

КАКАЯ НАМ ПОЛЬЗА ОТ ТЕСТА HPCG.

С.С. Андреев¹, С.А. Дбар¹, А.О. Лацис¹, Е.А. Плоткина¹

¹*Институт прикладной математики им. М. В. Келдыша РАН, lacis@kiam.ru*

Новый тест производительности HPCG совершенно не случайно появился в очень интересное и очень тяжелое для суперкомпьютерной отрасли время. Впервые в мировой истории высокопроизводительных вычислений осознание невозможности развития традиционным путем, возникшее совсем не вчера, сочетается с полным отсутствием общепринятых работающих альтернатив. Очень нужны новые архитектуры, причем, скорее всего, они будут проблемно-ориентированными, а значит – многочисленными. В этих условиях к общепринятому тесту производительности (или к системе тестов) предъявляются требования, которых не было еще, например, 7-8 лет назад.

Во-первых, тест должен «сравнивать несравнимое» - сопоставлять разные архитектуры, каждая из которых, скорее всего, хорошо справляется с одними задачами, и гораздо хуже – с другими.

Во-вторых, тест должен быть руководством по налаживанию межцехового диалога между системными инженерами, системными программистами и прикладными программистами, поскольку хорошие проблемно-ориентированные архитектуры не могут быть созданы ни силами одних системных инженеров (инженеры не знают методов), ни силами одних прикладных программистов (прикладные программисты не знают «железа»).

Словом, тест должен олицетворять собой конструктивный, систематический и междисциплинарный подход к сопоставлению архитектур.

Отвечает ли этим требованиям тест HPCG? С одной стороны, сам по себе – не отвечает. С другой стороны, как сам тест, так и приводимая его авторами мотивация, дают нам много ценных идей как для построения более полезной системы тестов, так и для налаживания на ее основе межцехового диалога.

На примере сопоставления старого (HPL) и нового (HPCG) тестов любопытно проследить диалектическую спираль требований к системе тестирования. Старый тест был предложен в то время, когда только появились процессоры с заметными объемами кэш-памяти, впервые позволившими значительно ускорять вычисления за счет локализации обработки, в блочных методах. Требовалось подчеркнуть важность локализации, и был предложен тест, позволявший это сделать. Причем сделать настолько хорошо, что никакая реальная программа по возможности такой локализации с предложенным искусственным тестом и близко сравниться не могла. За прошедшие с тех пор 20 с лишним лет усовершенствование процессоров общего назначения оставило системным архитекторам единственный резерв повышения производительности – ту самую локализацию (все остальные резервы давно вычерпаны). Старый тест, с его искусственными, неправдоподобно хорошими возможностями локализации обработки, из стимула прогресса стал его тормозом – теперь он провоцирует создание архитектур, способных эффективно выполнять старый тест HPL, и ничего больше. Ошибка исправлена

– новый, основанный на реальном численном методе, тест HPCG вообще не позволяет локализовать обработку данных в мелких блоках. И это – в то время, когда никаким другим путем ускориться в принципе нельзя! Виток спирали пройден и завел нас в тупик. Зачем же нам такой тест? Неужели только для демонстрации любимого некоторыми математиками (и, кстати, не вполне верного) утверждения, гласящего, что «распараллеливаются только самые плохие методы»?

Присмотримся подробнее не к букве, но к духу нового теста. В отличие от старого, он является «действующей моделью» реально интересного с практической точки зрения численного метода. Авторы теста предлагают для определения производительности хронометрировать не метод, а алгоритм (конкретную реализацию метода), но никто не заставляет нас поступать так же. Метод можно модифицировать, и в таком виде он становится пригоден и для мелко-блочной обработки, и для применения повышенной точности (то и другое повышает локализацию, а значит – может быть сильно ускорено на новых архитектурах). Критерием правомерности модификации служит ускорение сходимости (уменьшение числа итераций) на тяжелых, плохо обусловленных матрицах из флоридской коллекции. Отметим, что в исходной постановке матрица берется простая, настолько хорошо обусловленная, что для ее сходимости требуется буквально 1-2 итерации.

Модификация метода HPCG в сторону возможности мелко-блочной локализации вычислений интересна как пример того, какого рода работу приходится проделывать с методом, чтобы сделать его пригодным для новых архитектур. Этим обусловлена ценность теста с точки зрения налаживания междисциплинарного диалога. Но не только этим. HPCG является методом вычислительной линейной алгебры, а методы вычислительной линейной алгебры характерны не только своей чрезвычайной актуальностью в суперкомпьютерных приложениях. Они еще очень просто и хорошо классифицируются (в докладе рассказывается, как именно), причем типичные представители классов просты и равно понятны системным инженерам, системным и прикладным программистам, как и сами принципы классификации. Это позволяет нам легко построить систему тестов, обладающую внутренней стройностью, причем метод HPCG будет одним из тестов этой системы. Применяя эту систему тестов к различным архитектурам, мы сможем действительно много узнать об их (архитектур) применимости к различным классам реальных задач.