

ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ МНОГОЯДЕРНЫХ ПРОЦЕССОРОВ. ЧАСТЬ 2

МАКС ДОМЕЙКА (MAX DOMEIKA), специалист по программному обеспечению, Intel

Во второй части статьи (см. Часть 1 в ЭК10) представлены методы оценки и прогнозирования производительности разрабатываемого приложения на базе результатов проведенных тестов. Рассмотрены примеры измерения характеристик многоядерного процессора при работе в условиях конкретного приложения и методы анализа полученных результатов. Статья представляет собой сокращенный перевод [1].

ОЦЕНКА И ПРОГНОЗИРОВАНИЕ ПОВЕДЕНИЯ СИСТЕМЫ И ПРИЛОЖЕНИЯ

При анализе результатов тестирования перед разработчиком системы обычно встает вопрос: «Насколько верно результаты теста предсказывают производительность приложения, если использовать новый процессор или новый компилятор»? Другими словами, если новый процессор или компилятор увеличивают производительность теста X на Y%, то насколько этот процессор или компилятор улучшат разрабатываемое приложение?

Понятно, что ответ на этот вопрос зависит от ряда условий, ведь приложение не повторяет в точности тестовую программу. Кроме того, следует скептически относиться к тому утверждению, что новый процессор или компилятор так же повысят производительность системы, что была достигнута в тестовых данных. Обычно говорят, что следует ожидать некоторого улучшения производительности, и есть несколько статистических методов, которые помогут увеличить степень доверия к данным тестирования при оценке производительности разрабатываемого приложения. Для определения степени возможного улучшения производительности приложения на основе данных тестирования применяются три следующих подхода:

- предположение, что повышение производительности скоррелировано с улучшениями, полученными при тестировании;
- расчет коэффициента корреляции между приложением разработчика и некоторыми тестами, а затем использование данных о производительности системы из теста с наивысшей степенью корреляции;
- многомерный регрессионный анализ с использованием данных, полу-

ченных при ранее проведенном тестировании.

Первые два метода довольно легко реализовать. Третий метод использует многомерную регрессию для оценки производительности приложения с использованием в качестве исходных следующих видов данных:

- данные тестирования ряда ранее выпускавшихся процессоров и нового процессора;
- данные о производительности приложения, работающего на процессорах, для которых уже имеются тестовые данные.

Рассмотрим данные о производительности приложения, представленные в таблице 1. В колонках таблицы приведены данные о производительности в виде баллов для каждого теста из пакета SPEC CINT2000 для ряда процессоров, включая Intel Pentium 4, Intel Core Duo и Intel Core 2 Quad.

Колонка «Приложение разработчика» содержит гипотетические данные, которые характеризуют производительность приложения, полученную на каждом из перечисленных в таблице процессоров (это реальные данные, взятые из теста CINT2000/300 Base benchmark).

Колонка «Коэффициент корреляции» содержит отношение производительности, измеренной в каждом отдельном тесте, к данным из колонки «Приложение разработчика». Данные колонки «Оценка MPP» получены при многомерном регрессионном анализе, в котором данные для каждого теста CINT2000 являются независимыми переменными, а данные из колонки «Приложение разработчика» — зависимыми переменными.

Наконец, колонка «Базовый уровень» показывает общий балл тестов CINT2000. Если разработчик рассматривает возможность переноса приложения с процессора Dual-Core Intel Xeon

5160 с частотой 3 ГГц на процессор Intel Core 2 Extreme X6800 с частотой 2,93 ГГц, то в таблице 2 приведена оценка разницы в производительности системы на базе этих процессоров на основе данных из таблицы 1. В данном случае предполагается, что проверка приложения на базе процессора Intel Core 2 Extreme X6800 еще не была проведена, поэтому в колонке «Реальное значение» баллы для этого процессора отсутствуют (обозначено «????»). Процентная разница для колонки «Базовый уровень» предполагает, что приложение разработчика выигрывает 6,11% от перехода на новый процессор. Если использовать данные теста с наивысшей корреляцией (метод 2) с разрабатываемым приложением (тест 175 Base с корреляцией 0,994), то улучшение производительности составит 2,22%.

Если применить многомерную регрессию к представленному набору данных, производительность уменьшается на 1,68%. Реальная производительность уменьшается на 0,99%, и это говорит о том, что в данном примере многомерная регрессия показывает наиболее близкий результат. Конечно, нет гарантии, что использование коэффициента корреляции или многомерной регрессии приведет к лучшему прогнозированию производительности приложения для всех случаев. Однако сравнение трех вариантов оценки производительности обеспечивает большую степень достоверности.

Таким образом, вооружившись результатами проведенных тестов CPU2000 и данными тестирования разрабатываемого приложения на некоторых процессорах, можно сформировать еще две оценки ожидаемого улучшения производительности при переходе на новый процессор без запуска приложения на этом процессоре. Оценка на основе коэффициента корреляции между результатами отдельных тестов

Таблица 1. Оценка производительности приложения с помощью тестов SPEC CINT2000

Тип процессора	Тактовая частота, МГц	164 Base	175 Base	176 Base	181 Base	186 Base	197 Base	252 Base	253 Base	254 Base	255 Base	256 Base	Приложение разработчика	Оценка ММП	Базовый уровень
Intel Pentium 4	1300	488	277	544	463	430	425	573	613	631	656	380	377	366	474
Intel Pentium 4	1800	653	346	732	525	673	564	775	795	834	920	504	446	470	624
Intel Pentium 4	2000	718	334	692	511	644	580	880	905	884	845	503	464	453	636
Dual-Core Intel Xeon	2800	933	935	1654	1717	1041	1156	1769	1674	1519	2419	1050	1420	1397	1381
Dual-Core Intel Xeon LV 1,67 ГГц	1667	959	1128	1606	1680	1281	1182	1697	1558	1412	2219	1052	1712	1727	1417
Intel Pentium 4 631	3000	998	997	1683	1747	1099	1147	1982	1753	1658	2583	1104	1493	1477	1456
Intel Pentium D 930	3000	1005	1009	1715	1785	1119	1172	2002	1795	1676	2638	1113	1500	1510	1477
Dual-Core Intel Xeon LV 2,0 ГГц	2000	1147	1298	1892	1817	1537	1387	2036	1853	1660	2589	1228	2055	2032	1663
Intel Pentium 4	3800	1262	1031	2015	1415	1408	1442	2338	2171	1996	2746	1149	1855	1847	1663
Intel Pentium 4	3600	1207	1211	2046	1980	1342	1491	2250	2136	1947	3063	1289	1805	1835	1744
Intel Dual-Core Xeon 5080	3733	1261	1162	2101	1788	1377	1518	2355	2196	1943	2821	1301	1899	1896	1747
Intel Dual-Core Xeon 5080	3730	1282	1201	2101	1869	1415	1441	2547	2233	2041	3035	1329	1913	1922	1793
Intel Core Duo T2700	2333	1338	1511	2247	2215	1888	1646	2643	2225	1966	2984	1409	2489	2489	1987
Intel Core2 Duo T7400	2166	1345	1655	2578	3597	1972	1800	2820	2693	2182	3842	1736	2493	2481	2284
Intel Core 2 Duo T7600	2333	1441	1782	2784	3870	2114	1918	3034	2882	2332	4070	1850	2663	2695	2446
Intel Core2 Duo T7600	2333	1446	1781	2793	3852	2128	1933	3040	2892	2345	4096	1858	2687	2699	2455
Intel Core 2 Duo E6600	2400	1492	1899	2931	4316	2190	2010	3134	3017	2515	4451	1949	2776	2773	2588
Intel Core 2 Quad Extreme QX6700	2666	1653	2053	3209	4519	2426	2215	3477	3310	2742	4795	2136	3075	3057	2829
Intel Core 2 Duo E6700	2667	1663	2062	3225	4511	2437	2218	3488	3336	2738	4786	2150	3057	3086	2836
Intel Core 2 Duo E6700	2666	1654	2074	3223	4586	2432	2218	3478	3336	2738	4858	2149	3081	3070	2844
Intel Dual-Core Xeon 5160	3000	1816	2205	3070	4349	2827	2161	3456	3470	2878	4597	2244	3424	3424	2929
Intel Core 2 Extreme X6800	2933	1816	2254	3525	4921	2674	2432	3824	3666	2977	5280	2354	???	3367	3108
Коэффициент корреляции		0,981	0,994	0,984	0,944	0,988	0,989	0,983	0,980	0,975	0,974	0,984			0,990

CPU2000 и многомерной регрессии позволяет получить более точную информацию об ожидаемой производительности.

ИЗМЕРЕНИЯ ХАРАКТЕРИСТИК ВСТРАИВАЕМОЙ СИСТЕМЫ

Кроме чистой оценки производительности на основе полученных данных, тесты могут также дать характеристики многоядерного процессора в условиях конкретного приложения при различных параметрах системы. Для примера на рисунках 1 и 2 отражена производительность двух приложений, работающих на двухъядерном процессоре Intel Pentium D и использующих разное число параллельных рабочих элементов.

Как видно из рисунков, при выполнении вращения изображения может произойти резкое увеличение производительности системы, по всей вероятности, из-за работы кэш-памяти приложения. При выполнении быстрого преобразования Фурье (БПФ) использование множественных рабочих элементов увеличивает пропускную способность примерно на 20%. При параллельной обработке более двух наборов данных наблюдается небольшое снижение производительности системы.

Система на базе двух процессорных ядер — это самый простой случай. С помощью тестов можно снять характеристики системы с количеством ядер больше двух. Рисунки 3 и 4 отражают производительность системы на базе

Таблица 2. Сравнение общей производительности системы на базе двух процессоров

Процессор	175 Base	Базовый уровень	Оценка ММП	Реальные данные
Intel Dual-Core Xeon 5160	2205	2929	3424	3424
Intel Core 2 Extreme X6800	2254	3108	3367	???
Разница, %	2,22	6,11	-1,68	-0,99

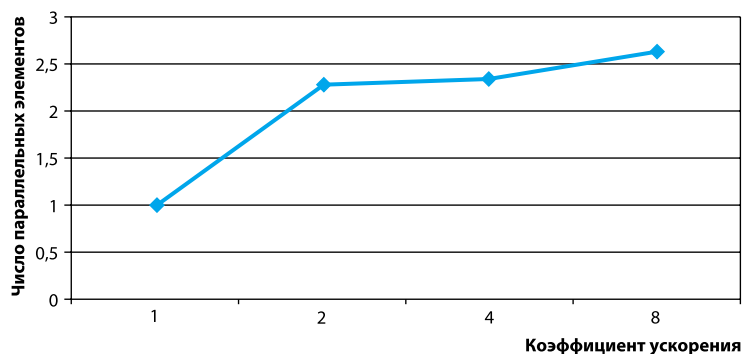


Рис. 1. Производительность системы при выполнении вращения изображения

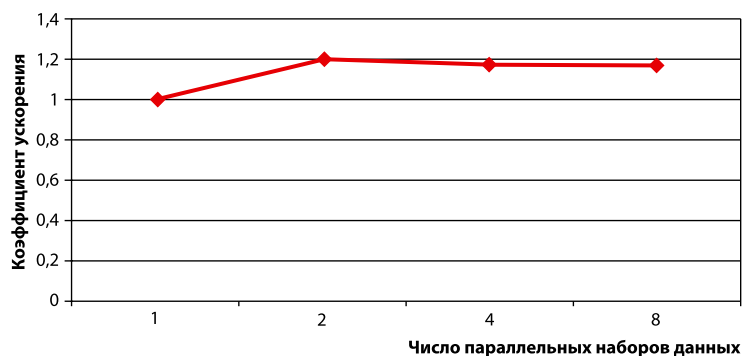


Рис. 2. Производительность системы при выполнении БПФ

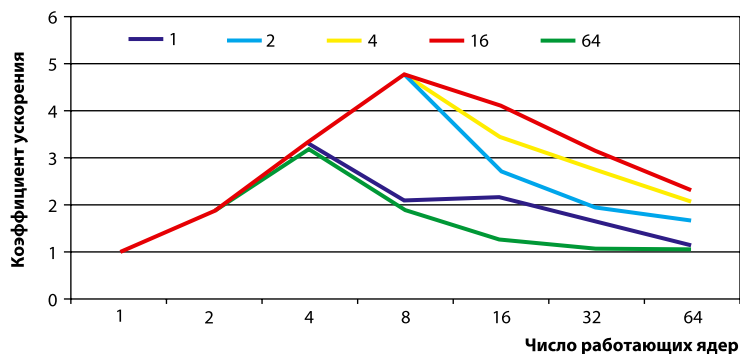


Рис. 3. Коэффициент ускорения при выполнении вращения изображения на 16-ядерной процессорной системе

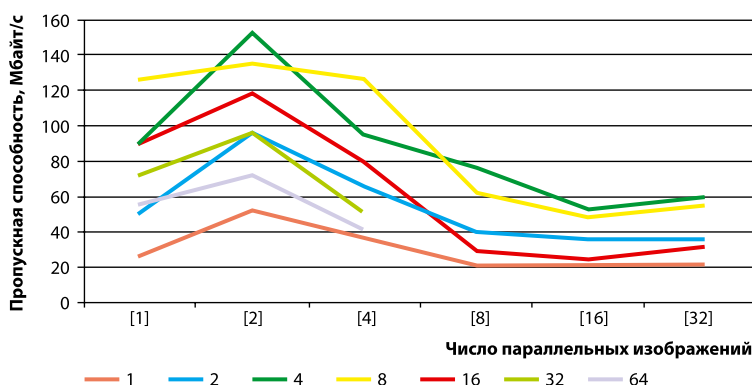


Рис. 4. Пропускная способность при выполнении вращения изображения на 16-ядерной процессорной системе

16 ядер при выполнении вращения изображения. Эти же рисунки показывают результаты выполнения теста EEMBC по вращению изображения в градациях серого размером 4 Мп. Рисунок 3 отражает эффект использования множественных ядер для ускорения обработки одного изображения, в то время как на рисунке 4 отображены результаты оптимизации пропускной способности системы (т.е. общего количества обработанных изображений).

Термин «размер среза данных» отражает глубину синхронизации разных ядер, обрабатывающих одну и ту же картинку, причем меньший размер среза данных требует большей глубины синхронизации. Рисунок 3 показывает, что при обработке одного изображения с использованием нескольких ядер со средней глубиной синхронизации работа системы может ускориться в 5 раз.

Заметим, что производительность растет практически линейно при 2—4 активных ядрах, однако когда активны все 16 ядер, производительность падает. Скорее всего, это связано с тем, что при работе 16 ядер затраты системных ресурсов на синхронизацию начинают превышать доступную вычислительную мощность.

Рисунок 4 показывает пропускную способность системы, причем каждая линия отражает число процессорных

ядер, работающих с любым отдельным изображением. Можно заметить улучшение пропускной способности более чем в 6 раз. Однако результаты все еще далеки от теоретической возможности 16-кратного ускорения 16-ядерной системы. Также интересно заметить, что работа восьми процессорных ядер обеспечивает наилучшую производительность, когда два изображения обрабатываются одновременно четырьмя процессорными ядрами.

АНАЛИЗ РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ

Тесты позволяют сделать оценку и сравнить производительность встраиваемой системы, однако часто на основе результатов тестирования могут быть сделаны недостаточно корректные и разумные выводы. Правильная оценка результатов зависит от параметров используемого теста. При анализе полученных результатов следует учитывать следующие факторы:

- конфигурацию системы;
- сертификацию теста;
- режимы выполнения тестирования (базовый или пиковый тест);
- анализ однопотоковых и многократных тестовых данных.

При сравнении производительности процессоров различных вендоров важно рассматривать систему в целом

и, в первую очередь, анализировать системы, наиболее сходные по конфигурации.

Во-вторых, следует определить, сертифицированы ли результаты теста (в случае EEMBC и BDTI), и опубликованы ли они официально (в случае SPEC). Сертифицированные результаты тестирования проверяются сторонними организациями, поэтому они вызывают большее доверие. Официально опубликованные результаты могут быть в дальнейшем пересмотрены другими компаниями.

В-третьих, необходимо учитывать условия проведения тестов. Например, SPEC использует два метода проведения тестов для одноядерных процессоров. Первоначальное тестирование, или базовые результаты, требуют ограниченного количества опций, которые должны быть одинаковыми для всех тестов в пакете. Второе тестирование, или пиковые результаты, допускают использование различных опций компилятора. Нецелесообразно сравнивать систему, имеющую базовые результаты тестирования, с системой, для которой были получены пиковые результаты.

Подобно SPEC, EEMBC также допускает два типа результатов: стандартные и оптимизированные. Стандартные результаты предполагают, что исходный код теста EEMBC не был модифицирован. Оптимизированные результаты допускают модификацию кода для использования функционально эквивалентных алгоритмов при выполнении тестов.

Тестовый пакет BDTI Benchmark Suites предназначен для растущего рынка приложений в области цифровой обработки сигнала. В таких приложениях программное обеспечение обычно тщательно отлажено для целевого процессора, поэтому результаты тестирования на BDTI Benchmark Suites основаны на полностью оптимизированном коде. Кроме того, требуется, чтобы вендоры получили разрешение от BDTI при распространении результатов сравнения продуктов с использованием их тестов.

Наконец, следует уделить внимание корректному использованию тестов. Недопустимо сравнивать производительность одноядерного и многоядерного процессоров, используя одноядерные тесты, например CPU2000 base, и утверждать, что многоядерный процессор не увеличивает вдвое производительность и, следовательно, совершенно не пригоден. Такие тесты нельзя использовать для анализа многоядерных процессоров.

ЛИТЕРАТУРА

1. Max Domeika. *Evaluating the performance of multi-core processors.*