

УДК 004.272.43

Грушко С.С.¹, Зеленьова І.Я.¹, Голуб Т.В.¹

¹доц. НУ «Запорізька політехніка»

ПРОГРАМОВАНІ ЛОГІЧНІ ІНТЕГРАЛЬНІ СХЕМИ ЯК ЕНЕРГОЕФЕКТИВНА ОСНОВА СТАЛОЇ ОБЧИСЛЮВАЛЬНОЇ ІНФРАСТРУКТУРИ

Вступ. Стрімке зростання обсягів цифрових даних та розвиток технологій штучного інтелекту призводять до безпрецедентного збільшення енергоспоживання обчислювальної інфраструктури. У 2024 році центри обробки даних спожили 415 ТВт·год електроенергії, що еквівалентно 1,5% світового споживання. Прогнози свідчать про подвоєння цього показника до 945 ТВт·год у 2030 році [8]. Найбільш динамічно зростає споживання серверів для задач штучного інтелекту: з 93 ТВт·год у 2025 році до 432 ТВт·год у 2030 році, тобто майже у п'ять разів. За таких умов питання енергоефективності апаратних платформ набуває стратегічного значення для забезпечення сталого розвитку цифрової економіки.

Серед сучасних обчислювальних платформ особливе місце посідають програмовані логічні інтегральні схеми (ПЛІС), які поєднують гнучкість програмного забезпечення з продуктивністю спеціалізованих апаратних рішень. На відміну від графічних процесорів, що домінують у сфері штучного інтелекту, ПЛІС дозволяють створювати оптимізовані обчислювальні структури під конкретні задачі, що забезпечує суттєві переваги в енергоефективності.

Аналіз енергоспоживання обчислювальних платформ. Сучасні графічні процесори для задач штучного інтелекту характеризуються високою розрахунковою тепловою потужністю. Флагманські моделі останнього покоління споживають до 700 Вт, попереднє покоління — від 250 до 400 Вт. Таке енергоспоживання зумовлене архітектурою, орієнтованою на максимальну паралельність обчислень та високу пропускну здатність пам'яті. Натомість ПЛІС-платформи для центрів обробки даних працюють у діапазоні від 75 до 225 Вт залежно від моделі та завантаження [2].

Порівняльні дослідження демонструють, що ПЛІС споживає в середньому лише 28% потужності графічного процесора при виконанні типових обчислювальних задач із порівнянною продуктивністю [6]. Ця перевага пояснюється можливістю створення спеціалізованих обчислювальних конвеєрів, що виключають надлишкові операції та мінімізують передачу даних між компонентами системи.

У сфері комп'ютерного зору перевага ПЛІС особливо виражена. Експериментальні дослідження засвідчують, що для простих операцій (фільтрація, перетворення кольорів) графічні процесори забезпечують кращу енергоефективність у 1,1–3,2 рази. Однак для складних обчислювальних конвеєрів, що включають послідовність операцій обробки зображень, ПЛІС демонструє перевагу в енергоефективності від 1,2 до 22,3 разів [4]. Важливо, що ця перевага зростає пропорційно до складності задачі: чим більше послідовних операцій виконується, тим ефективнішою стає ПЛІС-реалізація.

Аналогічна картина спостерігається у задачах цифрової обробки сигналів. При реалізації фільтрів із скінченною імпульсною характеристикою ПЛІС забезпечує швидкодію у 27 разів вищу за центральний процесор та удвічі вищу за графічний процесор, споживаючи при цьому лише 1,431 Вт [7]. Такі показники роблять ПЛІС незамінними для систем реального часу з обмеженим енергетичним бюджетом.

У контексті прискорення великих мовних моделей ПЛІС також демонструє значний потенціал. Хоча графічні процесори залишаються основною платформою для навчання таких моделей завдяки високій пропускній здатності, етап виконання запитів може бути ефективно реалізований на ПЛІС із суттєвим вигравом в енергоспоживанні [3, 5]. Це особливо актуально, оскільки саме виконання запитів становить до 90% енергоспоживання протягом життєвого циклу моделі.

Методи оптимізації енергоспоживання ПЛІС. Досягнення високої енергоефективності ПЛІС потребує застосування комплексу оптимізаційних методів на різних рівнях проектування [2].

На фізичному рівні застосовується динамічне масштабування напруги, що дозволяє знижувати напругу живлення у періоди низького навантаження. Також використовується відключення живлення неактивних блоків, що особливо ефективно для ПЛІС із великою кількістю ресурсів, частина яких може бути невикористаною у конкретному застосуванні.

На рівні регістрових передач ключовим методом є відключення тактового сигналу для блоків, що не виконують корисних обчислень у поточному такті. Це дозволяє суттєво зменшити динамічну складову споживання. Також важливу роль відіграє оптимізація маршрутизації з'єднань, оскільки міжз'єднання можуть споживати значну частку загальної потужності.

На рівні застосувань перспективним підходом є використання наближених обчислень. Багато задач, зокрема у сфері машинного навчання та обробки сигналів, допускають певну похибку результату без суттєвого впливу на якість. Використання знижених розрядностей даних, спрощених арифметичних операцій та інших методів апроксимації дозволяє зменшити енергоспоживання у кілька разів.

Сучасні ПЛІС-платформи для центрів обробки даних досягають показника 12,6 мВт на блок цифрової обробки сигналів [2], що відповідає найкращим галузевим стандартам та вимогам Європейського кодексу поведінки щодо енергоефективності центрів обробки даних.

Порівняння платформ та рекомендації щодо застосування. Узагальнення результатів аналізу дозволяє сформулювати критерії вибору обчислювальної платформи залежно від характеру задачі.

Графічні процесори залишаються оптимальним вибором для задач із високим ступенем паралелізму та регулярною структурою обчислень. Навчання великих моделей штучного інтелекту, обробка великих масивів однорідних даних, наукові розрахунки з інтенсивним використанням операцій із рухомою комою – усі ці задачі ефективно виконуються на графічних процесорах [3, 6]. Висока пропускна здатність пам'яті та розвинена програмна екосистема забезпечують швидку розробку та високу продуктивність.

ПЛІС доцільно обирати для задач із жорсткими вимогами до затримки, нерегулярною структурою обчислень або обмеженим енергетичним бюджетом [4, 5]. Системи реального часу, периферійні обчислення, спеціалізована обробка сигналів, прискорення виконання запитів до навчених моделей – це типові сфери застосування ПЛІС. Можливість створення оптимізованих апаратних структур забезпечує мінімальну затримку та максимальну енергоефективність.

Перспективним напрямом є створення гібридних архітектур, що поєднують переваги обох платформ. У такій конфігурації графічні процесори використовуються для навчання моделей, де важливою є пропускна здатність, а ПЛІС – для виконання запитів у виробничому середовищі, де важливі енергоефективність та низька затримка. Такий підхід дозволяє оптимізувати як процес розробки, так і експлуатаційні витрати.

Висновки. Проведений аналіз засвідчує, що програмовані логічні інтегральні схеми є перспективною платформою для побудови сталої обчислювальної інфраструктури. Їх ключові переваги включають високу енергоефективність (до 22 разів кращу за альтернативи для складних задач), реконфігурованість, що дозволяє амортизувати екологічний вплив виробництва, та подовжений життєвий цикл.

Водночас ПЛІС не є універсальним рішенням. Графічні процесори зберігають переваги для масово-паралельних задач та навчання великих моделей. Оптимальна стратегія розвитку обчислювальної інфраструктури передбачає диференційований підхід до вибору платформи залежно від характеру робочого навантаження та впровадження гібридних архітектур, що поєднують сильні сторони різних технологій.

Подальші дослідження доцільно спрямувати на розробку методів автоматичного розподілу обчислень між гетерогенними платформами та вдосконалення інструментів оцінки повного екологічного впливу обчислювальних систем.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Sudarshan C. C., Arora A., Chhabria V. A. GreenFPGA: Evaluating FPGAs as Environmentally Sustainable Computing Solutions. DAC '24. San Francisco, CA, USA, 2024. P. 1-6.
2. Tibaldi M., Pilato C. A Survey of FPGA Optimization Methods for Data Center Energy Efficiency. arXiv:2309.12884, 2023.
3. Koilia N., Kachris C. Hardware Acceleration of LLMs: A comprehensive survey and comparison. arXiv:2409.03384, 2024.
4. Qasaimeh M. et al. Comparing Energy Efficiency of CPU, GPU and FPGA Implementations for Vision Kernels. IEEE ISVLSI, 2019.
5. Giasemis F. I. et al. Comparative Analysis of FPGA and GPU Performance for Machine Learning-Based Track Reconstruction at LHCb. arXiv:2502.02304, 2025.
6. Vaithianathan M. et al. Comparative Study of FPGA and GPU for High-Performance Computing and AI. ESP IJACT, 2023. Vol. 1(1). P. 37-46.
7. Arucu M., Iliev T. Performance Evaluation of FPGA, GPU, and CPU in FIR Filter Implementation for Semiconductor-Based Systems. J. Low Power Electron. Appl., 2025. Vol. 15. P. 40.
8. IEA. Energy and AI. International Energy Agency, 2025. URL: <https://www.iea.org/reports/energy-and-ai>