

Библиотека PARMONC для решения «больших» задач по методу Монте-Карло*

М.А. Марченко

Институт вычислительной математики и математической геофизики СО РАН

В работе представлена библиотека PARMONC (сокращение от «PARallel MONte Carlo»), предназначенная для распараллеливания широкого круга приложений метода Монте-Карло, обладающих большой вычислительной трудоемкостью. Цель разработки библиотеки - создать простой в использовании программный инструмент для реализации распределенных вычислений, не требующий от пользователя знания языка MPI. «Ядром» библиотеки является тщательно протестированный, быстрый и надежный длиннопериодный генератор псевдослучайных чисел. С использованием библиотеки распараллеливание сложных программ статистического моделирования производится достаточно просто. В результате вычислительная нагрузка автоматически распределяется по процессорам кластера.

1. Введение

В настоящее время ведущими специалистами по вычислительной математике высказывается убеждение в том, что в ближайшем будущем в области компьютерного моделирования будут широко применяться вероятностные имитационные модели и методы Монте-Карло (методы численного статистического моделирования). С одной стороны, это убеждение основано на том, что вероятностные имитационные модели дают адекватное описание физических, химических, биологических и др. явлений при их рассмотрении «из первых принципов». С другой стороны, алгоритмы метода Монте-Карло, реализующие вероятностные модели, допускают возможность эффективного распараллеливания в виде распределенного статистического моделирования. Методы Монте-Карло перспективны в связи с появлением супер-ЭВМ экзафлопсного уровня в ближайшем будущем [1].

Упомянем одно из важных применений методов статистического моделирования. С течением химически реагирующих газов часто приходится сталкиваться при рассмотрении физических явлений, связанных с повышением температуры газовых смесей выше критической. К таким явлениям относятся процессы горения газовых топлив в камерах сгорания, соплах и горелках различных геометрических форм. Численное моделирование на компьютерах течений химически реагирующих газов по методу Монте-Карло является весьма трудоёмким, поскольку приходится учитывать многие физические и химические особенности протекающих при горении процессов. Как показывает анализ литературы, в последние десятилетия понимание химии горения (по крайней мере, с участием небольших молекул) и возможности моделирования процессов горения на больших компьютерах достигли такого уровня, который обеспечивает необходимую надежность результатов. Численное моделирование процессов горения требует значительных вычислительных ресурсов и относится к числу задач, для которых необходимо использование супер-ЭВМ. Такие потребности обусловлены большим средним временем расчета каждой реализации ансамбля тестовых частиц, большим числом независимых реализаций ансамбля (которое определяется из необходимой точности расчетов) и большим объемом используемой машинной памяти. Для такого рода задач применение описываемой в настоящей работе библиотеки PARMONC позволит существенно сократить трудоемкость расчетов.

*Работа проводилась в рамках реализации федеральной целевой программы "Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы", государственный контракт 07.514.11.4016, а также при финансовой поддержке грантов РФФИ №№ 10-07-00454, 12-01-00034, 12-01-00727.

В зависимости от условий задачи и параметров алгоритма статистического моделирования применяются разные методики параллельной реализации. Метод «распределенного статистического моделирования» состоит в распределении независимых реализаций ансамбля тестовых частиц по отдельным процессорам с периодическим осреднением полученных результатов по статистически эффективной формуле. При этом обмен между процессорами сводится к минимуму и в самом простом случае состоит лишь в начальном и финальном обмене данными. Распределенное статистическое моделирование является примером идеально масштабируемого параллельного алгоритма. Возможные отказы компонент вычислительной системы и связанные с этим потери данных приводят лишь к отсутствию соответствующих смоделированных выборочных значений, что компенсируется расчетами на других вычислительных узлах [2].

Следует упомянуть широкий класс приложений, например, методы прямого статистического моделирования в задачах газовой динамики и теории дисперсных систем, в которых ресурсов одного вычислительного узла недостаточно для моделирования реализации ансамбля и требуется распределение нагрузки между процессорами. При решении пространственно неоднородных кинетических уравнений для моделирования особо крупных ансамблей тестовых частиц целесообразно применять способы распараллеливания, основанные на пространственной декомпозиции расчетной области («параллельные алгоритмы статистического моделирования»). При этом следует применять способы динамической балансировки вычислительной нагрузки, основанные на специальных формулах прогноза времени вычислений каждым процессором [3].

2. Распределенное статистическое моделирование

Под численным статистическим моделированием обычно понимают реализацию с помощью компьютера вероятностной модели некоторого объекта с целью оценивания изучаемых интегральных характеристик на основе закона больших чисел [4]. При этом чем больше объем выборки, составленной из смоделированных независимых реализаций, тем выше точность оценивания, причем статистическая погрешность убывает обратно пропорционально квадратному корню от числа реализаций.

Как правило, при параллельной реализации необходимый объем выборки базовых случайных чисел очень велик, поэтому целесообразно использовать «длиннопериодные» псевдослучайные последовательности. А именно, для решения т.н. «больших задач» по методу Монте-Карло предлагается использовать генератор следующего вида [5]:

$$u_0 = 1, \quad u_n \equiv u_{n-1}A \pmod{2^{128}}, \quad \alpha_n = u_n 2^{-128}, \quad n = 1, 2, \dots,$$

где

$$A \equiv 5^{100109} \pmod{2^{128}}.$$

Последовательность чисел $\{\alpha_n\}$ является периодической, длина ее периода равна $L = 2^{126} \approx 10^{38}$.

Для распределения псевдослучайных чисел между разными вычислительными ядрами предлагается следующий порядок их использования. Последовательность $\{u_n\}$ предварительно разбивается на подпоследовательности длины μ , начинающиеся с чисел $u_{m\mu}$, $m = 0, 1, 2, \dots$. Разные подпоследовательности используются на разных ядрах. Значение «прыжка» генератора μ должно выбираться из соображений, чтобы μ псевдослучайных чисел хватало для моделирования на каждом ядре. Ясно, что для метода вычетов начальные значения $u_{m\mu}$ указанных подпоследовательностей получаются по формуле

$$u_{(m+1)\mu} \equiv u_{m\mu}A_\mu \pmod{2^{128}}, \quad A_\mu \equiv A^\mu \pmod{2^{128}}, \quad \alpha_{m\mu} = u_{m\mu} \cdot 2^{-128}. \quad (1)$$

Этот параллельный генератор был проверен с помощью содержательных статистических тестов [2] и успешно используется в ряде институтов СО РАН на протяжении последних десяти лет.

С целью понижения трудоемкости моделирование независимых выборочных значений можно распределить по разным вычислительным ядрам. Очевидно, что главным критерием осуществимости такой параллельной реализации является возможность «поместить» вычислительную программу моделирования выборочного значения в память каждого ядра.

При распределении вычислений по ядрам следует допускать возможность реализации различных объемов выборки на различных ядрах с использованием статистически оптимального способа осреднения результатов по формулам вида

$$\bar{\zeta} = \frac{\sum_{m=1}^M n_m \bar{\zeta}_m}{\sum_{m=1}^M n_m},$$

где M – число ядер, n_m – объем выборки для m -го процессора, $\bar{\zeta}_m$ – соответствующее среднее значение. Таким образом, можно добиться обратно пропорциональной зависимости величины трудоемкости случайной оценки от числа ядер (при условии, что используемые ядра имеют одинаковую производительность) [2].

Описанную методику параллельного моделирования можно легко модифицировать с целью эффективной оценки функционалов, зависящих от параметра $x \in X$:

$$\varphi(x) \approx E\zeta(x; \omega).$$

При этом моделируемое распределение случайной величины ω может зависеть, а может и не зависеть от параметра x . Аналогичную методику можно также использовать для оценки функционалов, возникающих в результате осреднения базовых функционалов по случайному параметру σ задачи, например по случайной плотности среды. Вероятностное представление при этом имеет вид двойного математического ожидания

$$\varphi \approx EE\zeta(\omega, \sigma).$$

Примеры подобных задач приведены в [2].

3. Библиотека PARMONC

С целью унификации реализации распределенного статистического моделирования при решении широкого круга задач методом Монте-Карло разработана и внедрена в широкое использование программная библиотека PARMONC (сокращение от PARallel MONte Carlo). Библиотека и примеры ее использования описаны в работах [6–8].

Возможность применения библиотеки PARMONC для распараллеливания широкого круга задач определяется естественной модульностью программ статистического моделирования. Общая структура такого рода программ, в упрощенном виде, следующая (в нотации языка C):

```

TypeRO RO, SUBT;
SUBT={0.0, .., 0.0};
for (i=0; i<L; i++){
    realization(RO);
    SUBT= SUBT+RO;
}
SUBT=SUBT/L;

```

Здесь L – общее число независимых реализаций случайного объекта, задаваемого композитным типом данных *TypeRO* (допускается поэлементное суммирование переменных такого типа). Реализации моделируются в результате работы процедуры (или блока операторов) *realization*. Полученная таким образом реализация *RO* (статистически независимая от других моделируемых реализаций) добавляется к «счетчику» *SUBT* и впоследствии осредняются, что дает статистическую оценку искомого математического ожидания случайного

объекта. В процессе вычислений используются потоки псевдослучайных чисел, получаемых в результате работы подпрограммы, которая в общем случае вызывается следующим образом:

```
a = rand();
```

Здесь a – очередное псевдослучайное число, равномерно распределенное в интервале от нуля до единицы. Для использования PARMONC определяется процедура *realization* (моделирующая подпрограмма), возвращающая реализацию RO . При этом считается, что моделирующая подпрограмма использует потоки псевдослучайных чисел, генерируемых внешней по отношению к ней подпрограммой. Описанный выше цикл по независимым реализациям и финальное осреднение заменяются вызовом библиотечной процедуры следующего вида:

```
parmoncc (realization, L, ...);
```

Здесь имя моделирующей подпрограммы и общее число независимых реализаций передаются в библиотечную процедуру *parmoncc* в качестве аргументов; для простоты остальные аргументы библиотечной процедуры опущены. Процедура *parmoncc* автоматически распределяет моделирование независимых реализаций по процессорам используемой многопроцессорной вычислительной системы. Все остальные операторы пользовательской программы остаются без изменений.

Библиотека PARMONC предназначена для использования на кластерах Сибирского суперкомпьютерного центра (ССКЦ КП СО РАН) для распараллеливания широкого круга трудоемких приложений метода Монте-Карло. «Ядром» библиотеки является тщательно протестированный, быстрый и надежный длиннопериодный генератор псевдослучайных чисел (1), разработанный в Лаборатории методов Монте-Карло ИВМиМГ СО РАН.

Библиотечные подпрограммы могут быть использованы в пользовательских программах, написанных на языках C, C++ и Fortran, причем от пользователя не требуется знание языка MPI. В процессе счета происходит автоматическое получение выборочных средних и границ погрешностей для статистических оценок, алгоритм моделирования которых задается в пользовательской подпрограмме. Имя такой подпрограммы передается в качестве аргумента в соответствующую библиотечную подпрограмму. В процессе счета результаты вычислений периодически сохраняются на жестком диске в удобном для дальнейшей обработки виде. Библиотечные подпрограммы автоматически распределяют вычислительную нагрузку по процессорам кластера. С помощью библиотеки PARMONC можно легко организовать продолжение ранее проведенных расчетов с автоматическим учетом их ре-

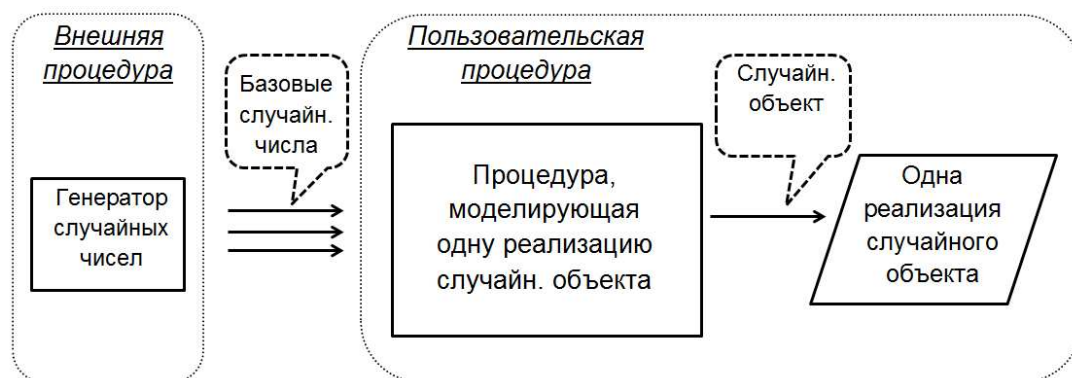


Рис. 1. Взаимосвязь объектов при статистическом моделировании

зультатов. Также с помощью библиотеки можно получать коррелированные статистические оценки различных функционалов.

Библиотека установлена на кластере НКС-30Т ССКЦ КП СО РАН в директории `/ifs/apps/parmonc/`. Библиотека PARMONC состоит из следующих подпрограмм и исполняемых файлов:

- *rnd128* - функция для получения одного случайного числа, равномерно распределенного в интервале от 0 до 1, с помощью параллельного генератора случайных чисел (1);
- *parmoncf* - процедура, осуществляющая распределенное статистическое моделирование (для программ на Fortran-e);
- *parmoncc* - процедура, осуществляющая распределенное статистическое моделирование (для программ на C);
- *manaver* - программа для осреднения выборочных средних, независимо рассчитанных на разных процессорах;
- *genparam* - программа для расчета параметров параллельного генератора случайных чисел.

Здесь *rnd128*, *parmoncf* и *parmoncc* являются библиотечными подпрограммами для использования в пользовательских программах на Fortran-e, C или C++, *genparam* и *manaver* являются исполняемыми файлами для запуска из командной строки. Объектные файлы PARMONC упакованы в статическую библиотеку *libparmonc.a*. Пользователь вставляет вызовы процедур *rnd128* и *parmoncf/parmoncc* в свои программы. Главная пользовательская программа, в которой находится вызов *parmoncf/parmoncc*, рассматривается компилятором как MPI-программа, несмотря на то, что в самой пользовательской программе нет явных вызовов директив MPI. Это означает, что такая программа должна компилироваться и собираться с использованием команд *mpiicc* или *mpii fort*. Результаты расчетов, выполненных с использованием процедур *parmoncf/parmoncc*, сохраняются в файлы, которые находятся в специальной поддиректории в рабочей директории пользователя. Все эти файлы обновляются всякий раз, когда центральный процессор получает данные с других процессоров, осредняет их и сохраняет на диск.

В файл `.bashrc`, который находится в домашней директории пользователя, необходимо добавить следующую строку:

```
source $/ifs/apps/parmonc/bin/parmoncvars.sh$
```

Тем самым объявляются три переменные окружения

```
$PRMCBIN , $PRMCLIB , $PRMCINC .
```

Эти переменные используются при компиляции и сборке приложений с помощью PARMONC, а также при запуске команд.

Для компиляции и сборки пользовательских программ с использованием библиотеки PARMONC следует использовать следующие команды (пример для программ на C):

```
$ mpiicc -o test -L$PRMCLIB -I$PRMCINC test.c -lparmonc
```

Здесь *test* - имя исполняемого файла, *test.f90* и *test.c* - пользовательские программы.

4. Заключение

Упомянем некоторые направления дальнейшего развития библиотеки PARMONC. Библиотека может стать базовым программным уровнем для масштабируемых параллельных приложений метода Монте-Карло, реализующих сложные вероятностные модели естествознания. Кроме того, целесообразно адаптировать библиотеку к современным супер-ЭВМ с гибридной архитектурой.

Литература

1. Глинский Б.М., Родионов А.С., Марченко М.А., Подкорытов Д.В. Винс Д.И. Агентно-ориентированный подход к имитационному моделированию суперЭВМ экзафлопсной производительности в приложении к распределенному статистическому моделированию // Вестник ЮУрГУ, серия "Математическое моделирование и программирование" – 2012 – 12 С. (принято в печать)
2. Марченко М.А., Михайлов Г.А. Распределенные вычисления по методу Монте-Карло // Автоматика и телемеханика. – 2007 – Вып. 5 – С. 157–170.
3. Marchenko M.A. Majorant frequency principle for an approximate solution of a nonlinear spatially inhomogeneous coagulation equation by the Monte Carlo method // Russ. J. Numer. Anal. Math. Modelling. – 2008. -- Vol. 21, N. 3. – P. 199-218.
4. Михайлов Г.А., Войтишек А.В. Численное статистическое моделирование. Методы Монте-Карло – Москва: Издательский центр «Академия», 2006.
5. Marchenko M.A. Parallel Pseudorandom Number Generator for Large-scale Monte Carlo Simulations // LNCS -- 2007 – Vol. 4671 – P. 276-282.
6. Marchenko M. PARMONC – A Software Library for Massively Parallel Stochastic Simulation // LNCS -- 2011 -- Vol. 6873 – P. 302-315.
7. Страница библиотеки PARMONC на сайте ССКЦ КП СО РАН:
<http://www2.sccc.ru/SORAN-INTEL/paper/2011/parmonc.htm>.
8. Ссылка на документацию к библиотеке PARMONC на сайте ССКЦ КП СО РАН:
<http://www2.sccc.ru/SORAN-INTEL/paper/2011/parmonc.pdf>.