

ПАРАЛЛЕЛЬНАЯ ВЕРСИЯ ПОЛУЛАГРАНЖЕВОЙ МОДЕЛИ ПЛАВ С ГОРИЗОНТАЛЬНЫМ РАЗРЕШЕНИЕМ ПОРЯДКА 20 КМ

М.А.Толстых^{1,2}, В.Г.Мизяк²

¹ *Институт вычислительной математики РАН*

² *Гидрометеорологический научно-исследовательский
центр Российской Федерации
tolstykh@metcom.ru, vmizyak@metcom.ru*

Введение

Известно, что модели атмосферы с грубым пространственным разрешением неточно, по сравнению с наблюдениями, описывают статистику синоптических процессов в тропиках и средних широтах. Более того, такие модели могут включать лишь параметрическое описание крайне важного процесса глубокой конвекции. Перспективные модели атмосферы, свободные от этих недостатков, должны будут иметь горизонтальное разрешение порядка 1–10 км. Глобальные модели прогноза погоды в ведущих мировых центрах имеют горизонтальное разрешение порядка 20 км [3]. Кроме того, оперативный прогноз налагает ограничение на допустимое время счета модели (как правило, не более 20 мин астрономического времени для прогноза на 24 часа). Таким образом, требуется эффективное применение параллельных вычислений на сотнях процессоров для ускорения расчетов.

1. Параллельная реализация модели ПЛАВ

Глобальная полулагранжева модель ПЛАВ [2] в 2009 году принята в качестве основного численного метода среднесрочного прогноза погоды. В настоящее время разрешение оперативной версии модели

составляет $0,9^\circ$ по долготе и $0,72^\circ$ по широте, 28 уровней по вертикали. По современным меркам данное разрешение является весьма грубым, поэтому коллективом разработчиков модели была реализована новая версия модели, имеющая горизонтальное разрешение порядка 20–25 км, а вертикальное – 51 уровень. Оказалось, что из-за ограниченной масштабируемости существующей параллельной реализации, время расчета прогноза по существовавшей на конец 2010 года параллельной реализации модели не удовлетворяет упомянутым выше требованиям оперативности. Имевшиеся результаты параллельного ускорения модели представлены на рис. 1. Видно, что модель могла эффективно использовать лишь около 200 процессоров.



Рис. 1. Параллельное ускорение экспериментальной версии модели (2010 год) с разрешением $0,225 \times 0,18^\circ$, 28 уровней по вертикали.

Имевшаяся параллельная реализация полулагранжевой модели ПЛАВ на основе технологии MPI была подробно представлена в [5], дальнейшее развитие программной реализации модели на основе сочетания технологий MPI и OpenMP (гибридной технологии) описывается в [1]. Отметим, что гибридный подход был впервые в России применен к реальному сложному программному комплексу.

Приведем краткое описание параллельной реализации модели (подробнее см. [1, 5]). Рассмотрим вычислительную структуру модели. Основная часть вычислений производится в сеточном пространстве, расчетная область разбивается по широте (рис. 2 слева).

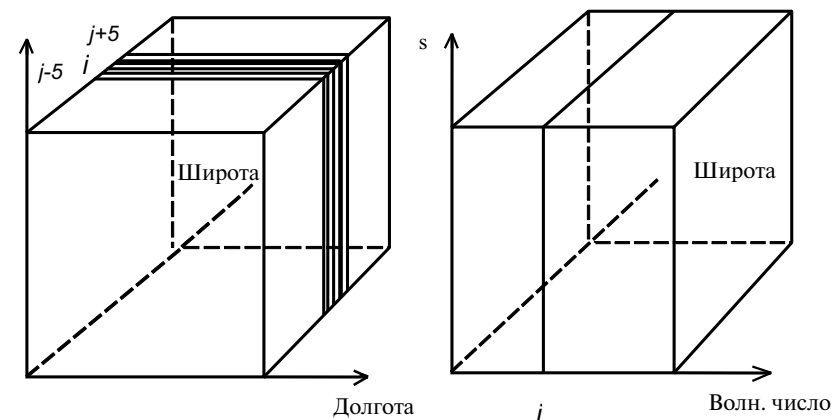


Рис. 2. Организация вычислений в сеточном пространстве (слева) и пространстве коэффициентов Фурье по долготе (справа).

Сначала вычисляются правые части прогностических уравнений (подробное описание дано в [2]), используя значения переменных на текущем шаге по времени. Эти правые части содержат тенденции прогностических переменных вследствие процессов подсеточного масштаба (солнечное излучение, глубокая конвекция и т.д.). Здесь мы имеем высокую степень параллелизма, так как эти вычисления для данной точки сетки не требуют значений переменных в смежных точках сетки в долготном и широтном направлениях, требуются только значения в вертикальных столбцах. Следует заметить, что эти вычисления занимают 61–70 % всего времени расчета. Затем вычисляются разностные аналоги дифференциальных операторов с помощью формул четвертого порядка. Эти вычисления для данной точки сетки требуют значения переменных из области. Процессоры обмениваются необходимыми значениями на границах с соседями перед тем, как эти вычисления начинаются. Там, где возможно, используются

асинхронные обмены, что позволяет одновременно выполнять вычисления и обмены данными.

Следующий этап вычислений — поиск исходных точек лагранжевых траекторий и интерполяция вычисленных слагаемых уравнений в эти точки. Эта часть вычислений занимает около 30 % машинного времени. Шаблон интерполяции составляет 4 точки по каждой координате. Исходная точка траектории может лежать за несколько интервалов сетки от конечной точки (которая всегда является точкой сетки в полулагранжевом методе). Максимальная ширина области зависимости вычислений для данной точки сетки зависит от размера шаблона, значения шага по времени и максимальной скорости ветра в атмосфере (около 110 м/с). Типичная половинная ширина области зависимости в нашей модели для горизонтального разрешения $0,7^\circ$ составляет 7 точек в каждом горизонтальном направлении. Таким образом, количество данных, которое передается между процессорами, существенно больше, чем для эйлеровых моделей.

После того, как интерполяции и прямые быстрые преобразования Фурье по долготе заканчиваются, нам необходимо перераспределить данные для вычислений в пространстве коэффициентов Фурье по долготе. Вычисления для каждого коэффициента требуют значения всех широт (рис. 2 справа), так что нам нужно обеспечить каждый процессор соответствующими данными, используя транспонирование данных. Вычисления в пространстве Фурье включают вычисление правой части и решение дискретного уравнения типа Гельмгольца, неявной горизонтальной диффузии четвертого порядка и вычисления горизонтальных компонент вектора скорости из дивергенции и завихренности. Эти вычисления выполняются параллельно для каждого коэффициента Фурье. После вычислений в пространстве Фурье нам нужно снова перераспределить данные между процессорами, так как для выполнения обратного преобразования Фурье по долготе нужны все коэффициенты.

Весь ввод-вывод выполняется нулевым процессом MPI, который затем рассылает необходимые данные остальным процессам. Примерно так же организована предварительная обработка результатов модели (интерполяция по вертикали с модельных сигма-уровней на стандартные изобарические поверхности).

При горизонтальном разрешении $0,18^\circ$ по широте модель теоретически может использовать до 1000 процессоров. Однако это будет требовать большого количества обменов данными между процессорами, поскольку необходимые данные будут размещены не только в соседних процессорах, но также в следующих за ними процессорах. Практически с помощью технологии MPI для упомянутого разрешения можно использовать до 125–150 процессоров (в зависимости от величины шага по времени модели).

В дополнение к распараллеливанию по технологии MPI было выполнено распараллеливание на основе OpenMP [1]. При этом с помощью OpenMP распараллеливались циклы по широте, т.е. те же, что и распараллеленные с помощью MPI. Тем самым заметно снижаются накладные расходы на организацию и синхронизацию.

Для повышения параллельной эффективности программного комплекса модели ПЛАВ было выполнено распараллеливание алгоритма постпроцессинга модели (расчета полей на стандартных изобарических поверхностях). Кроме того, все транспозиции данных были реализованы с помощью библиотеки подпрограмм SHMEM, имеющейся на вычислительных системах SGI, Cray и некоторых других. Эта библиотека реализует идеологию односторонних коммуникаций, которая также реализована в стандарте MPI-2. Кроме того, были выполнены оптимизации обменов, позволившие сократить объем пересылаемой информации.

2. Результаты расчетов

Были выполнены тестовые расчеты на вычислительной системе SGI Altix 4700, имеющей 1664 ядра процессоров Intel Itanium 9100, организованные в узлы по 128 ядер. Эта вычислительная система установлена в ГВЦ Росгидромета. Тестовые расчеты выполнялись для варианта модели с разрешением $0,225^\circ$ по долготе, $0,17-0,24^\circ$ по широте (примерно 20 км в средних широтах Северного полушария), 51 неравномерно расположенных вертикальных уровней в немонопольном режиме (запуск через систему очередей). Размерность сетки задачи — $1600 \times 865 \times 51$. Значения параллельного ускорения в зависимости от числа используемых вычислительных ядер приведены на рис. 3 и 4.

Шаги с расчетом радиации

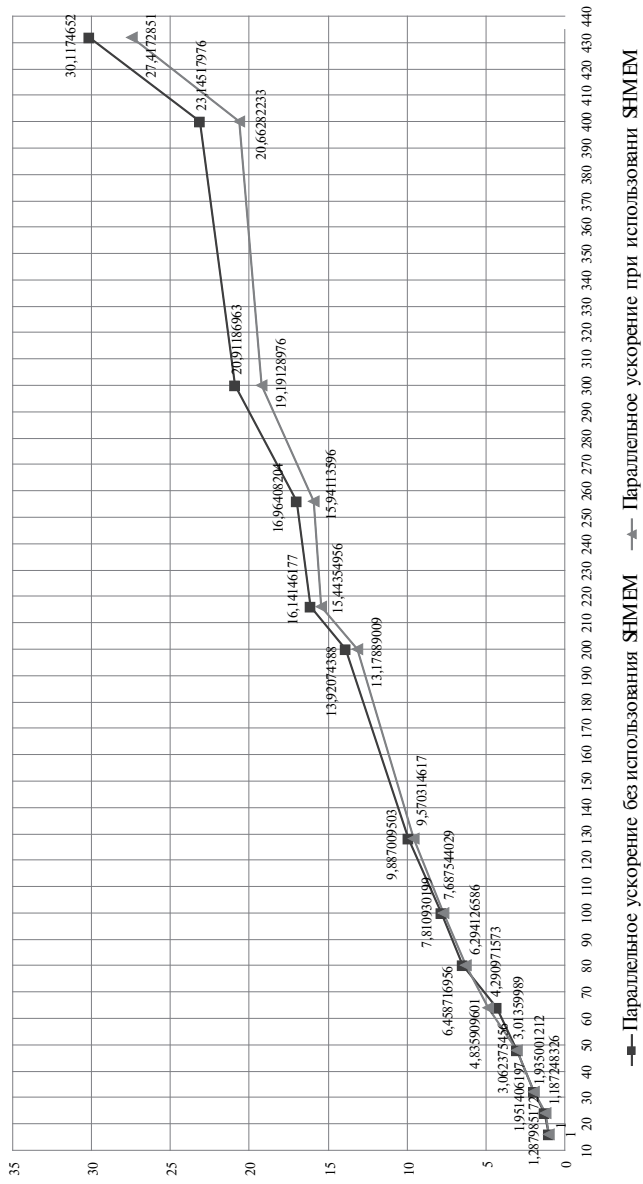


Рис. 3. Параллельное ускорение модели на вычислительной системе SGI Altix 4700 для шагов, в которых производится расчет радиации.

Шаги без расчета радиации

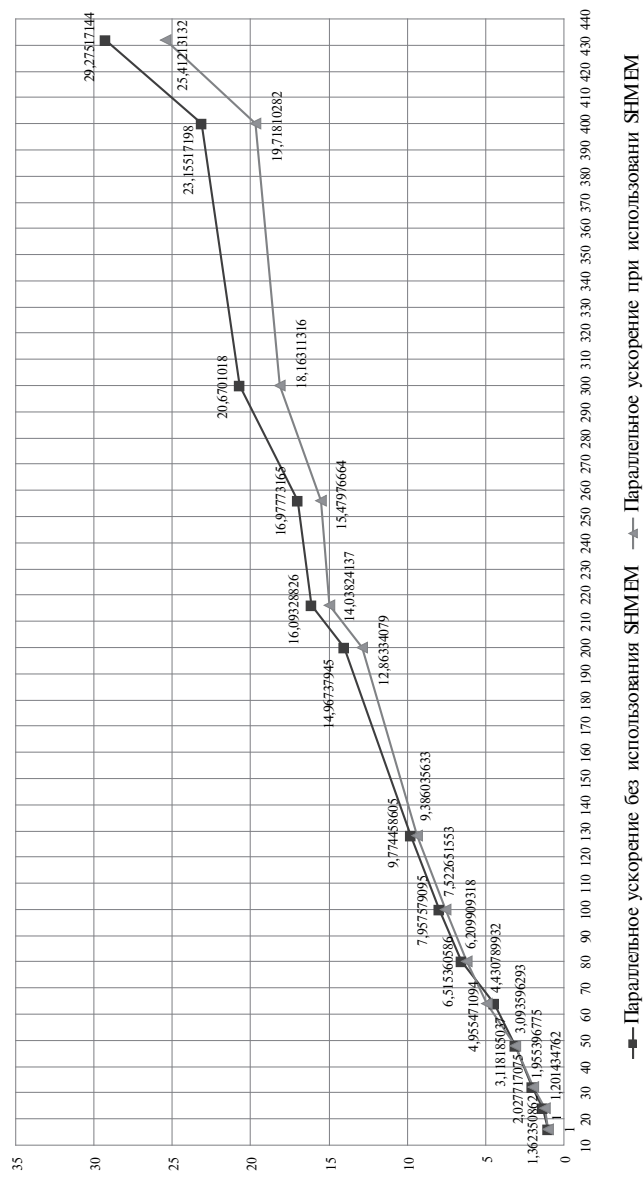


Рис. 4. Параллельное ускорение модели на вычислительной системе SGI Altix 4700 для шагов, в которых не производится расчет радиации.

На рисунках представлены графики параллельных ускорений для версий модели с использованием библиотеки SHMEM и без ее использования. Первый график отражает обратное отношение среднего времени выполнения одного шага модели, в котором не происходит вызова параметризации коротко- и длинноволновой радиации, к времени выполнения на 16 ядрах. Второй график построен по значениям времени для тех шагов, в которых радиация рассчитывается.

При небольшом количестве используемых процессорных ядер задача занимает слишком большой объем оперативной памяти и требует очень большого объема системных буферов MPI, поэтому минимальным количеством ядер, на котором можно запускать данную версию модели, является 16. По этой причине среднее время расчета одного шага по времени при различном количестве используемых ядер сравнивалось с временем расчета на 16 ядрах. Форма графиков практически одинакова на обоих рисунках, значения ускорений в обоих случаях слегка различаются.

В ходе численных экспериментов было выявлено, что оптимальным количеством OpenMP-нитей является четыре. Судя по всему, это обусловлено особенностями аппаратной реализации вычислительного комплекса Altix 4700, на котором производились запуски модели (4 ядра имеют физически общую память на той же плате, что и ядра).

Видно, что параллельное ускорение при количествах ядер, которым кратно число 864, заметно выше, чем при других исследованных значениях. При запуске на 432 ядрах наблюдается суперлинейное ускорение в случаях запуска без использования SHMEM и для шагов с расчетом радиации при запуске с использованием SHMEM. Это связано с тем, что модельная сетка без учета полюсов имеет 864 узла по широте, а декомпозиция расчетной области для распараллеливания по технологии MPI происходит по широтным полосам. Таким образом, при количествах ядер, являющихся делителями 864, наблюдается равномерная загруженность всех ядер вычислениями, и передачи сообщений лучше синхронизированы во времени. Еще одной причиной хорошего ускорения в этом случае является, по всей видимости, меньший объем данных, приходящийся на один процессор и, вследствие этого, лучшая их загрузка в кэш-память процессора. Это существенно увеличивает скорость доступа к данным в процессе вычислений.

Также можно заметить, что использование библиотеки SHMEM дает прирост ускорения, по сравнению с запуском без ее использования, только в случае расчета на 64 ядрах, как в случае шагов с расчетом радиации, так и в случае шагов без расчета радиации. Среднее время выполнения одного шага при использовании SHMEM увеличивается, по сравнению с запуском без использования библиотеки при использовании более 256 ядер. Скорее всего, это обусловлено тем, что при увеличении количества ядер сильно увеличивается время ожидания каждым процессом завершения всех односторонних коммуникаций. Технология SHMEM предусматривает блокировку процесса, пока он не окончил копирование из памяти другого процесса. Кроме того, перед началом обмена SHMEM предусмотрена барьерная синхронизация. В технологии MPI при использовании неблокирующих коммуникаций возможно выполнение арифметических операций на фоне обмена. Таким образом, при большом количестве процессов в варианте SHMEM тратится много времени на ожидание завершения всех односторонних вызовов.

Заключение

Для повышения параллельной эффективности программного комплекса модели ПЛАВ был выполнен ряд работ по оптимизации обменов между процессорами и распараллеливанию постпроцессинга, интегрированного в модель. В результате удалось заметно повысить эффективность параллельного программного комплекса модели. Прогноз на 24 часа по модели ПЛАВ в версии с размерностью сетки 1600'865, 51 уровнями по вертикали (разрешение во внетропической части Северного полушария не хуже 22 км) с параметризацией микрофизических процессов неконвективного характера ALARO-0 может быть рассчитан на 432 ядрах вычислительной системы Altix 4700 за 25 минут. Предполагается, что ряд запланированных оптимизаций модели (внедрение алгоритма дихотомии [4] вместо векторной прогонки с транспозициями данных, оптимизация процедур ввода-вывода) позволит сократить время расчета прогноза на сутки до 20 минут.

Работа выполнена при поддержке грантов РФФИ 10-05-01066, 12-09-12054 офи-м, Программ 14 и 17 Президиума РАН.

Список использованных источников

1. *Володин Е.М., Толстых М.А.* Параллельные вычисления в задачах моделирования климата и прогноза погоды // Вычислительные методы и программирование. – 2007. – Т. 8. – С. 113–122.
2. *Толстых М.А.* Глобальная полулагранжева модель численного прогноза погоды. – М.-Обнинск: ОАО ФОП, 2011. – 111 с.
3. Report of the twenty-fifth session of the Working Group on Numerical Experimentation (WGNE) (Offenbach, Germany, 26 November 2009). World Meteorological Organisation. Geneva. 2010. WMO/TD N 1524. – 25 p. – <http://www.wmo.int/pages/prog/wcrp/documents/wgne25rpt.pdf>.
4. *Terekhov A.V.* Parallel Dichotomy Algorithm for solving tridiagonal system of linear equations with multiple right-hand sides // Parallel Computing. – 2010. – V. 36. – P. 423–438.
5. *Tolstykh M.A., Gloukhov V.N.* Implementation of global atmospheric models on parallel computers // Вычислительные технологии. – 2002. – Т. 7. Спец. выпуск. – С. 101–109.

Поступила в редакцию 01.07.2011 г.