

УДК 658:004

Павлішин М.А.¹, Зеленцова І.Я.²

¹ студ. гр. КНТ-514м ЗНТУ

² канд. техн. наук, доц. ЗНТУ

ЗАСТОСУВАННЯ АПАРАТНИХ ПРИСКОРЮВАЧІВ НА FPGA ПРИ РОЗРОБЦІ НЕЙРОННИХ МЕРЕЖ

Актуальність даної теми обумовлена тим, що штучний інтелект безумовно, має перспективне майбутнє в галузі комп'ютерних обчислень, притому досить велика кількість досліджень проводиться у спробах створити надійний та корисний штучний інтелект. Останнім часом у світі стрімкими темпами розвивається нова область прикладної математики, що спеціалізується на штучних нейронних мережах (ШНМ). Актуальність досліджень у цьому напрямку підтверджується великою кількістю різних галузей використання ШНМ. Прикладами цих галузей є: автоматизація виробництва, апроксимація, прогнозування, створення експертних систем, організація асоціативної пам'яті, системи комп'ютерного зору, IoT, Blockchain та багато інших.

Сфери використання штучних нейронних мереж є досить різноманітними. При цьому така сфера задач ШНМ, як розпізнавання образів, є достатньо значимою у побудові комп'ютерного зору, але вона є досить трудомісткою, у зв'язку із природою зображень та швидкістю їх

надходження у систему. Частина досліджень ШНМ включає в себе вивчення області, що називається глибоке навчання (Deep Learning), вона є трудомісткою та потребує багато ресурсів і паралельної обробки. Наприклад, за допомогою ШНМ можна прогнозувати показники бірж, діагностувати хвороби або, наприклад, робити оцифровку рукописних документів, що мають важливе технічне значення або зберігшихся до наших днів давніх рукописів. Також процес оптичного розпізнавання символів прискорить введення персональних даних про громадян до систем обліку та баз даних і допоможе позбутися помилок, які притаманні людині. Це як мінімум, прискорить роботу та зробить сервіси більш зручними.

Необхідність використання апаратних прискорювачів в нейронних мережах зумовлена певними причинами. Оскільки у задачах штучного інтелекту грає роль швидкодія та ресурси доцільно обирати системи із використанням додаткових апаратних засобів для обчислення. Пов'язано це з тим, що обсяги оброблюваної інформації ростуть досить великими темпами, а час очікування щодо збільшення кількості ядер або частоти центральних процесорів є достатньо великою величиною (від року та більше, в залежності від виробника), тому є необхідність у використанні апаратних прискорювачів. Економічна складова також грає важливу роль, тому зменшення енерговитрат є однією з пріоритетних задач. Останнім часом у світі спостерігається бум нейронних мереж. Щоб навчання складних нейронних мереж відбувалось не тижнями/днями/часами, а годинами/хвилинами/секундами, без апаратних прискорювачів не обійтись. Тільки спеціалізоване апаратне забезпечення виконає, наприклад, паралельне множення високої точності із одночасними транзакціями до пам'яті. У майбутньому спеціалізовані ASIC для штучного інтелекту будуть всюди, причому міститимуть на одному кристалі як звичні процесори, так і частково реконфігуровувані блоки [1].

Актуальність проектування апаратних прискорювачів FPGA зумовлена тим, що на відміну від CPU і GPU, мікросхему FPGA можна перепрограмувати під конкретну задачу. Також одна із особливостей FPGA – зменшення енерговитрат у розрахунку на одиницю обчислювальної потужності. Ще однією важливою особливістю FPGA є відносно велика ступінь інтеграції, реконфігурованість та можливість перепрограмування. У свою чергу, мікросхеми FPGA мають відносно доступну ціну, що робить їх ще популярнішими. На рисунку 1 можна побачити вигляд FPGA – прискорювача.



Рисунок 1 – Приклад апаратного прискорювача на FPGA.

В навчальній лабораторії університету доступна плата DE0 із встановленою на ній мікросхемою Cyclone III FPGA. Вона вміщує у себе 200 тис. логічних елементів, 4 Мб пам'яті, 396 множників 18×18 , а також багату периферію (семисегментні індикатори, кнопки, ключі, VGA, PS/2, UART), що дозволяє спроектувати та відлагодити простий апаратний прискорювач, що може бути застосовано в процесі розробки нейронної мережі в рамках наукової роботи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Первый шаг к аппаратным ускорителям нейронных сетей для программистов лежит через изучение основ HDL, RTL и лаб на FPGA [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/post/349750/>

2. Тимошук П.В. Основи проектування нейронних мереж. Навчальний посібник. / П.В. Тимошук, М.В. Лобур // Львів: Видавництво Львівської політехніки, 2007. –328 с.