

Программа для сбора и анализа служебных данных работы квантовохимического пакета «ПРИРОДА»

Е.Ю. Панкратьев

Институт нефтехимии и катализа РАН, Уфа

1. Введение

Квантовохимический пакет «ПРИРОДА» хорошо зарекомендовал себя для исследования сложных молекулярных систем методом функционала плотности (DFT), а также методом теории возмущений Мёллера-Плессе второго порядка (MP2) [1]. Используемый в программе метод ускорения предоставляет возможность проводить расчёт больших металлокомплексных систем со сравнительно низкими вычислительными затратами [2, 3]. В алгоритм программы «ПРИРОДА» изначально была заложена возможность распараллеливания вычислений, что позволило использовать квантовохимический пакет как на персональных ЭВМ, так и на суперкомпьютерах.

Следует отметить, что проведение расчёта металлокомплексных систем квантовохимическими методами высокого уровня (например, MP2) в большинстве случаев уже невозможно на персональных ЭВМ [4]. Т.е. в качестве вычислительной системы возможно использование лишь суперкомпьютеров. Учитывая их высокую стоимость, возникает проблема либо выбора конфигурации при проектировании своего суперкомпьютера, либо оценки требуемых ресурсов в случае аренды процессорного времени. Программа «ПРИРОДА», как и большинство квантовохимических пакетов, до окончания расчёта не информирует пользователя о необходимых ресурсах, а в случае их нехватки происходит прекращение выполнения программы и расчёты во многих случаях требуется производить заново. В связи с этим актуальной задачей является разработка программы оценки требуемых вычислительных ресурсов для решаемых пользователем квантовохимических задач и составления отчётов об уже проведённых расчётах.

2. Постановка задачи

В целях эффективной организации квантовохимических расчётов и поддержания распределённой работы над задачей химического моделирования нескольких сотрудников в лаборатории структурной химии Института нефтехимии и катализа РАН разработана система обозначения файлов-заданий (входных файлов, IN) и выходных файлов (OUT) для квантовохимического пакета «ПРИРОДА».

Путь к директории, в которой по завершению квантовохимического расчёта размещаются для хранения файлы, строится по следующему алгоритму:

[Химическая_задача]\[Уникальное_соединение]\[Квантовохимическая_задача]

Здесь [Уникальное_соединение] обозначает, что для каждой не рассчитанной ранее структуры создаётся новая директория.

Имя файла расчёта составляется по следующему алгоритму:

[Название_соединения][Обозначение_конформации]_[Метод_расчёта]_[Тип_расчёта]

Например, [Метод_расчёта] может быть PBE_3z, riMP2_L1 и др. [Тип_расчёта] определяет последняя буква в имени файла, если это "o", то это задача оптимизации структуры молекулы; "h" – расчёт колебательной задачи; "n" – расчёт химических сдвигов ЯМР; "s" – последовательная оптимизация с несколькими замороженными геометрическими параметрами, значение которых постепенно наращивается (сканирование); "i0", "i1" – сканирование по координате реакции (спуск в сторону реагентов и продуктов).

Пример дерева директорий:

```
01_Task01\01_StructureA\1.optimize\input\structureAconf1_PBE_3z_o.in
01_Task01\01_StructureA\1.optimize\input\structureAconf2_PBE_3z_o.in
01_Task01\01_StructureA\1.optimize\structureAconf1_PBE_3z_o.out
```

```

01_Task01\01_StructureA\1.optimize\structureAconf2_PBE_3z_o.out
01_Task01\01_StructureA\2.vibration\input\structureAconf1_PBE_3z_h.in
01_Task01\01_StructureA\2.vibration\structureAconf1_PBE_3z_h.out
01_Task01\01_StructureA\3.nmr\input\structureAconf1_PBE_3z_n.in
01_Task01\01_StructureA\3.nmr\structureAconf1_PBE_3z_n.out
01_Task01\02_StructureB\1.optimize\input\structureB_PBE_3z_o.in
01_Task01\02_StructureB\1.optimize\structureB_PBE_3z_o.out

```

...

Когда расчёты для уникального соединения завершаются, дерево директорий [Уникальное_соединение] архивируется, сжимается и размещается в директории [Химическая_задача]. Когда все расчёты химической задачи решены, директория [Химическая_задача], содержащая архивы с расчётами также архивируется и сжимается.

Таким образом, в соответствии с нашими требованиями, утилита для сбора информации служебных данных квантовохимического пакета «ПРИРОДА» должна рекурсивно производить обход директорий, включая архивы, начиная с некоторой заданной пользователем директории. Внутри директории утилита должна находить OUT-файлы и если есть, соответствующие им IN-файлы.

Пример IN-файла:

```

$system memory=256 disk=-512 path=/tmp $end
$control theory=DFT basis="3z.bas" task=optimize $end
$dft functional=PBE $end
$optimize steps=200 $end

```

```
$molecule charge=0 mult=1 z-matrix
```

```
1
```

```
1 1 0.751
```

```
$end
```

Служебная информация в OUT-файле:

```

Priroda 6 (2006.08.20)
copyright (c) Dimitri Laikov

```

```
date: Mon Sep 14 22:30:06 2009
```

```
parallel execution on 4 processors
```

```
node 0: 'quant'
```

```
node 1: 'quant'
```

```
node 2: 'quant'
```

```
node 3: 'quant'
```

...

```

+-----+
| Number of atoms                2
| Charge of molecule             0
| Spin multiplicity              1
| Number of alpha electrons      1
| Number of beta electrons       1
| Number of core electrons       0
| Theoretical Method             DFT
| Approximation to E(xc)         PBE
| Dimension of wavefunction basis 12
| Dimension of auxiliary basis   22
| Hamiltonian: non-relativistic

```

```

+-----
...
=====

Memory used =      5384 KB
Memory used =      2004 KB
Memory used =       348 KB
Memory used =       348 KB
  Disk used =      5573 KB,          999 KB written,
12271 KB read
  Disk used =          0 KB,          0 KB written,
0 KB read
  Disk used =          0 KB,          0 KB written,
0 KB read
  Disk used =          0 KB,          0 KB written,
0 KB read
Message passing statistics:
node      sent KB    # msg    time          received      time
  0         11583     9751     0.1           707      3112     0.0
  1           691     3012     0.0          11561     9635     0.2
  2           651     2941     0.0          11516     9555     0.1
  3           651     2941     0.0          11516     9555     0.2

CPU  time =      0.46 sec = 0.01 min = 0.00 hr =   85.87%
CPU  time =      0.59 sec = 0.01 min = 0.00 hr =  110.51%
CPU  time =      0.49 sec = 0.01 min = 0.00 hr =   91.84%
CPU  time =      0.57 sec = 0.01 min = 0.00 hr =  106.03%
REAL time =      0.54 sec = 0.01 min = 0.00 hr
ratio      =  394.26%

date: Mon Sep 14 22:30:07 2009

```

Утилита должна считывать служебные данные выполнения программы и сохранять в базу данных.

3. Технические возможности

Программа реализована на языке сценариев PERL, для хранения данных используется реляционная СУБД SQLite. Для разархивации TAR, GZIP, BZIP2 и ZIP используются модули PERL, для RAR может быть использована внешняя программа. Программа реализована в виде двух подпрограмм: первая предназначена для произведения обхода директорий и файлов, и сбора данных; вторая предназначена для извлечения различных выборок из базы данных.

В базу данных записываются следующие параметры:

- Параметры молекулярной системы (брутто формула, количество атомов, количество электронов, спиновое число);
- Решаемая квантовохимическая задача (например, расчёт энергии молекулярной системы, оптимизация структуры, решение колебательной задачи и т.д.);
- Метод расчёта (используемая теория, для DFT метода функционал, используемый файл базисного набора, обозначение базиса, количество волновых функций основного и вспомогательного базиса для описания молекулярной системы, используе-

мый типа гамильтониана – нерелятивистский или релятивистский, тип решения SCF-задачи);

- Время запуска и завершения расчёта;
- Количество использованных процессоров;
- Затраченное процессорное время (реальное и с учётом простоя на операции ввода-вывода, коэффициент полезного использования процессоров);
- Количество использованной для расчёта оперативной и дисковой памяти (общее потребление, максимальное потребление на расчётный узел, в среднем на расчётный узел);
- Количество и размер отправленных и полученных сообщений MPI;

Извлечение данных производится с помощью задания в конфигурационном файле условия для SQL-запроса (например, “parameter1=value1 AND parameter2=value2”). Задан набор предустановленных конфигурационных файлов (например, для выборки всех расчётов только DFT методом, для расчёта суммы полного времени DFT расчётов и т.п.).

4. Применение утилиты

Возможны следующие приложения утилиты.

1. Оценка использованного процессорного времени для проведённых расчётов.

Например, было оценено, что для расчётов [5] было затрачено 1358.8 процессорочасов компьютера с конфигурацией Intel Core Quad Q6600 / 2ГБ ОЗУ / 250 ГБ RAID0 НЖМД, а для [6] – 911.2 процессорочасов.

2. Оценка вклада различных методов или типов задач в общее время.

Например, для [6] вклад DFT-расчётов составляет 14.3 процессорочасов, а одиночных MP2- и MP4-расчётов – 896.9 процессорочасов.

3. Так как потребляемые квантовохимической программой ресурсы не связаны с конкретными особенностями молекулярной системы, а определяются лишь используемым квантовохимическим методом и количеством базисных функций, необходимых для задания электронной структуры, то утилита может быть использована для оценки затрат новых, ещё не проведённых расчётов, на основе информации о схожих расчётах, присутствующих в базе данных.

4. Известны классические закономерности, показывающие рост вычислительных затрат в зависимости от роста размера молекулярной системы (для DFT $\sim O(N^3)$, для MP2 $\sim O(N^5)$). Однако применяемые в последнее время ускорения позволяют снизить степень зависимости роста ресурсов от роста размера молекулярной системы. Утилита может быть использована для составления таблицы, пригодной для аппроксимации затрат полиномом, с целью более удобного предсказания затрат.

Программу можно скачать по ссылке [7].

Благодарности

Авторы выражают благодарность администрации Уфимского государственного авиационного технического университета и лично Газизову Р. К., а также ООО «Ниагра Компьютерс» и лично Соколову А. за предоставленный доступ к вычислительным ресурсам.

Литература

1. Лайков Д. Н., Устынюк Ю. А. Система квантово-химических программ «ПРИРОДА-04». Новые возможности исследования молекулярных систем с применением параллельных вычислений. // Известия РАН, Серия химическая.– 2004.– V. 3.– N. P. 804-810.

2. Laikov D. N. Fast evaluation of density functional exchange-correlation terms using the expansion of the electron density in auxiliary basis sets. // *Chemical Physics Letters*.– 1997.– V. 281.– P. 151-156
3. Лайков Д.Н., Развитие экономного подхода к расчёту молекул методом функционала плотности и его применение к решению сложных химических задач. Дисс. ... к.х.н., Москва, МГУ, 2000, 102 с.
4. Панкратьев Е. Ю., Тюмкина Т. В., Хурсан С. Л., Губайдуллин И. М., Применение суперкомпьютеров в исследовании механизмов реакций металлокомплексного катализа. Параллельные вычислительные технологии (ПаВТ'2009): Труды международной научной конференции (Нижний Новгород, 30 марта - 3 апреля 2009 г.), Издание ЮУрГУ, Челябинск, 2009, С. 631-638.
5. Pankratyev E. Y., Tyumkina T. V., Parfenova L. V., Khalilov L. M., Khursan S. L., Dzhemilev U. M. DFT Study on Mechanism of Olefin Hydroalumination by XAlBu^i_2 in the Presence of Cp_2ZrCl_2 Catalyst. I. Simulation of Intermediate Formation in Reaction of HAlBu^i_2 with Cp_2ZrCl_2 . // *Organometallics*.– 2009.– V. 28.– №. 4.– PP. 968-977.
6. Панкратьев Е. Ю., Тюмкина Т. В., Хурсан С. Л., Парфёнова Л. В., Халилов Л. М., Джемилев У. М. Квантовохимическое исследование самоассоциации в триалкилаланах AlR_3 ($\text{R} = \text{Me}, \text{Et}$). // *Вестник Башкирского университета*.– Раздел «Химия».– 2008.– Т. 13.– №. 3(I).– С. 802-809.
7. <http://nwrr.ru/projects/util4qchem.html>