ПЕ макрочасткові результати за допомогою См підсумовуються за формулою:

$$Z_l = 2^{-s} Z_{l-1} + P_l,$$

де  $Z_0=0$ .

Пристрій паралельно-вертикального обчислення скалярного добутку працює за конвеєрним принципом і орієнтований на опрацювання неперервних потоків даних. Конвеєрний такт роботи такого пристрою

визначається так:

$$t_5 = t_{Pz} + t_{Cm},$$

де  $t_{Pz}$  – час спрацювання регістра,  $t_{Cm}$  – час додавання  $n$ -розрядних чисел. Обчислення скалярного добутку здійснюється за  $n$  конвеєрних тактів.

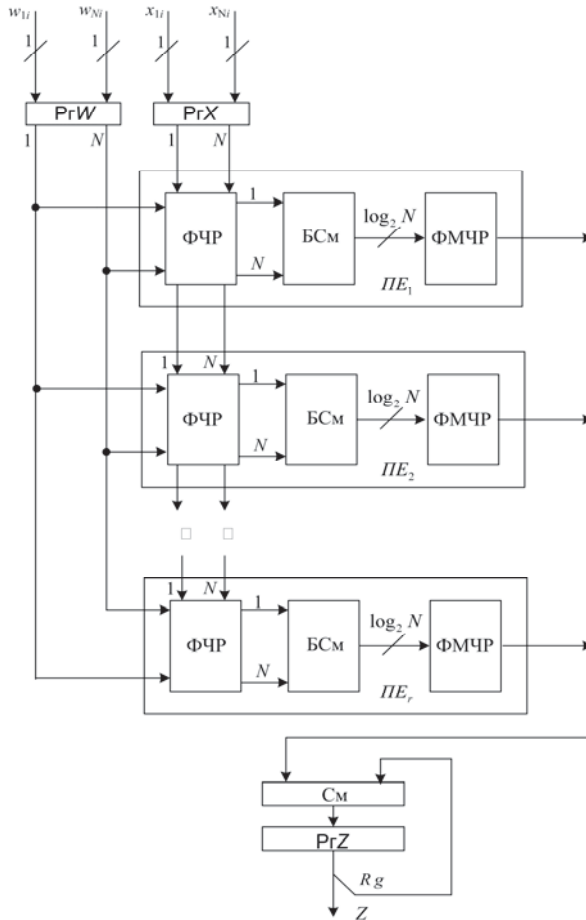


Рис. 5. Структура пристрою обчислення скалярного добутку з послідовним формуванням макрочасткових добутків

Апаратні затрати на реалізацію пристрою обчислення скалярного добутку з послідовним формуванням макрочасткових результатів, який працює за алгоритмом з аналізом одного розряду, визначаються так:

$$W_4 = 3W_{Pz} + nW_{PE} + W_{Cm},$$

де  $W_{PE}$ ,  $W_{Cm}$  – апаратні затрати на реалізацію відповідно процесорного елемента і  $n$ -розрядного суматора.



У випадку формування часткових результатів з аналізом двох розрядів множника ( $g=2$ ) використовується модифікований алгоритм Бута. Цей алгоритм передбачає розбиття множника  $X_j$  на групи з трьох розрядів, так що сусідні групи мають один спільний розряд. Розбиття множника  $X_j$  на групи починають з молодших розрядів, причому молодший розряд наймолодшої групи завжди доповнюється нулем.

### **Висновки.**

1. Формування частин макрочасткових результатів зводиться до інтегральної однорозрядної макрооперації групового підсумовування  $N$  розрядів часткових результатів і їх підсумовування за допомогою накопичуючого суматора.

2. Кількість макрочасткових результатів залежить від розрядності множників  $X_j$  та кількості розрядів множника, які аналізуються для отримання часткового результату

3. Розроблений нейроелемент забезпечує однотипність і простоту реалізації, зменшення кількості зовнішніх виводів і досягнення високої продуктивності за рахунок глибокого розпаралелювання до бітового рівня та використання конвеєризації. Підвищити ефективність нейроелемента можна роздільним або комплексним використанням методів, які дадуть змогу зменшити кількість макрочасткових результатів, час їхнього формування та підсумовування.

1. *Haykin S. Neural networks and learning machines. Third Edition. / S. Haykin. – New York: Prentice Hall, 2009. – 936 p.*
2. *Руденко О.Г., Бодяньський Є.В. Штучні нейронні мережі / О.Г. Руденко, Є.В. Бодяньський. – Харків: ТОВ «Компанія СМІТ», 2006. – 404 с.*
3. *Цмоць І.Г. Модель та НВІС-структури формального нейрона паралельно-вертикального типу з використанням мультиплексування шин / І.Г.Цмоць, О.В. Скорохода, Б.І. Балич // Збірник наукових праць Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова – Львів. – 2013. – Випуск № 67. – С. 160-166.*
4. *Цмоць І.Г. Методи та НВІС-структури пристроїв паралельно-вертикального обчислення сум парних добутків / І.Г. Цмоць, Б.І. Балич, О.В. Скорохода // Відбір і обробка інформації. – Львів, 2011. – № 33 (109). – С. 109–116.*
5. *Цмоць І.Г. Інформаційні технології та спеціалізовані засоби обробки сигналів і зображень у реальному часі / І.Г. Цмоць. – Львів: УАД, 2005. – 227 с.*
6. *Параллельная обработка информации: Т.4. Высокопроизводительные системы параллельной обработки информации / под ред. В.В. Грицька. – Киев: Наук. думка, 1988. – 272 с.*
7. *Шальто А.А. Методы аппаратной и программной реализации алгоритмов / А.А. Шальто. – СПб.: Наука, 2000 – 780 с.*
8. *Кун С. Матричные процессоры на СБИС / С. Кун. – М.: Мир, 1991. – 672 с.*

*Поступила 7.10.2013р.*