

УДК 004.032.26(063); 004.383.8.032.26(063)

О.К. КОЛЕСНИЦЬКИЙ\*

## АНАЛІТИЧНИЙ ОГЛЯД АПАРАТНИХ РЕАЛІЗАЦІЙ СПАЙКОВИХ НЕЙРОННИХ МЕРЕЖ

\*Вінницький національний технічний університет, Вінниця, Україна

**Анотація.** У статті проведено аналітичний огляд сучасних проектів з апаратної реалізації спайкових нейромереж. Виділено їх головні недоліки та запропоновано варіант апаратної реалізації оптоелектронної спайкової нейронної мережі, вільний від цих недоліків завдяки використанню оптичних сигналів для зв'язків між нейронами та організації навчання за допомогою апаратних засобів. Оцінено основні технічні параметри запропонованої спайкової нейромережі.

**Ключові слова:** спайкові нейронні мережі, апаратна реалізація, нейрокомп'ютер.

**Аннотация.** В статье проведен аналитический обзор современных проектов по аппаратной реализации спайковых нейросетей. Выделены их главные недостатки и предложен вариант аппаратной реализации оптоэлектронной спайковой нейросети, свободный от этих недостатков благодаря использованию оптических сигналов для связей между нейронами и организации обучения с помощью аппаратных средств. Оценены основные технические параметры предлагаемой спайковой нейросети.

**Ключевые слова:** спайковые нейронные сети, аппаратная реализация, нейрокомпьютер.

**Abstract.** The analytic overview of modern projects on spiking neural networks hardware implementation are conducted in the article. There major disadvantages are highlighted. The variant of optoelectronic spiking neural network hardware implementation is proposed. It is free of these disadvantages by using optical signals for communication between neurons and the organization of learning through hardware. The main technical parameters of the proposed spiking neural network are estimated.

**Keywords:** spiking neural networks, hardware implementation, neurocomputer.

### 1. Вступ

Сьогодні на питання «Які засоби краще розв'язують важкоформалізовані і неформалізовані задачі?» відповідь очевидна – нейрокомп'ютери. Нейрокомп'ютер – це інформаційна система, основним процесорним ядром якої є штучна нейронна мережа (на відміну від мікропроцесора), а основним принципом функціонування є навчання на прикладах (на відміну від програмування) [1,2]. Більшість сучасних нейрокомп'ютерних засобів існують у вигляді програмних або програмно-апаратних реалізацій, але загальновідомим є той факт, що максимум переваг від застосування нейрокомп'ютерів можна отримати саме при їх апаратній реалізації [2]. На даний час, на жаль, ще не створено ефективною апаратної реалізації нейрокомп'ютера. Ефективною вважається така апаратна реалізація нейрокомп'ютера, яка містить максимально можливу кількість нейронів (в ідеалі – близьку до кількості нейронів у мозку людини:  $5 \times 10^{10}$ ) і при цьому займає мінімальний об'єм і споживає мінімум енергії. В останні роки спостерігається впевнений перехід від традиційних нейронних мереж на основі бінарних та аналогових нейронів з потенційними сигналами до так званих спайкових нейромереж з імпульсними сигналами [3, 4]. Спайкові нейромережі є більш подібними до своїх біологічних прототипів, а тому мають більші потенційні можливості в досягненні адекватного відтворення інтелектуальних функцій мозку. Реалізація подібних мозку нейрокомп'ютерів дозволить вирішити дві взаємно пов'язані задачі: 1) створення «розумних» комп'ютерів для виконання складних когнітивних неформалізованих завдань; 2) розкриття

таємниць роботи мозку шляхом його зворотного конструювання технічними засобами. Найближчою стратегічною метою є розробка нейроморфних ядер (чипів апаратних нейромереж), які зацікавлені дослідники зможуть використовувати для перевірки своїх власних гіпотез і теорій щодо принципів роботи кори мозку і для побудови на їх основі різноманітних нейрокомп'ютерних засобів для практичних застосувань.

*Мета цієї статті* – проаналізувати сучасні проекти з апаратної реалізації спайкових нейромереж, визначити їх переваги та недоліки й запропонувати шляхи усунення цих недоліків.

## 2. Обґрунтування вибору спайкових нейронних мереж для побудови операційного блоку нейрокомп'ютерів

В останні роки спостерігається інтенсифікація наукових досліджень у галузі спайкових нейронних мереж (англ. Spiking neural networks). Докладно це питання розглядається в [4, 5], але, якщо коротко, то, по-перше, ці мережі є універсальними і їх структура не залежить від розв'язуваної задачі, як у традиційних мережах зі статичними сигналами. Тобто не потрібно цілого «зоопарку» нейронних мереж для задоволення всіх практичних потреб.

Структурно-функціональна модель спайкової нейронної мережі, запропонована в [3], будується, на відміну від проблемно-орієнтованих мереж, на принципах створення динамічних систем у комбінації зі статистичною теорією навчання. Структура спайкової нейронної мережі зображена на рис.1 і містить вхідні нейрони (вхідний шар), інтернейрони (проміжковий шар, або мікромережа) та вихідні нейрони (вихідний шар). Така структура перекликається зі структурою мозку, де є аферентні (сенсорні), проміжкові і еферентні (ефекторні) нейрони. Кількість і ваги синаптичних зв'язків кожного нейрона в такій мережі вибираються на основі даних нейрофізіологічних досліджень (тобто за аналогією з біологічними нейронними мережами). Випадковість вибору нейронів, зв'язаних з будь-яким нейроном у мережі, приводить до виникнення багатоконтурних зворотних зв'язків, тобто такі спайкові нейронні мережі є рекурентними. Нейрони в такій мережі повинні бути двох типів: 1) збуджувальні (видають на виході електричний імпульс додатної полярності) та 2) гальмівні (видають на виході електричний імпульс від'ємної полярності).

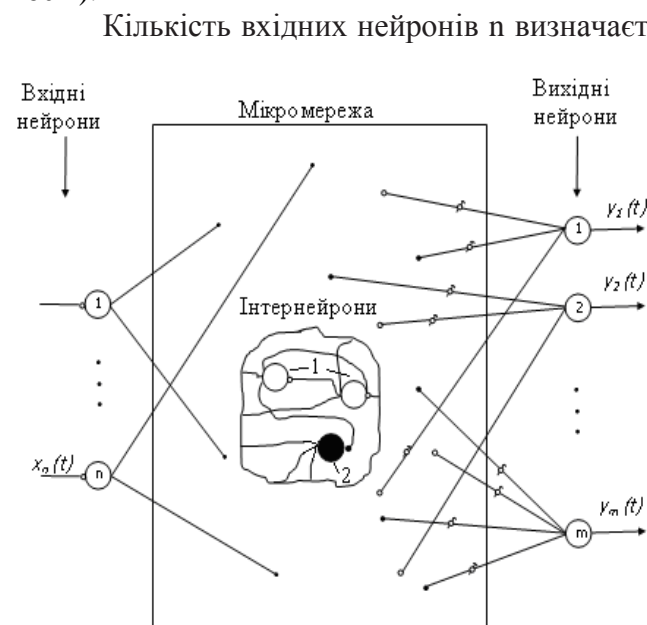


Рис. 1. Структура спайкової нейронної мережі: 1 – збуджувальні нейрони, 2 – гальмівні нейрони

Кількість вхідних нейронів  $n$  визначається розв'язуваною задачею. Кількість вихідних нейронів  $m$  дорівнює кількості можливих класів вхідних образів. Кількість зв'язків одного нейрона з іншими (наприклад, з 10...15 % інших), дальність їх поширення, співвідношення кількості збуджувальних і гальмівних зв'язків (наприклад, відповідно 80 % і 20 %) може визначатися даними нейрофізіологічних досліджень [3]. Як видно зі структури спайкової нейронної мережі (рис. 1), лише вихідні нейрони повинні мати можливість підстроювати свої ваги синаптичних зв'язків у процесі навчання. Це відрізняє вихідні нейрони від інтернейронів, які мають постійні ваги зв'язків, що задаються при формуванні мережі.

Спайкові нейронні мережі, завдяки нейроморфності, мають перед

традиційними, крім універсальності, ще й такі переваги [3–5]:

- 1) розпізнавання динамічних образів (мова, динамічні зображення та ін.);
- 2) багатозадачність (інформація про вхідні потоки циркулює в рекурентній нейронній мережі і на вихід одночасно можуть бути подані результати різних задач за допомогою різних груп вихідних нейронів, навчених на виконання тієї чи іншої задачі);
- 3) розпізнавання з передбаченням (будь-який динамічний процес може бути розпізнаний навіть по неповній інформації про нього, тобто навіть раніше, ніж він закінчиться);
- 4) простота процедури навчання (навчаються не всі нейрони мережі, а тільки вихідні зчитувальні нейрони);
- 5) підвищена продуктивність обробки інформації та завадостійкість завдяки імпульсному представленню інформації.

Саме завдяки цим перевагам спайкових нейронних мереж вони є найперспективнішою архітектурою для побудови операційного ядра нейрокомп'ютерів.

### 3. Огляд відомих проектів з апаратної реалізації спайкових нейронних мереж

Розглянемо, які є на теперішній час найпотужніші проекти в галузі створення (апаратної реалізації) спайкових нейромереж. Їх є багато в різних країнах світу, але можна виділити п'ять основних, представлених у табл. 1.

Найбільш відомим є Human Brain Project [6] (табл. 1), який почався у 2005 р. під назвою Blue Brain Project, а з 2012 р. має назву Human Brain Project. Проект присвячений моделюванню фактичної роботи мозку на цифрових суперкомп'ютерах. Моделювання охоплює різні рівні організації мозку – від окремих нейронів до кори головного мозку в цілому. Мета проекту полягає в тому, щоб розгадати принципи роботи мозку і секрети виникнення складних когнітивних функцій мозку із, здавалося б, простого обміну імпульсами (спайками) між групами нейронів, і, як наслідок, в отриманні нових інформаційних технологій безпосередньо з архітектури мозку. У даному проекті використовується суперкомп'ютер «Blue Gene» продуктивністю 22,8 тріллионів оп/с з програмним забезпеченням «NEURON». Хоча проект і не стосується безпосередньо проблем ефективного апаратної реалізації спайкових нейромереж, все ж він базується саме на застосуванні спайкових нейронів та структур нейронів, аналогічних біологічним прототипам, і має велике значення для апробації принципів структурної організації та методів навчання спайкових нейронних мереж, які потім можна буде використати при побудові апаратних реалізацій.

До недоліків та слабких місць проекту можна віднести:

- 1) це комп'ютерне моделювання, а не апаратне;
- 2) низька швидкодія моделювання: 1 секунда роботи реального нейрона забирає 80 хвилин машинного часу (для моделювання людського мозку в цілому потрібен комп'ютер у тисячу разів потужніший, ніж найпотужніші суперкомп'ютери, доступні сьогодні);
- 3) велика споживана потужність: на детальне моделювання 1 нейрона потрібно близько 200 Вт, тому моделювання цілого мозку є проблемою;
- 4) через великі потреби в ресурсах комп'ютера та велику споживану потужність детально моделюються лише групи нейронів, які є активними.

Проект SpiNNaker [7] (табл. 1) присвячений суто апаратній реалізації спайкових нейромереж за допомогою добре розвинутої на сьогоднішній день елементної бази мікропроцесорів. У проекті використовуються багатоядерні мікропроцесори (МП) ARM968 (в 1 чипі 20 МП, причому 1 МП моделює 1000 нейронів, коштує близько 2\$ та споживає менше 1 Вт). Таким чином, маємо 20000 нейронів на чіп. На кінець 2011 року було об'єднано у систему 50 чіпів. Це близько 1 млн (950 тис.) нейронів:  $10^6$ .

Таблиця 1. Найпотужніші проекти в галузі створення слайкових нейромереж

Назва проекту	Human Brain Project (до цього Blue Brain Project)	Spinnaker	SyNAPSE (DARPA)	Neurogrid (Brains in silicon group)	BrainScaleS (продовж. FACETS 2005-2010)
Розшифровка	Human Brain Project	Spiking Neural Network Architecture	Systems of Neuromorphic Adaptive Plastic Scalable Electronics	Grid technology for neurosciences	Brain-inspired multiscale computation in neuromorphic hybrid systems
Країна	Швейцарія	Великобританія	США	США	Євросоюз
Організація	ІВМ та Швейцарський Федеральний Технічний Інститут Лозанни (EPFL).	Манчестерський ун-т, Університет Саутгемптона	1) ІВМ, 2) провідні університети (Колумбійський та ін.), 3) Hughes Research Labs, HP	1) ІВМ, 2) Стенфордський ун-т (Stanford University) та ін.	18 дослідницьких груп (у т.ч. 13 університетів) з 10 європейських країн
Керівник	Проф. Генрі Маркрам (Henry Markram)	Професори Стів Фарбер (Steve Furber) та Ендрю Браун (Andrew Brown)	Дхармендра Мода (Dharmendra Modha), ІВМ	Проф. Кwabена Боен (Kwabena Boahen), ун-тет Стенфордський ун-т	Проф. Карлхайнц Майер (Karlheinz Meier), ун-тет Гейдельберг, Німеччина
Джерело фінансів	Швейцарський уряд та грант Євросоюзу	EPSRC (Engineering and Physical Sciences Research Council)	1) DARPA (Defense Advanced Research Projects Agency), 2) ІВМ	NSF (National Science Foundation), NIH (National Inst of Health)	Грант Євросоюзу
Обсяг фінансування	1 млрд євро на 2012–2023	£2.5m на 2005–2010 рр. і ще £2.5m на продовж	2009–2013 – Σ:\$102,6 млн	\$4,9 млн	€ 8,5 млн спочатку, плюс € 700 000 у продовження
Початок	2005-2008 Blue Brain Prj, 2008–2012 – C2S2. 3 2012 р. – Human Brain Prj	2005	2008	2006	2005–2010 – FACETS. 3 2011 – BrainScaleS
Закінчення	2023	2014...	2016	2013	2014
Електронна база	Суперкомп'ютер «Blue Gene» з програмним забезпеченням «NEURON», яке написано Майклом Хайнсом (Michael Hines)	Мікропроцесори (МП) ARM968 (130 нм, 1 чіп – 20 МП, 1 МП=1000 нейронів). 20000 нейронів на чіп	Цифрові НВІС (45 нм). У чіпі (4,2 мм <sup>2</sup> ) 256 слайкових нейронів на основі цифрових суматора та компаратора. Один нейрон 35x95 μm <sup>2</sup> (для порівняння біологічний нейрон 4-100 μm в Ø)	Аналогові НВІС на основі к-МДН. Нейроморфний чіп Neucosoc містить 65536 (256x256) штучних нейронів на площі 11,9x13,9 мм <sup>2</sup>	Аналогові НВІС (180 нм) на основі к-МДН. Один чіп (5x10 мм <sup>2</sup> ) реалізує до 512 слайкових нейронів. Нейрони аналогові, а синапси і сигнали зв'язку між чипами цифрові
К-ть нейронів на сьогодні	У 2011 – мережа зі 100 колонок неокортекса з 1 млн нейронів	На середину 2012 року – 50 чіпів, у кожному чіпі по 20 МП (1 МП=1000 нейронів). Це 1 млн (864 тис.) нейронів	До 2013 р. 1 млн нейронів	Прототип Neurogrid складається з 16 "нейродер" (65536 штучних нейронів кожне), тобто містить 1 млн нейронів	Кожна підкладка Si Ø20 см містить 384 чіпи, у чіпі до 512 слайкових нейронів (близько 200 000 нейронів на підкладку)

Назва проекту	Human Brain Project (продовж. Blue Brain Project)	Spinnaker	SyNAPSE (DARPA)	Neurogrid (Brains in silicon group)	BrainScaleS (продовж. FACETS 2005-2010)
К-ть нейронів у перспективі	До 2014 р. планується змоделювати мозок пацюка (100 млн нейронів). Мозок людини (100 млрд нейронів) – до 2023 р.	У перспективі 500 чіпів і наразі 5000 чіпів (10 <sup>8</sup> нейронів – майже 1% мозку людини) до 2014	Довготривала мета створення системи з 10 <sup>10</sup> нейронів і 10 <sup>14</sup> синапсів, що займає об'єм < 2 літрів та споживає 1 кВт	Наступне покоління Neurogrid повинно мати вже 64 млн кремнієвих нейронів, що приблизно відповідає мозку миші	1,6 млн нейронів і 400 млн синапсів (підкладки каскадуються)
К-ть синапсів	3×10 <sup>7</sup> синапсів (2006 р.). 3000 синапсів на 1 нейрон	1000 синапсів на 1 нейрон	1280 бінарних синапсів/нейрон (262 тис. незмінних та 6536 змінних)	6 млрд синапсів (~6000 синапсів на 1 нейрон)	256 синапсів на 1 нейрон (49 млн синапсів на підкладку)
Модель нейронів	Пірамідальний нейрон	Слайковий нейрон Іжикевича	LIF (leaky integrate-and-fire neuron)	Quadratic integrate-and-fire neuron (QI&F)	Exponential integrate and fire neuron (AdExp)
Тип мережі	Неокортикальні колонки	Повнзв'язна	Hopfield-like, Restricted Boltzmann machine (RBM)	2-D рекурентна мережа зі ступенем зв'язності «мексиканський капелюх»	Задається програмно
Продуктивність	1,25x10 <sup>4</sup> імпл/с (при продуктивності комп. 22,8x10 <sup>12</sup> оп/с)	Час передачі імпульсу 1 мс, 10 <sup>9</sup> імпл/с	Дискретний часовий крок 1 мс, 10 <sup>9</sup> імпл/с	9,1x10 <sup>7</sup> імпл/с	10 <sup>11</sup> імпл/с
Слож. потужність	<17 Вт/нейрон>	1 Вт на 1000 нейронів (1 мВт/нейрон)	0,5 мВт/нейрон при 0,53В пороговій напр. МДН	3,1 Вт на 10 <sup>6</sup> нейронів 3,1 мкВт/нейрон	1 кВт на підкладку, тобто 5 мВт/нейрон
Недоліки	Це комп'ютерна, а не апаратна реалізація	Немає відокремлених нейронів	Синапси бінарні	Обмежена зв'язність («мексикан. капелюх»)	Не буває на підкладці 100% придатних чіпів
Відеутні безпосередні зв'язки між нейронами, оскільки технологічно неможливе створення великої кількості електричних ліній зв'язку в площині напівпровідникового кристалу. Безпосередні зв'язки замінюються організацією штучних програмованих чи спеціально кодованих протоколів обміну між групами нейронів, які дозволяють зменшити кількість електричних ліній зв'язку у площині напівпровідникового кристалу, але порушуються принципи роботи біологічного мозку					
Синхронні системи передачі імпульсів (у мозку – асинхронна передача)					
Навчання нейронної мережі відбувається за допомогою цифрових комп'ютерів та спеціального програмного забезпечення, а не за допомогою власних непрограмних механізмів і засобів, не пов'язаних з обчисленнями					
1 с реального нейрона=80 хвилин моделювання			Нейрони цифрові		
Велика споживана потужність			Велика споживана потужність		

Системна архітектура SpiNNaker виконана у вигляді двовимірної тороїдальної мережі (рис. 2) мікропроцесорів (Chip Multiprocessor, CMP), зв'язаних між собою за допомогою Ethernet. Ця мережа має трикутні комірки, щоб можна було утворювати шляхи для обходу дефектних вузлів. Кожен вузол (рис. 3) CMP складається з двох чипів: система на кристалі SpiNNaker MPSoC та гігабітна мікросхема пам'яті DDR SDRAM. У свою чергу, кожен чип SpiNNaker MPSoC містить 20 ядер ARM968, зв'язаних в мережу двома (один на вхід, інший на вихід) контролерами NoC (мережевий модуль Network-on-Chip, NoC) з внутрішньою системою синхронізації.

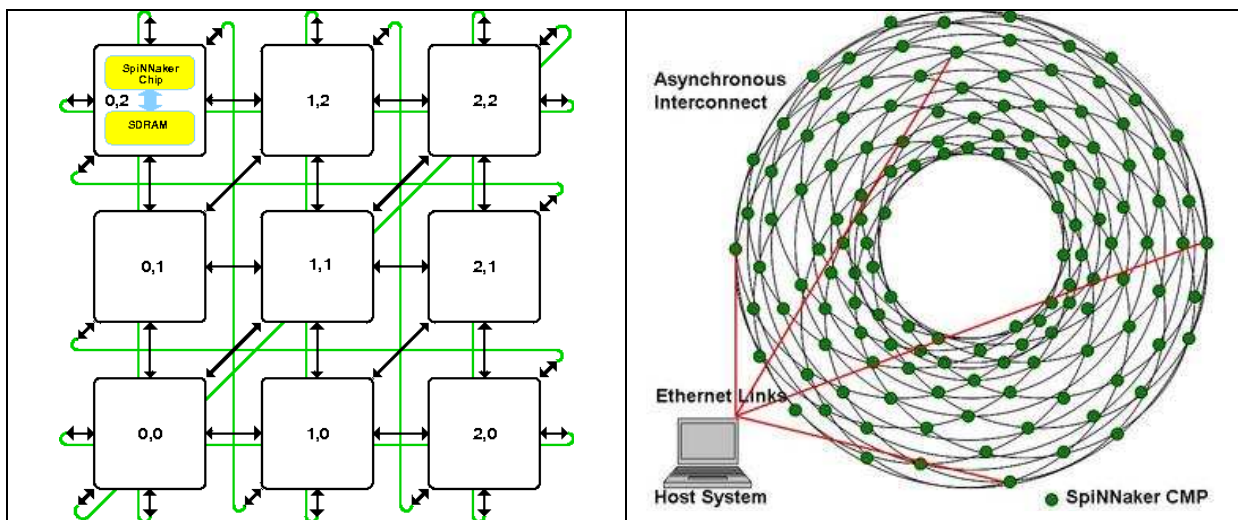


Рис. 2. Системна архітектура SpiNNaker

Проблема узгодженої роботи сотень тисяч вузлів мережі SpiNNaker ускладнена тим, що це однорідна система з високим ступенем розподілу, яка не має ніякого явного способу синхронізації. Проблема вирішується шляхом виділення одного з 20 ядер вузла і передачі йому функцій управління і моніторингу процесу. У процедурі обміну даними між вузлами в мережі використовуються багатоадресні повідомлення mc (multicast).

Кожний МП має 64 кБ кеш-пам'яті для даних і 32 кБ кеш-пам'яті для інструкцій, а також комунікаційний контролер, який буде імітувати нервові імпульси за допомогою пакетів даних. Також пристрій має контролер пам'яті, який дозволить взаємодіяти з 1 ГБ пам'яті, розташованої в корпусі чипа. У системі SpiNNaker імітовані нейрони зможуть передавати імпульси до будь-яких інших імітованих нейронів у межах системи, що містить мільйон процесорів. При цьому час передачі імпульсу складе близько однієї мілісекунди, що приблизно відповідає швидкості передачі нервових імпульсів реальними нейронами в мозку людини.

Система SpiNNaker все-таки не є мережею штучних нейронів, а є масово-паралельною багатопроцесорною обчислювальною системою типу MIMD. Вона містить до 1036800 ARM9 ядер і 7 ТБайт оперативної пам'яті, розподіленої по всій системі. Ці факти і визначають її недоліки та слабкі місця:

1) у цій системі немає штучних нейронів як таких. Нейрон замінено мікропроцесором, який виконує обчислення за формулами (використовується математична модель нейрона Іжикевича). Тому важко говорити про нейроморфність і очікувати допомоги у розгадці таємниць мозку шляхом його зворотного конструювання;

2) у біологічному мозку при передачі інформації між нейронами відсутня синхронізація (тактування), тобто це є асинхронна система. Але у SpiNNaker існує синхронізація. Всередині вузли працюють за традиційними для мікропроцесорів синхронними принципа-

ми. Між собою вузли обмінюються мікропакетами у квантовані проміжки часу. Це знову ж таки не говорить про нейроморфність і допомогу у розгадці таємниць мозку шляхом його зворотного конструювання;

3) низька швидкодія: час передачі імпульсів (1 мс) хоч і відповідає швидкодії реальних нейронів, але від надпродуктивної електронної елементної бази варто чекати більшого;

4) відносно велика споживана потужність: 1 мВт на нейрон (для моделювання мозку ( $5 \times 10^{10}$  нейронів) знадобиться 50 мегаВт).

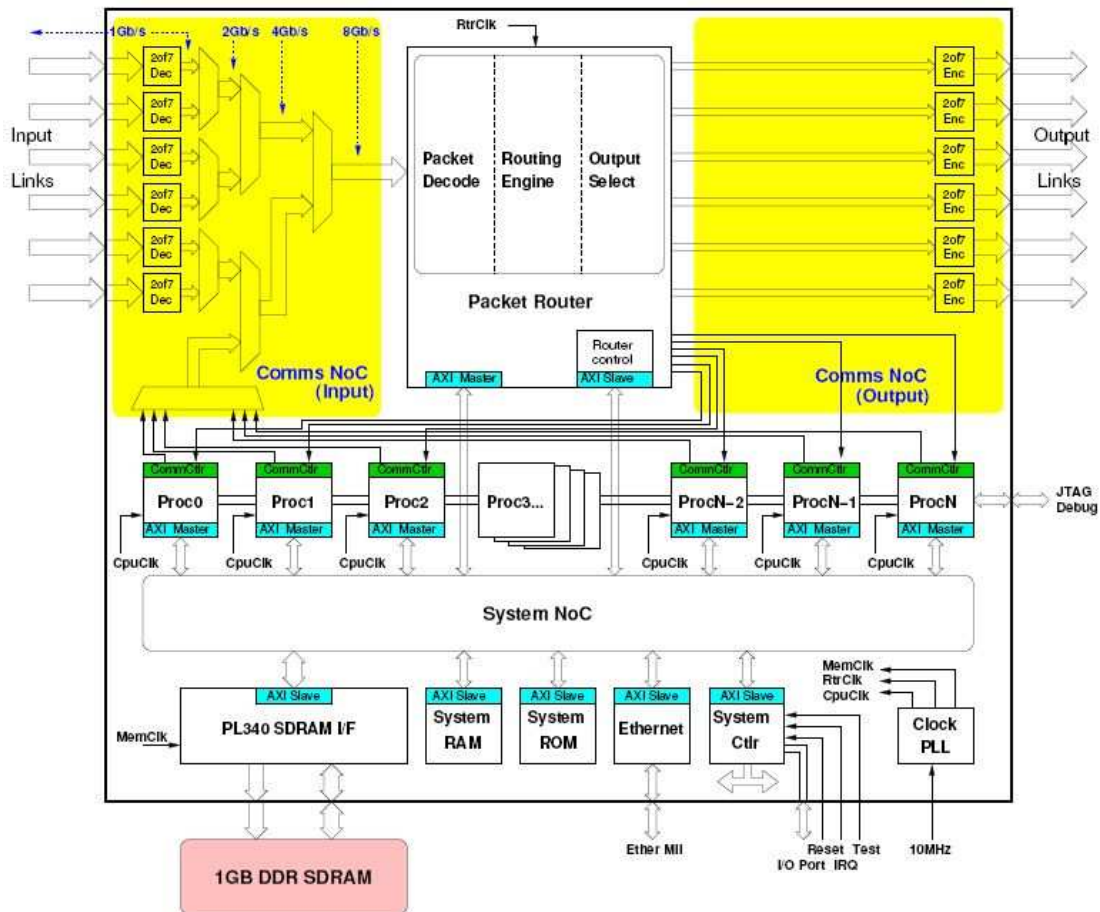


Рис. 3. Схема вузла системи SpiNNaker

Метою програми SyNAPSE (DARPA) [8] (табл. 1) є створення технології побудови електронних нейроморфних машин, які відповідали б масштабам біологічних прототипів. З цією метою було розроблено та виготовлено ключові будівельні блоки модульної нейроподібної архітектури – нейросинаптичні ядра (рис. 4), кожне з яких має 256 спайкових (integrate-and-fire) нейронів (на цифрових елементах) і решітку  $1024 \times 256$  біт SRAM пам'яті для організації синапсів. Для цього було використано НВІС-технологію «кремній-на-ізоляторі» фірми ІВМ з технологічною нормою 45 нм. Повністю цифрова реалізація забезпечує однозначну відповідність між апаратним і програмним забезпеченням та має ультра-низьку споживану потужність: 45 пДж/імпульс. Створене нейросинаптичне ядро є повністю налаштованим з точки зору параметрів нейронів (рис. 5), типів аксонів і ваг синапсів, а тому потенційно має широкий спектр застосування.

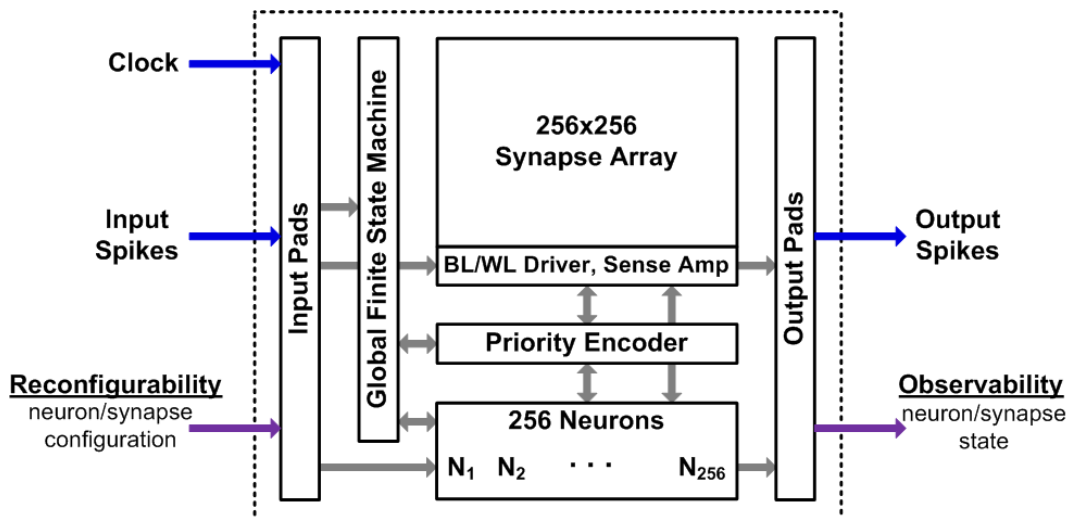


Рис. 4. Схема нейросинаптичного ядра системи SyNAPSE

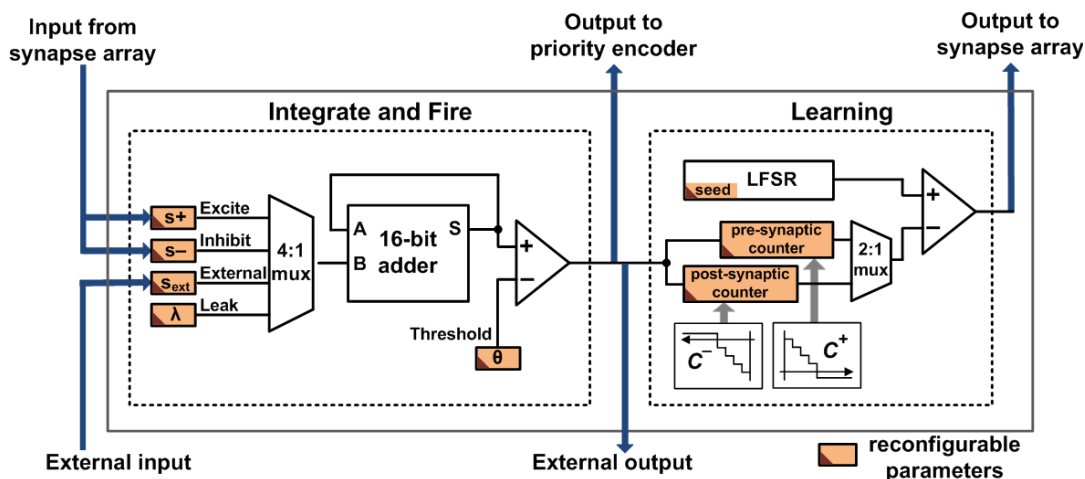


Рис. 5. Схема цифрового к-МДН нейрона системи SyNAPSE, який має засоби реконфігурації

Система SyNAPSE (DARPA), порівняно із системою SpiNNaker, у більшій мірі адекватна біологічним нейронним мережам, тому що складається із штучних нейронів (нейроелементів), а не з процесорів. Але неадекватним біологічним прототипам залишається організація міжнейронного обміну спайками (технологія address-event representation) та організація початкового завантаження параметрів і навчання, які відбуваються за посередництвом цифрового комп'ютера через USB-порт за допомогою спеціальних програм. Технологія «address-event representation» передбачає, що імпульси, згенеровані на аксонах нейронів у поточному часовому інтервалі (~1 мс), будуть передані на відповідні дендрити інших нейронів у наступному часовому інтервалі.

Зазначені факти визначають такі недоліки та слабкі місця системи SyNAPSE (DARPA):

1) штучні нейрони в цій системі виконано на цифровій елементній базі, що спрощує їх технологічну реалізацію, але не відповідає організації природних нейронів, які є аналоговими системами, що працюють за принципами накопичення аналогових іонних струмів та порогового спрацьовування. Це знижує адекватність системи біологічним прототипам та зменшує шанси на достовірність моделювання за їх допомогою механізмів роботи мозку;



2) значення синаптичних коефіцієнтів є бінарними, що не відповідає біологічним прототипам, де синаптичні коефіцієнти мають неперервні значення;

3) використовує програмоване віртуальне з'єднання нейронів. У біологічному мозку при передачі інформації між нейронами відсутня синхронізація (тактування), тобто це є асинхронна система. Але у SyNAPSE існує синхронізація при передачі імпульсів (імпульси передаються у тактовані проміжки часу із затримкою на 1 такт). Це знову ж таки не говорить про нейроморфність SyNAPSE і зменшує шанси на достовірність моделювання за допомогою SyNAPSE механізмів роботи мозку;

4) низька швидкодія: інтервал передачі імпульсів 1 мс хоч і відповідає швидкодії реальних нейронів, але від надпродуктивної електронної елементної бази варто чекати більшої швидкодії.

Проект NeuroGrid [9] (табл. 1) за метою та підходом схожий до програми SyNAPSE (DARPA). Тобто його метою також є створення технології побудови електронних нейроморфних машин, які відповідали б масштабам біологічних прототипів. Так само в NeuroGrid було розроблено та виготовлено ключові будівельні блоки модульної нейроподібної архітектури – нейросинаптичні ядра (рис. 6 а). Але різниця полягає в тому, що нейрони побудовані не на цифрових елементах (суматорах та лічильниках), а на аналоговій елементній базі з використанням к-МДН-транзисторів (рис. 7). Нейроморфний чип NeuroCore (рис. 6 а) містить 65536 (256x256) штучних нейронів на площі 11,9x13,9 мм<sup>2</sup>. Прототип Neurogrid складається з 16 таких "нейроядер", які зібрані на платі 6,5x7,5 дюймів (рис. 6 б), тобто містить 1 мільйон нейронів. Ці нейроядра з'єднані у мережу вигляду «бінарне дерево» і мають пікову швидкість 91 Мегазв'язків/с. Імпульси передаються в таблицю маршрутизації (дочірня плата на рис. 6 б), яка підтримує програмовані зважені з'єднання. Тобто імпульси передаються від чипа до чипа не безпосередньо, а у спосіб, що передбачає відправлення адреси, яка вказує, в якому чипі нейрон якого рядка та стовпця згенерував імпульс і в який момент часу. Вся меганейронна система споживає 3,1 Вт.

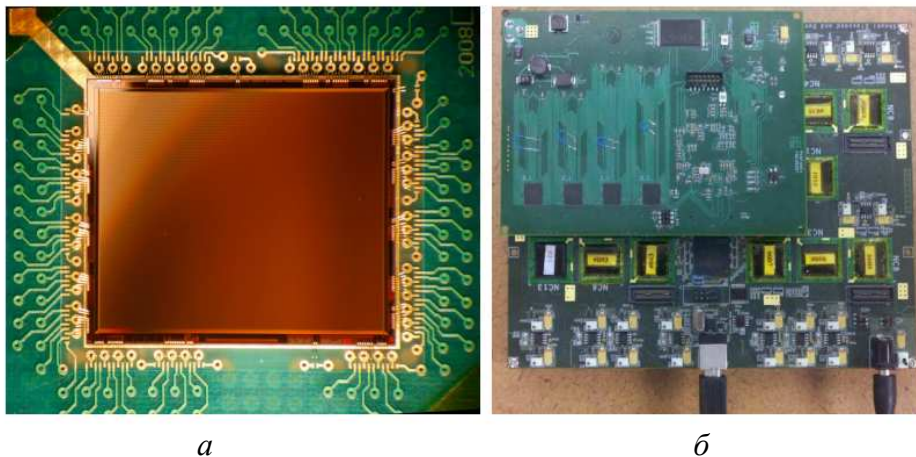


Рис. 6. Вид чипа нейросинаптичного ядра NeuroCore з 256x256 нейронами (а) та плата NeuroGrid, яка містить 16 ядер NeuroCore (б)

Перевагами системи NeuroGrid є велика кількість нейронів у чипі і низька питома споживана потужність, але недоліки та слабкі місця полягають в такому:

1) використовує програмоване віртуальне з'єднання нейронів, тобто імпульси передаються від нейрона до нейрона не безпосередньо, а у спосіб, що передбачає відправлення адреси, яка вказує в якому чипі нейрон якого рядка та стовпця згенерував імпульс і в який момент часу. Це не говорить про нейроморфність NeuroGrid і зменшує шанси на достовірність моделювання за допомогою NeuroGrid механізмів роботи мозку;

2) ініціалізація початкових значень ваг синапсів та зміна їх значень при навчанні здійснюються за допомогою цифрового комп'ютера та спеціального програмного забезпечення. Тобто відсутнє так зване on-line on-chip навчання, що знижує нейроморфність системи NeuroGrid. До речі, цей недолік властивий і всім попереднім розглянутим системам;

3) хоча кількість нейронів у чипі величезна, мережа нейронів не є вільнозв'язна, а є 2-D мережею зі ступенем зв'язності «мексиканський капелюх» (локальне збудження і дистальне гальмування), що не являється універсальним, а тому знижує функціональні можливості.

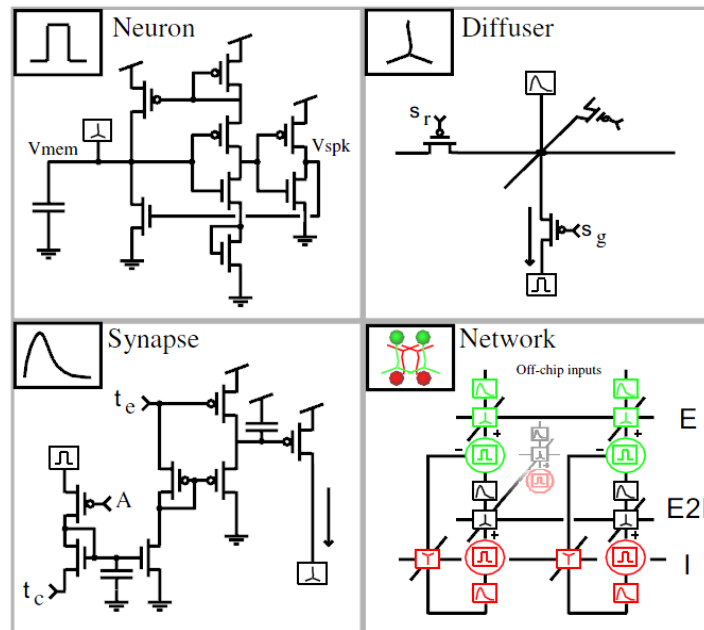


Рис. 7. Транзисторна реалізація в системі NeuroGrid аналогового к-МДН нейрона, дифузера та синапса. Схема «Network» показує, як ці окремі блоки організовані в нейронну мережу

Проект BrainScaleS [10] (табл. 1) має на меті зрозуміти принципи обробки інформації в мозку на різних рівнях, починаючи від окремих нейронів до цілих функціональних ділянок мозку. Дослідження включають в себе три підходи: (1) in vivo біологічні експерименти; (2) моделювання на суперкомп'ютерах з продуктивністю порядку  $10^{15}$  Flops; (3) побудова нейроморфних процесорів. Завдання полягає у тому, щоб визначити загальні теоретичні засади функціонування мозку і використовувати ці знання для побудови штучних когнітивних систем.

Нейроподібне апаратне забезпечення реалізується на аналогових НВІС (180 нм) на основі к-МДН масштабу всієї підкладки. Кожна підкладка кремнію діаметром 20 см (рис. 8) містить 384 чипи, кожен з яких реалізує 128 000 синапсів і до 512 спайкових нейронів. В сумі це дає близько 200 000 нейронів і 49 мільйонів синапсів на підкладку. Нейрони – аналогові, а синапси і сигнали зв'язку між чипами – цифрові. НВІС нейрони працюють значно швидше, ніж біологічні оригінали. Це дозволяє моделювати нейронні мережі, які працюють в десятки тисяч разів швидше, ніж реальний час. Підкладка підтримується алюмінієвою пластиною, яка також служить для відведення тепла. Багатошарова друкована плата встановлюється на поверхню підкладки і служить інтерфейсом вхід/вихід для нейронної схеми. Більші мережі можна побудувати з'єднанням декількох підкладок у модуль.

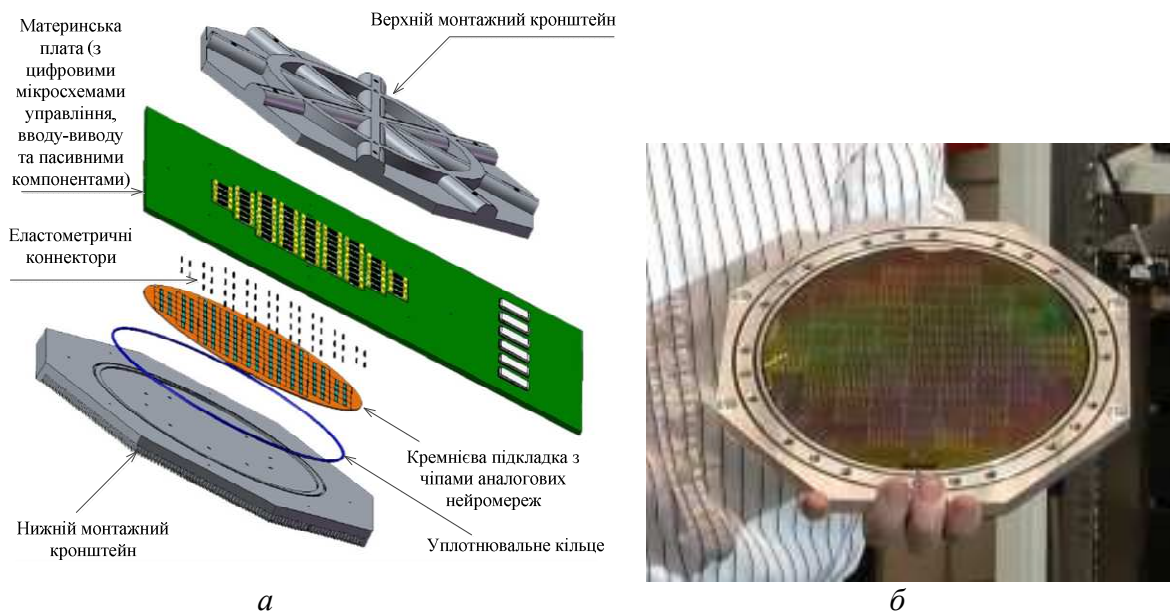


Рис. 8. Конструкція (а) та вигляд (б) модуля підкладки нейроморфних чипів системи BrainScaleS

Зазначені факти визначають такі недоліки та слабкі місця системи BrainScaleS:

1) імпульси передаються від нейрона до нейрона не безпосередньо, а у спосіб, що передбачає відправлення адреси, яка вказує, в якому чипі нейрон якого рядка та стовпця згенерував імпульс і в який момент часу. Це не говорить про нейроморфність BrainScaleS і зменшує шанси на достовірність моделювання за допомогою NeuroGrid механізмів роботи мозку;

2) ініціалізація початкових значень ваг синапсів та зміна їх значень при навчанні здійснюється за допомогою цифрового комп'ютера та спеціального програмного забезпечення. Тобто відсутнє так зване on-line on-chip навчання, що також знижує нейроморфність системи BrainScaleS;

3) майже не буває при виробництві напівпровідникових підкладок випадку, коли 100% чипів є якісними і функціонально бездоганними, що знижує фактичне число нейронів і ускладнює роботу з підкладкою.

#### 4. Недоліки відомих проектів з апаратної реалізації спайкових нейронних мереж та шляхи їх подолання

Як видно з аналітичного огляду відомих проектів, всі вони мають два головних недоліки:

1) відсутні безпосередні зв'язки між нейронами, оскільки технологічно неможливе створення великої кількості електричних ліній зв'язку у площині напівпровідникового кристалу. Безпосередні зв'язки замінюються організацією штучних програмованих чи спеціально кодованих протоколів обміну між групами нейронів, які дозволяють зменшити кількість електричних ліній зв'язку у площині напівпровідникового кристалу, але при цьому порушуються принципи роботи біологічного мозку;

2) навчання нейронної мережі відбувається за допомогою цифрових комп'ютерів та спеціального програмного забезпечення, а не за допомогою власних непрограмних механізмів і засобів, не пов'язаних з обчисленнями.

Якщо усувати перший недолік шляхом формування ліній безпосереднього електричного зв'язку між електронними схемами нейронів, то займаний ними простір, з урахуванням простору, необхідного для ізоляції однієї лінії зв'язку від іншої, може стати настільки великим, що на пластині кремнію не залишиться місця для розміщення самих схем нейроелементів. Водночас, сполучення нейроелементів за допомогою світлових променів не ви-

магає ізоляції між сигнальними шляхами, світлові потоки можуть проходити один через інший без взаємного впливу. Більш того, сигнальні шляхи можуть бути розташовані у трьох вимірах. Щільність оптичних шляхів передачі обмежена тільки розмірами джерел світла, їх дивергенцією і розмірами детектора. Крім того, всі сигнальні шляхи можуть працювати одночасно, тим самим забезпечуючи величезний темп передачі даних.

Тому перший недолік можна виправити за допомогою таких архітектур спайкових нейромереж, які розраховані на оптоелектронну реалізацію, при якій зв'язки між нейронами будуть здійснюватись за допомогою оптичних сигналів. У перспективі можливо буде використовувати і наноелектронну елементну базу при винайденні способів тривимірного конструювання наноелементів. Другий недолік може бути виправлений шляхом розробки таких архітектур спайкових нейромереж, які мають у своєму складі засоби для організації навчання власними схемотехнічними засобами без використання обчислювальних процедур (за аналогією з біологічними нейромережами).

### 5. Архітектура апаратної реалізації оптоелектронної спайкової нейронної мережі

Структура оптоелектронної спайкової нейронної мережі (СНМ), запропонована автором [11, 12], представлена на рис. 9. Вона складається з  $n$  вхідних нейронних елементів  $EI_1 \dots EI_n$ , на які поступають вхідні сигнали  $x_1 \dots x_n$ ,  $N$  інтернейронів  $E_1 \dots E_N$ , матриці світловипромінювачів (світлодіоди або напівпровідникові лазери), транспаранта  $T$  (виконаного, наприклад, у вигляді фотопластини, коефіцієнти прозорості локальних ділянок якої відповідають значенням вагових коефіцієнтів зв'язку інтернейронів), оптично керованого транспаранта (ОКТ) [13] із системою з  $2m$  пар смугових вертикальних електродів,  $m$  вихідних нейронних елементів  $EO_1 \dots EO_m$ . У ОКТ непарні пари електродів відповідають збуджувальним (Excitatory) зв'язкам і мають виводи  $Y_i^E$  і  $Y_i^{E'}$ , а парні пари електродів відповідають гальмівним (Inhibitory) зв'язкам і мають виводи  $Y_i^I$  і  $Y_i^{I'}$  ( $i=1 \dots m$ ). Транспарант  $T$  і ОКТ утворюють загальну матрицю вагів зв'язків усіх нейронів СНМ.

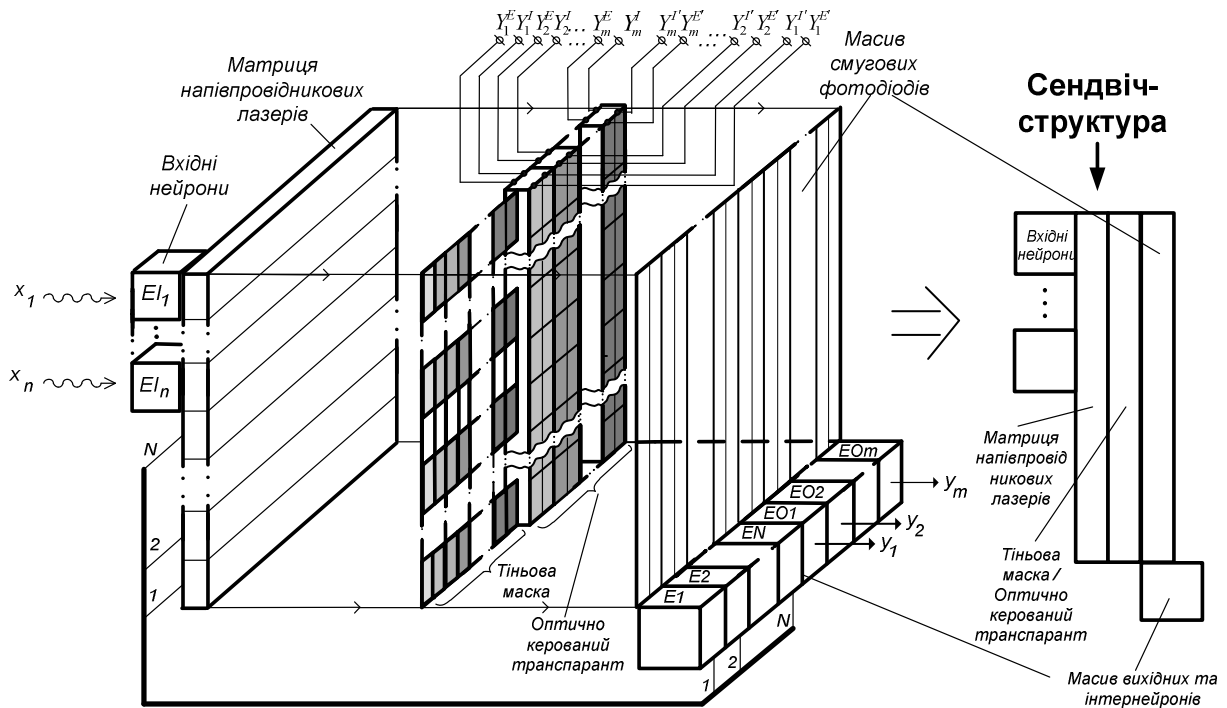


Рис. 9. Структура оптоелектронної спайкової нейронної мережі

Запропонована оптоелектронна спайкова нейронна мережа працює таким чином.

У режимі розпізнавання вхідні оптичні сигнали  $x_1(t) \dots x_n(t)$  поступають на входи відповідних сенсорних нейронів  $EI_1 \dots EI_n$ , які перетворюють ці сигнали у послідовності імпульсів  $\delta(t-t_i)$  з частотою, пропорційною амплітуді сигналу  $x_i(t)$ . Імпульси мають фіксовану амплітуду і тривалість та керують одразу цілим рядком матриці світловипромінювачів, з виходу яких оптичні імпульси надходять на  $i$ -й рядок встановлених поряд транспарантів Т і ОКТ. У режимі розпізнавання між всіма електродами  $Y_i^E$  і  $Y_i^{E'}$ , також, як і між всіма електродами  $Y_i^I$  і  $Y_i^{I'}$  ( $i=1 \dots m$ ) – нульові різниці потенціалів, при яких коефіцієнти пропускання локальних областей ОКТ, встановлені в режимі навчання, залишаються незмінними. Оптичні імпульси, проходячи через комірки транспарантів Т і ОКТ, що мають різні коефіцієнти пропускання, рівні вагам синаптичних зв'язків нейронів, змінюють свою амплітуду (відбувається множення амплітуди імпульсу на вагу зв'язку). Імпульси, що пройшли через транспаранти Т і ОКТ, оптично збираються (просторово інтегруються) за допомогою масиву смугових фотодіодів зі стовпців цих транспарантів. Масив смугових фотодіодів формує електричні сигнали, які керують світло випромінювачами інтернейронів  $E_1 \dots E_N$  (вектор  $XN$ ) і подаються на них за допомогою джгута. В інтернейронах сигнали вектора  $XN$  (серед яких є і збуджувальні і гальмівні сигнали) піддаються часовій інтеграції, внаслідок чого формуються сигнали  $YN$ , які знов поступають через транспаранти Т і ОКТ на входи інтернейронів  $E_1 \dots E_N$ . Функціонування здійснюється повністю асинхронно, на відміну від традиційних нейронних мереж на формальних нейронах Маккалока-Пітса з потенційними виходами.

Оптичні сигнали, зібрані масивом смугових фотодіодів зі стовпців транспаранта ОКТ, є вхідними сигналами (вектор  $XO$ ) для вихідних нейронів  $EO_1 \dots EO_m$ . Останні здійснюють часову інтеграцію збуджувальних і гальмівних сигналів, що містяться у векторі  $XO$ , і формують вихідні сигнали СНМ (вектор  $YO$ ). Найявність максимально інтенсивної імпульсації на  $j$ -тому вихідному нейроні свідчить про те, що аналізовані вхідні сигнали максимально схожі на  $j$ -тий еталонний образ. По відношенню середньої за період вимірювання частоти імпульсації цього вихідного нейрона до максимально можливої частоти можна судити про міру відповідності аналізованого і еталонного образів.

У режимі навчання [14] необхідно послідовно подавати на входи еталонні набори сигналів  $x_1^s(t) \dots x_n^s(t)$  ( $S=1 \dots s$ ), відповідні кожному з  $S$  еталонів розпізнаваних образів, і відповідним чином міняти керувальні потенціали між електродами  $Y_i^E$  і  $Y_i^{E'}$ , а також між електродами  $Y_i^I$  і  $Y_i^{I'}$  ( $i=1 \dots m$ ). Так, спочатку на входи нейронів  $EI_1 \dots EI_n$  подають сукупність еталонних сигналів першого образу, на електроди  $Y_1^E$  і  $Y_1^{E'}$  подають від блоку управління (на рис. 9 не показаний) позитивну різницю потенціалів  $+U$ ; на електроди  $Y_i^I$  і  $Y_i^{I'}$ , а також на електроди  $Y_i^E$  і  $Y_i^{E'}$  ( $i=1 \dots m, i \neq 1$ ) – нульову різницю потенціалів, а на електроди  $Y_i^I$  і  $Y_i^{I'}$  ( $i=1 \dots m, i \neq 1$ ) – від'ємну різницю потенціалів  $-U$ . Це приводить до того, що оптичні імпульси, що поступають через ОКТ на збуджувальний вхід першого вихідного нейрона  $EO_1$ , будуть збільшувати вагу його зв'язку з тим нейроном, від якого прийшов імпульс (правило Хебба). А імпульси, що поступають на гальмівні входи інших нейронів, збільшують ваги гальмівних зв'язків, що в режимі розпізнавання викликати загальмування цих нейронів неінцидентними їх збудженому стану імпульсами. Ваги зв'язків гальмівного входу першого нейрона і збуджувальних входів інших нейронів залишаються незмінними.

Далі процедуру повторюють для кожного  $s$ -го еталонного образу. При цьому на виході нейронів  $EI_1 \dots EI_n$  подають сукупність еталонних сигналів  $s$ -го образу, на електроди  $Y_s^E$  і  $Y_s^{E'}$  подають від блока управління (на рис. 9 не показаний) позитивну різницю потенціалів  $+U$ ; на електроди  $Y_s^I$  і  $Y_s^{I'}$ , а також на електроди  $Y_i^E$  і  $Y_i^{E'}$  ( $i = 1 \dots m, i \neq s$ ) – нульову різницю потенціалів, а на електроди  $Y_i^I$  і  $Y_i^{I'}$  ( $i = 1 \dots m, i \neq s$ ) – негативну різницю потенціалів  $-U$ .

Після цього слід протестувати роботу мережі, подаючи на вхід по черзі еталонні образи і контролюючи, щоб максимальна інтенсивність імпульсації при  $i$ -тому еталоні була саме на  $i$ -тому виході. Якщо ця вимога виконується, то закінчити навчання. Інакше – змінити випадковим чином порядок представлення еталонів і повторити процес навчання.

Таким чином, процес навчання здійснюється за допомогою власних схемотехнічних засобів (оптично керованих транспарантів ОКТ) без використання обчислювальних процедур (за принципом адаптивного зворотного зв'язку), а тому відпадає потреба у зовнішній цифровій пам'яті для зберігання синаптичних коефіцієнтів навчених нейронів.

## 6. Оцінка технологічних параметрів оптоелектронної спайкової нейронної мережі

Запропонована реалізація оптоелектронної спайкової нейронної мережі виконана у гібридному вигляді, тобто поєднує оптичні двовимірні просторово-неперервні структури (ОКТ [13]) і електронні (НВІС) компоненти. Як видно з рис. 9, дану СНМ можна виготовити у вигляді «сандвіч-структури». Так, ОКТ являє собою шар рідких кристалів, що знаходиться між скляними пластинами. Між цими ж пластинами поруч з шаром рідких кристалів можна розташувати транспарант  $T$  у вигляді фотопластинки. До цієї твердої конструкції з одного боку «приклеюється» матриця лазерів, на яку за технологією гібридних ІС встановлено напівпровідниковий чип з лінійкою вхідних нейронів  $EI_1 \dots EI_n$ . З іншого боку до цієї твердої конструкції «приклеюється» підкладка з масивом смугових фотодіодів, на якій також встановлено напівпровідниковий чип з інтернейронами  $E_1 \dots E_N$  та вихідними нейронами  $EO_1 \dots EO_m$ , зовнішні елементи нейронів (конденсатори) та контакти для джгута дротів, що йде на матрицю лазерів.

Для оцінки розміру площі такої «сандвіч-структури», що приходить на 1 нейрон, треба визначити роздільну здатність усіх складових. Так, напівпровідникові лазери виготовляються з апертурою 8–10 мкм [15], ОКТ має роздільну здатність 100–200 лін/мм (5–10 мкм) [13], смугові фотодіоди можна виконати шириною 6–8 мкм [16]. Найбільшим значенням є 10 мкм. Якщо вважати технологічно нормальним розміром для оптоелектронних пристроїв апертуру величиною 3 см, то на ній можна розташувати 3000 пікселів. Таким чином, на сьогоднішній день є реальним виготовлення апаратних реалізацій СНМ з кількістю нейронів близько 3000. Причому це будуть модулі СНМ, які можна каскадувати за допомогою оптичних засобів і отримувати СНМ з більшою кількістю елементів.

Порівняння параметрів запропонованої оптоелектронної реалізації СНМ з відомими [6–10] наведено у табл. 2, звідки видно, що запропонована СНМ майже за всіма кількісними параметрами не поступається відомим, а по більшості якісних показників (зв'язність, окремі нейрони, безпосередні зв'язки, асинхронність, on-chip навчання і т.п.) навіть переважає їх.

Таблиця 2. Порівняння параметрів запропонованої оптоелектронної реалізації СНМ з відомими

	SpiNNaker	Synapse (DARPA)	Neurogrid	BrainScaleS	Запропонована система
Елементна база	Мікропроцесори	Цифрові НВІС	Аналогові НВІС	Аналогові НВІС	Оптоелектроніка НВІС
Нейронів на чип	20000	256	65536 (256x256)	512	3000*
	Зв'язність нейронів не обмежена(+)		К-ть нейронів велика, але їх зв'язність обмежена (-)		Зв'язність нейронів не обмежена (+)
	Немає фізично окремих нейронів (-)		Є фізично відокремлені нейрони (+)		
	Немає безпосереднього зв'язку між нейронами (-)			Є безпосередній зв'язок між нейронами (+)	
Розмір чипа	19x19 мм <sup>2</sup> (в корп.)	4,2 мм <sup>2</sup>	11,9x13,9 мм <sup>2</sup>	5x10 мм <sup>2</sup>	30x30 мм <sup>2</sup> *
	Чип встановлюється в корпус, а корпус – на плату з додатковими обслуговуючими електронними схемами. Тому розміри системи в цілому великі (-)				
К-ть нейронів у мережі на сьогодні	1 млн (950 тис.)	До 2013 р 1 млн	1 млн	200 000	1 млн*
	Необхідність синхронізації (-)				
К-ть синапсів на 1 нейрон	Комунікаційний контролер	1280 синапсів. Всі синапси бінарні	6000	256	Асинхронна система (+)
	Синапси цифрові (-)				
Продуктивність	10 <sup>9</sup> імпл/с	10 <sup>6</sup> імпл/с	9,1x10 <sup>7</sup> імпл/с	10 <sup>11</sup> імпл/с	(0,2...1)x10 <sup>11</sup> *
Споживана потужність на 1 нейрон	1 мВт	0,5 мВт (45pJ/spike)	3,1 мкВт	5 мВт	10..20 мкВт*
	Навчання здійснюється за допомогою цифрового комп'ютера (-)				Оп-чип навчання (+)

\* - оціночні параметри;

(-) – позначає недолік;

(+) – позначає перевагу

## 7. Висновки

У результаті аналітичного огляду сучасних проектів з апаратної реалізації спайкових нейромереж було з'ясовано, що на сьогодні вони всі використовують технологію електронних надвеликих інтегральних схем (НВІС), яка є добре розвинуеною і апробованою, а тому зру-

чною. Деякі проекти використовують як елементну базу мікропроцесори, інші – цифрові НВІС, деякі – аналогові НВІС.

Було виділено два головних недоліки всіх проектів: 1) відсутність безпосередніх зв'язків між нейронами, оскільки технологічно неможливе створення великої кількості електричних ліній зв'язку у площині напівпровідникового кристалу (безпосередні зв'язки замінюються організацією штучних програмованих чи спеціально кодованих протоколів обміну між групами нейронів, які дозволяють зменшити кількість електричних ліній зв'язку у площині напівпровідникового кристалу, але втрачається відповідність принципам роботи біологічного мозку); 2) навчання спайкових нейронних мереж відбувається за допомогою цифрових комп'ютерів та спеціального програмного забезпечення, а не за допомогою власних не програмних механізмів і засобів, не пов'язаних з обчисленнями.

Запропонований у статті варіант апаратної реалізації оптоелектронної спайкової нейронної мережі вільний від цих недоліків. Перший недолік усунуто завдяки використанню оптичних сигналів для організації зв'язків між нейронами, оскільки світлові промені не вимагають ізоляції між сигнальними шляхами, можуть проходити один через інший без взаємного впливу, можуть розташовуватись у трьох вимірах та працювати одночасно, забезпечуючи величезний темп передачі даних. Другий недолік виправлено завдяки організації навчання запропонованої спайкової нейронної мережі за допомогою апаратних засобів (ОКТ) без використання обчислювальних процедур, причому також існує здатність мережі донавчатись і перенавчатись (адаптивність).

## СПИСОК ЛІТЕРАТУРИ

1. Різник О.М. Нейрокомп'ютери. Прихована науково-технічна революція [Електронний ресурс] / О.М. Різник // Світогляд. – 2010. – № 4. – С. 8 – 14. – Режим доступу: <http://www.mao.kiev.ua/jscans/2010-4-riznik.pdf>.
2. Галушкин А.И. Нейрокомпьютеры: учебн. пособ. для вузов / Галушкин А.И. – М.: ИПРЖР, 2000. – Кн. 3. – 528 с.
3. Maass W. Networks of spiking neurons: the third generation of neural network models / W. Maass // Neural Networks. – 1997. – N 10. – P. 1659 – 1671.
4. Бардаченко В.Ф. Перспективи застосування імпульсних нейронних мереж з таймерним представленням інформації для розпізнавання динамічних образів / В.Ф. Бардаченко, О.К. Колесницький, С.А. Василецький // УСiМ. – 2003. – № 6. – С. 73 – 82.
5. Бардаченко В.Ф. Таймерні нейронні елементи та структури / Бардаченко В.Ф., Колесницький О.К., Василецький С.А. – Вінниця: УНІВЕРСУМ-Вінниця, 2005. – 126 с.
6. Markram H. The blue brain project EPFL [Електронний ресурс] / H. Markram. – Режим доступу: <http://bluebrain.epfl.ch>.
7. Furber S. SpiNNaker – A Universal Spiking Neural Network Architecture [Електронний ресурс] / S. Furber, A. Brown // SpiNNaker Overview. – Режим доступу: <http://apt.cs.man.ac.uk/projects/SpiNNaker>.
8. Мода Д. DARPA SyNAPSE Program [Електронний ресурс] / Д. Мода. – Режим доступу: <http://www.artificialbrains.com/darpa-synapse-program#256processor>.
9. Simpson A. NeuroGrid Project. Grid technology for neuroscience [Електронний ресурс] / A. Simpson. – Режим доступу: <http://www.neurogrid.ac.uk>.
10. Meier K. BrainScaleS – Neuromorphic processors [Електронний ресурс] / K. Meier. – Режим доступу: <http://www.artificialbrains.com/brainscales/>.
11. Колесницький О.К. Компактна оптоелектронна реалізація імпульсної нейронної мережі / О.К. Колесницький, І.В. Бокоцей // Оптико-електронні інформаційно-енергетичні технології. – 2010. – № 2. – С. 54 – 62.
12. Kolesnytskyj O.K. Optoelectronic Implementation of Pulsed Neurons and Neural Networks Using Bispin-Devices / O.K. Kolesnytskyj, I.V. Bokotsey, S.S. Yaremchuk // Optical Memory & Neural Networks (Information Optics). – 2010. – Vol.19, N 2. – P. 154 – 165.
13. Пространственные модуляторы света / А.А. Васильев, Д. Касасент, И.Н. Компанец, А.В. Пар-



фенов. – М.: Радио и связь, 1987. – С. 125 – 183.

14. Колесницкий О.К. Метод распознавания многомерных временных рядов при помощи импульсных нейронных сетей / О.К. Колесницкий, Самра Муавия Хамо // Інформаційні технології та комп'ютерна інженерія. – 2006. – № 2 (6). – С. 86 – 93.

15. Princeton Optonics' Advanced High-Power Diode Lasers [Електронний ресурс]. – Режим доступу: <http://www.princetonoptonics.com/pdfs/HighPowerVCSELarrays.pdf>.

16. Стемпковский А. КМОП-фотодиодные СБИС – перспективная элементная база однокристалльных систем приема и обработки информации / А. Стемпковский, В. Шилин // ЭЛЕКТРОНИКА: наука, технология, бизнес. – 2003. – № 2. – С. 14 – 20.

*Стаття надійшла до редакції 27.02.2014*