

ИСПОЛЬЗОВАНИЕ MPI FOR PYTHON ДЛЯ ОРГАНИЗАЦИИ ОЧЕРЕДИ ИСПОЛНЕНИЯ ПРИ СКАНИРОВАНИИ ПОВЕРХНОСТЕЙ ПОТЕНЦИАЛЬНОЙ ЭНЕРГИИ

А. В. Романов^{1,*}, *С. Д. Кургалин*^{1,**}, *С. В. Борзунов*^{1,***},
А. Г. Потапов^{1,****}

¹ Воронежский государственный университет, Воронеж, Россия

С использованием MPI for Python разработан эффективный параллельный алгоритм сканирования поверхности потенциальной энергии при диффузии атомарного водорода, обеспечивающий запуск, распараллеливание и контроль вычислений. Расчеты организованы в виде процессов, взаимодействующих через общий набор файлов, что уменьшает накладные расходы при интенсивном обмене данными между процессами, минимизирует потери производительности и упрощает постановку задачи в очередь исполнения. Применение алгоритма будет способствовать созданию новых конструкционных материалов для атомной промышленности и водородной энергетики.

Using MPI for Python, an efficient parallel algorithm for scanning the potential energy surface during diffusion of atomic hydrogen has been developed, providing launch, parallelization and control of calculations. Calculations are organized as processes interacting through a common set of files, which reduces overhead costs during intensive data exchange between processes, minimizes performance losses and simplifies setting the task in the execution queue. The use of the algorithm will contribute to the creation of new structural materials for the nuclear industry and hydrogen energy.

PACS: 07.05.Tr; 02.60.-x

Параллелизм на уровне процессов является неотъемлемой частью расчетов на крупных вычислительных кластерах [1]. Несмотря на то, что математические библиотеки для научных вычислений традиционно реализуются на языках C/C++ и Fortran, логика верхнего уровня, позволяющая гибко настраивать параметры исследуемой модели и программировать этапы компьютерного моделирования, все чаще создается на языке Python. За два последних десятилетия в данной парадигме выпущен

* E-mail: alphard.rm@gmail.com

** E-mail: kurgalin@bk.ru

*** E-mail: borzunov@cs.vsu.ru

**** E-mail: artem2011.ap@inbox.ru

ряд программных пакетов, завоевавших большую популярность, таких как PyTorch, TensorFlow, ASE и GPAW. Низкий порог вхождения в язык и простота сопровождения программного кода на языке Python позволяют достаточно просто программировать логику сложных расчетов, без погружения в детали архитектуры используемого аппаратного окружения [2], что особенно актуально в случае гибридных вычислительных систем. Более того, современные реализации Python включают в себя модули очередей, управление процессами и распределенные вычисления. Эти и другие особенности языка объясняют продолжающийся быстрый дрейф в сторону его использования в разработке инструментария для научных вычислений [3–5]. Класс задач, решаемых с помощью такого подхода, обширен; Python-интерфейсы уже разработаны для задач молекулярной динамики, ядерной физики, физики элементарных частиц, квантовой химии и др.

Примером применения MPI for Python для решения одной из актуальных задач, возникающих в процессе создания новых конструктивных материалов для атомной промышленности и водородной энергетики, является организация очереди исполнения в задаче сканирования поверхности потенциальной энергии процесса диффузии атомарного водорода по поверхности твердой фазы материала. Сканирование поверхности потенциальной энергии является удобным инструментом исследования и требует существенных вычислительных ресурсов и практически всегда выполняется в параллельном режиме на высокопроизводительных компьютерных системах [6]. Тема водородных технологий имеет здесь особое значение. Несмотря на десятилетия изучения взаимодействия водорода с различными материалами и накопленный богатый практический опыт, эти исследования не потеряли своей актуальности и быстро набирают популярность, особенно в свете курса на отказ от углеводородного ископаемого топлива. Хотя натурный эксперимент и доказал свою эффективность при разработке новых материалов и сплавов, устойчивых, например, к водородной коррозии, теоретическое обоснование процессов с участием водорода остается неполным. Моделирование диффузии атома водорода по поверхности твердой фазы и через объем исследуемого образца, содержащего поверхностные дефекты, представляет интерес не только для создания новых перспективных материалов, но и с точки зрения особенностей программно-аппаратной реализации такого моделирования.

Сканирование поверхности потенциальной энергии можно представить в виде набора процессов, когда атомы водорода находятся в узлах прямоугольной решетки, наложенной на поверхность исследуемого материала. Данные для расчетов формируются из пула точек, являющихся узлами решетки сканирования. Такие расчеты хорошо распараллеливаются с помощью модуля mpi4py языка Python и при этом не требуют высокоскоростного обмена данными между узлами вычислительного кластера. Таким образом, можно использовать более дешевые комплексы на базе персональных компьютеров или применять распределенные вычисления.

Подобные расчеты, тем не менее, имеют свои сложности. В частности, каждый расчет может выполняться в течение разных, сильно различающихся по продолжительности промежутков времени, и в ходе сканирования неизбежно появляются процессы-лидеры и процессы-отстающие, что снижает эффективность вычислений в случае распараллеливания по рядам и блокам сетки. Ситуация усугубляется масштабами вычислений, так как в зависимости от размеров модели количество структур в задаче оптимизации геометрии системы может достигать нескольких тысяч, а при сканировании поля потенциальной энергии — десятков тысяч структур.

Решением проблемы является использование модуля очередей языка Python. Штатная очередь исполнения, как правило, имеет ограниченные возможности контроля подобных вычислений. В этом случае на главный процесс возлагается функция распределения начальных геометрий системы из общего пула данных. Главный процесс ждет передачи ранга от остальных процессов и выдает каждому из них по одной геометрии за один раз. По завершении расчета процессы снова посылают свои ранги главному процессу до исчерпания всей очереди данных. Этот вариант хорошо работает на небольших системах, когда достаточно одного процессорного ядра для выполнения расчетов за разумное время, и позволяет избежать потерь производительности.

Существует и другой подход к решению данной задачи. Вычисления можно ускорить, используя электронную плотность многоатомной системы, рассчитанную для предыдущей геометрии. В этом случае вычисления начинаются в произвольных узлах координатной сетки с последующим переходом в ближайший незанятый узел. Очевидно, что максимальная производительность алгоритма в этом случае может быть достигнута только на машинах с общей памятью. Тем не менее падение скорости вычислений за счет передачи информации об электронной плотности многоатомной системы через довольно медленный интерконнект до некоторой степени нивелируется, так как это происходит только в момент перехода вычислительного процесса на новый маршрут ветвления, когда рядом не остается непересчитанных узлов.

Разработанный алгоритм прошел детальную апробацию на высокопроизводительном вычислительном кластере Суперкомпьютерного центра Воронежского государственного университета [7]. Его применение будет способствовать созданию новых конструкционных материалов для атомной промышленности и водородной энергетики.

Работа является продолжением цикла исследований, представленных на конференции «GRID'2023» и опубликованных в журнале «ЭЧАЯ» [8].

СПИСОК ЛИТЕРАТУРЫ

1. Kurgalin S., Borzunov S. A Practical Approach to High-Performance Computing. Springer Intern. Publ., 2019.

2. Романов А. В., Кургалин С. Д., Петрищев К. О. Использование языка программирования Python в разработке высокопроизводительных математических кодов // Актуальные проблемы прикладной математики, информатики и механики: Тр. междунар. науч. конф. Воронеж, 2024 г. С. 15–18.
3. Борзунов С. В., Кургалин С. Д. Языки программирования. Python: решение сложных задач. СПб.: Лань, 2023.
4. Kurgalin S., Borzunov S. Algebra and Geometry with Python. Springer, 2021.
5. Борзунов С. В., Кургалин С. Д., Чуракова Т. А. Python: алгоритмы для физиков и математиков. Воронеж: Изд. дом ВГУ, 2024.
6. Kataoka Y., Haruyama J., Sugino O., Shiga M. Predictive Evaluation of Hydrogen Diffusion Coefficient on Pd(111) Surface by Path Integral Simulations Using Neural Network Potential // Phys. Rev. Res. 2024. V. 6. P. 043224.
7. Борзунов С. В., Кургалин С. Д. Использование ресурсов Суперкомпьютерного центра Воронежского государственного университета в учебном процессе и научных исследованиях // Суперкомпьютерные дни в России: Тр. междунар. конф. М.: Изд-во Моск. ун-та, 2018. С. 972–977.
8. Борзунов С. В., Романов А. В., Кургалин С. Д., Петрищев К. О. Высокопроизводительные вычисления на математических сопроцессорах и графических ускорителях с использованием Python // ЭЧАЯ. 2024. Т. 55, № 3. С. 615–618.