РЕАЛИЗАЦИЯ НА ПРОЦЕССОРАХ С ТЕХНОЛОГИЕЙ CUDA УНИВЕРСАЛЬНОЙ ОПЕРАЦИИ НАД МАТРИЧНЫМИ СТРУКТУРАМИ

Исламов Г.Г.
Удмуртский государственный университет, г. Ижевск
тел.: 8 (3412) 91-60-90 e-mail: ggislamov@udm.net

Профессор Вышинский В.А. в своём пленарном докладе [1] при анализе восьмого признака отличия одного поколения вычислительной техники от другого показал, что основой вычислительных машин будущего поколения должна стать *универсальная алгоритмическая матрично-алгебраическая система*. В новой машинной технологии обработка сложных структур данных сведётся к обработке матриц, при этом будет использоваться математический аппарат регулярного матричного представления для реализации отвечающей решаемой проблеме числовой алгебры. Аппаратурная реализация принципиально нового языка высокого уровня позволит существенно упростить и удешевить программное обеспечение создаваемой машины будущего поколения. При этом, удастся заменить маломощную машинную арифметику машинной алгеброй и, тем самым, добиться устранения противоречия, возникшего между интеграцией вычислительных схем в аппаратуре вычислительных систем и отсутствием адекватной интеграции в вычислительном технологической процессе. Кроме того, откроется прямой путь повышения производительности вычислений не за счёт повышения тактовой частоты генератора и последующего преодоления «теплового барьера», а за счёт извлечения преимущества из сверхбольшой интеграции.

В докладе будет показано, что практическую реализацию упомянутой выше универсальной алгоритмической матрично-алгебраической системы можно осуществить на базе технологии CUDA [2] многопоточного программирования на процессорах компании NVIDIA и *универсальной операции над матричными структурами* [3-5]. Целесообразность и плодотворность такого подхода к решению различного типа вычислительных задач отмечалась в работах [4-7].

Приводимый ниже фрагмент программы на языке C для CUDA описывает универсальную операцию над матричными структурами для графического устройства, поддерживающего технологию многопоточного программирования:

```
__global__ void Pivot(float * Ad, float * Bd, int p, int q, int M)
{
  float a,b,c,d;
  int i = threadIdx.x;
  int j = threadIdx.y;
  a = Ad[i+M*j]; b = Ad[i+M*q];
  c = Ad[p+M*j]; d = Ad[p+M*q];
  if ((i == p) && (j == q)) Bd[i+M*j] = 1 / d;
     else if (i == p) Bd[i+M*j] = -a / d;
             else if (j == q) Bd[i+M*j] = a / d;
                     else Bd[i+M*j] = a - b * c /d;
  __syncthreads();
}
```

Последний параметр этой процедуры задаёт размеры преобразуемой матрицы, которая определяется первым параметром. Выбор графического устройства сильно влияет на точность и быстродействие вычислений, производимых с помощью процедуры `Pivot`. Наиболее перспективными для наших целей должны оказаться графические устройства на базе процессоров Fermi компании NVIDIA [2]. Этот процессор с 960 ядрами имеет унифицированную 64-разрядную схему адресации памяти. Такая адресация позволяет процессору Fermi работать с адресным пространством объемом до 1 Тб. Отметим другие

значимые для высокопроизводительных вычислений особенности архитектуры этого процессора:

- восьмикратное увеличение пиковой производительности при вычислениях с плавающей запятой двойной точности (по сравнению с GT200);
- унифицированное адресное пространство с полной поддержкой языка C++;
- IEEE 754-2008 32-разрядная и 64-разрядная точность;
- значительно повышенная производительность при работе с атомарными операциями;
- параллельная обработка программных ядер (kernels).

Ниже приводится очередное применение универсальной операции над матричными структурами при расчёте вектора-столбца гонораров и налоговых отчислений, а также вектора-строки приоритетов в задаче государственного управления финансовым рынком. Финансовый рынок служит важной, но не определяющей, составляющей экономики государства. Заложенные исторически правила его функционирования не являются совершенными, что показывает затянувшийся мировой финансовый кризис. Обладая ограниченным в объёме универсальным ресурсом, выражающим всеобщий эквивалент труда, финансовый рынок искусственно создаёт предпосылки для безграничного печатания денег как необходимого условия своего существования. Это приводит к хронической инфляции, последствия которой оказываются катастрофическими как для населения государства, так и для отдельных отраслей экономики. Известно, что финансовый рынок обязан стимулировать инновации в сфере производства материальных благ и воспроизводства интеллектуальных ресурсов, финансировать планомерное развитие потребностей людей и бесперебойное потребление материальных и информационных ресурсов. Это значит, что требуется эффективное управление финансовым рынком со стороны государства. Предлагаемая нами математическая модель государственного управления финансовым рынком показывает, что финансовый рынок может находиться в одном из трёх состояний: *деградация, кризис и стабильность*. *Первое* состояние свидетельствует о крайне неэффективном управлении и отвечает процессу разорения финансового рынка. *Второе* состояние связано с необходимостью включения печатного станка для обеспечения экспоненциального роста денежной массы в зависимости от номера временного периода. *Третье* состояние соответствует степенному росту денежной массы. Всё это есть следствие иерархической структуры финансового рынка и проявление закона его эволюции, которые задаются матрицей A процентов взаимных платежей участников финансового рынка и правилом внешнего управления со стороны государства.

Структура матрицы A описана в работе [8]: имеется m участников финансового рынка, образующих иерархическую структуру с выделенным корнем. Единственный путь, ведущий из этого корня к любой другой вершине, определяет полное имя участника этого рынка. Каждый из участников финансового рынка обладает уникальной способностью увеличивать за указанный период времени (час, неделя, месяц) имеющийся у него капитал на определенное число процентов, не зависящее от объема капитала и порядкового номера периода, но своё для каждого участника. В конце любого периода каждый родитель перечисляет на счет своего дочернего участника определенный процент (для каждого дочернего участника свой) от той суммы, которой родитель располагал в начале временного периода. В это же время все участники финансового рынка (включая и корневого) перечисляют на счёт корневого участника сумму, которая была у них на начало периода, и весь доход, который был получен в течение текущего периода от использования этой суммы. Принятая в модели иерархия финансового рынка обеспечивает его мобильность и управляемость. Однако эти положительные качества могут быть использованы в разрушительных целях, придавая финансовому рынку уязвимость при неблагоприятных внешних воздействиях: интервенция и вывоз капитала, и т. д.

Рассматриваемая модель эволюции финансового рынка описывается дискретным управляемым процессом $x[k+1] = Ax[k] + u[k],\ k = 0,1,2,...,$ где управление со стороны государства формируется по правилу $u[k] = -p \cdot qx[k]$. Здесь $p = colon(p_1,...,p_m)$ есть столбец гонораров и штрафов, установленных государством. Положительная компонента p_i выражает величину гонорара $i-$го участника финансового рынка, тогда как абсолютное значение отрицательной компоненты p_i определяет величину налога, которым облагается участник i в конце планового периода k. Строка $q = (q_1,...,q_m)$ содержит государственные приоритеты, а столбец $x[k] = colon(x_1[k],...,x_m[k])$ - капиталы участников рынка.

В силу иерархической структуры возможна такая нумерация участников финансового рынка, при которой номер дочернего участника больше номера родителя. Мы предполагаем, что такая нумерация уже имеет место. В таком случае матрица A, содержащая проценты перевода денежного капитала среди участников рынка представима в виде суммы нижнетреугольной матрицы с нулями на основной диагонали и одноранговой матрицы. Нетрудно построить подпространства Крылова для этой матрицы. Как известно, каждому базису такого подпространства отвечает фробениусова форма матрицы и отвечающая ей преобразующая матрица, которые могут быть эффективно построены. Можно показать, что каждое ненулевое собственное значение матрицы A имеет геометрическую кратность, равную единице. В силу теоремы двойственности работы [9] за счёт выбора одноранговой матрицы $K = p \cdot q$ можно получить наперёд заданный спектр возмущённой матрицы $A - K$ без учёта нуля, и, тем самым, обеспечить заданный темп роста (убывания) капитала каждого из участников финансового рынка. Мы видим, что рассматриваемая модель финансового рынка приводит к задаче модального управления, в которой требуется построить одноранговую матрицу K, гарантирующую заданный темп роста (убывания) капитала. Анализ доказательства теоремы двойственности работы [9] показывает, что можно указать эффективный алгоритм построения такой матрицы K. Таким образом, за счёт выбора управляющих векторов p, q можно обеспечить любое из указанных выше трёх состояний финансового рынка.

Ниже описывается начальная информационная структура высокопроизводительного алгоритма построения управляющих векторов. Все вычисления организованы в виде двумерной таблицы, первая из которых имеет вид

$$\begin{array}{cccccc} & & -s_1 & +s_2 & \dots & (-1)^n s_m \\ t_1 & b_1 & d_{11} & d_{12} & \dots & d_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t_m & b_m & d_{m1} & d_{m2} & \dots & d_{mm} \end{array}$$

В первой строке находятся s-переменные, которые обозначают коэффициенты характеристического многочлена матрицы A. В первом столбце находятся вспомогательные t-переменные, которые будут обмениваться с s-переменными на каждом шаге применения универсальной операции над матричными структурами [3]. Стратегия применения этой операции, правило пересчёта таблицы и математическое обоснование необходимых для наших целей свойств можно найти в работе [5]. Следующие столбцы исходной таблицы составлены из координат соответственно векторов $A^m e, A^{m-1} e,..., Ae, e$, где e есть порождающий вектор базиса Крылова. Они, без учёта первого вектора, образуют *матрицу перехода* от единичного базиса к базису Крылова. После m-кратного применения универсальной операции получим последнюю таблицу, в первом столбце которой окажутся все s-переменные, во втором столбце будут находиться коэффициенты характеристического многочлена. В последних столбцах находится информация для получения *преобразующей матрицы*, которая вместе с

матрицей перехода необходима для построения управляющих векторов p и q. Тем самым обмен информацией между центральным процессором и графическим устройством происходит лишь на этапе ввода указанной выше двумерной таблицы и вывода конечного результата. Сама двумерная таблица формируется на центральном процессоре.

Литература

1.Вышинский В.А. Решение фундаментальной проблемы современного развития вычислительной техники // Труды Первой международной конференции «Трёхмерная визуализация научной, технической и социальной реальности. Кластерные технологии моделирования, Т. 1. Ижевск, УдГУ (4-6 февраля 2009 г.), С. 27-36.

2.NVIDIA CUDA Programming Guide 3.0 // www.nvidia.com.

3.Исламов Г.Г. Универсальная операция над матричными структурами // Современные проблемы вычислительной математики и математической физики: Международная конференция, Москва, МГУ имени М.В. Ломоносова, 16-18 июня 2009 г. Тезисы докладов. – М.: Издательский отдел факультета ВМК МГУ имени М.В. Ломоносова; Макс Пресс, 2009. – 396 с.

4.Исламов Г.Г. Факторы процесса внедрения высокопроизводительных вычислений в учебно-научный и производственный потенциал Удмуртской Республики // Труды Первой международной конференции «Трёхмерная визуализация научной, технической и социальной реальности. Кластерные технологии моделирования, Т. 1. Ижевск, УдГУ (4-6 февраля 2009 г.), С. 139-143.

5.Исламов Г.Г., Коган Ю.В. Об одном алгоритме поиска базисного минора матрицы // Вестник удмуртского университета. – 2006. - № 1 : Математика. – С. 63 – 70.

6.Исламов Г.Г., Исламов А.Г. Гибридные вычисления при расчёте балансовой модели экономики // Новые образовательные технологии в вузе (НОТВ-2010): седьмая междунар. науч.-метод. конф., 8-10 февр. 2010 г.; сб. материалов / отв. ред. А.В. Цветков. – Екатеринбург, 2010. – Ч. 2. – С. 254-258.

7.Исламов Г.Г., Исламов А.Г. О минимальной запаздывающей и корректирующей связи для балансовой модели производства // Параллельные вычислительные технологии 2010 : третья междунар. науч. конф., 29 марта – 2 апреля 2010 г., сб. материалов. – Уфа.

8.Исламов А.Г., Исламов Г.Г., Лукин О Л. Об одной модели финансового рынка. Режим доступа: http://www.iton.mfua.ru/tesis/islamov2.doc (дата обращения: 15.03.2010)

9. Исламов Г.Г. Об управлении спектром динамической системы // Дифференц. уравнения. 1987. Т. 23, № 8. – С. 1299-1302.