

УДК 519.687.1

© А. К. Новиков

**БАЛАНСИРОВКА ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ
ДЛЯ МНОГОЯДЕРНЫХ КЛАСТЕРОВ¹**

На примере конечно-элементного приложения рассматривается балансировка вычислительной нагрузки для многоядерных кластеров. Балансировка основывается на многоуровневом отображении расчетных данных и учитывает, что латентность и пропускная способность средств коммуникации между ядрами процессоров, между процессорами и между вычислительными узлами существенно отличаются.

Ключевые слова: вычислительный кластер, многоядерность, балансировка вычислительной нагрузки, метод конечных элементов.

Современные многопроцессорные вычислительные системы строятся на основе многоядерных процессоров и многопроцессорных вычислительных узлов. Существенное различие в латентности и пропускной способности коммуникационных средств при обмене данными между процессами, размещенными на разных ядрах и разных вычислительных узлах, подтверждается результатами теста производительности межпроцессорных обменов [1], проведенного на МВС-100k (Межведомственный суперкомпьютерный центр РАН, г. Москва). Латентность при однонаправленном обмене средствами MPI внутри вычислительного узла составляет 0.3-0.6 мкс, между вычислительными узлами 3-4 мкс. Пропускная способность внутри вычислительного узла в 1.5-2 раза выше.

Применительно к параллельному конечно-элементному приложению учесть многоядерность вычислительной системы можно разделяя расчетную сетку на подобласти одним из следующих способов.

1. Выбрать число подобластей сетки равным числу ядер процессоров.
2. Выполнить многоуровневое разделение сетки, создав дерево данных с корнем — исходной сеткой и числом листьев равным количеству ядер процессоров. В обоих случаях выделение вычислительных процессов и обмен осуществимы средствами стандарта MPI-1.
3. Выбрать число подобластей сетки равным количеству вычислительных узлов. Процессы внутри конечно-элементного приложения порождать

¹Работа выполнена при финансовой поддержке РФФИ (грант 06-07-89015).

динамически средствами MPI-2 или OpenMP (распараллеливание циклов при решении системы уравнений МКЭ).

Для многоуровневого разделения число подобластей (первый уровень) сетки берется равным числу вычислительных узлов. Полученные подобласти распределяются по вычислительным узлам и там разделяются на подобласти второго уровня (по числу ядер на узле). Рассматриваются следующие варианты обменов данными между подобластями второго уровня.

1. Обмены между подобластями второго уровня с разных вычислительных узлов происходят непосредственно.

2. Подобласти второго уровня обмениваются через подобласти первого уровня. Для этого выбираются процессы, которые или имеют минимальную нагрузку, или соответствуют двусвязным подобластям второго уровня.

Приводятся результаты исследований на MBC-100k, в ходе которых методом конечных элементов решается задача теории упругости, а балансировка вычислительной нагрузки осуществляется на основе приведенных выше подходов. При разделении исходной сетки и подобластей применяются алгоритмы разделения неструктурированных сеток [2]. Сравнение результатов подтверждает преимущество многоуровневого подхода.

* * *

1. Тест производительности межпроцессорных обменов. НИВЦ МГУ.
<http://parallel.ru/testmpi/transfer.html>
2. Копысов С. П., Новиков А. К. Параллельные алгоритмы адаптивного перестроения и разделения неструктурированных сеток // Матем. моделир. 2002. Т. 14, № 9. С. 91–96.

Поступила в редакцию 15.02.08

A. K. Novikov

Load balancing for multicores clusters

The paper presents a study of multilevel load balancing in the finite element method application for multicores clusters.

Новиков Александр Константинович
Институт прикладной
механики УрО РАН,
426067, Россия, г. Ижевск,
ул. Т.Барамзиной, 34
E-mail: sc_work@mail.ru