

Методы и критерии сравнительной оценки микропроцессоров

Практически все существующие на сегодняшний день методы оценки производительности процессоров фактически основаны на измерении времени, затрачиваемого вычислителем на решение некоторой детерминированной пользовательской задачи. Считается, что чем меньше вычислитель затрачивает времени, тем более производительным он является. Однако, оценка таких временных затрат является весьма сложной задачей. Связано это во многом с наличием дополнительных накладных временных расходов, возникающих в процессе измерений, связанных с необходимостью выполнения системного кода и осуществления операций ввода/вывода. Особенно серьезные трудности возникают при выполнении измерений с последующим сравнением результатов на разных платформах и операционных системах. В связи с этим в методах оценки производительности в зависимости от целевой задачи используют разное время: астрономическое, показывающее временной интервал между двумя событиями; время выполнения, показывающее время, требуемое для выполнения определенного алгоритма; время ответа, показывающее временной интервал, проходящий между окончанием некоторого запроса и началом ответа на него вычислительной системой; или прошедшее время, представляющее собой общее время выполнения задания и включающее в себя время работы процессора, время на обращение к оперативной памяти, время на операции ввода/вывода, накладные расходы на выполнение системного кода. Для описания времени, затрачиваемого процессором на выполнение заданного алгоритма, используется термин время центрального процессора. Из этого времени исключено время ожидания выполнения операций ввода-вывода и время выполнения других программ. Необходимость применения этого параметра обусловлена тем, что в многозадачном режиме происходит совмещение выполнения разных задач, использующих различные ресурсы.

Особенностью процессора как объекта измерения является то, что скорость взаимодействия функциональных устройств процессора чаще всего не зависит от динамических характеристик этих устройств, а задается частотой генератора тактовых импульсов. Поэтому время центрального процессора может быть вычислено двумя способами: умножением количества тактов синхронизации необходимых для выполнения заданного алгоритма на длительность такта синхронизации, либо делением количества тактов синхронизации, необходимых для выполнения заданного алгоритма, на частоту синхронизации. Важной характеристикой процессора является среднее количество тактов синхронизации, необходимых для выполнения одной команды (CPI clock cycles per instruction). Этот параметр позволяет оценить время процессора, необходимое для выполнения заданного алгоритма, зная количество выполняемых команд.

В итоге производительность процессора определяется тремя параметрами: тактовой частотой; средним количеством тактов на команду; количеством выполняемых команд. При сравнении производительности процессоров необходимо рассматривать все три компонента. Исходя из определения производительности, как скорости выполнения операций можно сказать, что интегральными критериями оценки производительности процессора будут являться количество выполненных операций за единицу времени или количество времени, затрачиваемое на выполнение заданного числа операций.

В процессе поиска стандартной метрики для оценки производительности исторически было выбрано несколько единиц измерения. Одной из наиболее распространенных среди них является IPS (instructions per second), которая показывает количество инструкций, выполняемых процессором в одну секунду. На первый взгляд метрика проста, но ее использование не позволяет учесть особенности выполняемого кода и архитектуру процессора. С ее помощью невозможно сравнивать процессоры, имеющие разные системы команд и архитектуры. Серьезные проблемы возникают также в связи с использованием сопроцессоров для выполнения команд с плавающей запятой и оптимизирующих компиляторов. Все это не дает возможности выполнять объективное сравнение с помощью метрики IPS разных процессоров.

Для задач, связанных с научно-техническими расчетами, требующими арифметики с плавающей точкой, производительность процессоров оценивается в показателе FLOPS (floating point operations per second), который показывает, сколько операций с плавающей запятой выполняет

Секция 12. Построение и анализ радиотехнических систем

в секунду данный процессор. В реальности FLOPS также является достаточно плохой мерой производительности, поскольку неоднозначным является уже само его определение, а получаемые результаты существенно зависят от конкретного выполняемого кода и архитектуры процессора. Другая проблема заключается в том, что время выполнения разных операций с плавающей запятой может существенно отличаться, в результате рейтинг FLOPS существенно зависит от состава тестовой смеси. Поэтому результаты, полученные на одной и той же системе с помощью различных тестов, могут существенно различаться.

Кроме того, IPS и FLOPS подвержены влиянию очень многих факторов, напрямую не связанных с производительностью самого процессора, например: пропускная способность каналов связи с окружением процессора, производительность основной памяти, синхронность работы кэш-памяти и других. Все это приводит к получению разных количественных результатов даже при проведении серии однотипных измерений на одной и той же вычислительной системе.

Несмотря на большое число существенных недостатков, IPS и FLOPS продолжают использоваться для оценки производительности процессоров. Причины такой популярности обусловлены тем, эти метрики являются абсолютными величинами, кроме того подавляющее большинство вычислительных систем построены по классической архитектуре с использованием стандартных процессоров, что позволяет использовать для сравнения общепринятые тесты.

С целью получения интегрального показателя производительности и локализации измерительного алгоритма в классе предполагаемых решаемых задач было разработано большое количество пакетов синтетических и натуральных тестов, содержащих наборы команд специально подобранных тестовых фрагментов алгоритмов, максимально приближенных к предметной области применения вычислителя. С учетом области применения настоящей работы к таким тестам в первую очередь можно отнести LINPACK, тесты SPEC, смеси AIM, Whetstone/Dhrystone, Nbench и CoreMark. В области цифровой обработки сигналов наиболее авторитетным специализированным пакетом тестов является тест VTDMark2000. Синтетические и натуральные тесты не могут служить в качестве настоящих тестовых пакетов для оценки производительности, поскольку они не могут точно моделировать среду конечного пользователя и оценивать производительность всех компонентов системы, участвующих в конкретном вычислительном процессе. Без такой гарантии точность результатов измерения производительности остается под вопросом.

Более того, результаты тестов обычно представляются в виде некоторых относительных индексов, позволяющих провести сравнение процессоров между собой по различным параметрам, включая производительность. При этом остается совершенно не ясно, сможет ли процессор, обладающий некоторым значением индекса, удовлетворить требованиям по выполнению целевой задачи за заданное время, следовательно, вопрос, может ли быть тот или иной процессор применен для построения пользовательской системы, остается в итоге открытым.