УДК 681.325

РЕАЛІЗАЦІЯ ПАРАЛЕЛЬНИХ АЛГОРИТМІВ ЛІНІЙНОЇ АЛГЕБРИ НА ОСНОВІ ТЕХНОЛОГІЇ CUDA

В. П. Семеренко, к.т.н., доцент,
В. В. Ткаченко, студент,
Вінницький національний технічний університет
vpsemerenko@mail.ru

В останні роки бурхливо розвивається новий напрямок паралельних обчислень – неграфічні (універсальні) обчислення на графічних процесорах (GPGPU – General-Purpose Computation on GPU).

Якщо програмування перших поколінь GPU було достатньо трудомістким, то ситуація кардинально змінилась з появою в 2007 році програмно-апаратної технології CUDA (Compute Unified Device Architecture). Ця технологія дозволяє ефективно реалізовувати паралельні алгоритми розширеною мовою С на графічних процесорах компанії NVIDIA восьмого покоління (GeForce 8) і старше. За допомогою CUDA можна прискорити обчислення в сотні разів, але тільки в том випадку, якщо задача може бути розбита на велику кількіть однакових підзадач.

Графічний процесор в CUDA має SIMD-архітектуру, коли одна й та же операція застосовується одночасно до множини даних, розподілених по незалежним потокам.

Саме така модель обчислень часто використовується в задачах завадостійкого кодування і криптоаналізу. Зокрема, актуальним є дослідження коректуючої здатності циклічних кодів в двійкових і недвійкових полях Галуа.

Для отримання об'єктивної характеристики коректуючої здатності циклічного коду достатньо провести аналіз його графової моделі.

Якщо для представлення циклічного коду використати теорію лінійних послідовнісних схем (ЛПС), тоді як графову модель таких кодів доцільно вибрати діаграму переходів (ДП) цього автомату.

Для циклічного (n,k)-коду з непримітивним породжувальним багаточленом над полем $GF(2)$ ДП представляє собою багаторівневий орієнтований граф G_{FA}, який складається із сукупності нульових циклів (НЦ) розмірності не більше n, утворених нульовими дугами. Ці НЦ упорядковані по таким рівням. На нульовому рівні розташований тривіальний НЦ, який складається лише з однієї вершини v_0. На першому рівні знаходиться основний НЦ (ОНЦ) розмірності n, який зв'язаний з вершиною v_0 парою протилежно направлених одиничних дуг. Далі розташовуються периферійні НЦ (ПНЦ). На $(i+1)$-му рівні кожний ПНЦ має $(i+1)$ пар протилежно направлених одиничних дуг с ПНЦ i-го рівня та відсутні одиничні дуги з ПНЦ рівнів $(i-1)$ і менше (i =1,2,3, …). Якщо G_{FA} має τ рівнів, причому на τ-му рівні є C_n^τ НЦ, (где C_n^τ – число сполучень із τ по n), тоді відповідний йому циклічний код может виправити τ випадкових помилок.

ЛПС можна також розглядати і з позицій автоматної моделі. Послідовність векторів внутрішніх станів ЛПС, які відповідають вершинам одного циклу в графі G_{FA}, також утворюють цикл. Оскільки сукупність циклів із векторів станів має таку ж структуру, що і сукупність циклів із

вершин, тому для характеристики циклів із векторів станів будемо використовувати такі ж терміни: НЦ, ОНЦ і ПНЦ.

Якщо графова модель ЛПС зручна для наочного представлення, то всі подальші обчислення доцільно вести в рамках автоматної моделі. Розрахунок автоматного НЦ розмірності n здійснюється по рекурсивній формулі

$$S(t+1) = A \times S(t), \quad GF(q), \ i = 1 \div n, \qquad (1)$$

де A – $((n-k)\times(n-k))$-матриця ЛПС,

$S(t), S(t+1)$ – стани ЛПС в моменти часу t і $t+1$.

Обчислення всіх автоматних НЦ здійснюється за допомогою ітеративної процедури протягом τ ітерацій. На i-й ітерації із m_i НЦ i-го рівня формується m_{i+1} НЦ $(i+1)$-го рівня $(m_{i+1} \geq m_i)$. Це еквівалентно ієрархічній побудові всіх НЦ в рамках графової моделі.

Всі потоки в моделі обчислень CUDA також утворюють ієрархію, тільки іншого типу:

потоки $\rightarrow$ блоки $\rightarrow$ мережа.

Ієрархія потоків дуже добре узгоджується з ієрархічною процедурою обчислення НЦ. На нижньому рівні ієрархії кожний потік (thread) обчислює множину станів по формулі (1).

Оскільки n потоків об’єднуються в один блок, тому потоки одного блоку виконують роботу по формуванню із одного НЦ i-го рівня нових НЦ $(i+1)$-го рівня протягом однієї ітерації.

На наступній ітерації обчислень отримані НЦ стають базовими для формування нових НЦ, тому вони повинні переміститися на нижню ступінь в ієрархії обчислень. Таке переміщення в моделі обчислень CUDA реалізується передачею даних між ступенями в ієрархії

потоків через спільну память, що розділяється (shared memory).

Серед обчислених n нових НЦ на кожній ітерації є велика кількість однакових, вони мають бути вилучені і не приймати участь в подальших операціях. Таким чином, між основними ітераціями, на яких обчислюються НЦ, мають бути допоміжні ітерації, на яких відбувається мінімізація кількості НЦ. Порівняння НЦ між собою вимагає значних витрат часу, тому тут також доцільно максимально використовувати паралельну обробку даних.

Отже, для побудови повної графової моделі для циклічного (n,k)-коду, який виправляє τ випадкових помилок, необхідно $(2\tau-1)$ ітерацій в моделі обчислень CUDA.

Циклічні коди над полем $GF(q), q>2$, до яких належать коди Ріда-Соломона, мають складнішу графову модель. З кожної вершини графа G_{FA} над полем $GF(q)$ входять та виходять по одній нульовій дузі і по n ненульових дуг. Ієрархія НЦ в графі G_{FA} залишається такою ж, тільки їх колькість багатократно збільшується. Для співставлення ієрархії НЦ з ієрархією потоків зручно розглядати блок потоків як дво- або тривимірну структуру.