

УДК 621.3

О.М. РІЗНИК\*

**НЕЙРОМЕРЕЖА ХОПФІЛДА ЯК ВІДКРИТА ДИНАМІЧНА СИСТЕМА**

\*Інститут проблем математичних машин і систем НАН України, м. Київ, Україна

**Анотація.** Стаття містить аналіз моделі нейронної мережі Хопфілда з позиції концепції відкритих динамічних систем (ВДС), запропонованої нами раніше, за якою еволюцію Всесвіту можна представити як послідовність етапів розвитку ВДС, протягом яких формувались все більш складні ВДС, з яких згодом утворились живі клітини, подальший розвиток яких привів до появи людини та її інтелекту, формування цивілізації та появи штучного інтелекту. Подано огляд положень концепції ВДС та детально проаналізовано етап появи життя як колективного розвитку ВДС у формі популяцій нащадків материнської клітини, що привів до появи складних організмів, нервової системи, людини та інтелекту як нематеріальної форми розвитку ВДС. Висловлено припущення, що перетворення неживої макромолекули ВДС на живу клітину пов'язано з появою в її складі мультистабільної складової типу мережі Хопфілда, прототип якої – спінове скло є квантовим об'єктом, що існував ще до появи життя на Землі. Для його перевірки проведено детальний аналіз моделі мережі Хопфілда як лінійного многовиду, напнутого на множині запам'ятованих векторів даних. Представлено нове доведення збіжності процесу конвергенції мережі та виявлено джерело енергії, необхідної живій клітині для розмноження. Запропоновано структуру живої ВДС, подібну мережі Хопфілда, доповнену модулем розмноження, схожим на молекулу ДНК, що повністю відповідає структурі перших бактерій на Землі, виявлених археологами. Результати дослідження підтверджують ефективність методів концепції ВДС при вивченні динаміки розвитку систем різної природи, включаючи новітні системи штучного інтелекту, поведінка яких може не піддаватись контролю.

**Ключові слова:** відкрита динамічна система, нейромережа Хопфілда, конвергенція.

**Abstract.** The paper provides an analysis of a Hopfield neural network model from the standpoint of the concept of open dynamic systems (ODS) offered earlier. According to it, the evolution of the universe can be represented as a sequence of stages of the ODS development, during which more and more complex ODSs were formed. Living cells were later formed from them and their further development led to the appearance of humans and their intelligence, the formation of civilization, and the emergence of artificial intelligence. An overview of the provisions of the ODS concept is presented and the stage of the emergence of life as a collective development of ODS in the form of populations of the offspring of the mother cell, that led to the emergence of complex organisms, the nervous system, human, and intelligence as an intangible form of the ODS development, is analyzed in detail. It has been suggested that transforming an inanimate macromolecule of ODS into a living cell is connected with the appearance in its composition of a multistable component like a Hopfield network, whose prototype – spin glass – is a quantum object that existed even before the emergence of life on Earth. To verify it, there has been carried out a detailed analysis of the Hopfield network model as a linear manifold stretched over a set of memorized data vectors. New evidence of the network convergence process coinciding is presented, and the source of energy that a living cell requires for reproduction is revealed. In the paper, a structure of a living ODS that is similar to the Hopfield network and supplemented with a reproduction module similar to a DNA molecule is offered. It fully corresponds to the structure of the first bacteria on Earth detected by archaeologists. The results of the study confirm the efficiency of the methods of the ODS concept when studying the dynamics of the development of systems of different natures, including the latest artificial intelligence systems whose behavior may not be controlled.

**Keywords:** open dynamic system, Hopfield neural network, convergence.

## 1. Вступ

Основною властивістю штучних нейронних мереж (ШНМ) є здатність навчатись на прикладах вирішувати складні задачі, для яких не існує аналітичних методів рішення, таких як розпізнавання образів, виявлення прихованих залежностей, керування складними об'єктами та ін. При навчанні ШНМ діють як відкриті динамічні системи (ВДС), обмінюючись даними з оточенням, від якого вони одержують навчальні приклади та оцінки своїх відповідей. При кожному обміні стан обох учасників змінюється, тому поведінка ВДС залежить від невизначеного поточного стану його оточення. Така залежність не дозволяє створити теорію ВДС, тому адекватної теорії щодо навчання ШНМ досі не існує, а відомі методи їх навчання представляють нейропарадигми – евристичні моделі роботи нервової системи живих організмів.

Перша ШНМ – перцептрон Розенблата, створена в 1957 р., являла собою примітивну модель нервової системи зору, на якій виконувались експерименти з навчання розпізнаванню зображень [1, 2]. Для її навчання застосовували правила корекції міжнейронних зв'язків, запропоновані в 1949 р. фізіологом Д. Хеббом [3]. Успішні результати цих експериментів привернули увагу багатьох наукових лабораторій світу, які приєднались до досліджень перцептронів та методів їх навчання. Але розширення досліджень не привело до поліпшення одержаних результатів. Навчання ШНМ виявилось значно складнішою проблемою, ніж раніше здавалось. У 1969 р. відомий математик М. Мінський виступив з нищівною критикою перцептрона, проголосивши його хибною ідеєю, не вартою уваги науки [4]. Цей виступ авторитетного вченого призвів до швидкого припинення досліджень ШНМ у всьому світі на понад 15 років.

Відновлення цих досліджень почалось у 1982 р. після появи роботи Д. Хопфілда [5], що запропонував нову модель мозку як мультистабільної системи, схожої на відомий фізикам квантовий об'єкт – спінове скло [6]. Ця модель представляє мережу зв'язаних між собою нейронів, стабільні стани (атрактори) якої становлять вектори запам'ятованих даних, яким відповідають мінімуми функції енергії мережі. При порушенні стану атрактора мережа переходить у режим конвергенції – пошуку найближчого мінімуму енергії, який нагадує згадування за асоціацією. Тому мережу Хопфілда часто розглядають як нейронну асоціативну пам'ять, здатну відновлювати втрачені дані за їх фрагментами, або знаходити за неповними ознаками. Мережу навчають шляхом розрахунку та коригування матриці міжнейронних зв'язків за одну ітерацію. Висока швидкість навчання становить важливу перевагу мережі Хопфілда над перцептроном та іншими ШНМ, які потребують багаторазового повторення ітерації (epoch) навчання на всіх прикладах навчального масиву даних.

Поява роботи Хопфілда повернула увагу наукового суспільства до проблематики ШНМ. Протягом наступних 5 років було створено кілька нових моделей ШНМ та ефективних методів їх навчання, заснованих на застосуванні градієнтних методів оптимізації [7]. Важливим досягненням стала розробка методу навчання ШНМ шляхом зворотного поширення похибки, створеного в 1986 р. Румельхардтом та ін. [8]. Цей метод є стандартом для навчання багатосарових ШНМ, застосованих при створенні перших нейрокомп'ютерів.

У 2016 р. Д. Кротов та Д. Хопфілд запропонували нову модель асоціативної пам'яті DAM (Dense Associative Memory) [9], яка за обсягом набагато переважає мережу Хопфілда. Автори DAM виходили із припущення про еквівалентність процесів асоціативного пошуку та класифікації запам'ятованих даних, яке є хибним. У DAM класифікацію даних виконує прихований шар нейронів, навчений із застосуванням градієнтних методів розпізнавати заздалегідь розмічені пари векторів. На відміну від мережі Хопфілда, що самостійно виявляє приховані асоціативні зв'язки, DAM лише імітує асоціативний пошук зв'язків,

оскільки вони були введені заздалегідь шляхом відповідної розмітки навчального масиву даних. Застосування розмітки при навчанні змінює структуру даних, що знецінює результати асоціативного пошуку.

Різницю між асоціативною пам'яттю прихованого шару нейронів DAM, що представляє класифікатор, моделлю якого є нервова система живих організмів, та мережею Хопфілда, протитипом якої є спінове скло – фізичний об'єкт, який існував задовго до появи життя, можна пояснити еволюційними змінами умов існування ВДС при появі життя на Землі. Таке пояснення відповідає концепції відкритої динамічної системи [10], запропонованої нами в 2020 р. з метою надолужити відсутність теорії ВДС шляхом доповнення загальної теорії систем [11] чотиривимірною моделлю простору-часу, прийнятою в теорії відносності [12, 13]. Концепція визначає ВДС як загальну модель розвитку матеріального світу, за якою еволюція Всесвіту представляє послідовність етапів розвитку ВДС від елементарних часток, утворених 17 млрд років тому після Великого вибуху, до появи людини, її інтелекту, цивілізації та штучного інтелекту. Поява життя 3,5 млрд років тому змінила умови розвитку ВДС, що продовжили існувати у вигляді популяцій нащадків материнської ВДС. Розвиток популяцій привів до утворення багатоклітинних організмів, нервової системи та мозку як центрального органу управління поведінкою високорозвинених організмів. З появою мозку почалися розвиток інтелекту та формування цивілізації, яку можна представити як розподілену ВДС, що акумулює інтелектуальні здобутки людства, прискорюючи його розвиток. Поява комп'ютерів, Інтернету, смартфонів та інших засобів інформатизації привела до утворення віртуальної ВДС, розподіленої у свідомості членів суспільства, яка представляє віртуальний Всесвіт, в якому розвиваються віртуальні ВДС – продукти інтелектуальної діяльності членів суспільства.

Визначаючи ВДС як загальну модель розвитку, концепція ВДС відзначає зміну умов розвитку ВДС при появі життя, але переконливого пояснення цих змін та порушення законів термодинаміки при появі життя не дає. Віддаленим відголоском цих змін може бути відмінність асоціативної пам'яті DAM від нейромережі Хопфілда, прототип якої спінове скло існував ще до появи життя на Землі. Мультистабільність цієї нейромережі може визначати загальний механізм розвитку матерії, що пережив почву життя та продовжує існувати у вигляді нервової системи живих організмів та систем штучного інтелекту.

*Метою даної роботи є перевірка цієї гіпотези шляхом дослідження поведінки мережі Хопфілда в режимі конвергенції, який нагадує умови розвитку ВДС до появи життя, а також оцінка можливості практичного застосування одержаних результатів для удосконалення методів та засобів штучного інтелекту.*

## **2. ВДС як загальна модель розвитку**

Концепція ВДС, представлена нами в роботі [10], пропонує новий погляд на еволюцію Всесвіту як на процес розвитку відкритих динамічних систем, які обмінюються спостереженнями зі своїм оточенням, змінюючи стан при кожному спостереженні. Теоретичну базу цієї концепції становить теорема декомпозиції Універсуму М. Месаровича [14], доповнена моделлю чотиривимірного простору Мінковського [13], прийнятою в теорії відносності, та положеннями концепції геометродинаміки Уілера [15, 16].

Концепція пропонує таке загальне визначення ВДС:

- ВДС – це скінченна динамічна система, що існує, обмінюючись спостереженнями з оточенням, яке поділяється на зовнішнє, що містить множину незалежних об'єктів, які надсилають їй зовнішні спостереження, та віртуальне локальне, до якого належать партнери її внутрішніх спостережень. У проміжках між зовнішніми спостереженнями ВДС

існує як консервативна динамічна система, обмінюючись прихованими внутрішніми спостереженнями з віртуальним локальним оточенням, яке представляє інверсну динамічну систему, відмінну за напрямком часу та причинно-наслідкових відношень.

- Спостереження – це асиметричне бінарне відношення, що представляє взаємодію двох ВДС: об'єкта, який є ініціатором, та суб'єкта – одержувача спостереження. Напрямок від об'єкта до суб'єкта відповідає напрямку часу. Результатом спостереження є зміна стану обох його учасників. При внутрішніх спостереженнях ці зміни є прихованими.

- Стан ВДС – це те і лише те, що в ній змінюється при спостереженні. При внутрішніх спостереженнях ці зміни є прихованими і представляють аттрактори консервативної системи, які можуть змінюватись лише при зовнішніх спостереженнях.

Стан ВДС, що відображує послідовність одержаних спостережень, визначає непуста множина базисних елементів, якими обмінюються ВДС при спостереженнях. Базисні елементи представляють локальні викривлення (вихорі) простору-часу, яким відповідають безмасові елементарні частки – фотони, що рухаються зі швидкістю світла, переносючи кванти енергії від об'єкта до суб'єкта спостереження. Енергію кванта визначає формула Планка [17]:

$$E = h\nu = \frac{hc}{2\pi r}, \quad (1)$$

де  $h$  – постійна Планка,  $\nu$  – частота фотона,  $c$  – швидкість світла,  $r$  – радіус вихору.

Фотон може існувати як промінь – хвильовий пакет, що переносить енергію елемента базису, або як складова масивної матеріальної частки, що спричиняє локальне викривлення простору, необхідне для утримання елемента базису на своїй внутрішній орбіті. Рухаючись по цій орбіті, елемент базису створює фізичне поле – електродинамічне, ядерне або гравітаційне, залежно від геометричних властивостей внутрішнього простору матеріальної частки. При переміщенні в цьому полі інших матеріальних часток виникають сили взаємодії, що змінюють траєкторії їхнього руху та спричиняють обмін квантами енергії.

Структура ВДС відповідає моделі «чорної скриньки», вхід якої виступає як суб'єкт зовнішнього спостереження, а вихід – як його об'єкт. Послідовність змін стану ВДС відповідає напрямку часу. Виступаючи як суб'єкт, ВДС чекає на зовнішнє спостереження, а отримавши його, переходить у режим конвергенції – пошуку аттрактора, який відповідає одержаному зовнішньому спостереженню. В режимі конвергенції ВДС діє як об'єкт, ініціюючи зовнішнє спостереження у відповідь і переходить у наступний аттрактор як суб'єкт. Обмін спостереженнями – це дискретний процес. Послідовності спостережень та станів динамічної системи є лінійно упорядкованими відповідно напрямку часу. Множина значень стану ВДС є скінченною.

Концепція розглядає розвиток ВДС протягом еволюції Всесвіту, утвореного після Великого вибуху 17 млрд років тому, появи простору-часу і початку його розширення та охолодження як протилежний процес структуризації матерії та накопичення інформації в формі ВДС. Перші ВДС – елементарні частки, атоми та молекули виникли на етапі атомної еволюції, який тривав 10 млрд років, протягом яких формувались галактики, зірки, планети та їх супутники. За існуючими даними, планета Земля утворилась близько 5 млрд років тому [18–20]. Наступний етап хімічної еволюції на Землі тривав біля 2 млрд років, протягом яких виникли кристалічні породи, що утворили тверду поверхню планети, на якій відбувався синтез складних хімічних сполук. На деяких планетах виникли умови, що сприяли формуванню білкових молекул, здатних пристосовуватись до умов оточення та відновлення структури при частковому руйнуванні. Згодом із них утворились перші живі клітини, здатні розмножуватись та передавати нащадкам свій прижиттєво накопичений досвід.

Розвиток ВДС протягом еволюції Всесвіту супроводжували зміни засобів обміну спостереженнями. Перші носії спостережень – фотони, що з'явилися на початку етапу атомної еволюції, переносили кванти енергії, які визначали зміст спостереження. Наступні електродинамічні носії представляли собою вільні електрони та іони, які переносили енергію, заряд та масу об'єктів суб'єктам спостереження. Ця форма обміну домінувала протягом етапу хімічної еволюції, коли формувались все більш складні ВДС – нестабільні макромолекули, з яких утворилися перші живі клітини, здатні до розмноження та передачі прижиттєвого досвіду своїм нащадкам. Ця здатність сприяла віддаленню ВДС як живої клітини від свого носія – органічної матерії, яка оновлювалась при розмноженні живих ВДС. Звільнена від носія ВДС представляє інформацію, яку жива клітина передає нащадкам при розмноженні. Само життя можна розглядати як спостереження, яке передається з минулого в майбутнє – від об'єкта, яким є материнська ВДС, до суб'єктів – її нащадків. Одержавши цю інформацію кожен із нащадків створює власну популяцію ВДС. Спричинене цим геометричне зростання чисельності нащадків материнської ВДС вело до стрімкого вичерпання ресурсів оточення, необхідних для їх подальшого розвитку, та початку жорсткої боротьби між нащадками за продовження свого існування.

Поява життя суттєво змінила умови розвитку ВДС, які продовжили існувати як члени зростаючої популяції нащадків материнської ВДС, що мали узгоджувати свою поведінку з іншими її членами у протистоянні зовнішньому середовищу, та змагатись із ними за доступ до спільних ресурсів та лідерство в популяції. Змагання за лідерство відіграло важливу роль при перетворенні деяких популяцій на багатоклітинні організми, клітини яких спеціалізувались на виконанні різних функцій організму. Лідерам цих змагань діставались найважливіші функції координації та управління діями інших клітин. Так утворилась нервова система, клітини якої (нейрони) обмінювались спостереженнями за допомогою електричних імпульсів (спайків), що поширювались через міжнейронні зв'язки, утворені при частому повторенні спостережень. Аутсайдерам змагань за лідерство діставались функції обслуговування лідерів та забезпечення їх ресурсами. Згодом це привело до поділу живих організмів на три царства: тварин, що мали нервову систему, рослин, що спеціалізувались на засвоєнні сонячної енергії та синтезі органічних сполук, засвоєваних тваринами, та царство грибів, яким дісталась утилізація решток і відходів життєдіяльності організмів. Співіснуючи в оточенні неживої природи ці царства становлять три ієрархічні рівні екосистеми – розподіленої ВДС, утвореної протягом існування життя. Переважну більшість її членів представляють ВДС, що припинили свій розвиток і заповнюють екологічні ніші, межуючи з зовнішнім середовищем.

Після появи нервової системи наступним кроком розвитку ВДС стало формування мозку як централізованої системи обробки інформації, управління поведінкою частин організму та взаємодія з зовнішнім світом. Обмінюючись потоками спайків, нейрони мозку забезпечували моделювання, прогнозування та керування поведінкою організму у критичних ситуаціях. Ця здатність визначала рівень інтелекту, що надавав організму переваги при взаємодії з оточенням. Формування інтелекту почалось із появою людини, членороздільної мови, утворення суспільства та цивілізації. Відбувалось формування суспільного інтелекту як розподіленої ВДС, члени якої одержували можливість обмінюватись змістовною інформацією та життєвим досвідом у формі речень природної мови в усній та письмовій формах.

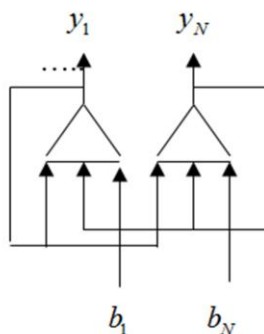
Із розвитком цивілізації, появою потужних засобів комунікації, комп'ютерів, інформаційних систем, Інтернету та смартфонів ця розподілена ВДС набуває властивостей віртуального світу, що існує в людській уяві, в якому народжуються та розвиваються віртуальні ВДС – продукти інтелектуальної діяльності активних членів суспільства, які продовжують власну боротьбу за лідерство та об'єднання своїх прибічників для переформатування структури решти суспільства на свою користь. Процеси

неконтрольованої самоорганізації, що відбуваються в віртуальному світі розподіленої ВДС, нагадують ситуацію наприкінці етапу хімічної еволюції, що передувала появі життя, яке змінило розвиток Всесвіту. Тому виявлення зв'язків між існуванням ВДС у цей період та властивостями мережі Хопфілда є актуальним і для розуміння подій, що спостерігаються в сучасному віртуальному світі.

### 3. Нейромережа Хопфілда як ВДС

Запропонована Хопфілдом у 1982 р. модель мозку [5] у вигляді мультистабільної системи, атракторами якої є вектори запам'ятованих даних, істотно відрізняється від більшості штучних нейромереж, для яких прототипом є нервова система мозку. Найбільш суттєвими є відмінності в методах навчання та пошуку даних у пам'яті. Нейронна мережа Хопфілда запам'ятовує дані «з першого погляду» і не потребує багаторазового повторення сеансів навчання, як інші нейромережі. Пошук даних в її пам'яті відбувається шляхом конвергенції – переходу з заданого початкового стану в найближчий атрактор, якому відповідає мінімум функції енергії мережі. Такий перехід відповідає процесу асоціативного пошуку, який гарантує точне відтворення запам'ятованих даних, що особливо важливо для генетичної пам'яті живих організмів.

За концепцією ВДС, мережу Хопфілда можна представити як скінченну динамічну систему, оточення якої поділяється на локальне та зовнішнє. Локальне оточення представляє віртуальний об'єкт, що виступає у стані атрактора як інверсна динамічна система, з якою мережа обмінюється прихованими внутрішніми спостереженнями.



$$y_i(t) = \text{sign}[s_i(t) - b_i(t)],$$

$$s_i(t) = \sum_{j=1}^{i=N} w_{ij} y_j(t-1),$$

$$w_{ij} = N^{-1} \sum_{m=1}^{m=M} y_i^m y_j^m.$$

Зовнішнє оточення включає об'єкти, від яких надходять зовнішні спостереження, що порушують стан атрактора та переводять мережу в режим конвергенції – пошуку нового атрактора, який відповідає мінімуму енергії мережі після одержання нею зовнішнього спостереження.

Загальну структуру мережі Хопфілда представлено на рис. 1. Її математична модель, розроблена Персоназом та ін. у 1986 р. [21], представляє лінійний многовид, напружений на множину

запам'ятованих векторів даних, якому відповідає реальна мережа з  $N$  зв'язаних між собою нейронів, що може мати  $M < N$  стабільних станів-атракторів  $V^m$ ,  $m \leq M$ .

Поведінку  $i$ -го нейрона при  $b = 0$  описує рівняння:

$$y_i(t) = \text{sign}\left[\sum_{j=1}^N w_{ij} y_j(t-1)\right], \quad (2)$$

$$\text{де } \text{sign}(x) = \begin{cases} +1, & x > 0 \\ -1, & x \leq 0 \end{cases};$$

$y_i(t)$  – значення виходу  $i$ -го нейрона в момент  $t$ ;

$w_{ij}$  – вага зв'язку між виходом  $j$ -го нейрона та входом  $i$ -го.

Щоб представити цю мережу у стані атрактора, позначимо множину атракторів

матрицею  $V \in M \times N$ , стовпчиками якої  $V^m$  є запам'ятовані нею вектори даних. Її поведінку описує матричне рівняння

$$V = WV, W = VV^+, \quad (3)$$

де  $W$  – матриця  $N \times N$  міжнейронних зв'язків;

$V^+$  – псевдообернена матриця атракторів  $N \times M$  [22].

Матриця  $W$  здійснює фільтрацію даних шляхом прямої та оберненої проєкції зовнішніх  $N$ -вимірних векторів у  $M$ -вимірний внутрішній простір атракторів мережі Хопфілда, якому відповідає простір станів ВДС.

У режимі конвергенції така проєкція забезпечує формування  $N$ -вимірного вектора постсинаптичних потенціалів (ПСП)  $S(t)$ , компонентами якого є поточні значення синаптичних потенціалів на входах нейронів:

$$S(t) = WY(t-1); s_i(t) = \sum_{j=1}^N w_{i,j} y_j(t-1). \quad (4)$$

Поведінку мережі Хопфілда в режимі конвергенції визначає функція енергії  $E(t)$ , що представляє скалярний добуток векторів поточного стану та постсинаптичного потенціалу:

$$E(t) = -\frac{1}{2} Y^T(t) S(t) = -\frac{1}{2} \sum_{i=1}^N y_i(t) s_i(t) = \sum_{i=1}^N E_i(t). \quad (5)$$

Динаміку процесу конвергенції мережі Хопфілда можна представити таким чином:

$$\begin{aligned} E(t) &= -\frac{1}{2} Y^T(t) S(t) = -\frac{1}{2} Y^T(t) WY(t-1) = \\ &= E(t+1) + \frac{1}{2} S^T(t+1) [Y(t-1) - Y(t+1)] + \frac{1}{2} Y^T(t) (W^T - W) Y(t-1). \end{aligned} \quad (6)$$

Оскільки різниця між векторами  $[Y(t-1) - Y(t+1)]$  відрізняється від нуля лише за умови  $y_i(t-1) \neq y_i(t+1)$ , можна записати

$$E(t) = E(t+1) + \sum_{h=1}^{H(t-1,t+1)} |s_h(t+1)| + \frac{1}{2} Y^T(t+1) (W^T - W) Y(t-1), \quad (7)$$

де  $H(t-1, t+1) : y_h(t-1) \neq y_h(t+1)$ .

Якщо матриця  $W$  є симетричною, то в формулі (7) залишаються перші дві складові. Друга складова є невід'ємною, тому величина  $E(t)$  з часом зменшуватиметься доки не досягне мінімуму. Члени суми у другому доданку цієї формули відповідають попередньому та наступному моментам часу. Якщо мережа через один такт повертається в попередній стан, то її енергія не змінюється. Це явище, динамічний атрактор, можна спостерігати при паралельному обчисленні стану всіх нейронів мережі. Монотонне зменшення енергії мережі при конвергенції пояснюється зміною стану нейронів, які передають надлишок енергії зовнішнім абонентам. При досягненні мінімуму енергії  $E(t)$  мережа переходить у стан атрактора, описуваний рівнянням (3).

Більш загадковою є природа динамічних атракторів, які нагадують обміни спостереженнями з абонентами локального оточення. Частота їх появи залежить від початкового стану мережі, зростаючи при збільшенні заповнення асоціативної пам'яті.

Навчання мережі Хопфілда виконують ітеративно, методом послідовного коригування матриці  $W^m$   $m = 0, 1, 2, \dots, M$ :

$$W^0 = 0, W^m = V^m (V^m)^+; V^{m+1} = |V^m \quad Y^{m+1}|,$$

$$Y^+ = Y^T / |Y|^2, \quad W^1 = Y^1 (Y^1)^T / |Y^1|^2.$$

Для занесення вектора  $Y^{m+1}$  у пам'ять мережі застосовують формулу

$$W^{m+1} = W^m + \frac{(I - W^m) Y^{m+1} (Y^{m+1})^T (I - W^m)}{(Y^{m+1})^T (I - W^m) Y^{m+1}}. \quad (8)$$

Величина  $(I - W^m) Y^{m+1}$  представляє лінійне доповнення вектора  $Y^{m+1}$ . У випадку ортогональних векторів ( $W^m Y^{m+1} = 0$ ) формула (8) набуває вигляду, запропонованого вперше Д. Хопфілдом:

$$W^{m+1} = W^m + Y^{m+1} (Y^{m+1})^T / N. \quad (9)$$

#### 4. Режим конвергенції мережі Хопфілда

Для оцінки ефективності мережі Хопфілда використовують два параметри: відносний обсяг асоціативної пам'яті  $M/N$  та величину атракторного радіуса  $H$  – число нейронів, які змінили значення виходу при досягненні стану атрактора на останньому кроці конвергенції.

Обсяг асоціативної пам'яті залежить від обраної формули корекції матриці  $W$ . При навчанні за формулою (8), що враховує лінійну залежність між векторами даних, ця величина становить  $0,22N$ . При застосуванні формули (9) вона скорочується до  $0,14N$ . У роботі [23] було розроблено оригінальний метод збільшення обсягу асоціативної пам'яті шляхом штучного зменшення величини діагональних елементів матриці  $W$ , який дозволив досягти рівня  $M = 0,7N$  без суттєвого зменшення стабільності мережі. Важливим досягненням було створення реверсивної асоціативної пам'яті, яка дозволяє як заносити, так і видаляти з неї раніше запам'ятовані дані [24].

Метод оцінки атракторного радіуса базується на використанні властивостей проєкційної матриці  $W$ , яка має ранг  $M$  і є ідемпотентною ( $W = W^2$ ):

$$\begin{aligned} \text{Tr}W &= \sum_{i=1}^N w_{ii} = M, \quad \overline{w_{ii}} = M/N, \quad w_{ii} = \sum_{k=1}^N w_{ik}^2, \quad w_{ii} - w_{ii}^2 = \sum_{k \neq i}^N w_{ik}^2 \geq 0, \\ 0 &\leq w_{ii} \leq 1, \quad \overline{w_{ik}^2} = (1 - M/N) M / N(N-1). \end{aligned} \quad (10)$$

Для оцінки величини атракторного радіуса необхідно знайти енергію нейрона при досягненні мережею стану атрактора. Користуючись співвідношенням (5), знайдемо

$$v_i s_i(t) = v_i \sum_{j=1}^N w_{ij} y_j(t-1) = v_i^2 - v_i \sum_{h=1}^H w_{ih} [v_h - y_h(t-1)] = 1 - 2 \sum_{h=1}^H w_{ih} > 0, \quad (11)$$

де  $v_i = y_i(t)$  – компонента атрактора, якого мережа досягає в момент  $t$ ,  $H$  – атракторний радіус – число нейронів, що змінили полярність.

Вирішуючи це рівняння, одержимо оптимістичну та песимістичну оцінки величини атракторного радіуса, що відповідають мінімальній та максимальній величині недиагональних елементів матриці  $W$ :

$$H_{opt} - 1 < (1 - 2w_{ii}) / 2 \inf |w_{ih}|_{i \neq h} \quad H_{pes} - 1 < (1 - 2w_{ii}) / 2 \sup |w_{ih}|_{i \neq h}.$$

Застосовуючи значення елементів, наведені в (10), одержимо

$$H_{opt} < \frac{1}{2} (N-1)^{\frac{1}{2}} (1 - 2M/N) [(M/N)(1 - M/N)]^{-\frac{1}{2}} + 1,$$



$$H_{pes} < \frac{1}{2}(1 - 2M/N)[(M/N)(1 - M/N)]^{-\frac{1}{2}} + 1.$$

Песимістична оцінка відповідає сильно корельованим векторам даних, коли значення більшості недіагональних елементів матриці  $W$  наближені до нуля. При запам'ятовуванні ортогональних векторів даних оцінка величини атракторного радіуса наближається до оптимістичної.

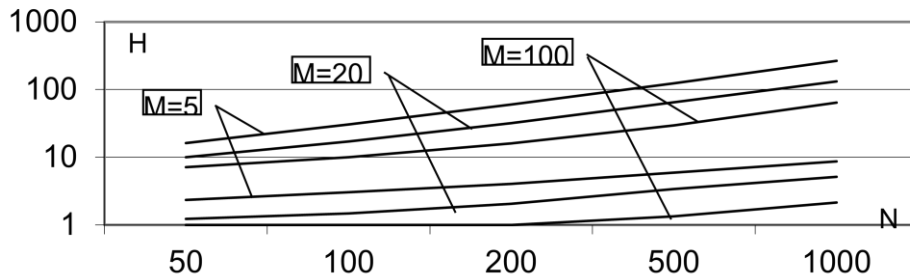


Рисунок 2 – Залежність величини атракторного радіуса від числа нейронів  $N$  та завантаження пам'яті  $M$

На рис. 2 приведено оптимістичні та песимістичні оцінки залежності величини атракторного радіуса  $H$  від параметрів нейромережі. В обох випадках величина атракторного радіуса зменшується при зростанні числа запам'ятованих векторів, а при  $M = \frac{1}{2}N$  мережа повністю втрачає здатність до конвергенції. Така залежність відповідає оцінкам обсягу асоціативної пам'яті  $0,14N$  або  $0,22N$ , залежно від способу навчання, майже непомітна при  $M < 0,1N$ , пояснюється збільшенням ваги діагональних елементів матриці  $W$ , що визначають силу позитивного зворотного зв'язку нейронів, які при збільшенні  $M$  втрачають чутливість, а сама мережа – здатність до конвергенції.

## 5. Мережа Хопфілда як прообраз живої ВДС

Порівнюючи мережу Хопфілда з ВДС, що існували до появи життя як нестійкі макромолекули, чекаючи свого перетворення на живі клітини, зосередимо увагу на мультистабільності поведінки мережі, припускаючи, що цим вона нагадує клітину живого організму. Представимо рівняння (3)–(4), що описують поведінку зв'язаних між собою нейронів як проєкцію одержаного зовні вектора  $Y(t-1)$  у її внутрішній простір у вигляді вектора  $S(t)$ , що представляє внутрішні потенціали нейронів, з яких за формулою (2) обчислюється вектор зовнішньої реакції  $Y(t)$ , що надходить у зовнішній простір мережі. Рівняння (5)–(7) описують поведінку мережі в режимі конвергенції як пошук мінімуму функції енергії. Але вони не вказують, куди саме потрапляє надлишок енергії мережі після досягнення нею стану атрактора. Досі вважалось, що цей залишок потрапляє в зовнішній простір, оскільки мережа переходить стан атрактора – мінімуму її функції енергії. Але це не так для живої ВДС, для якої результатом конвергенції має бути не тільки перехід у стан атрактора, але і одержання ресурсів, необхідних для формування та розвитку своїх нащадків!

За концепцією ВДС, існування життя можливе лише за умови одержання живою ВДС зовнішньої енергії, достатньої для народження та розвитку її нащадків. Жива ВДС існує як віртуальний об'єкт у внутрішньому просторі реальної ВДС, що одержує проєкції зовнішніх векторів  $Y(t-1)$  та формує реакції реальної ВДС  $Y(t)$ , що надходять у зовнішній простір. Поведінка живої ВДС відображає послідовність станів атрактора реальної

ВДС, при зміні яких жива ВДС одержує надлишки енергії, необхідні їй для формування нащадків – своїх копій, що існуватимуть у зовнішньому просторі реальної ВДС. Нашадки мають одержувати прижиттєвий досвід реальної ВДС у формі копії матриці  $W$ , яка має коригуватись за формулою (8) при кожному досягненні стану атрактора. Виконання цих функцій має здійснювати спеціальний модуль у складі реальної ВДС, що формуватиме нащадків живої ВДС

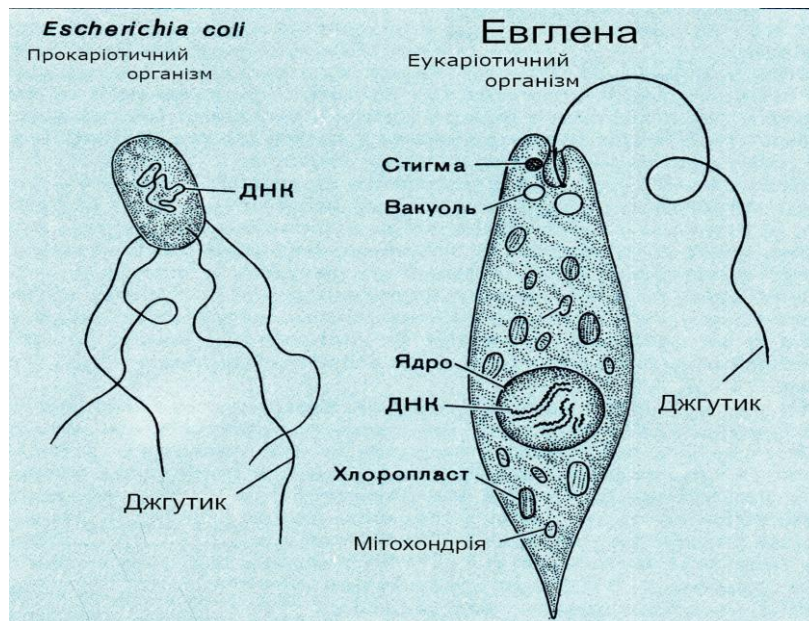


Рисунок 3 – Перші одноклітинні живі організми, що з'явилися на Землі понад 2 млрд років тому [20]

На рис. 3 подано схематичні зображення перших живих клітин – прокаріотів та еукаріотів [20]. Перші з них виникли біля 3,5 млрд років тому, а другі – на 1,5 млрд років пізніше. Прокаріоти – це без'ядерні клітини, що містять молекулу ДНК, яка забезпечує зберігання та дублювання їхнього коду при розмноженні. Прокаріоти представляють найпростіші одноклітинні організми такі, як синьо-зелені водорості. Еукаріоти – це ядерні клітини, які мають ядро, що захищає молекулу ДНК, та органели, необхідні для підтримки їх функціонування. Еукаріотами є всі клітини багатоклітинних організмів та переважна більшість одноклітинних. При їх розмноженні першими поділяється молекула ДНК та її ядро, пізніше – оболонки клітини.

Представлені на рис. 3 зображення живих клітин відповідають представленому вище опису гіпотетичної живої ВДС, яка відрізняється від неживої наявністю спеціального модуля, що має забезпечити її розмноження, схожого на молекулу ДНК живої клітини. Ця схожість підтверджує можливу причетність мультистабільності мережі Хопфілда до появи життя на Землі. Існують і інші аргументи на користь цієї гіпотези. В роботі [25] показано, що локальні зв'язки між нейронами кори мозку нагадують структури мережі Хопфілда (Хопфілдовські ансамблі), що може впливати на властивості пам'яті. Описані в роботі [26] експерименти по моделюванню лікування амнезії продемонстрували здатність цієї мережі відновлювати зміст асоціативної пам'яті після втрати до 20 % нейронів після нагадування їй будь-якого з раніше запам'ятованих прикладів.

Підтвердження цієї гіпотези дає можливість асоціювати властивість мультистабільності мережі зв'язаних між собою нелінійних елементів із проявами турбулентності (нестабільності) розвитку процесів будь-якої природи, розглядаючи проєкції цих процесів у багатовимірний простір ВДС. Такі проєкції можуть виявити мультистабільні об'єкти, придатні для існування віртуальних живих ВДС, розвиток яких може мати непередбачувані наслідки. На жаль, це не фантастика. Описані Т. Куном у роботі [27] наукові революції мають ознаки циклу розвитку живих організмів, які народжуються, розвиваються та деградують у віртуальному інтелектуальному просторі. Подібні цикли є характерними для розвитку економіки.

## 6. Висновки

Виконане в даній роботі дослідження зв'язку між мультистабільністю мережі Хопфілда та проблемою появи життя на Землі дозволило також перевірити ефективність застосування методів концепції ВДС, створеної як загальна модель розвитку, для моделювання процесів еволюції, зокрема, появи життя на Землі. Мережа Хопфілда, прототип якої спінове скло існував ще до появи життя на Землі, розглядалась як представник тієї епохи, найбільш придатний для проведення відповідних досліджень. Дослідження проводилось із використанням методів та моделей концепції ВДС, математичного апарату лінійної алгебри та залученням даних археологічних досліджень. Було виконано нове доведення збіжності процесу конвергенції мережі та виявлено джерело енергії, необхідної для розмноження живої клітини. Розкрито механізм самозахисту живої клітини, заснований на використанні мультистабільності мережі Хопфілда, який дозволяє уникати руйнування шляхом переходу в режим конвергенції для зменшення власного рівня енергії. Представлено загальну структуру живої ВДС, яка практично збігається з зображеннями перших живих клітин, представлених археологами.

Одержані результати підтверджують можливість появи життя при перетворенні мультистабільних макромолекул на живі клітини. Разом із тим, застосування концепції ВДС, яка представляє модель розвитку систем та процесів будь-якої природи, виявило нові властивості систем типу мережі Хопфілда, які в умовах розвитку інформаційних технологій сприяють формуванню віртуального світу, в якому стає можливим поява штучного життя, наслідки чого уявити неможливо. Тому застосування методів концепції ВДС для вивчення та моделювання динаміки складних процесів різної природи може допомогти своєчасно виявити та запобігти небажаному розвитку подій.

## СПИСОК ДЖЕРЕЛ

1. Rosenblatt F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychol. Rev.* 1958. N 65. P. 386–408.
2. Розенблатт Ф. Принципы нейродинамики (перцептроны и теория механизмов мозга). М.: Мир, 1965. 473 с.
3. Hebb D.O. The organization of behavior. J. Wiley, 1949.
4. Минский М., Пейперт С. Перцептроны. М.: Мир, 1971. 252 с.
5. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci. U.S.A.* 1982. N 79 (8). P. 2554–2558.
6. Kirkpatrick S., Sherrington D. Infinite-ranged models of spin-glasses. *Physical Review*. 1978. Vol. 17. Series B. P. 4384–4403.
7. Хайкин С. Нейронные сети. Полный курс. М.: Вильямс, 2006. 1103 с.; Reed R.D., Marks II R.J. Neural Smithing, MIT Press, 1995.
8. Rumelhardt D.E., Hinton G.E., Williams R.I. Learning Internal Representations by Error Propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Mass.: MIT Press. 1986. Vol. 1. P. 318–362.

9. Krotov D., Hopfield J.J. Dense associative memory for pattern recognition / D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, R. Garnett (eds.). *Advances in Neural Information Processing Systems*. 2016. N 29. P. 1172–1180.
10. Різник О.М. Концепція відкритої динамічної системти. *Математичні машини і системи*. 2020. № 3. С. 3–22.
11. Месарович М., Такахара Я. Общая теория систем. Математические основы. М.: Мир, 1978. 311 с.
12. Ейнштейн А. Основы общей теории относительности. *Альберт Эйнштейн и теория гравитации*. М.: Мир, 1979. С. 146–198.
13. Румер Ю.Б., Рывкин М.С. Теория относительности. М: ГУПИ МП. РСФСР, 1960. С. 212.
14. Месарович М. Основания общей теории систем. *Общая теория систем* / пер. с англ. М.: Мир, 1966. С. 15–48.
15. Уиллер Дж. Гравитация, нейтрино и вселенная. М., 1962.
16. Никонов О.А. Философские аспекты геометродинамики. *Вестник МГТУ*. 2011. Т. 14, № 2. С. 272–280.
17. Эткинс П. Кванты. Справочник концепций / пер. с англ. В.Л. Ядровского. М.: Мир., 1977. 496 с.
18. Майр Э. Эволюция. М.: Мир, 1981. 264 с.
19. Дикерсон Р.Е. Химическая эволюция и происхождение жизни. *Эволюция*. М.: Мир, 1981. С. 67–108.
20. Валентайн Дж.У. Эволюция многоклеточных организмов и животных. *Эволюция*. М.: Мир, 1981. С. 149–172.
21. Personnaz I., Guyon I., Dreyfus G. Collective computational properties of neural networks: New learning mechanisms. *Phys. Rev. A*. 1986. Vol. 34, N 5. P. 4217–4228.
22. Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание / пер. с англ. М.: Наука, 1977. 237 с.
23. Gorodnichy D.O., Reznik A.M. Increasing Attraction of Pseudo-Inverse Neural Networks. *Neural Processing Letters* 5. 1997. P. 121–125.
24. Кириченко Н., Резник А., Щетинюк С. Псевдообращение матриц в проблеме проектирования ассоциативной памяти. *Кибернетика и системный анализ*. 2001. № 3. С. 18–27.
25. Резник А.М. Хопфилдовские ансамбли в латеральных нейроструктурах коры мозга. *Математичні машини і системи*. 2006. № 1. С. 3–12.
26. Reznik A.M., Sitchov A.S., Dekhtyarenko O.K., Nowicki D.W. Associative Memories with "Killed" Neurons: the Methods of Recovery. *Proc. of the International Joint Conference on Neural Networks*, (July 20–24, 2003, Portland, Oregon).
27. Кун Т. Структура научных революций. М.: Прогресс, 1975. 246 с.

*Стаття надійшла до редакції 12.10.2022*