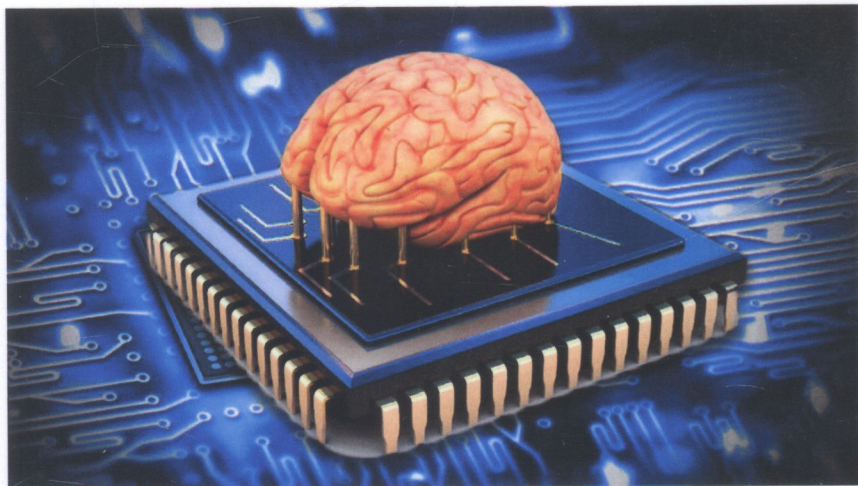


О. К. Колесницький,
В. І. Месюра

**НЕЙРОМЕРЕЖЕВІ
МОДЕЛІ ТА ТЕХНОЛОГІЇ
ОБЧИСЛЮВАЛЬНОГО ІНТЕЛЕКТУ.
НЕЙРОКОМП'ЮТЕРИ**

Частина I



Міністерство освіти і науки України
Вінницький національний технічний університет

О. К. Колесницький,
В. І. Месюра

НЕЙРОМЕРЕЖЕВІ МОДЕЛІ ТА ТЕХНОЛОГІЇ
ОБЧИСЛЮВАЛЬНОГО ІНТЕЛЕКТУ.
НЕЙРОКОМП'ЮТЕРИ

Частина I

Вінниця
ВНТУ
2021

УДК 004.93
К60

Рекомендовано до друку Вченою радою Вінницького національного технічного університету Міністерства освіти і науки, України (протокол № 13 від 29.04.2021 р.)

Рецензенти:

А. Я. Кулик, доктор технічних наук, професор

Т. Б. Мартинюк, доктор технічних наук, професор

Л. М. Куперштейн, кандидат технічних наук, доцент

Колесницький, О. К.

К60 **Нейромереві моделі та технології обчислювального інтелекту. Нейрокомп'ютери. Частина I : навчальний посібник / О. К. Колесницький, В. І. Месюра. – Вінниця : ВНТУ, 2021. – 66 с.**

ISBN ISBN 978-966-641-871-8

Посібник присвячено матеріалам лекційного курсу з дисципліни «Нейромереві моделі та технології обчислювального інтелекту» для здобувачів III (освітньо-наукового) рівня вищої освіти підготовки докторів філософії денної і заочної форм навчання спеціальності 122 «Комп'ютерні науки».

У посібнику розглянуто загальні принципи побудови архітектури нейрокомп'ютерів. Наведено обґрунтування вибору спайкінгових нейронних мереж як операційного блоку нейрокомп'ютерів. Розглянуто моделі та структури спайкінгових нейронних елементів та мереж, а також методи нейрокомп'ютерної обробки інформації на їх основі. Наведено аналітичний огляд міжнародних науково-технічних проектів з апаратної реалізації спайкінгових нейронних мереж.

УДК 004.93

ISBN 978-966-641-871-8

© ВНТУ, 2021

ЗМІСТ

ВСТУП.....	4
1 ПРИНЦИПИ ПОБУДОВИ АРХІТЕКТУРИ НЕЙРОКОМП'ЮТЕРІВ.....	5
1.1 Розвиток поняття нейрокомп'ютер.....	5
1.2 Огляд принципів побудови архітектури нейрокомп'ютера.....	8
1.3 Обґрунтування вибору спайкінгових нейромереж як операційного блока нейрокомп'ютера.....	11
1.4 Удосконалення принципів побудови архітектури нейрокомп'ютера.....	13
1.5 Контрольні запитання.....	16
2 АРХІТЕКТУРА СПАЙКІНГІНГОВОЇ НЕЙРОННОЇ МЕРЕЖІ І НЕЙРОМЕРЕЖЕВИЙ МЕТОД РОЗПІЗНАВАННЯ БАГАТОВИМІРНИХ СИГНАЛІВ.....	17
2.1 Моделі спайкінгових нейронних елементів.....	17
2.1.1 Формальні моделі спайкінгових нейронних елементів.....	19
2.1.2 Модель спайкінгового нейронного елемента з роздільними входами.....	23
2.2 Структура спайкінгової нейронної мережі.....	28
2.2.1 Аналіз моделі і структури спайкінгової нейронної мережі.....	28
2.2.2 Вдосконалення структури спайкінгової нейронної мережі.....	31
2.3 Метод навчання спайкінгових нейронних мереж.....	33
2.4 Метод розпізнавання багатовимірних сигналів з використанням спайкінгової нейронної мережі.....	38
2.5 Контрольні запитання.....	42
3 ПРОЕКТИ З АПАРАТНОЇ РЕАЛІЗАЦІЇ СПАЙКІНГОВИХ НЕЙРОННИХ МЕРЕЖ.....	44
3.1 Огляд відомих проектів з апаратної реалізації спайкінгових нейронних мереж.....	44
3.2 Недоліки відомих проектів з апаратної реалізації спайкінгових нейронних мереж та шляхи їх подолання.....	54
3.3 Архітектура апаратної реалізації оптоелектронної спайкінгової нейронної мережі.....	55
3.4 Оцінення технологічних параметрів оптоелектронної спайкінгової нейронної мережі.....	58
3.5 Перспективи подальших досліджень.....	58
3.6 Контрольні запитання.....	61
ЛІТЕРАТУРА.....	62

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 у разі їх апаратної реалізації [2]. На сьогодні, на жаль, не створено ефективної 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: 5×10^{10}) і водночас займає мінімальний об'єм і споживає мінімум енергії. Протягом останніх років спостерігається впевнений перехід від традиційних нейронних мереж на основі бінарних та аналогових нейронів з потенційними сигналами до так званих спайкінгових нейромереж з імпульсними сигналами [3, 4]. Спайкінгові 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 2 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 та для побудови на їх основі різноманітних нейрокомп'ютерних засобів для практичних застосувань.

1 ПРИНЦИПИ ПОБУДОВИ АРХІТЕКТУРИ НЕЙРОКОМП'ЮТЕРІВ

Великим завданням нейрокомп'ютерної техніки є розробка гнучких, подібних мозку, архітектур нейрокомп'ютерів, здатних до широкого спектра реально-часових застосувань. Водночас потрібно прагнути до максимально можливої кількості нейронів (в ідеалі – як у мозку людини – 10^{11}) та до ультра низьких потужностей споживання і компактного розміру, як у біологічних нейронних систем. Створення подібних мозку нейрокомп'ютерів дозволить вирішити 2 взаємно пов'язані задачі: 1) створення «розумних» комп'ютерів для виконання складних когнітивних неформалізованих завдань; 2) розкриття таємниць роботи мозку людини шляхом його зворотного конструювання технічними засобами. Найближчою стратегічною метою є створення нейроморфних ядер (чипів апаратних нейромереж), які інші вчені зможуть використовувати для перевірки своїх власних теорій щодо принципів роботи кори мозку і для побудови на їх основі різноманітних нейрокомп'ютерних систем для практичних застосувань. Ці нові знання можуть бути потім використані в наступному поколінні нейроморфних ядер.

У науковій літературі існує велика кількість інформації про різноманітні архітектури побудови нейрокомп'ютерів. Оскільки всі відомі нейрокомп'ютери можна розділити на 3 великих класи (програмні, програмно-апаратні та апаратні), то і архітектури також можна розглядати для програмних [1, 2], програмно-апаратних [3,4] та апаратних нейрокомп'ютерів [4]. Оскільки максимум переваг нейрокомп'ютерів перед традиційними комп'ютерами можна отримати тільки у випадку їх апаратної реалізації [4,5], то будемо розглядати тільки архітектури апаратних нейрокомп'ютерів. Для аналогії можна згадати широко відомі принципи побудови цифрових комп'ютерів, сформульовані свого часу Джоном фон Нейманом [6].

1.1 Розвиток поняття нейрокомп'ютер

Схемотехнічний (конструктивно-технологічний) та системотехнічний (архітектурний) аспекти розробки засобів комп'ютерної техніки тісно пов'язані один з одним, що відіграє важливу роль під час вибору оптимальної стратегії проектування комп'ютерів [7]. Це положення повністю стосується і нейрокомп'ютерів. Тут навіть важко визначити, що є первинним, а що вторинним. З одного боку, нові архітектурні рішення стимулюють розвиток технології, а з іншого боку, досягнення технології наштовхують на зміну архітектурних рішень і так далі; ці два процеси циклічно розвиваються по зростаючій спіралі.

Поняття «нейрокомп'ютер» на сьогоднішній день не має усталеного означення. Смысл цього поняття змінювався із розвитком нейро-

комп'ютерної техніки. Докладно історія та етапи розвитку нейрокомп'ютерів описано в роботі [4], тому проаналізуємо їх коротко (рис. 1.1).

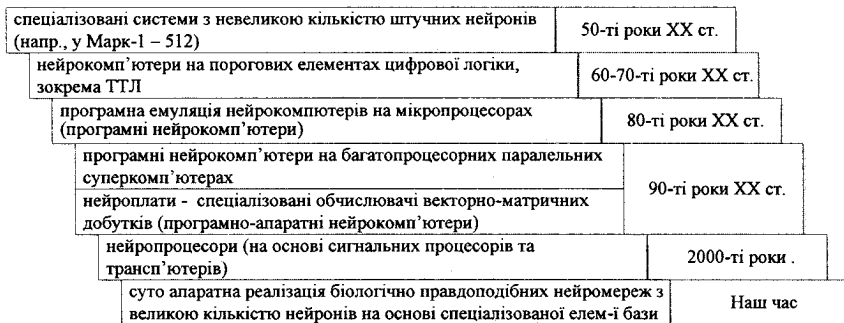


Рисунок 1.1 – Етапи розвитку нейрокомп'ютерів

Перші нейрокомп'ютери, такі як Марк-1, Адалін, Мадалін та ін. були створені у 50-х роках 20-го століття і були спеціалізованими системами з невеликою кількістю штучних нейронів (наприклад у Марк-1 – 512 нейронів). Вони розв'язували вузьке коло завдань із розпізнавання нескладних образів. Елементна база цих нейрокомп'ютерів була досить примітивною: від електромеханічних інтеграторів до магнітних, електрохімічних та конденсаторних.

Далі у 60-х – 70-х роках 20-го століття в зв'язку із розвитком технології цифрових інтегральних схем нейрокомп'ютери почали реалізовувати на порогових елементах цифрової логіки, зокрема ТТЛ. Але виникла суттєва проблема ефективних засобів для зміни вагових коефіцієнтів нейронів в процесі навчання та перекомутації топології їх зв'язків у випадку зміни типу мережі під конкретну задачу. Апаратними засобами ця проблема вирішувалась досить складно, тому почали застосовувати для вказаної мети універсальні цифрові комп'ютери, які до того часу вже набули достатнього розвитку. Так з'явилися гібридні нейрокомп'ютери, які мали у своєму складі блок нейронів та універсальний цифровий комп'ютер, що виконував функції введення, виведення, навчання, управління та індикації.

Подальший розвиток мікропроцесорної техніки зробив неефективним використання спеціалізованих систем. Більш ефективною стала програмна емуляція нейромережевих принципів обробки інформації на стандартних мікропроцесорах. Тоді з'явилися так звані програмні нейрокомп'ютери, тобто пакети прикладних програм, які дозволяли моделювати різні типи нейронних мереж за кількості нейронів – сотні-тисячі.

Але програмна емуляція нейромереж з великою кількістю нейронів потребувала використання значної частки ресурсів комп'ютера (великих обсягів оперативної пам'яті та значного завантаження процесора). Це відбувалось через необхідність виконання великої кількості операцій векторно-матричного множення. Для вирішення цієї проблеми було визначено 2

шляхи: 1) моделювати нейромережі на багатопроцесорних паралельних суперкомп'ютерах (програмні нейрокомп'ютери) та 2) розвантажити центральний процесор шляхом введення в комп'ютер додаткового апаратного забезпечення – нейроплати, яка є спеціалізованим обчислювачем векторно-матричних добутоків (програмно-апаратні нейрокомп'ютери).

Що стосується нейроплат, то спочатку їх виконували на серійних мікросхемах, а потім почали розробляти спеціальні нейропроцесори, використовувати сигнальні процесори та трансп'ютери. Але всі ці намагання не дали бажаних результатів через декілька причин. Першою причиною є те, що таке моделювання нейромереж, коли асинхронний обмін нервовими імпульсами між біологічними нейронами подається у вигляді синхронізованої обробки цифрових масивів та послідовним обміном кодами по шинах даних, не характеризується адекватністю біологічним прототипам. А тому неможливо буде реалізувати складні когнітивні функції біологічних нейромереж. Другою причиною є те, що кількість штучних нейронів в таких реалізаціях ($10^3 \dots 10^4$) є дуже далекою від кількості нейронів в людському мозку (10^{11}), а підвищити цю кількість суттєво не вдасться через обмеження технології виробництва планарних напівпровідникових чипів.

В кінці 90-х років пройшла також і пора повального захоплення програмною емуляцією нейрокомп'ютерних технологій на персональних ЕОМ, робочих станціях і навіть суперЕОМ. Стало зрозумілим, що тільки створення спеціалізованої елементної бази (як тоді здавалося, НВІС-нейрочипів) для суто апаратної реалізації біологічно правдоподібних нейромереж з великою кількістю нейронів дозволить значно підвищити відношення продуктивності до вартості в процесі реалізації нейрокомп'ютерних засобів обробки інформації. Однак реалізації на основі електронних НВІС мають низку суттєвих недоліків, пов'язаних з неможливістю організації в площині напівпровідникового кристала великої кількості ліній електричного зв'язку між нейронами. Через це замість асинхронної передачі імпульсів між нейронами (як в природі) використовується синхронізована передача кодованих цифрових пакетів даних [8] між групами нейронів, що спотворює самі принципи організації природних нейроструктур (нейроморфність).

Проблемою сьогодення є створення апаратних нейрокомп'ютерів з великою кількістю енергоекономних нейронів та нейроморфними принципами організації: безпосередні зв'язки між фізично відокремленими нейронами, адаптивне навчання без використання обчислювальних процедур.

Для визначення шляхів подолання цієї проблеми потрібно відповісти на такі основні питання:

1. Яка має бути архітектура нейрокомп'ютера?
2. Яка має бути структура його процесорного ядра – нейронної мережі?
3. За якою технологією реалізовувати апаратно цю нейронну мережу?
4. Які методи та засоби застосувати для адаптивного навчання нейромережі без використання обчислювальних процедур?

1.2 Огляд принципів побудови архітектури нейрокомп'ютера

Щоб розглянути архітектуру нейрокомп'ютерів, потрібно спочатку визначитись із самим поняттям архітектури. Наприклад, досі серед фахівців не існує одностайного означення поняття архітектури комп'ютера (обчислювальної машини). Розглянемо два найбільш вживані означення:

Архітектура комп'ютера — це концептуальна структура комп'ютера [7], яка визначає проведення обробки інформації і містить методи перетворення інформації в дані та принципи взаємодії технічних засобів і програмного забезпечення.

Архітектура комп'ютера – це логічна організація, структура та ресурси комп'ютера, які може використовувати програміст. Архітектура визначає принципи дії, інформаційні зв'язки та взаємне з'єднання основних логічних вузлів комп'ютера [9].

Об'єднує ці два означення таке: архітектура комп'ютера має у своєму складі як структуру, що віддзеркалює склад комп'ютера, так і програмно-математичне забезпечення. Структура комп'ютера – це сукупність елементів і зв'язків між ними.

Основи вчення про архітектуру цифрових комп'ютерів були закладені Джоном фон Нейманом [6]. Сукупність цих принципів породила класичну (фон-нейманівську) архітектуру ЕОМ. Фон Нейман не тільки сформулював фундаментальні принципи логічної будови ЕОМ, але і запропонував її структуру, зображену на рис. 1.2.

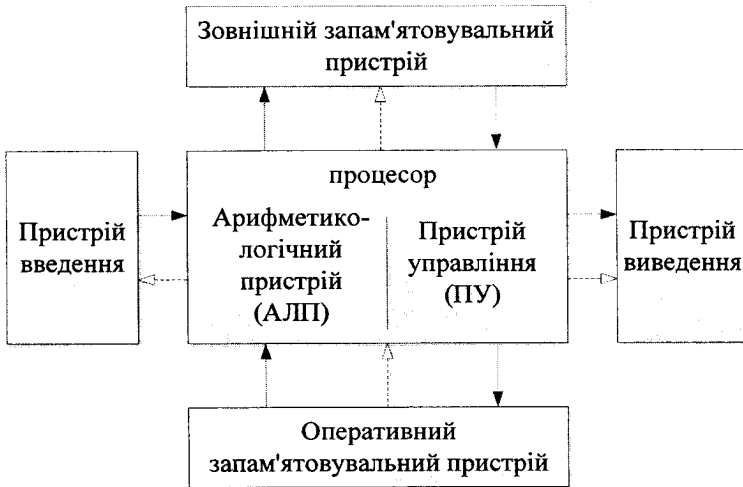


Рисунок 1.2 – Архітектура цифрового комп'ютера, побудованого на принципах фон Неймана. Суцільні лінії – потоки інформації, пунктирні – потоки сигналів управління

Згадаємо принципи фон Неймана:

1. *Принцип двійкового кодування*: вся інформація в комп'ютері подається у двійковому вигляді, сполучення 0 і 1.

2. *Принцип однорідності пам'яті*: й програми, й дані зберігаються в одній і тій самій пам'яті, тому комп'ютер не розпізнає, що зберігається у цій комірці пам'яті, а там можуть зберігатися цифри, текст, команда і т. д. Над командами можна здійснювати такі самі дії, що і над даними.

3. *Принцип адресованості пам'яті*: структурно основна пам'ять (ОП) складається з пронумерованих комірок, центральному процесору (ЦП) в будь-який момент часу доступна будь-яка комірка пам'яті. Тому можливо присвоювати імена блокам пам'яті для більш зручної взаємодії ОП і ЦП.

4. *Принцип послідовного програмного управління*: програма складається із сукупності команд, які виконуються ЦП автоматично одна за одною у певній послідовності.

5. *Принцип умовного переходу*: не завжди відбувається так, що команди виконуються одна за одною, тому можлива наявність команд умовного переходу, які змінюють послідовність виконання команд залежно від значення даних, що зберігаються.

Дамо означення архітектури нейрокомп'ютера за аналогією з цифровим комп'ютером.

Архітектура нейрокомп'ютера – це сукупність концептуальної структури нейрокомп'ютера та фундаментальних принципів функціонування і взаємодії його складових частин, які містять методи подання та обробки інформації, навчання та організації зручного інтерфейсу з користувачем.

Якою має бути архітектура нейрокомп'ютера? В науково-технічній літературі дуже мало публікацій, що стосуються архітектури абстрактного нейрокомп'ютера та його концептуальної узагальненої структури.

Є багато інформації щодо структур і архітектур різноманітних типів нейронних мереж (багатошаровий перцептрон, мережа Хопфілда, мережа Хеммінга та ін.). Також є структури і архітектури конкретних (а не абстрактних) нейропроцесорів (напр., на основі векторно-матричних операцій), нейроплат або «векторно-матричних» нейрокомп'ютерів [3, 4], орієнтованих на конкретну елементну базу (напр., ПЛІС) або конкретну структуру нейронної мережі на формальних чи аналогових нейронах.

Як приклад узагальненої структури нейрокомп'ютера, яка корелюється з іншими подібними роботами, можна навести схему, подану на рис. 1.3 [10].

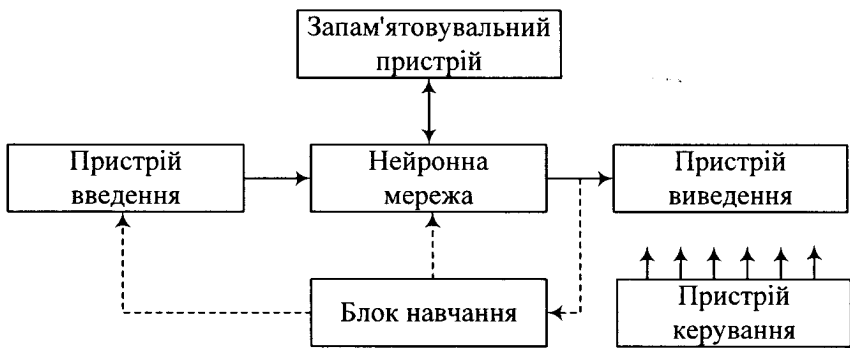


Рисунок 1.3 – Узагальнена структурна схема абстрактного нейрокомп'ютера

У цій самій роботі [10] сформульовано основні архітектурні принципи побудови абстрактного нейрокомп'ютера:

1. Основним операційним блоком нейрокомп'ютера, його процесором є штучна нейронна мережа, яка являє собою сукупність формальних нейронів, з'єднаних каналами передачі інформації.

2. Нейронна мережа не виконує обчислень, а трансформує вхідний сигнал (образ) у вихідний відповідно до своєї топології та значень коефіцієнтів міжнейронних зв'язків.

3. У запам'ятовувальному пристрої нейрокомп'ютера зберігається програма змінення коефіцієнтів зв'язку між нейронами.

4. Пристрої введення і виведення інформації виконують ті самі функції, що і в структурі фон Неймана, а пристрій керування виконує синхронізацію роботи всіх структурних блоків нейрокомп'ютера під час вирішення конкретної задачі.

5. Нейрокомп'ютер працює в двох режимах – навчання та функціонування. Процес навчання – це розв'язання задачі оптимізації, метою якої є мінімізація функції помилки (або нев'язки) на цій множині прикладів навчальних пар вхідного та вихідного образів шляхом підбору коефіцієнтів міжнейронних зв'язків. В режимі функціонування навчальний блок відключено і на вхід нейрокомп'ютера подаються сигнали з шумом, які необхідно розпізнати.

Як бачимо, в [10] принципи сформульовано дещо аморфно, не в такому визначеному стилі, як принципи фон Неймана. Ця структура і сформульовані принципи потребують нині уточнень і більш чітких формулювань, що викликано досягненнями наукових досліджень останніх років, зокрема, досягненнями в галузі спайкінгових нейронних мереж [5,8].

1.3 Обґрунтування вибору спайкінгових нейромереж як операційного блока нейрокомп'ютера

Протягом останніх років спостерігається інтенсифікація наукових досліджень в галузі спайкінгових нейронних мереж (англ. – *Spiking neural networks*), в яких інформаційними сигналами є імпульси (спайки), а не статичні сигнали певного рівня. Чим це викликано? Докладно це питання розглядається в [4, 5, 8], але якщо коротко, то, по-перше, тим, що ці мережі є нейроморфними (більш подібними до мозку, ніж традиційні) та універсальними (їх структура не залежить від розв'язуваної задачі, як у традиційних мереж зі статичними сигналами). Тобто не потрібно цілого «зоопарку» структур нейронних мереж для задоволення всіх практичних потреб.

Структурно-функціональна модель спайкінгової нейронної мережі, запропонована в [11], будується, на відміну від проблемно-орієнтованих мереж, на принципах створення динамічних систем в комбінації зі статистичною теорією навчання. Структура спайкінгової нейронної мережі зображена на рис. 1.4 і містить вхідні нейрони (вхідний шар), інтернейрони (проміжковий шар або – мікромережа) та вихідні нейрони (вихідний шар).

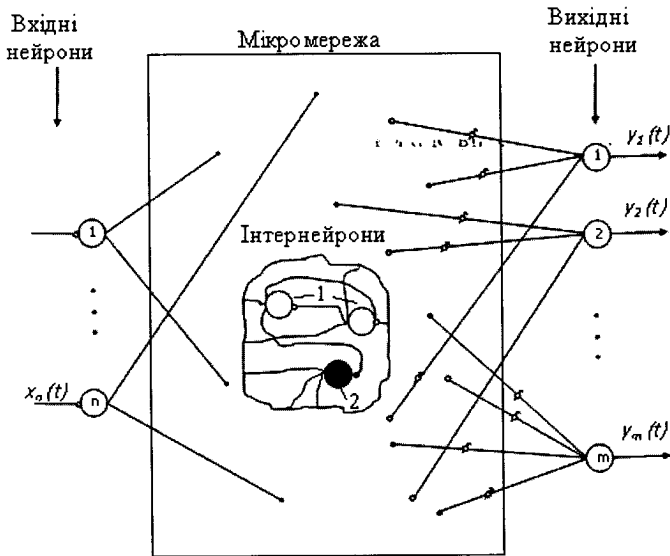


Рисунок 1.4 – Структура спайкінгової нейронної мережі
1 – будувальні нейрони, 2 – гальмівні нейрони

Така структура перекликається із структурою мозку, де є аферентні (сенсорні), проміжкові та еферентні (ефекторні) нейрони. Питома 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, які зв'язані з будь-яким нейроном в мережі, приводить до виникнення багатоконтурних зворотних зв'язків, тобто такі спайкінгові нейронні мережі є рекурентними. Нейрони в такій мережі мають бути двох типів: 1) збуджувальні (видають на виході електричний імпульс додатної полярності) та 2) гальмівні (видають на виході електричний імпульс від'ємної полярності).

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 визначитися даними нейрофізіологічних досліджень [8, 11]. Як видно із структури спайкінгової нейронної мережі (див. рис. 1.4), лише вихідні нейрони мають мати можливість підстроювати свої ваги синаптичних зв'язків в процесі навчання. Це відрізняє вихідні нейрони від інтернейронів, які мають постійні ваги зв'язків, що задаються під час формування мережі.

На рис. 1.5 зображено абстрактну модель спайкінгової нейронної мережі [11] у вигляді автомата з «плаваючими» станами. Її «плаваючий» високорозмірний аналоговий стан $x(t)$ змінюється безперервно в часі. І хоча динаміка такого автомата в загальному випадку дуже складна, немає необхідності визначати його передавальну функцію з цієї динаміки, оскільки можливе відновлення інформації, яка містилася в $x(t)$, безпосередньо з поточного стану автомата, навіть якщо він спотворений деяким шумом.

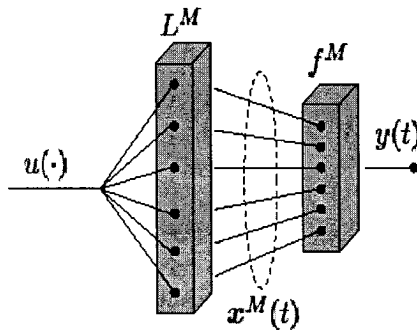


Рисунок 1.5 – Структура автомата з плаваючими станами (АПС)

Формально такий автомат M складається з фільтра L^M (тобто, функція, яка відображає вхідні потоки $u(\cdot)$ в потоки $x(\cdot)$), причому $x(t)$ може залежати не тільки від $u(t)$, але і абсолютно довільно нелінійно від попередніх входів $u(s)$: $x(t) = L^M(u(t))$, і з функції зчитування без запам'ятовування f^M , яка відображає в будь-який момент часу t вихід фільтра $x(t)$ («плаваючий стан»)

в деякий цільовий вихід $y(t)$. Загалом, такий автомат реалізує фільтр, який відображає $u(\cdot)$ в $y(\cdot)$.

Рекурентно з'дана нейромережа може розглядатися в першому наближенні як реалізація такого фільтра загального призначення L^M (наприклад, деяка незміщена аналогова пам'ять), з якого різні вихідні нейрони витягують і комбінують різні компоненти інформації, що містилася в попередньому вхідному сигналі $u(\cdot)$. Якщо цільовий вихід $y(t)$ приймає аналогові значення, то можна використовувати замість одного зчитувального нейрона ансамбль зчитувальних нейронів, імпульсна активність яких в момент часу t представляє рівень $y(t)$ у просторово-частотному коді. Фактично, ці зчитувальні нейрони мають пам'ять, але їхня стала часу мембрани істотно коротша, ніж тривалість періоду, протягом якого необхідно накопичення інформації, що потребується в більшості когнітивних задач.

Спайкінгові нейронні мережі завдяки нейроморфності мають перед традиційними, крім універсальності, ще й такі переваги [5, 8, 11]:

1) *розпізнавання динамічних образів* (мова, динамічні зображення та ін.);

2) *багатозадачність* (інформація про вхідні потоки циркулює в рекурентній нейромережі 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 за неповною інформацією 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 (спайкінговому) поданню інформації.

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

1.4 Удосконалення принципів побудови архітектури нейрокомп'ютера

Структура нейрокомп'ютера (див. рис. 1.3) і наведені під рис. 1.3 принципи побудови нейрокомп'ютера [10] потребують на сьогоднішній день уточнень, викликаних досягненнями наукових досліджень останніх років та результатами робіт [5, 12–15]. Так, в п. 1 замість «формальних нейронів» потрібно зазначити «біологічно-реалістичних спайкінгових нейронів», а замість «які з'єднані каналами передачі інформації» потрібно «які з'єднані безпосередньо». Це пояснюється просто: щоб розкрити таємниці людського мозку, потрібно максимально реалістично його імітувати, тобто замість формальних нейронів з потенційним виходом використовувати

спайкінгові нейрони, а зв'язки між нейронами все-таки робити безпосередніми, а не заміняти процес асинхронної передачі імпульсів процесом синхронізованої передачі пакетів цифрових кодів. Якщо цього не зробити, то є ризик не отримати потрібного ефекту виникнення надзвичайних властивостей (усвідомлення, мислення, емоції та ін.) у мережі звичайних інформаційних одиниць, якими є нейрони.

Із п. 2 можна повністю погодитись. У п. 3 викликає сумнів той факт, що в природній нейромережі (у мозку) немає ніяких програм, зокрема і програм навчання (змінення коефіцієнтів зв'язку між нейронами). Механізм навчання закладено у саму структурно-функціональну організацію мозку. І цей факт підтверджує «правило Хебба» [3, 8, 10], отримане із результатів нейрофізіологічних досліджень. Достеменно механізми роботи мозку поки не вивчені, тому чітко і однозначно скоригувати структуру за рис. 1.3 поки неможливо, але однозначно механізм навчання має бути закладений в організацію нейрокомп'ютерного ядра (нейромережі). І саме запам'ятовування також має відбуватися у нейронній мережі (принцип «пам'ять в процесорі»). Але окремий запам'ятовувальний пристрій може існувати для запам'ятовування другорядної інформації з метою розвантаження нейронної мережі.

Із п. 4 можна погодитись частково. Що стосується пристроїв введення, то давно існує думка, що нейрокомп'ютер має мати сенсорні поля, аналогічні п'яти відчуттям людини (тобто розуміти людську мову, сприймати зображення, відчувати запах, смак та мати тактильні сенсорні). А що стосується пристроїв виведення, то відповідно нейрокомп'ютер має генерувати людську мову, формувати зображення та мати виконуючі механічні органи для руху в просторі та орієнтації своїх сенсорних полів. Нейрофізіологи визначили, що мозок людини складається із окремих мікромереж нейронів (неокортикальні колонки) і в мозку є ділянки, що відповідають за певні функції. Щодо пристрою управління, то можливо на початкових етапах він є потрібним, але, із просуванням наших знань про мозок, функції управління буде закладено також в одну з підсистем нейронного ядра нейрокомп'ютера.

Із п. 5 також можна погодитись частково, з єдиним уточненням, що таке розділення на 2 режими (навчання та функціонування) є досить умовним. Адже відомо, що в біологічному мозку ці 2 режими існують паралельно і нема ніякої зовнішньої системи, що їх перемикає.

Крім цих пунктів, варто було б ще додати пункт про доукомплектування нейрокомп'ютера звичайним цифровим комп'ютером, який використовувався б для розв'язання формалізованих завдань. Тобто не цифровий комп'ютер має управляти блоками нейрокомп'ютера, а нейрокомп'ютер має управляти цифровим комп'ютером (ставити йому завдання, з якими той справляється краще і швидше) та використовує результати цих завдань.

Таким чином, з урахуванням наведених доводів, структура абстрактного нейрокомп'ютера може бути подана схемою згідно з рис. 1.6; а. архітектурні принципи його побудови сформульовано так:

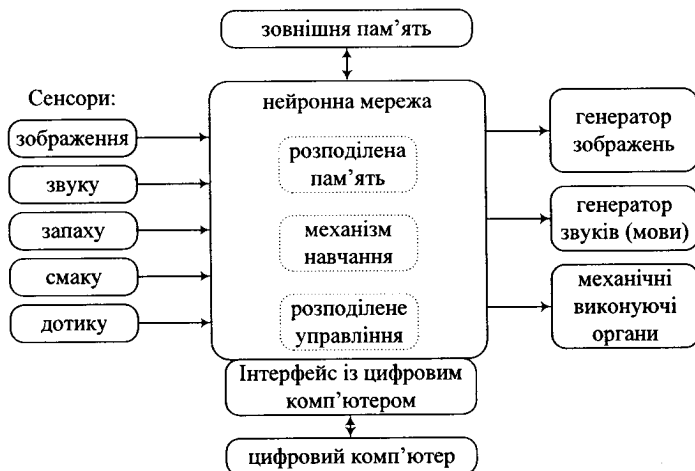


Рисунок 1.6 – Удосконалена узагальнена структурна схема абстрактного нейрокомп'ютера

1. *Принцип імпульсного кодування:* вся інформація в нейрокомп'ютері подається в імпульсній (спайкінговій) формі (інформація кодується моментами появи імпульсів).

2. *Принцип асоціативності обробки:* основним процесорним ядром нейрокомп'ютера є апаратно реалізована мережа біологічно-реалістичних спайкінгових нейронів максимально можливої кількості, які зв'язані між собою безпосередньо. Нейронна мережа не виконує обчислень, а трансформує вхідний сигнал (образ) у вихідний відповідно до своєї топології та значень коефіцієнтів міжнейронних зв'язків.

3. *Принцип автономності та адаптивності навчання:* навчання нейрокомп'ютера на виконання заданого відображення має відбуватися автономно на основі адаптації, а механізм навчання має бути закладений в організацію його процесорного ядра (нейронної мережі).

4. *Принцип розподіленості пам'яті:* функції запам'ятовування у нейрокомп'ютері реалізуються його процесорним ядром (нейронною мережею), а окрема зовнішня пам'ять має допоміжне значення.

5. *Принцип природності інтерфейсу:* нейрокомп'ютер має мати звичний для природи людини інтерфейс, тобто розуміти і генерувати мову, зображення і мати сенсорні поля, аналогічні п'яти відчуттям людини, а також мати виконувальні механічні органи для руху у просторі та орієнтації своїх сенсорних полів.

6. *Принцип однорідності обробки і управління*: функції управління функціональними частинами нейрокомп'ютера реалізуються його процесорним ядром (нейронною мережею).

7. *Принцип підсилення цифровим комп'ютером*: нейрокомп'ютер має мати у своєму складі цифровий комп'ютер, який використовується для вирішення формалізованих завдань. Нейрокомп'ютер управляє цифровим комп'ютером (ставить йому формалізовані завдання, з якими той справляється краще і швидше) через відповідний інтерфейс.

Відповідно до цих принципів можна дати таке означення нейрокомп'ютера.

Нейрокомп'ютер – це комплекс технічних засобів для обробки інформації, основним процесорним блоком якого є апаратно реалізована великомасштабна мережа спайкінгових нейронів; який має звичні для природи людини вербально-візуальний інтерфейс і сенсорні поля та виконує когнітивні функції, властиві мозку людини.

Нейрокомп'ютер відповідно до цього означення суттєво відрізняється від всіх попередніх програмних, програмно-апаратних та апаратних нейрокомп'ютерів на основі цифрових чи аналогових потенційних штучних нейронів. Його можна відокремити від них, надавши йому таку характеристику, як «природній нейрокомп'ютер», тому що він має бути так само інтелектуальним, творчим і самосвідомим, як людина. Жодна така машина поки ще не була побудована, але це лише тема часу.

1.5 Контрольні запитання

1. Які засоби краще розв'язують важкоформалізовані і неформалізовані задачі?
2. Дайте означення нейрокомп'ютера.
3. На які 3 класи поділяють всі нейрокомп'ютери?
4. Охарактеризуйте етапи розвитку апаратних нейрокомп'ютерів.
5. Які основні вимоги сьогодення до апаратних нейрокомп'ютерів?
6. Дайте означення поняття архітектура нейрокомп'ютера.
7. Наведіть структуру цифрового комп'ютера та сформулюйте фундаментальні принципи логічної будови ЕОМ фон Неймана.
8. Обґрунтуйте вибір спайкінгових нейромереж як операційного блока нейрокомп'ютера.
9. Наведіть структуру спайкінгової нейронної мережі та поясніть її.
10. Наведіть абстрактну модель спайкінгової нейронної мережі та поясніть її.
11. Які переваги мають спайкінгові нейронні мережі перед традиційними?
12. Наведіть структуру нейрокомп'ютера та сформулюйте фундаментальні принципи його побудови.

2 АРХІТЕКТУРА СПАЙКІНГІНГОВОЇ НЕЙРОННОЇ МЕРЕЖІ І НЕЙРОМЕРЕЖЕВИЙ МЕТОД РОЗПІЗНАВАННЯ БАГАТОВИМІРНИХ СИГНАЛІВ

2.1 Моделі спайкінгових нейронних елементів

Спайкінгові нейронні елементи [8, 16] є фізичними моделями біологічних нейронів (3-тє покоління нейронних елементів) і за своїми властивостями і характеристиками набагато ближчі до своїх прототипів (біологічних нейронів), ніж бінарні нейрони (формальні нейрони Мак-Каллока і Пітса – 1-є покоління нейронних елементів) і аналогові нейрони з потенційним виходом (2-ге покоління нейронних елементів).

Щоб правильно вибрати модель і розробити апаратну реалізацію спайкінгового нейронного елемента, необхідно знати принципи функціонування біологічного нейрона. Найважливішими рисами функціонування біологічного нейрона [17–20] є:

- частотний вихід – вихідна інформація про рівень збудження кодується в серії нервових імпульсів з відповідною частотою;
- амплітуда і тривалість окремих нервових імпульсів, які проходять по одному і тому ж волокну, постійні, а частота і кількість нервових імпульсів в послідовності залежать від інтенсивності збудження (закон «все або нічого»). Такий спосіб передачі інформації є найбільш завадостійким, тобто в широких межах не залежить від стану волокон, що проводять нервові імпульси.

Біологічний нейрон має такі функціональні можливості [18, 19]:

- просторове підсумовування – відбувається впродовж одночасного збудження різних дендритів, які ведуть до одного нейрона;
- часове підсумовування – відбувається впродовж серії частих збуджень нижче порогу;
- підпорогове підсумовування – якщо подразнювати нейрон короткочасними імпульсами з певною частотою і амплітудою, що перевищує поріг за напругою, то потенціал дії виникає не одразу. Це пояснюється тим, що енергії окремого імпульсу недостатньо для того, щоб перевищити поріг за енергією. Проте, кожен стимул віддає свою енергію нервовій клітині і зменшує значення порогу енергії. Тим самим виконується умова підпорогового підсумовування, якщо стимули не відстають значно один від одного в часі;
- залежність «сила–тривалість» – залежність між амплітудою і тривалістю струму, необхідною для досягнення порогового стану (рис. 2.1) [20]. Ця залежність пов'язана з багатьма чинниками, і, насамперед, із збудливістю: чим вона вища, тим менший струм необхідний для збудження. Основні параметри залежності такі: реобазис – найменша сила постійного електричного струму, який у разі достатньої тривалості дії викликає збудження в

нейроні (рис. 2.1); характеризує збудливість нейрона; хронаксія – найменший час, впродовж якого постійний електричний струм величиною в дві реобазис, діючи на нейрон, викликає виникнення вихідного імпульсу (рис. 2.1);

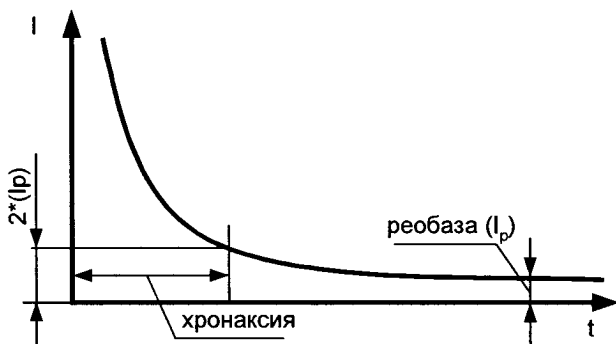


Рисунок 2.1 – Залежність «сила – тривалість»

- залежність частоти ритмічного збудження від амплітуди сигналу подразнення спочатку має лінійно зростаючу ділянку, а потім ділянку насичення;

- трансформація ритму – нервові клітини мають властивість змінювати частоту імпульсів, які передаються, тобто властивість трансформації ритму. Характер розряду залежить, по-перше, від властивостей подразника і, по-друге, від мембранного заряду, збудливості, лабільності нейрона. У звичайних умовах чим сильніше збудження, тим частіше виявляється відповідна імпульсація. У випадку високої збудливості нейрона може виникати підвищення частоти імпульсації (мультиплікація ритму). У разі низького рівня збудливості і лабільності нейрона частіше відбувається розрідження ритму, тому що декілька вхідних імпульсів мають підсумовуватися, аби нарешті досягти порогу виникнення потенціалу дії. Як результат, на декілька вхідних імпульсів нейрон відповідає лише одним імпульсом;

- залпова активність – відповідь нервової клітини може виникнути у формі одиночного потенціалу дії, серії імпульсів з частотою, що зменшується, а також у вигляді пачок імпульсів, які з'являються через певні інтервали;

- фоновая активність – нейрони мають здатність до постійної імпульсної активності. Їх мембранний потенціал періодично коливається, то збільшуючись, то зменшуючись. Фонову активність можуть викликати випадкові імпульси, що циркулюють в нервовій системі, але вона може існувати і за відсутності збуджень, які приходять до нейрона, як результат зміни іонних та обмінних процесів, коливань температури й інших причин. У

останньому випадку її називають спонтанною. Фонова активність відіграє важливу роль у діяльності цілого організму. По-перше, імпульси, які приходять аферентними шляхами, легше збуджують клітини із спонтанною активністю, ніж бездіяльні. По-друге, клітини, які мають постійний ритм фонові активності, можуть задавати цей ритм іншим нервовим клітинам (бути пейсмейкером). Водночас виникають великі комплекси синхронно працюючих клітин, в яких завдяки загальному ритму особливо добре виражена взаємодія між окремими нейронами. По-третє, деякі такі клітини виконують роль лічильників часу в центральній нервовій системі. Їх діяльність допомагає підтримувати певний ритм рухів, суб'єктивно оцінювати інтервали часу, через певні терміни включати необхідні функції і тому подібне;

- акомодация – пристосування нервової тканини до дії збудження, повільно наростаючого за силою. За безперервної стимуляції кількість медіатора, що звільняється в синапсі, поступово зменшується до тих пір, поки запаси медіатора не будуть виснажені, тоді подальша передача сигналів гальмується. Це запобігає пошкодженню ефекторів через перезбудження;
- рефрактерність – відсутність чутливості нейрона до вхідних збуджень під час формування потенціалу дії. У момент виникнення високовольтної частини – спайка нервова клітина не може відповісти на збудження новим потенціалом дії, тобто є абсолютно незбудливою (абсолютна рефрактерна фаза). Потім збудливість нейрона поступово поновлюється до початкового рівня (відносна рефрактерна фаза) і навіть деякий час може його перевищувати (екзальтаційна фаза).

Відзначимо, що всі біологічні нейрони виробляють імпульс лише одного знаку. Хоча вибір найбільш важливої властивості нейрона є неоднозначним, саме наявність частотного багатозначного виходу нейрона є визначальним під час побудови моделей нейронних елементів.

У п. 2.1.1 наведено відому модель спайкінгового нейрона, показано її обмеження і недоліки, а в п. 2.1.2 запропоновано удосконалену модель спайкінгового нейрона.

2.1.1 Формальні моделі спайкінгових нейронних елементів

Детальні моделі нейрона, основані на провідності, можуть відтворити електрофізіологічні властивості нейрона з високою мірою точності, але через властиву їм складність ці моделі важко аналізувати і апаратно реалізувати. Тому прості феноменологічні спайкінгові моделі нейрона надзвичайно популярні для вивчення нейронного кодування, пам'яті, динаміки нейронних мереж, а також для побудови апаратних реалізацій нейронів. Найпоширенішими є формальні порогові моделі імпульсації нейрона [21]. Імпульси генеруються кожного разу, коли мембранний потенціал u досягає деякого порога ϑ . Момент досягнення порогу визначає час генерації імпульсу $t^{(j)}$:

$$t^{(f)} : u(t^{(f)}) = \vartheta \text{ i } \left. \frac{du(t)}{dt} \right|_{t=t^{(f)}} > 0. \quad (2.1)$$

Для математичного і комп'ютерного моделювання спайкінгових нейронних мереж (СНМ) використовуються формальні порогові моделі імпульсації нейрона [21]: LIF-модель (leaky integrate-and-fire model), NIF-модель (nonlinear integrate-and-fire model), SRM-модель (spike response model) та ін.

Найпоширенішою є модель накопичувального спайкінгового нейрона з витоком (leaky integrate-and-fire neuron – LIF-нейрон [21]) – рис. 2.2: базову схему зображено в пунктирному колі. Струм $I(t)$ заряджає RC -коло. Напряга $u(t)$ на ємності порівнюється з порогом ϑ . Якщо $u(t)=\vartheta$ у момент часу $t_i^{(f)}$, то генерується вихідний імпульс $\delta(t-t_i^{(f)})$. Пресинаптичний імпульс $\delta(t-t_j^{(f)})$ фільтрується в синапсі за НЧ і генерує вхідний імпульсний струм $\alpha(t-t_j^{(f)})$.

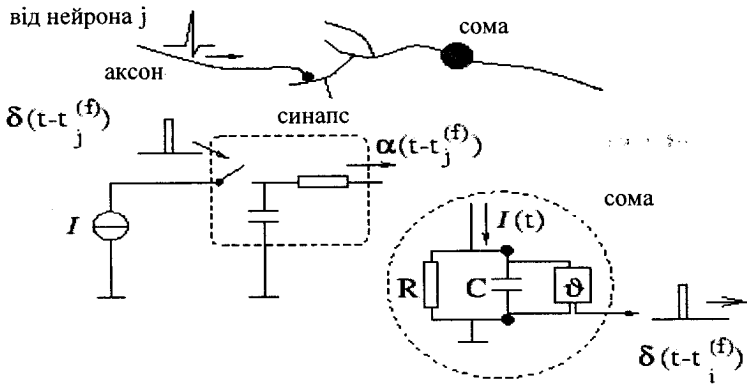


Рисунок 2.2 – Схематична діаграма накопичувальної спайкінгової моделі нейрона з витоком (LIF-моделі)

Базова схема моделі LIF-нейрона складається з паралельно з'єднаних ємності C і резистора R , керованих струмом $I(t)$ (рис. 2.2). Вхідний струм $I(t)$ може бути розділений на дві складові: $I(t) = I_R + I_C$. I_R – струм через резистор R розраховується за законом Ома як $I_R = u/R$, де u – напруга на резисторі. Друга складова I_C заряджає ємність C . З означення ємності $C = q/U$ (де q – заряд і U – напруга) ми знаходимо струм через ємність $I_C = C(du/dt)$. Тому

$$I(t) = \frac{u(t)}{R} + C \frac{du}{dt}. \quad (2.2)$$

Помноживши (2.2) на R і ввівши сталу часу $\tau_m = RC$ інтегратора з витоком, отримуємо стандартну форму математичної моделі LIF-нейрона

$$\tau_m \frac{du}{dt} = -u(t) + RI(t). \quad (2.3)$$

Напругу u називатимемо мембранним потенціалом, а τ_m – сталою часу мембрани нейрона.

У моделі LIF-нейрона потенціал дії явно не описується. Вихідні імпульси є формальними подіями, що характеризуються моментом імпульсації $t^{(j)}$, який визначається пороговим критерієм

$$t^{(j)}: \quad u(t^{(j)}) = \mathcal{G}. \quad (2.4)$$

Відразу ж після $t^{(j)}$ потенціал скидається до нового значення $u_r < \mathcal{G}$,

$$\lim_{t \rightarrow t^{(j)}, t > t^{(j)}} u(t) = u_r. \quad (2.5)$$

Для $t > t^{(j)}$ динаміка знову описується (2.3), доки порогове значення не буде досягнуте знову. Комбінація інтегратора з витоком (2.3) і скидання (2.5) визначає базову модель LIF-нейрона. Оскільки мембранний потенціал ніколи не перевищує порогу, порогова умова (2.1) зводиться до критерію (2.4), тобто, умова du/dt може бути опущена.

Це загальна версія моделі. Окрім цього модель LIF-нейрона може також містити період абсолютної рефрактерності (несприйнятливості). В цьому випадку міркують таким чином: якщо u досягає порогу у момент часу $t = t^{(j)}$, ми перериваємо динаміку (2.3) на час періоду абсолютної рефрактерності Δ^{abs} і відновлюємо інтеграцію в час $t^{(j)} + \Delta^{abs}$ з новою початковою умовою u_r .

Наприклад, якщо LIF-нейрон, визначений у (2.3) – (2.5), стимулюється постійним вхідним струмом $I(t) = I_0$, а потенціал скидання для простоти $u_r = 0$, то середня частота імпульсів моделі LIF-нейрона з періодом абсолютної рефрактерності Δ^{abs} , що стимулюється струмом I_0 , знаходиться так:

$$v = \left[\Delta^{abs} + \tau_m \ln \frac{RI_0}{RI_0 - \vartheta} \right]^{-1} \quad (2.6)$$

На рис. 2.3, а) показано часову залежність мембранного потенціалу LIF-нейрона, що збуджується постійним вхідним струмом $I_0 = 1,5$. На рис. 2.3, б) показано передавальну функцію LIF-нейрона (частота генерації ν без (суцільна лінія) і з (пунктирна лінія) абсолютною рефрактерністю $\Delta^{abs} = 4$ мс як функція постійного збуджувального струму I_0 за $I_v = 1$, $R = 1$, $\tau_m = 10$ мс, $u_r = 0$).

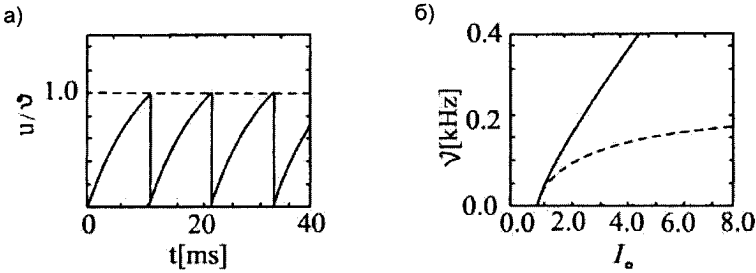


Рисунок 2.3 – а) часова залежність мембранного потенціалу LIF-нейрона, що збуджується постійним вхідним струмом $I_0 = 1,5$;
б) передавальна функція LIF-нейрона

Результати попереднього прикладу можуть бути узагальнені на довільні умови збудження і довільне значення напруги скидання $u_r < \vartheta$. Якщо імпульс з'являється у момент часу \hat{t} , то для $t > \hat{t}$ стимулювальний струм $I(t)$ і значення u_r розглядатимуться як початкова умова для інтеграції за (2.3), тобто

$$u(t) = u_r \exp\left(-\frac{t-\hat{t}}{\tau_m}\right) + \frac{1}{C} \int_0^{t-\hat{t}} \exp\left(-\frac{s}{\tau_m}\right) I(t-s) ds \quad (2.7)$$

Цей вираз описує мембранний потенціал для $t > \hat{t}$ і дійсний аж до моменту наступного досягнення порогу. Коли $u(t) = \vartheta$, мембранний потенціал знову скидається до u_r і інтегрування запускається знову. На рис. 2.4 показано напругу $u(t)$ моделі LIF-нейрона (вгорі), керованого вхідним струмом $I(t)$ (показано внизу). Вхідний струм $I(t)$ складається з накладення чотирьох

синусоїдальних компонентів з випадково вибраними частотами плюс позитивний струм зсуву $I_0 = 1,2$, який управляє мембранним потенціалом, зміщуючи його до порогу.

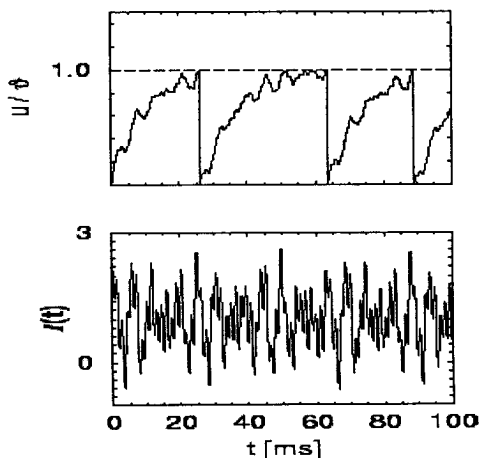


Рисунок 2.4 – Напряга $u(t)$ моделі LIF-нейрона (вгорі), керованого вхідним струмом $I(t)$ (внизу)

Формальна модель LIF-нейрона має цілу низку недоліків і відмінностей від біологічного нейрона. По-перше, в моделі LIF-нейрона передбачається, що збуджувальні імпульси мають мати позитивну полярність (заряджають ємність мембрани – рис. 2.2), а гальмівні імпульси мають мати негативну полярність (розряджають ємність мембрани). Це не відповідає роботі біологічного нейрона, оскільки в ньому всі імпульси мають однакову позитивну полярність. По-друге, через перший недолік в штучній мережі на LIF-нейронах необхідно мати нейрони двох типів: 1) збуджувальні, які видають імпульси позитивної полярності і 2) гальмівні, які видають імпульси негативної полярності. А це ускладнює мережу в цілому, оскільки збільшує необхідну для розв'язання конкретного завдання кількість нейронів. З огляду на це, було розроблено модель спайкінгового нейрона, вільну від вказаних недоліків.

2.1.2 Модель спайкінгового нейронного елемента з роздільними входами

Для усунення вказаних вище недоліків формальної моделі LIF-нейрона, яку можна схематично зобразити позначенням, наведеним на рис. 2.5, а), пропонується такий підхід.

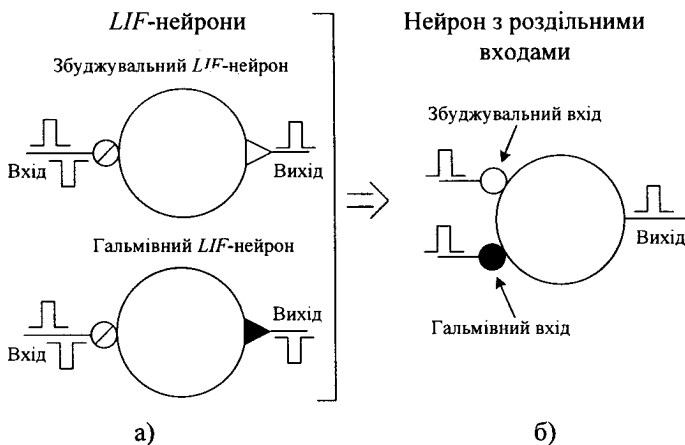


Рисунок 2.5 – Відмінності відомих LIF-нейронів від запропонованого нейрона з розділними входами

Біологічний нейрон має на дендритному дереві і збуджувальні синапси, і гальмівні. Імпульси, що приходять на збуджувальні синапси, збільшують заряд (і потенціал) мембрани, а імпульси, що приходять на гальмівні синапси, зменшують заряд (і потенціал) мембрани. Тобто всі вхідні сигнали конкретного біологічного нейрона можна умовно розділити на дві групи: 1) збуджувальні сигнали і 2) гальмівні сигнали. Тому можна умовно вважати, що нейрон має два входи – збуджувальний і гальмівний, причому на першому вході відбувається просторово-часове підсумовування всіх сигналів від збуджувальних синапсів, а на другому вході відбувається просторово-часове підсумовування всіх сигналів від гальмівних синапсів. Просторово-часове підсумовування легко реалізується для оптичних сигналів шляхом їх подачі на загальну апертуру будь-якого фотоприймального пристрою (наприклад, фотодіода). Тому запропонована до використання модель нейрона має мати роздільні входи для збуджувальних і гальмівних імпульсів і подається умовним зображенням, показаним на рис. 2.5, б). Цю модель легко реалізувати за допомогою оптоелектронної елементної бази.

Схематична діаграма спайкінгової моделі нейрона з розділними входами так само, як і схематична діаграма накопичувальної спайкінгової моделі нейрона з витокон (LIF-моделі), показана на рис. 2.2, має мати конденсатор, що імітує ємність мембрани, також має мати пороговий елемент, що спрацьовує у випадку досягнення потенціалом на конденсаторі порогового рівня, але вхідна частина відрізняється і являє собою два джерела струму – $I_{зб}$ і $I_{глім}$ з ключовими елементами, як показано на рис. 2.6.

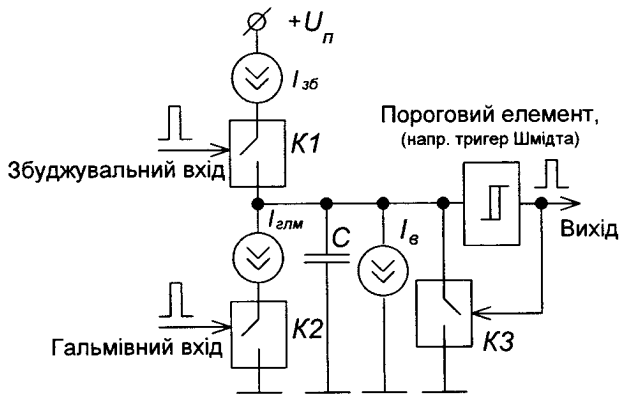


Рисунок 2.6 – Схематична діаграма спайкінгової моделі нейрона з роздільними входами

Ця модель працює таким чином: з приходом збуджувального імпульса ключ $K1$ замикається на час дії імпульса і конденсатор C заряджається струмом I_b протягом часу тривалості імпульсу. Водночас різниця потенціалів на ньому збільшується. З приходом гальмівного імпульса ключ $K2$ замикається на час дії імпульса і конденсатор C розряджається струмом $I_{элм}$ протягом часу тривалості імпульса. За такої умови різниця потенціалів на ньому зменшується. Таким чином, здійснюється просторово-часове алгебраїчне підсумовування вхідних стимулів нейрона, а точніше – знаходиться різниця між сумою збуджувальних стимулів і сумою гальмівних стимулів. Єдина особливість цього процесу полягає в тому, що ця різниця є так званою «обмеженою різницею». «Обмежена різниця» означає те, що в разі відсутності на конденсаторі C заряду, імпульси, що приходять на гальмівний вхід, не зменшують різницю потенціалів на конденсаторі нижче нуля (тобто в негативну зону). Якщо за якийсь період часу збуджувальних імпульсів було більше, ніж гальмівних, причому на стільки, що різниця потенціалів на конденсаторі C досягла порогового значення, за якого спрацьовує пороговий елемент, то у момент досягнення цього порогового значення вихідний сигнал порогового елемента різко зростає, формуючи передній фронт вихідного імпульсу. Водночас замикається ключ $K3$ і розряджає конденсатор C , внаслідок чого потенціал на вході порогового елемента зменшується і його вихідний сигнал різко знижується до нуля, формуючи задній фронт вихідного імпульсу. Потім описані процеси повторюються. Джерело струму I_b імітує витік заряду з ємності мембрани, що властиво і для LIF-нейрона. Інерційність спрацьовування порогового елемента і час розряду ємності C через ключ $K1$ визначає період рефрактерності нейронного елемента.

Схематичній діаграмі спайкінгової моделі нейрона з роздільними входами з рис. 2.6 відповідає апаратна реалізація нейронного елемента на біспін-приладі [22], запропонована в [16, 23, 24], схема якої показана на рис. 2.7, де $VD1$, $VD2$ – фотодіоди, що приймають сигнали збудження і гальмування, відповідно; VB – біспін-прилад.

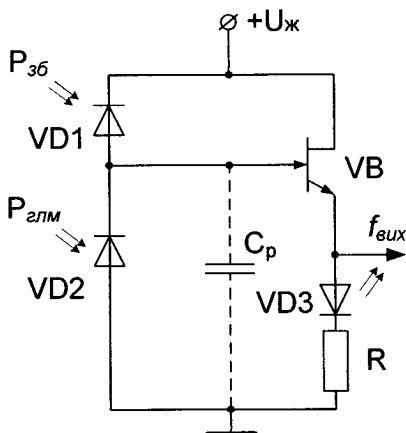


Рисунок 2.7 – Інтернейрон для СНМ на основі біспін-приладу

Математична модель спайкінгового нейрона з рис. 2.7 також розроблена в [16, 23, 24] і має такий вигляд:

$$\tau = \frac{C_p (U_{\text{ж}} - 0,35)}{S_{\phi 1} \cdot P_{\text{зб}} - S_{\phi 2} \cdot P_{\text{гльм}}}, \quad (2.8)$$

де τ – період повторення вихідних імпульсів нейрона;

$S_{\phi 1}$, $S_{\phi 2}$ – інтегральні чутливості фотодіодів $VD1$ і $VD2$, відповідно;

$P_{\text{зб}}$, $P_{\text{гльм}}$ – оптичні потужності на збуджувальному і гальмівному входах, відповідно;

$U_{\text{ж}}$ – напруга живлення;

C_p – ємність підкладки біспін-приладу.

У [16, 23] доведено, що спайкінговий нейронний елемент на біспін-приладі (математична модель (2.8)) має майже всі властивості біологічного нейрона, а значить може успішно використовуватися в спайкінгових нейронних мережах.

Проте, в моделі (2.8) не враховано 2 чинники:

1) наявність деякого струму витоку з підкладки через запірний контакт біспін-приладу на загальну шину (I_b);

2) наявність фази абсолютної рефрактерності Δ^{abs} – невеликий інтервал часу після формування вихідного імпульсу, впродовж якого ємність підкладки не заряджається навіть за наявності вхідного збуджувального стимулу (тобто нейрон нечутливий в цей час до вхідних дій).

Тому удосконалимо модель (2.8), розроблену в [16, 23, 24], з урахуванням вказаних вище чинників.

На рис. 2.6 показано вже уточнену порівняно з [16, 22, 23] схему заміщення спайкінгового інтернейрона на біспін-приладі згідно з рис. 2.7. У цій схемі $I_{зб} = S_{\phi 1} \cdot P_{зб}$; $I_{гкм} = S_{\phi 2} \cdot P_{гкм}$, а наявність струму витоку I_v через запірний контакт біспін-приладу на загальну шину враховано введенням генератора струму величиною I_v .

За цією схемою заміщення можна скласти таку математичну модель, врахувавши наявність періоду абсолютною рефрактерності додаванням величини Δ^{abs} до виразу (2.8), а наявність струму витоку – відніманням величини I_v від алгебраїчної суми струмів в знаменнику виразу (2.8). Як результат отримуємо такий вираз, що описує залежність періоду τ генерації імпульсів спайкінговим нейроном від величин оптичних потужностей $P_{зб}$ і $P_{гкм}$ відповідно на збуджувальному і гальмівному входах нейрона:

$$\tau = \Delta^{abs} + \frac{C_p (U_{ж} - 0,35)}{S_{\phi 1} \cdot P_{зб} - S_{\phi 2} \cdot P_{гкм} - I_v} \quad (2.9)$$

Переходячи від періоду до частоти, отримуємо таку математичну модель нейрона на біспін-приладі, що описує залежність частоти ν генерації імпульсів нейроном від величин оптичних потужностей $P_{зб}$ і $P_{гкм}$ відповідно на збуджувальному і гальмівному входах нейрона:

$$\nu = \left[\Delta^{abs} + \frac{C_p (U_{ж} - 0,35)}{S_{\phi 1} \cdot P_{зб} - S_{\phi 2} \cdot P_{гкм} - I_v} \right]^{-1} \quad (2.10)$$

Цю математичну модель використовуватимемо для комп'ютерного моделювання спайкінгових нейронних мереж, а також для оцінення адекватності використовуваної моделі спайкінгового нейрона та формальної моделі LIF-нейрона.

Оскільки ми змінили (удосконалили) модель самого нейрона, застосувавши в ній два роздільні входи (один – для збуджувальних сигналів, а другий – для гальмівних), і перейшовши від використання імпульсів різної полярності до імпульсів однакової полярності для збуджувальних та галь-

мівних сигналів, то також зміниться і організація та структура зв'язків, і кількість нейронів мережі. Тому необхідно розробити нову вдосконалену структурну організацію спайкінгової нейронної мережі в цілому.

2.2 Структура спайкінгової нейронної мережі

2.2.1 Аналіз моделі і структури спайкінгової нейронної мережі

Спайкінгові нейронні мережі (СНМ) [8, 11, 25–28] складаються з спайкінгових нейронних елементів [8, 16, 23]. Нейроморфність (схожість з біологічними нейронами) спайкінгових нейронних елементів полягає насамперед в тому, що на своєму виході (аксоні) вони формують не логічні рівні (як формальні нейрони Мак-Каллока і Пітса) і не безперервні рівні з діапазону $[0, 1]$ або $[-1, 1]$ як аналогові нейронні елементи, а послідовність імпульсів, миттєва частота яких визначається мірою збудження нейрона у поточний момент часу.

СНМ не мають певних правил з'єднання нейронів між собою, як, наприклад, широко відомі мережі Хопфілда, Хеммінга, багатoshарові перцептрони та інші. Вони є сукупністю нейронів, з'єднаних між собою випадковим чином, так, що є як прямі, так і зворотні зв'язки. Кожен нейрон мережі пов'язаний не зі всіма іншими нейронами, а лише з деякими. Під час формування конкретної СНМ використовують результати нейрофізіологічних досліджень, якими, наприклад, встановлено, що, по-перше, в мозку людини сукупність нейронів в усіх масштабах 1 мм^3 формує відносно незалежну локальну мікромережу, що несе певне функціональне навантаження, по-друге, серед всіх зв'язків одного нейрона приблизно 20 % є гальмівними та ін. Тому такі СНМ часто називають генеричними рекурентними спайкінговими нейронними мережами.

Функціонування рекурентних мереж спайкінгових нейронів [26, 28, 29] ґрунтується на ідеї так званого «рідкого комп'ютера» («liquid computer» [11]), яка полягає в тому, що рідина деякий час зберігає інформацію про проведені над нею зовнішні дії у вигляді динамічно змінного стану її поверхневого хвилювання. Математичною моделлю «рідкого комп'ютера» є «автомат з рідкими станами» (за аналогією з кінцевими автоматами).

Концептуальна структура автомата з рідкими станами (АРС) полегшує аналіз можливостей реально-часових обчислень моделей нейронних мікромереж. Ця модель не потребує побудови задачезалежної структури нейронної мережі і може використовуватися для аналізу обчислень довільно «знайдених» або побудованих моделей нейронних мікромереж. Ця модель також не потребує жодного попереднього рішення відносно «нейронного коду», яким інформація подана в мережі.

Основна ідея полягає в тому, що нейронна (рекурентна) мікромережа може слугувати як незміщена аналогова (короткочасна) пам'ять (неформально названа як «рідка») про поточні і попередні вхідні сигнали мережі.

«Рідким станом» $x(t)$ (рис. 2.8) називають вектор вкладів всіх нейронів мікромережі в мембранний потенціал вихідного нейрона у час t . Відзначимо, що це є вся інформація про стан мікромережі, до якої вихідний нейрон підключений. На відміну від певних станів кінцевого автомата, «рідкий стан» АРС не потрібно технічно визначати (будувати) для специфічного завдання. Передбачається, що воно безперервно змінюється в часі, є достатньо чутливим і таким, що має високу розмірність для того, щоб містити всю інформацію, яка може бути необхідною для специфічних завдань.

Рекурентна мікромережа спайкінгових нейронів перетворює входні сигнали в стан $x(t)$, який розшифровується за допомогою вихідних функцій без пам'яті f_1, \dots, f_n у вихідні сигнали $f_1(x(t)), \dots, f_n(x(t))$. «Рідкий стан» $x(t)$ нейронної мікромережі може бути перетворений у будь-який час t картою вихідних нейронів в деякий цільовий вихід $f(x(t))$ (який в загальному випадку видається із специфічним поданням або нейронним кодом). Структуру АРС показано на рис. 2.8.

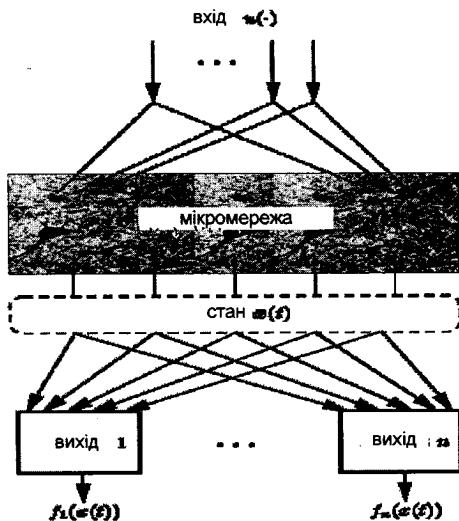


Рисунок 2.8 – Автомат з рідкими станами (АРС)

У [11] показано, що лише синапси вихідних нейронів мають бути налаштовані для конкретного обчислювального завдання. Це потребує, щоб будь-які дві різні входні часові $u_1(s), s \leq t$ і $u_2(s), s \leq t$, які мають викликати різні вихідні сигнали в деякий подальший момент часу t , переводили рекурентну мережу в два істотно різних стани $x_1(t)$ та $x_2(t)$ у час t (*властивість розділення*). Іншими словами: поточний стан $x(t)$ мікромережі під час t має містити всю інформацію про попередні входні сигнали. Щоб мати універсальні обчислювальні можливості, АРС має задовольняти окрім вла-

стивості розділення, ще і *властивості апроксимації*: вихідні нейрони мають бути здатні апроксимувати будь-яку задану неперервну функцію f , яка відображає поточний рідкий стан $x(t)$ в поточний вихід $v(t)$.

Якщо АРС характерні ці властивості, то можливо навчити вихідні нейрони без пам'яті, щоб видавати бажаний вихідний сигнал у час t . Якщо час t змінюється, то можна використовувати такі ж принципи, аби видавати як вихідний сигнал бажані часові серії імпульсів або функції часу t на тому самому виході. Це приводить до такої (автономної) процедури навчання вихідних нейронів, щоб виконати поставлене завдання:

1. Визначити нейронну мікромережу, яка має аналізуватися.
2. Записати стани $x(t)$ мікромережі в різні моменти часу у відповідь на багаточисельні різні (навчальні) вхідні сигнали $u(t)$.
3. Застосувати алгоритм навчання з учителем до набору навчальних прикладів у формі $[x(t), u(t)]$, щоб навчити вихідну функцію f , яка визначає фактичний вихідний сигнал $f(x(t))$, приймати значення, які були б якомога ближчими до цільового виходу $y(t)$.

Одна з переваг цього підходу полягає в тому, що немає необхідності брати до уваги будь-які часові аспекти задачі навчання, оскільки вся часова обробка зроблена неявно в рекурентній мережі. До того ж, жодне попереднє рішення не потрібне відносно нейронного коду, яким інформація про попередні вхідні сигнали кодується в поточному рідкому стані мережі. Відмітимо також, що АРС може легко здійснювати декілька обчислень паралельно, використовуючи одну й ту саму рекурентну мережу. Просто доведеться навчати для кожного цільового виходу окремий вихідний нейрон, який може повністю використовувати ту саму рекурентну мережу.

Згідно з теоретичним аналізом цієї обчислювальної моделі [26, 27, 30, 31] немає жодних апріорних обмежень для здатності цієї моделі здійснювати в реальному часі обчислення з короткочасною пам'яттю. Природно, що для здійснення завдань, які потребують великий обсяг пам'яті або велику стійкість до шуму у випадку розділення образів, знадобиться мережа з великою кількістю нейронів.

Проте, в розробленій нині парадигмі рекурентних мереж спайкінгових нейронів є деякі недосліджені моменти і недоліки:

- 1) не розроблено критерії, за якими можна було б оцінити мінімальну кількість нейронів мікромережі, необхідну для розв'язання конкретного завдання, а також мінімально необхідну кількість нейронів, з якими має бути пов'язаний кожен з входів мережі і кожен вихідний нейрон мережі;
- 2) не досліджено питання швидкості навчання, збіжності і достовірності роботи таких мереж у випадку використання різних алгоритмів навчання;
- 3) актуальними залишаються питання ефективної технічної реалізації [8] таких мереж великих об'ємів, особливо питання організації зв'язків з різними вагами між віддаленими нейронами, а також можливість здійснення *on-line* навчання з використанням лише внутрішньочипових можливостей;

4) для реалізації використовуються моделі спайкінгових нейронів, які не дозволяють в одньому нейронному елементі мати і збуджувальні, і гальмівні входи, тому часто замість одного нейрона використовується пара нейронів (один – збуджувальний, який формує на виході позитивний імпульс струму, інший – гальмівний, який формує на виході негативний імпульс струму)

Проблеми 1) і 2) не розглядаються в рамках цього навчального посібника.

Деякі питання проблеми 3) розглядаються в розділі 3.

Вирішити проблему 4) можна у разі реалізації спайкінгових нейронів на основі біспін-приладу за схемою згідно з рис. 2.7.

2.2.2 Вдосконалення структури спайкінгової нейронної мережі

Структура СНМ, що розглядається в посібнику для розпізнавання багатоканальних сигналів, відрізняється від запропонованої в [8], по-перше, тим, що доповнена входними нейронами, які перетворюють входні сигнали, що є функціями часу в пропорційну значенню функції частоту імпульсів. По-друге, як інтернейрони (середній шар мережі) використовуються нейронні елементи одного типу, які мають як збуджувальні, так і гальмівні входи, керовані імпульсами позитивної полярності, та які на виході видають імпульси лише позитивної полярності.

На відміну від цього, в структурі мережі [8] використовуються два типи нейронів: збуджувальні нейрони, які видають імпульс струму позитивної полярності, і гальмівні нейрони, які видають імпульс струму негативної полярності. Тобто, для досягнення заданої обчислювальної потужності відома мережа має мати більшу кількість нейронів. З урахуванням вказаних змін структура СНМ буде мати вигляд, показаний на рис. 2.9.

Як інтернейрони в цій структурі використовуються запропоновані спайкінгові нейрони з роздільними входами згідно з моделлю – рис. 2.5, б) і рис. 2.6. Кількість нейронів вибирається пропорційно об'єму розпізнаваних образів і міри їх коррельованості (схожості). Кількість зв'язків одного нейрона з іншими (наприклад, 10...15 %), дальність їх поширення, співвідношення кількості збуджувальних і гальмівних зв'язків (наприклад, відповідно 80 % і 20 %) може визначатися даними нейрофізіологічних досліджень [8].

Як видно із структури спайкінгової нейронної мережі (рис. 2.9), лише вихідні нейрони мають мати можливість підстроювати свої ваги синаптичних зв'язків в процесі навчання. Це відрізняє вихідні нейрони від інтернейронів (див. рис. 2.7), які мають постійні ваги зв'язків, що задаються під час формування мережі. У разі апаратної реалізації спайкінгових нейронних мереж у вигляді НВІС для організації змінних в процесі навчання вагів зв'язків використовують спеціальні схемотехнічні засоби [32]. Такими засобами можуть бути лінійні і нелінійні електронні компоненти з парамет-

рами, що адаптуються. Це, наприклад, варикапи, МДН-транзистори з плаваючим затвором, варистори, резистори на основі оксидів висути або аморфного полікристалічного кремнію, які програмуються за допомогою імпульсів різної полярності [32] та інші адаптивні електронні, оптоелектронні і магнітоелектронні компоненти.

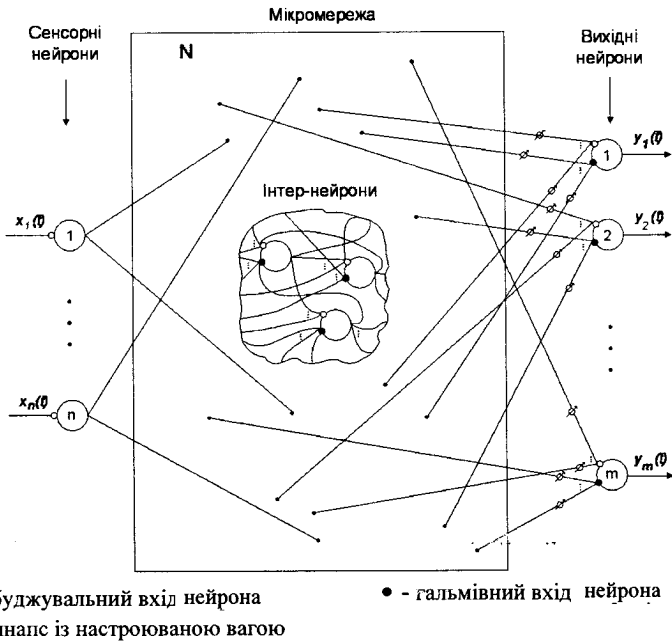


Рисунок 2.9 – Структура спайкінгової нейронної мережі

У розділі 3 посібника розглядається апаратна реалізація спайкінгової нейронної мережі на основі оптоелектронної елементної бази. У ній схемотехнічні засоби для організації підстроювання вагів зв'язків не входять до складу кожного окремого вихідного нейрона, а винесені окремо та реалізовані у вигляді оптично керованого транспаранта з пам'яттю [33]. Тому за такої реалізації схема і математична модель вихідного нейрона не відрізняються від схеми і моделі інтернейрона.

Запропонована СНМ з рис. 2.9 простіша за відомі СНМ на базі, наприклад, LIF-нейронів за рахунок скорочення загальної кількості нейронів мережі завдяки застосуванню оптоелектронних нейронних елементів з однаковою полярністю сигналів для збуджувальних і гальмівних імпульсів. Пояснимо цей факт. У відомих реалізаціях СНМ використовуються окремо збуджувальні нейрони (рис. 2.5, а) зверху), які формують вихідний імпульс позитивної полярності, і окремо гальмівні нейрони, які формують вихідний

імпульс негативної полярності (рис. 2.5, а) знизу). Процентне співвідношення їх кількості в мережі визначається результатами нейрофізіологічних досліджень і найчастіше використовується 80 % збуджувальних нейронів та 20 % гальмівних нейронів. Запропонована в [34] реалізація спайкінгових нейронних елементів на основі біспін-приладів відрізняється тим, що поєднує в одному нейроні збуджувальні і гальмівні входи, тобто під час побудови спайкінгових нейронних мереж на основі таких нейронних елементів можна досягти скорочення апаратних витрат принаймні на 20 %.

2.3 Метод навчання спайкінгових нейронних мереж

Теорія навчання розглядає три фундаментальні властивості, пов'язані з навчанням за прикладами: ємність, складність зразків і обчислювальна складність. Під ємністю розуміється кількість зразків, які може запам'ятати мережа, і які функції та межі ухвалення рішень можуть бути на ній сформовані. Складність зразків визначає число навчальних прикладів, необхідних для досягнення здатності мережі до узагальнення. Дуже велика кількість прикладів може викликати «перенавченість» мережі, коли вона добре функціонує на прикладах навчальної вибірки, але погано – на тестових прикладах, що належать до того самого статистичного розподілу. Відомо чотири основних типи правил навчання: корекція за помилкою, машина Больцмана, правило Хебба і навчання методом змагання.

1. *Правило корекції за помилкою.* У разі навчання з учителем для кожного вхідного прикладу задано бажаний вихід d . Реальний вихід мережі y може не збігатися з бажаним. Принцип корекції за помилкою під час навчання полягає у використанні сигналу $(d - y)$ для модифікації вагів, що забезпечує поступове зменшення помилки. Навчання має місце лише у разі, коли нейронна мережа помиляється. Відомі різні модифікації цього алгоритму навчання [35–37].

2. *Навчання Больцмана* є стохастичним правилом навчання, яке впливає з інформаційних теоретичних і термодинамічних принципів [35, 36]. Метою навчання Больцмана є таке налаштування вагових коефіцієнтів, за якого стани видимих нейронів задовольняють бажаний розподіл імовірності. Навчання Больцмана може розглядатися як спеціальний випадок корекції за помилкою, в якому під помилкою розуміється розбіжність кореляцій станів в двох режимах.

3. *Правило Хебба.* Найстарішим навчальним правилом є постулат навчання Хебба [35, 36]. Хебб спирався на такі нейрофізіологічні спостереження: якщо нейрони з обох боків синапсу активізуються одночасно і регулярно, то сила синаптичного зв'язку зростає. Важливою особливістю цього правила є те, що зміна синаптичної ваги залежить лише від активності нейронів, які пов'язані цим синапсом. Це істотно спрощує «ланцюжки» навчання в реалізації HBIC (VLSI).

4. *Навчання методом змагання*. На відміну від навчання Хебба, в якому безліч вихідних нейронів можуть збуджуватися одночасно, у змагальному навчанні вихідні нейрони змагаються між собою за активацію. Це явище відоме як правило «переможець забирає все». Подібне навчання має місце в біологічних нейронних мережах. Навчання за допомогою змагання дозволяє кластеризувати вхідні дані: подібні приклади групуються мережею відповідно до кореляцій і подаються одним елементом.

Можна відмітити, що мережа ніколи не перестане навчатися, якщо параметр швидкості навчання не дорівнює 0. Деякий вхідний зразок може активізувати інший вихідний нейрон на подальших ітераціях в процесі навчання. Це ставить питання про стійкість навчальної системи. Система вважається стійкою, якщо жоден з прикладів навчальної вибірки не змінює своєї належності до категорії після певної кількості ітерацій навчального процесу. Один із способів досягнення стабільності полягає в поступовому зменшенні до 0 параметра швидкості навчання. Проте, це штучне гальмування навчання викликає іншу проблему, звану пластичністю і пов'язану зі здатністю адаптації до нових даних. Ці особливості навчання методом змагання відомі під назвою дилеми стабільності–пластичності Гроссберга.

Для навчання рекурентних мереж спайкінгових нейронів [8, 11, 25–27] можуть використовуватися будь-які алгоритми навчання з учителем, оскільки необхідно навчити лише вихідні (зчитувальні) нейрони мережі, які проєкціюють «рідкі стани» мікромережі в бажану сукупність вихідних сигналів мережі (цільові виходи). Відзначимо, що в біологічному контексті цільові виходи можуть бути забезпечені серцевинищем (наприклад, в разі завдань передбачення або для апроксимації функцій в підкріплюваному навчанні) так, що навчальний алгоритм контролюється вчителем лише з погляду локальної мережі, але не контролюється з погляду системи. Фактично, для мереж спайкінгових нейронів можуть бути запропоновані нові алгоритми на додаток до безлічі вже відомих (доступних) алгоритмів [8, 37].

1. Дельта-правило – один з найстаріших і найвивченіших алгоритмів нейронного навчання. Він розроблений для навчання одиночного персептрона (порогового вентиля) певному завданню. Через двійковий вихід персептрон обмежений у разі розв'язання завдань класифікації. Привабливі особливості цього алгоритму – його простота і діалоговий характер, який означає, що навчання може бути проведене від прикладу до прикладу і немає необхідності мати весь набір даних навчальних прикладів, доступних одночасно. Недоліком дельта-правила є те, що він не сходиться до стабільного рішення, якщо не існує ідеального розв'язку задачі (тобто, якщо дані лінійно нероздільні).

2. Машина опорних векторів (support vector machine – SVM) – метод мінімізації структурного ризику. SVM (в разі нелінійного SVM вхід спочатку нелінійно перетворюється в багатовимірний простір ознак, де потім визначається оптимальний вектор вагів) – більш прогресивний алгоритм

для знаходження параметрів (вагів) одиночного перцептрона, застосований до віртуального нелінійного проєкціювання даних в багатовимірний простір. Основна ідея – вибрати серед всіх можливих вагових наборів один, який з найменшою вірогідністю є перенавченим. Перевага SVM у тому, що він гарантує знаходження оптимального набору вагів (за допомогою квадратичного програмування). Ціна, яку доводиться платити за це, – великий час обчислень, якщо розмірність і обсяг навчальних даних істотний.

3. *P*-дельта-правило навчання. *P*-дельта-правило є узагальненням дельта-правила, яке навчає популяцію перцептронів, щоб налаштувати заданий відгук популяції (в термінах числа перцептронів, які перевищують поріг). На відміну від інших розподілених алгоритмів нейронного навчання він потребує набагато менше комунікацій між нейронними модулями. Результуюча вихідна функція може також бути реалізована як масив незв'язаних спайкінгових нейронів. Вочевидь, що n перцептронів можуть виводити не лише бінарні рівні, але і числа в інтервалі між 0 і 1 (у дискретних рівнях, кратних $1/n$). Насправді, масиви перцептронів мають універсальну властивість апроксимації, тобто вони можуть апроксимувати будь-яку задану неперервну функцію на будь-якій обмеженій області з будь-якою бажаною мірою точності. *P*-дельта-алгоритм навчання в деякому розумінні подібний SVM-алгоритму, оскільки він явно намагається уникати перенавчання. Одним з можливих недоліків є те, що користувачеві доведеться тонко розуміти взаємний вплив численних параметрів, які йому доведеться встановлювати. Проте, застосовуючи параметри, використовувані за замовчуванням, можна отримати вельми гарні результати.

4. Лінійна регресія (метод найменших квадратів). Лінійна регресія є, ймовірно, одним з найпростіших «навчальних» алгоритмів. З цього алгоритму потрібно починати під час дослідження різних алгоритмів навчання, оскільки він вельми швидкодіє, немає жодних параметрів для налаштування, кожен знає точно, що обчислюється, і якщо є досить даних, то перенавчання малоімовірно. До того ж він може бути застосований як до бінарних (вихід просто піддається пороговій обробці), так і до аналогових цільових функцій.

5. Алгоритм зворотного поширення. Алгоритм зворотного поширення – найпопулярніший алгоритм для навчання багатошарових штучних нейронних мереж з прямими зв'язками. Подібно до *p*-дельта-правила він має перевагу в тому, що може, в принципі, апроксимувати дуже складні цільові функції, якщо вибрано правильну архітектуру мережі (що не є тривіальним завданням). Проте потрібно бути обережним, щоб не використовувати дуже великі мережі і уникнути перенавчання. Детальніше алгоритм зворотного поширення описаний в [35].

Який алгоритм навчання потрібно використовувати? Відповідь значною мірою залежить від типу мікромережі, завдання, яке необхідно реалізувати, і від вигляду результатів, які має видавати мережа. Якщо гарні результати можуть бути отримані з дуже простим виходом, яким є, напри-

клад, простий одиночний перцептрон (наприклад, навчений за дельта-правилом або лінійною регресією), то можна зробити висновок, що мікромережа буде не лише здійснювати всю необхідну часову інтеграцію, але і проєкціюватиме свою інформацію про попередні вхідні сигнали в досить високовимірний простір, щоб полегшити подальше лінійне розпізнавання образу (іншими словами, в цьому випадку мікромережа реалізує додатково досить нелінійне і високороздільне ядро (операцію фільтрації), яке відіграє роль, подібну до ядра SVM).

Потрібно зазначити, що всі вищеперелічені правила навчання використовуються у випадку програмної або програмно-апаратної реалізації нейронних мереж, коли процес навчання здійснюється шляхом виконання обчислень за певними формулами на цифровому процесорі універсального комп'ютера або спецобчислювача. У разі апаратної реалізації нейронних мереж особливе значення має так зване on-line (або on-chip) навчання, тобто навчання за допомогою спеціальних схемотехнічних засобів, які змінюють ваги синаптичних зв'язків нейронних елементів без застосування будь-яких обчислювальних операцій на цифрових процесорах. Як такі схемотехнічні засоби можуть застосовуватися лінійні і нелінійні електронні компоненти з параметрами, що адаптуються. Це, наприклад, варикапи, МДН-транзистори з плаваючим затвором [32], варистори, вольфрам-танталові резистори та інші адаптивні електронні, оптоелектронні й магнітоелектронні схемотехнічні засоби. Оскільки саме апаратні реалізації нейронних мереж найбільш ефективні з погляду швидкодії, габаритів, споживаної потужності та інших показників, у п. 3.3 запропоновано варіант апаратної реалізації спайкінгової нейронної мережі на оптоелектронній елементній базі. А тут ми розглянемо запропоноване в [34] правило навчання для оптоелектронної спайкінгової нейронної мережі, яке можна використовувати зокрема в разі використання спайкінгової нейронної мережі для розпізнавання багатовимірних сигналів.

Саме під час розробки правил навчання апаратних нейронних мереж доречно буде пригадати правило навчання Хебба [10], з якого вирости більшість сучасних алгоритмів навчання. Запропонована ним модель навчання без учителя, в якій синаптична сила (вага) зростає, якщо активовано обидва нейрони, джерело і приймач. Таким чином, часто використовувані шляхи в мережі посилюються і феномен звички та навчання через повторення отримує пояснення.

У штучній нейронній мережі, що використовує навчання за Хеббом, нарощування вагів визначається рівнем збудження передавального і приймального нейронів. Це можна записати як

$$w_{ij}(n+1) = w_{ij}(n) + \alpha OUT_i \cdot OUT_j, \quad (2.11)$$

де $w_{ij}(n+1)$ – значення ваги від нейрона i до нейрона j до підстроювання;

$w_{ij}(n)$ – значення ваги від нейрона i до нейрона j після підстроювання;

α – коефіцієнт швидкості навчання;

OUT_i – вихід нейрона i і вхід нейрона j ,

OUT_j – вихід нейрона j .

У [34] пропонується таке модифіковане правило навчання рекурентної спайкінгової нейронної мережі:

1. Всім вагам зв'язків синапсів вихідних нейронів мережі присвоїти невеликі випадкові значення (наприклад, в діапазоні $0 \dots 0,15$ для збуджувальних і $(-0,15 \dots 0$ – для гальмівних).

2. Номер кроку $i := 0$.

3. $i := i + 1$. Встановити на електродах управління всіх збуджувальних синапсів i -го вихідного нейрона потенціал, який визначає режим, коли кожен імпульс, що прийшов на відповідний синапс, збільшуватиме величину коефіцієнта передачі цього синапсу.

4. Встановити на електродах управління всіх гальмівних синапсів інших $(m - 1)$ нейронів (окрім i -го) потенціал, який визначає режим, коли кожен імпульс, що прийшов на відповідний синапс, збільшуватиме величину негативного коефіцієнта передачі цього синапсу (тобто збільшувати тим самим гальмівний ефект).

5. На вхідні нейрони подати i -й еталонний n -вимірний сигнал тривалістю T_0 . Імпульси, які з'являтимуться на входах збуджувальних синапсів i -го вихідного нейрона, збільшать ваги його зв'язків з тими нейронами мікромережі, які генерують саме за дії на мережу i -го еталонного n -вимірного сигналу. Це приводить до того, що після навчання у разі подачі на вхід СНМ n -вимірного сигналу, схожого на i -й еталон, саме i -й вихідний нейрон видасть максимальну кількість імпульсів на інтервалі T_0 . А імпульси, які з'являтимуться на входах гальмівних синапсів інших вихідних нейронів (окрім i -го), збільшать їх гальмування по зв'язках з тими нейронами мікромережі, які генерують саме за дії на мережу i -го еталонного n -вимірного сигналу.

6. Якщо $i \neq m$, то перейти до п. 3. Інакше – всі сигнали управління з синапсів вихідних нейронів зняти. Протестувати роботу мережі, подаючи на вхід по чергово еталонні n -вимірні сигнали і контролюючи, щоб максимальна кількість імпульсацій за i -го еталону була саме на i -му виході. Якщо ця вимога виконується, то перейти до п. 7. Інакше – змінити випадковим чином черговість подачі еталонів і повторити процес навчання з п. 2.

7. Процес навчання закінчений.

Таким чином, завдяки використовуваному методу розпізнавання на основі СНМ і запропонованому методу навчання, процес навчання в запропонованій нейромережеві системі здійснюється значно простіше. Спрощення виникає, по-перше, за рахунок використання схемотехнічних засобів для навчання і відсутності складних обчислювальних процедур під час навчання нейронної мережі і пристроїв для їх здійснення, по-друге, навчаються не всі нейрони мережі, а лише вихідні нейрони.

2.4 Метод розпізнавання багатовимірних сигналів з використанням спайкінгової нейронної мережі

Головна мета функціонування інтелектуальних систем – вирішувати когнітивні задачі не гірше за людину.

Людина сприймає образи за допомогою органів чуття (очі, вуха, ніс і т. д.), які є багатоконпонентними рецепторами. Ці багатоконпонентні рецептори видають багатоканальну інформацію у вигляді сукупності сигналів (імпульсних). Центральна нервова система (ЦНС) людини має справу не з самим образом, а з результатом дії цього образу на рецептори. Для ЦНС образом є сукупність імпульсних сигналів, які генеруються рецепторами, а потім проходять складну обробку. Тому в процесі побудови нейроморфних інтелектуальних систем нових поколінь велике значення має вирішення проблеми розпізнавання багатовимірних сигналів.

Постановка задачі:

- необхідно розпізнати n динамічних сигналів (сигнали мультисенсорів газів, антенних ґрат, параметри складних динамічних об'єктів і т. ін.);
- всі n сигналів надходять одночасно (паралельно) і обмежені на часовому відрізку;
- результат розпізнавання необхідно отримувати у реальному часі – відразу після закінчення сигналів.

Варіанти вирішення:

- *кореляційні методи* – не підходять, оскільки необхідно багато паралельно працюючих кореляторів;
- *традиційні нейронні мережі* – не підходять, оскільки часто для отримання результату необхідно декілька ітерацій і, крім цього, вони працюють із відліками вхідних сигналів або їх ознаками;
- *спайкінгові нейронні мережі* – підходять, оскільки: 1) пристосовані до розпізнавання динамічних образів без їх попереднього перетворення у вектор ознак; 2) багатозадачні; 3) розпізнають із передбаченням; 4) мають просту процедуру навчання.

Загальна постановка задачі розпізнавання багатовимірних сигналів (часових рядів): є n часових рядів, кожен з яких є залежністю якоїсь величини S_i ($i=1...n$) від часу t : $S_1(t), S_2(t), \dots, S_n(t)$. Необхідно визначити, чи є ці часові ряди на інтервалі часу $T_0 = t_n - t_x$ схожими із заданою мірою точності на один з m n -вимірних еталонних часових рядів, відомих заздалегідь.

Здавалося б, цю задачу можна розбити на n задач з розпізнавання одиничних сигналів. Проте, для того, щоб розпізнавати ці сигнали по черзі на одному розпізнавальному пристрої, їх необхідно запам'ятати, а потім по чергово обробляти. По-перше, у випадку запам'ятовування в цифровому вигляді вони піддадуться дискретизації за часом, квантуванню, а потім ві-

дновленню, що неминуче призведе до спотворення інформації. По-друге, за великого значення n загальна тривалість процесу розпізнавання буде неприйнятно великою. Якщо ж розпізнавати ці сигнали паралельно в часі на n розпізнавальних пристроях, а потім усереднювати результати, то за великого значення n апаратурні витрати виявляться неприйнятно великими. Можливий і третій варіант – розпізнавати з усереднюванням послідовність векторів розмірності n , складених з дискретних в часі відліків цих сигналів. Проте, в цьому випадку час розпізнавання все одно перевищить T_0 і виникнуть спотворення сигналів, пов'язані з їх дискретизацією. Окрім залежності від t ці сигнали в загальному випадку можуть залежати від набору динамічних параметрів C будь-якого процесу або явища, тобто є взаємозв'язаними (корельованими) і якщо піддаються завадам, то цей вплив також корельований. Тому вельми важливо проводити розпізнавання цих сигналів паралельно в часі, а не послідовно або із затримкою на дискретизацію і обробку відліків.

Для розпізнавання багатовимірних сигналів можна було б застосувати відомі методи розпізнавання сигналів, оснований, наприклад, на перетворенні Фур'є [38] або на вейвлет-перетворенні [39] у поєднанні з кореляційними або нейронно-мережевими методами розпізнавання за ознаками. Ці методи є добре дослідженими і широко застосовуються. Але мають 2 істотних недоліки: 1) після закінчення сигналів потрібен деякий час, щоб, поперше, отримати самі коефіцієнти розкладання сигналів в ряд, а, по-друге, ще деякий час аби нейронна мережа ітеративно зійшла до якого-небудь рішення; 2) у випадку переходу від аналогових сигналів до коефіцієнтів розкладання в ряд Фур'є або вейвлет-перетворення втрачається частина корисної інформації, через що з низькою точністю розпізнаватимуться сильно корельовані образи.

Тому в ідеальному випадку для розпізнавання багатовимірних сигналів підійшло б інтелектуальне середовище, в яке можна одночасно ввести сигнали від усіх каналів багатовимірного сигналу в реальному масштабі часу і після закінчення цих сигналів (або навіть до їх закінчення) отримати результат класифікації.

Як таке інтелектуальне середовище можна використовувати так звані спайкінгові нейронні мережі (СНМ) [8, 11, 25–28]. Вони побудовані за аналогією з мережами біологічних нейронів мозку тварин і людини та складаються зі спайкінгових нейронних елементів [8, 16, 23] (3-тє покоління нейронних елементів), які ближчі до біологічних нейронів, ніж бінарні нейрони Мак-Каллока і Пітса (1-ше покоління нейронних елементів) та аналогові нейрони з потенційним виходом (2-ге покоління нейронних елементів).

Враховуючи модифіковану структуру спайкінгової нейронної мережі (див. рис. 2.9), а також адаптоване під використання оптоелектронних схемотехнічних засобів правило навчання вихідних нейронів спайкінгової

нейронної мережі (п. 2.3), принципи функціонування і математичні моделі спайкінгових нейронів з роздільними входами (п. 2.1.2), можна сформулювати метод розпізнавання багатовимірних сигналів з використанням спайкінгової нейронної мережі.

Цей метод полягає в такому:

1. Створити (згенерувати) рекурентну спайкінгову нейронну мережу, складену з інтернейронів, в кількості $N \geq 15m$, де m – кількість еталонних образів (в цьому випадку багатовимірних сигналів), які мережа має «запам'ятати». З'єднання нейронів в мережі виконати за даними нейрофізіологічних досліджень. Ваги зв'язків нейронів вибрати випадковими.

2. Сформулювати n вхідних нейронів за кількістю каналів багатовимірною сигналу. З'єднати кожен з них випадковим чином з не менше, ніж k нейронами мікромережі ($n < k < N$). Ваги зв'язків вибрати випадковими.

3. Сформулювати m вихідних нейронів за кількістю образів, які необхідно розпізнавати. З'єднати кожен з них випадково з не менше, ніж l нейронами мікромережі ($m < l < N$). Ваги зв'язків вибрати випадковими.

4. Застосувати алгоритм навчання, описаний в попередньому пункті (п. 2.3). Водночас підстроюються лише ваги зв'язків кожного з m вихідних нейронів. Як ідеальний цільовий вихід $y(t)$ може виступати імпульсний сигнал з постійною (максимальною) частотою імпульсів ν_{max} , що дорівнює лабільності вихідного нейрона. Фактичний вихідний сигнал $f(x(t))$ буде послідовністю імпульсів з довільними часовими проміжками між ними. Вхідні сигнали мережі тим ближчі до еталонного образу, чим більша середня за період розпізнавання частота імпульсів вихідного нейрона, що відповідає цьому еталонному образу.

5. Подати на вхід мережі досліджуваний n -вимірний сигнал і зафіксувати, який з m вихідних нейронів видаватиме максимальну кількість імпульсів. Саме цей нейрон і визначає еталонний образ, якому максимально відповідають вхідні сигнали мережі. Як оцінка міри схожості вхідних сигналів і еталонного образу може використовуватися відношення середньої за період розпізнавання частоти імпульсів вихідного нейрона до ν_{max} .

Розглянемо, яким чином в запропонованому методі розпізнавання багатовимірних сигналів здійснюється підвищення швидкодії порівняно з традиційними нейромережевими методами. Звичайним нейромережевими системам [3, 11, 35] розпізнавання потрібен після закінчення періоду вимірювання багатовимірних сигналів деякий час на видачу результату розпізнавання. За цей час система має виділити вектор ознак, подати його в нейронну мережу і почекати, поки мережа зійдеться до якогось результату. Зазвичай, це час від декількох мілісекунд до одиниць секунд залежно від обсягу мережі і продуктивності використовуваних обчислювальних засобів. Тому для традиційних нейромережесистем (на основі багатошарових перцептронів, мереж Кохонена, ART-мереж і ін.) загальний час розпізнавання знаходиться за формулою:

$$T_{3AG} = T_{CM} + T_{\Phi O} + T_{OO}, \quad (2.12)$$

де T_{CM} – період тривалості багатовимірних сигналів;
 $T_{\Phi O}$ – час формування ознак для розпізнавання;
 T_{OO} – час обробки ознак розпізнавання нейронною мережею (рис. 2.10).

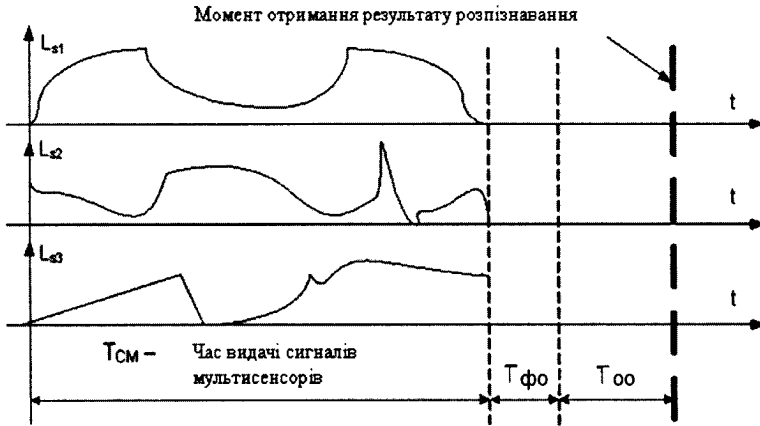


Рисунок 2.10 – Часова діаграма, що пояснює процес розпізнавання багатовимірних сигналів традиційною нейронною мережею

У запропонованій нейромережевій системі результат розпізнавання можна отримати ще до закінчення періоду вимірювання багатовимірних сигналів за найінтенсивнішою імпульсацією на одному з вихідних нейронів. Тому для запропонованої нейромережевої системи на основі спайкінгових нейронних мереж загальний час розпізнавання

$$T'_{3AG} \leq T_{CM}. \quad (2.13)$$

Як оцінку виграшу за швидкодією запропонованої нейромережевої системи можна використовувати коефіцієнт k , який дорівнює відношенню часу розпізнавання багатовимірних сигналів нейромережевими системами на основі традиційних нейронних мереж T_{3AG} до часу розпізнавання багатовимірних сигналів запропонованою нейромережевою системою на основі спайкінгових нейронних мереж T'_{3AG} . Таким чином, час розпізнавання в запропонованій нейромережевій системі на основі спайкінгових нейронних мереж в k разів менший, ніж в традиційних нейромережевих системах:

$$k \geq \frac{T_{3AG}}{T'_{3AG}} = 1 + \frac{T_{\Phi O} + T_{OO}}{T_{CM}}. \quad (2.14)$$

Наприклад, за $T_{CM} = 1$ с, а $(T_{\Phi 0} + T_{00}) = 0,1 \dots 0,5$ с, $k \geq 1,1 \dots 1,5$, тобто виграш за швидкодією становить 10...50 %.

Запропонований метод має такі переваги:

- 1) сигнали подаються на розпізнавання одночасно без розділення в часі або в просторі, без дискретизації, а також без попереднього запам'ятовування, що перешкоджає втраті інформації, а тому – підвищує точність і достовірність розпізнавання (особливо у разі розпізнавання сильно скорельованих образів);
- 2) результат розпізнавання може бути оцінений ще до завершення періоду дії самих сигналів за інтенсивною імпульсацією на відповідному виході (розпізнавання з передбаченням), що підвищує швидкодію;
- 3) використовувана нейронна мережа має неважку процедуру навчання;
- 4) підвищено завадостійкість за рахунок подання інформації в частотно-імпульсній формі.

2.5 Контрольні запитання

1. Назвіть функціональні можливості біологічного нейрона.
2. Охарактеризуйте залежність сила–тривалість нейрона.
3. Дайте означення параметрів реобазиса та хронаксія нейрона. Що вони характеризують?
4. Опишіть схему та математичну модель формального LIF-нейрона.
5. Назвіть недоліки формальної моделі LIF-нейрона та його відмінності від біологічного нейрона.
6. Наведіть схему спайкінгового нейронного елемента з розділними входами та опишіть її роботу.
7. Наведіть схему та математичну модель інтернейрона для СНМ на основі біспін-приладу.
8. Проаналізуйте модель і структуру спайкінгової нейронної мережі.
9. Які є недосліджені моменти і недоліки в розробленій нині парадигмі рекурентних мереж спайкінгових нейронів?
10. Наведіть вдосконалену структуру спайкінгової нейронної мережі.
11. Які переваги має вдосконалена структура спайкінгової нейронної мережі перед традиційною?
12. Назвіть три фундаментальні властивості навчання за прикладами.
13. Охарактеризуйте чотири основних типи правил навчання.
14. Що таке дилема стабільності–пластичності Гроссберга?

15. Назвіть відомі алгоритми навчання для мереж спайкінгових нейронів.
16. Що таке on-line (або on-chip) навчання нейронних мереж?
17. Сформулюйте та запишіть математичну модель правила навчання Хебба.
18. Сформулюйте модифіковане правило навчання рекурентної спайкінгової нейронної мережі.
19. Сформулюйте постановку задачі та назвіть основні варіанти вирішення задачі розпізнавання багатовимірних сигналів.
20. Сформулюйте метод розпізнавання багатовимірних сигналів з використанням спайкінгової нейронної мережі.
21. Яким чином в запропонованому методі розпізнавання багатовимірних сигналів здійснюється підвищення швидкодії порівняно з традиційними нейромережевими методами?
22. Які переваги має запропонований метод перед традиційними нейромережевими методами?

3 ПРОЕКТИ З АПАРАТНОЇ РЕАЛІЗАЦІЇ СПАЙКІНГОВИХ НЕЙРОННИХ МЕРЕЖ

Як було показано у п. 1.3 спайкінгові нейронні мережі є найперспективнішою архітектурою для побудови операційного ядра нейрокомп'ютерів. Розглянемо, які є нині найпотужніші проекти в галузі створення (апаратної реалізації) спайкінгових нейромереж. Їх є багато в різних країнах світу, але можна виділити 5 основних (табл. 3.1).

3.1 Огляд відомих проектів з апаратної реалізації спайкінгових нейронних мереж

Розглянемо, які є нині найпотужніші проекти в галузі створення (апаратної реалізації) спайкінгових нейромереж. Їх є багато в різних країнах світу, але можна виділити 5 основних, які подано у табл. 3.1.

Найбільш відомим є Human Brain Project [40] (табл. 3.1), 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 цьому проекті використовується суперкомп'ютер «Blue Gene» продуктивністю 22,8 трильйонів операцій/с з програмним забезпеченням «NEURON». Хоча проект і не стосується безпосередньо проблем ефективної апаратної реалізації спайкінгових нейромереж, все ж він базується саме на застосуванні спайкінгових нейронів та структур нейронів, аналогічних біологічним прототипам, і має велике значення для апробації принципів структурної організації та методів навчання спайкінгових нейронних мереж, які потім можна буде використати під час побудови апаратних реалізацій.

До недоліків та слабких місць проекту можна віднести таке:

- 1) це все-таки комп'ютерне моделювання, а не апаратне;
- 2) низька швидкодія моделювання: 1 с роботи реального нейрона забирає 80 хвилин машинного часу (для моделювання людського мозку в цілому потрібно комп'ютер в тисячу разів більш потужний, 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.

Таблиця 3.1 – Найпотужніші проекти в галузі створення спайкінгових нейромереж

Назва проекту	Human Brain Project (до цього <i>Blue Brain Project</i>)	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 neuroscinces	Brain-inspired multiscale computation in neuromorphic hybrid systems
Країна	Швейцарія	Великобританія	США	США	Євросоюз
Організація	ІВМ та Швейцарський Федеральний Технічний Інститут Лозанни	Манчестерський ун-т, Університет Саутгемптона	1) ІВМ, 2) НР 3) провідні університети (Колумбійський та ін.)	1) ІВМ, 2) Стенфордський ун-т та ін.	18 дослідницьких груп (зокрема 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. Квабена Боен (Kwabena Boahen), Стенфордський ун-т	Проф. Карлхайнц Майер (Karlheinz Meier), ун-тет Гейдельберг, Німеччина
Джерело фінанс.	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 фінанс.	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 в продовження
Початок	2005-2008 Blue Brain Prj, 2012 р - Human Brain Prj	2005	2008	2006	2005-2010 - FACETS, 3 2011 - BrainScaleS
Закінч	2023	2014...	2016...	2013...	2014 ...
Ел. база	Суперкомп'ютер «Blue Gene» з програмним забезпеченням «NEURON», яке написано Майклом Хайнсом (Michael Hines).	Мікропроцесори (МП) ARM968 (130 нм, 1 чип – 20 МП, 1 МП=1000 нейронів). 20000 нейронів на чип.	Цифрові НВІС (45 нм). У чипі (4,2 мм ²) 256 спайк-х нейронів на осн. цифрових суматора та компаратора. Один нейрон 35×95 μm ² (для порівняння біологічний нейрон 4÷100 μm в Ø).	Аналогові НВІС на основі к-МДН. Нейроморфний чип Neurocore містить 65536 (256х256) штучних нейронів на площі 11,9×13,9 мм ² .	Аналогові НВІС (180 нм) на основі к-МДН. Один чип (5×10 мм ²) реалізує до 512 спайкінгових нейронів . Нейрони аналогові, а сигнали зв'язку між чипами і сигналами 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 чипів, в кожному чипі 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 чипи, у чипі 512 спайк. нейронів (близько 200 000 нейронів на підкладку).

Продовження таблиці 3.1

Назва проекту	Human Brain Project (продовж. Blue Brain Proj)	SpiNNaker	SyNAPSE (DARPA)	Neurogrid (Brains in silicon group)	BrainScaleS (продовж. FACETS 2005-2010)
К-ть нейронів в пер-спект.	2014 р змодельовано мозок пацюка (100 млн нейронів). Мозок людини (100 млрд нейронів) - до 2023 р.	В перспективі 500 чипів, і наразті 5000 чипів (10 ⁸ нейронів – майже 1% мозку людини) до 2014	Довготривала мета - створення системи з 10 ¹⁰ нейронів і 10 ¹⁴ синапсів, що займає об'єм < 2 літрів та споживає 1 кВт.	Наступне покоління Neurogrid має 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 ⁷ синапсів (2006 р.) 3000 синапсів на 1 нейрон	1000 синапсів на 1 нейрон	1280 бінарних синапсів/нейрон (262 тис. незмінних та 65536 змінних)	6 млрд синапсів (~6000 синапсів на 1 нейрон)	256 синапсів на 1 нейрон (49 млн синапсів на підкладку)
Модель нейрона	Пірамідальний нейрон	Спайкінговий 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	Повнозв'язна	Hopfield-like, Restricted Boltzmann machine (RBM)	2-D рекурентна мережа зі ступенем зв'язності «мексиканський капелюх»	Здається програмно
Продукт-ть	1,25×10 ⁴ імп/с (за продуктивності комп. 22,8×10 ¹² оп/с)	час передачі імпульсу 1 мс, 10 ⁹ імп/с	дискретний часовий крок 1 мс, 10 ⁹ імп/с	9,1×10 ⁷ імп/с	10 ¹¹ імп/с
Спож. Потуж.	<17 Вт/нейрон>	1 Вт на 1000 нейронів (1мВт/нейрон)	0,5 мВт/нейрон за 0,53В поогової напр. МДН	3,1 Вт на 10 ⁶ нейронів 3,1 мкВт/нейрон	1 кВт на підкладку, тобто 5 мВт/нейрон
Недоліки	Це все ж комп'ютерна, а не апаратна реалізація,	Немає відокремлених нейронів	Синапси бінарні	Обмежена зв'язність («мексиканс. капелюх»)	Не буває на підкладці 100% придатних чипів
	Відсутні безпосередні зв'язки між нейронами, оскільки технологічно неможливо створення великої кількості електричних ліній зв'язку в площині напівпровідникового кристала. Безпосередні зв'язки замінюються організацією штучних програмованих чи спеціально кодованих протоколів обміну між групами нейронів, які дозволяють зменшити кількість електричних ліній зв'язку в площині напівпровідникового кристала, 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 спожив. потужн.					велика спож. потужність

Проект SpiNNaker [41] (див. табл. 3.1) вже присвячений суто апаратній реалізації спайкінгових нейромереж за допомогою добре розвиненої на сьогодні елементної бази мікропроцесорів. В 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 чип. На кінець 2011 року було об'єднано у систему 50 чипів. Це близько 1 млн (950 тис) нейронів: 10^6 .

Системна архітектура SpiNNaker виконана у вигляді двовимірної тороїдальної мережі (рис. 3.1) мікропроцесорів (Chip Multiprocessor, CMP), пов'язаних між собою за допомогою Ethernet. Ця мережа має трикутні комірки, щоб можна було утворювати шляхи для обходу дефектних вузлів. Кожен вузол (рис. 3.2) CMP складається з двох чипів: система на кристалі SpiNNaker MPSoC та гігабітна мікросхема пам'яті DDR SDRAM. Так само 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.

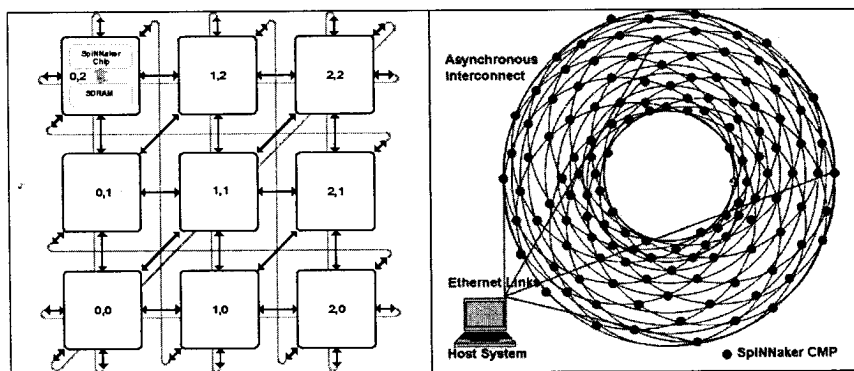


Рисунок 3.1 – Системна архітектура SpiNNaker

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 та 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 в межах системи, що містить мільйон процесорів. Водночас час передачі імпульсу становитиме близько однієї мілісекунди, що приблизно відповідає швидкості передачі нервових імпульсів реальними нейронами в мозку людини.

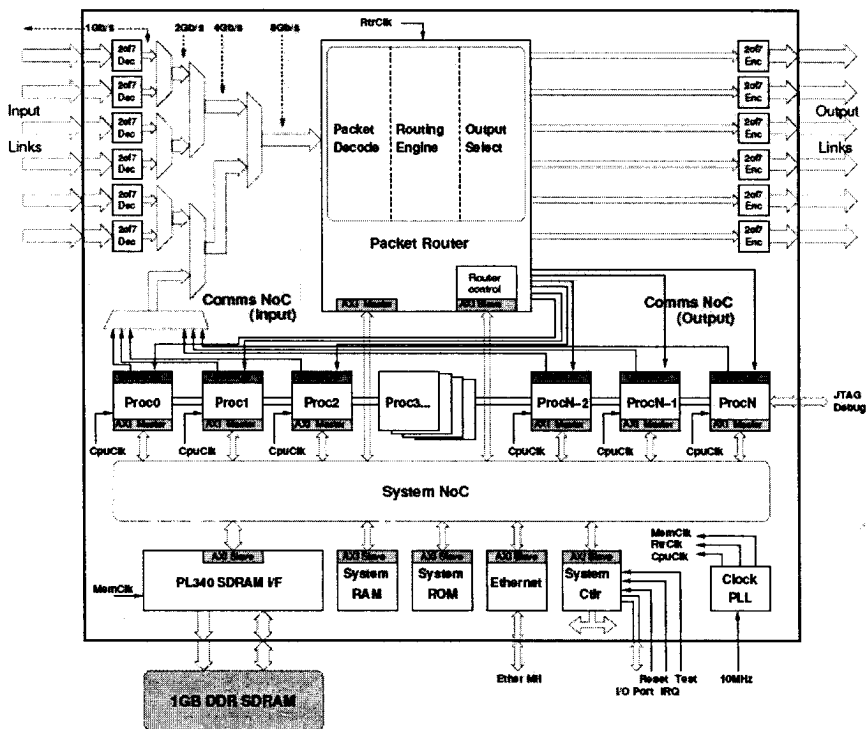


Рисунок 3.2 – Схема вузла системи SpiNNaker

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

2) у біологічному мозку під час 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 принципами, 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 (5×10^{10} нейронів) знадобиться 50 мегаВт).

Метою програми SyNAPSE (DARPA) [42] (див. табл. 3.1) 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. 3.3), кожне з яких має 256 спайкінгових (integrate-and-fire) нейронів (на цифрових елементах) і решітку 1024×256 біт SRAM пам'яті для організації синапсів. Для цього було використано NBIC технологію «кремній-на-ізоляторі» фірми IBM 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 погляду параметрів нейронів (рис. 3.4), типів аксонів і ваг синапсів, а тому потенційно має широкий спектр застосування.

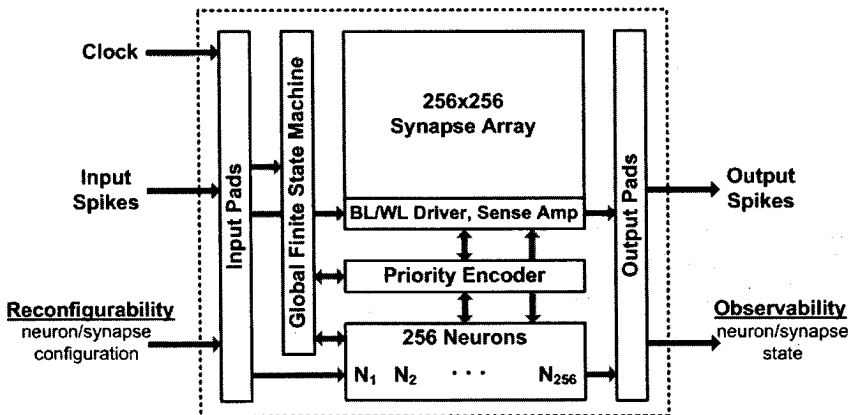


Рисунок 3.3 – Схема нейросинаптичного ядра системи SyNAPSE

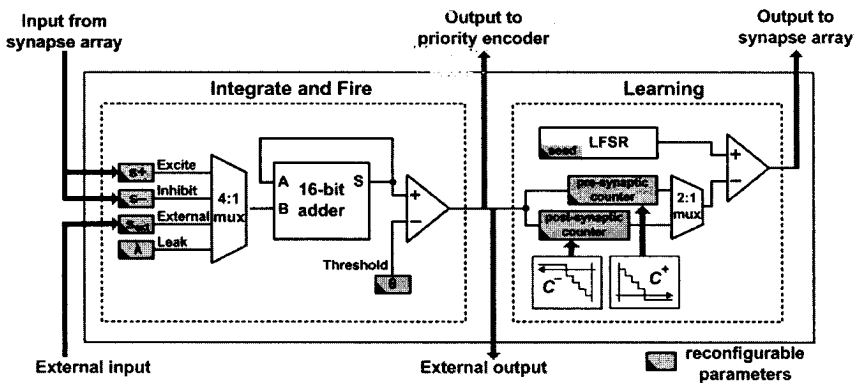


Рисунок 3.4 – Схема цифрового к-МДН нейрона системи SyNAPSE, який має засоби реконфігурації

Система SyNAPSE (DARPA) порівняно із системою SpiNNaker більшою мірою адекватна біологічним нейронним мережам, тому що складається із штучних нейронів (нейроелементів), а не із 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 в поточному часовому інтервалі (~1 мс), будуть передані на відповідні дендрити інших нейронів в наступному часовому інтервалі. Зазначені факти і 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 спрацювання. 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 у разі передачі інформації між нейронами відсутня синхронізація (тактування), тобто це є асинхронна система. Але у SyNAPSE існує синхронізація під час передачі імпульсів (імпульси передаються у тактовані проміжки часу із затримкою на 1 такт). Це знову ж та-

ки 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 [43] (див. табл. 3.1) схожий за метою та підходом 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. 3.5, 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. 3.6). Нейроморфний чип Neurocore (рис. 3.5, а) містить 65536 (256×256) штучних нейронів на площі 11,9×13,9 мм². Прототип Neurogrid складається з 16 таких «нейроядер», які зібрані на платі 6,5×7,5 дюймів (рис. 3.5, 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. 3.5, 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.

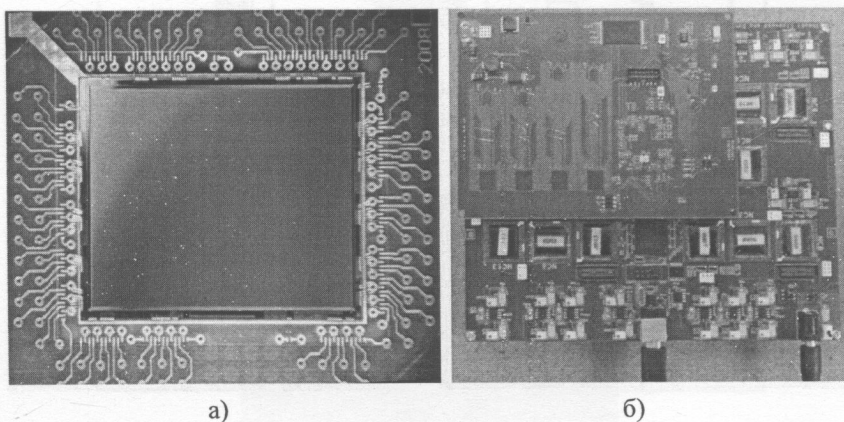


Рисунок 3.5 – Вигляд чипа нейросинаптичного ядра NeuroCore з 256×256 нейронами (а) та плата NeuroGrid, яка містить 16 ядер NeuroCore (б)

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 під час навчання здійснюється 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 в чипі величезна, мережа нейронів не є вільнозв'язною, а є 2-D мережею зі ступенем зв'язності «мексиканський капелюх» (локальне збудження і дистальне гальмування), що не є універсальним, а тому знижує функціональні можливості.

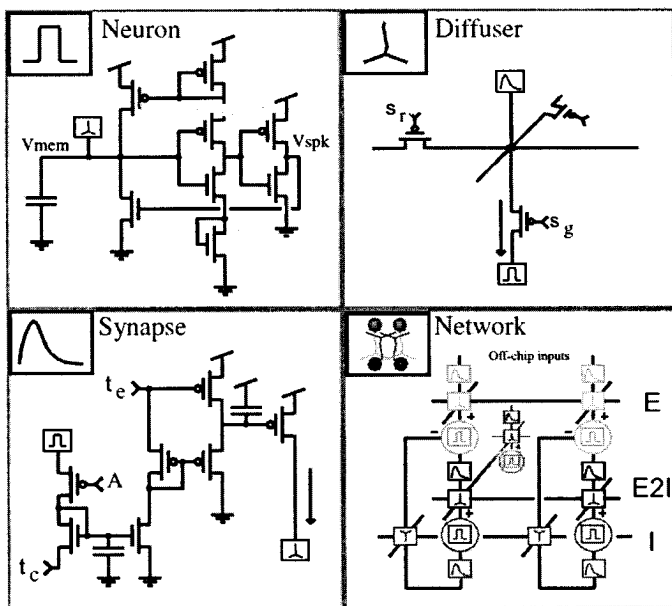
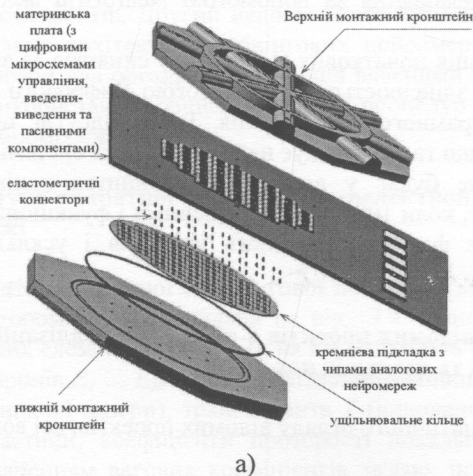


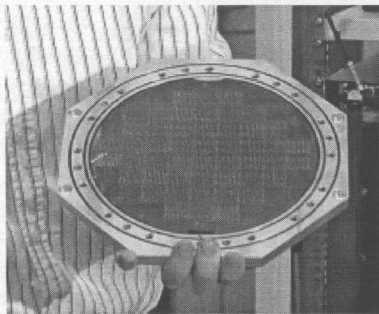
Рисунок 3.6 – Транзисторна реалізація в системі NeuroGrid аналогового к-МДН нейрона, дифузера та синапса. Схема «Network» показує, як ці окремі блоки організовані в нейронну мережу

Проект BrainScaleS [44] (див. табл. 3.1) має на меті зрозуміти принципи обробки інформації в мозку на різних рівнях, починаючи від окремих нейронів до цілих функціональних ділянок мозку. Дослідження містять в своєму складі 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. 3.7) містить 384 чипи, кожен з яких реалізує 128 000 синапсів і до 512 спайкінгових нейронів. Це дає в сумі близько 200 000 нейронів і 49 мільйонів синапсів на підкладку.



а)



б)

Рисунок 3.7 – Конструкція (а) та вигляд (б) модуля підкладки нейроморфних чипів системи BrainScaleS

Нейрони – аналогові, а синапси і сигнали зв'язку між чипами – цифрові. НВІС нейрони працюють значно швидше, ніж біологічні оригінали. Це дозволяє моделювати нейронні мережі, що працюють в десятки тисяч разів швидше, ніж реальний час. Підкладка підтримується алюмінієвою пластиною, яка також слугує для відведення тепла. Багатошарова друкована плата встановлюється на поверхню підкладки і слугує інтерфейсом вхід/вихід для нейронної схеми. Більші мережі можна побудувати з'єднанням декількох підкладок у модуль.

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 під час навчання 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 у випадку виробництва напівпровідникових підкладок випадку, коли 100% чипів є якісними і функціонально бездоганними, що знижує фактичну кількість нейронів і ускладнює роботу з підкладкою.

3.2 Недоліки відомих проектів з апаратної реалізації спайкінгових нейронних мереж та шляхи їх подолання

Як видно з аналітичного огляду відомих проектів, всі вони мають 2 головних недоліки:

1) відсутні безпосередні зв'язки між нейронами, оскільки технологічно неможливо створення великої кількості електричних ліній зв'язку в площині напівпровідникового кристала. Безпосередні зв'язки замінюються організацією штучних програмованих чи спеціально кодованих протоколів обміну між групами нейронів, які дозволяють зменшити кількість електричних ліній зв'язку в площині напівпровідникового кристала, але в такому разі 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 потребує ізоляції між сигнальними шляхами, світлові потоки можуть проходити один через одного без взаємного впливу. Більш того, сигнальні шляхи можуть бути розташовані в 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 спайкінгових нейромереж, які розраховані на оптоелектронну реалізацію, за якої зв'язки між нейронами будуть здійснюватись за допомогою оптичних сигналів. В перспективі можливо буде використовувати і наноелектронну елементну базу в процесі винайдення способів тривимірного конструювання наноелементів. Другий недолік може бути виправлений шляхом розробки таких архітектур спайкінгових нейромереж, які мають у своєму складі засоби для організації навчання власними схемотехнічними засобами без використання обчислювальних процедур (за аналогією з біологічними нейромережами).

3.3 Архітектура апаратної реалізації оптоелектронної спайкінгової нейронної мережі

Структура оптоелектронної спайкінгової нейронної мережі (СНМ), запропонована автором [13, 14], подана на рис. 3.8. Вона складається з n вхідних нейронних елементів $E_1 \dots E_n$, на які надходять 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) [33] 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 ваг зв'язків всіх нейронів СНМ.

Запропонована оптоелектронна спайкінгова нейронна мережа працює таким чином. У режимі розпізнавання вхідні оптичні сигнали $x_i(t) \dots x_n(t)$ надходять на входи відповідних сенсорних нейронів $E_1 \dots E_n$, які перетворюють ці сигнали в 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 T і ОКТ. У режимі розпізна-

вання між всіма електродами Y_i^E і $Y_i^{E'}$, також як і між всіма електродами Y_i^I і $Y_i^{I'}$ ($i=1...m$) – нульові різниці потенціалів, за 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 та дорівнюють 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 із стовпців цих транспарантів. Масив смугових фотодіодів формує електричні сигнали, які керують світловипромінювачами інтернейронів $E_1...E_N$ (вектор XN) і подаються на них за допомогою згута. У інтернейронах сигнали вектора XN (серед яких є і збуджувальні, і гальмівні сигнали) піддаються часовій інтеграції, внаслідок чого формуються сигнали YN , які знову надходять через транспаранти Т і ОКТ на входи інтернейронів $E_1...E_N$. Функціонування здійснюється повністю асинхронно, на відміну від традиційних нейронних мереж на формальних нейронах Маккалока-Пітса з потенційними виходами.

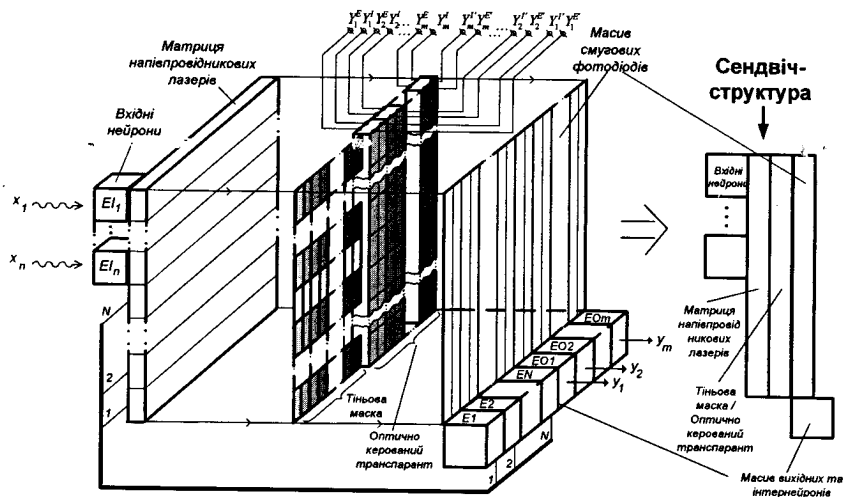


Рисунок 3.8 – Структура оптоелектронної слайкінгової нейронної мережі

Оптичні сигнали, зібрані масивом смугових фотодіодів із стовпців транспаранта ОКТ, є вхідними сигналами (вектор XO) для вихідних нейронів $EO_1...EO_m$. Останні здійснюють часову інтеграцію збуджувальних і гальмівних сигналів, що містяться у векторі XO , і формують вихідні сиг-

нали СНМ (вектор \bar{Y}_0). Наявність максимально інтенсивної імпульсації на j -тому вихідному нейроні свідчить про те, що аналізовані вхідні сигнали максимально схожі на j -тий еталонний образ. За відношенням середньої за період вимірювання частоти імпульсації цього вихідного нейрона до максимально можливої частоти можна робити висновок про ступінь відповідності аналізованого і еталонного образів.

У режимі навчання [12] необхідно послідовно подавати на входи еталонні набори сигналів $x^s_j(t) \dots x^s_n(t)$ ($s=1 \dots S$), відповідні кожному з S еталонів розпізнаваних образів і відповідним чином міняти керувальні потенціали між електродами Y^E_i і $Y^{E'}_i$, 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^E_i і $Y^{E'}_i$ подають від блока управління (на рис. 3.8 не показаний) позитивну різницю потенціалів $+U$; на електроди Y^I_i і $Y^{I'}_i$, а також на електроди Y^E_i і $Y^{E'}_{i'} (i=1 \dots m, i \neq i')$ – нульову різницю потенціалів, а на електроди Y^I_i і $Y^{I'}_{i'} (i=1 \dots m, i \neq i')$ – від'ємну різницю потенціалів $-U$. Це приводить до того, що оптичні імпульси, які надходять 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 надходять на гальмівні входи інших нейронів, збільшують ваги гальмівних зв'язків, що в режимі розпізнавання викликатиме загальмовування 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. Водночас на входи нейронів $EI_1 \dots EI_n$ подають сукупність еталонних сигналів S -го образу, на електроди Y^E_s і $Y^{E'}_s$ подають від блока управління (на рис. 3.8 не показаний) позитивну різницю потенціалів $+U$; на електроди Y^I_s і $Y^{I'}_s$, а також на електроди Y^E_i і $Y^{E'}_{i'} (i=1 \dots m, i \neq s)$ 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 необхідно протестувати роботу мережі, подаючи на вхід по чергово еталонні образи і контролюючи, щоб максимальна інтенсивність імпульсації за i -того еталону було саме на i -тому виході. Якщо ця вимога виконується, то закінчити навчання. Інакше – змінити випадковим чином порядок подання 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 зовнішніх цифровій пам'яті для зберігання синаптичних коефіцієнтів навчених нейронів.

3.4 Оцінення технологічних параметрів оптоелектронної спайкінгової нейронної мережі

Запропонована реалізація оптоелектронної спайкінгової нейронної мережі виконана у гібридному вигляді, тобто поєднує оптичні двовимірні просторово-неперервні структури (ОКТ [33]) і електронні (НВІС) компоненти. Як видно з рис. 3.8, цю 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 Т 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 $E_1 \dots E_n$. 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 жгута дротів, що йде на матрицю лазерів.

Для оцінення розміру площі такої «сандвіч-структури», що припадає на 1 нейрон, потрібно визначити роздільну здатність всіх складових. Так, напівпровідникові лазери виготовляються з апертурою 8–10 мкм [45], ОКТ має роздільну здатність 100–200 лін/мм (5–10 мкм) [33], смугові фотодіоди можна виконати шириною 6–8 мкм [46]. 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 [40–44] наведено у табл. 3.2, звідки видно, що запропонована СНМ майже за всіма кількісними параметрами не поступається відомим, а за більшістю якісних показників (зв'язність, окремі нейрони, безпосередні зв'язки, асинхронність, on-chip навчання і под.) навіть переважає їх.

3.5 Перспективи подальших досліджень

Як результат аналітичного огляду сучасних проектів з апаратної реалізації спайкінгових нейромереж було з'ясовано, що на сьогодні вони всі використовують технологію електронних надвеликих інтегральних схем (НВІС), яка є добре розвиненою і апробованою, а тому зручною. Деякі проекти використовують як елементну базу мікропроцесори, деякі – цифрові НВІС, деякі – аналогові НВІС.

Таблиця 3.2 – Порівняння параметрів оптоелектронної реалізації СНМ з іншими 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 (256×256)	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	19×19 мм ² (в корп.)	4,2 мм ²	11,9×13,9 мм ²	5×10 мм ²	30×30 мм ² *
	Чип встановлюється в корпус, а корпус – на плату з додатковими обслуговуючими електронними схемами. Тому розміри системи, загалом, великі (-)				Плата з додатковими електронними схемами не потрібна. Тому розміри системи відносно чипа не збільшуються (+)
К-ть нейронів в мережі	1 млн (950 тис)	До 2013 р 1 млн	1 млн	200 000	1 млн*
	необхідність синхронізації (-)				асинхронна система (+)
К-ть синапсів на 1 нейрон	Комунікаційний контролер	1280 синапсів. Всі синапси бінарні	6000	256	3000*
	синапси цифрові (-)	синапси бінарні (-)	синапси неперервнозначні (+)		
Продуктивність	10 ⁹ імп/с	10 ⁹ імп/с	9,1×10 ⁷ імп/сек	10 ¹¹ імп/с	(0,2...1)×10 ¹¹ *
Споживана потужність на 1 нейрон	1 мВт	0,5 мВт (45рJ/spike)	3,1 мкВт	5 мВт	10..20 мкВт*
	навчання здійснюється за допомогою цифрового комп'ютера (-)				on-chip навчання (+)

* – оціночні параметри
 (-) – позначає недолік
 (+) – позначає перевагу

Було виділено 2 головних недоліки всіх проектів: 1) відсутність безпосередніх зв'язків між нейронами, оскільки технологічно неможливо 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 в площині напівпровідникового кристалу, але втрачається відповідність принципам роботи біологічного мозку); 2) навчання спайкінгових нейронних мереж відбувається за допомогою цифрових комп'ютерів та спеціального програмного забезпечення, а не за допомогою власних непрограмних механізмів і засобів, не пов'язаних з обчисленнями.

Варіант апаратної реалізації оптоелектронної спайкінгової нейронної мережі за рис. 3.8, запропонований в [47], вільний від цих недоліків. Перший недолік усунуто завдяки використанню оптичних сигналів для організації зв'язків між нейронами, оскільки світлові промені не потребують ізоляції між сигнальними шляхами, можуть проходити один через одного без взаємного впливу, можуть розташовуватись у трьох вимірах та працювати одночасно, забезпечуючи величезний темп передачі даних. Другий недолік виправлено завдяки організації навчання запропонованої спайкінгової нейронної мережі за допомогою апаратних засобів (ОКТ) без використання обчислювальних процедур, причому також існує здатність мережі донавчатись і перенавчатись (адаптивність).

Варіант апаратної реалізації оптоелектронної спайкінгової нейронної мережі за рис. 3.8 побудовано на основі підходу, що передбачає наявність одновимірних вхідного та вихідного сигналів та двовимірної матриці вагових коефіцієнтів у вигляді двовимірного фотографічного транспаранта (для фіксованих зв'язків) та ОКТ (для зв'язків, що модифікуються).

Більш перспективним є підхід, який передбачає наявність двовимірних вхідного та вихідного сигналів і чотирирівірної матриці вагових коефіцієнтів у вигляді голографічного середовища, в якому записуються динамічно модифіковані тривимірні голограми.

Ще однією перспективною технологією апаратної реалізації нейронних мереж, яка згодом прийде на зміну оптики та оптоелектроніки, є нанотехнологія, основана на молекулярному складанні квантових елементів (квантових польових транзисторів, резонансно-тунельних діодів, одновимірних квантових провідників, квантових точок і под.) за допомогою модернізованого сканувального тунельного мікроскопа (СТМ).

Нанотехнологія здатна забезпечити не тільки фізично гранично можливий ступінь інтеграції субмолекулярних квантових елементів (СМКЕ, розміри СМКЕ 1...10 нм, а міжатомні відстані 0,3...0,5 нм) із фізично гра-

нично можливою швидкістю 10...100 ТГц, але і так необхідну для апаратної реалізації нейронної мережі тривимірну архітектуру.

Нанотехнологія складання СМКЕ за допомогою модернізованого СТМ не є інтегральною технологією. Субмолекулярні квантові елементи складаються зондом СТМ (тактова частота 1 МГц) послідовно один за одним і тут-таки тестуються. Якщо СМКЕ, що тестується, «не вийшов», його можна одразу розібрати, скласти наново і знову протестувати. Таким чином можна забезпечити 100%-й вихід придатних виробів.

Наростивши на підкладку один двовимірний шар СМКЕ, що апаратно реалізує двовимірну матрицю входів або нейронів, можна його замаскувати та виростити другий двовимірний шар СМКЕ, що апаратно реалізує другу двовимірну матрицю нейронів. Під час вирощування маскувального шару, в нього можна вбудувати чотиривимірну систему одновимірних квантових провідників, що апаратно реалізують чотиривимірну систему синапсів з ваговими коефіцієнтами. Одновимірні квантові провідники занадто тонкі (\varnothing 1...3 нм), щоб заважати один одному в тривимірній товщі маскувального шару. Повторюючи описані вище операції, можна виростити тривимірний пристрій, який апаратно реалізує нейронну мережу з двовимірними матрицями входів та виходів.

Нанотехнологія, безумовно, згодом вплине на всі інші технології і, можливо, витіснить їх. Однак субмолекулярна квантова елементна база тільки починає створюватись і ще має безліч недоліків, які потрібно буде подолати, перш ніж вона набуде практичного значення. Суттєвим недоліком, наприклад, поки що є швидка деградація параметрів створюваних елементів (від кількох годин до кількох тижнів).

3.6 Контрольні запитання

1. Охарактеризуйте проект Human Brain Project.
2. Охарактеризуйте проект SpiNNaker.
3. Охарактеризуйте проект SyNAPSE.
4. Охарактеризуйте проект Neurogrid.
5. Охарактеризуйте проект BrainScaleS.
6. Назвіть недоліки відомих проектів з апаратної реалізації спайкінгових нейронних мереж та шляхи їх подолання.
7. Опишіть архітектуру оптоелектронної спайкінгової нейронної мережі та принципи її функціонування.
8. Оцініть технологічні параметри оптоелектронної спайкінгової нейронної мережі.
9. Опишіть перспективи подальших досліджень з апаратної реалізації спайкінгових нейронних мереж

ЛІТЕРАТУРА

1. Миркес Е. М. Нейрокомпьютер. Проект стандарта / Миркес Е. М. – Новосибирск : Наука, Сибирская издательская фирма РАН, 1998. – 188 с.
2. Концепция и архитектура программного нейрокомпьютера SNC : материал технической информации / В. И. Гриценко, И. С. Мисуну, Д. А. Рачковский и др. // Управляющие системы и машины. – 2004. – № 3. – С. 3–14.
3. Круг П. Г. Нейронные сети и нейрокомпьютеры : учебное пособие / Круг П. Г. - М. : Издательство МЭИ, 2002. – 176 с.
4. Галушкин А. И. Нейрокомпьютеры. Кн. 3: учебное пособие для вузов / Галушкин А. И. - М. : ИПРЖР, 2000. - 528 с.
5. 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.
6. Von Neumann, J. The General and Logical Theory of Automata / J. Von Neumann // in Cerebral Mechanisms in Behavior (ed. L. A. Jeffress). – New York : Wiley, 1951. – pp. 1–41.
7. Мотоока Т. Калькуляторы на СБЭС: в 2-х кн. / Мотоока Т., Томита С., Танака Х. ; пер. с япон. – М. : Мир, 1988. – 392 с.
8. Pulsed Neural Networks / W. Maass and C. M. Bishop, editors. - MIT Press : Cambridge. - 2001. - 377 p.
9. Архитектура ЭВМ. Курс дистанционного обучения [Электронный ресурс]. - Режим доступа: <http://www.lessons-tva.info/edu/e-inf1/e-inf1-2-2.html>
10. Комарцова Л. Г. Нейрокомпьютеры : учебное пособие для вузов / Л. Г. Комарцова, А. В. Максимов. – [2-е изд., перераб. и доп.]. – М. : Изд-во МГТУ им. Н. Э. Баумана. – 2004. – 400 с.
11. Natschläger T. The «liquid computer»: A novel strategy for real-time computing on time series / T. Natschläger, W. Maass, H. Markram // Special Issue on Foundations of Information Processing of TELEMATIK, v.8, No1: pp.39-43, 2002.
12. Колесницький О. К. Метод розпізнавання багатовимірних часових рядів при допомозі імпульсних нейронних мереж / О. К. Колесницький, Самра Муавія Хамо // Інформаційні технології та комп'ютерна інженерія. – 2006. - № 2(6). - С. 86–93.

13. 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. – № 2. – P.154–165.

14. 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.

15. Колесницький О. К. Математична модель імпульсного нейроелемента на біспін-приладі / О. К. Колесницький, І. В. Бокоцей // *Вісник Хмельницького національного університету*. – 2011. – № 5. – С. 141–149.

16. Бардаченко В. Ф. Таймерні нейронні елементи та структури : монографія / В. Ф. Бардаченко, О. К. Колесницький, С. А. Василецький. – Вінниця : УНІВЕРСУМ-Вінниця, 2005. – 126 с.

17. Большая медицинская энциклопедия / [Гл. ред. Петровский Б. В.]. – М. : Советская энциклопедия, 1981. – Т. 16. – 512 с.

18. Антомонов Ю. Г. Принципы нейродинамики / Ю. Г. Антомонов, Котова А. Б. – К. : Наукова думка, 1974. – 194 с.

19. Антомонов Ю. Г. Введение в структурно-функциональную теорию нервной клетки / Ю. Г. Антомонов, А. Б. Котова. – К. : Наукова думка, 1976. – 263 с.

20. Кузьмин И. В. Основы теории информации и кодирования / И. В. Кузьмин, В. А. Кедрус. – К. : Вища школа, 1986. – 238 с.

21. Spiking Neuron Models [Електронний ресурс] / Wulfiram Gerstner and Werner M. Kistler // *Single Neurons, Populations, Plasticity* (Cambridge University Press, 2002). – Режим доступу : <http://diwww.epfl.ch/~gerstner/BUCH.html>

22. Кнаб О. Д. Биспин — новый тип полупроводниковых приборов / О. Д. Кнаб // *Электронная промышленность*. – 1989. – № 8. – С. 3–8.

23. Василецький С. А. Оптоелектронні частотно-динамічні нейронні елементи: дис. канд. техн. наук : 05.13.05 / Василецький Сергій Анатолійович. — Вінниця, 2003. – 146 с.

24. Василецький С. А. Математичне та комп'ютерне моделювання оптоелектронного частотно-динамічного нейронного елемента / С. А. Василецький, О. К. Колесницький // *Оптоелектронні інформаційно-енергетичні технології*. – 2001. – № 2. – С. 96–101.

25. What makes a dynamical system computationally powerful? / [R. Legenstein, W. Maass, S. Haykin and oth.] // *New Directions in Statistical Signal Processing: From Systems to Brain*. – Cambridge, MIT Press, 2007. – P. 127–154.

26. Maass W. Real-time computing without stable states: A new framework for neural computation based on perturbations / W. Maass, T. Natschläger, H. Markram // *Neural Computation*. – 2002. – Vol. 14(11). – P. 2531–2560.

27. Natschläger T. Computer models and analysis tools for neural microcircuits / W. Maass, T. Natschläger, H. Markram ; in edit. R. Kötter // *Neuroscience Databases. A Practical Guide*. Chapter 9. – Boston : Kluwer Academic Publishers, 2003. – P. 123–138.

28. Jaeger H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication / H. Jaeger, H. Haas // *Science*. – 2004. – Vol. 5. – P. 78–80.

29. Maass W. Fading memory and kernel properties of generic cortical microcircuit models / W. Maass, T. Natschläger, and H. Markram // *Journal of Physiology*. – 2004. Vol. 98(4–6). – P. 315–330.

30. Maass W. Temporal integration in recurrent microcircuits / W. Maass and H. Markram ; in edit. M. A. Arbib // *The Handbook of Brain Theory and Neural Networks*. – Cambridge : MIT Press, 2003. – P. [2] 1159–1163.

31. Волоконная оптика и приборостроение / [Бутусов М. М., Галкин С. Л., Оробинский С. П., Пал Б. П.]. – Л. : Машиностроение, 1987. – 327 с.

32. Уоссермен Ф. Нейрокомпьютерная техника / Уоссермен Ф. – М. : Мир, 1992. – 240 с.

33. Пространственные модуляторы света / [А. А. Васильев, Д. Касасент, И. Н. Компанец, А. В. Парфенов]. – М. : Радио и связь, 1987. – 356 с.

34. Колесницький О. К. Методи і засоби розпізнавання сигналів мультисенсорів газів на основі імпульсних нейронних мереж : монографія / Колесницький О. К. – Вінниця : ВНТУ, 2011. – 120 с.

35. Саймон Хайкин. Нейронные сети: полный курс / Саймон Хайкин ; [2-е изд.; пер. с англ.] – М. : Издательский дом «Вильямс», 2006. – 1104 с.

36. Ротштейн А. П. Проектирование бездефектных человеко-машинных технологий / А. П. Ротштейн, П. Д. Кузнецов. – К. : Техніка, 1992. – 180 с.

37. Hertz J. Introduction to the Theory of Neural Computation / J. Hertz, A. Krogh, and R. G. Palmer. – Boston : Addison-Wesley Reading, Mass., 1991. – 432 p.

38. Сергиенко А. Б. Цифровая обработка сигналов : учебник для вузов / Сергиенко А. Б. – С-Пб. : Питер, 2006. – 307 с.

39. Новиков Л. В. Основы вейвлет-анализа сигналов : уч. пособие / Новиков Л. В. – С-Пб. : ООО «Модус+», 1999. – 152 с.

40. THE BLUE BRAIN PROJECT EPFL [Электронный ресурс] / Henry Markram. – Режим доступа: <http://bluebrain.epfl.ch/>

41. SpiNNaker - A Universal Spiking Neural Network Architecture. SpiNNaker Overview [Электронный ресурс] / Steve Furber, Andrew Brown. – Режим доступа: <http://apt.cs.man.ac.uk/projects/SpiNNaker>.

42. DARPA SyNAPSE Program [Электронный ресурс] / Дхармендра Моха. – Режим доступа: <http://www.artificialbrains.com/darpa-synapse-program#256processor>

43. NeuroGrid Project. Grid technology for neuroscience [Электронный ресурс] / Andrew Simpson. – Режим доступа: <http://www.neurogrid.ac.uk/>

44. BrainScaleS – Neuromorphic processors [Электронный ресурс] / Karlheinz Meier. – Режим доступа: <http://www.artificialbrains.com/brainscales/>

45. Princeton Optronics' Advanced High-Power Diode Lasers [Электронный ресурс] – Режим доступа: <http://www.princetonoptronics.com/pdfs/HighPowerVCSELarrays.pdf>

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

47. Neurocomputer architecture based on spiking neural network and its optoelectronic implementation / Oleh K. Kolesnytskyj; Vladislav V. Kutsman; Krzysztof Skorupski; Mukaddas Arshidinova, Proc. SPIE 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019, 1117609 (6 November 2019); doi: 10.1117/12.2536607

Навчальне видання

Олег Костянтинович Колесницький
Володимир Іванович Месюра

**НЕЙРОМЕРЕЖЕВІ МОДЕЛІ ТА ТЕХНОЛОГІЇ
ОБЧИСЛЮВАЛЬНОГО ІНТЕЛЕКТУ.
НЕЙРОКОМП'ЮТЕРИ**

Частина I

Навчальний посібник

Рукопис оформлено *О. Колесницьким*

Редактор *Т. Старічек*

Оригінал-макет підготувала *Т. Криклива*

Підписано до друку 13.09.2021.
Формат 29,7×42 ¼. Папір офсетний.
Гарнітура Times New Roman.
Друк різнографічний. Ум. друк. арк. 3,96.
Наклад 50 (1-й запуск 1-21) пр. Зам. № 2021-093.

Видавешь та виготовлювач
Вінницький національний технічний університет,
інформаційний редакційно-видавничий центр.

ВНТУ, ГНК, к. 114.

Хмельницьке шосе, 95,

м. Вінниця, 21021.

Тел. (0432) 65-18-06.

press.vntu.edu.ua;

Email: irvc.vntu@gmail.com

Свідоцтво суб'єкта видавничої справи
серія ДК № 3516 від 01.07.2009 р.