

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ НЕЙРОННИХ МЕРЕЖ В ГІБРИДНИХ СИСТЕМАХ НА ОСНОВІ APU

Зимогляд А. С.

Науковий керівник – проф., д.т.н. Яровий А. А.

Останнім часом технології нейронних мереж все більше актуалізуються для вирішення задач у різноманітних галузях життєдіяльності людини. Разом з тим спостерігається динамічний розвиток методів організації та навчання штучних нейронних мереж, а також комп'ютерних засобів їх реалізації. Зважаючи на це, актуальною задачею є правильний вибір програмно-апаратної платформи для ефективної реалізації штучних нейронних мереж.

Метою дослідження є підвищення швидкодії штучних нейронних мереж на основі їх реалізації в гібридних обчислювальних системах на базі APU (Accelerated Processing Unit).

Нейронні мережі в своїй реалізації мають високу здатність до розпаралелення, оскільки базуються на матричних операціях. Одним з оптимальних сучасних засобів виконання масштабних паралельних обчислень є GPGPU-обчислення (неграфічні обчислення на графічних процесорах). Проте в класичних обчислювальних системах з GPU-прискорювачами (CPU+GPU) значну частину часу займають операції копіювання даних між оперативною пам'яттю та відеопам'яттю. Цього недоліку позбавлені гібридні системи на основі APU, в яких як CPU, так і GPU може використовувати спільний адресний простір в оперативній пам'яті.

Слід зазначити, що використання APU є раціональним в програмах, де обмін даними між CPU і GPU відбувається часто, а масштаби обчислень не дуже великі (оскільки за обчислювальними можливостями сучасні APU не можуть зрівнятися з сучасними дискретними GPU). Проте, системи на APU набагато дешевші, ніж аналогічні за продуктивністю традиційні системи.

На сьогоднішній день оптимальними APU, в яких реалізовано спільний адресний простір для пам'яті CPU і GPU (технологія hUMA – heterogeneous uniform memory access) є процесори AMD з архітектурою Kaveri.

Для написання паралельних програм під AMD APU використовується мова програмування OpenCL (Open Computing Language). Оскільки OpenCL розроблялась, як платформонезалежна мова для паралельних обчислень, то програми написані нею під AMD APU можна перекомпілювати майже без змін під будь-яку іншу платформу (наприклад, на Nvidia CUDA).

Реалізація нейронних мереж в гібридних системах на основі APU дозволяє досягти високої продуктивності (на основі паралельних обчислень та спільного доступу до пам'яті) при мінімальній собівартості апаратної платформи.