

Параллельный эволюционный алгоритм в задаче оптимизации функции большой размерности

Котряхова С. Н., Степанова М. М.
СПбГУ, Физический Факультет
Дубна, 2016

Постановка задачи

Разработка программы оптимизации потенциала ReaxFF.

ReaxFF (Reactive Force Field) — это эмпирическое силовое поле, позволяющее моделировать химические и физические процессы методами молекулярной динамики.

Общий вид потенциала:

$$E_{ReaxFF} = E_{bond} + E_{lp} + E_{over} + E_{under} + E_{val} + E_{pen} + \\ + E_{coa} + E_{C2} + E_{tors} + E_{conj} + E_{H-bond} + E_{vdWaaLs} + E_{Coulomb}$$

$$E_{ReaxFF} = f(p_1, \dots, p_N)$$

МД-потенциал ReaxFF

Целевая функция:

$$Error = \sum_{k=1}^L \sigma_k |U_k^{QC} - U_k^{ReaxFF}| + \sum_{k=1}^L \sigma_{L+k} \sqrt{\sum_{\alpha=1}^A \sum_{i=1}^3 (F_{k\alpha i}^{QC} - F_{k\alpha i}^{ReaxFF})^2}$$

Здесь U_k – потенциальные энергии моделей оптимизирующего набора,
 $F_{k\alpha i}$ – компоненты сил, действующих на атомы каждой модели,
 L – число моделей оптимизирующего набора,
 A – число атомов в моделях,
 σ – весовые множители.

Индексы QC и $ReaxFF$ – означают, соответственно, характеристики, полученные из квантовых расчетов и расчетов ReaxFF

Особенности задачи

- Большая размерность целевой функции оптимизации потенциала ReaxFF
- Характерная особенность ландшафта целевой функции
- Значительное время расчета значения целевой функции (для расчета используется МД-симулятор LAMMPS)

Особенности ландшафта целевой функции

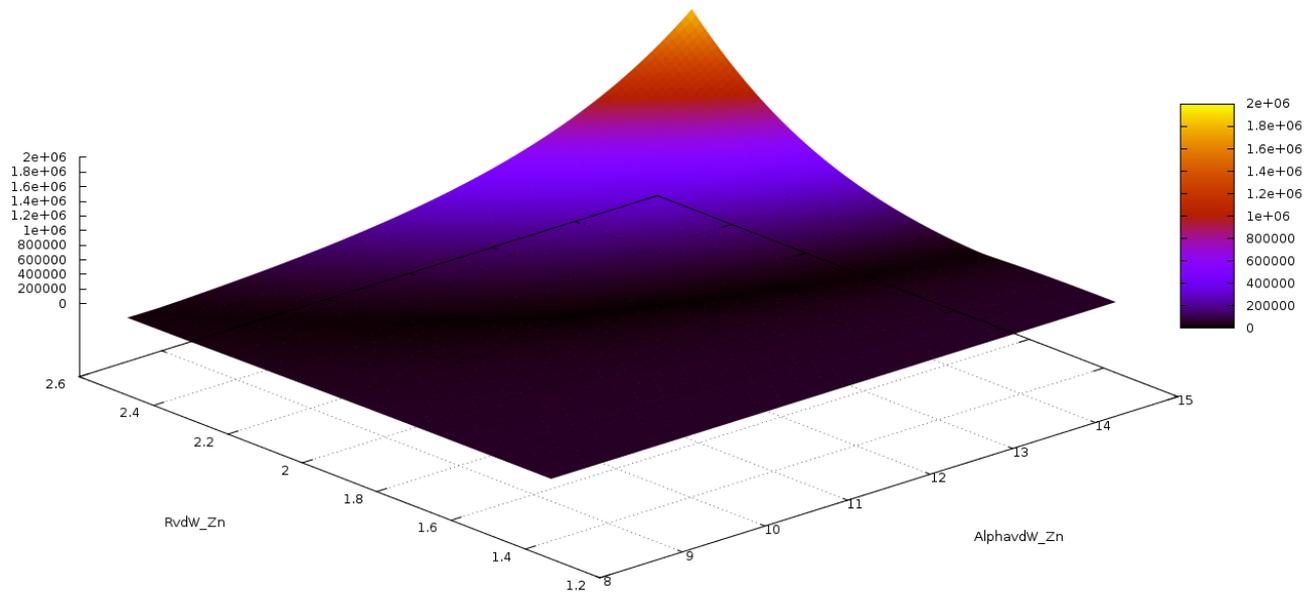


График зависимости синтетической целевой функции от параметров RvdW_Zn и AlphavdW_Zn

Особенности ландшафта целевой функции

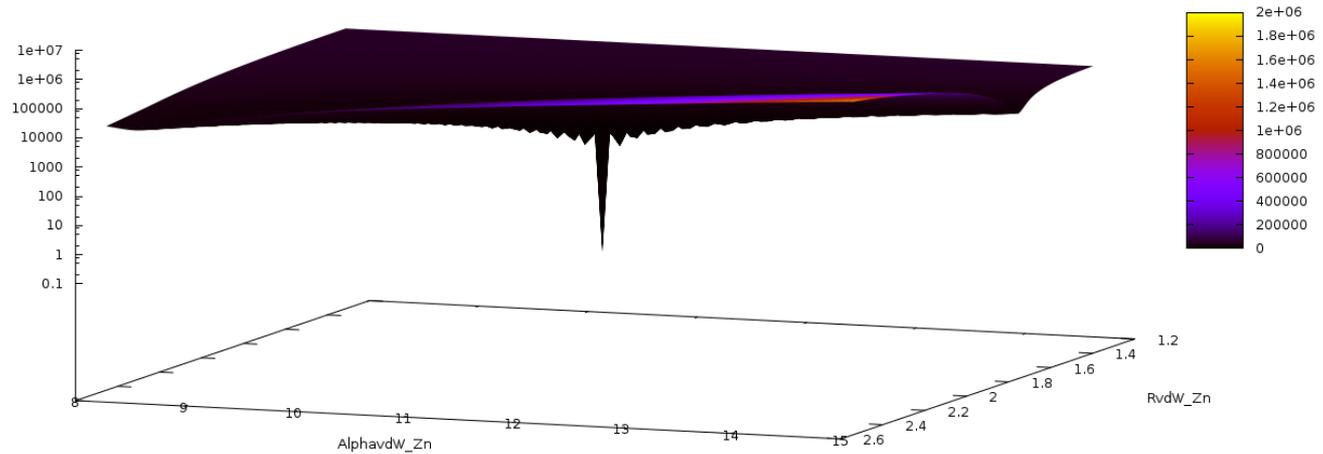


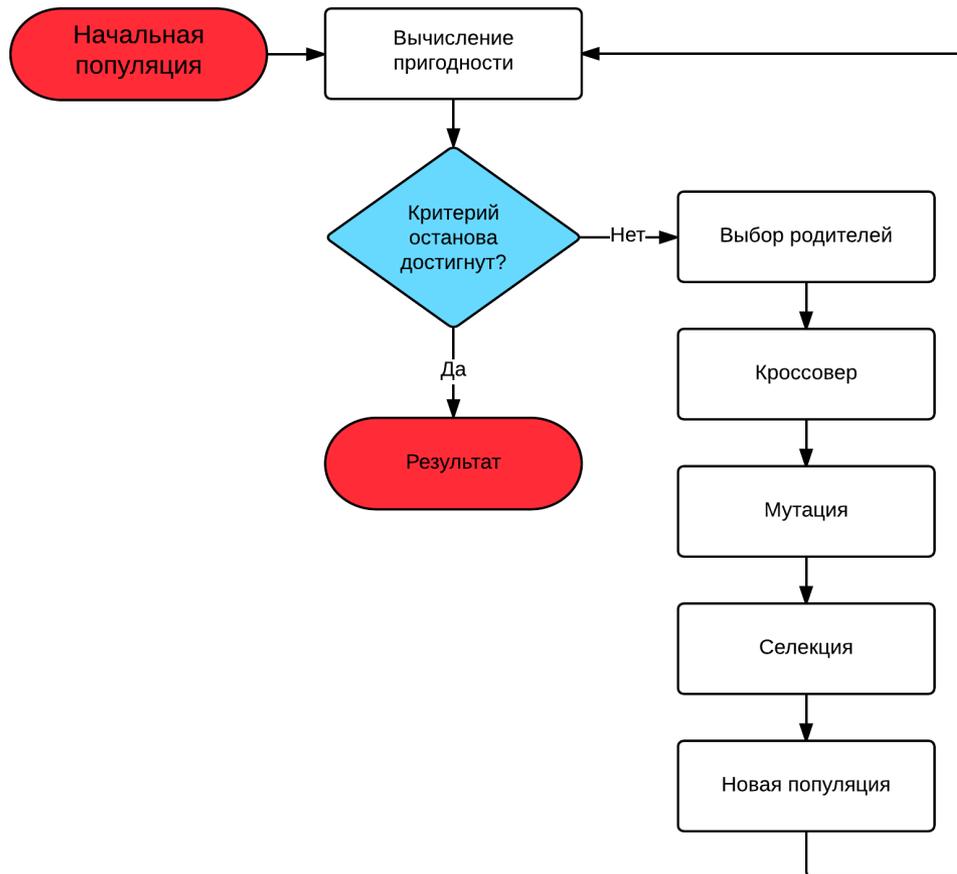
График зависимости синтетической целевой функции от параметров RvdW_Zn и AlphavdW_Zn в логарифмическом масштабе

Метод решения

Разработан параллельный гибридный алгоритм, включающий в себя:

- Генетический алгоритм (Genetic Algorithm) – эвристический метод оптимизации и моделирования, имитирующий принципы естественного отбора
- Оптимизацию роя частиц (Particle Swarm Optimization) – стохастический алгоритм оптимизации, основанный на идее самоорганизации роя частиц

Генетический алгоритм



Основные операции:

- Выбор родителей – фенотипный аутбридинг
- Скрещивание – одноточечный кроссовер
- Мутации – при наличии у родителей совпадающих генов
- Селекция

Оптимизация роя частиц

Популяция рассматривается как рой частиц.

$X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ – координата i -частицы

$V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ – скорость i -частицы

$P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ – лучшее положение i -частицы

$P_g = (p_{g1}, p_{g2}, \dots, p_{gD})$ – лучшее положение роя

Оптимизация роя частиц

Скорости и координаты обновляются по следующему правилу:

$$v_{id}^{t+1} = w^t v_{id}^t + c_1 r_1^t (p_{id}^t - x_{id}^t) + c_2 r_2^t (p_{gd}^t - x_{id}^t)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}$$

где $d \in \{1, 2, \dots, D\}, i \in \{1, 2, \dots, N\}$

N - размер популяции, индекс t - номер итерации,

w - весовой коэффициент инерции,

r_1 и r_2 случайные числа в диапазоне $[0, 1]$,

c_1, c_2 - параметры алгоритма характеризующие рой частиц

Гибридный алгоритм

В основе гибридного алгоритма лежит поочередное использование генетического алгоритма и метода оптимизации роя частиц для формирования популяции.

Реализация позволяет провести оптимизацию параметров алгоритма для ускорения сходимости на конкретной выбранной целевой функции.

Оптимизация параметров алгоритма

Нахождение оптимального соотношения числа итераций GA и PSO.

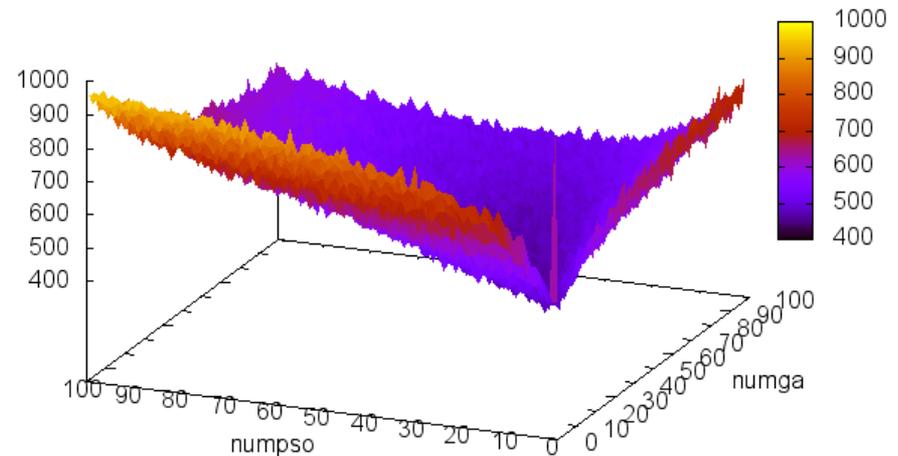
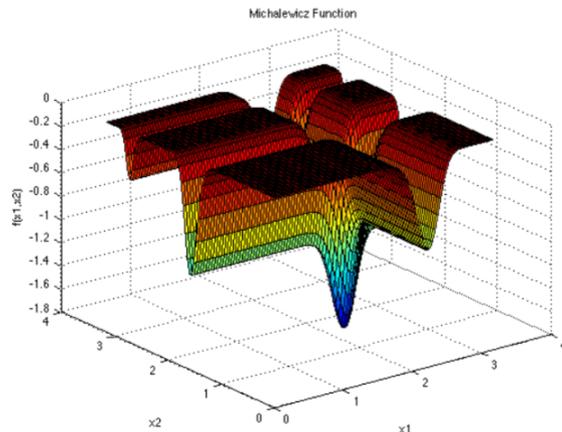
Функция Михалевича.

$$f(x) = - \sum_{i=1}^n \sin(x_i) \cdot \left(\sin\left(\frac{i \cdot x_i^2}{\pi}\right) \right)^{2m},$$

$$0 \leq x_i \leq \pi, i \in [1; n]$$

при $n = 2$: $f(x^*) = -1.8013$,

где $x^* = (2.20, 1.57)$



Зависимость скорости сходимости от числа итераций GA и PSO

GA	PSO	# итераций
25	20	439

Гибридный алгоритм

- Подходит для решения задач оптимизации функций многих параметров
- Справляется даже со сложным ландшафтом целевой функции
- Работает на мультимодальных функциях
- Хорошо локализует минимум

Проверка алгоритма на классических тестовых функциях

Функция Растригина:

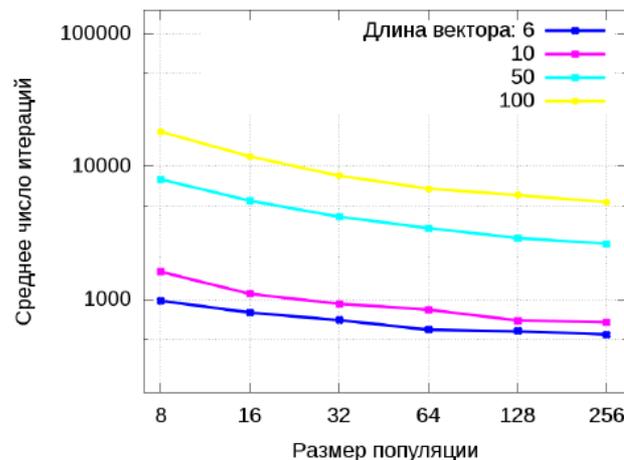
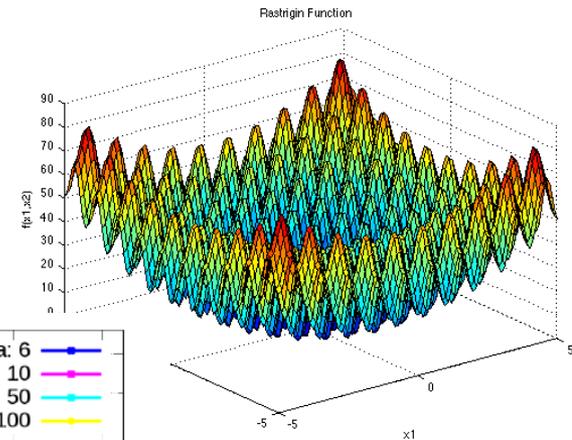
$$f(x) = 10 \cdot n + \sum_{i=1}^n \left(x_i^2 - 10 \cdot \cos(2 \cdot \pi \cdot x_i) \right),$$

Область определения:

$$-5.12 \leq x_i \leq 5.12, i \in [1; n]$$

Глобальный минимум:

$$f(x^*) = 0, \text{ где } x^* = (0, \dots, 0)$$



← Зависимость числа итераций от размера популяции для векторов разной длины

Проверка алгоритма на классических тестовых функциях

Функция Розенброка:

$$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$$

Область определения:

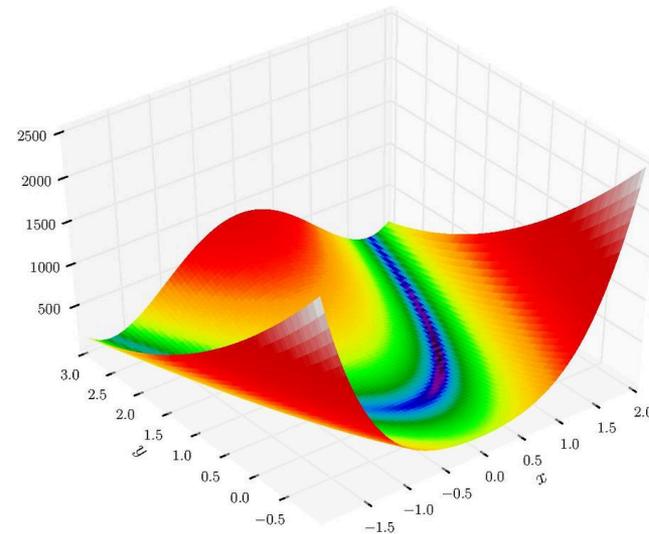
$$-5 \leq x_i \leq 10, i \in [1; n]$$

Глобальный минимум:

$$f(x^*) = 0, \text{ где } x^* = (1, \dots, 1)$$

Зависимость числа итераций от размера популяции

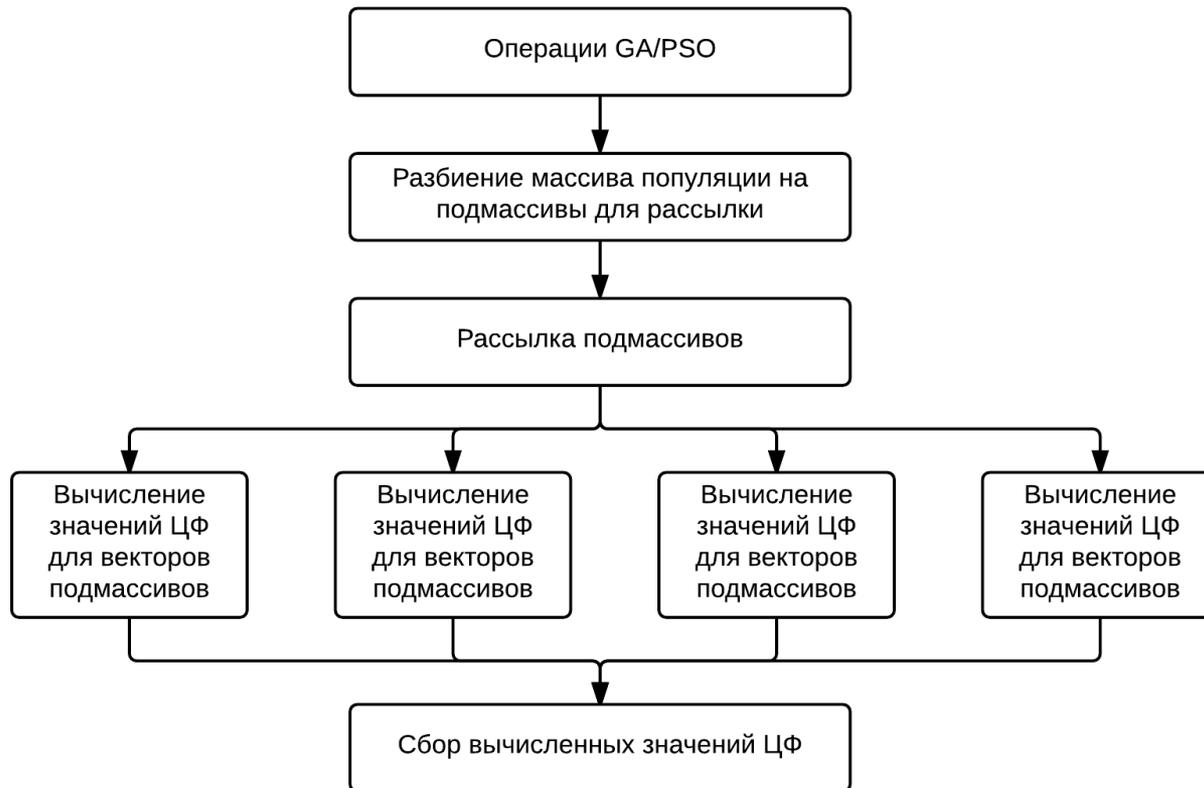
Размер популяции	Число итераций	Число обработанных векторов
8	203811	3260976
16	79509	2544288
32	30831	1973184
64	11140	1425920
128	5351	1369856
256	2606	1334272



Реализация параллельной работы алгоритма

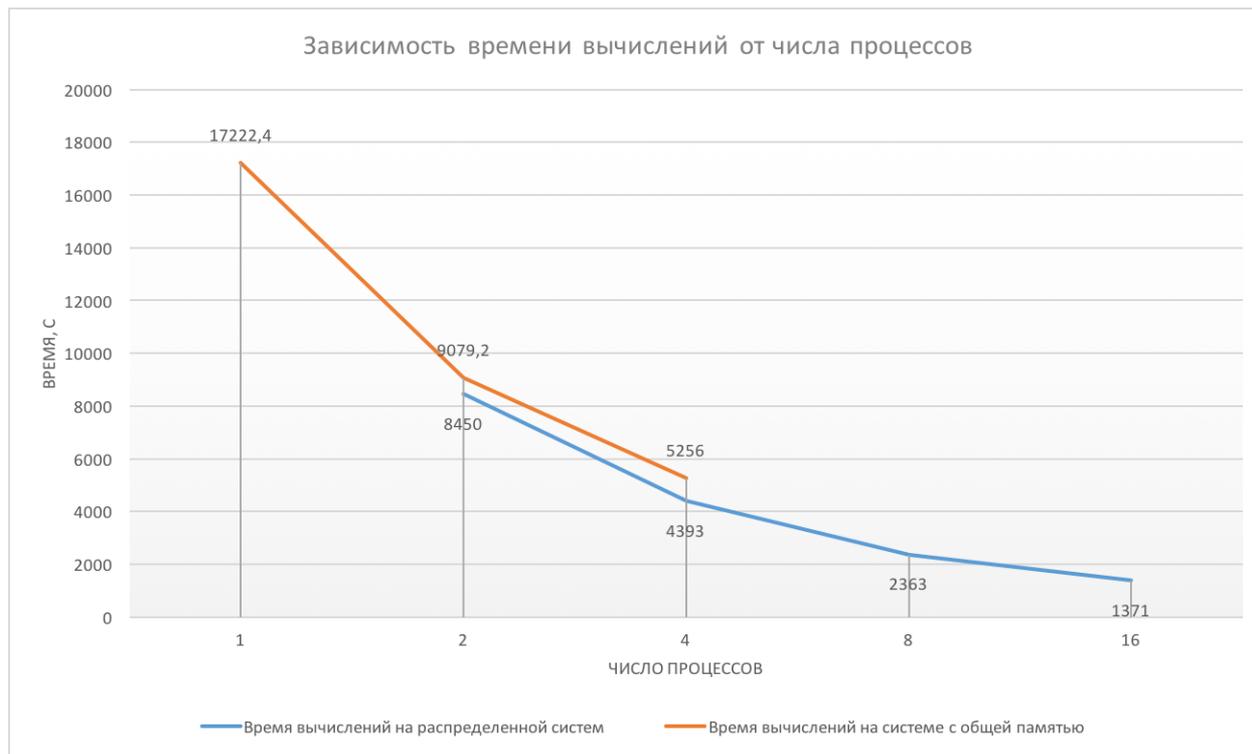
- Параллельная работа алгоритма реализована по схеме Master-Slave с использованием MPI (OpenMPI)
- Распараллелена самая трудозатратная часть – вычисления значений целевой функции (ЦФ)
- Отсутствует межпроцессный обмен (обмен производится только между Master'ом и Slave'ми)

Тело основного цикла



Зависимость времени вычислений от числа процессов

Целевая функция – минимизация потенциала ReaxFF.
Число итераций – 100.



Результаты оптимизации потенциала ReaxFF

Получен набор оптимизированных параметров для систем Zn-O-H, позволяющий проводить молекулярно-динамическое моделирование*.

- Оптимизация 24 параметров,
- Размер популяции – 128 векторов,
- Результат достигнут за 3361 итерацию
- Затрачено 179 часов при работе 4 процессов на Intel Core i5 CPU 3.00GHz, RAM 8Gb

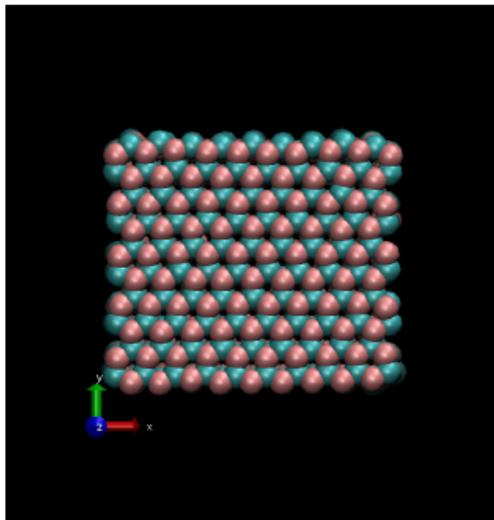
**Сравнение с другими методами оптимизации см. стендовый доклад
“Parametrization of the Reactive MD Force Field for Zn-O-H Systems”*

Моделирование кристаллов

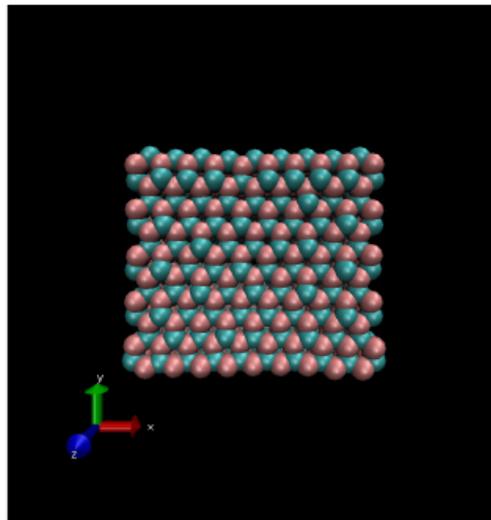
ZnO со структурой типа вюрцит:

Параметры моделирования:

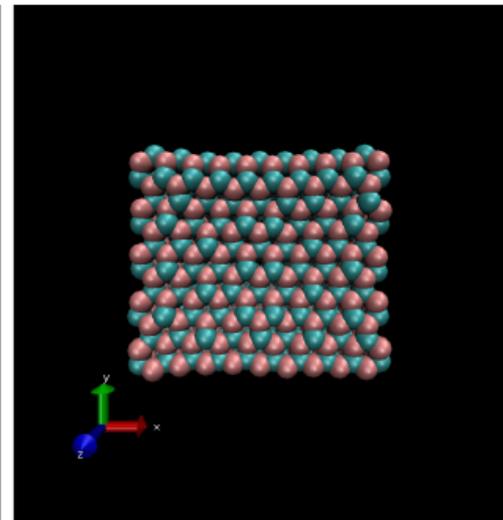
2090 атомов, куб $30 \times 30 \times 30 \text{ \AA}$ в вакууме при $T = 300 \text{ K}$



LAMMPS



GSA



GA

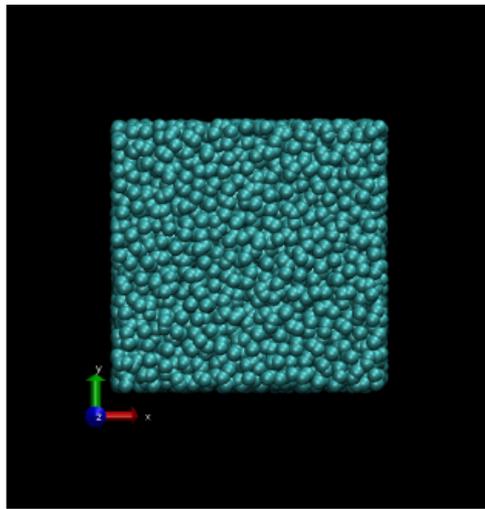
Моделирование кристаллов

ZnO со структурой типа каменной соли:

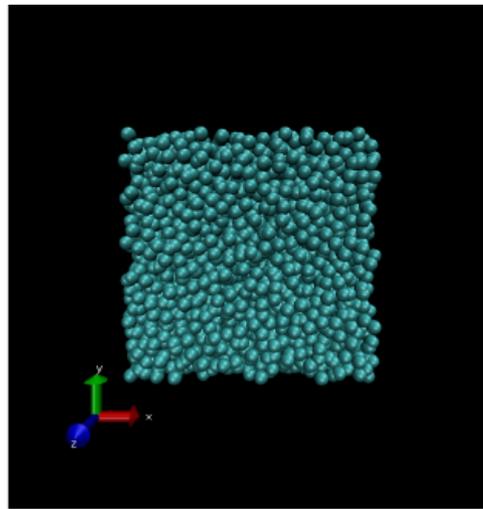
Существует при высоком давлении.

Параметры моделирования:

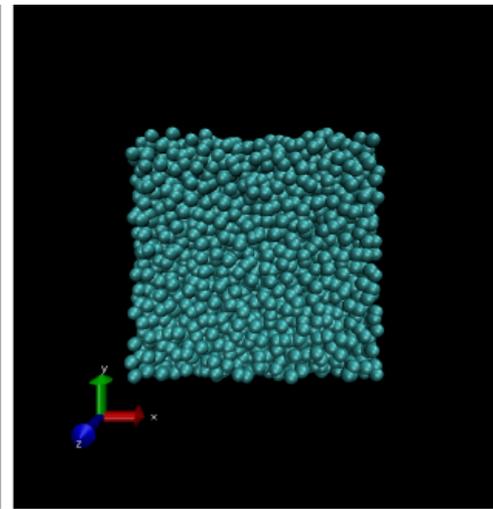
2200 атомов оксида, куб $30 \times 30 \times 30 \text{ \AA}$ в кислородной атмосфере, давление 4.6 ГПа, $T = 300 \text{ K}$



LAMMPS



GSA



GA

Заключение

- Разработана оригинальная реализация гибридного эволюционного алгоритма
- Алгоритм эффективен для задач оптимизации функций большой размерности обладающих сложным ландшафтом
- Параллельная реализация эффективна при работе с ресурсоемкими целевыми функциями
- Получены хорошие результаты для задачи оптимизации потенциала ReaxFF

Спасибо за
внимание!
