Моделювання багатопортової гетероасоціативної пам’яті (БГАП) на основі еквівалентністних моделей, що реалізуються на векторно-матричних перемножувачах

https://doi.org/10.31713/MCIT.2023.037

Володимир Красиленко
Вінницький національний аграрний університет
ВНАУ
м. Вінниця, Україна
krasvg@i.ua

Ілля Чіков
Вінницький національний аграрний університет
ВНАУ
м. Вінниця, Україна
ilya95chikov@gmail.com

Діана Нікітович
Вінницький національний аграрний університет
ВНАУ
м. Вінниця, Україна
diananikitovych@gmail.com

Анотація—**Робота присвячена розгляду питань стосовно асоціативної обробки інформації з метою апаратно-програмної реалізації відповідних математичних, імітаційних та фізичних моделей асоціативної пам’яті. На основі огляду відомих публікацій, в тому числі тих, у яких були розглянуті еквівалентністні моделі асоціативної та гетероасоціативної памяті та показані їх переваги, обгрунтовано необхідність подальших досліджень таких моделей АП чи ГАП та проведення імітаційних модельних експериментів в плані пошуку їх ефективних реалізацій. Наведені результати імітаційного моделювання двох реалізацій багатопортової гетероасоціативної памяті (БГАП) на основі векторно-матричних перемножувачів та векторно-матричних еквіваленторів, які, як прискорювачі, адаптовані до еквівалентністних моделей. Результати моделювання процесів гетеро-асоціативного розпізнавання кортежу букв, виконані у Mathcad для двох варіантів реалізації БГАП, підтверджують їх правильне функціонування, оскільки на виходах-портах БГАП сформовані моделями правильні кортежі всіх вихідних 100 букв, попарно асоційованих зі 100 вхідними буквами.**

Ключові слова—багатопортова асоціативна пам'ять, еквівалентністна модель, нейронна мережа, гетероасоціативна пам'ять, ємність асоціативної пам'яті. векторно-матричний перемножувач.

I. ВСТУП, ОГЛЯД ТА АНАЛІЗ ПУБЛІКАЦІЙ

Актуальність та необхідність дослідження процесів природнього інтелекту, асоціативної обробки інформації з метою апаратно-програмної реалізації відповідних математичних, імітаційних та фізичних моделей асоціативної пам’яті (АП) обумовлена широким застосуванням в останні десятиріччя самих новітніх біологічно натхненних та нейрофізіологічних принципів в нейрокібернетиці, в сучасних інтелектуальних системах підтримки прийняття рішень, системах дистанційного моніторингу, в робото-технічних системах розпізнавання та ідентифікації образів самого різноманітного типу та для самих різноманітних проблемно-орієнтованих сфер застосування. Відома ціла низка нейронних моделей та мереж, що реалізують авто-асоціативну та гетеро-асоціативну пам’ять (ГАП) [1-11]. Але аналіз значної частки подібних за тематичним спрямуванням публікацій, включно з вищевказаними [5-14], та більш нових та недавніх робіт [15-21] показує, що ємність таких моделей АП, яка визначається кількістю запам’ятованих та успішно асоційованих (розпізнаних) образів, не перевищує (0,14-0,60)*N, де N-кількість нейронів у моделі АП. В той же час, ще більш як 15 років тому, появились роботи [12-14], в яких були запропоновані та досліджені, так звані авторами, еквівалентністні моделі (ЕМ) нейромереж (НМ) та АП (ГАП), а в них експериментально були підтверджені їх суттєві переваги у порівнянні з іншими відомими моделями та нейронними парадигмами, особливо стосовно їх ємності та більш зручного відображення на апаратні паралельні процесори. Ємність таких ЕМ АП може, як мінімум у 4-5 разів перевищувати N, хоч часткові модельні експерименти дозволять стверджувати про ще більшу на порядки їх ємність. Крім того, ЕМ АП допускають значно більшу потужність завад, що перекручують та змінюють вхідні образи, при яких якісно та правильно формуються асоціативні відгуки. Все це обумовлює цікавість до подальших досліджень таких ЕМ АП чи ГАП в плані пошуку їх ефективних реалізацій. В роботах [12-14, 22] були промодельовані та досліджені чотири еквівалентністні моделі, в результаті чого було

зроблено висновки, що такі моделі успішно можуть бути застосовані для побудови не лише однопортової але і багатопортової, більш загальної, ГАП для обробки та розпізнавання образів з сильною кореляцією та значним їх ушкодженням завадами. В роботах [22, 23] були запропоновані та висвітлені деякі можливі апаратні реалізації таких ЕМ моделей, модифікованих під обробку 1-D чи 2-D образів. Але більшість цих пропозицій стосувалися лише однопортових реалізацій АП, що базувались на використанні оптичних систем з просторовим та часоімпульсним інтегруванням [22. 23]. Однак такі реалізації є суто аналоговими, а тому не допускають ні суттєвого збільшення динамічного діапазону значень оброблюваних сигналів і точності обчислювальних процедур, ні збільшення розмірності образів (векторів чи матриць), що запам'ятовуються в АП. Крім того, в тих же роботах на основі моделей, процедур обробки було визначено, що застосування найбільш узагальнених еквівалентністних моделей для реалізації багатопортової ГАП (БГАП) потребує векторно-матричних чи матрично-тензорних процедур з операціями еквівалентності (нееквівалентності), або спеціалізованих пристроїв-прискорювачів, так званих векторно-матричних еквіваленторів (ВМЕ). Оскільки, як було показано авторами роботи [22], ВМЕ реалізуються на основі двох векторно-матричних перемножувачів (ВМП), то це дозволяє застосовувати високопродуктивні швидкодіючі векторно-матричні спеціалізовані обчислювачі, матричні та систолічні процесори для побудови БГАП.

Вище наведений огляд та аналіз публікацій дозволяє визначити, як одне з важливих завдань, необхідність розроблення та моделювання таких ЕМ БГАП, які були б найкраще пристосованими для реалізації на базі ВМП чи ВМЕ, та визначення характеристик та показників таких моделей та реалізацій.

II. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ ЕМ БГАП НА БАЗІ ВМП ТА ВМЕ

Теоретичні основи проектування БГАП на основі ЕМ були розроблені та наведені в роботах [12-14, 22, 23]. А тому тут, з метою верифікації таких моделей та якості функціонування запропонованих БГАП на їх основі, ми розглядаємо лише аспекти моделювання вищезазначених об'єктів та аналізуємо отримані результати. Для імітаційного моделювання та проведення експериментів у Mathcad ми створили навчальну вибірку зі взаємно пов'язаних (асоційованих між собою) різних образів у вигляді векторів, які є елементами деякої множини. Для кращої наочності, візуалізації та прискореного сприйняття і порівняння, в одному з експериментів в якості образів було вибрано кодові вектори символів-літер ASC11, при цьому використано здебільше саме букви. Кожен кодовий вектор букви, чи вхідної чи з навчального набору, є чотирикратним повторенням восьми-бітового двійкового коду Грея (байту), числовим еквівалентом якого є число з діапазону від 0 до 255, яке є кодом символу. У процедурах перетворення традиційних двійкових кодів у двійкові коди Грея нами у Mathcad використовувалися моделі на базі векторизованих операцій еквівалентності (нееквівалентності). А чотирикратне повторення байтів кодових векторів нами використовувалося для збільшення розмірності векторів-образів та для можливості перевіряти завадостійкість при розпізнаванні образів у БГАП і в умовах розширеного діапазону потужності завад, яка пропорційна кількості змінених біт у кодових векторах.

Процедура введення 128 символів (фрагмент з частиною введених різних букв чи символів) показано на рисунку 1 (а). Кожен символ у відповідності до прийнятої системи кодування кодується байтом. На цьому ж фрагменті показана процедура введення 100 англійських букв чи символів, що будуть відповідати 100 вхідним портам, на кожен з яких подається кодовий вектор відповідної букви чи відповідного символу. Для формування пар гетеро-асоційованих образів ми за допомогою циклічного зсуву ставили у відповідність кожній букві зі створеного 100-буквенного набору наступну за нею букву. На рисунку 2 показані бінарне 2-градаційне зображення розмірністю 100×32 елементів (матриця INPX) та контрастне обернене зображення (матриця INPXN), що відповідають 100 входам БГАП. На рисунку 2 також показані у вигляді таких же двох бінарних матриць TX та TXN розмірністю 32×128 елементів 128 векторів навчальної вибірки, в яку введені 128 різних символів та букв з клавіатури ПЕОМ після їх перетворення у 32-х компонентні бітові вектори. Використовуючи формули, що наведені внизу рис. 2, нами була сформована у відповідності до кожного з 128 навчальних векторів (матриця TX та TXN) сукупність з 128 асоційованих навчальних векторів у вигляді матриць TY та TYN (на рис.3 вони показані транспонованими). Тобто першій букві з набору TX ставилася у відповідність друга буква, кожній подальшій букві – наступна, а останній – перша. На рис.3. (зліва) показані формули, що були нами використані для моделювання процедур знаходження за допомогою двох векторно-матричних перемножувачів (ВМП) необхідних матриць (HN та HNN) нормованих еквівалентностей та нееквівалентностей на першому кроці. Там же показані фрагменти вікон з результатами моделювання процедур обчислення вихідних матриць (OUTY іі OUTYN) нормованих еквівалентностей та нееквівалентностей на другому кроці.

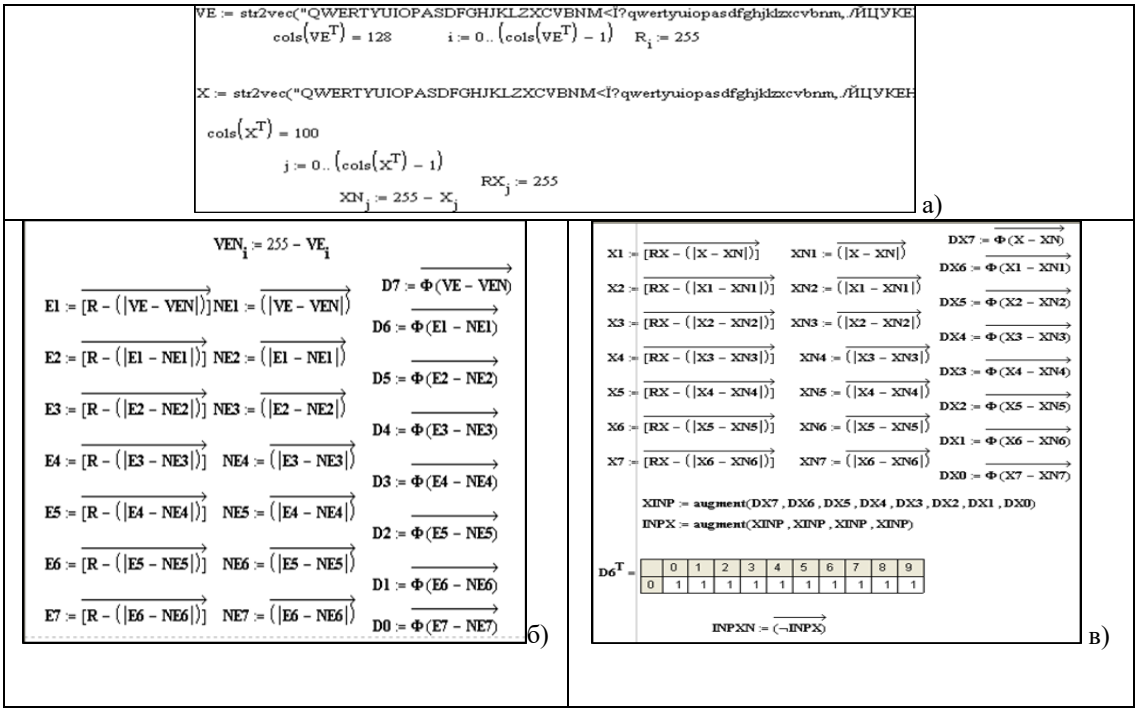
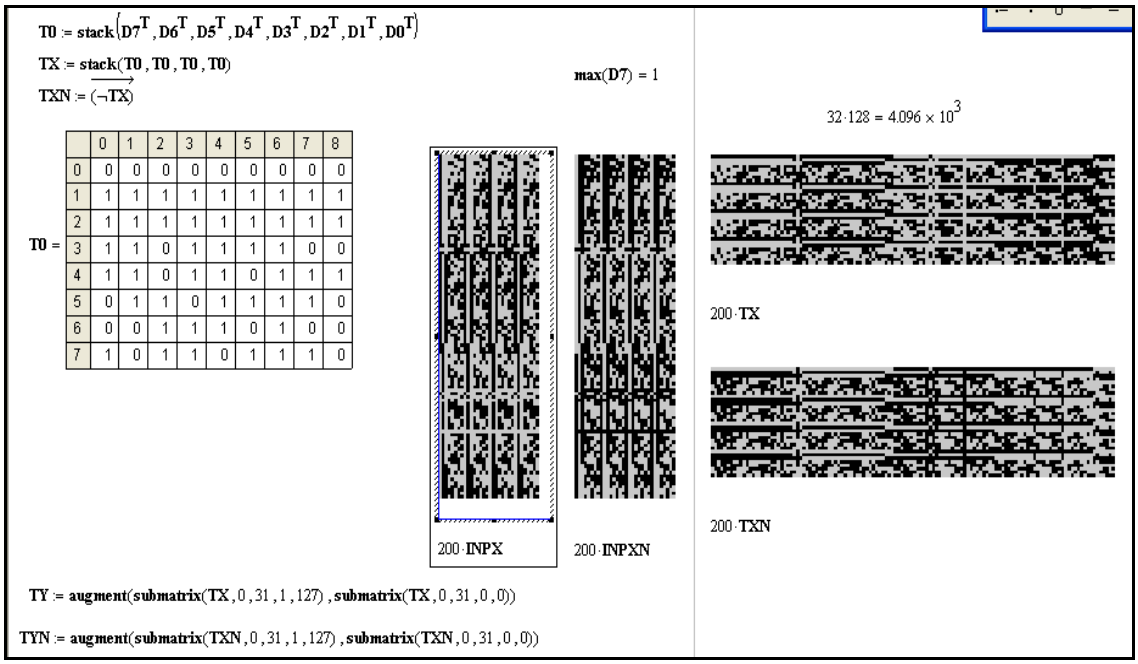


Рисунок 1. Вигляд фрагментів з вікон Mathcad, які описують процедури (а) введення набору вхідних векторів та формування з неї навчальної вибірки, формули (б) для кодування та формування матриці навчальних векторів, формули (в) для аналого-цифрового перетворення та кодування набору вхідних образів.



Рисуснок 2. Зображення матриць, що відповідають вхідному багатовекторному входута першій половині навчальної вибірки

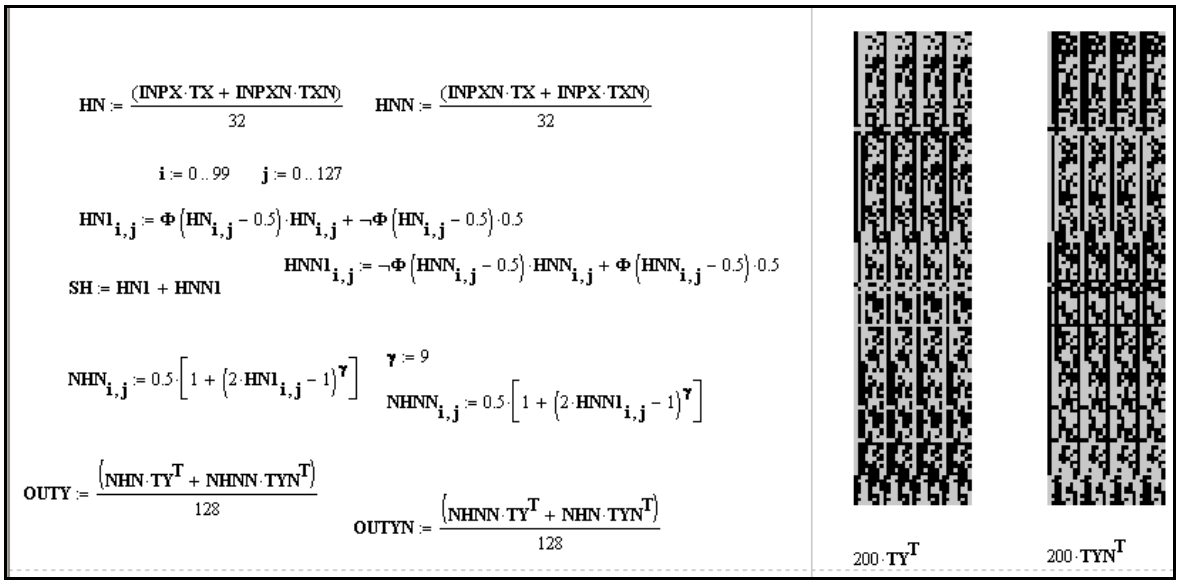


Рисунок 3. Зображення другої (вихідної) половини навчальної вибірки у вигляді зображень матриць (праворуч) та формули векторно-матричних еквівалентністних та нелінійних перетворень сигналів нейронів прихованого прошарку (ліворуч).

Крім того, на рис. 3 показані формули нелінійного перетворення сигналів нейронів прихованого прошарку з коефіцієнтом нелінійності γ, що відповідають по-компонентним нелінійним перетворенням матриць NHN та NHNN. Як видно з рисунка 4, використання операцій та векторизованих перетворень, що відтворюють порогову по-компонентну обробку суб-векторів та формування масиву вихідних векторів-відгуків на виході БГАП з наступним їх перетворенням у вихідні букви-символи, дозволили правильно сформувати необхідний кортеж букв, в якому всі 100 пар правильно асоційовано розпізнано. Вхідний та утворений вихідний кортежі букв показані на рисунку 4 внизу. Таким чином, отримані результати моделювання підтверджують правильне функціонування БГАП, оскільки на виходах-портах БГАП утворюється кортеж вихідних 100 букв, попарно асоційованих зі 100 вхідними буквами. Як бачимо, всі 100 букв на виході БГАП розпізнані у відповідності до сформованих навчальною вибіркою гетеро-асоціацій, як при використанні для моделювання моделі БГАП на базі ВМЕ (результати показані на рис.5), так і при використанні у модельних експериментах моделі БГАП на базі ВМП (результати показані на рис.6). Утворені кортежі символів-букв, що відображені також на рис.5, свідчать про успішні відтворення всіх асоційованих пар, а саме формування для кожної букви з вхідного кортежу відповідної їй наступної

.

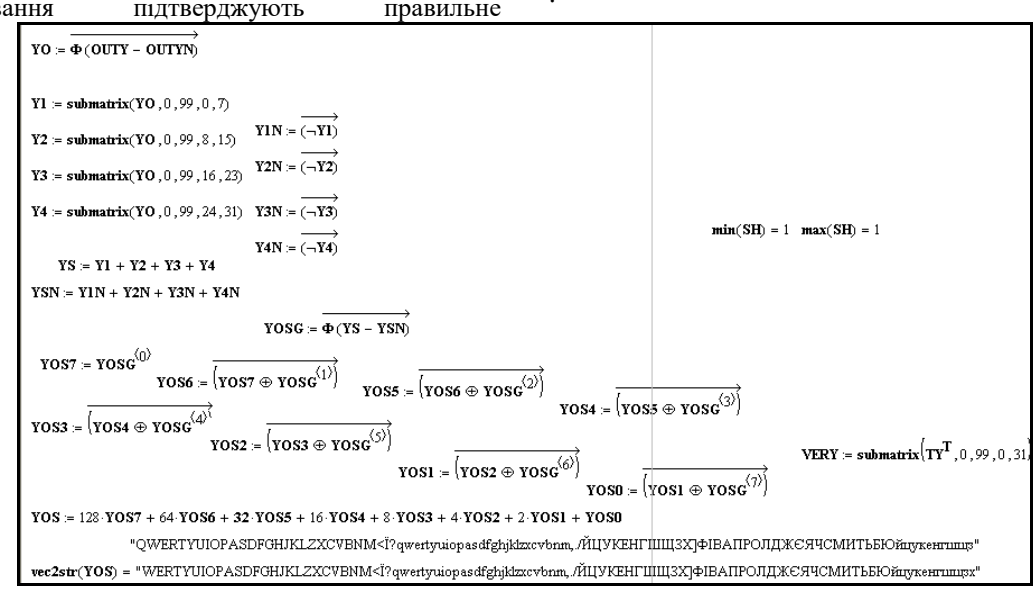


Рисунок 4. Вікно Mathcad з формулами для моделювання функцій активації та порогової обробки при формуванні масиву відгуків БГАП у вигляді вихідних векторів

Додаткові експерименти показали, що промодельовані можливі варіанти реалізації БГАП допускають у допустимих межах ушкодження завадами кодових векторів, яке не спричиняє до порушень правильного функціонування БГАП. Про ці додаткові аспекти отриманих результатів буде повідомлено у доповіді.

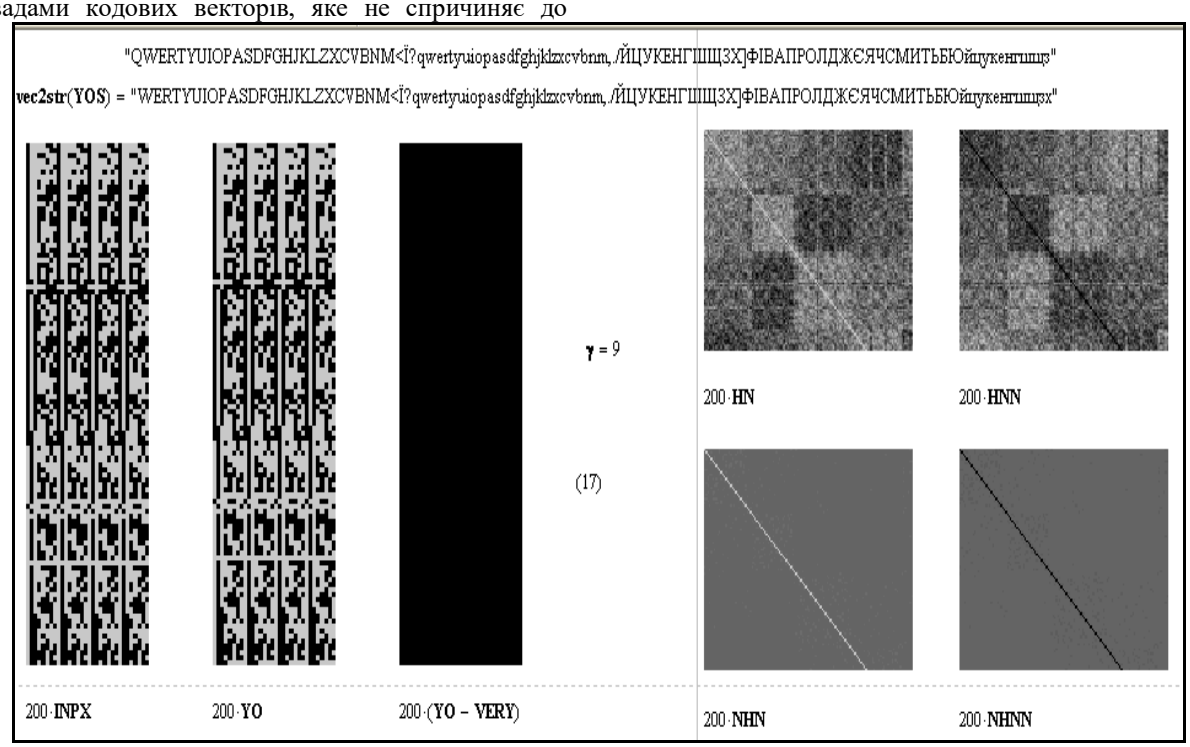


Рисунок 5. Результати моделювання БГАП на основі ВМЕ у вигляді результуючихзображень та різницевого (третє зліва у вигляді нульової матриці) зображення,що підтверджує успішне асоціювання усіх 100 векторів.

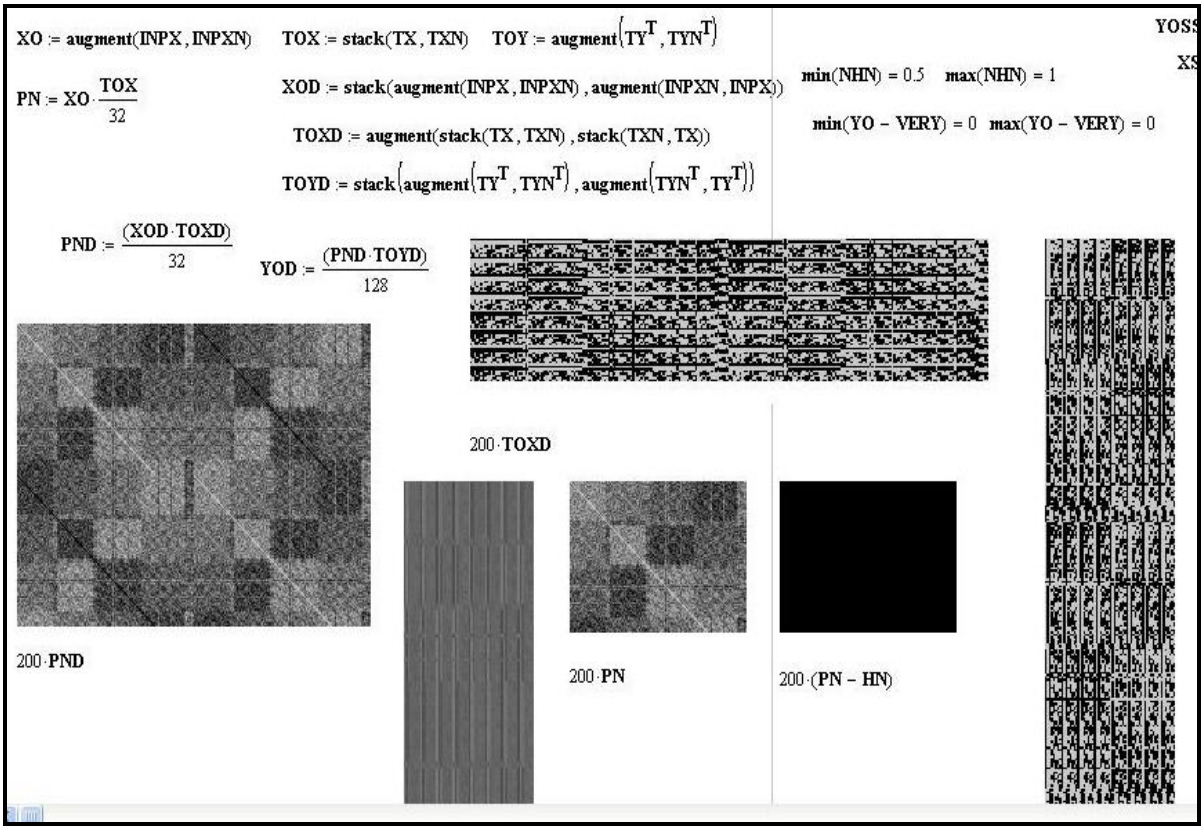


Рисунок 6. Результати моделювання БГАП на основі ВМП

III. ВИСНОВКИ

В результаті розроблення та моделювання БГАП, що функціонують на основі еквівалентністних моделей , було підтверджено можливості реалізацій БГАП на базі таких апаратно-програмних прискорювачів з паралельною обробкою, як векторно-матричні перемножувачі та векторно-матричні еквівалентори (по суті 2 перемножувача), які додатково до виконання ними паралельно лінійно-алгебраїчних процедур- операцій були б наділені можливістю виконання ними паралельно по-компонентних нелінійних перетворень.

ЛІТЕРАТУРА

[1] M.T. Hagan, B.D. Howard and M.B. Hudson, “Neural Network Design. Chapter 13,” PWS Publ. Company, 1996, pp. 13-1–13-37.

[2] J.A. Anderson, “An Introduction to Neural Networks. Chapter 6-7,” MIT Press, 1997, pp. 143–208.

[3] S.O. Haykin, “Neural Networks and Learning Machines (3rd Edition),” Prentice Hall, 2009.

[4] S. Grossberg “Nonlinear neural networks: Principles, mechanisms, and architectures,” Neural Networks, vol. 1, 1988, pp. 17–61.

[5] O. Rudenko, and E.V. Bodyansky, “Artificial neuron networks: Handbook,” Kharkiv: SMIT, 2005.

[6] J.J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” Proc. Natl. Acad. Sci., vol. 79, 1982, pp. 2554–2558.

[7] B. Kosko, “Bi-directional associative memories,” IEEE Transactions on Systems, Man and Cybernetics, vol. 18(1), 1988, pp. 49–60.

[8] B. Kosko, “Constructing an associative memory,” vol. 12, 1987, pp. 137–144.

[9] B. Kiselyov, N. Kulakov, A. Mikaelian, and V. Shkitin, “Optical associative memory for high-order correlation patterns,” Opt. Eng., vol. 31(4), 1995, pp. 764–767.

[10] A.A. Frolov, D. and Husek, I.P., “Muraviev Informational efficiency of sparsely encoded Hopfield-like associative memory,” Optical Memory & Neural Networks, vol. 12(3), 2003, pp. 177–197.

[11] A.A. Frolov, D.A. Rachkovskij, and D. Husek, On information characteristics of Willshaw-like auto-associative memory. Neural Network World, vol. 12(2), 2002, pp. 141–157.

[12] V.G. Krasilenko, and A.T. Magas, “Multiport optical associative memory based on matrix-matrix equivalentors,” Proceedings of SPIE (SPIE, Bellingham, WA 1997), vol. 3055, 1997, pp. 137–146.

[13] V.G. Krasilenko, A.I. Nikolskyy, R.A. Yatskovskaya, and V.I. Yatskovsky, “The concept models and implementations of multiport neural net associative memory for 2D patterns,” in Optical Pattern Recognition XXII, Proceedings of SPIE (SPIE, Bellingham, WA 2011), vol. 8055, 2011, 80550T.

[14] V.G. Krasilenko, A. A. Lazarev, and S.K. Grabovlyak, “Design and simulation of a multiport neural network heteroassociative memory for optical pattern recognitions,” SPIE, vol. 8398, 2012, 83980N.

[15] M. Demircigil, J. Heusel, M. Lowe, S. Upgang, and F. Vermet, “On a model of associative memory with huge storage capacity,” J. Stat. Phys, vol. 168(2), 2017, pp. 288–299.

[16] V. Gripon, J. Heusel, M. Lowe, and F. Vermet, “A comparative study of sparse associative memories,” Journal of Statistical Physics, vol. 164, 2016, pp. 105–129.

[17] V. Gripon, M. Lowe, and F. Vermet, “Associative memories to accelerate approximate nearest neighbor search,” Appl. Sci., vol. 8(9), 2018, 1676.

[18] D. Krotov, and J.J. Hopfield, “Dense associative memory for pattern recognition,” Advances in Neural Information Processing Systems 29 (NIPS 2016), 2016, pp. 1172–1180.

[19] D. Krotov, and J.J. Hopfield, “Dense associative memory is robust to adversarial inputs,” Neural Computation, vol. 30(12), 2018, pp. 3151–3167.

[20] A. Mazumdar, and A.S. Rawat, “Associative memory using dictionary learning and expander decoding,” Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 2017, pp. 267–273.

[21] N. Onizawa, H. Jarollahi, T. Hanyu, and W.J. Gross, “Hardware implementation of associative memories based on multiple-valued sparse clustered networks,” IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 6(1), 2016, pp. 13–24.

[22] V.G. Krasilenko, A.E. Nikolsky, and Y.A. Flavitskay, “The structures o Optical Neural Nets Based on New Matrix – Tensor Equivalental Models (MTEMS) and Results of Modeling,” Optical Memory and Neural Networks (Information Optics), vol. 19, 2010, pp. 31–38.

[23] V.G. Krasilenko, A.A. Lazarev, and D.V. Nikitovich, “Design and simulation of optoelectronic neuron equivalentors as hardware accelerators of self-learning equivalent convolutional neural structures (SLECNS),” Proc. SPIE 10689, Neuro-inspired Photonic Computing, 106890C (May 21, 2018). https://doi.org/10.1117/12.2316352