

Анализ и моделирование производительности кластерных ЭВМ на задаче аэродинамики*

К.С. Солнушкин

Проведен анализ производительности ряда кластерных ЭВМ на основе современных центральных процессоров производства Intel и AMD (двух- и четырехядерных) на типовой задаче аэродинамики с использованием программной системы FLUENT. Предложены модели производительности.

1. Введение

Основным подходом к повышению производительности ЭВМ является *согласование архитектуры ЭВМ и структуры алгоритмов*, которые преимущественно будут выполняться на этой ЭВМ. В ряде случаев применяемые алгоритмы заданы исходно и не могут быть изменены. Это происходит, например, при расчетах с использованием готовых программных систем, таких, как САЕ-система “FLUENT” компании “ANSYS” для расчетов в области аэрогидродинамики.

В таких случаях единственным способом повышения производительности является выбор такой архитектуры ЭВМ, которая хорошо согласуется с применяемыми алгоритмами. Под архитектурой ЭВМ здесь понимается многоуровневая иерархическая структура аппаратных ресурсов и программного обеспечения вычислительной системы.

Для кластерных ЭВМ характерны типовые способы построения архитектуры. Наибольшее разнообразие вариантов архитектуры реализуется на системотехническом уровне, который определяет структуру вычислительной системы: состав и способ связи модулей ЭВМ. При проектировании кластерных ЭВМ на этом уровне задают такие характеристики, как тип и число центральных процессоров вычислительного узла, количество вычислительных узлов, вид высокоскоростной вычислительной сети, объединяющей узлы («интерконнекта»). Кроме того, на производительность кластерной ЭВМ важное влияние оказывает применяемая операционная система и реализация коммуникационной библиотеки (например, MPI).

Таким образом, возникает задача целенаправленного проектирования кластерных ЭВМ с целью достижения наибольшей производительности на имеющихся алгоритмах; эта задача решается преимущественно путем выбора оптимальной структуры ЭВМ. При этом большую роль играет *анализ производительности* этих алгоритмов на имеющихся ЭВМ, а также прогнозирование производительности для ЭВМ произвольной структуры на основе математических моделей, то есть *моделирование производительности*.

Для построения математических моделей производительности применяют статистическую обработку экспериментальных данных на основе методов нелинейной регрессии [12], широко используется имитационное моделирование с применением потактовых эмуляторов центральных процессоров (cycle-by-cycle simulation) [10], в том числе с последующей обработкой результатов с помощью нейронных сетей [9].

В ряде случаев модель производительности можно построить, применяя знания о структуре используемых алгоритмов [1]. Существует подход к моделированию производительности, основанный на сравнении характеристик исследуемого приложения и известных программ оценочного тестирования [11]. Если требуется учитывать случайные составляющие в работе ЭВМ, например, отказы оборудования, применяют вероятностные модели производительности [5].

При моделировании производительности на основе обработки экспериментальных данных основными препятствиями являются недостаточность и неточность имеющихся данных. Так, в данной работе исследуется производительность программной системы “FLUENT” на типовой задаче аэродинамики – обтекании корпуса автомобиля с учетом турбулентности; обработаны имеющиеся в открытом доступе экспериментальные данные для 14 различных конфигураций

* Работа поддержана грантом Правительства Санкт-Петербурга для молодых ученых за 2008 г.

кластерных ЭВМ с числом ядер до 128 [8]. Для каждой конфигурации доступны данные о времени вычислений при различном числе задействованных ядер, на основе именно этих данных происходит поиск зависимостей между конфигурацией ЭВМ и ее производительностью.

С одной стороны, в опубликованных экспериментальных данных о поведении системы “FLUENT” уделено недостаточно внимания детальному описанию исследуемых ЭВМ, например, иногда не указываются конкретные модели ЦП, кроме того, не приведены данные о конфигурации и характеристиках оперативной памяти, тогда как известно, что при различной конфигурации ОЗУ разница в производительности может достигать 15 % [6]. Вместе с тем, существует значительное число факторов, влияющих на производительность, но практически не поддающихся учету. Например, применение библиотеки MPI, рассчитанной на работу с сетью Infiniband, обычно дает прирост производительности до 15 % по сравнению с библиотекой, рассчитанной на сеть Gigabit Ethernet, даже в том случае, если вычисления производятся в рамках одного узла кластера, когда сеть не используется вовсе [6] (sic!). Приведенные примеры говорят о том, что зависимости, построенные по опубликованным экспериментальным данным, будут описывать производительность ЭВМ со значительными погрешностями, с чем придется смириться.

В данной работе предложена модель, в которой производительность кластерной ЭВМ произвольной конфигурации дается произведением двух компонент: производительности в последовательном режиме R_{ser} и эффективности распараллеливания Eff . Первая компонента аппроксимируется линейной зависимостью от тактовой частоты центрального процессора (ЦП), коэффициент пропорциональности определяется типом ЦП. Вторую компоненту – эффективность распараллеливания – находят из условия принадлежности исследуемой конфигурации к одной из пяти групп; для каждой группы функция эффективности задана табличным способом, значения функции получены путем усреднения экспериментальных данных. Несмотря на простоту, модель демонстрирует хорошее соответствие с экспериментальными данными.

Модели производительности находят применение в синтезе кластерных ЭВМ. Сопоставив каждой конфигурации ЭВМ значения производительности и совокупной стоимости владения [2], находят значение критерия «цены [единицы] производительности» [4]. Это – интегральный критерий качества ЭВМ. Синтез ЭВМ, то есть выбор ее конфигурации, осуществляют на основе поиска минимума указанного критерия на множестве возможных конфигураций ЭВМ.

2. Характеристики вычислительного процесса в параллельном режиме счета

Рассмотрим величины, характеризующие вычислительный процесс. Под ЭВМ мы будем понимать программно-аппаратный комплекс, состоящий из аппаратного обеспечения, операционной системы и, в случае параллельных вычислений, библиотеки передачи сообщений (MPI). Первой из рассматриваемых величин является *производительность ЭВМ в последовательном режиме*, которая зависит от быстродействия аппаратных компонентов ЭВМ: центрального процессора (ЦП), оперативного запоминающего устройства (ОЗУ) и др., а также от применяемой операционной системы.

Производительность, разумеется, зависит и от решаемой задачи, а именно, от применяемых математических моделей и методов и их характеристик, а также от входных данных задачи. Так, для широкого круга задач вычислительной аэрогидродинамики, заключающихся в решении уравнений Навье-Стокса, к числу факторов, влияющих на производительность, относятся: вычислительная схема (явная, неявная), шаг по времени, применяемые модели турбулентности и многие другие факторы.

Второй характеристикой является *производительность в параллельном режиме*. Третьей характеристикой вычислительного процесса при работе ЭВМ в параллельном режиме является *ускорение счета* (англ. “speedup”). Ускорение определяют как частное от производительности на N потоках к производительности в последовательном режиме.

Четвертая характеристика вычислительного процесса – степень близости достигнутого ускорения к линейному. Ее называют «*эффективность распараллеливания*», находят как частное между достигнутым ускорением и линейным и выражают в процентах. Высокая эффективность

распараллеливания (близкая к 100 % даже при больших N) свидетельствует об удачном сочетании задачи и архитектуры ЭВМ, а также является косвенным признаком высокого качества проектирования ЭВМ (сбалансированного подбора компонентов). В данной работе именно эффективность распараллеливания применяется для анализа производительности в параллельном режиме.

Для пользователей кластерной ЭВМ наибольший интерес представляет производительность ЭВМ именно при вычислениях в параллельном режиме, а не в последовательном. Поэтому возникает вопрос, можно ли по производительности ЭВМ в последовательном режиме обоснованно прогнозировать производительность в параллельном режиме. Оказывается, что это невозможно. Рассмотрим в качестве примера две кластерных ЭВМ, на основе ЦП типов “А” и “В” соответственно. На одной из задач из пакета оценочного тестирования системы “FLUENT” [7] в последовательном режиме первая ЭВМ достигла производительности в 59 задач в сутки против 39 у второй ЭВМ; в основном за счет более высокой тактовой частоты ЦП типа “А”. Отсюда можно сделать неверный вывод о том, первая ЭВМ и в параллельном режиме будет демонстрировать более высокую производительность.

Однако при вычислениях в параллельном режиме с ростом числа подзадач N эффективность распараллеливания первой ЭВМ резко падает, и при $N=32$ составляет 45 % против 74 %. (Это связано, по-видимому, с недостаточным быстродействием ОЗУ первой ЭВМ, ограничивающим производительность). Производительность ЭВМ при $N=32$ потока составила, соответственно, 848 и 919 задач в сутки, то есть вторая ЭВМ оказалась в параллельном режиме быстрее первой.

Таким образом, по имеющимся данным о производительности ЭВМ в последовательном режиме *невозможно* сделать надежного прогноза о производительности в параллельном режиме, и это очень важный вывод. Это предопределяет необходимость специального анализа данных о производительности в параллельном режиме, чему посвящена вторая часть данной статьи.

3. Анализ и моделирование производительности в последовательном и параллельном режиме счета

Проведем анализ производительности ряда кластерных ЭВМ на основе современных двух- и четырехядерных ЦП при численном решении типовой задачи аэродинамики – обтекании корпуса автомобиля с учетом турбулентности. Используются экспериментальные данные о производительности (в задачах в сутки), опубликованные в открытом доступе, для 14 различных конфигураций кластерных ЭВМ с числом ядер до 128 [8]. Эти данные не приведены в статье из-за их значительного объема.

Конфигурации кластерных ЭВМ (тип ЦП, число ядер одного ЦП, общее число ядер ЭВМ, тактовая частота, тип операционной системы (ОС) и вычислительной сети) сведены в табл. 1.

Таблица 1. Конфигурации кластерных ЭВМ

№ п/п	Производитель	Тип ЦП	Число ядер		Тактовая частота ЦП, ГГц	Тип ОС	Тип вычислительной сети
			в ЦП	в ЭВМ			
1	Bull	Intel Woodcrest	2	128	3	Linux	IB
2	HP	Intel Woodcrest	2	64	3	Linux	IB
3	HP	Intel Woodcrest	2	32	3	WinCCS	IB
4	IBM	Intel Woodcrest	2	128	3	Linux	GigE
5	IBM	Intel Woodcrest	2	128	3	Linux	IB
6	IBM	Intel Woodcrest	2	128	3	Linux	Myrinet
7	SGI	Intel Harpertown	4	128	2,8	Linux	IB
8	SGI	Intel Woodcrest	2	128	3	Linux	IB
9	Tyan	Intel Woodcrest	2	16	2,3	Linux	GigE
10	Tyan	Intel Woodcrest	2	16	2,3	WinCCS	GigE
11	HP	AMD Opteron	2	64	2,2	WinCCS	IB
12	HP	AMD Opteron	2	128	2,6	Linux	IB

13	IBM	AMD Opteron	2	128	2,8	Linux	GigE
14	IBM	AMD Opteron	2	128	2,8	Linux	IB

В таблице применены следующие сокращения: “WinCCS” – ОС “Microsoft Windows Compute Cluster Server 2003”, “IB” – вычислительная сеть “Infiniband”, “GigE” – вычислительная сеть “Gigabit Ethernet”.

Производительность каждой конкретной ЭВМ имеет случайный характер. Действительно, для двух ЭВМ с одинаковой конфигурацией (например, п.п. 1, 2 и 5) значения производительности в последовательном режиме отличаются на величину до 15 %, а в параллельном режиме при 64 потоках – до 9 %. Это объясняется тем, что кроме перечисленных в таблице факторов, на производительность влияют и прочие, неучтенные факторы.

Предложим зависимость производительности в последовательном режиме от тактовой частоты ЦП. На производительность в последовательном режиме влияют тип ЦП, его тактовая частота и тип применяемой ОС. Будем исходить из упрощающего предположения, что производительность пропорциональна тактовой частоте. В таком случае желательно проанализировать производительность, разбив множество ЭВМ на группы по двум оставшимся факторам: тип ЦП и тип ОС. Типов ЦП имеется три: “Intel Woodcrest” двухядерные, “Intel Harpertown” четырехядерные, “AMD Opteron” двухядерные. Типов ОС имеется два: “Linux” и “WinCCS”. Таким образом, множество исследуемых ЭВМ разбивается на 6 групп, однако в этом случае из-за недостатка экспериментальных данных невозможно распределить ЭВМ по группам равномерно; поэтому мы будем применять разбиение только по типам ЦП, а тип ОС при построении модели учитывать не будем.

Исходя из сделанного нами предположения о том, что производительность в последовательном режиме пропорциональна тактовой частоте ЦП, зависимость будет определяться формулой $R_{ser} = k_{CPU} f_{CPU}$, где R_{ser} – производительность (англ. “serial” – последовательный), в задачах в сутки; f_{CPU} – тактовая частота ЦП, ГГц; k_{CPU} – искомый коэффициент пропорциональности, свой для каждого типа ЦП. Более совершенные ЦП при той же тактовой частоте обеспечивают большую производительность ЭВМ, поэтому им соответствует большее значение k_{CPU} .

Коэффициент k_{CPU} также зависит от параметров решаемой задачи; так, при прочих равных факторах на задачах большей размерности ЭВМ достигает меньшей производительности, поэтому для таких задач значение k_{CPU} меньше.

Для каждой ЭВМ мы находим k_{CPU} как частное от деления производительности в последовательном режиме на тактовую частоту ЦП, а затем усредняем найденные значения по ЭВМ в каждой группе. Значения k_{CPU} для разных типов ЦП приведены в табл. 2. (По четырехядерным ЦП “Intel” данные имеются только для одной ЭВМ, поэтому представленное значение k_{CPU} нельзя считать достаточно надежным).

Таблица 2. Значения коэффициента пропорциональности для производительности в последовательном режиме

Тип ЦП	k_{CPU}
AMD Opteron, двухядерные	40,3
Intel Woodcrest, двухядерные	49,0
Intel Harpertown, четырехядерные	57,0

Отношение значений коэффициентов k_{CPU} для двух ЦП показывает, во сколько раз отличается производительность в последовательном режиме для ЭВМ, основанных на указанных ЦП, на данной задаче. Так, применение двухядерных ЦП “Intel Woodcrest” обеспечивает производительность примерно на 20 % выше, чем двухядерных ЦП “AMD Opteron”, при одинаковой тактовой частоте.

Предложим теперь зависимость для описания производительности в параллельном режиме. При вычислениях в параллельном режиме к числу факторов, определяющих производительность ЭВМ, добавляется быстроедействие вычислительной сети и библиотеки “MPI”. Эти факторы определяют достижимую эффективность распараллеливания: высокоскоростные сети позволяют достичь большей эффективности, а значит и большего ускорения счета по сравнению с последовательным режимом.

Зная эффективность распараллеливания $Eff = \varphi(N)$, достижимую на данной ЭВМ при N потоках вычислений, можно, исходя из определения эффективности, получить ускорение счета по сравнению с последовательным режимом: $S = Eff \cdot N$. Тогда производительность в параллельном режиме можно найти по формуле $R_{par} = S \cdot R_{ser} = Eff \cdot N \cdot R_{ser}$; здесь R_{ser} – производительность в последовательном режиме, зависимость ее от тактовой частоты мы нашли ранее.

Анализ экспериментальных данных показал, что эффективность распараллеливания можно прогнозировать для разных типов вычислительных сетей; следовательно, производительность ЭВМ в параллельном режиме также можно прогнозировать, применяя указанные формулы.

В качестве примера на рис. 1 приведены графики эффективности распараллеливания $Eff = \varphi(N)$ для четырех ЭВМ из табл. 1; ЭВМ № 5 и 8 основаны на двухядерных ЦП “Intel Woodcrest”, а № 12 и 14 – на двухядерных ЦП “AMD Opteron”. Как видно из графика, обе ЭВМ на основе ЦП “AMD Opteron” вплоть до 32 потоков вычислений сохраняют значения эффективности распараллеливания, близкие к единице.

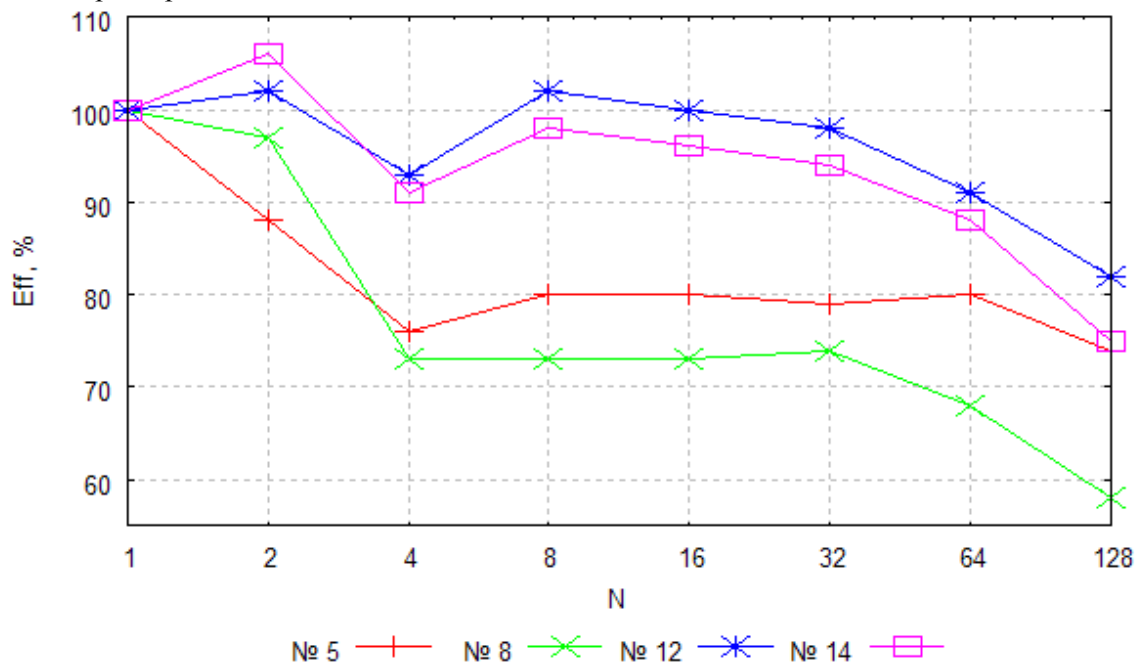


Рис. 1. Эффективность распараллеливания для четырех кластерных ЭВМ (N – количество потоков вычислений; нумерация ЭВМ соответствует табл. 1. ЭВМ № 5 и 8 демонстрируют один шаблон поведения эффективности с ростом числа потоков, а ЭВМ № 12 и 14 – другой шаблон)

Множество исследуемых ЭВМ можно разбить на группы в соответствии с шаблонами поведения эффективности распараллеливания при росте числа потоков. Всего было идентифицировано пять непересекающихся групп (в скобках указаны номера ЭВМ в соответствии с табл. 1):

1. ЭВМ с вычислительной сетью “Gigabit Ethernet” (№ 4, 9, 10, 13);
2. ЭВМ с двухядерными ЦП “Intel Woodcrest” и вычислительной сетью “Infiniband” (№ 1-3, 5, 8);
3. То же, с четырехядерными ЦП “Intel Harpertown” (№ 7);
4. ЭВМ с двухядерными ЦП “AMD Opteron”, вычислительной сетью “Infiniband” и ОС “Linux” (№ 12, 14);

5. То же, с ОС “WinCCS” (№ 11).

Во всех группах найдем для каждого значения N (числа потоков) среднее значение эффективности распараллеливания для ЭВМ этой группы.

В табл. 3 приведены модельные значения эффективности распараллеливания Eff , которые можно использовать для прогнозирования производительности в параллельном режиме (группы с номерами 3 и 5 состоят всего из одной ЭВМ, поэтому модельные данные нельзя считать достаточно надежными).

Таблица 3. Модельные значения эффективности распараллеливания Eff по группам ЭВМ, %

Число потоков N	Номер группы				
	1	2	3	4	5
2	94	94	97	104	100
4	79	74	81	92	98
8	78	75	51	100	89
16	71	75	52	98	88
32	75	75	55	96	82
64	52	74	58	89	73
128	37	68	50	79	63

Приведем примеры использования найденных зависимостей (моделей производительности). Допустим, требуется найти производительность ЭВМ на основе “AMD Opteron” с тактовой частотой 3,2 ГГц в последовательном режиме. Для этого необходимо умножить значение коэффициента пропорциональности $k_{CPU} = 40,3$ на тактовую частоту ЦП: $R_{ser} = k_{CPU} f_{CPU} = 128,96$.

В случае применения ЦП с тактовой частотой 2,8 ГГц производительность по модели составит 112,8; тогда как экспериментальное значение – 106,0; таким образом, относительная погрешность модели составляет порядка 6 %.

Пусть требуется отыскать производительность кластерной ЭВМ на основе ЦП “AMD Opteron” с тактовой частотой 2,8 ГГц, в параллельном режиме, при 64 потоках, при применении ОС “Linux” и вычислительной сети “Gigabit Ethernet”. Данная ЭВМ относится к группе 1, для которой из табл. 3 находим, что при распараллеливании на 64 потока достигается эффективность $Eff=52$ %. Тогда ускорение счета $S = Eff \cdot N = 0,52 \cdot 64 = 33,28$. Производительность в последовательном режиме известна из эксперимента и составляет 106,0; тогда производительность в параллельном режиме, согласно найденной нами зависимости, составит $R_{par} = S \cdot R_{ser} = 33,28 \cdot 106,0 = 3527,7$. Погрешность относительно имеющихся данных эксперимента составляет около 5 %. В случае применения сети “Infiniband” ЭВМ относится к группе 4. Тогда $Eff=89$ %, $S = 56,96$, $R_{par} = 6037,8$. Погрешность составляет в этом случае около 6 %.

4. Выводы

В работе предложены простые зависимости, позволяющие определить производительность кластерных ЭВМ на основе современных двух- и четырехядерных ЦП при вычислениях на типовой задаче аэродинамики при использовании САЕ-системы “FLUENT”. Предложенные модели обладают приемлемыми значениями погрешностей относительно имеющихся экспериментальных данных.

Дальнейшие направления работ связаны с проверкой установленных закономерностей для новых классов вычислительных задач.

Литература

1. Бухановский А.В., Зильберштейн О.И., Иванов С.В., Ковальчук С.В., Лопатухин Л.И., Попов С.К. Моделирование экстремальных явлений в атмосфере и океане как задача высокопроизводительных вычислений // Вычислительные методы и программирование: новые вычислительные технологии.– М.: Изд-во НИВЦ МГУ им. М.В.Ломоносова, 2008.– №9.– С. 141-153.
2. Солнушкин К.С. Моделирование совокупной стоимости владения вычислительной системой // Научно-технические ведомости СПбГПУ.– СПб: Изд-во СПбГПУ, 2008.– №6(69).– С. 130-135.[<http://konstantin.solnushkin.ru>], 14.12.2008.
3. Солнушкин К.С. О значении терминов "производительность" и "быстродействие" в применении к ЭВМ / К.С.Солнушкин // Научно-технические ведомости СПбГПУ.– СПб.: Изд-во СПбГПУ, 2008.– №3(59).– С. 297-300
4. Солнушкин К.С. Постановка задачи синтеза кластерных ЭВМ, оптимальных по критерию цены производительности // Труды XV Всероссийской научно-методической конференции Телематика-2008. СПб.: Изд-во СПбГУ ИТМО, 2008.
5. Черкесов, Г.Н. Вероятностные модели производительности контролируемых вычислительных систем: Учеб. пособие / Г. Н. Черкесов; Ленинградский политехнический институт им. М.И.Калинина.– Л.: ЛПИ, 1983.– 79 с.
6. Douglas M. Pase, Hari Reddy. Tuning The Performance of ANSYS FLUENT on IBM BladeCenter HS21 and HS21 extended memory. February 2007. [ftp://ftp.software.ibm.com/eserver/benchmarks/wp_Fluent_021307.pdf], 14.12.2008.
7. FLUENT Performance Benchmarks. [<http://www.fluent.com/software/fluent/benchmarks.htm>], 14.12.2008.
8. FLUENT. Exterior Flow Around a Passenger Sedan. [http://www.fluent.com/software/fluent/fl5bench/flbench_6.3/problems/fl512.htm], 14.12.2008.
9. İpek, E., McKee, S., de Supinski, B., Schulz, M., Caruana, R. Efficiently exploring architectural design spaces via predictive modeling // Proceedings Architectural Support for Programming Languages and Operating Systems (ASPLOS XII), October 2006. ACM Press: New York, NY.
10. Joshua J. Yi, David J. Lilja, "Simulation of Computer Architectures: Simulators, Benchmarks, Methodologies, and Recommendations" // IEEE Transactions on Computers, vol. 55, no. 3, pp. 268-280, Mar. 2006
11. Kenneth Hoste, Aashish Phansalkar, Lieven Eeckhout, Andy Georges, Lizy K. John and Koen De Bosschere. Performance prediction based on inherent program similarity // Proceedings of the 15th International conference on Parallel architectures and compilation techniques (PACT'2006), p. 114-122.
12. Strohmaier, E. Statistical performance modeling: Case study of the NPB 2.1 results // Proceedings of Euro-Par'97 Parallel Processing, p. 985-992.