Федеральное государственное автономное образовательное учреждение
высшего профессионального образования «Московский физико-технический
институт (государственный университет)»

Выпускная квалификационная работа на степень бакалавра

Применение алгоритмов машинного обучения для предсказания кристаллических структур

Студент: Долгирев П.Е., 128 группа.

Научный руководитель: Оганов А.Р.
Dr. Habilitation, PhD.

Москва 2015

Содержание

1 Введение

Долгое время считалось, что невозможно будет предсказать атомную структуру кристаллического тела, зная только формулу вещества (см. [1]). При заданных $P-T$ условиях стабильная структура отвечает минимуму потенциала Гиббса: $G = E - TS + PV$. Энтропия - это одно из самых трудных слагаемых и обычно для упрощения задачи кладут $T = 0$: $G = E + PV$. Потенциал Гиббса является функцией $3N + 3$ координат ($3N + 6 - 3$, где $3N$ - число координат атомов; 6 - параметры ячейки; -3 соответствует тому, что положение одного из атомов может быть фиксировано) и исследование всего $(3N+3)$-мерного пространства вычислительно невозможно. Однако оказывается, что для поиска стабильной структуры и нет необходимости исследовать все пространство.

Сложность вычислений состоит также и в том, что для того, чтобы вычислить энергию E структуры (здесь подразумевается, что структура определяется положением своих атомов, а ее энергия отвечает основному состоянию электронной конфигурации), нужно решить, в общем случае, квантово-механическое уравнение (уравнение Кона-Шэма и их теория функционала плотности (DFT) - см. [2]). В частности, пакет VASP (см. [3]) умеет релаксировать данную структуру, то есть VASP считает энергию и силы, что позволяет из данной структуры (по сути градиентным спуском) получить структуру, отвечающей какому-то минимуму потенциала Гиббса. Этот потенциал, к сожалению, имеет огромное количество минимумов, и поэтому все равно трудно найти наиболее стабильную структуру. Тем не менее, оказывается, что эти минимумы расположены близко друг к другу (см. [4, 12]).

1.1 USPEX

USPEX - это эволюционный алгоритм предсказания кристаллических структур. Это означает, что этот алгоритм симулирует дарвиновскую эволюцию, только здесь в отборе принимают участия различные структры, отвечающие наперед заданной химической формуле. Так, например, USPEX может использовать VASP: в данном поколении структуры сначала оптимизируются с помощью VASP, а далее участвуют в отборе для следующего поколения... (см. детали в [4, 5]). Одно из важнейших свойств USPEX в том, что по сути он сужает $(3N + 3)$-мерное пространство до относительно небольшой области, где сосредоточены минимумы, а дальше исследует в основном эту область.

Этот алгоритм продемонстрировал выдающиеся результаты и сейчас используется более, чем 2000 учеными для теоретического предсказания структур[1]. В частности, в работе [6] сравнивались эксперементальные данные с результатами USPEX по предсказанию устойчивых структур Na_xCl_y под давлением. Полученные результаты великолепно согласовывались с экспериментом.

[1] см. сайт http://uspex.stonybrook.edu/index.html

1.2 Основная задача

Во время эволюции получается, что часто рождаются "похожие"структуры, а так как подобные вычисления дорогие, то мы предполагаем, что методы аппроксимации (в частности, методы машинного обучения) могут существенно ускорить работу USPEX. То есть в данной работе мы применяли методы машинного обучения для ускорения работы USPEX.

2 Машинное обучение

Машинное обучение — это довольно широкий класс алгоритмов (см. подробнее [7]) и в данной работе мы будем рассматривать регрессионные алгоритмы для задачи аппроксимации. Пусть про функцию f известны ее значения на некоторой выборке $\{x_i \in X^l, y_i = f(x_i)\}$; задача аппроксимации состоит в подходящем выборе весов ω для некоторой фиксированной функции-модели $F(\omega, x)$ так, чтобы эта функция приближала f. Это сводится к минимизации целевой функции: $Q(\omega, X^l, y) \to \min$. В качестве целевой функции мы выбирем следующее приближение: $Q(\omega, X^l, y) = \sum_{i=1}^{l} \Lambda(\omega, x_i; y_i)$, где $\Lambda(\omega, x_i; y_i) = (F(\omega, x_i) - y_i)^2$ есть величина ошибки приближения на элементе x_i. Минимизация целевой функции основана на градиентном спуске и конкретный алгоритм минимизации зависит уже от специфики задачи (это подробнее будет обсуждаться ниже по отношению к предсказанию кристаллических структур). Сразу отметим, что выбор модели и целевой функции критически влияет на скорость сходимости и на качество сходимости будущего алгоритма. Так, многие алгоритмы машинного обучения отличаются между собой выбором модели.

В данной работе мы будем использовать машинное обучение для аппроксимации энергии кристалла, как функцию координат атомов. Идея применения в кристаллохимию машинного обучения не нова и наш метод основан на многих идеях, описанных в [8–11] (в течении всего текста мы будем сравнивать наш подход и результаты с методами и результатами других научных групп, тем самым демонстрируя основные преимущества и недостатки нашего метода).

2.1 Нейронные сети

В качестве алгоритма машинного обучения мы используем нейронные сети (NN) — см. [7, 10, 13–15] — потому что этот алгоритм является наиболее общим[2] и обучение также интуитивно понятно. В статье [10] описаны основные способы

[2] в [13] сформулирована теорема Колмогорова: любая непрерывная функция n аргументов на единичном кубе $[0,1]^n$ представима в виде суперпозиции одного аргумента и операции сложения: $f(x_1, x_2, \cdots, x_n) = \sum_{k=1}^{2n+1} h_k\Big(\sum_{i=1}^{n} \phi_{ik}(x_i)\Big)$. Видно, что данная формула напоминает структуру нейронной сети (более того, в ней написано достаточное число узлов). Тем не менее, в этой теореме функции ϕ, h зависят от f, а в нейронных сетях активационные функции выбираются заранее. Теорема Колмогорова в данном случае дает просто интуитивное понимание о важности нейронных сетей. В этой же статье подробней обсуждается, почему нейронные сети являются универсальными аппроксиматорами функций.

применения нейронных сетей к современным расчетам в химии: к кристаллам, к поверхностям, к кластерам и линейным цепочкам. В статьях [13–15] описаны алгоритм нейронных сетей и важнейшие эвристики, направленные на создание оптимального алгоритма.

На рис.1 изображена простейшая архитектура $(n - H - M)$ нейронной сетки с одним скрытым слоем: n нейронов (узел нейронной сетки называется нейроном) в входном слое, H нейронов — в скрытом слое и M нейронов — на выходном слое. Вектор $x = (x_1, ..., x_n)$ называют также вектором признаков (feature-vector) — такое название удачно в нашем случае, потому что в качестве x_i мы будем брать переменные, которые будут как-то описывать кристаллическую структуру. На нейроны скрытого слоя u^h передается информация от входного слоя под действием линейного преобразования и последующего действия активационной функции:

$$u^h = \sigma_h(\sum_{j=0}^{n} w_{jh} x^j).$$

Аналогичным образом информация передается на выходной слой a^m:

$$a^m = \sigma_m(\sum_{h=0}^{H} w_{hm} u^h).$$

Задача сводится к тому, чтобы оптимизировать весовые параметры w_{jh} и w_{hm} — это делается стандартным алгоритмом обратного распространения ошибок (back-prop — см. ниже 2.2).

Написанное может быть легко обобщено на большее количество скрытых слоев. Подходящая архитектура сетки зависит от кокретной задачи и в случае аппроксимации непрерывной функции достаточно 1-2 скрытых слоя. Количество узлов в слоях часто подбирается экспериментально, то есть тестируются много различных архитектур сеток и выбирается оптимальная.

Обычно в качестве активационных функций берется сигмоида или гиперболический тангенс, мы же выбрали:

$$\mathrm{th}(x) + \gamma x,$$

что обеспечивает лучшую сходимость алгоритма и решает проблему "паралича"нейронной сети - см. [13, 14].

2.2 Back-Prop

Этот алгоритм приведен много где и суть его в следующем: оказывается, что количество операций, которые нужно совершить для вычисления градиента сравнимо с вычислением самой функции. Пусть $x_i \in X^l$, а $y_i = (y_1, ..., y_M)$ — вектор значений, соответствующий x_i. Мы хотим минимизировать следующую функцию:

$$Q(w) = \frac{1}{2} \sum_{m=1}^{M} (a^m(x_i) - y_i^m)^2.$$

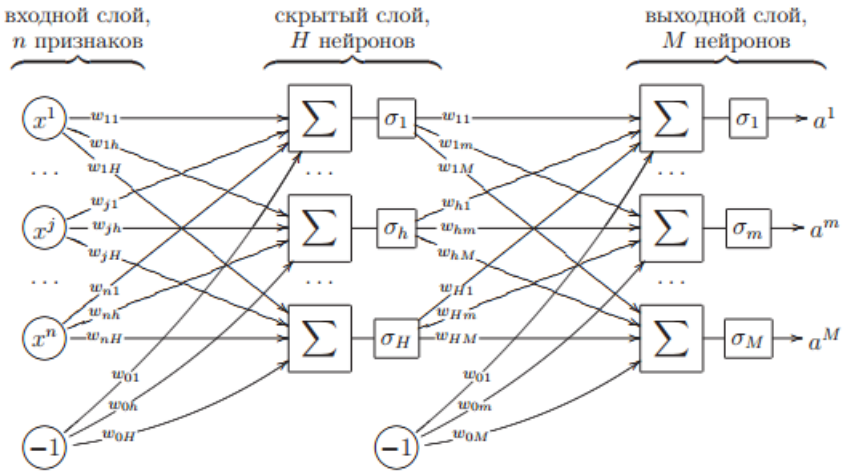


Рис. 1: Структура многослойной нейронной сети.

Посчитаем сначала следующие частные производные:

$$\frac{\partial Q}{\partial a^m} = a^m(x_i) - y_i^m = \epsilon_i^m;$$

$$\frac{\partial Q}{\partial u^h} = \sum_{m=1}^{M}(a^m(x_i) - y_i^m)\sigma'_m w_{hm} = \sum_{m=1}^{M}\epsilon_i^m \sigma'_m w_{hm} = \epsilon_i^h.$$

Величину ϵ_i^m естественно назвать величиной ошибки на узле m выходного слоя; под величиной ϵ_i^h мы будем понимать величину ошибки в скрытом слое. Вычисление ϵ_i^h из ϵ_i^m похоже на то, как если бы мы запустили сеть в обратном направлении — отсюда название алгоритма (обратное распространение ошибок).

Теперь ключевые формулы:

$$\frac{\partial Q(w)}{\partial w_{hm}} = \frac{\partial Q(w)}{\partial a^m}\frac{\partial a^m}{\partial w_{hm}} = \epsilon_i^m \sigma'_m u^h;$$

$$\frac{\partial Q(w)}{\partial w_{jh}} = \frac{\partial Q(w)}{\partial u^h}\frac{\partial u^h}{\partial w_{jh}} = \epsilon_i^h \sigma'_h x_i^j.$$

Так же разделяют методы градиетного спуска: если менять веса сетки, тренируясь только на одной структуре, то такой алгоритм называется стохастическим градиентным спуском. Он применяется обычно в таких задачах, где данных очень много и нужно быстро найти подходящую нейронную сетку. Когда данных относительно мало, то используется одновременно вся выборка в одной итерации — берут средний градиент по всей выборке. Этот алгоритм называется batch gradient descent и именно его мы использовали в данной работе. Преимещество batch gradient descent еще в том, что сходимость этого алгоритма легко контролируется, а также для него написаны ряд методик, ускоряющих эту сходимость (в частности, метод сопряженных градиентов работает только с batch).

3 Структурные суммы

Здесь мы сформулируем один метод, который будет применен совместно с нейронными сетями (см. ниже).

Предположим, что атомы в кристалле взаимодействуют классически, т.е. во взаимодействии участвуют 2 атома, которое может быть описано посредством потенциала. Если в кристалле K-типов атомов, то количество потенциалов будет $\frac{1}{2}K(K+1)$, каждый из которых описывает свое взаимодействие. Предположим также, что каждый из потенциалов имеет вид:

$$\phi(r) = \sum_{k=1}^{k_{max}} \frac{A_k}{r^k}.$$

Теперь посчитаем энергию кристалла (энергия, приходящаяся на элементарную ячейку), которая в данном случае равна:

$$E_{2body} = \sum_{l=0}^{\infty} \sum_{i,j=1}^{N} \sum_{k=1}^{k_{max}} \frac{A_{i,j}^k}{r_{i,j}^k(l)},$$

где суммирование происходит по всему кристаллу; мы как-то занумеровали все элементарные ячейки, причем $l = 0$ для ячейки, энергию которой мы вычисляем; i-й атом в нулевой ячейке, а j-й в l-й; $A_{i,j}^k$ — k-й коэффициент в потенциале взаимодействия между атомом i и атомом j, который зависит только от типов этих атомов. Подобные выражения

$$\sum_{l=0}^{\infty} \sum_{i,j=1}^{N} \sum_{k=1}^{k_{max}} \frac{1}{r_{i,j}^k(l)}$$

называются структурными суммами.

Слагаемые с $k = 1, 2, 3$ отвечают дальнодействующим взаимодействиям в том смысле, что при этих k интегралы $\int_1^{\infty} \frac{r^2 dr}{r^k}$ расходятся, и трудность для этих слагаемых состоит в том, что для того, чтобы посчитать их вклад в полную энергию, нельзя будет суммировать внутри сферы большого радиуса, вместо суммирования по всему кристаллу. Кулоновское слагаемое люди уже давно умеют считать и для этого существует метод Эвальда (см. [16, 17]), в котором происходит также суммирование и по обратной ячейке. Эту же идею мы использовали, чтобы посчитать слагаемые как с $k = 2, 3$, так и для того, чтобы посчитать слагаемые с $k = 4, 5$. Вывод формул громоздок и почти ничем не отличается от того, который описан в [17]. Ниже приводятся сами формулы. Слагаемые с $k \geq 6$ мы суммируем внутри большой сферы (типично с радиусом в 10Å). Для того, чтобы посчитать $k = 4, 5$ с той же точностью, что и слагаемые с $k \geq 6$, нам бы пришлось суммировать в сфере очень большого радиуса; по этой причине мы их считаем так сложно.

Мы получили, что для каждого из 5-ти слагаемых вместо суммирования по кристаллу, мы теперь суммируем, как по прямой решетке, так и по обратной.

И суммировать теперь достаточно в сферах относительно небольшого радиуса. Однако теперь нужно выбрать радиусы сфер R_{max}, G_{max} и параметр g так, чтобы в итоге совершить минимальное количество операций и добиться высокой точности. В коде GULP[3] используются оптимальные параметры для кулоновского слагаемого; небольшой анализ показывает, что эти параметры дают ту же точность и для $k = 2, 3, 4, 5$, и поэтому эти же параметры были использованы для расчетов старших слагаемых.

Отметим теперь, что суммарный заряд ячейки должен равняться нулю (это очевидно из физических соображений, но так же это видно из того, что в соответствующей формуле мы получим бесконечность). Оказывается, что аналогичное можно сказать и для слагаемых $k = 2, 3$ (см. формулы ниже): $\sum\limits_{i,j=1}^{N} A_{i,j}^{k} = 0$. Мы заодно получаем замечательный вывод о том, что в кристаллах, где всего один тип атомов, нет дальнего взаимодействия. А поэтому такие кристаллы принципиально проще тех, где типов атомов больше. Пока предлагаемый ниже нами метод работает для кристаллов с одним типом атомов.

Энергию кристалла можно рассматривать, как функцию от коэффициентов $A_{i,j}^{k}$; очевидно, что эта функция линейна. А это означает, что, зная энергии нескольких кристаллических структур, мы можем восстановить коэффициенты (с помощью линейной регрессии). По этой причине мы надеемся, что этот метод структурных сумм поможет в случае, когда есть дальнее взаимодействие.

Далее мы провели 2 теста для проверки нашего алгоритма. В первом тесте мы симулировали Леннард-Джонсовское взаимодействие. Общий вид потенциала Леннарда-Джонса следующий:

$$\phi(r) = \epsilon\Big[\Big(\frac{r_m}{r}\Big)^{12} - 2\Big(\frac{r_m}{r}\Big)^{6}\Big],$$

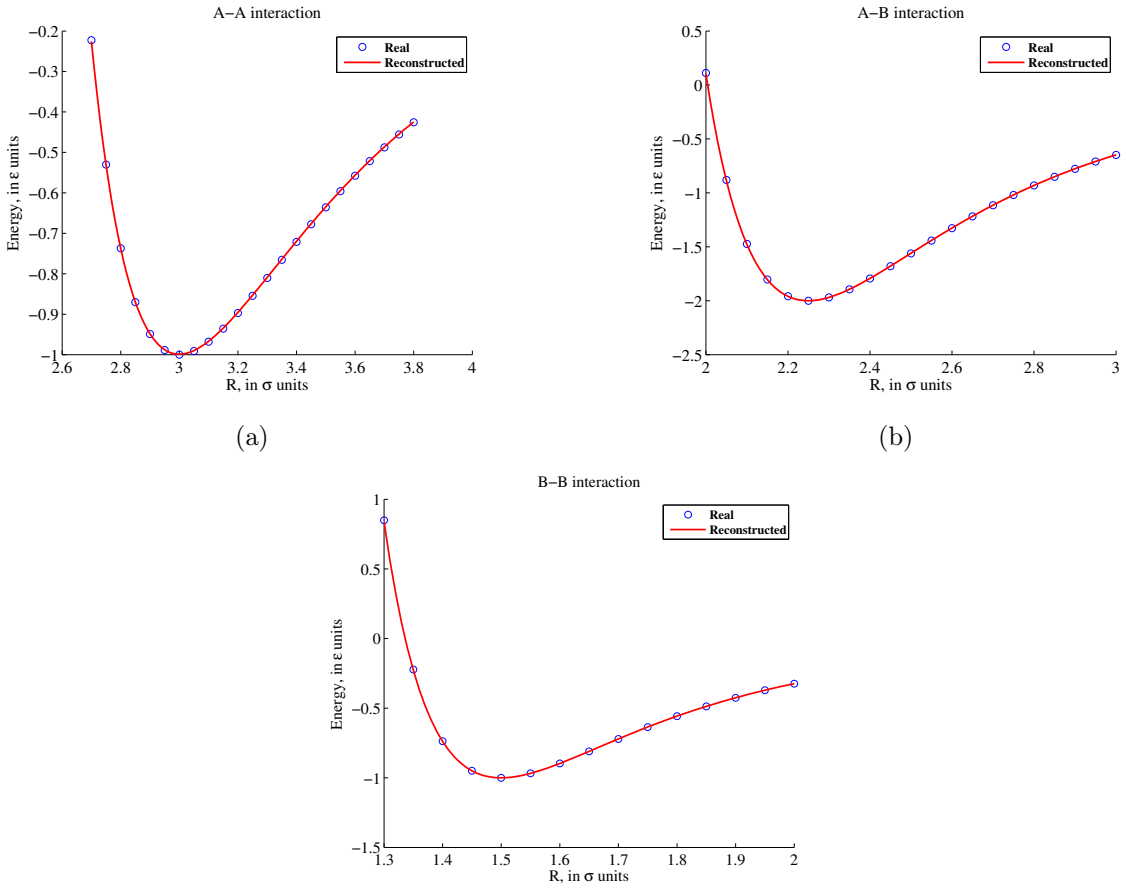
где ϵ — глубина потенциала, а r_m соответствует позиции минимума. Мы рассмотрели систему A_xB_y, где потенциалы имеют Леннард-Джонсовский тип со следующими коэффициентами: $\epsilon_{AA} = \epsilon, r_{AA} = 3\sigma; \epsilon_{AB} = 2\epsilon, r_{AB} = 2.25\sigma; \epsilon_{BB} = \epsilon, r_{BB} = 1.5\sigma$ — это подразумевает, что мы работаем в единицах ϵ и σ. Используя USPEX, мы сгенерировали 10000 рандомных (с произвольным числом атомов в ячейке) структур, а далее, используя GULP с нашими коэффициентами, мы расчитали энергии этих структур. Зная эти энергии, мы восстановили исходные потенциалы и на Рис. 2 изображены исходные потенциалы и восстановленые. Энергии структур были в промежутке $[-310; -19]$ в единицах ϵ. Полученная квадратичная величина ошибки (RMSE) была меньше, чем 0.025ϵ; коэффициент Пирсона равен 1; восстановленные коэффициенты почти совпадали с исходными (настолько, что потенциалы не различимы — см. Рис. 2).

Второй тест был немного сложнее: к потенциалу Леннарда-Джонса мы добавили кулоновское взаимодействие. Мы взяли ту же систему AB, но на этот раз с фиксированной пропорцией атомов A_nB_n, что позволило зафиксировать заряды на атомах: каждый A атом имел заряд $1e$ в условных единицах и, соот-

[3]см. мануал GULP на сайте: https://nanochemistry.curtin.edu.au/local/docs/gulp

ветственно, каждый B атом в каждой структуре имел заряд $-1e$. После того, как такие структуры были сгенерированы USPEXом и их энергия была посчитана GULPом, мы восстановили потенциалы. Результаты были даже лучше прежних; более того, удалось восстановить заряд $1e$ на каждом A атоме (если говорить более строго, то мы восстановили квадрат заряда).

Тем самым мы разработали новый метод, который позволяет восстанавливать потенциалы взаимодействий. В кристалле 2-х частичный вклад в энергию существенно больше многочастичного, и поэтому этот метод может быть использован для грубого предсказания энергий.



(a) (b)

(c)

Рис. 2: Восстановленный и исходный потенциалы в системе A_xB_y с Леннард-Джонсовским взаимодействием: (a) A-A потенциал; (b) A-B потенциал; (c) B-B потенциал.

$$E_{1/r} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{q_iq_j}{r_{ij}(l)} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{q_iq_j}{r_{ij}(l)}\operatorname{erfc}(g\cdot r_{ij}(l)) - \sum_{i}^{N}q_i^2\frac{g}{\sqrt{\pi}} + \\ + \frac{2\pi}{V}\sum_{i,j}^{N}\sum_{\vec{G}\neq 0}q_iq_j\exp(i\vec{G}\vec{r}_{ij})\frac{\exp(-\frac{G^2}{4g^2})}{G^2} \qquad (1)$$

$$E_{1/r^2} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{A_{ij}}{r_{ij}^2(l)} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}A_{ij}\frac{\exp(-r_{ij}^2(l)\cdot g^2)}{r_{ij}^2(l)} - \frac{1}{2}\sum_{i}^{N}A_{ii}g^2 + \\ + \frac{\pi^2}{V}\sum_{i,j}^{N}\sum_{\vec{G}\neq 0}A_{ij}\exp(i\vec{G}\vec{r}_{ij})\frac{\mathrm{erfc}(\frac{G}{2g})}{G} \tag{2}$$

$$E_{1/r^3} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{A_{ij}}{r_{ij}^3(l)} = \sum_{l=0}^{\infty}\sum_{i,j}^{N}A_{ij}\Big(\frac{\mathrm{erfc}(g\cdot r_{ij}(l))}{2r_{ij}^3(l)} + \frac{g}{\sqrt{\pi}}\frac{\exp(-r_{ij}^2(l)\cdot g^2)}{r_{ij}^2(l)}\Big) - \\ - \frac{2}{3\sqrt{\pi}}\sum_{i}^{N}A_{ii}g^3 + \frac{\pi}{V}\sum_{i,j}^{N}\sum_{\vec{G}\neq 0}A_{ij}\exp(i\vec{G}\vec{r}_{ij})\Big(-\mathrm{Ei}\Big(-\frac{G^2}{4g^2}\Big)\Big) \tag{3}$$

$$E_{1/r^4} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{A_{ij}}{r_{ij}^4(l)} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}A_{ij}\frac{\exp(-r_{ij}^2(l)\cdot g^2)}{r_{ij}^4(l)})(g^2\cdot r_{ij}^2(l)+1) + \\ + \frac{\pi^2}{2V}\sum_{i,j}^{N}\sum_{\vec{G}\neq 0}A_{ij}\exp(i\vec{G}\vec{r}_{ij})\Big(\frac{2g}{\sqrt{\pi}}\exp\Big(-\frac{G^2}{4g^2}\Big) - G\,\mathrm{erfc}\Big(\frac{G}{2g}\Big)\Big) + \\ + \frac{\pi^{3/2}}{V}\sum_{i,j}^{N}A_{ij}g - \frac{1}{4}\sum_{i}^{N}A_{ii}g^4 \tag{4}$$

$$E_{1/r^5} = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{A_{ij}}{r_{ij}^5(l)} = \sum_{l=0}^{\infty}\sum_{i,j}^{N}A_{ij}\Big(\frac{g\exp(-r_{ij}^2(l)\cdot g^2)}{3\sqrt{\pi}r_{ij}^4(l)}(2r_{ij}^2(l)\cdot g^2+3) + \\ + \frac{\mathrm{erfc}(g\cdot r_{ij}(l))}{2r_{ij}^5(l)}\Big) - \frac{4}{15\sqrt{\pi}}\sum_{i}^{N}A_{ii}g^5 + \frac{2\pi}{3V}\sum_{i,j}^{N}A_{ij}g^2 + \\ + \frac{2\pi}{3V}\sum_{i,j}^{N}\sum_{\vec{G}\neq 0}A_{ij}\exp(i\vec{G}\vec{r}_{ij})\Big(g^2\exp\Big(-\frac{G^2}{4g^2}\Big) + \frac{G^2}{4}\,\mathrm{Ei}\Big(-\frac{G^2}{4g^2}\Big)\Big) \tag{5}$$

4 Основной метод и результаты на Al

Наша модель состоит в том, что мы хотим описывать систему 2-х частичным взаимодействием, но при этом учесть многочастичные вклады, поэтому мы представим энергию в следующем виде:

$$E = E_{2body} + E_{manybody}, \tag{6}$$

где 2-х частичный вклад представляет описание потенциалом, а второе слагаемое описывает многочастичные поправки. Представим 2-х частичный вклад так: $E_{2body} = E_0 + struct.sums$. Написанное подразумевает, что в общем случае потенциал может зависеть от плотности, и этот вопрос будет рассмотрен ниже в Разделе 5. Первое слагаемое будем аппроксимировать с помощью структурных сумм, описанных выше, а второе — с помощью нейронных сетей. Настройку коэффициентов мы делаем итеративно: сначала подбираем коэффициенты 2-х частичного взаимодействия, а после с помощью нейронной сетки приближаем разность; когда сетка натренируется, то приближаем с помощью структурных сумм разность между полной энергией и той, что дает сетка и т.д.

Чтобы проверить наше предположение, что 2-х частичный вклад наиболее важен, мы вычисляем следующее:

$$\eta_{2body} = \left\langle \frac{E_{2body}}{E} \right\rangle,$$

где усреднение происходит по всем структурам выборки. Если это число большое, то система может быть описана потенциалом с хорошей точностью.

Дальше будем подразумевать, что у нас один тип атомов; мы проводили тест в основном на структурах Al.

4.1 Стратегия и данные

Выбор данных - это один из основных ключевых моментов метода. Сначала с помощью USPEX мы генерируем 10000-20000 случайных структур и на них обучаемся, что уже нормально настраивает коэффициенты. Затем мы ставим полный расчет USPEX и собираем данные 10-20 поколений; на этих 10-20 поколениях мы продолжаем тренировать уже только нейронную сеть. Натренированная сетка уже может быть использована для предсказания/релаксации. Мы используем классификатор, который будет судить о том, насколько данная структура новая для машинного обучения; если новая, то мы перетренируем сетку с того момента, как собрали данные первых поколений. Мы используем простейший классификатор, который описан в [11]: если X — это тренировочные данные, то x_{min}, x_{max} — это вектора отвечающая минимальным и максимальным значениям каждого из признаков; структура x будет новой, если не выполнено условие: $x_{min} \leq x \leq x_{max}$.

Отметим теперь, что другие научные группы [8, 11] используют в качестве данных — данные, полученные с помощью молекулярной динамики; это в частности означает, что каждая из структур уже изначально находится близко к

оптимальной. В этом смысле наш подход лучше тем, что мы работаем с любыми структурами; молекулярная динамика, в отличие от USPEX, не является универсальной для предсказания кристаллических структур.

4.2 Вектор-признаков

Теперь займемся описанием кристаллической структуры. Мы хотим, чтобы это описание было однозначным, т.е. два вектора-признаков должны быть различными для двух различных структур. Так, например, было показано (см. в моем предыдущем НИРе), что фингерпринт-функция (см. [18,19]) не обладает таким свойством. Помимо однозначности, нужно, чтобы описание было независимо по отношению к трансляциям и поворотам кристалла, к выбору элементарной ячейки и перестановкам порядка атомов.

В вектор-признаков мы включаем следующие слагаемые, отвечающие 2-частичному распределению (по сути интегралы от фингерпринт-функции):

1 объем элементарной ячейки

2 структурные суммы, начиная с $k \geq 4$.

3 суммы, следующего типа:

$$\sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{1}{r_{ij}(l)}\operatorname{erfc}(g\cdot r_{ij}(l)); \sum_{l=0}^{\infty}\sum_{i,j}^{N}\frac{\exp(-r_{ij}^2(l)\cdot g^2)}{r_{ij}^2(l)}$$

для некоторых g. Эти члены возникают в формулах для структурных сумм при $k = 1, 2$. Важно отметить, что мы суммируем в сфере такого радиуса, чтобы суммы уже сошлись.

4 параметр порядка, описанный в [18,19] и который пропорционален $\int F^2 dR$, где F — фингерпринт-функция. Оказывается, что для оптимальных структур чем выше параметр порядка, тем ниже энергия — см. [19].

Мы так же включили слагаемые, связанные с 3-х частичными вкладами — это слагаемое полностью заимствовано из [8]:

$$\sum_{j,k\neq i}\Big(\frac{1+\lambda\cos\theta_{jik}}{2}\Big)^{\xi}\exp(-\eta(R_{ij}^2+R_{ik}^2+R_{kj}^2))\cdot f_c(R_{ij})\cdot f_c(R_{ik})\cdot f_c(R_{kj})$$

$$f_c(x)=\begin{cases}0.5\cdot[\cos(\frac{\pi x}{R_c})+1] & x\leq R_c\\ 0 & R_c\leq x\end{cases},$$

где i-й атом внутри ячейки, а j-й и k-й атомы внутри сферы небольшого радиуса.

4.3 Результаты на Al

Мы тестировали много различных архитектур и пришли к заключению, что одного скрытого слоя достаточно. В таблице 1 можно увидеть качество обучения нейронных сетей с одним слоем в зависимости от размера скрытого слоя. В качестве данных мы использовали почти 11784 структур для обучения (что составляло где-то 80% всех данных) и 4242 структур для теста (20% от всех данных).

Таблица 1: Нейронные сетки с одним скрутым слоем в зависимости от размера этого слоя. Было использовано 17000 тренировочных структур и 4242 структур для теста. Коэффициета Пирсона везде около 99.9%.

H:	30	40	50	60	70	80
RMSE train:	0.0456	0.0463	0.0446	0.0444	0.0441	0.0429
RMSE test:	0.0480	0.0489	0.0472	0.0485	0.0479	0.0482

В итоге, мы использовали архитектуру $(57 - 5 - 1)$. Далее были собраны данные для структур Al: 11784 тренировочных структур и 2946 структур для теста. Это те самые данные, которые мы используем для начальной настройки: они были сгенерированы и обсчитаны с помощью USPEX, причем энергии менялись в промежутке $[-30, 0]eV$. Мы получили, что RMSE на обучении 0.055 $eV/atom$, а на тесте = 0.056 $eV/atom$. На рис. 4.3 изображен восстановленый потенциал взаимодействия атомов Al; видна осцилляция — это согласуется с теорией Фриделя об осцилляциях потенциала в металлах. Вклад 2-х частичного взаимодействия в полную энергию оказался равным 92%, что согасуется с начальным предположением о значении 2-х частичного вклада.

Далее мы собрали 15 поколений при расчете USPEX системы Al — вышло 692 структур, из которых 554 были использованы для обучения. Из 138 тестовых структур классификатор обнаружил всего 2 новые. Результаты следующие: RMSE on train = 14.3 $meV/atom$; RMSE on test = 17.6 $meV/atom$.

Подобные результаты говорят в пользу того, что они могут быть использованы для предсказания и/или релаксации; по крайней мере, эта стратегия может давать хорошие стартовые значения для дальнейших квантово-механических расчетов.

5 Приложения

Здесь рассмотрены следующие системы: потенциалы Al при разных плотностях; потенциалы для C и B систем, как системы, где ожидаются большой вклад многочастичных взаимодействий; некоторое благородные газы, так как там предполагается Леннард-Джонсовское взаимодействие. Для того, чтобы восстановить эти потенциалы, мы обычно генерировали и обсчитывали около 25000 рандомных стуктур с помощью USPEXa, а после применяли комбинированную схему.

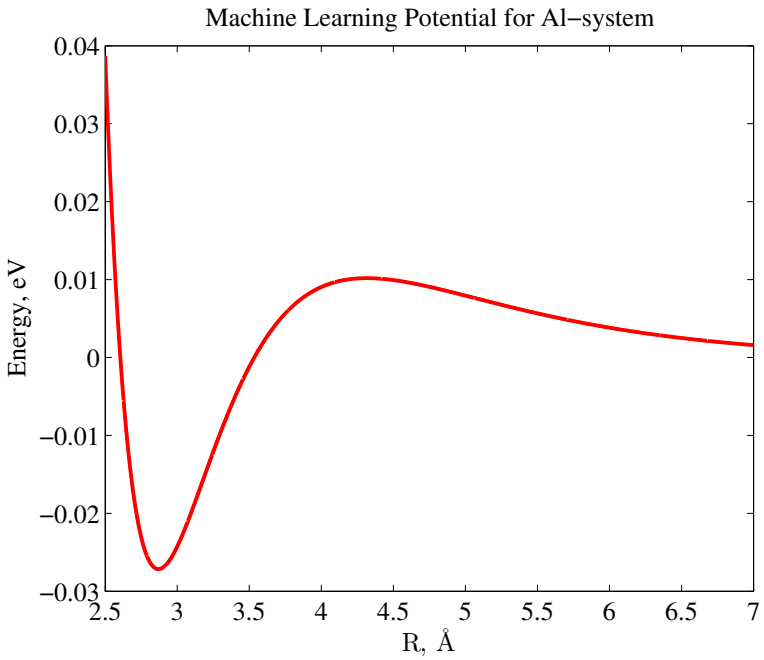


Рис. 3: Потенциал взаимодействия атомов Al, который получился в результате совместной настройки параметров сетки и весов при структурных суммах.

5.1 Al при разных плотностях

Как было написано выше, потенциал взаимодействия зависит от плотности и, применив, комбинированную схему мы рассмотрели эту зависимость — см. Рис. 5.1. Мы использовали следующие плотности: $8.62 \cdot 10^{-2}$ Å^{-3}, $7.09 \cdot 10^{-2}$ Å^{-3}, $6.02 \cdot 10^{-2}$ Å^{-3}, $5.23 \cdot 10^{-2}$ Å^{-3} и $4.63 \cdot 10^{-2}$ Å^{-3}, соответственно. Это потенциалы имеею похожую форму: у каждого есть фриделевская осцилляция; растояние, соответствующее минимуму, у всех одно и то же, что еще раз подтверждает теорию Фриделя. Это растояние близко к 2.86 Å — такое же расстояние между близжайшими соседями в сср структуре из атомов Al. Так же отметим, что мы получили, чем более плотная система, тем выше расположен потенциал.

5.2 Потенциалы для C и B систем

На Рис. 5.2 и Рис. 5.2 мы построили потенциалы для систем C и B. Потенциалы усреднены по различным плотностям (так как рандомные структуры, на которых восстанавливались потенциалы, имеют различные плотности). C-потенциал на Рис. 5.2 был интуитивно ожидаем, так как кратчайшее расстояние между атомами C в алмазе 1.54 Å. Для B системы мы рассмотрели различное число атомов в элементарной ячейке: 8, 10 и 12. Можно видеть, что эти 3 потенциала очень похожи и это свидетельствует о том, что потенциал B приблизительно универсален. Так же сам потенциал соответствует чистому отталкиванию, а это значит, что в этой системе все удерживается за счет многочастичных взаимодействий. Вклады 2-х частичного слагаемого в полную энергию оказались всего 75% для C-системы и около 80% для B-системы.

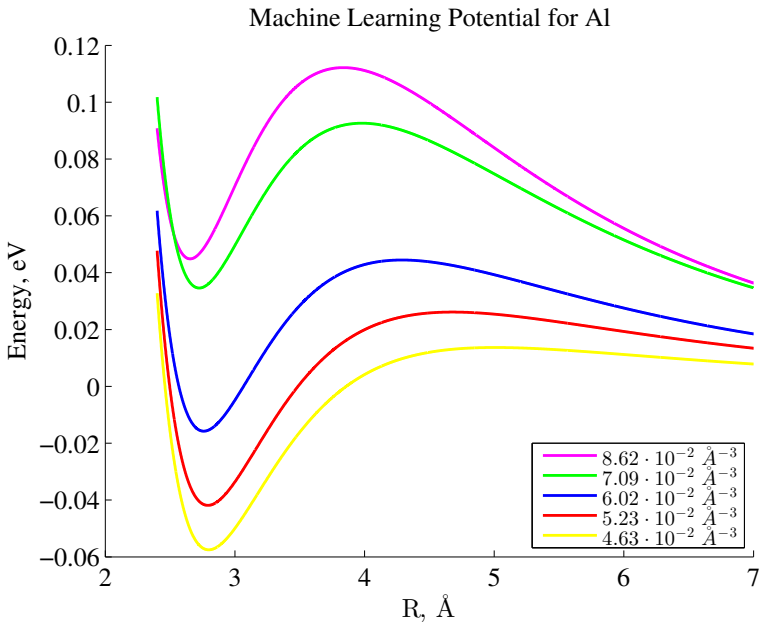


Рис. 4: Потенциалы Al в зависимости от плотности.

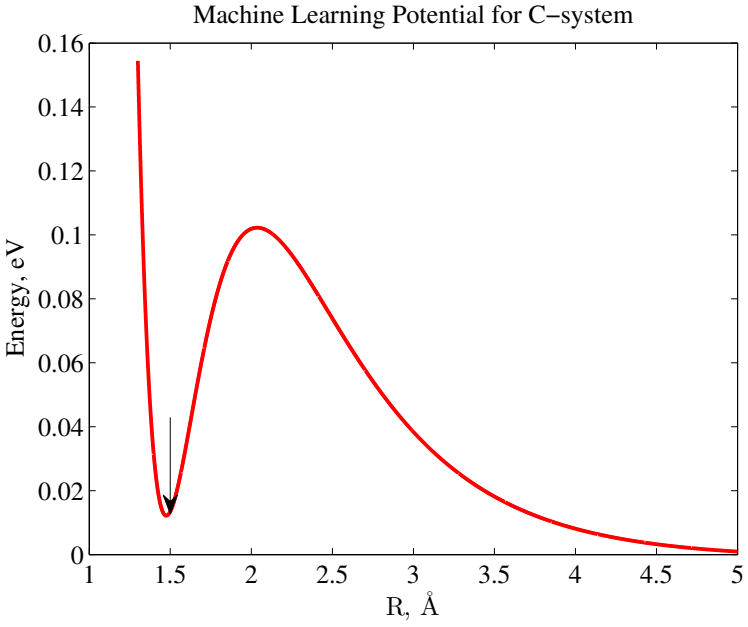


Рис. 5: C-потенциал. Стрелочка соответствует растоянию между соседями в алмазе.

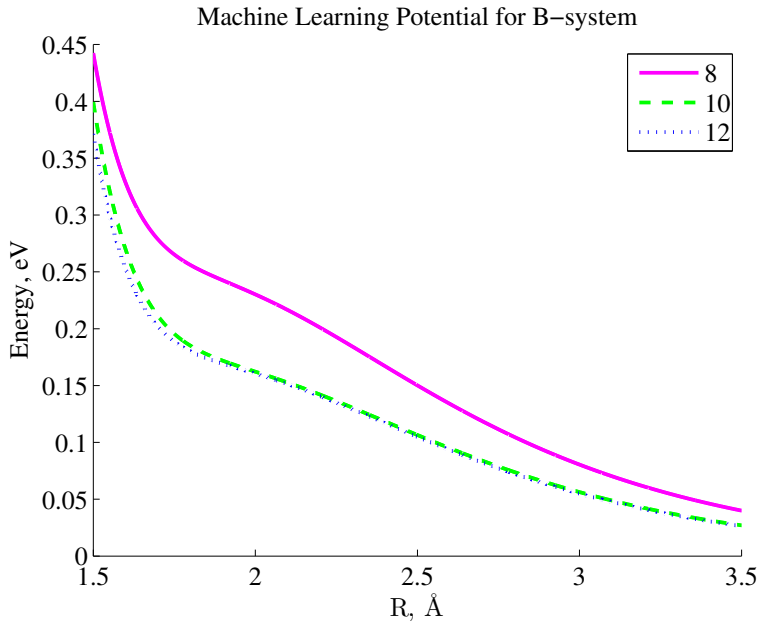


Рис. 6: B-потенциалы с различным числом атомов в элементарной ячейке: 8, 10, 12.

5.3 Потенциалы благородных газов

На Рис. 5.3 и Рис. 5.3 построены восстановленные потенциалы He и Xe. Обычно потенциалы благородных газов могут быть описаны классическим Леннард-Джонсовским взаимодействием, однако на графиках все же эти потенциалы отличается от чистого Леннарда-Джонса, так как при больших R значения потенциалов положительны. Тем не менее, полученный результат так же говорит в пользу того приближения, что при низких температурах He ведет себя, как жидкость: абсолютное значение потенциала при $R \geq 4$ Å меньше $5 \cdot 10^{-4}$ eV. В обоих случаях вклад двухчастичного взаимодействия оказался выше 95%. Более того, величина RMSE в системе He была меньше 0.003 $eV/atom$, что означает, что эта система может быть очень хорошо описана нашим методом.

6 Заключение

Была предложена и запрограммирована стратегия машинного обучения для предсказания кристаллических струкур, используя возможности USPEX. Метод работает пока для систем с одним типом атомов и показал хорошие результаты на системе Al. Попутно был предложен метод структурных сумм и с помощью этого метода был восстановлен потенциал в системах Al, B, C, He и Xe. Оказалось, что для Al 2-х частичный вклад в полную энергию составляет около 90%; для таких систем, как B и C — $70 - 80$%; для He и Xe выше 95%. Метод структурных сумм может быть использован, если нужно грубо оценить энергию; сам потенциал тоже содержит много информации о взаимодействии внутри кристалла.

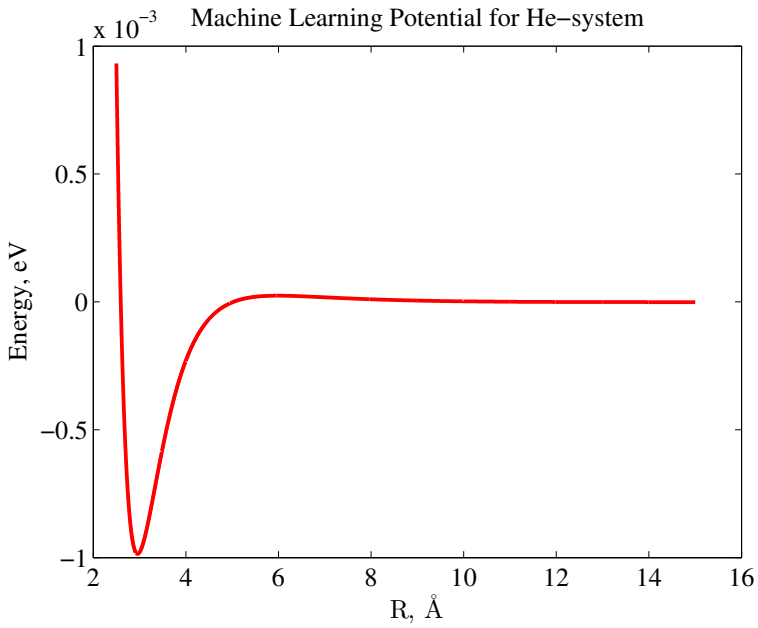


Рис. 7: Потенциал He.

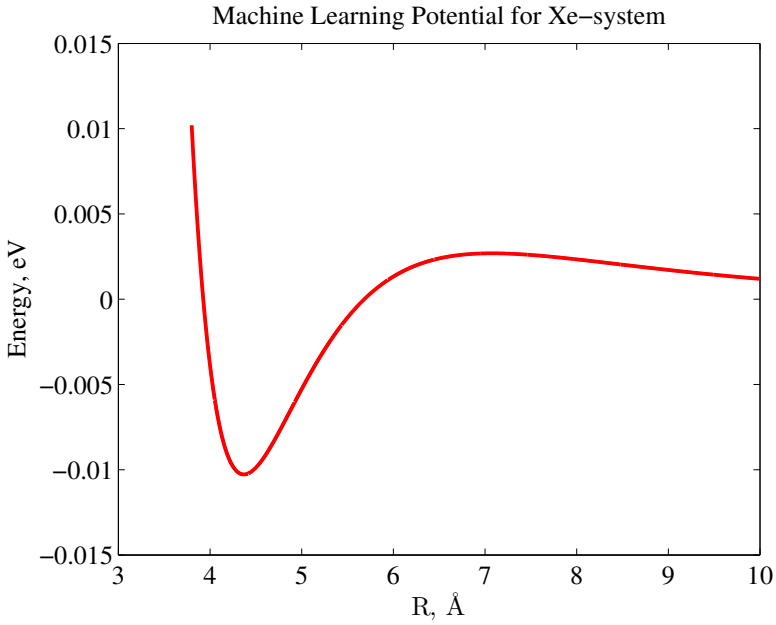


Рис. 8: Потенциал Xe.

В дальнейшем мы планируем расширить основной алгоритм, чтобы его можно было применять в системах, где есть дальнее взаимодействие.

Список литературы

1. J. Maddox, “Crystals from first principles”, Nature (London) **335**, 201 (1988).

2. W. Kohn and L. Sham, “Self-Consistent Equations Including Exchange and Correlation Effects”, Phys. Rev. **140**, A1133 (1965).

3. G. Kresse and J. Furthmüller, “Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set”, Phys. Rev. B **54**, 11169 (1996).

4. A.R. Oganov, C.W. Glass, “Crystal structure prediction using ab initio evolutionary techniques: Principles and applications”, J. Chem. Phys. 124, art. 244704 (2006).

5. C.W. Glass, A.R. Oganov, N. Hansen, “USPEX—Evolutionary crystal structure prediction”, Comp. Phys. Comm. 175 (2006).

6. W.W. Zhang, A.R. Oganov, A.F. Goncharov , Q. Zhu, S.E. Boulfelfel, A.O. Lyakhov, E. Stavrou, M. Somayazulu, V.B. Prakapenka, Z. Konopkova, “Unexpected Stable Stoichiometries of Sodium Chlorides”, Science **342**, 1502-1505 (2013).

7. A. Zielesny, "From Curve Fitting to Machine Learning Intelligent Systems Reference Library, Vol. 18, Springer (2011).

8. J. Behler and M. Parrinello, “Generalized Neural-Network Representation of High-Dimensional Potential-energy Surfaces”, PRL 98, 146401 (2007).

9. N. Artrith, et al., “High-dimensional neural-network potentials for multicomponent systems: Application to zinc oxide”, PR B 83, 153101 (2011).

10. J. Behler, "Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations”, PCCP **13**, 17930-17955 (2011).

11. Vonkatesh Botu and Rampi Ramprasad, “Ab-Initio Molecular Dynamics Acceleration Scheme with an Adaptive Machine Learning Framework”, arXiv:1410.3353v1, Oct 2014.

12. A.R. Oganov and M. Valle, “How to quantify energy landscapes of solids”, J. Chem. Phys. **130**, 104504 (2009).

13. К.В. Воронцов, “Лекции по искусственным нейронным сетям”.

14. Y. LeCun, L. Bottou, G.B. Orr and K. Muller, “Efficient BackProp”.

15. Y. LeCun, J.S. Denker and S.A. Solla, "Optimal Brain Damage", AT&T Bell Laboratories, Holmdel, N. J. 07733.

16. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale", Ann. Phys. **369** (3): 253–287; (1921).

17. M. Dove, "Introduction to lattice dynamics".

18. M. Valle and A.R. Oganov, "Crystal fingerprint space — a novel paradigm for studying crystal-structure sets", Acta Crystallographia Section A, vol. 66, pp. 507-517, Sept. 2010.

19. A.R. Oganov and M. Valle, "How to quantify energy landscapes of solids", J. Chem. Phys. **130**, 104504 (2009).