

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

На правах рукописи

РАБА Никита Олегович

ПРОГРАММНЫЙ КОМПЛЕКС КОМПЬЮТЕРНОГО
ИССЛЕДОВАНИЯ АТМОСФЕРНЫХ ПРОЦЕССОВ
ДЛЯ МНОГОЯДЕРНЫХ ПРОЦЕССОРОВ

05.13.11 — Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Санкт-Петербург
2012

Работа выполнена на кафедре информатики математико-механического факультета Санкт-Петербургского государственного университета.

Научный руководитель: кандидат физико-математических наук,
доцент АМПИЛОВА Наталья Борисовна.

Официальные оппоненты: доктор физико-математических наук,
профессор ФЛЕГОНТОВ Александр Владимирович
(Российский государственный педагогический
университет им. А.И. Герцена),

доктор физико-математических наук,
профессор АНДРИАНОВ Сергей Николаевич
(Санкт-Петербургский государственный
университет)

Ведущая организация: Санкт-Петербургский государственный
политехнический университет.

Защита диссертации состоится 1 марта 2012 года в 13 часов 30 минут на заседании совета Д 212.232.51 по защите докторских и кандидатских диссертаций при Санкт-Петербургском государственном университете по адресу: 198504, Санкт-Петербург, Старый Петергоф, Университетский пр., 28, математико-механический факультет Санкт-Петербургского государственного университета, ауд. 405.

С диссертацией можно ознакомиться в Научной библиотеке им. М. Горького Санкт-Петербургского государственного университета по адресу: 199034, Санкт-Петербург, Университетская наб. 7/9.

Автореферат разослан «___» _____ 2012 года.

Ученый секретарь диссертационного
совета, доктор физико-математических
наук, доцент

Кривулин Н.К.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность задачи. Механизм переноса тепла и формирования осадков в конвективных облаках является интересной и важной проблемой, как с научной, так и с практической точек зрения, поскольку именно с развитием конвективных облаков связаны такие опасные природные явления как гроза, смерч, ливень, град. Прогноз возникновения и эволюции таких облаков, а также оценка воздействия на них (например, для борьбы с градом, предотвращения гроз, или, наоборот, для искусственного вызывания осадков в засушливых районах) являются важнейшими задачами, стоящими перед исследователями. Одним из наиболее эффективных инструментов изучения конвективных облаков и механизмов воздействия на них является компьютерное исследование.

Для целей оперативного прогноза опасных конвективных явлений, например в метеорологических центрах аэропортов, где для расчетов могут использоваться не суперкомпьютеры, а обычные ПК, необходимы такие представления, которые не требуют больших вычислительных ресурсов. Наилучшими, с этой точки зрения, являются так называемые полуторамерные представления (облако образуется в области, имеющей форму цилиндра, все параметры усредняются по сечению цилиндра, скорость потока имеет дополнительную горизонтальную составляющую через границу цилиндра), которые позволяют достаточно быстро делать достоверные прогнозы на краткосрочный период и исследовать активные воздействия на метеорологические процессы.

Полуторамерные представления, используемые в настоящий момент, обладают следующими недостатками: в развитии облака не учитывается стадия диссипации (рассеяния), применяется не спектральное, а упрощенное параметрическое описание микрофизических процессов, не учитывающее размеры частиц и содержащее только один тип ледяных частиц. Поэтому актуальной является задача разработки и реализации такого представления, которое устраняет эти недостатки, т.е. улучшает точность прогноза, и при этом не намного увеличивает время расчета. Решение этой задачи требует разработки и реализации более эффективных алгоритмов.

Одним из способов ускорения вычислений является их распараллеливание. Современные графические процессоры (GPU — Graphics

Processing Unit) имеют гораздо больше вычислительных ядер по сравнению с центральными процессорами (CPU — Central Processing Unit), поэтому еще один способ ускорения — перенос вычислений на GPU.

Чтобы полностью реализовать преимущество графических процессоров, нужны специальные алгоритмы. Они должны быть «существенно параллельными», т.е. их выполнение должно разбиваться на тысячи потоков, которые выполняют одни и те же действия над разными данными. Также есть дополнительные ограничения: действия должны быть достаточно простыми, должно быть как можно меньше ветвлений, синхронизаций потоков и обращений к глобальной памяти. Разработка и реализация таких алгоритмов позволяет в десятки раз увеличить скорость прогноза.

Целью работы является разработка и реализация программного комплекса, позволяющего эффективно и точно исследовать развитие конвективного облака.

Основные задачи:

- разработать и реализовать программный комплекс для компьютерного исследования развития конвективного облака на основе расширенного представления;
- разработать квадратичные по времени алгоритмы для расчета характеристик облака;
- разработать модификации этих алгоритмов для многоядерных процессоров;
- разработать модификации этих алгоритмов для графических процессоров.

Основные результаты:

- разработан и реализован программный комплекс, который позволяет эффективно и точно исследовать как развитие конвективного облака, так и активные воздействия на него;
- разработаны новые квадратичные по времени алгоритмы расчета коагуляции для представлений со спектральной жидкой и смешанной микрофизикой;
- разработаны методы распараллеливания вычислений между ядрами процессора: распараллеливание по экспериментам и по пространству с использованием потоков; приведен способ балансировки нагрузки; для 4-х ядерного процессора распараллеливание позволило сократить время расчета в 2.7 раза;

- разработан многопоточный алгоритм расчета коагуляции для GPU с использованием технологии CUDA, представлен способ разбиения вычислений на несколько этапов, приведен и проанализирован способ распределения расчетов по потокам, описан метод предотвращения коллизий записи при обращении к памяти несколькими потоками одновременно, показаны способы оптимизации доступа к памяти; разработанный алгоритм позволил сократить время расчета в 15–20 раз.

Научная новизна:

- разработаны более совершенные комбинированные методы представления конвективного облака, включающие учет жидких и твердых частиц различных размеров, основанные на так называемых полуторамерных моделях;
- разработан алгоритм расчета столкновения и слияния частиц нескольких типов (коагуляции) для графических процессоров, основанный на специальном представлении данных.

Практическая значимость

Разработка алгоритма расчета коагуляции для GPU позволила сократить время вычислений в 15–20 раз и обеспечить оперативность расчетов. Этот алгоритм является универсальным и допускает применение в представлениях размерности 2 и 3.

Разработанный программный комплекс предназначен для оперативного прогноза эволюции конвективных облаков и возникновения опасных конвективных явлений, а также для планирования мер активного воздействия на облака (определения времени и места наиболее эффективного воздействия).

Апробация работы. Результаты работы докладывались:

- на научной конференции институтов Росгидромета, посвящённой 50-летию отдела физики облаков Главной геофизической обсерватории им. А.И. Воейкова «Теоретические и экспериментальные исследования конвективных облаков» (Санкт-Петербург, 18–20 ноября, 2008);
- на 5-ой Международной конференции «Dynamical Systems and Applications» (Констанца, Румыния, 15–18 июня, 2009);
- на International Conference on Computational Science, ICCS 2010 (Амстердам, Нидерланды, 31 мая – 2 июня, 2010);
- на International Conference on Computational Science and Its Applications, ICCSA 2010 (Фукуока, Япония, 23–26 марта, 2010);

- на 9-ой Международной научно-практической конференции «Исследование, разработка и применение высоких технологий в промышленности» (Санкт-Петербург, 22–23 апреля, 2010);
- на International Conference on Computational Science and Its Applications, ICCSA 2011 (Сантандер, Испания, 20–23 июня, 2011);
- на 64-ой научной конференции «Герценовские чтения» (Санкт-Петербург, 11–16 апреля, 2011).

Публикации. Основные результаты, полученные при выполнении диссертационной работы, опубликованы в 11 работах автора, перечисленных в прилагающемся списке. Работы [1–3] опубликованы в изданиях по перечню ВАК. В работах [3–10] автору принадлежит разработка и реализация представлений облака, алгоритмов распараллеливания на ядра процессора, тестирование, а соавторам — общие постановки задач.

Структура и объем работы. Диссертационная работа объемом 160 машинописных страниц состоит из введения, пяти глав, четырех приложений и списка литературы, включающего 64 наименования, и содержит 42 рисунка и 24 таблицы.

СОДЕРЖАНИЕ РАБОТЫ

Введение содержит обзор современного состояния предметной области и обоснование актуальности диссертационной работы. Сформулирована цель работы и представлены выносимые на защиту положения. Также рассмотрена структура и краткое содержание диссертации по главам.

В **главе 1** приведено описание представлений конвективного облака, разработанных автором, основанных на полуторамерном представлении [12], подробно описанном в приложении Б.

Традиционное представление (полуторамерная модель), не требует больших вычислительных ресурсов, однако способно воспроизводить те характеристики облака, которые могут стать предикторами для существующих методов прогноза опасных конвективных явлений. Представления с большей размерностью нуждаются в значительно больших (на несколько порядков) вычислительных ресурсах по сравнению с полуторамерными.

При достаточной простоте, традиционное представление обладает следующими недостатками, ограничивающими область его применения: отсутствует компенсирующий нисходящий поток (поэтому в развитии облака не воспроизводится стадия диссипации), используется сильно упрощенный параметрический (а не спектральный) микрофизический блок (что снижает достоверность прогнозов).

Автором были разработаны и реализованы усовершенствованные представления, основанные на традиционном, в которых область конвективных течений представляется двумя концентрическими цилиндрами, благодаря чему учитывается нисходящий поток и моделируется стадия диссипации.

Представление 1 содержит такой же микрофизический блок, как и традиционное представление, т.е. исправляет только первый недостаток.

В **представлении 2** используется спектральное описание микрофизических процессов для жидкой фазы, что увеличивает точность компьютерного исследования. В каждом узле сетки хранится информация о распределениях по размерам (спектрах) капель и частиц аэрозоля. Весь диапазон масс капель делится на N интервалов. Каждый интервал характеризуется средней массой капли в нем (для i -го интервала масса капли m_i). Спектр представляется в виде набора значений f_i , i от 1 до N , где f_i — концентрация капель в i -ом интервале в единице объема. Чем больше размер капли, тем больше скорость падения. Маленькие капли, скорость падения которых почти равна нулю, образуют облако, крупные капли выпадают в виде осадков. Учитываются следующие микрофизические процессы: нуклеация (образование новых капель на частицах аэрозоля), конденсация, испарение, коагуляция (столкновение и слияние частиц друг с другом). Возникновение осадков связано главным образом с процессом коагуляции, т.к. благодаря коагуляции возникают капли таких размеров, при которых они способны падать со скоростью, большей, чем скорость восходящих потоков.

Представление 3 добавляет к представлению 2 спектральную микрофизику, включающую и твердую фазу. Дополнительно учитываются градины, снежинки, крупа и 3 типа ледяных кристаллов. Частицы разных типов имеют разные устоявшиеся скорости падения относительно потока. Микрофизический блок включает следующие процессы: нуклеацию, замерзание капелек, таяние ледяных частиц, конденсацию на поверхности частиц, испарение с поверхности частиц, коагуляцию. При столкновении и слиянии

двух частиц друг с другом образуется частица, тип которой зависит от типов столкнувшихся частиц, их масс, температуры. Представление 3 является наиболее совершенным, в нем отсутствуют все недостатки, присущие традиционному представлению, но оно же является самым сложным с точки зрения вычислений.

Сравнение разработанных представлений с традиционным и проверка их адекватности приведены в приложении В.

Глава 2 посвящена разработке и реализации алгоритмов расчета коагуляции. В представлениях со спектральной микрофизикой основная часть компьютерных ресурсов тратится на вычисление микрофизических процессов. Расчет коагуляции (особенно в представлениях со смешанной фазой, т.е. с жидкой и твердой) занимает в десятки раз больше времени по сравнению с расчетом других процессов, поэтому оптимизация расчета коагуляции требуется в первую очередь.

Уравнение для изменения концентрации f капель массы m (стохастическое уравнение коагуляции) имеет вид:

$$\partial f(m) / \partial t = \frac{1}{2} \int_0^m K(m-m', m') f(m-m') f(m') dm' - f(m) \int_0^\infty K(m, m') f(m') dm',$$

где $K(m, m')$ — ядро коагуляции (вероятность столкновения и слияния капель с массами m и m').

Формула перераспределения капель, вызванного коагуляцией, для спектра в виде набора f_i была предложена Коветцом и Олундом [13]:

$$\partial f_i / \partial t = \sum_{j=1}^{i-1} \sum_{k=j+1}^i B_{jki} K_{jk} f_j f_k - f_i \sum_{j=1}^N K_{ij} f_j,$$

где $K_{jk} = K(m_j, m_k)$,

$$B_{jki} = \begin{cases} (r_j^3 + r_k^3 - r_{i-1}^3) / (r_i^3 - r_{i-1}^3), & \text{если } r_{i-1}^3 < r_j^3 + r_k^3 < r_i^3, \\ (r_{i+1}^3 - r_j^3 - r_k^3) / (r_{i+1}^3 - r_i^3), & \text{если } r_i^3 \leq r_j^3 + r_k^3 < r_{i+1}^3, \\ 0, & \text{в остальных случаях,} \end{cases}$$

r_i — радиус капель массы m_i . Для расчета всех новых значений f_i , $i = 1, \dots, N$ такой алгоритм требует $O(N^3)$ операций.

Автором был разработан следующий алгоритм. Вместо трехмерного массива B_{jki} заполняется матрица весовых коэффициентов A_{jk} и матрица индексов I_{jk} : $A_{jk} = (m_{i+1} - m_j - m_k) / (m_{i+1} - m_i) = B_{jki}$, $I_{jk} = i$, если $m_i \leq m_j + m_k < m_{i+1}$ (если такого i не находится, то $I_{jk} = 0$). Заметим, что $B_{jki} = A_{jk}$, если $i = I_{jk}$, $B_{jki} = 1 - A_{jk}$, если $i = I_{jk} + 1$, и $B_{jki} = 0$ в остальных случаях.

За счет столкновения капель из интервалов j и k концентрация капель в интервалах j и k уменьшится, в интервалах i и $i + 1$ увеличится ($m_i \leq m_j + m_k < m_{i+1}$, т.е. $i = I_{jk}$; капли, образовавшиеся при слиянии перераспределяются между интервалами i и $i + 1$ с весовыми коэффициентами A_{jk} и $1 - A_{jk}$):

$$\begin{aligned}\partial f_j / \partial t|_{jk} &= -K_{jk} f_j f_k, \quad \partial f_k / \partial t|_{jk} = -K_{jk} f_j f_k, \\ \partial f_i / \partial t|_{jk} &= K_{jk} f_j f_k A_{jk}, \quad \partial f_{i+1} / \partial t|_{jk} = K_{jk} f_j f_k (1 - A_{jk}), \\ \partial f_x / \partial t|_{jk} &= 0, \text{ для } x, \text{ не равных } j, k, i \text{ и } i + 1,\end{aligned}$$

где $\partial f_y / \partial t|_{jk}$ — изменение концентрации капель в интервале y за счет столкновения капель из интервалов j и k . Общая масса капель при столкновении не меняется.

Изменение концентрации капель в интервале y после коагуляции:

$$\partial f_y / \partial t = \sum_{j=1}^{N-1} \sum_{k=j+1}^N \partial f_y / \partial t|_{jk}.$$

Расчет изменения концентрации во всех интервалах после коагуляции (используются дополнительные массивы S^- и S^+ размерности N):

1. Инициализировать S^- и S^+ : $S^-_i := 0$, $S^+_i := 0$, для i от 1 до N .
2. Перебрать все пары (j, k) интервалов сталкивающихся частиц, $j < k$, и для каждой пары, для которой $I_{jk} \neq 0$, выполнить: $i := I_{jk}$, $F := K_{jk} f_j f_k$, $S^-_j := S^-_j + F$, $S^-_k := S^-_k + F$, $S^+_i := S^+_i + F A_{jk}$, $S^+_{i+1} := S^+_{i+1} + F (1 - A_{jk})$.
3. Вычислить $\partial f_i / \partial t := S^+_i - S^-_i$, для i от 1 до N .

В представлении со смешанной фазой алгоритм коагуляции намного более сложный, т.к. сталкиваться могут частицы разных типов, а тип получающейся частицы зависит как от типов и масс сталкивающихся частиц, так и от окружающей температуры. Уравнение для изменения концентрации f_i частиц i -го типа массы m :

$$\begin{aligned}\partial f_i(m) / \partial t &= \sum_{j=1}^{N_{PT}} \sum_{k=1}^{N_{PT}} C_{ijk}(T) \int_0^{m/2} K_{jk}(m - m', m') f_j(m - m') f_k(m') dm' - \\ &- f_i(m) \sum_{j=1}^{N_{PT}} \int_0^\infty K_{ij}(m, m') f_j(m') dm',\end{aligned}$$

где N_{PT} — количество типов частиц, $C_{ijk}(T)$ — коэффициент, который равен 1, если при слиянии частицы j -го типа с меньшей по массе частицей k -го типа при температуре T образуется частица i -го типа, и равен 0 в противном случае, $K_{ij}(m, m')$ — ядро коагуляции (вероятность столкновения и слияния частицы i -го типа с массой m и частицы j -го типа с массой m').

Автором был разработан следующий алгоритм расчета изменения концентрации частиц всех типов после коагуляции (дополнительные массивы

S^- и S^+ размерности $N_{PT} \times N$; f_{ni} — концентрация частиц типа n в интервале i ; интервалы для частиц всех типов берутся одинаковыми):

1. Инициализировать S^- и S^+ : $S^-_{ni} := 0$, $S^+_{ni} := 0$, для n от 1 до N_{PT} , i от 1 до N .
2. Перебрать все пары (j, k) интервалов сталкивающихся частиц, и для каждой пары, для которой $I_{jk} \neq 0$, перебрать типы сталкивающихся частиц m (тип частиц из интервала j), n (тип частиц из интервала k), $m \leq n$. Для каждой четверки (m, n, j, k) , для которой либо $j < k$, либо $m \neq n$, определить p — тип частицы, образующийся при слиянии (используется рассчитанная заранее матрица переходов), и выполнить: $i := I_{jk}$, $F := K_{mnjk} f_{mj} f_{nk}$, $S^-_{mj} := S^-_{mj} + F$, $S^-_{nk} := S^-_{nk} + F$, $S^+_{pi} := S^+_{pi} + F A_{jk}$, $S^+_{p i+1} := S^+_{p i+1} + F (1 - A_{jk})$.
3. Вычислить $\partial f_{ni} / \partial t := S^+_{ni} - S^-_{ni}$ для n от 1 до N_{PT} , i от 1 до N .

Такой алгоритм универсален: он может быть применен к представлениям с различным количеством учитываемых частиц (N_{PT}), а также использоваться в представлениях с большей размерностью (а не только в полуторамерных).

Доказано следующее

Утверждение. Сложность разработанного алгоритма расчета коагуляции для представления с жидкой микрофизикой $O(N^2)$ (т.е. меньше, чем алгоритм Коветца-Олунда), для представления со смешанной микрофизикой $O(N^2_{PT} \cdot N^2)$.

В главе 3 приведены методы распараллеливания для разработанных представлений.

Так как большинство современных ПК имеют многоядерные процессоры, допускающие параллельное выполнение расчетов, то одним из способов ускорения вычислений является их распараллеливание. Рассматриваются основные стратегии распараллеливания (по пространству, по экспериментам), их достоинства и недостатки, возможность их использования в приведенных представлениях.

В разработанных представлениях параметры в текущий момент времени зависят только от параметров на предыдущем шаге и не зависят от параметров в соседних узлах сетки на текущем шаге, поэтому можно рассчитывать параметры в разных узлах сетки на разных ядрах процессора (распараллеливание по пространству). Параллельные вычисления реализуются с помощью потоков. Создаются потоки, которым передаются ссылки на данные, рассчитанные на предыдущем шаге. Каждый поток выполняет вычисления в определенных узлах сетки. Переход к следующему шагу происходит, когда все потоки выполняются.

Запуск потоков, а также переключение между потоками, выполняемыми на одном ядре, требует некоторого времени, поэтому следует уменьшить их количество. Для K -ядерного процессора, ядра которого могут оптимизировать выполнение T потоков ($T = 2$ для современных процессоров с поддержкой технологии Hyper-Threading), оптимально запускать $T \cdot K$ потоков.

Для наилучшей загрузки ядер процессора необходимо примерно равное время выполнения потоков (балансировка нагрузки). Расчет параметров в различных узлах сетки требует разного времени (например, в тех узлах, в которых нет частиц и влажность меньше 100%, расчет микрофизических процессов проводить не нужно). Так как спектры частиц и влажность меняются незначительно от узла к узлу, то расчет параметров в соседних узлах сетки занимает примерно одинаковое время. Поэтому для балансировки нагрузки можно распределять узлы сетки по $T \cdot K$ потокам так, чтобы соседние узлы рассчитывались в разных потоках. В первом потоке рассчитываются все динамические и микрофизические процессы в узлах с номерами $1 + i \cdot T \cdot K$ ($i = 0, 1, 2, \dots$), во втором — с номерами $2 + i \cdot T \cdot K, \dots$, в $T \cdot K$ — с номерами $T \cdot K + i \cdot T \cdot K$.

В главе 4 проанализированы современные технологии для переноса вычислений с центрального процессора на видеокарту. Приведены алгоритм расчета коагуляции на видеокарте (для представления со смешанной микрофизикой) и его оптимизации, дана оценка сложности.

Современные графические процессоры (GPU) имеют гораздо больше вычислительных ядер (несколько сотен) по сравнению с CPU (2–8). Чтобы полностью реализовать преимущество GPU, нужны специальные «существенно параллельные» алгоритмы (их выполнение должно разбиваться на тысячи потоков, которые выполняют одни и те же действия над разными данными). Также есть дополнительные ограничения: действия должны быть достаточно простыми, должно быть как можно меньше ветвлений, синхронизаций потоков и обращений к глобальной памяти.

В настоящее время существует несколько технологий для использования графических процессоров для вычислений общего назначения (GPGPU — General-Purpose computing on Graphics Processing Units): CUDA, ATI Stream, OpenCL, DirectCompute. Одной из самых популярных технологий является CUDA фирмы NVidia. Она позволяет наиболее полно использовать возможности GPU, а в ее основе лежит язык C/C++ с некоторыми расширениями и ограничениями, что упрощает написание программ.

Программа на CUDA задействует и CPU и GPU. На CPU выполняются последовательные вычисления, а на GPU массивно-параллельные. Эти параллельные вычисления представляются в виде одновременно выполняющихся потоков.

Представлен следующий способ вычисления. Каждый поток обрабатывает столкновение частицы типа m из массовой категории j с частицами разных типов ($n \geq m$) из массовой категории k , т.е. один поток обрабатывает максимум N_{PT} столкновений (или меньше, например, когда $k \leq j$). Значения j , m и k зависят от индексов потока $threadIdx$ и блока $blockIdx$ (в CUDA потоки объединяются в блоки, блоки в сетки, потоки и блоки имеют свои многомерные индексы, потоки внутри одного блока могут синхронизироваться и имеют доступ к быстрой разделяемой памяти).

Одновременно несколько потоков могут пытаться изменить значения в одной и той же ячейке памяти (например, в одной и той же ячейке массива S^+), при этом неизвестно, какое значение там будет в результате.

Для предотвращения таких коллизий записи был разработан следующий алгоритм. Для каждой пары интервалов (j, k) сталкивающихся частиц типов m и n , создаются 4 отдельно выделенные ячейки памяти: для уменьшения концентрации частиц типа m в интервале j и типа n в интервале k , и для увеличения концентрации частиц типа p в интервалах i и $i + 1$ ($i = I_{jk}$). Для этого используется «развернутый» массив MP (вместо массивов S^- и S^+). Для каждой четверки (m, n, j, k) хранятся 4 соответствующих индекса для этого массива: ind_{0mnjk} , ind_{1mnjk} , ind_{2mnjk} , ind_{3mnjk} . Индексы рассчитываются таким образом, чтобы ячейки памяти, относящиеся к увеличению (или уменьшению) концентрации частиц одного типа находились рядом друг с другом (т.е. образовывали последовательные области в массиве MP). Более того, ячейки памяти, относящиеся к увеличению (или уменьшению) концентрации частиц одного типа в одном массовом интервале тоже должны находиться рядом друг с другом. Размеры таких областей (количество ячеек) должны быть кратны RBS (значение этого параметра объяснено далее). Некоторые ячейки могут остаться незаполненными, для них может не быть соответствующего индекса ind . Индексы концов этих последовательных областей: pos_{0ni} — индекс последнего элемента в массиве MP , относящегося к уменьшению концентрации частиц типа n в интервале i , pos_{1ni} — индекс последнего элемента в массиве MP ,

относящегося к увеличению концентрации частиц типа n в интервале i . Приводится способ вычисления индексов ind и pos и размерности массива MP .

Расчет изменения концентрации происходит в 3 этапа. На первом каждый поток для каждого столкновения (соответствующего ему) рассчитывает $F = K_{mnjk} f_{mj} f_{nk}$ и заполняет четыре ячейки массива MP : $MP[ind_{0mnjk}] = -F$, $MP[ind_{1mnjk}] = -F$, $MP[ind_{2mnjk}] = F A_{jk}$, $MP[ind_{3mnjk}] = F (1 - A_{jk})$. После расчета всех столкновений, массив MP оказывается полностью заполненным (он изначально заполняется 0; ячейки, для которых нет индекса ind , остаются 0). На втором и третьем этапах происходит суммирование отдельных областей массива, относящихся к одним и тем же интервалам (например, элементы массива MP с индексами от $pos_{0ni-1} + 1$ до pos_{0ni} находятся в области, относящейся к уменьшению концентрации частиц типа n в интервале i). На втором этапе суммируются элементы блоков массива MP размерностью RBS . Это производится методом параллельной редукции, поэтому параметр RBS (reduction block size) изначально выбирается равным степени 2. Результат суммирования сохраняется в массиве MP в ячейках с индексами, кратными RBS . На третьем этапе рассчитывается окончательная сумма по областям (в каждой области суммируются только элементы, индексы которых кратны RBS , каждый поток производит суммирование в своей области). Пусть M — результат суммирования для области от $pos_{0ni-1} + 1$ до pos_{0ni} , P — для области от $pos_{1ni-1} + 1$ до pos_{1ni} , тогда $\partial f_{ni}/\partial t = M + P$.

Доказано следующее

Утверждение. Количество обращений к памяти у приведенного алгоритма для GPU не более чем в 1.5 раза превышает количество обращений к памяти у алгоритма для CPU.

Производительность данных алгоритмов ограничивается в основном пропускной способностью шины памяти. Так как пропускная способность шины памяти современных видеокарт в десятки раз больше, чем у оперативной памяти компьютера, то приведенный алгоритм позволяет существенно (см. главу 5) ускорить выполнение расчета.

В главе 5 показаны результаты сравнения быстродействия различных алгоритмов вычисления коагуляции. В следующей таблице приведено время расчета коагуляции в зависимости от количества интервалов N (все остальные параметры одни и те же) на CPU (Intel Core 2 Duo E6400) и на GPU (NVidia GTX470), а также ускорение вычислений на GPU:

N	64	128	256	512
CPU (сек.)	3.34	13.72	55.48	225.02
GPU (сек.)	0.203	0.688	2.81	11.53
Ускорение (раз)	16.5	19.9	19.7	19.5

В **заключении** приведены основные результаты, полученные в работе.

Приложение А содержит классификацию представлений конвективных облаков. Представления облаков можно классифицировать: по размерности пространства (одномерные, полуторамерные, двухмерные и трехмерные), по способу представления спектра частиц (с параметрической и со спектральной микрофизикой), по учету ледяных частиц (с жидкой и со смешанной фазой). В представлениях с параметрической микрофизикой переходы между паром, облачными каплями, дождевыми каплями и градинами задаются параметрически и никак не зависят от реальных размеров частиц. Представления со спектральной микрофизикой учитывают размеры частиц и наиболее полно отражают природные процессы. В представлениях с жидкой фазой ледяные частицы (твердое состояние) не рассматриваются, в отличие от представлений со смешанной фазой.

В **приложении Б** описывается традиционное полуторамерное представление с параметрической микрофизикой, впервые предложенное в [12]. В данном представлении предполагается, что облако образуется в области, имеющей форму цилиндра. Все параметры (скорость, температура, давление, плотность, содержание пара, облачных капель, дождевых капель, града) усредняются по сечению цилиндра и меняются только по вертикали. Вокруг цилиндра находится неподвижная атмосфера с постоянными параметрами, с которой происходит взаимодействие через боковую поверхность цилиндра.

В **приложении В** проведено сравнение поведения облака в представлениях с одним и двумя цилиндрами. Приведенные результаты показывают присутствие фазы диссипации в жизненном цикле облака в представлении с двумя цилиндрами.

Приложение Г содержит описание интерфейса программного комплекса и входных данных.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в журналах, рекомендованных ВАК:

[1] Раба Н.О. Оптимизация алгоритмов расчета физических процессов в модели облака с детальным описанием микрофизики // Вест. СПбГУ. Сер. 10, Прикладная математика. Информатика. Процессы управления.– 2010.– Вып. 3.– С. 121–126.

[2] Раба Н.О. Разработка и реализация алгоритма расчета коагуляции в модели облаков со смешанной фазой с использованием технологии CUDA // Вестн. СПбГУ. Сер. 10, Прикладная математика. Информатика. Процессы управления. – 2011. – Вып. 4. – С. 94–104.

[3] Раба Н.О., Станкова Е.Н. Исследование влияния компенсирующего нисходящего потока на жизненный цикл конвективного облака с помощью численной полутримерной модели с двумя цилиндрами // Труды Главной геофизической обсерватории им. А.И.Воейкова.– 2009.– Вып. 559.– С. 192–209.

Другие публикации:

[4] Станкова Е.Н., Раба Н.О., Ампилова Н.Б. Численное моделирование микрофизических процессов в смешанных конвективных облаках. Сравнение результатов расчётов с данными натуральных экспериментов // Научная конференция институтов Росгидромета «Теоретические и экспериментальные исследования конвективных облаков». – СПб., 2008. – С. 90–91.

[5] Станкова Е.Н., Раба Н.О., Ампилова Н.Б. О влиянии алгоритма определения высоты нижней границы облака на прогнозирование конвективных явлений с помощью численных моделей // Научная конференция институтов Росгидромета «Теоретические и экспериментальные исследования конвективных облаков». – СПб., 2008. – С. 43–44.

[6] Raba N., Stankova E., Ampilova N. One-and-a-half-dimensional Model of Cumulus Cloud with Two Cylinders. Research of Influence of Compensating Descending Flow on Development of Cloud // Ovidius University Annals. Series: Civil Engineering. – Constantza: Ovidius University Press, 2009. – Vol. 1, Special Issue 11. – P. 93–101.

[7] Раба Н.О., Станкова Е.Н., Ампилова Н.Б. Распараллеливание вычислений при моделировании конвективных облаков // Высокие технологии, исследования, промышленность. Сборник трудов 9-ой международной научно-практической конференции «Исследование, разработка и применение высоких

технологий в промышленности». – СПб.: Изд-во Политехн. ун-та, 2010. – Т. 1. – С. 213–216.

[8] *Raba N., Stankova E., Ampilova N.* On Investigation of Parallelization Effectiveness with the Help of Multi-core Processors // *Procedia Computer Science*. – 2010. – Vol. 1, Issue 1. – P. 2757–2762.

[9] *Raba N., Stankova E.* On the Possibilities of Multi-Core Processor Use for Real-Time Forecast of Dangerous Convective Phenomena // *Lecture Notes in Computer Science. Computational Science and Its Applications – ICCSA 2010*. – 2010. – Vol. 6017. – P. 130–138.

[10] *N. Raba, E. Stankova* On the Problem of Numerical Modeling of Dangerous Convective Phenomena: Possibilities of Real-Time Forecast with the Help of Multi-core Processors // *Lecture Notes in Computer Science. Computational Science and Its Applications – ICCSA 2011*. – 2011. – Vol. 6786. – P. 633–642.

[11] *Раба Н.О.* Использование технологии CUDA для расчета коагуляции в модели облаков // *Некоторые актуальные проблемы современной математики и математического образования. Герценовские чтения – 2011*. – СПб., 2011. – С. 198–205.

Цитированная литература

[12] *Shiino J.* A Numerical Study of Precipitation Development in Cumulus Clouds // *Papers in Meteorology and Geophysics*. – 1978. – Vol. 29, № 4. – P. 157–194.

[13] *Kovetz A., Olund B.* The Effect of Coalescence and Condensation on Rain Formation in a Cloud of Finite Vertical Extent // *J. Atm. Sci.* – 1969. – Vol. 26. – P. 1060–1065.