Инженерный вестник Дона, №7 (2023)
ivdon.ru/ru/magazine/archive/n7y2023/8577

Оптимизация пакета Quantum Espresso для работы на GPU Nvidia с помощью технологии CUDA

Д.Л. Абдрахманов, Д.Н. Жариков, О.И. Жарикова, Д.В. Завьялов

Волгоградский государственный технический университет»

Аннотация: В данной статье исследуется процесс оптимизации пакета Quantum Espresso для эффективного использования графического процессора (GPU) от Nvidia с помощью технологии CUDA. Quantum Espresso является мощным инструментом для квантово-механического моделирования и расчета свойств материалов. Однако, оригинальная версия пакета не была разработана для использования на GPU, поэтому требуется оптимизация для достижения наилучшей производительности.
Ключевые слова: Quantum Espresso, GPU, CUDA, ускорение вычислений.

1. Введение

В последние десятилетия современная наука и технологии достигли значительного прогресса в области квантово-механического моделирования и расчета свойств материалов. Одним из важных инструментов для этих расчетов является пакет программного обеспечения Quantum Espresso [1]. Он предоставляет средства для выполнения сложных квантово-механических расчетов, включая электронные структуры, оптические свойства, термодинамические характеристики и многое другое.

С появлением графических процессоров (GPU) и развитием технологии параллельных вычислений, возникла возможность эффективно использовать GPU для ускорения научных расчетов. Графические процессоры от Nvidia с их высокой вычислительной мощностью и параллельными архитектурами стали особенно популярными в научном сообществе.

Quantum Espresso, как и многие другие программные пакеты, изначально разрабатывался для выполнения на центральных процессорах (CPU) и не был оптимизирован для работы на GPU. Однако, благодаря технологии вычислений общего назначения (GPGPU) и инструментарию CUDA от Nvidia, стало возможно оптимизировать Quantum Espresso для работы на GPU Nvidia и повысить его производительность.

2. Установка необходимых инструментов CUDA и Nvidia HPC SDK

Для начала, требуется установить пакет CUDA от Nvidia, который предоставляет средства для разработки программного кода, оптимизированного для работы на GPU [2, 3].

После успешной установки CUDA Toolkit, необходимо также установить Nvidia HPC SDK [4], который предоставляет компиляторы и другие инструменты для разработки программного обеспечения, оптимизированного для работы на GPU Nvidia [5].

3. Настройка компиляции Quantum Espresso

Основная часть исходного кода Quantum Espresso написана на языке Fortran, поэтому важно указать путь, по которому система сможет найти компилятор Fortran. На рис.1 указаны пути для версии Nvidia HPC 23.1.

```
F90=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/compilers/bin/pgf90
CC=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/compilers/bin/pgcc
F77=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/compilers/bin/pgf77
```

Рис. 1. – Путь к компилятору Fortran, пакета Nvidia HPC SDK

Кроме того, для предотвращения ошибок в сборке необходимо изменить файл `.bashrc` пользователя, добавив формулировку, представленную на рис.2.

```
export PATH="/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/compilers/bin/:$PATH"
```

Рис. 2. – Содержимое файла пользователя .bashrc

4. Указание параметров CUDA и версии библиотек

Quantum Espresso также требует указания параметров CUDA и версии библиотек для корректной работы на GPU Nvidia.

На рис.3 представлены параметры CUDA в команде конфигурации.

```
--with-cuda=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/cuda/ --with-cuda-runtime=11.0 --with-cuda-cc=5.0
```

Рис. 3. – Параметры CUDA и версии библиотек

5. Использование MPI и интеграция с GPU Nvidia

Quantum Espresso использует интерфейс MPI для параллельного выполнения задач. Для интеграции с GPU Nvidia необходимо явно указать используемую версию MPI и использовать реализацию от Nvidia, вместо системного MPI [6].

Параметры MPI в команде конфигурации показаны на рис.4. Мы отключим использование системного MPI и будем использовать реализацию от компании Nvidia, которая может работать с GPU. Это позволит Quantum Espresso использовать параллельные возможности GPU Nvidia.

```
--enable-openmp --disable-parallel --with-cuda-mpi=yes
```

Рис. 4. – Подключение MPI от Nvidia

6. Сборка и результаты

Итоговая формулировка команды конфигурации представлена на рис.5.

```
(base) citrullux@node47:~/q-e-qe-7.2$ ./configure F90=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/compilers/bin/pgf90
CC=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/compilers/bin/pgcc F77=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/compilers/
bin/pgf77 --with-cuda=/opt/nvidia/hpc_sdk/Linux_x86_64/23.1/cuda/ --with-cuda-runtime=11.0 --enable-openmp --dis
able-parallel
```

Рис. 5. – Команда конфигурации

После успешной настройки компиляции Quantum Espresso с поддержкой GPU Nvidia, можно перейти к процессу сборки пакета. Используя команду `makeall -j4`, можно запустить сборку на 4 потоках центрального процессора.

По завершении сборки, в директории `bin/` будут находиться собранные пакеты Quantum Espresso, готовые к использованию на вычислительных ресурсах (рис.6).

```
alpha2f.x               dvscf_q2r.x         gww.x               ld1.x                   ph.x          projwfc.x         q2r.x                   turbo_lanczos.x
average.x               dynmat.x            head.x              manycp.x                plan_avg.x    pw2bgw.x          rism1d.x                turbo_magnon.x
band_interpolation.x    epa.x               hp.x                manypw.x                plotband.x    pw2critic.x       scan_ibrav.x            turbo_spectrum.x
bands.x                 epsilon.x           ibrav2cell.x        matdyn.x                plotproj.x    pw2gw.x           simple_bse.x            wannier90.x
bse_main.x              ev.x                initial_state.x     molecularnexafs.x       plotrho.x     pw2wannier90.x    simple_ip.x             wannier_ham.x
cell2ibrav.x            fermi_proj.x        kcwpp_interp.x      molecularpdos.x         pmw.x         pw4gww.x          simple.x                wannier_plot.x
cppp.x                  fermi_velocity.x    kcwpp_sh.x          neb.x                   postahc.x     pwcond.x          spectra_correction.x    wfck2r.x
cp.x                    fqha.x              kcw.x               open_grid.x             ppacf.x       pwi2xsf.x         sumpdos.x               wfdd.x
dist.x                  fs.x                kpoints.x           path_interpolation.x    pprism.x      pw.x              turbo_davidson.x        xspectra.x
dos.x                   gww_fit.x           lambda.x            phcg.x                  pp.x          q2qstar.x         turbo_eels.x
```

Рис. 6. – Собранные пакеты Quantum Espresso

Данная конфигурация (рис.5) позволила успешно собрать пакет Quantum Espresso версии 6.8 для использования на вычислительных ресурсах кластера Волгоградского государственного технического университета [7-9].

7. Тестирование и анализ производительности

В целях измерения производительности была протестирована задача моделирования вещества методом молекулярной динамики [10]. Чтобы добиться успешного выполнения, систему требовалось привести к равновесному состоянию, из которого начинается моделирование вещества со свойствами, приближенными к реальным.

Из-за того, что метод Кара-Парринелло включает в себя динамику волновых функций, не достаточно просто найти оптимальные позиции атомов, нужно также минимизировать энергию электронов.

На самом первом этапе, который не входит в представленные ниже результаты, находим основное состояние электронного облака (состояние с минимальной энергией функционала Кона-Шема) с неподвижными ионами.

В дальнейшем на протяжении 7500 шагов симуляции рассматривается динамика ионов для нахождения их оптимальных позиций - при таких позициях силы воздействия на ионы будут минимальны. Для каждого шага динамики ионов один раз вычисляется приближенное значение основного состояния электронного облака.

Последний этап в ходе 5000 шагов включает те же действия, что были в предыдущем, но выполняется 10 итераций приближения в расчёте основного электронного состояния.

Результаты тестирования производительности обновленного пакета Quantum Espresso с поддержкой вычислений на GPU Nvidia представлены в таблице №1.

Мы запускали нашу тестовую задачу на разных аппаратных платформах.

Таблица № 1

Время расчета задачи релаксации ионов + электронов

Ресурсы	время CPU+GPU, мин	время Wall, мин
7500 шагов (1 шаг электронов за 1 шаг ионов)		
Xeon E5-2650v4 (24 п.)	397	425
Xeon Gold-6130 (16п.)	342	371
Xeon Gold-6130 (16п.)+Tesla V100	259	289
5000 шагов (10 шагов электронов за 1 шаг ионов)		
Xeon E5-2650v4 (24 п.)	265	284
Xeon Gold-6130 (16п.)	225	253
Xeon Gold-6130 (16п.)+Tesla V100	174	205

Использование GPU как сопроцессоров на отдельных вычислительных узлах дает нам прирост производительности и уменьшает время расчета нашей задачи в среднем на 25 процентов.

Заключение

Оптимизация Quantum Espresso для работы на GPU Nvidia с помощью технологии CUDA открывает новые возможности для ускорения квантово-механических расчетов, позволяет эффективно использовать возможности графических процессоров и повысить производительность расчетов.

Литература

1. User’s Guide for Quantum ESPRESSO URL: quantum-espresso.org/Doc/user_guide_PDF/user_guide.pdf (дата обращения: 23.04.2023)

2. Cheng John, Grossman Max, McKercher Ty Professional CUDA® C Programming. – John Wiley & Sons Inc., 2014.- 528 p.: ISBN: 9781118739327

3. Сидорович М.А. Применение технологии cuda для повышения скорости обработки эмпирических данных // Научно-техническое и экономическое сотрудничество стран АТР в XXI веке. 2021. Т. 1. С. 263-267.

4. NVIDIA HPC SDK Version 23.1 Documentation URL: docs.nvidia.com/hpc-sdk/archive/23.1/index.html (дата обращения: 23.04.2023)

5. Wen-mei Hwu, Kirk David, El Hajj Izzat Programming Massively Parallel Processors. A Hands-on Approach. 4th Edition. - Morgan Kaufmann, 2022. – 580 p.

6. Абдрахманов Д.Л., Жариков Д.Н., Завьялов Д.В. Средства ускорения выполнения задач с большим объёмом операций ввода/вывода в гетерогенной вычислительной системе. // Инженерный вестник Дона, 2022, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2022/7688/.

7. Андреев А.Е., Егунов В.А., Завьялов Д.В., Жариков Д.Н. О применении высокопроизводительных вычислений в фундаментальных исследованиях, прикладных и образовательных проектах ВолгГТУ // Параллельные вычислительные технологии (ПаВТ'2021). Короткие статьи и описания плакатов. XV международная конференция. Челябинск, 2021. С. 131-142.

8. Андреев А.Е., Егунов В.А., Завьялов Д.В., Жариков Д.Н. Развитие направления параллельных и высокопроизводительных вычислений в ВолгГТУ // Параллельные вычислительные технологии (ПаВТ'2021). Короткие статьи и описания плакатов. XV международная конференция. Челябинск, 2021. С. 151-161

9. Алексеев И.А., Егунов В.А., Панюлайтис С.В., Чекушкин А.А. Методы и средства балансировки нагрузки в неоднородных вычислительных системах // Инженерный вестник Дона, 2020, № 11 URL: ivdon.ru/ru/magazine/archive/n11y2020/6667.

10. Шеин Д.В., Завьялов Д.В., Жариков Д.Н. Моделирование фосфорена методом классической молекулярной динамики с использованием глубокого обучения // Физика. Технологии. Инновации. ФТИ-2022: тез. докл. IX Междунар. молодеж. науч. конф., посвящ. 100-летию со дня рожд. проф. С. П. Распопина (г. Екатеринбург, 16-20 мая 2022 г.) - С. 330-331.

References

1. User’s Guide for Quantum ESPRESSO URL: quantum-espresso.org/Doc/user_guide_PDF/user_guide.pdf (data obrashheniya: 23.04.2023).

2. Cheng John, Grossman Max, McKercher Ty Professional CUDA® C Programming. John Wiley & Sons Inc., 2014. 528 p.

3. Sidorovich M.A. Nauchno-texnicheskoe i e`konomicheskoe sotrudnichestvo stran ATR v XXI veke. 2021. T. 1. pp. 263-267.

4. NVIDIA HPC SDK Version 23.1 Documentation URL: docs.nvidia.com/hpc-sdk/archive/23.1/index.html (data obrashheniya: 23.04.2023)

5. Wen-mei Hwu, Kirk David, El Hajj Izzat Programming Massively Parallel Processors. A Hands-on Approach. 4th Edition. Morgan Kaufmann, 2022. 580 p. ISBN 9780323912310

6. Abdraxmanov D.L., Zharikov D.N., Zav`yalov D.V. Inzhenernyj vestnik Dona, 2022, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2022/7688/.

7. Andreev A.E., Egunov V.A., Zav`yalov D.V., Zharikov D.N. Parallel`ny`e vy`chislitel`ny`e texnologii (PaVT'2021). Korotkie stat`i i opisaniya plakatov. XV mezhdunarodnaya konferenciya. Chelyabinsk, 2021. pp. 131-142.

8. Andreev A.E., Egunov V.A., Zav`yalov D.V., Zharikov D.N. Parallel`ny`e vy`chislitel`ny`e texnologii (PaVT'2021). Korotkie stat`i i opisaniya plakatov. XV mezhdunarodnaya konferenciya. Chelyabinsk, 2021. pp. 151-161.

9. Alekseev I.A., Egunov V.A., Panyulajtis S.V., Chekushkin A.A. Inzhenernyj vestnik Dona, 2020, № 11. URL: ivdon.ru/ru/magazine/archive/n11y2020/6667.

10. Shein D.V., Zav`yalov D.V., Zharikov D.N. Fizika. Texnologii. Innovacii. FTI-2022: tez. dokl. IX Mezhdunar. molodezh. nauch. konf., posvyashh. 100-letiyu so dnya rozhd. prof. S. P. Raspopina (g. Ekaterinburg, 16-20 maya 2022 g.) pp. 330-331.