# Раздел VII. Технические средства защиты информации

УДК 004.382.2

**А.И. Дордопуло, И.И. Левин, Д.А. Сорокин**

## ОПТИМИЗАЦИЯ ВЫЧИСЛИТЕЛЬНОЙ СТРУКТУРЫ ЗАДАЧ С ПЕРЕМЕННОЙ ИНТЕНСИВНОСТЬЮ ПОТОКОВ ДАННЫХ НА РЕКОНФИГУРИРУЕМЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ*

*Статья посвящена применению методов оптимизации и адаптации архитектуры реконфигурируемой вычислительной системы под структуру задач с переменной интенсивностью потоков данных, к числу которых относятся задачи криптографии, молекулярного конструирования лекарств, транскодирования видеоизображений. Отличительной особенностью описываемого решения по сравнению с известными реализациями является функционально завершенное решение полной задачи, обеспечивающее согласованность функционирования всех фрагментов в едином вычислительном контуре.*

*Аппаратная реализация; докинг; реконфигурируемые вычислительные системы; криптография.*

**A.I. Dordopulo, I.I. Levin, D.A. Sorokin**

## OPTIMIZATION OF COMPUTING STRUCTURE OF TASKS WITH VARIABLE DATA FLOW DENSITY ON RECONFIGURABLE COMPUTER SYSTEMS

*The paper is devoted to use of methods of optimization and adaptation of architecture of a reconfigurable computer system to the structure of the solving task with variable data flow density like tasks of cryptography, molecular drug design, image transcoding. In comparison with existing realizations, the distinctive feature of the viewed solution is all-in-one solution of complete problem of docking on reconfigurable computer system, providing coordinated functioning of all fragments of the task in a single computer system.*

*Hardware realization; docking; reconfigurable computer systems; cryptography.*

**Введение.** Необходимость качественного решения вычислительно-трудоемких задач в различных областях науки и техники, таких как информационная безопасность, криптоанализ, молекулярное моделирование, требует применения новых подходов к повышению производительности реконфигурируемых вычислительных систем. Основная проблема при решении таких задач на РВС состоит в том, что размер потока данных в их вычислительной структуре заранее не определен и зависит от самих обрабатываемых данных. Так, для задач криптоанализа размер потока данных в разных местах вычислительной структуры отличается до десяти раз, а для задач молекулярного моделирования – на 2–4 десятичных порядка.

Рассмотренные в статье [1] методы оптимизации фрагментов задач с переменной интенсивностью потоков данных и средства адаптации архитектуры реконфигурируемой вычислительной системы (РВС) под структуру решаемой задачи обеспечивают согласованную аппаратную реализацию всех фрагментов задачи в едином вычислительном контуре.

* Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации.

Применение этих методов (редукция вычислительной структуры графа фрагмента задачи, использование предвычисленных массивов для хранения результатов повторяющихся операций, согласованное распараллеливание подграфов графа задачи, использование специальной структуры хранения данных) позволяет получить качественно новое решение задач с существенно переменной интенсивностью потоков данных, обеспечивающее многократный выигрыш (не менее 10 раз) по скорости решения задачи на РВС по сравнению с вычислительными системами традиционной архитектуры при пересчете на один процессор. Рассмотрим оптимизацию фрагмента задачи с переменной интенсивностью потоков данных на примере задачи молекулярного моделирования – докинга [2, 3].

**Оптимизация вычислительно-трудоемких фрагментов расчета внутренней энергии лиганда задачи докинга.** Внутренняя энергия лиганда $E_{inner}$ учитывает энергию Ван-дер-Ваальса $E_{lig-VdW}$, электростатическую энергию $E_{lig-ES}$ и энергию торсионного взаимодействия лиганда $E_{lig-tors}$ [2]. Расчет внутренней энергии лиганда состоит из разнородных по плотности потока обрабатываемых данных процедур.

Так, расчет $E_{lig\text{-}tors}$ выполняется однократно, время обработки одного лиганда зависит от числа торсионных связей $N_{tors}$, времени на обработку одной связи $t_{tors}$ и вычисляется по формуле

$$t_{Etors} = N_{tors} \times t_{tors}.$$

Согласно приведенной в работе [2] математической модели, максимальное число $N_{\text{tors}}$ не превысит 500, время обработки одной связи $t_{\text{tors}}$=1 такту работы ПВМ. Тогда можно определить время обработки для расчета торсионной энергии при структурной реализации $E_{\text{lig-tors}}$, которое составит $t_{\text{Etors}}$=500 тактов. Полученное значение не превышает значение времени обработки для фрагментов *R* и *RT*, поэтому структурная реализация процедуры расчета $E_{lig\text{-}tors}$ не требует распараллеливания для согласованной работы в едином вычислительном контуре. Вместе с тем, необходимый для её реализации аппаратный ресурс также был сокращен при помощи методов оптимизации 1 и 2, что позволило сократить число устройств с 68 до 51.

При вычислении значений энергии Ван-дер-Ваальса $E_{lig\text{-}VDW}$ и электростатического взаимодействия $E_{lig\text{-}ES}$ используется принцип суперпозиции полей [2, 3], когда для текущего атома необходимо учитывать влияние всех остальных атомов лиганда, в результате чего многократно (до двух десятичных порядков) возрастает поток промежуточных данных. Время расчета значений этих энергетических составляющих выполняется за время, определяемое по формуле:

$$t_{Evdw\_es} = \frac{N_{atom}}{2} \cdot (t_{vdw} + t_{es}), \quad (3)$$

где $t_{vdw}$ – время расчета одного прохода $E_{lig\text{-}VDW}$, $t_{es}$ – время расчета одного прохода $E_{lig\text{-}ES}$.

Значения $E_{lig\text{-}VDW}$ и $E_{lig\text{-}ES}$ зависят от одной общей вычисляемой переменной, поэтому целесообразно объединить структурные реализации одной ступени расчета $E_{lig\text{-}VDW}$ и $E_{lig\text{-}ES}$ в единый фрагмент, чтобы сократить вдвое время обработки. Формула (3) в этом случае примет вид

$$t_{Evdw\_es} = \frac{N_{atom}}{2} \cdot t_{vdw\_es} = \frac{N_{atom} \cdot (N_{atom} - 1)}{2} \cdot t_a, \quad (4)$$

где $t_a$ – время обработки одной пары атомов, которое составляет 1 такт работы ПВМ.

Для лиганда максимального размера с числом атомов $N_{atom}$=200 время расчета электростатической $E_{lig\text{-}ES}$ и Ван-дер-Ваальсовой $E_{lig\text{-}VDW}$ составляющих общей энергии составит 19 900 тактов, что существенно превышает максимальное из достигнутых при реализации фрагментов $R$, $RT$, $E_{\text{lig-prot}}$ значение в 950 тактов и требует распараллеливания для согласования скорости обработки информации разных фрагментов задачи в едином вычислительном контуре. Степень распараллеливания $n$ в данном случае можно определить как

$$n = \frac{t_{Evdw\_es}}{t_{\ddot{I}}}, \tag{5}$$

где$t_{\Pi}$ – наибольшее время выполнения других фрагментов задачи, равное 950 тактам. При $t_{Evdw\_es}$=19 900 и $t_{Elig\text{-}prot}$=950 необходимая степень распараллеливания $n$ составит примерно 21.

При определении необходимой степени распараллеливания должны учитываться и ограничения по занимаемому ресурсу, которые определяются как

$$\lim_{t_{Evdw\_es} \to t_{\ddot{I}}} \frac{V(n)}{V_0} \le 1, \tag{6}$$

где $V_0$ – объем свободного аппаратного ресурса, $V(n)$ – объем аппаратного ресурса, необходимого для построения вычислительной структуры $E_{lig\text{-}VDW\_ES}$ с учетом распараллеливания.

Число устройств, необходимых для реализации одной ступени фрагмента $E_{lig\text{-}VDW\_ES}$, равно 33, поэтому при $n$=21 получим $V_{\min}(n)$=693 устройства.

Свободный аппаратный ресурс ПВМ 16V5-75 для реализации вычислительной структуры $E_{lig\text{-}VDW\_ES}$ после размещения вычислительных структур $R$, $RT$ и $E_{lig\text{-}prot}$ для реализации вычислительной структуры $E_{lig\text{-}VDW\_ES}$ можно оценить по формуле

$$V_0 = c \times V_p - V_g - V_R - V_{RT} - V_{Etors} - V_{Elig-prot},$$

где$V_p$ – объем ресурса, эквивалентный всему доступному ресурсу платформы;

$c$ – коэффициент, характеризующий долю затрат на блоки объединения и согласования потоков данных между вычислительными структурами и внутри них, для рассматриваемого случая составляет 0,71;

$V_g$ – объем ресурса на реализацию фрагментов *GEN* и *MPS*; $V_g$=95;

$V_R$ – объем ресурса на реализацию фрагмента $R$; $V_R$ =440;

$V_{RT}$ – объем ресурса на реализацию фрагмента $RT$; $V_{RT}$=17;

$V_{Etors}$ – объем ресурса на реализацию фрагмента $E_{lig\text{-}tors}$; $V_{Etors}$=51;

$V_{Elig\text{-}prot}$ – объем ресурса на реализацию фрагмента $E_{lig\text{-}prot}$; $V_{Elig\text{-}prot}$=51.

С помощью несложных вычислений получим $V_0$=438 устройств.

Поскольку условие (6) не выполняется, то с помощью методов 2 и 3 объем затрачиваемого на реализацию вычислительной структуры $E_{lig\text{-}VDW\_ES}$ аппаратного ресурса сократим в 1,5 раза – с 33 до 21 устройства. Затем, поскольку дальнейшая редукция вычислительного графа невозможна, степень распараллеливания необходимо сократить до значения $n$=20, что обеспечит выполнение условия (6). Время обработки лиганда с числом атомов $N$=200 в этом случае составит

$$t'_{Evdw\_es} = \frac{t_{Evdw\_es}}{n} = 995 \text{ тактов.}$$

Таким образом, имеющиеся в структуре задачи фрагменты по скорости обработки данных можно разделить на две группы:

- фрагменты $R$, $RT$ и $E_{\text{lig-tors}}$ (не более 600 тактов);
- фрагменты $E_{lig\text{-}VDW\_ES}$, $E_{lig\text{-}prot}$ (не более 1000 тактов).

Для согласования скорости обработки между этими группами наиболее простым решением, которое и было реализовано, явилось замедление скорости входного потока данных в соответствии со скоростями обработки на самом медленном участке задачи $E_{lig\text{-}VDW\_ES}$.

**Фрагменты задачи докинга в едином вычислительном контуре.** В результате была синтезирована структура вычислительного конвейера, обеспечивающего полное решение задачи на PBC, представленная на рис. 3.

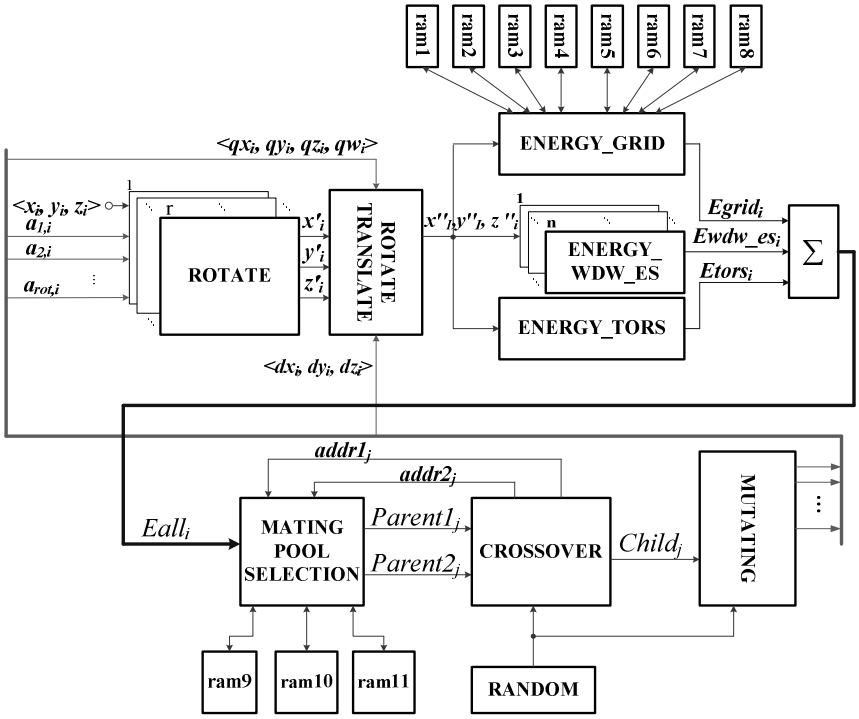


*Рис. 3. Структура вычислительного конвейера*

Разработанная вычислительная структура с учетом архитектурных особенностей ПВМ 16V5-75 была отображена на структуру вычислительного поля 16V5-75, результат отображения представлен на рис. 4.

Общий объем задействованного оборудования на построение вычислительных блоков составил 1 078 устройств, реализующих 32-разрядные математические операции в стандарте IEEE754. В результате применения методов оптимизации было достигнуто 4-кратное сокращение объема используемого оборудования при сохранении заданной скорости решения задачи.

**Результаты вычислительных экспериментов.** Для проверки, оценки и обобщения результатов реализации задачи докинга на PBC был проведен ряд вычислительных экспериментов для четырех тестовых выборок, две из которых содержат молекулы с одинаковым числом атомов (48 и 52) и разным количеством торсионных степеней свободы, третья выборка состоит из молекул с одинаковым количеством торсионных степеней свободы, равным 7, но разным числом атомов, а четвертая выборка состоит из молекул с большим количеством атомов (до 198) и большим числом торсионных степеней свободы (до 18) [5, 6].

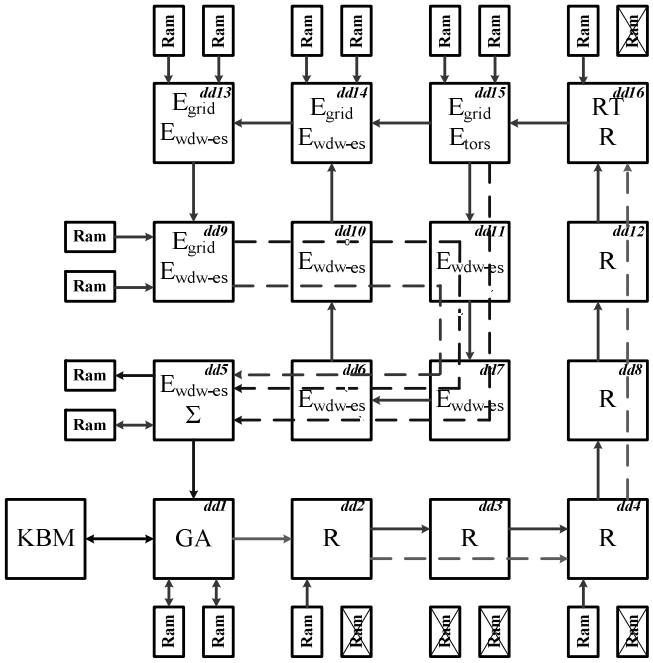


*Рис. 4. Отображение основных фрагментов задачи докинга на структуру ПВМ 16V5-75*

Оценка проводилась сравнением времени выполнения задачи молекулярного докинга на нескольких вычислительных системах:

- на персональных компьютерах, оснащенных процессорами Intel Core2Duo 3 ГГц, оперативной памятью объемом 2 Гбайт;
- одном стандартном вычислительном узле суперкомпьютера СКИФ МГУ «Чебышев» (2 четырехъядерных процессора Intel Xeon E5472 3.0 ГГц, имеющих общую оперативную память объемом 8 Гбайт), (НИВЦ МГУ имени М.В. Ломоносова);
- нескольких (до 32 процессоров) вычислительных узлах суперкомпьютера СКИФ МГУ «Чебышев» (НИВЦ МГУ имени М.В. Ломоносова);
- ПВМ 16V5-75.

Для экспериментов по сравнению времени работы программы на РВС и кластерной системе стандартной архитектуры использовалась параллельная версия программы докинга SOL.

При докинге молекул тестовых выборок на РВС (ПВМ 16V5-75) была достигнута реальная производительность 125,7 Гфлопс, что составляет 89,8 % от пиковой производительности.

В результате выигрыш ПВМ 16V5-75 во времени докинга для одного поколения из 30 000 особей по сравнению с Intel Core2Duo 3 ГГц в зависимости от числа атомов и торсионных связей составил от 35 до 90 раз; выигрыш по сравнению со стандартным вычислительным узлом суперкомпьютера СКИФ МГУ «Чебышев» – от 6 до 15 раз; по сравнению с 32-х процессорами суперкомпьютера СКИФ МГУ «Чебышев» – от 3 до 5 раз.

**Заключение.** Особенности организации вычислений в параллельной версии программы докинга SOL приводят к интенсивному обмену MPI-сообщениями, в результате чего на скорость решения задачи, в основном, начинает влиять латент-

ность MPI-системы, а не скорость передачи данных. Для числа процессов больше, чем 30, время, затрачиваемое на синхронизацию, оказывается сравнимым со временем, затрачиваемым на расчеты, поэтому оптимальное число процессоров для одного независимого запуска программы составляет 32, а максимальное значение ускорения, достигаемое на 32-х процессорах, – 5 раз. Дальнейшее увеличение количества процессоров не повлияет на ускорение одного независимого запуска, а с учетом дополнительной нагрузки на MPI-сеть в ходе обмена информацией между процессами будет наблюдаться даже замедление работы программы.

Значение ускорения выполнения программы на РВС существенно зависит от размера рассчитываемой молекулы: чем больше количество атомов и торсионных степеней свободы, тем значительнее выигрыш во времени решения задачи на РВС по сравнению с вычислительными системами традиционной архитектуры. Это связано, прежде всего, с гораздо более эффективным способом параллельной организации вычислений и меньшими накладными расходами на организацию вычислительного процесса на специализированных вычислительных системах, чем на кластерах стандартной архитектуры. ПВМ 16V5-75 обеспечивает многократный выигрыш в скорости докинга, доходящий до 70 раз для одного поколения популяции, который практически линейно увеличивается с увеличением числа атомов в лиганде и числа торсионных связей до их максимальных значений.

Полученные экспериментальные результаты позволяют сделать вывод о том, что использование РВС для решения задач с переменной интенсивностью потоков данных обеспечивает существенное ускорение: не менее чем на один десятичный порядок при пересчете на один процессор по сравнению с вычислительными системами традиционной архитектуры. Таким образом, применение РВС для таких задач, как молекулярный докинг, линейный или дифференциальный анализ стойкости криптографических систем, требующих обработки больших потоков данных с переменной интенсивностью, позволяет существенно сократить как время решения, так и материальные затраты.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Левин И.И., Дордопуло А.И., Сорокин Д.А.* Реализация докинга для молекулярного моделирования на реконфигурируемых вычислительных системах // Известия ЮФУ. Технические науки. – 2011. – № 7 (120). – С. 217-224.
2. Молекулярная стыковка: http://ru.wikipedia.org/wiki/Докинг.
3. *Романов А.Н., Кондакова О.А., Григорьев Ф.В. и др.* Компьютерный дизайн лекарственных средств: программа докинга SOL // Вычислительные методы и программирование. – 2008. – Т. 9. – С. 213-233.
4. Генетический алгоритм http://ru.wikipedia.org/wiki/Генетические_алгоритмы.
5. AutoDock. http://autodock.scripps.edu/.
6. *Каляев И.А., Левин И.И., Семерников Е.А., Шмойлов В.И.* Реконфигурируемые мультиконвейерные вычислительные структуры. – 2-е изд., перераб. и доп. / Под общ. ред. И.А. Каляева. – Ростов-на-Дону: Изд-во ЮНЦ РАН, 2009. – 344 с.

Статью рекомендовал к опубликованию д.т.н., профессор Я.Е. Ромм.

**Левин Илья Израилевич**
Научно-исследовательский институт многопроцессорных вычислительных систем им. академика А.В. Каляева федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет».
E-mail: levin@mvs.tsure.ru.
347922, г. Таганрог, ул. Ленина, д. 224/1, кв. 65.
Тел.: 88634623226.
Зам. директора по науке; д.т.н.

**Сорокин Дмитрий Анатольевич**
347922, г. Таганрог, пер. Украинский, д. 21, кв. 30.
E-mail: jotun@inbox.ru.
Тел.: 88634393820.
Научный сотрудник.

**Дордопуло Алексей Игоревич**
Учреждение Российской академии наук «Южный научный центр РАН».
E-mail:scorpio@mvs.tsure.ru.
347900, г. Таганрог, 10-й переулок, 114/1, кв. 6.
Тел.: 88634368651.
К.т.н.; с.н.с. отдела ИТ и ПУ.

**Levin Ilya Israilevich**
Kalyaev Scientific Research Institute of Multiprocessor Computer Systems at Southern Federal University.
E-mail: levin@mvs.tsure.ru.
224/1, Lenin Street, Ap. 65, Taganrog, 347922, Russia.
Phone: +78634623226.
Deputy Director of Science; Dr. of Eng. Sc.

**Sorokin Dmitry Anatolievich**
E-mail: jotun@inbox.ru.
21, Ukrainskiy Lane, Ap. 30, Taganrog, 347922, Russia.
Phone: +78634393820.
Scientific Associate.

**Dordopulo Alexey Igorevich**
Southern Scientific Centre of the Russian Academy of Sciences.
E-mail: scorpio@mvs.tsure.ru.
114/1, 10th Lane, Ap. 6, Taganrog, 347900, Russia.
Phone: +78634368651.
Senior Staff Scientist; Cand. of Eng. Sc.

УДК 004.08

**А.М. Максимов, Е.Н. Тищенко**

# ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ НОСИТЕЛЕЙ ИНФОРМАЦИИ В ЗАЩИЩЁННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

*В современных защищённых информационных системах общий уровень защищенности определяется уровнем защищенности самого слабого звена. Одним из таких звеньев являются накопители данных. Распространение, удешевление и увеличение объёмов накопителей вынуждает делать больше затрат для контроля за носителями данных, чтобы сохранить защищённое состояние информационной системы в целом. Дополнительные проблемы в этом направлении создаются с появлением, развитем и распространением накопителей данных, созданных по новым технологиям. К таким накопителям уже мало применимы, или же в принципе не применимы подходы, использующиеся в настоящее время, что требует поиска новых решений по защите информационных систем.*

*Информационная система; накопители данных; магнитные накопители данных; твердотельные накопители данных; программно-техническая экспертиза.*