

О реализации вихревых методов вычислительной гидродинамики с использованием технологии CUDA

С.Р. Гречкин-Погребняков, И.К. Марчевский
МГТУ им. Н.Э. Баумана

Исследована возможность ускорения вычислений в методе вихревых элементов — бессеточном лагранжевом методе вычислительной гидродинамики — за счет использования графических ускорителей. Реализован алгоритм, основанный на авторской модификации метода вихревых элементов, позволяющий проводить все необходимые вычисления непосредственно на графическом ускорителе с использованием технологии CUDA. Скорость решения типичной задачи на одном ускорителе GeForce GTX 970 или Tesla C2050 оказывается соизмеримой со скоростью решения аналогичной задачи на кластере, содержащем 30-40 ядер, при использовании технологии MPI. Полученные результаты говорят о высокой эффективности применения графических ускорителей при решении задач гидродинамики вихревыми методами.

1. Введение

Вычислительный эксперимент в механике сплошной среды занимает сегодня центральное место в исследовании физических процессов и технических систем. Все возрастающие требования к точности моделирования при сохранении приемлемого времени выполнения расчетов служат на протяжении последних десятилетий мощным стимулом развития как численных методов, так и вычислительной техники.

Задачи вычислительной гидро- и газодинамики, а также сопряженные задачи гидроаэроупругости являются одними из наиболее сложных с вычислительной точки зрения в силу нелинейности исходных математических моделей и особенностей их численного анализа (в частности, необходимости расчета течений в пограничных слоях, где на малом расстоянии происходит значительное изменение параметров течения). Выбор подхода к решению задачи вычислительной гидромеханики не всегда является очевидным; существует значительное количество различных методов, которые упрощенно можно разделить на 3 класса: эйлеровы, в которых расчет ведется на неподвижной сетке; лагранжевы, в которых сетка движется вместе с частицами среды (либо вообще отсутствует в явном виде); и гибридные эйлерово-лагранжевые подходы, в определенной мере соединяющие в себе преимущества методов обоих классов. При решении задач гидроупругости предпочтительными являются лагранжевы и гибридные методы, поскольку область, занятая жидкостью или газом, изменяется на каждом шаге расчета: ее форма зависит от деформации и перемещения обтекаемых тел, которые, в свою очередь, определяются действующими на них гидродинамическими нагрузками.

Существует весьма узкий, однако важный с практической точки зрения класс задач, предполагающий моделирование течений с малыми дозвуковыми скоростями, когда влиянием сжимаемости среды можно пренебречь. Если область, в которой завихренность отлична от нуля, сравнительно невелика (что обычно наблюдается при моделировании внешнего обтекания тел), то для решения таких задач могут быть особенно эффективны бессеточные лагранжевы вихревые методы: в случае решения плоских задач — метод дискретных вихрей [1, 2] и метод вязких вихревых доменов [3]; для моделирования пространственных течений — метод вихревых рамок [2, 4], метод вихревых отрезков [5] и их модификации.

Вопросам ускорения вычислений в вихревых методах посвящено достаточное количество работ; основные пути повышения скорости счета — это использование приближенных быстрых методов [6] и применение параллельных вычислительных алгоритмов с исполь-

зованием технологии MPI [7, 8]. Получаемые результаты расчетов показывают хорошую масштабируемость алгоритмов вихревых методов. В работе [8] показано, что при распараллеливании всех основных операций удается достичь ускорения в 20...30 раз при проведении расчетов на 64-ядерном кластере. В то же время публикации, посвященные систематическим исследованиям в области проведения расчетов вихревыми методами с использованием графических ускорителей и технологии nVidia CUDA, практически отсутствуют; отдельные сведения приведены в [9, 10]. Отметим, что в активно развиваемых за рубежом гибридных методах расчеты на графических ускорителях применяются весьма успешно [11].

Целью настоящей работы является разработка модификации алгоритма метода вязких вихревых доменов, адаптированной под использование графических ускорителей и предназначенной для расчета двумерных течений несжимаемой среды и моделирования обтекания профилей, и исследование эффективности ее применения.

2. Постановка задачи

Рассматривается двумерная задача о моделировании внешнего обтекания жесткого неподвижного профиля потоком вязкой несжимаемой среды постоянной плотности, движение которой описывается уравнением неразрывности и уравнением Навье — Стокса:

$$\nabla \cdot \vec{V} = 0, \quad (1)$$

$$\frac{\partial \vec{V}}{\partial t} + (\vec{V} \cdot \nabla) \vec{V} = -\nabla \left(\frac{p}{\rho} \right) + \nu \Delta \vec{V}. \quad (2)$$

Здесь $\vec{V} = \vec{V}(\vec{r}, t)$ — скорость среды, $p = p(\vec{r}, t)$ — давление, $\rho = \text{const}$ — плотность, $\nu = \text{const}$ — кинематическая вязкость; ∇ и Δ — дифференциальные операторы Гамильтона и Лапласа. Начальное распределение скоростей в среде полагается известным; в качестве граничных условий используются условие затухания возмущений на бесконечности

$$\vec{V}(\vec{r}, t) \rightarrow \vec{V}_\infty, \quad p(\vec{r}, t) \rightarrow p_\infty \quad \text{при} \quad |\vec{r}| \rightarrow \infty$$

и условие прилипания на границе K профиля

$$\vec{V}(\vec{r}, t) = 0 \quad \text{при} \quad \vec{r} \in K.$$

При проведении практических расчетов в зависимости от решаемой задачи могут представлять интерес как нестационарное поле скоростей среды и поле давления в ней, так и распределение давления по обтекаемому профилю, а также интегральные величины действующих на профиль со стороны потока нагрузок — лобового сопротивления, подъемной силы и аэродинамического момента.

3. Краткое описание вихревых методов вычислительной гидродинамики

Использование вихревых методов предполагает переход к завихренности $\vec{\Omega} = \nabla \times \vec{V}$ как к первичной расчетной величине; при этом в плоских задачах гидродинамики вектор $\vec{\Omega}$ имеет лишь одну ненулевую компоненту. Поле скоростей может быть восстановлено по известному распределению завихренности при помощи закона Био — Савара

$$\vec{V}(\vec{r}, t) = \vec{V}_\infty + \iint_S \frac{\vec{\Omega}(\vec{\xi}, t) \times (\vec{r} - \vec{\xi})}{|\vec{r} - \vec{\xi}|^2} dS_\xi, \quad (3)$$

где интеграл берется по области течения S ; при этом уравнение неразрывности (1) выполняется автоматически. Для расчета давления наиболее удобен аналог интегралов Бернулли

и Коши — Лагранжа [12], при этом для вычисления интегральных значений гидродинамических нагрузок, действующих на обтекаемый профиль (главного вектора сил и момента), можно использовать весьма простые выражения, также приведенные в [12].

В терминах завихренности уравнение (2) принимает простой вид

$$\frac{D\vec{\Omega}}{Dt}\Big|_{\vec{v}} = \nu\Delta\vec{\Omega}, \quad (4)$$

где $\frac{D}{Dt}\Big|_{\vec{v}} = \frac{\partial}{\partial t} + (\vec{V} \cdot \nabla)$ — субстанциональная (материальная) производная. В случае моделирования течения идеальной (невязкой) несжимаемой среды уравнение (4) выражает собой закон Гельмгольца, в соответствии с которым завихренность «вморожена» в жидкость, а эволюция завихренности, таким образом, представляет собой ее перенос вместе с частицами жидкости.

Отметим, что во многих инженерных приложениях, в частности, при моделировании обтекания профилей, имеющих острые кромки и угловые точки, при проведении моделирования течений и расчете нагрузок можно ограничиться моделью идеальной жидкости. Данный подход положен в основу известного метода дискретных вихрей (МДВ) [1, 2], активно развиваемого еще с 60-х годов и хорошо зарекомендовавшего себя в целом ряде отраслей.

Если влияние вязкости существенно и требуется ее корректный учет в расчете, возможен расчет так называемой диффузионной скорости $\vec{W} = -\nu\frac{\nabla\Omega}{\Omega}$, эффективный способ вычисления которой разработан в методе вязких вихревых доменов (ВВД) [3, 12]. С учетом диффузионной скорости уравнение (4) можно привести к виду

$$\frac{D\vec{\Omega}}{Dt}\Big|_{\vec{v}+\vec{W}} = 0, \quad (5)$$

где по аналогии с субстанциональной производной введено обозначение

$$\frac{D}{Dt}\Big|_{\vec{v}+\vec{W}} = \frac{\partial}{\partial t} + ((\vec{V} + \vec{W}) \cdot \nabla).$$

Механический смысл уравнения (5) заключается в том, что в области течения S происходит перенос имеющейся завихренности со скоростью $\vec{V} + \vec{W}$, при этом «новая» завихренность внутри области течения не образуется. Генерация завихренности происходит лишь на границе области течения, т. е. на обтекаемом профиле.

Вихревые методы относятся к так называемым методам частиц (particle methods) [13], при этом в качестве таких частиц выступают элементарные поля завихренности — вихревые элементы, при помощи которых моделируется распределение завихренности в области течения. Общее количество N вихревых элементов может быть достаточно большим и достигать десятков-сотен тысяч и даже миллионов; сами вихревые частицы в методах МДВ и ВВД представляют собой круглые вихри постоянного достаточно малого радиуса ε .

Положение каждого вихревого элемента в пространстве характеризуется радиус-вектором \vec{r}_i , а содержащаяся в нем завихренность — циркуляцией Γ_i , $i = 1, \dots, N$. При этом влияние вихревых элементов на скорость среды в произвольной точке \vec{r} вычисляется в соответствии с дискретным аналогом закона Био — Савара (3)

$$\vec{V}(\vec{r}, t) = \vec{V}_\infty + \sum_{i=1}^N \frac{\Gamma_i}{2\pi} \frac{\vec{k} \times (\vec{r} - \vec{r}_i)}{\max\{|\vec{r} - \vec{r}_i|^2, \varepsilon^2\}}, \quad (6)$$

где \vec{k} — единичный вектор, перпендикулярный плоскости течения.

Основной смысл вихревых методов состоит в переходе от решения уравнения Навье — Стокса (2), или, после преобразований, уравнения в частных производных (4) или (5),

к решению системы обыкновенных дифференциальных уравнений, описывающих движение вихревых элементов. Действительно, описанный выше «механический смысл» уравнения (5) позволяет считать, что циркуляции вихревых элементов остаются постоянными, а их перемещение происходит со скоростью $\vec{V} + \vec{W}$:

$$\begin{cases} \frac{d\Gamma_i}{dt} = 0, \\ \frac{d\vec{r}_i}{dt} = \vec{V}(\vec{r}_i) + \vec{W}(\vec{r}_i), \end{cases} \quad i = 1, \dots, N. \quad (7)$$

Здесь $\vec{V}(\vec{r}_i)$ — скорость среды в точке расположения i -го вихревого элемента, вычисляемая по формуле (6), $\vec{W}(\vec{r}_i)$ — диффузионная скорость в этой точке.

Отметим, что возможен и альтернативный подход — считать положения вихревых элементов неизменными, а их циркуляции пересчитывать так, чтобы моделировать перенос завихренности. На такой интерпретации основан класс численных методов, называемых в литературе методами «вихрь в ячейке» (vortex-in-cell) [13].

Граничное условие на бесконечности в вихревых методах выполняется автоматически, при этом не требуется искусственно ограничивать расчетную область, как это делается в сеточных методах. Выполнение граничного условия прилипания на профиле (или условия непротекания, если моделируется течение невязкой среды) обеспечивается генерацией на нём вихревого слоя на каждом шаге расчета по времени, также моделируемого набором вихревых элементов. Этот вихревой слой является присоединенным, если обтекание невязкое, т. е. вихревые элементы, его моделирующие, остаются на своих местах и заменяются новыми на следующем шаге расчета, и является свободным при обтекании профиля вязкой жидкостью. Последнее означает, что вихревые элементы «сходят в поток» и становятся частью вихревого следа, образующегося вблизи обтекаемого профиля и позади него.

Для определения интенсивности вихревого слоя на профиле и последующего расчета циркуляций вихревых элементов можно применять различные методы и подходы, сводящиеся к решению системы линейных алгебраических уравнений, размерность которой соответствует густоте дискретизации профиля. Наибольшую эффективность, особенно применительно к методу ВВД, показал метод, основанный на обеспечении равенства нулю касательной компоненты скорости на профиле [14]; он во многих случаях превосходит по точности «классический» подход, обычно используемый в МДВ, на несколько порядков при одинаковой дискретизации профиля.

4. Программная реализация и оценка вычислительной сложности вихревых методов

Моделирование течений вихревыми методами включает в себя последовательное выполнение ряда операций алгоритма на каждом шаге расчета (рис. 1). Обычно для установления режима обтекания профиля требуется выполнение нескольких тысяч и даже десятков тысяч временных шагов. Это количество определяется формой профиля, его ориентацией по отношению к набегающему потоку, тщательностью дискретизации профиля (более тщательная дискретизация требует более мелкого шага во времени) и рядом других параметров.

Особенности решаемых задач и конкретных применяемых алгоритмов могут влиять на реализацию операций из указанных блоков (в частности, некоторые из блоков могут отсутствовать). Будем считать, что на каждом шаге расчета на профиле генерируется n вихревых элементов, общее количество вихрей в области течения равно N ($n \ll N$). В практических расчетах n обычно составляет от нескольких сотен до нескольких тысяч, N может сильно варьироваться от задачи к задаче и составляет от нескольких десятков тысяч до нескольких сотен тысяч. Отметим, что вычисление скорости, генерируемой одним вихревым элементом в произвольной точке области течения, как следует из анализа формулы (6), требует выполнения 6 арифметических операций умножения/деления.

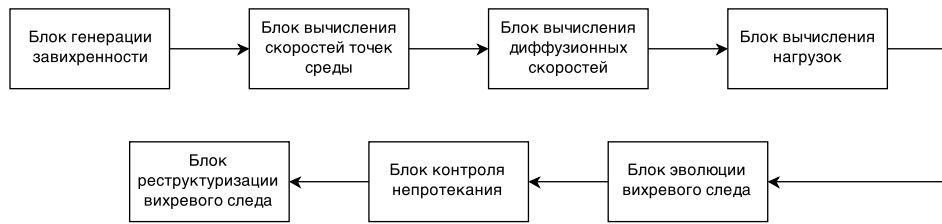


Рис. 1. Блок-схема программы

Далее оценим трудоемкости отдельных блоков алгоритма, указанных на рис. 1.

Блок генерации завихренности. В данном блоке производится решение системы линейных алгебраических уравнений и определяются циркуляции вихревых элементов, генерируемых на данном шаге расчета. Матрица системы вне зависимости от используемого метода определяется лишь формой обтекаемого профиля, поэтому в случае обтекания недеформируемого профиля она остается постоянной. Формирование матрицы системы и ее обращение производятся однократно перед началом счета, поэтому трудоемкость этих операций в расчет не применяется. На каждом шаге производится лишь вычисление правой части системы, основанное на определении скоростей среды в n точках профиля. Отыскание решения системы с матрицей $n \times n$ сводится к умножению обратной матрицы на вектор правой части. Таким образом суммарная трудоемкость блока генерации завихренности приближенно равна $(6N + n) \cdot n \approx 6N \cdot n$.

Блок вычисления скоростей точек среды является наиболее трудоемким; здесь необходимо вычислить скорости всех вихревых элементов, находящихся в области течения. Данная задача аналогична задаче N тел, ее вычислительная сложность при непосредственном вычислении всех парных влияний вихревых элементов друг на друга составляет $6N^2$.

Блок вычисления диффузионных скоростей. Вычислительную сложность блока вычисления диффузионных скоростей оценить труднее, поскольку она зависит от особенностей реализации алгоритма и конкретных значений параметров расчетной схемы, однако поскольку диффузионные скорости также определяются парными влияниями вихревых элементов и влиянием на них обтекаемого профиля, вычислительная сложность алгоритма в первом приближении также пропорциональна N^2 . Результаты практических расчетов показывают, что данный блок является в 1,5...2,0 раза более трудоемким, чем предыдущий.

Блок вычисления нагрузок. Как упоминалось выше, в зависимости от решаемой задачи могут применяться различные алгоритмы вычисления гидродинамических нагрузок. В простейшем случае, когда вычисляются лишь интегральные нагрузки, зависящие только от распределения давления на профиле, сложность данного алгоритма крайне низка и пропорциональна n ; учет силы вязкого трения требует дополнительных затрат, при этом сложность становится пропорциональной n^2 . Расчет распределения давления и касательных напряжений на профиле будет иметь сложность, пропорциональную $N \cdot n$.

Блок эволюции вихревого следа реализует смещение всех вихревых элементов со скоростями $\vec{V}_i + \vec{W}_i$, найденными ранее. Таким образом, его вычислительная трудоемкость низка и пропорциональна N .

Блок контроля непротекания необходим, поскольку в ходе расчета отдельные вихревые элементы могут проникать внутрь профиля. В данном блоке производится их программное отыскание и «извлечение» из профиля. Вычислительная сложность данной операции также в первом приближении пропорциональна n^2 .

Блок реструктуризации вихревого следа обеспечивает объединение близкорасположенных вихревых элементов в один, а также исключение из расчетной схемы вихревых элементов, удалившихся от обтекаемого профиля на значительное расстояние (обычно 10...20 характерных размеров профиля), поскольку они практически не оказывают влияния на параметры течения вблизи профиля. Вычислительная сложность алгоритма, как

показывают расчеты, пропорциональна N^2 , при этом коэффициент пропорциональности сравнительно мал (меньше значения 6, соответствующего второму блоку).

5. Параллельная реализация вихревых методов

Созданный к настоящему времени программный комплекс POLARA [15] позволяет производить расчеты обтекания профилей произвольной формы потоком вязкой несжимаемой среды. Разработанные параллельные реализации основных операций алгоритма позволили существенно сократить время выполнения расчетов на многопроцессорных вычислительных комплексах с использованием библиотеки MPI.

Для повышения эффективности работы программного комплекса на системах с общей памятью была исследована возможность применения для этих же целей технологии OpenMP. С ее помощью были распараллелены те же 3 наиболее трудозатратные операции, что и при использовании MPI. При проведении расчетов на вычислительных машинах с общей памятью на 2–8 вычислительных ядрах было установлено, что выигрыш по времени, достигаемый за счет использования OpenMP, не превышает нескольких процентов по сравнению с применением MPI. В связи с этим данная возможность практически не использовалась, поскольку применение MPI позволяет сделать программу универсальной — она может работать как на системах с общей, так и с распределенной памятью (кластерах). В последнем случае эффективность распараллеливания, очевидно, снижается, но такая возможность необходима для решения сложных задач, когда в расчетной схеме присутствуют десятки-сотни тысяч вихревых элементов.

Примером такой вычислительно сложной задачи может служить тестовая задача Блазиуса о расчете обтекания тонкой пластинки вязким потоком. Данная задача (для случая бесконечно тонкой и бесконечно длинной пластинки) имеет точное аналитическое решение, с которым можно сравнить результаты расчета. Характерный вид вихревого следа в пограничном слое на пластинке показан на рис. 2 [16].

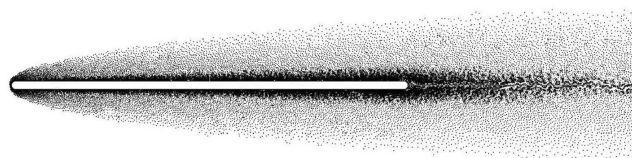


Рис. 2. Вид вихревого следа вблизи пластинки. Точки — положения вихревых элементов

В расчетах рассматривалась пластинка с относительной толщиной 2% и полукруглыми законцовками; число Рейнольдса, вычисленное по длине пластинки, составляло 10^3 , вихревой след в установившемся режиме состоял примерно из 120 000 вихревых элементов. Результаты расчета компонент вектора скорости в сечении пограничного слоя, отстоящем от переднего края пластинки на одну треть ее длины, в сравнении с точным аналитическим решением показаны на рис. 3. Отметим, что аналитическое решение является автомодельным, поэтому положение сечения может быть выбрано произвольным образом; необходимо лишь, чтобы оно отстояло на некоторое расстояние от задней кромки пластинки конечной длины, рассматриваемой в расчете.

В качестве путей дальнейшего ускорения вычислений можно отметить два направления. Первое — это использование приближенных быстрых методов расчета вихревого влияния [6], для которых построена весьма точная оценка вычислительной сложности [17], позволяющая оптимальным образом подобрать параметры алгоритма и снизить вычислительную сложность вычисления вихревого влияния с $O(N^2)$ до $O(N \log N)$. Оценки и результаты практических расчетов [18] показывают, что ускорение за счет использования быстрого метода достигает нескольких десятков раз при сохранении приемлемой точности.

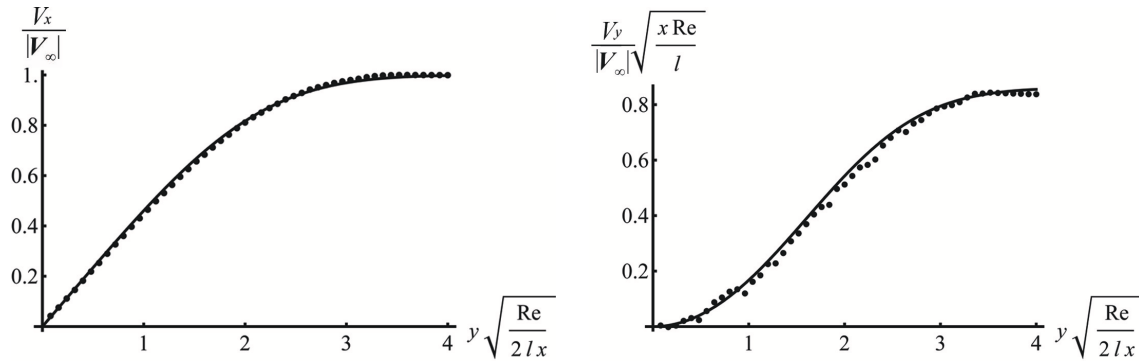


Рис. 3. Распределение продольной (слева) и поперечной (справа) компоненты скорости в пограничном слое в сечении $x = 0,33$. Сплошная линия — точное аналитическое решение

Второй подход предполагает использование для проведения расчета возможностей графических ускорителей и технологии nVidia CUDA. Отметим, что данные подходы не являются взаимоисключающими; применительно к гравитационным задачам N тел известны эффективные реализации приближенных быстрых методов на GPU [19]. В настоящей работе ограничимся переносом на GPU алгоритма, в котором парные влияния вихревых элементов рассчитываются непосредственно.

6. Особенности использования технологии CUDA

Важной особенностью вихревых методов, позволяющей эффективно использовать возможности графических ускорителей (GPU), являются малые затраты оперативной памяти при проведении расчета, поскольку в данном случае не требуется строить сетку и хранить значения неизвестных во всех ее узлах. Оценим необходимые затраты оперативной памяти для работы алгоритма на рассмотренном выше примере решения задачи Блазиуса.

Для хранения обратной матрицы для расчета циркуляций генерируемых на профиле вихрей требуется около 50 Мб памяти (при генерации 2600 ВЭ на профиле на каждом шаге). В разработанном алгоритме для каждого вихревого элемента требуется хранить его циркуляцию, координаты, вектор скорости, а также характерное расстояние от него до ближайших к нему вихрей, следовательно, требуется $48 \cdot N$ байт оперативной памяти, где N — количество ВЭ в области течения. В тестовых расчетах N не превышало трехсот тридцати тысяч, следовательно, затраты памяти на их хранение составляют около 15 Мб. При выполнении каждого шага в памяти хранятся как «старые» так и «новые» параметры вихрей, поэтому требуется дополнительно 15 Мб памяти. Если для интегрирования системы (7), описывающей движение вихрей, будет использоваться многостадийный метод Рунге — Кутты, то это потребует дополнительных 15 Мб оперативной памяти в расчете на каждую промежуточную стадию. В данном примере использовался явный метод Эйлера.

Некоторое количество оперативной памяти требуется также на хранение параметров обтекаемого профиля (не более нескольких сотен килобайт), а также «служебной информации». Порядка 30 Мб в памяти GPU резервируется системой при запуске любой программы, использующей технологию CUDA. Таким образом, для успешной работы программы в данном примере достаточно менее 120 Мб свободной оперативной памяти. Отметим, что приведенные оценки являются по сути оценками сверху: значения $n = 2600$ и $N = 330000$ являются избыточными для большинства практических задач, поэтому реальные затраты оперативной памяти на практике скорее всего окажутся существенно ниже. В таблице 1 показано сравнение полученных оценок потребления оперативной памяти с ее реальным расходом при работе программы, полученном с помощью утилиты MSI Afterburner.

Таблица 1. Затраты оперативной памяти

Назначение	Оценка	Экспериментальные данные
Матрица системы	51 Мб	52 Мб
Параметры ВЭ	2×15 Мб	2×16 Мб
Системные затраты при запуске	30 Мб	30 Мб
Пиковое потребление	111 Мб	114 Мб

Таким образом, оценка потребности в оперативной памяти

$$Mem = 8 \cdot n^2 + 2 \cdot 48 \cdot N \quad (\text{байт})$$

является весьма точной.

Столь низкие затраты оперативной памяти упрощают пересылку данных между оперативной памятью компьютера и оперативной памятью графического ускорителя, однако время на их выполнение все равно оказывается существенным, поэтому была разработана реализация алгоритма вихревого метода, в которой все вычисления полностью перенесены на GPU, а пересылки данных между GPU и компьютером происходят лишь при сохранении промежуточных результатов.

На практике сохранение текущего состояния вихревого следа в файл производится обычно раз в несколько десятков или даже сотен шагов; значения гидродинамических нагрузок, действующих на тело, вычисляются на каждом шаге, однако они представляют собой лишь три действительных числа, пересылка которых не вызывает проблем. Различная статистика расчета и прочая «телеметрическая» информация в силу ее небольшого объема (обычно суммарно несколько десятков-сотен килобайт, реже — мегабайты) может сохраняться в оперативной памяти GPU и пересылаться в память компьютера для дальнейшего сохранения в файл один раз в конце расчета, либо регулярно, но редко, скажем, 1 раз за 1000 шагов.

Анализ используемых алгоритмов показал, что для их эффективного исполнения на GPU требуется не слишком значительная их переработка; в частности, практически все операции были реализованы без использования операторов условного перехода `if...else`. Кроме того, в целях оптимизации программы все выполняемые операции, ранее разделенные на 7 блоков (рис. 1), были сведены в 3 крупных блока, названных по наиболее трудоемкой операции: блок генерации, блок скоростей, блок реструктуризации.

Определение скоростей, т. е. расчеты парных влияний вихревых элементов (ВЭ) друг на друга могут выполняться независимо, что делает очевидным подход к распараллеливанию данной операции. Каждое CUDA-ядро вычисляет скорость одного ВЭ, используя разделяемую память, в которую поочередно помещаются блоки из N_b ВЭ. При этом общее количество ВЭ увеличивается до кратного N_b путем добавления «фиктивных» ВЭ с нулевой циркуляцией. Аналогично производится расчет диффузионных скоростей.

При **генерации** ВЭ на профиле каждое CUDA-ядро вычисляет скорость среды в точке рождения аналогично вычислению парных влияний ВЭ.

При **реструктуризации** вихревого следа (включающей в себя перемещение ВЭ, контроль непротекания и собственно реструктуризацию — объединение близкорасположенных ВЭ в один) для каждого вихревого элемента необходимо проверять несколько условий и в случае их выполнения проводить над данным ВЭ некоторые операции. Использование оператора ветвления в программах, использующих GPGPU, значительно снижает скорость выполнения расчета, поэтому для минимизации негативных последствий необходимо,

во-первых, категорически избегать вложенных ветвлений, а, во-вторых, проектировать алгоритм так, чтобы вероятность выполнения одной из веток была значительно выше другой и вычислительная сложность менее вероятного варианта была многократно меньшей по сравнению с основным. Также ветвления алгоритма не критичны в случае крайне низкой сложности каждого из путей.

В связи с указанными особенностями **контроль непротекания** реализован в 2 этапа: на первом этапе для каждого ВЭ проверяется условие попадания в габаритный прямоугольник профиля (прямоугольник со сторонами, параллельными осям координат, в который вписан обтекаемый профиль); попавшие в него запоминаются в отдельном массиве — таким образом вероятность непопадания в прямоугольник выше вероятности попадания, а вычислительная сложность обеих веток незначительна, что соответствует приведенным критериям эффективного применения операторов ветвления в CUDA-программах. Второй этап выполняется лишь для ВЭ, оказавшихся внутри габаритного прямоугольника; для каждого из них проверяется пересечение траектории его движения на последнем шаге с панелями профиля — участками ломаной, аппроксимирующей профиль.

Операция **объединения близкорасположенных вихревых элементов** из тех же соображений разделена на два этапа. На первом этапе для каждого ВЭ происходит поиск первого подходящего для объединения вихревого элемента, и номера ВЭ, образующих эту пару, запоминается в отдельном массиве. Таким образом, для каждого ядра при выполнении этого этапа возможны два состояния — либо продолжается поиск пары, либо вычисления завершены, причем первый вариант более вероятен. Это приводит лишь к небольшому замедлению всего алгоритма. Второй этап объединения предполагает внесение изменений в параметры ВЭ, поэтому для соблюдения условия консистентности данных он проводится в последовательном режиме (но выполняется на GPU) и состоит в объединении пар ВЭ из служебного массива.

На заключительной стадии блока реструктуризации из расчета исключаются ВЭ, удалившиеся на значительное расстояние от профиля и происходит перенумерация массива ВЭ для вытеснения из него удаленных ранее вихрей. Этот этап также проводится в последовательном режиме и имеет сравнительно малую трудоемкость.

Операция **вычисления гидродинамических нагрузок** при проведении расчета на GPU получается пренебрежимо малозатратной, поскольку все основные вычисления (при вычислении интегральных нагрузок на профиль и силы вязкого трения) производятся одновременно с вычислением парных влияний ВЭ и их диффузионных скоростей.

Таким образом, операция вычисления скоростей ВЭ исключительно эффективно распараллеливается на GPU, эффективность реализации остальных операций несколько ниже; незначительная часть кода, имеющая минимальную вычислительную сложность, выполняется на GPU в последовательном режиме.

В настоящей работе расчеты на GPU проводились с использованием графических ускорителей Tesla C2050 и GeForce GTX 970, имеющих 448 и 1644 ядра соответственно и по 14 мультипроцессоров. Результаты расчетов показывают, что оптимальным размером блока для ускорителя Tesla C2050 является значение $N_b = 128$. Для ускорителя GeForce GTX 970 размер блока практически не влияет на время счета, что можно объяснить особенностью архитектуры последнего поколения GPU: общая DRAM-память является кэшируемой.

Ниже в таблицах 2 и 3 приведены затраты времени на проведение 100 шагов одного и того расчета с одинарной и двойной точностью; на обтекаемом профиле на каждом шаге рождалось $n = 2576$ вихревых элементов, их количество в следе нарастало от 0 до 51 500 и в среднем составляло около 40 000.

Анализ данных, приведенных в таблицах 2 и 3, показывает также, что в графических ускорителях GeForce последнего поколения обработка ветвлений реализована более эффективно (блок Б.3), однако они уступают сравнительно старым ускорителям Tesla при выполнении блока Б.2, где ветвления практически отсутствуют.

Таблица 2. Время расчета на GPU Tesla C2050 / GeForce GTX 970, одинарная точность

	Время счета (в секундах) при различном размере блока				
	16	32	64	128	256
Блок генерации (Б.1)	4,4/1,2	2,4/1,0	2,4/1,0	3,0/1,0	3,8/1,0
Блок скоростей (Б.2)	241,6/33,8	124,9/18,0	87,7/17,8	85,4/17,7	89,3/17,6
Блок реструкт-и (Б.3)	189,7/69,9	108,6/46,1	70,1/34,9	64,3/34,8	68,7/35,6
Общее время расчета	435,7/104,9	235,9/65,1	160,2/53,7	152,7/53,5	161,8/54,2

Таблица 3. Время расчета на GPU Tesla C2050 / GeForce GTX 970, двойная точность

	Время счета (в секундах) при различном размере блока				
	16	32	64	128	256
Блок генерации (Б.1)	4,4/2,9	2,2/2,9	2,4/2,9	2,7/2,9	3,0/2,9
Блок скоростей (Б.2)	241,0/162,1	125,2/161,0	87,5/164,3	85,5/160,6	89,2/162,7
Блок реструкт-и (Б.3)	189,3/54,8	108,9/54,5	70,0/50,0	64,2/50,0	68,3/50,6
Общее время расчета	434,7/219,8	236,3/218,4	159,9/217,2	152,4/213,5	160,5/216,2

В результате решения описанной выше задачи Блазиуса по моделированию обтекания пластинки получены следующие соотношения трудоемкостей основных операций (рис. 4), отражающее сделанные выше оценки.

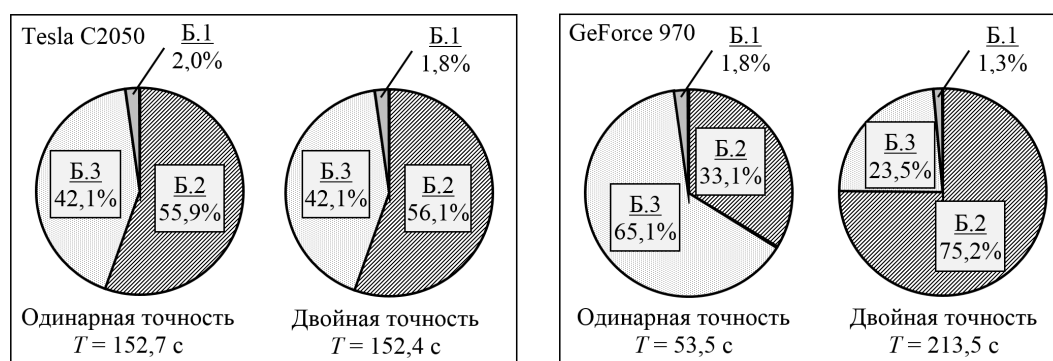


Рис. 4. Соотношение трудоемкостей укрупненных блоков при проведении расчета на GPU Tesla C2015 (слева) и GeForce GTX 970 (справа) с одинарной и двойной точностью

Затраты времени на выполнение одного шага расчета при $N \approx 50\,000$ в последовательном режиме на CPU, на 4-ядерном процессоре с использованием MPI, на 16-ядерном кластере и с использованием графических ускорителей приведены в таблице 4.

Таблица 4. Время проведения расчета

CPU	CPU	Кластер, 4 узла	Tesla C2050	GeForce 970	GeForce 970
(1 ядро)	(4 ядра)	(CPU, 16 ядер)	(double, float)	(double)	(float)
211,2 с	57,1 с	17,6 с	3,01 с	4,25 с	0,98 с

Как видно из представленных результатов, использование возможностей графического ускорителя Tesla C2050 позволяет сократить время решения задачи более, чем в 5 раз по сравнению с расчетом на 16-ядерном кластере и в 70 раз по сравнению с последовательным расчетом на CPU (характеристики CPU и кластера приведены в [7]). При этом расчеты с одинарной и двойной точностью занимают одинаковое время.

При проведении расчета на графическом ускорителе GeForce GTX 970 с двойной точностью производительность падает в 1,4 раза, несмотря на существенно большее количество CUDA-ядер. В то же время при расчетах с одинарной точностью производительность повышается втрое. Данный эффект объясняется тем, что ускорители Tesla оптимизированы для выполнения расчетов, а видеокарты GeForce относятся к «бытовым» и их производительность при двойной точности существенно (более, чем в 4 раза) снижается.

Отметим, что при проведении расчетов вихревыми методами одинарной точности в ряде случаев может быть недостаточно; при реализации отдельных операций требуется двойная точность, поэтому на практике оптимальным является «смешанный» алгоритм.

7. Выводы

Рассмотрен подход к ускорению вычислений в вихревых методах при моделировании обтекания профиля несжимаемой средой за счет применения технологии nVidia CUDA. Разработан алгоритм, позволяющий проводить все вычисления на графическом ускорителе. Получены и подтверждены оценки затрат оперативной памяти и трудоемкостей отдельных операций алгоритма. Достигнутые значения ускорения позволяют существенно расширить применение вихревых методов и повысить точность решения задач путем повышения количества вихревых элементов при сохранении сравнительно малого времени расчета.

Проведено решение тестовой задачи Блазиуса о моделировании обтекания тонкой пластинки; результаты расчетов хорошо согласуются с точным аналитическим решением и известными результатами вычислительных экспериментов.

Литература

1. Belotserkovskii S.M., Lifanov I.K. *Methods of Discrete Vortices*. Boca Raton: CRC Press, 1994. 454 p.
2. Лифанов И.К. *Метод сингулярных уравнений и численный эксперимент (в математической физике, аэродинамике, теории упругости и дифракции волн)*. М.: ТОО «Янус», 1995. 520 с.
3. Дынникова Г.Я. *Вихревые методы исследования нестационарных течений вязкой несжимаемой жидкости: Дис. ... докт. физ.-мат. наук. М., 2011. 269 с.*
4. Апарин В.А., Дворак А.В. *Метод дискретных вихрей с замкнутыми вихревыми рамками // Труды ВВИА им. проф. Н.Е. Жуковского, 1986. Вып. 1313. С. 424–432.*

5. Марчевский И.К., Щеглов Г.А. Модель симметричного вортон-отрезка для численного моделирования пространственных течений идеальной несжимаемой среды // Вестник МГТУ им. Н.Э. Баумана. Серия: Естественные науки. 2008. № 4. С. 62–71.
6. Дынникова Г.Я. Использование быстрого метода решения «задачи N тел» при вихревом моделировании течений // Журнал вычислительной математики и математической физики. 2009. Т. 49, № 8. С. 1458–1456.
7. Лукин В.В., Марчевский И.К., Морева В.С., Попов А.Ю., Шаповалов К.Л., Щеглов Г.А. Учебно-экспериментальный вычислительный кластер. Ч. 2. Примеры решения задач // Вестник МГТУ им. Н.Э. Баумана. Естественные науки. 2012. № 4. С. 82–102.
8. Марчевский И.К., Щеглов Г.А. Применение параллельных алгоритмов при решении задач гидродинамики методом вихревых элементов // Вычислительные технологии и программирование. 2010. Т. 11. С. 105–110.
9. Дынникова Г.Я., Сыроватский Д.А. Трехмерное бессеточное моделирование нестационарных течений жидкости на суперкомпьютерах гибридной архитектуры с графическими ускорителями // Ломоносовские чтения. Тезисы. М., 2012. С. 71–72.
10. Дынникова Г.Я., Сыроватский Д.А. Численное моделирование трехмерного нестационарного обтекания идеальной жидкостью тонких несущих поверхностей бессеточным методом дипольных доменов // Современные проблемы аэрогидродинамики: Тезисы докладов XVII школы-семинара. Сочи, 2014. С. 55.
11. Rees W.M., Rossinelli D., Hadjidoukas P., Koumoutsakos P. High performance CPU/GPU multiresolution Poisson solver // Adv. in Parallel Computing. 2014. V. 25. P. 481–490.
12. Андронов П.Р., Гувернюк С.В., Дынникова Г.Я. Вихревые методы расчета нестационарных гидродинамических нагрузок. М.: Изд-во МГУ, 2006. 184 с.
13. Li S., Liu W.K. Meshfree Particle Methods. Berlin: Springer-Verlag, 2007. 502 p.
14. Moreva V.S., Marchevsky I.K. Vortex element method for 2D flow simulation with tangent velocity components on airfoil surface // ECCOMAS 2012 — 6th European Congress on Computational Methods in Applied Sciences and Engineering: Book of proceedings. Vienna, 2012. 14 p.
15. Марчевский И.К., Морева В.С. Параллельный программный комплекс POLARA для моделирования обтекания профилей и исследования расчетных схем метода вихревых элементов // Параллельные вычислительные технологии: труды Международной научной конференции. Новосибирск, 2012. С. 236–247.
16. Макарова М.Е., Марчевский И.К., Морева В.С. Моделирование обтекания тонкой пластинки с использованием модифицированной схемы метода вихревых элементов // Наука и образование: электронное научно-техническое издание. 2013. № 9. С. 233–242.
17. Кузьмина К.С., Марчевский И.К. Оценка трудоемкости быстрого метода расчета вихревого влияния в методе вихревых элементов // Наука и образование: электронное научно-техническое издание. 2013. № 10. С. 399–414.
18. Кузьмина К.С., Марчевский И.К. Об ускорении вычислений при решении двумерных сопряженных задач гидроупругости вихревыми методами // Вестник Пермского Национального исследовательского политехнического университета. Аэрокосмическая техника. 2014. № 39. С. 145–163.
19. Yokota R., Barba L.A. Hierarchical N -body simulations with auto-tuning for heterogeneous systems // Computing in Science and Engineering. 2012. V. 14. P. 30–39.