

КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

ИНСТИТУТ ФИЗИКИ

*Лаборатория компьютерного дизайна новых материалов и
машинного обучения*

ИЗУЧЕНИЕ ФИЗИЧЕСКИХ СВОЙСТВ
МАТЕРИАЛОВ С ПОМОЩЬЮ
МАШИННО-ОБУЧЕННЫХ ПОТЕНЦИАЛОВ,
СОЗДАНЫХ В ПРОГРАММНОМ ПАКЕТЕ
MEDEA®

УДК 538.9
ББК 22.37

*Утверждено на заседании учебно-методической комиссии Института физики
Казанского федерального университета
Протокол № 10 от 13 июня 2024 года
Протокол № 10 заседания кафедры общей физики от 4 июня 2024 года*

Рецензент:

Кандидат физико-математических наук, научный сотрудник лаборатории дизайна материалов, Сколковский институт науки и технологий **И. В. Чепкасов**

Изучение физических свойств материалов с помощью машинно-обученных потенциалов, созданных в программном пакете MedeA[®] [Текст]: учебное пособие / Р.Ф. Ахмеров, И.И. Гумарова, Р.М. Бурганова, О.В. Недопекин. – Казань: КФУ, 2024. – 62 с.

Теоретическое изучение физической природы и свойств неорганических и органических соединений как в кристаллическом (моно-, поликристаллы, наноструктуры), так и в аморфном состоянии (полимеры, композиты), в зависимости от их химического, изотопного состава, температуры и давления – является важным этапом исследования новых материалов.

В частности, особый интерес представляют расчеты электронной зонной структуры, динамики решетки и кристаллической структуры, построения фазовых диаграмм сложных многокомпонентных систем. Ставший в последние десятилетия популярным теоретический подход, основанный на приближенных решениях уравнения Шредингера (первопринципные методы), способен очень точно описывать свойства конденсированных сред. Однако теоретическое описание их свойств ограничено вследствие необходимости привлечения больших расчетных мощностей.

Сегодня, благодаря развитию компьютерных технологий и алгоритмов машинного обучения, стало возможным применять методы на основе силовых полей, например, молекулярную динамику, но обладающих точностью первопринципных методов. Это возможно благодаря тому, что машинное обучение происходит на основе первопринципных расчетов. Такой подход позволяет описывать сложные многокомпонентные системы за разумное время с хорошей точностью.

Настоящее учебное пособие призвано дать обзор подходов машинного обучения для создания потенциалов межатомного взаимодействия, принципов составления базы данных структур на основе расчетов из первых принципов, использования сгенерированных потенциалов в коде LAMMPS, реализующих алгоритм молекулярной динамики, для предсказания широкого диапазона свойств материалов. Настоящее пособие адресовано, в первую очередь, магистрам специальности «Физика перспективных материалов», обучающихся на курсе «Компьютерный дизайн новых материалов», а также широкому кругу читателей, интересующихся указанной проблемой.

Оглавление

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------|----|
| Введение | 4 |
| 1. Обзор методов моделирования материалов | 4 |
| 2. Основы теории машинного обучения | 10 |
| 2.1. Введение | 10 |
| 2.2. Машинно-обучаемые потенциалы | 12 |
| 3. Программный модуль MLP-G: создание машинно-обучаемых потенциалов на основе <i>ab initio</i> данных | 18 |
| 3.1. Создание обучающего набора | 18 |
| 3.2. Настройка и создание машинно-обучаемого потенциала | 19 |
| 4. Практические задания | 27 |
| 4.1. Использование готового машинно-обученного потенциала на примере предсказания свойств никеля (Ni) в коде LAMMPS | 27 |
| 4.2. Создание обучающего набора и генерация машинно-обученного потенциала для титана (Ti). Исследование фазовых переходов в титане | 32 |
| 4.3. Создание обучающих наборов и машинно-обучаемых потенциалов на примере поверхности Nb | 42 |

Введение

В учебном пособии представлен обзор возможностей машинного обучения для получения потенциалов межатомных взаимодействий и предсказания с их помощью свойств материалов. Приводится описание программного модуля MLPG (Machine Learning Potential Generator – генератор потенциала взаимодействия с помощью машинного обучения), интегрированного в программный пакет MedeA[®] [1]. Основная функция этого модуля – создание собственных межатомных потенциалов взаимодействия с помощью машинного обучения (machine learning potential – MLP). В дальнейшем созданные потенциалы могут быть использованы для моделирования методом классической молекулярной динамики с точностью, сопоставимой с расчетами из первых принципов, сокращая время расчетов.

В пособии приведены рекомендации по построению обучающего набора, выполнению вычислений методом теории функционала плотности ТФП (density functional theory – DFT) для всех структур, включенных в обучающий набор, и эффективному хранению соответствующих данных для использования в программном модуле MLPG. В первой части пособия представлен обзор методов моделирования, далее приведены основы теории машинного обучения для дизайна новых материалов, а также описаны имеющиеся в программном обеспечении машинно-обученные потенциалы и, наконец, пошагово разобраны примеры применения возможностей модуля MLPG для предсказания свойств физических систем.

1. Обзор методов моделирования материалов

Рассмотрим различные современные методы моделирования материалов. На Рис. 1 представлены методы, охватывающие различные временные и пространственные масштабы, включая метод конечных элементов и инженерное проектирование. Также на рисунке отмечены методы, реализованные в настоящее время в программном пакете MedeA[®]:

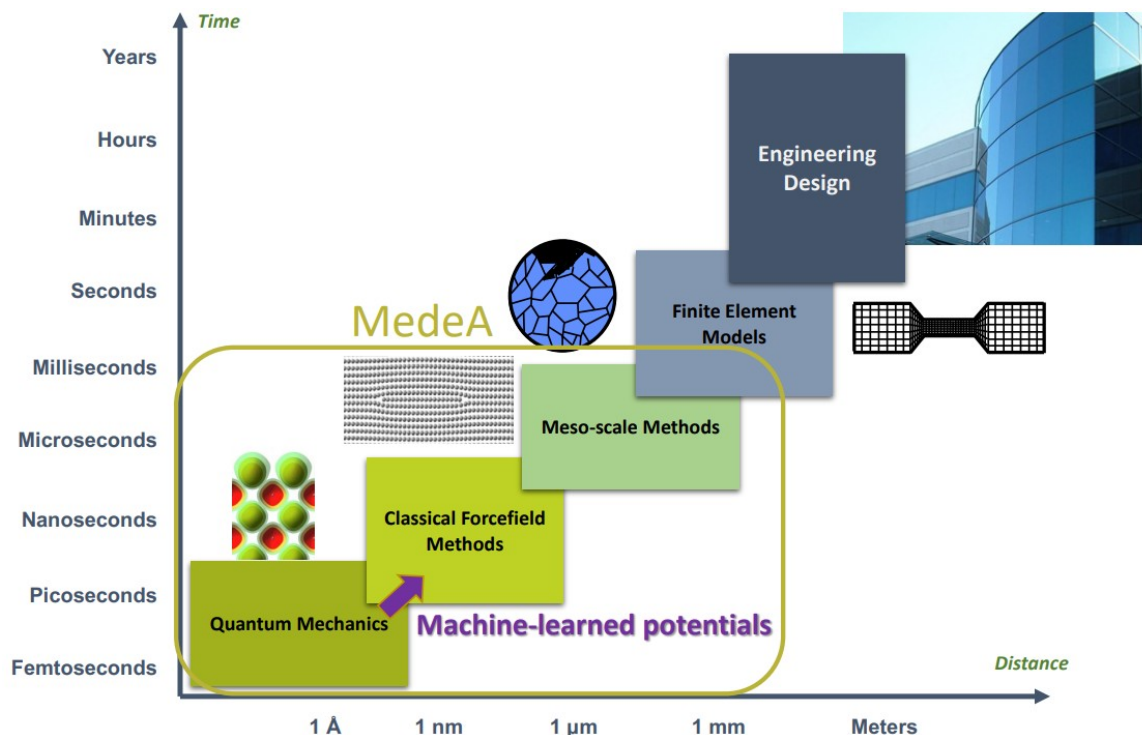


Рис. 1. Современные методы моделирования материалов

1. Для решения задач на очень больших временах и масштабах используются методы инженерного проектирования, самым распространённым из которых является метод конечных элементов.
2. Для решения задач на малых временных и пространственных масштабах используются первопринципные квантово-механические методы (*Ab initio quantum mechanics*). Они способны предсказать электронную конфигурацию заданной структуры, определяемую положениями атомов. На основании электронной конфигурации можно вычислить полные энергии и силы, а также предсказать многие другие свойства системы.
3. Для систем, определяемых на больших временных и пространственных масштабах, применяются классические методы с использованием силовых полей (*Forcefield Methods*). В этом подходе электроны или электронные конфигурации не учитываются, и модель упрощается до рассмотрения атомов и ионов. То есть в данном подходе электроны явно не моделируются, но можно, например, учитывать поля переноса заряда и реактивные силы.
4. Для моделирования материалов и процессов на еще больших временных и пространственных масштабах используется мезомасштабное моделирование, в котором вместо атомов и ионов рассматриваются сегменты структуры, например, повторяющиеся фрагменты блок-сополимеров или поверхностно-активных веществ. По сути, мезомасштабное моделирование связывает дизайн материалов в атомистическом масштабе с методами конечных элементов. Причем точность выбора форм наноразмерных сегментов и их взаимодействия играют решающую роль в определении свойств материалов.

Каждый из этих методов имеет свои преимущества и применяется в зависимости от конкретных задач и требуемой точности моделирования.

В приближении Борна-Оппенгеймера, в котором ядра атомов и электроны системы описываются отдельно, так как характерные времена изменения состояния ядер и электронов существенно различаются, энергия атомов, зависящая от положений ядер, известна как поверхность потенциальной энергии. Знание поверхности потенциальной энергии позволяет выполнять расчеты различных характеристик исследуемой системы, включая свободную энергию и другие термодинамические величины, моделировать эволюцию системы во времени, выявлять устойчивые и метастабильные атомные конфигурации, а также производить расчет колебательных мод и частот.

Наиболее точные представления о поверхности потенциальной энергии получаются из квантово-механических расчетов. Однако для большинства реальных систем получить точные решения уравнения Шредингера невозможно из-за чрезвычайной сложности таких вычислений, а использование методов, позволяющих приближенно решать уравнения Шредингера, ограничено доступными вычислительными ресурсами. Одним из наиболее распространенных методов приближенного решения уравнения Шредингера является теория функционала плотности (ТФП) (*Density Functional Theory – DFT*). Метод ТФП стал популярным благодаря компромиссу между скоростью расчетов и их эффективностью. Однако вычислительные затраты растут кубически с увеличением числа различных электронов в системе $\mathcal{O}(N^3)$ [2]. Поэтому ТФП редко используется для расчета термодинамических средних значений, или свойств систем, содержащих более 10^6 атомов, или для моделирования молекулярной динамики продолжительностью более 10 нс. На практике большинство вычислений ТФП ограничиваются меньшими временными и пространственными масштабами. Другие же квантово-механические методы обычно хуже масштабируются на

большие системы [3] или требуют привлечения дополнительных экспериментальных или предварительно вычисленных данных [4].

Квантово-механические методы моделирования основаны на фундаментальных законах физики и поэтому предоставляют точные результаты для большинства систем. Однако многие системы и явления, представляющие исследовательский интерес, требуют анализа на временных и пространственных масштабах, которые часто оказываются недоступными для квантово-механических расчетов из-за их высокой вычислительной сложности. Примерами таких задач можно назвать, например, процесс затвердевания металлических сплавов. В этом случае свойства материала зависят не только от его наноструктуры, но и от макроструктуры, и анализ должен охватывать временные масштабы, соответствующие производственному процессу, который определяет формирование конечного сплава.

Поэтому для подобных сложных задач были разработаны альтернативные методы для вычисления поверхности потенциальной энергии, которые обладают значительно более низкими вычислительными затратами. Эти методы известны как модели межатомного потенциала или силовые поля. В таких моделях взаимодействие между атомами описывается функцией, которая описывает их взаимодействие, и вычисление потенциальной энергии системы в этом случае масштабируется линейно с числом атомов в системе. Примеры простых межатомных потенциалов включают кулоновский потенциал [5], в котором ядра рассматриваются как точечные заряды, взаимодействующие по законам электростатики. Другим примером является потенциал Леннард-Джонса [6], который учитывает дисперсионное взаимодействие атомов на больших расстояниях, он пропорционален r^{-6} . Можно также выделить несколько следующих важных моделей. Одна из них – это потенциал Букингема [7], который представляет собой скорректированный потенциал Леннард-Джонса, включающий экспоненциальное отталкивание на коротких расстояниях. Еще одной моделью является потенциал Стиллингера–Вебера [8], специально разработанный для точного воспроизведения равновесного угла связи в кристаллическом кремнии и материалах с алмазоподобной структурой.

Другой более поздний подход представлен моделью погруженного атома (Embedded Atom Model – EAM) [9], которая базируется на теории приближения сильной связи (Tight-Binding model, TB) [10]. Метод погруженного атома учитывает не только парное взаимодействие атомов, но и распределение электронной плотности в системе. Работа Бреннера продемонстрировала идентичность потенциалов взаимодействия, полученных на основе концепции порядка связей (потенциалы порядка связей), с потенциалами в модели погруженного атома [11]. Эти потенциалы порядка связей, включая взаимодействия, обусловленные углами связи, известны как потенциалы Терзоффа [12]. Подобным образом, включение углов связи в метод погруженного атома привело к разработке модифицированной модели погруженного атома (Modified Embedded Atom Model – MEAM) [13], которая учитывает угловую зависимость электронной плотности в потенциале взаимодействия.

Среди более продвинутых методов стоит отметить оптимизированный по заряду многочастичный потенциал (Charge Optimized Many Body potential – COMB) [14]. Этот подход включает модель уравнивания зарядов, что позволяет успешно моделировать структуру сложных материалов, характеризующихся сосуществованием нескольких типов связей, таких как металлические, ковалентные, ионные и остаточные. Кроме того, существует еще один широко используемый метод, объединяющий несколько вышеупомянутых моделей, именуемый реактивным силовым полем (Reactive Force Field, ReaxFF) [15]. Этот метод способен учитывать динамическое образование и разрыв связей, а также поляризационные эффекты. Для более подробного изучения этих и других методов можно обратиться к соответствующим обзорам в научной литературе, например [16].

Модели межатомных потенциалов, построенные на физических принципах взаимодействия атомов и часто называемые «классическими» или «эмпирическими», использу-

ются для моделирования систем на особенно больших временных и пространственных масштабах. Эти методы позволяют современным процессорам проводить вычисления энергии со скоростью порядка 1 мкс/атом [17]. Однако несмотря на универсальность этих моделей потенциалов и применимость к различным системам, их точность в описании поверхности потенциальной энергии сильно ограничена. Кроме того, для каждой конкретной системы потенциал необходимо параметризовать. Более точные потенциалы, такие как SOMB и ReaxFF, обычно содержат сотни параметров, которые должны быть определены для каждого типа системы атомов. Следовательно, эти методы требуют больших вычислительных ресурсов по сравнению с более простыми моделями межатомными потенциалами [18]. Поэтому в настоящее время активно исследуются возможности применения машинного обучения для систематического определения оптимальных параметров таких потенциалов. Данная проблема подробно рассмотрена в обзоре, доступном в работе [19].

Так как квантово-механические расчеты представляют собой наиболее точный метод получения данных о системах, все чаще используется подход, заключающийся в подгонке моделей межатомных потенциалов к данным, полученным с использованием теории функционала плотности или других квантово-механических методов. Эрколесси и Адамс продемонстрировали, что такая подгонка потенциала к силам, вычисленным с использованием ТФП, является эффективной, поскольку расчеты с помощью теории функционала плотности предоставляют как энергию заданной конфигурации, так и, с небольшими дополнительными вычислительными затратами, силы, действующие на каждый атом. Это позволяет собрать обширный набор обучающих данных, содержащих в общей сложности $3N + 1$ точку данных (при условии отсутствия ограничений симметрии) для системы с N атомами [20].

Первоначально данный подход в основном применялся для параметризации существующих моделей потенциалов. Однако со временем он стал основой для современных методов, связанных с машинным обучением, таких как машинно-обучаемые потенциалы (Machine Learning Potentials – MLP). В данном подходе форма потенциала определяется прежде всего используемым методом машинного обучения, а не является упрощенным выражением основных физических взаимодействий.

Методы расчета энергии системы, основанные на машинном обучении, имеют долгую историю применения. Так, еще в 1995 г. Бланк и др. [21] продемонстрировали свой потенциал использования нейронных сетей с прямой связью для моделирования глобальных свойств поверхности потенциальной энергии на основе данных, полученных из ограниченного числа конфигураций изучаемой системы. В их исследовании было показано, что предложенный метод, основанный на нейронных сетях, позволяет быстрее получать межатомный потенциал. Более того, расчеты, проведенные с использованием этого потенциала, обладают более высокой точностью по сравнению с использованием исходного эмпирического потенциала. Этот пример, примененный к диффузии молекул CO на поверхности Ni, иллюстрирует эффективность методики на основе нейронных сетей в повышении точности и скорости расчетов энергии системы.

Так, машинно-обучаемые потенциалы межатомного взаимодействия предоставляют возможность выполнения расчетов на больших масштабах пространства и времени, сохраняя при этом точность, сопоставимую с квантово-механическим описанием. Важно отметить, что потенциал, полученный машинным обучением, не учитывает электронные конфигурации, но строится таким образом, чтобы точно воспроизводить поверхность потенциальной энергии на квантово-механическом уровне. Это позволяет рассчитывать силы, действующие на атомы в системе, и выполнять расчеты методом молекулярной динамики, что делает данный подход мощным инструментом для изучения и моделирования сложных систем. Поэтому моделирование с использованием классических методов и потенциалов, полученных с помощью машинного обучения, обеспечивает высокую скорость

и точность расчетов. Кроме того, возможность создания индивидуальных потенциалов для конкретных исследуемых систем на основе набора обучающих данных после нескольких симуляций демонстрирует гибкость этого подхода.

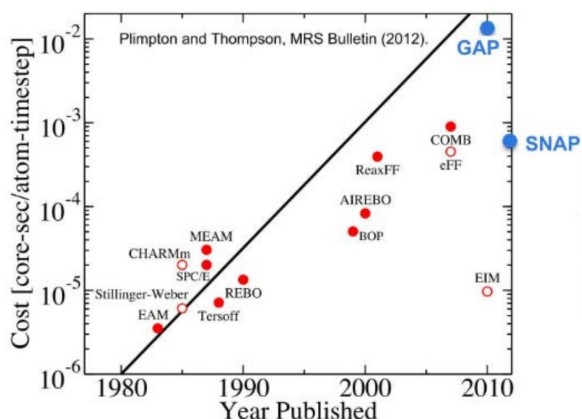


Рис. 2. Вычислительные затраты различных потенциалов на современном процессоре и годы их публикации. Линия соответствует удвоению вычислительных затрат каждые два года, что сходно закону Мура для эволюции сложностей аппаратных средств

На Рис. 2 проиллюстрирована растущая вычислительная сложность многочастичных потенциалов с течением времени. Так, до 80-х годов широко использовались потенциалы парного взаимодействия, такие как потенциал Леннарда-Джонса. Однако с развитием вычислительных возможностей стали разрабатываться более сложные межатомные потенциалы, включая потенциалы гауссовой аппроксимации (Gaussian Approximation Potential – GAP) и потенциалы спектрального анализа соседей (Spectral Neighbour Analysis Potential – SNAP), которые широко применяются в машинном обучении. Подобно известному закону Мура для экспоненциального роста количества транзисторов в полупроводниковых чипах [22], представленный график иллюстрирует рост вычислительной сложности в области многочастичных потенциалов с течением времени.

Точность расчетов, проводимых с помощью эмпирических потенциалов, можно сравнить с расчетами методом ТФП. На Рис. 3 отображена зависимость ошибки при расчетах с использованием различных потенциалов относительно ТФП расчетов в зависимости от вычислительных затрат. Более простые потенциалы, несмотря на их меньшие вычислительные затраты, обеспечивают лишь качественное описание систем. С увеличением сложности формы потенциала относительная ошибка уменьшается, и точность приближается к результатам, полученным методом квантовой механики. Методы машинного обучения требуют больших вычислительных затрат по сравнению с другими моделями потенциалов, однако они достигают наивысшей точности, близкой к результатам квантово-механических расчетов.

Итак, можно заключить, что одним из наиболее перспективных направлений применения машинного обучения в области дизайна материалов является создание быстрых и точных моделей межатомного потенциала. Эти модели позволяют значительно ускорить расчеты многокомпонентных систем и при этом обеспечивают высокую точность в описании межатомных взаимодействий, что делает их важным инструментом для исследований и разработки новых материалов.

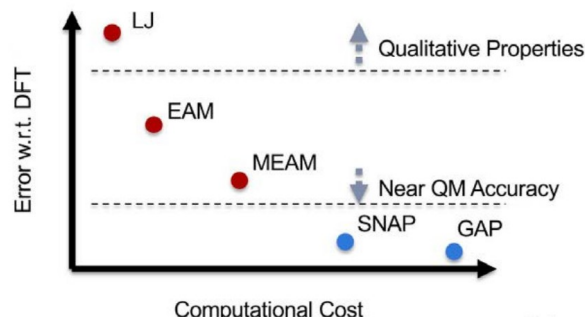


Рис. 3. Относительная ошибка различных потенциалов по сравнению с теорией функционала плотности в зависимости от вычислительных затрат

Контрольные вопросы и дополнительные задания:

1. Перечислите и кратко опишите существующие методы моделирования материалов. Для каких систем и решения каких задач используется каждый из них?
2. Что означает термин *ab initio*?
3. Почему метод ТФП получил такое широкое распространение по сравнению с другими методами расчетов? В чем основное преимущество ТФП перед другими квантово-механическими методами моделирования?
4. В чем заключаются суть приближения Борна-Оппенгеймера? Когда оно применимо?
5. Что представляет собой потенциал межатомного взаимодействия?
6. Какие преимущества и недостатки существуют у классических силовых полей?
7. Каковы преимущества и недостатки машинно-обучаемых потенциалов перед другими известными потенциалами (эмпирическими и полуэмпирическими)?

2. Основы теории машинного обучения

2.1. Введение

Разработка моделей межатомного потенциала относится к области машинного обучения, известной как машинное обучение с учителем [23]. Целью машинного обучения с учителем является определение функции f , которая точно предсказывает значения y на основе наборов входных данных x . В контексте моделей межатомного потенциала, x можно представить как тип атомов и координаты ядер, y – энергии этого состояния, а f – модель межатомного потенциала, которую нужно обучить. В общем случае, выбор подхода к машинному обучению с учителем состоит из трех основных шагов:

1. Определение пространства гипотез. Это пространство функций, из которых будет выбрана функция f . Поскольку пространство всевозможных функций бесконечно и может включать в себя множество функций, способных идеально подходить под обучающие данные, необходимо установить ограничения на это пространство гипотез. Например, ограничение может заключаться в использовании ограниченного набора базисных функций [24]. Такие функции, которые, вероятно, не будут обладать высокой предсказательной способностью (например, из-за их сложности или физической неприменимости), могут быть исключены из рассмотрения с использованием процедуры регуляризации.
2. Определение целевой функции. Эта функция будет использоваться для оценки того, насколько хорошо каждая модель из пространства гипотез соответствует требуемым данным. В контексте подгонки моделей межатомного потенциала типичной целевой функцией является квадрат ошибки в предсказании нормализованных значений энергии, сил и/или вириальных напряжений по сравнению с обучающими данными. Часто в эту целевую функцию также включается регуляризация.
3. Выбор метода поиска оптимальной функции в пространстве гипотез. В зависимости от характера пространства гипотез, метод выбора наилучшей функции может различаться. Например, если пространство гипотез состоит из нейронных сетей, то для нахождения наилучших весов в модели может применяться метод обратного распространения ошибки [25, 26]. Для пространства гипотез, представляющего собой линейную комбинацию базисных функций, часто можно использовать методы линейной алгебры для нахождения оптимальной функции.

Для построения модели межатомного потенциала с использованием методов машинного обучения, на пространство гипотез обычно накладывают три ограничения. Эти ограничения предназначены для обеспечения согласованности функции в этом пространстве с физическими принципами:

1. Атомы взаимодействуют только на коротких расстояниях. Это позволяет снизить вычислительные затраты и сложность модели межатомного потенциала, а также обеспечивает возможность линейного масштабирования потенциала относительно числа атомов. Обычно используется максимальный радиус (радиус отсечки) между взаимодействующими атомами, который составляет около 4–6 Å [27–29], хотя иногда этот радиус может быть больше [30]. Дальнодействующие силы, такие как кулоновское взаимодействие или дисперсионные силы, могут быть явно добавлены к короткодействующему потенциалу [31–35].

2. Потенциальная энергия должна быть инвариантна к перестановкам среди атомов одного и того же типа [36] и изометрическим преобразованиям системы (т. е. поворотам, отражению, перемещению и их комбинациям).
3. Поверхность потенциальной энергии должна меняться плавно в зависимости от положения атомов [37].

На практике, в рамках метода создания машинно-обучаемых потенциалов, результаты *ab initio* вычислений сводятся к установлению наиболее фундаментальной связи между структурой и свойствами (или значениями на поверхности потенциальной энергии). Другими словами, расчеты из первых принципов необходимы для создания наборов свойств, соответствующих разным структурам. Для того чтобы эта информация стала доступной для машинного обучения, необходимо уникальное, гибкое и эффективное описание расположения атомов в структурах обучающей выборки. В широко используемом подходе машинно-обучаемых потенциалов это реализуется путем построения так называемых дескрипторов (также известных как «отпечатки пальцев») локальной среды вокруг каждого атома в пределах радиуса отсечки, которые инвариантны к перестановкам среди атомов одного и того же типа и изометрии системы [29, 36, 38–46]. Затем алгоритм машинного обучения ищет функцию этих дескрипторов, оптимизируя целевую функцию.

На Рис. 4 показан пример дескрипторов атомной структуры, так называемые центрированные по атому функции симметрии [29]. Суть данного представления состоит в разделении структуры на локальные сферические кластеры с центром на данном атоме и заданным радиусом отсечки [29].

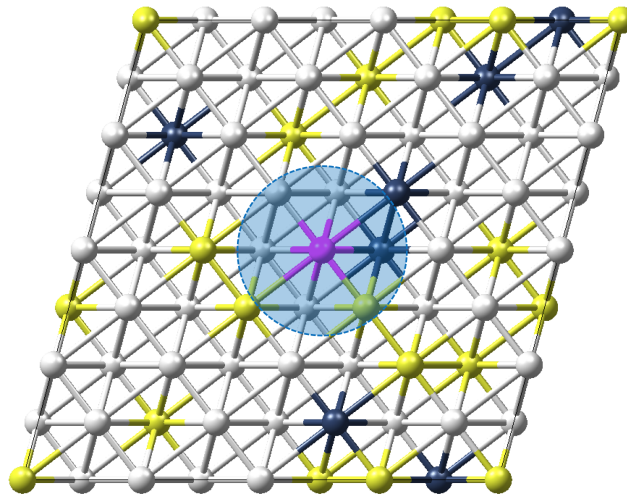


Рис. 4. Иллюстрация зависимости энергии атома от окружения для системы с периодическими граничными условиями. Энергетический вклад атома, выделенного розовым цветом, зависит от положения всех атомов внутри заштрихованной сферы, определяемой радиусом отсечки r_c функции симметрии

Примеры хорошо изученных дескрипторов включают в себя биспектральные компоненты [44], кулоновские матрицы [38] и метод гладкого перекрытия атомных позиций (Smooth Overlap of Atomic Positions – SOAP) [37]. Такие дескрипторы обычно используются в различных подходах к созданию машинно-обучаемых потенциалов, включая потенциалы нейронных сетей (Neural Network Potentials – NNP) через подход Белера-Парринелло [29], а также потенциалы гауссовой аппроксимации (GAP) [44] и потенциалы спектрального анализа соседей (SNAP) [45, 46], упомянутые в прошлом разделе. В последние годы также появилось несколько новых подходов, таких как потенциалы тензора

момента [24], сети передачи сообщений [32, 47–49] и символьная регрессия [45]. Дополнительные детали исследований, связанных с упомянутыми дескрипторами и методами для построения машинно-обучаемых потенциалов, могут быть найдены в соответствующей литературе. В данном пособии мы остановимся лишь на методах, реализованных в программном пакете MedeA[®].

2.2. Машинно-обучаемые потенциалы

Машинно-обучаемые потенциалы (Machine Learning Potentials – MLP) представляют собой некоторые универсальные функции, содержащие несколько десятков или даже сотен параметров. То есть машинно-обучаемые потенциалы часто имеют гораздо больше параметров, чем традиционные силовые поля. Уровень вычислительной эффективности машинно-обучаемых потенциалов зависит от количества параметров: чем меньше параметров, тем быстрее вычисления, но тем более ограничено его применение вне обучающего набора конфигураций исследуемого материала. Наоборот, машинно-обучаемые потенциалы с большим числом параметров могут описывать более широкий диапазон конфигураций материала с улучшенной точностью, что максимизирует ценность больших обучающих наборов, созданных с использованием *ab initio* вычислительных методов.

Машинно-обучаемые потенциалы с большим набором параметров требуют значительных вычислительных ресурсов как для их разработки, так и для их использования. Кроме того, универсальность машинно-обучаемых потенциалов и их функциональная форма, которая может быть оптимизирована автоматизированными методами, предполагает отказ от традиционных силовых полей, где отдельные компоненты потенциала имеют непосредственное физическое значение. Только весь машинно-обучаемый потенциал, взятый как единое целое, имеет физический смысл. Это означает, что машинно-обучаемый потенциал, оптимизированный для определенного обучающего набора конфигураций материала, будет отличаться от любого другого потенциала, оптимизированного для другого набора данных или с другими критериями подгонки. Добавление новых структурных конфигураций, новых компонентов материала или изменение критериев подгонки потребует совершенно новой оптимизации машинно-обучаемого потенциала.

На практике общий рабочий процесс создания машинно-обучаемого потенциала обычно включает в себя два шага, которые иногда могут повторяться:

1. Генерация большого набора данных об атомных структурах и связанных с ними энергиях, силах, тензорах напряжений и других свойств, полученных в результате вычислений с помощью теории функционала плотности.
2. Генерация потенциала с точки зрения эффективного и подходящего набора дескрипторов атомной структуры с помощью подгонки к данным ТФП. Этот шаг лучше всего описывается формулой $y = f(x, w)$, где x представляет собой набор дескрипторов атомных структур, y представляет собой набор свойств, который должен быть согласован в процессе подгонки, w – дополнительные параметры, повышающие гибкость подхода, f – машинно-обучаемая функция.

Программный пакет MedeA[®] поддерживает два машинно-обучаемых потенциала: потенциал нейронной сети (NNP) и потенциал спектрального анализа соседей (SNAP).

2.2.1. Потенциал нейронной сети

Рассмотрим подход, основанный на нейронных сетях, где в качестве дескрипторов атомной структуры используются центрированные по атому функции симметрии (Neural Network Potentials – NNP) через подход Белера-Парринелло [29].

Нейронные сети (НС) основаны на принципах работы нейронов мозга. Используя количество точек, в которых значение функции известно, параметры НС оптимизируются для воспроизведения входных данных в процессе «обучения», а затем используются для оценки функции в другом месте. Для представления поверхности потенциальной энергии расчеты ТФП обычно используются для предоставления набора обучающих данных. После обучения координаты атомов передаются в нейронную сеть, на выходе получается потенциальная энергия, из которой также могут быть аналитически рассчитаны силы [21].

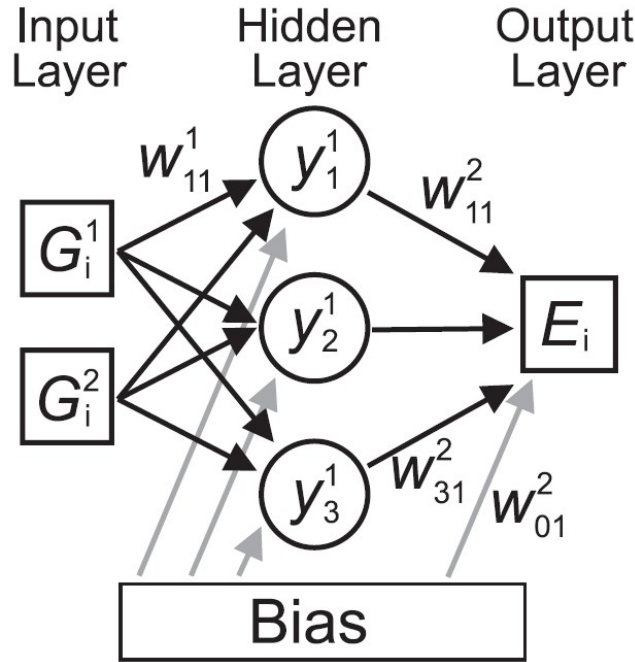


Рис. 5. Пример стандартной нейронной сети, используемой для подбора поверхностей потенциальной энергии [21, 50]. Узел в выходном слое дает энергию E_i , которая в данном случае зависит от значений двух входных узлов, G_i^1 и G_i^2 . Между входным и выходным слоями есть скрытый слой с тремя узлами, представленными кружками. Стрелки соответствуют 13 весовым параметрам w_{ij}^k , которые соединяют узел j в слое k с узлом i в слое $k - 1$. Узел смещения используется для адаптации области нелинейности функций активации. Функциональная форма этой небольшой сети дана в уравнении (1)

Структура простой нейронной сети, которая до сих пор использовалась для представления поверхности потенциальной энергии, схематично показана на Рис. 5 для двумерного случая. В узлах входного слоя заданы две обобщенные координаты G_i^1 и G_i^2 , определяющие энергию конфигурации i . Узел в выходном слое дает соответствующую энергию E_i . Между входным и выходным слоями находится один или несколько «скрытых слоев», каждый из которых имеет определенное количество узлов. Все узлы в каждом слое связаны с узлами в соседних слоях вещественными весовыми параметрами, которые первоначально выбираются случайным образом. Затем для заданного набора координат выход нейронной сети задается выражением

$$E_i = f_a^2 \left[w_{01}^2 + \sum_{j=1}^3 w_{01}^2 f_a^1 \left(w_{0j}^1 + \sum_{\mu=1}^2 w_{\mu j}^1 G_i^\mu \right) \right], \quad (1)$$

здесь w_{ij}^k – весовой параметр, соединяющий узел j в слое k с узлом i в слое $k - 1$, а w_{0j}^k – вес смещения, который используется в качестве регулируемого смещения для функций активации f_a^k . Функции активации, как правило, представляют собой нелинейные функции,

которые вводят возможность подгонки нелинейных функций к нейронной сети [21, 50]. В настоящей работе в качестве функции активации в скрытых слоях использовался гиперболический тангенс, а в выходном слое – линейная функция. Поскольку весовые параметры изначально выбираются случайным образом, выход нейронной сети не соответствует правильной полной энергии, но поскольку последняя известна для набора точек из вычислений ТФП, можно построить функцию ошибок и минимизировать ее для оптимизации весовых параметров итерационным способом. Затем полученный оптимизированный набор весов можно использовать для расчета потенциальной энергии для нового набора координат.

Эта структура нейронной сети имеет ряд недостатков, которые препятствуют ее применению для расчета поверхности потенциальной энергии большой размерности. Поскольку все веса, как правило, различны, порядок, в котором координаты конфигурации вводятся в нейронную сеть, не является произвольным, и замена координат двух атомов изменит общую энергию, даже если два атома относятся к одному и тому же типу. Другим ограничением, связанным с фиксированной структурой сети, является тот факт, что нейронная сеть, оптимизированная для определенного числа степеней свободы, т. е. числа атомов, не может быть использована для предсказания энергий для другого размера системы, поскольку оптимизированные веса действительны только для фиксированного числа входных узлов. Таким образом, чтобы представить поверхности потенциальной энергии, пригодные для систем всех размеров, необходимо ввести новую топологию нейронной сети.

Основная идея состоит в том, чтобы представить полную энергию E системы как сумму атомных вкладов E_i , подход, который обычно также используется в эмпирических потенциалах

$$E = \sum_i E_i. \quad (2)$$

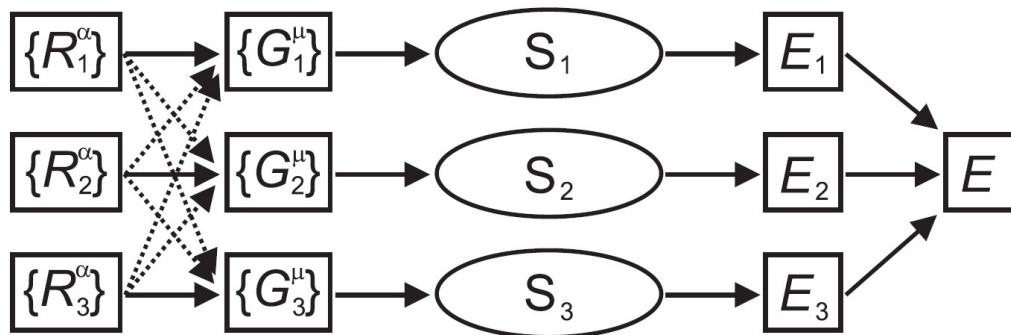


Рис. 6. Общая структура нейронной сети, содержащей три атома Рис. 6. Декартовы координаты атома i задаются R_i^α . Они преобразуются в набор значений μ функции симметрии G_i^μ , описывающих локальное геометрическое окружение атома i , которое зависит от положений всех атомов в системе, как показано пунктирными стрелками. Значения функции симметрии атома i затем попадают в подсеть S_i , что дает энергетический вклад E_i атома i в полную энергию системы E . Структура подсетей соответствует нейронной сети, показанной на Рис. 5

Общая структура этой новой сетевой топологии схематично показана на Рис. 6 для системы, состоящей из трех атомов и всех связанных с ними степеней свободы. R_i^α представляют декартовы координаты α атома i . На первом этапе эти координаты преобразуются в набор значений функции симметрии G_i^μ для каждого атома i . Эти значения функции симметрии описывают энергетически значимое локальное окружение каждого атома и впоследствии используются в качестве входных данных для нейронной сети. Они зависят от положения всех атомов в системе, как показано пунктирными стрелками.

Теперь для каждого атома в системе существует «стандартная» нейронная сеть (см. Рис. 5), которую мы называем подсетью S_i и которая после весовой оптимизации дает энергетический вклад E_i в общую энергию E . Сумма всех энергетических вкладов дает полную энергию системы. Для обеспечения инвариантности полной энергии по отношению к обмену двумя атомами структура всех подсетей и значения весовых параметров должны быть одинаковыми в каждом S_i .

В этом подходе ключевым моментом является введение нового типа функции симметрии. В то время как другие типы функций симметрии использовались ранее [21], тут значения функции симметрии каждого атома отражают локальное окружение, определяющее его энергию; т. е. две структуры с разной энергией должны давать разные наборы значений функции симметрии, а идентичные локальные среды должны давать один и тот же набор. Кроме того, значения функции симметрии должны быть инвариантны по отношению к вращению или перемещению системы. Наконец, количество функций симметрии не должно зависеть от координации атома, потому что координационное число атома может измениться при молекулярно-динамическом моделировании, а структура подсетей не должна изменяться, если нейронная сеть остается применимой в целом.

Функции симметрии могут быть построены из позиций атомов аналогично эмпирическим потенциалам. Но если в последнем случае эти члены используются для непосредственного построения полной энергии системы, то в случае нейронной сети они используются только для описания структуры. Назначение энергий структурам выполняется на втором этапе нейронной сети.

Чтобы определить энергетически значимое локальное окружение, используется функция отсечки f_c межатомного расстояния R_{ij} , которая имеет вид

$$f_c(R_{ij}) = \begin{cases} \frac{1}{2} \left[1 + \cos \left(\frac{\pi R_{ij}}{R_c} \right) \right] & R_{ij} \leq R_c; \\ 0 & R_{ij} > R_c. \end{cases} \quad (3)$$

При межатомных расстояниях больше порога R_c эта функция дает нулевое значение и наклон. Отсечка должна быть достаточно большой, чтобы включать несколько ближайших соседей. В данной работе использовалась отсечка в 7 \AA .

Радиальные функции симметрии строятся в виде суммы гауссиан с параметрами η и R_s ,

$$G_i^1 = \sum_{j \neq i}^{all} e^{-\eta(R_{ij}-R_s)} f_c(R_{ij}), \quad (4)$$

где суммирование по всем соседям j обеспечивает независимость координационного номера.

Угловые члены строятся для всех троек атомов путем суммирования значений косинусов углов $\theta_{ijk} = \frac{\mathbf{R}_{ij}\mathbf{R}_{ik}}{R_{ij}R_{ik}}$ с центром в атоме i , с $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$,

$$G_i^2 = 2^{1-\zeta} \sum_{j,k \neq i}^{all} (1 + \lambda \cos \theta_{ijk})^\zeta e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}), \quad (5)$$

с параметрами $\lambda = (+1, -1)$, η и ζ . Умножение на три функции обрезания и на гауссиан обеспечивает плавный спад до нуля в случае больших межатомных расстояний. Отметим, что G_i^μ в уравнениях (4) и (5) зависят от всех положений атомов внутри радиуса отсечки. Используется несколько функций каждого типа с разными значениями параметров. Выбор функций симметрии и их параметров не уникален и не должен быть уникальным, и можно использовать многие типы функций, если набор значений функций подходит для описания окружения атома.

2.2.2. Потенциал спектрального анализа соседей

Альтернативная схема, предложенная Бартоком и его коллегами, позволяет определить расположение атомов в каждом локальном кластере с точки зрения их плотности, как показано в уравнении (1) [44, 45]. В данном подходе δ -распределение устанавливает соседние положения в локальной среде. (В рамках подхода SOAP каждое δ -распределение заменяется распределением Гаусса, чтобы плотность выражалась плавно изменяющейся функцией). Затем плотность раскладывается по базису сферических гармоник, как показано в уравнении (2) [51, 52]. Однако прямое разложение по сферическим гармоникам не было бы инвариантным относительно поворотов. Эта проблема решается введением «биспектра» сферических гармоник, который обладает вращательной инвариантностью и вместе с тем включает в себя всю информацию об углах, содержащуюся в описываемых конфигурациях материала [44, 45]. Кроме того, радиальная зависимость атомной плотности внутри каждого кластера выражается фиктивным «углом» θ_0 , что позволяет представить атомную плотность как разложение по математически полному набору четырехмерных сферических гармоник, где θ и ϕ — обычные углы, а θ_0 содержит радиальную информацию, масштабированную до радиуса отсечки, деленного на π . Таким образом, для данного распределения атомов в локальном кластере это четырехмерное представление плотности полностью определяется коэффициентами $u_{m,m'}^j$, входящими в уравнение (2). Наконец, это представление может быть преобразовано в выражение для коэффициентов биспектра B , как показано в уравнении (3), которые содержат всю информацию о распределении атомов, необходимую для построения машинно-обучаемого потенциала. В уравнении (3) C — это известные коэффициенты Клебша-Гордана, представляющие собой интегралы произведений трехмерных сферических гармоник на их комплексно-сопряженные формы.

$$\rho_i(r) = \delta(r) + \sum_{r_{ij} < r_c} f_c(r_{ij}) w_j^{atom} \delta(r - r_{ij}), \quad (1)$$

$$\rho_i(r) = \sum_{j=0}^{\infty} \sum_{m'=-j}^j u_{m,m'}^j U_{m,m'}^j(\theta, \phi, \theta_0), \quad (2)$$

$$B_{j_1, j_2, j} = \sum_{m_1, m'_1 = -j_1}^{j_1} \sum_{m_2, m'_2 = -j_2}^{j_2} \sum_{m, m' = -j}^j (u_{m,m'}^j) C_{j_1 m_1 j_2 m_2}^{j m} \times u_{j_1 m'_1, j_2 m'_2}^{j m'} u_{m_1, m_1}^{j_1} u_{m_2, m_2}^{j_2}. \quad (3)$$

Когда коэффициенты биспектра определены, полные энергии и силы, рассчитанные для структур обучающего набора, записываются в виде суммы по атомам, как указано в уравнении (4) и (5) [44, 45]. Опять же, начиная с первоначальных предположений, параметры β затем корректируются до тех пор, пока полная энергия, определенная из этих значений, не согласуется со значением, полученным в результате соответствующего *ab initio* расчета с заданной точностью. В конце концов, эта процедура приводит к потенциалу гауссовой аппроксимации (Gaussian approximation potential – GAP), который был первоначально предложен Бартоком и его коллегами и использует байесовскую регрессию для определения значений параметров [44].

$$E_{SNAP} = \sum_{i=1}^N \beta_0^{\alpha_i} + \sum_{k=1}^K \beta_k^{\alpha_i} B_k^i, \quad (4)$$

$$F_{SNAP}^j = - \sum_{i=1}^N \beta^{\alpha_i} \frac{\partial B^i}{\partial r_j}. \quad (5)$$

Напротив, потенциал спектрального анализа ближайших соседей (spectral neighbor analysis potential – SNAP), представленный Томпсоном и соавторами [45, 53], используя то же представление локального окружения, что и потенциал гауссовой аппроксимации, пытается определить параметры β с помощью линейной регрессии, а не регрессии гауссовского процесса. Этот выбор дает явные преимущества. В частности, по мнению авторов, этот шаг делает процесс подгонки потенциала более надежным, для которого скорость расчетов методами молекулярной динамики не зависит от размера обучающей выборки. Линейная регрессия также подходит для больших обучающих наборов и, наконец, позволяет проводить прямой анализ чувствительности метода.

На практике генерация потенциалов спектрального анализа соседей происходит следующим образом:

1. Строится обучающий набор структур, который используется для расчета полных энергий, сил и напряжений с помощью *ab-initio* расчетов.
2. Локальная структура каждого атома каждой структуры из набора представляется с помощью четырехмерного разложения по сферическим гармоникам в терминах биспектральных коэффициентов B , которые играют роль локальных структурных дескрипторов.
3. Линейная регрессия используется для получения значений коэффициентов β , входящих в разложение полной энергии в локальных структурных дескрипторах B .
4. Наконец, гиперпараметры, такие как количество структурных дескрипторов, радиусы отсечки локальных атомных кластеров и веса атомных типов, оптимизируются с использованием числового пакета, такого как программное обеспечение DAKOTA, или любого другого кода оптимизации, например, код NEWUOA Пауэлла [54].

Контрольные вопросы и дополнительные задания:

1. В чем заключается суть машинного обучения с учителем?
2. Что такое дескрипторы локальной среды?
3. В чем разница между классическими силовыми полями и машинно-обучаемыми потенциалами?
4. Преимущества и недостатки больших обучающих наборов.
5. Из каких практических шагов состоит процесс создания машинно-обучаемого потенциала?
6. Какой подход в построении дескрипторов локальной среды реализуется при генерации потенциала нейронной сети? Опишите суть данного подхода.
7. Как происходит генерация потенциала нейронной сети?
8. Как называется и как реализуется подход в построении дескрипторов локальной среды в подходе спектрального анализа соседей?
9. Как происходит генерация потенциала спектрального анализа ближайших соседей?

3. Программный модуль MLPG: создание машинно-обучаемых потенциалов на основе *ab initio* данных

Модуль генератора машинно-обучаемого потенциала (MLPG) использует рассчитанные из первых принципов энергии, силы и напряжения для создания машинно-обучаемых потенциалов (MLP), которые, как и классические силовые поля, могут быть использованы в коде LAMMPS для расчета интересующих свойств методами классической молекулярной динамики. Хорошо обученный машинный потенциал может быть использован для произвольных конформаций, если они покрываются обучающим набором. При этом точность сравнима с расчетами из первых принципов, используемых для создания потенциала, но с гораздо меньшими вычислительными затратами. Однако точность предсказания машинно-обученного потенциала может уменьшиться, если исследуемая конфигурация системы находится вне диапазона, охватываемого обучающей выборкой.

В генераторе машинно-обучаемого потенциала программного пакета MedeA[®] реализованы два подхода: потенциал спектрального анализа соседей (SNAP) и потенциал нейронных сетей (NNP). Потенциал нейронной сети генерируется с использованием кода n2p2, разработанного Синграбером и соавторами [55] с весами и смещениями нейронной сети, состоящей из двух скрытых слоев, каждый из которых содержит определяемое пользователем количество нейронов, оптимизированных с помощью подхода затухающей памяти с фильтром Калмана. Чтобы использовать сгенерированный таким образом потенциал нейронной сети, необходим интерфейс n2p2 LAMMPS [55]. Потенциал спектрального анализа соседей создается с использованием кода FitSNAP [45, 46] в сочетании с оптимизатором без использования производной Пауэлла (Powell's Derivative Free Optimizer) для подбора так называемых гиперпараметров, таких как, например, радиусы отсечки и веса различных типов атомов [54].

Генерация и оптимизация машинно-обученного потенциала обычно выполняется в три этапа. Во-первых, создается обучающая выборка на основе расчетов из первых принципов для ряда различных систем, различающихся по структуре и, при необходимости, по химическому составу. Во-вторых, машинно-обученный потенциал получается путем минимизации разности между энергиями, силами и напряжениями, полученными из первых принципов и машинно-обученным потенциалом. Наконец, машинно-обученный потенциал проверяется путем прогнозирования выбранных свойств систем, не включенных в обучающий набор, и сравнения с результатами, полученными из первых принципов.

3.1. Создание обучающего набора

Первая задача при создании машинно-обученных потенциалов состоит в том, чтобы определить и создать подходящий начальный набор обучающих данных. Этот набор данных остается статичным, и его состав определяет природу материальной системы и физические/химические свойства, которые могут быть описаны с помощью машинно-обученного потенциала. Примеры свойств материала, на которые можно ориентироваться, включают параметры решетки, коэффициенты упругости, свободную энергию объемных структур, а также фазовую стабильность, энергию поверхности, энергию образования дефектов (например, вакансий и междоузлий), фононную дисперсию, коэффициенты диффузии и коэффициенты теплового расширения. Эта базовая информация служит основой для создания структурных моделей, которые будут включены в обучающую выборку.

Обычно помимо структур основного состояния необходимо учитывать и деформированные структуры. Деформированные структуры могут быть получены путем изотропного или одноосного сжатия или расширения, или за счет деформаций сдвига. Дополнительные структуры, которые могут потребоваться, включают структуры с ваканси-

ями и/или межузельными атомами. Наконец, следует также включать структуры, полученные моделированием методом молекулярной динамики из первых принципов (*ab initio* Molecular Dynamics – *ab initio* MD). В идеале, структуры, полученные методом *ab initio* MD, должны создаваться при температурах, охватывающих диапазон, описываемый машинно-обученным потенциалом.

После создания структур обучающей выборки с помощью инструментов, предоставляемых программным обеспечением MedeA[®], выполняются расчеты из первых принципов. На этом этапе важно использовать одни и те же вычислительные параметры для всех расчетов, чтобы гарантировать согласованность данных обучающего набора, то есть энергий, сил и напряжений. Согласованность вычислительных параметров явно снижает систематические ошибки, присущие выбранному функционалу ТФП. Результаты всех этих расчетов сохраняются в виде обучающего набора, который создается с помощью MedeA Fitting Data Manager.

Важно! Все структуры в обучающем наборе должны иметь симметрию P1, чтобы генератор машинно-обучаемого потенциала их учитывал.

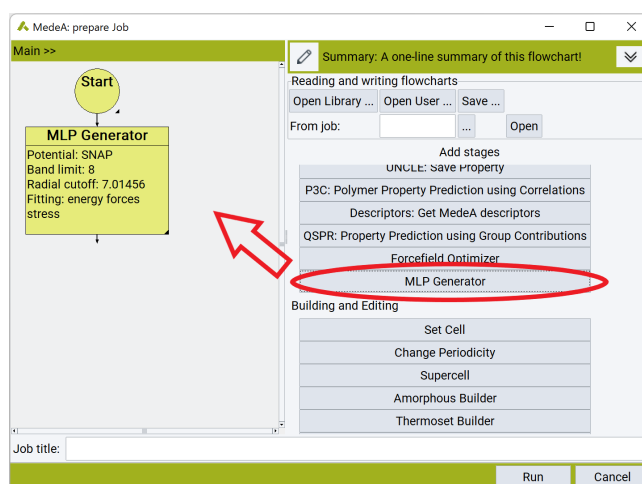
Всегда проверяйте, чтобы вычислительные параметры всех расчетов ТФП, используемых для обучающего набора, особенно PAW-потенциалы, соответствовали энергии отсечки плоской волны и k-сетки.

Оптимальное значение энергии отсечки плоской волны зависит от используемых PAW потенциалов. Для расчетов должно быть выбрано самое большое значение из всех потенциалов, используемых для обучающей выборки.

Если изучаемая система является металлом, то следует использовать шаг k-сетки меньше 0.3 1/Å.

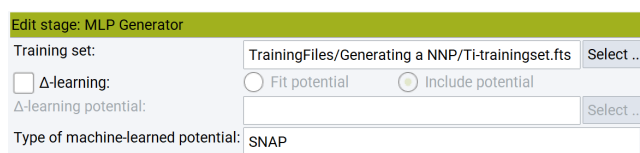
3.2. Настройка и создание машинно-обучаемого потенциала

Доступ к генератору машинно-обучаемого потенциала осуществляется с помощью интерфейса MedeA flowchart. Это можно сделать, открыв Jobs » New Job... и добавив этап моделирования MLP Generator в пустую блок-схему.



Настройте генератор машинно-обучаемого потенциала, дважды щелкнув на блок MLP Generator. Будут предложены различные варианты.

Первый шаг состоит в выборе обучающего набора **Training set**, который будет использоваться для создания машинно-обучаемого потенциала. Используйте кнопку **Выбрать**, чтобы открыть диалоговое окно выбора файла.



Флажок **Δ-learning** может использоваться для запроса установки или использования Δ -потенциала. Если флажок установлен, то кнопки **Fit potential** и **Include potential** становятся активными. Кнопку **Fit potential** можно выбрать, если заданный выше обучающий набор содержит данные для подгонки Δ -потенциала. В этом случае будет создан Δ -потенциал. Уже существующий Δ -потенциал можно использовать в подгонке, выбрав кнопку **Include potential**. Поле ввода **Δ-learning potential** станет активным и потребует выбора файла силового поля с Δ -потенциалом.

Затем пользователь может выбрать тип машинно-обучаемого потенциала с помощью **Type of machine-learning potential**. В настоящее время MedeA[®] поддерживает потенциалы SNAP и NNP. Для каждого из этих двух вариантов осуществляется настройка собственных параметров.

3.2.1. Потенциал спектрального анализа соседей

Когда выбран машинно-обучаемый потенциал SNAP, становятся доступными две вкладки: **Parameters for SNAP** и **Advanced**.

1. Вкладка **Parameters for SNAP**.

| Parameters for SNAP | | Advanced | | |
|---------------------|--------------------------------------------|--------------------------------------------|--------------------------------------------|--|
| Band limit: | 8 | | | |
| Radial cutoff: | 7.01456 | | | |
| Element | Relative radius | Weight | Energy shift | |
| Ti | 0.5 | 1.0 | 0.0 | |
| Fit: | <input checked="" type="checkbox"/> Energy | <input checked="" type="checkbox"/> Forces | <input checked="" type="checkbox"/> Stress | |
| Weights: | 1.0 | 0.01 | 1.0e-06 | |

На этой вкладке устанавливаются следующие параметры:

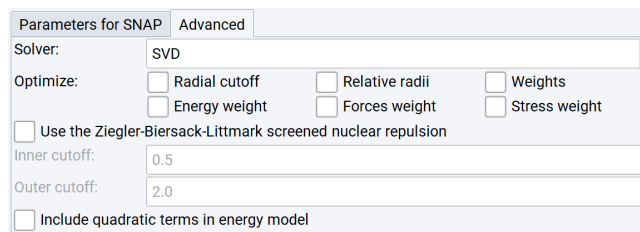
Band limit относится к максимальному угловому моменту, используемому в расширении сферических гармоник локальных атомных дескрипторов. Допускаются только четные значения.

Radial cutoff — это коэффициент, объединенный с полем ввода для **Relative radius** и умножающий сумму двух конкретных относительных радиусов атомов разного типа для определения радиуса отсечки, используемого для конкретного атома в локальном атомном кластере другого атома. Следовательно, радиус отсечки, применяемый для конкретного локального атомного кластера, зависит от типов как центрального атома, так и любого другого атома в этом кластере.

Для каждого элемента в обучающем наборе можно указать **Relative radius**, **Weight** и **Energy shift**. В то время как **Weight** позволяет различать разные типы атомов в расширении плотности соседних атомов в локальном атомном окружении, **Relative radius** определяет вклад атома в **Radial cutoff**, который зависит от типа атома и получается путем сложения относительных радиусов центрального атома локального атомного окружения и конкретного атома окружения с последующим умножением на **Radial cutoff**. Наконец, **Energy shift** определяет энергетический сдвиг, применяемый к определенному типу атома до фактической процедуры подгонки.

Пользователь может поставить галочки на **Energy**, **Forces** и **Stress** для учета при подгонке. При включении их веса могут быть определены. Хотя установка определенного веса равным нулю будет иметь тот же эффект, что и снятие флажка с соответствующего места, использование флажка более эффективно и надежно, поскольку уменьшает размер базовой числовой задачи.

2. Вкладка **Advanced**.



На вкладке **Advanced** можно настроить следующие параметры, требующие более детальных знаний о настройке SNAP:

Поле со списком **Solver** позволяет пользователю выбрать алгоритм для решения линейных уравнений. Возможные варианты: **SVD**, **LASSO**, **RIDGE**, и **ELASTIC**.

Шесть флажков с метками **Radial cutoff**, **Relative radius**, **Weights**, **Energy weight**, **Forces weight** и **Stress weight** могут быть установлены для выполнения оптимизации соответствующих параметров вместо их определения вручную. Поскольку оптимизация этих параметров требует большего количества подгонок (порядка 100), время выполнения при установке будет значительно увеличено. Тем не менее, настоятельно рекомендуется оптимизировать первые три из этих так называемых гиперпараметров.

При использовании **Use the Ziegler-Biersack-Littmark screened nuclear repulsion** потенциал Циглера-Бирсака-Литтмарка может быть добавлен к машинно-обучаемому потенциалу. Внутреннюю и внешнюю обрезку потенциала Циглера-Бирсака-Литтмарка можно контролировать с помощью переменных **Inner cutoff** и **Outer cutoff**.

Флажок **Include quadratic terms in energy model** можно использовать для расширения SNAP квадратичными членами.

3.2.2. Потенциал нейронной сети

При выборе потенциала нейронной сети (NNP) становятся доступными три вкладки: **Parameters for NNP**, **Advanced** и **Symmetry function**:

1. **Parameters for NNP**. На этой вкладке контролируются основные параметры подгонки NNP. Можно установить следующие переменные:

Number of epochs определяет общее число эпох. Это общее количество просмотров всей обучающей выборки.

Cutoff for symmetry functions определяет радиус отсечки набора функций симметрии по умолчанию. Эта переменная отбрасывается, когда определяется пользовательский набор функций симметрии на вкладке **Symmetry function**.

Fraction of data to use as test set устанавливает долю структур, которые будут случайным образом перенесены из обучающего набора в проверочный набор.

Когда флажок **Manual seed for random numbers** установлен, начальное значение для генератора случайных чисел определяется пользователем, в противном случае в качестве начального значения используется случайное значение.

Флажок **Fit Forces** определяет, будут ли учитываться силы при подгонке NNP в дополнение к энергиям. Вес этих сил может быть установлен в соответствующем поле ввода **Weight**.

2. На вкладке **Advanced** можно настроить следующие параметры, требующие более детальных знаний о настройке NNP:

| Parameters for NNP | Advanced | Symmetry functions |
|-----------------------------------------|----------|--------------------|
| Number of nodes per hidden layer: | | 20 |
| Fraction of energy to update: | | 1.0 |
| Fraction of force to update: | | 0.03 |
| Number of bins: | | 200 |
| Prune threshold for symmetry functions: | | 0.01 |

Number of nodes per hidden layer определяет количество узлов в обоих скрытых слоях нейронной сети.

Поля **Fraction of energy to update** и **Fraction of force to update** определяют долю энергии и принудительных обновлений в эпоху.

Number of bins определяет количество интервалов для гистограмм функций симметрии.

Prune threshold for symmetry functions устанавливает порог диапазона обрезки. Все функции симметрии с диапазоном активации ниже этого значения будут удалены.

3. **Symmetry function.** На этой вкладке можно настроить функции симметрии. Любые определенные пользователем функции симметрии переопределяют настройку радиуса отсечки в разделе **Parameters for NNP**. В противном случае используется текущий набор дескрипторов по умолчанию – 10 смещенных функций радиальной симметрии на сетке, определяемых методом, предложенным Имбальдзано и соавторами [56], т.е. $n_r = 10$; центрированные узкие и широкие угловые функции, сгенерированные для $\lambda = \{-1, +1\}$, $\zeta = \{1, 3, 12\}$ на сетке, предложенной Гастеггером и соавторами [57], $n_\alpha = 4$, $r_{low} = 1$.

| Parameters for NNP | Advanced | Symmetry functions |
|--------------------------------------|----------------------------------------------------------|------------------------------|
| <input type="checkbox"/> Use default | <input type="checkbox"/> Use weighted symmetry functions | |
| Symmetry functions | Mode | R low R cutoff N points Zeta |
| Grid radial | Shift | 1.0 6.0 10 1 3 12 |
| Grid narrow angular | Center | 1.0 6.0 4 1 3 12 |
| Grid wide angular | Center | 1.0 6.0 4 1 3 12 |

Для большинства пользователей достаточно оставить набор функций симметрии по умолчанию с флажком **Use default**. Если этот флажок снят, набор функций симметрии становится определяемым пользователем. Переключайте **Use weighted symmetry functions** для выбора между взвешенными и невзвешенными функциями симметрии [57].

Наборы функций симметрии можно добавлять и удалять с помощью кнопок **Add** и **Delete**. Первая колонка в таблице **Symmetry function** позволяет задать определение функций симметрии как **Grid radial** для определения радиальных функций на сетке [56], или как **Grid narrow angular** для узких угловых функций на

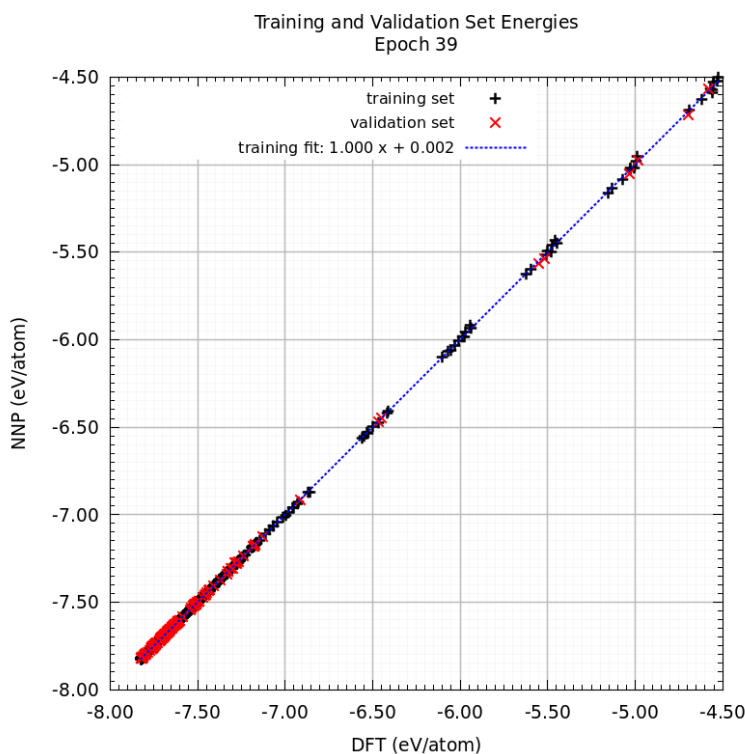
сетке, или как **Grid wide angular** для функций широкой угловой симметрии на сетке. Оба определения угловой сетки основаны на работе Гастеггера [57]. Вторая колонка **Mode** позволяет выбрать centrosymmetric или смещенную функцию симметрии. В колонке **R low** устанавливается нижняя граница, используемая для определения функций угловой симметрии на сетке, в колонке **R cutoff** указывается радиус отсечки для всех функций симметрии (это перезаписывает параметр отсечки, указанный в разделе **NNP**). Колонка **N points** настраивает количество узлов сетки, на которых задан данный набор функций симметрии. Последняя колонка **Zeta** определяет показатели степени функций угловой симметрии, которые могут быть указаны в виде списка, разделенного пробелами. Наименьшее значение показателя должно быть равно или больше единицы, и этот столбец активен только для функций угловой симметрии.

3.2.3. Оценка качества машинно-обученного потенциала, созданного с помощью программного модуля MLPG

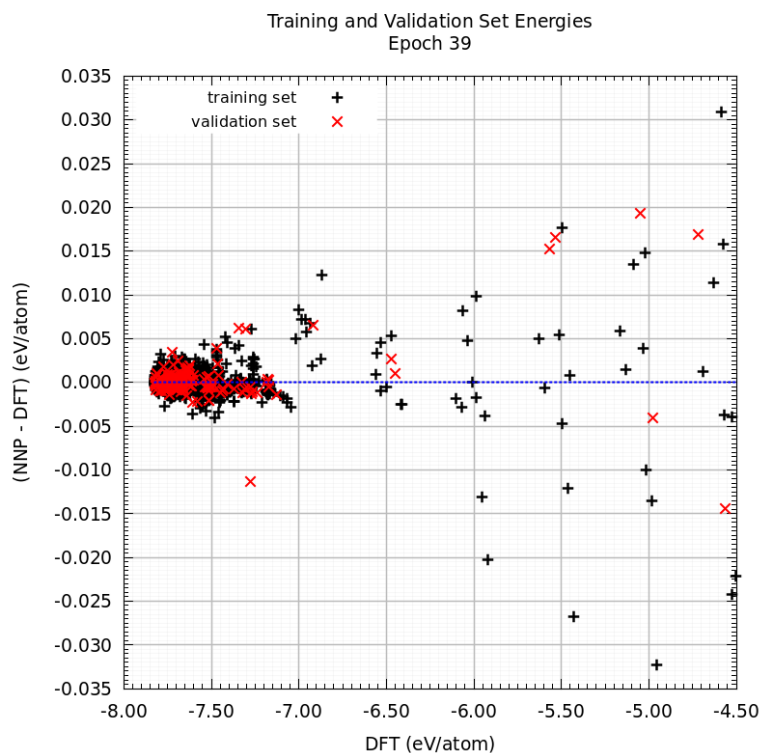
Файл Job.out генератора машинно-обучаемого потенциала предоставляет информацию о качестве произведенной подгонки. В нем представлен отчет об отклонениях свойств, полученных с помощью MLP, от свойств, полученных из предыдущих расчетов из первых принципов, с данными для энергий, сил и напряжений. Детали, представленные в файле Job.out, зависят от использованного способа генерации машинно-обученного потенциала (SNAP или NNP).

В дополнение к статистической информации, содержащейся в Job.out, доступен графический вывод в следующих файлах:

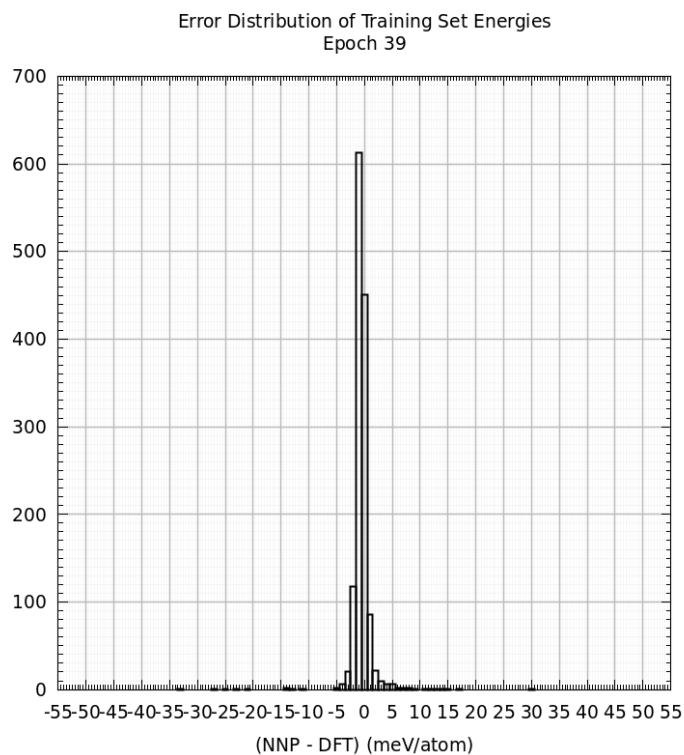
- Energies_DFT_vs_MLP.png показывает корреляцию между энергиями, рассчитанными с помощью ТФП (DFT), и энергиями, предсказанными MLP, как для обучающего (черные крестики), так и для проверочного набора (красные крестики). Обратите внимание, что данные для проверки отображаются только при создании потенциала нейронной сети (NNP):



- Energies_Difference_DFT_vs_MLP.png показывает отклонение предсказанных энергий машинно-обученным потенциалом от соответствующих значений ТФП (DFT):

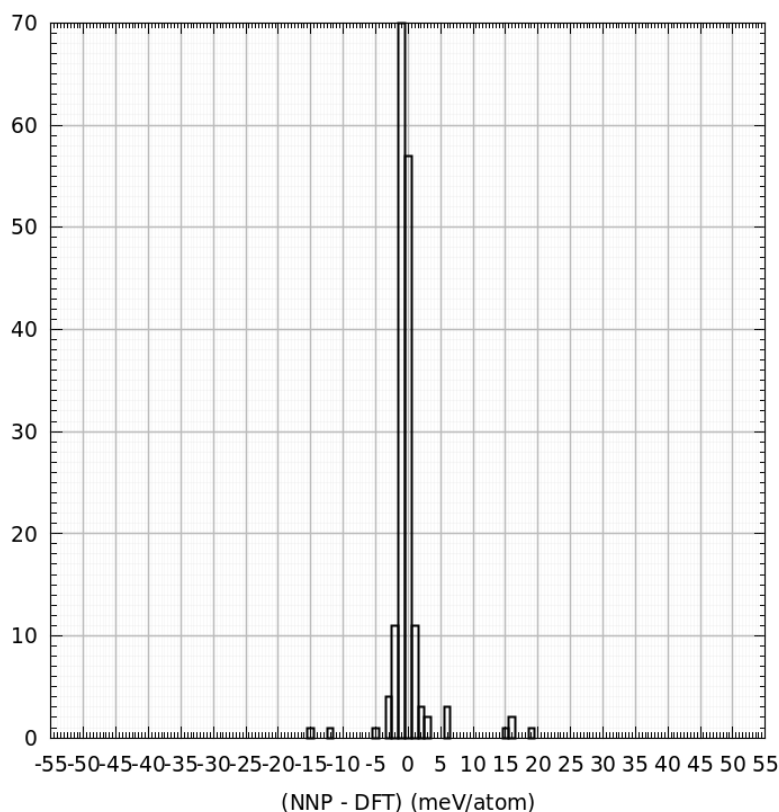


- Energies_Error_Distribution_TrainingSet.png содержит гистограмму ошибок прогнозируемых энергий машинно-обученного потенциала для обучающего набора:

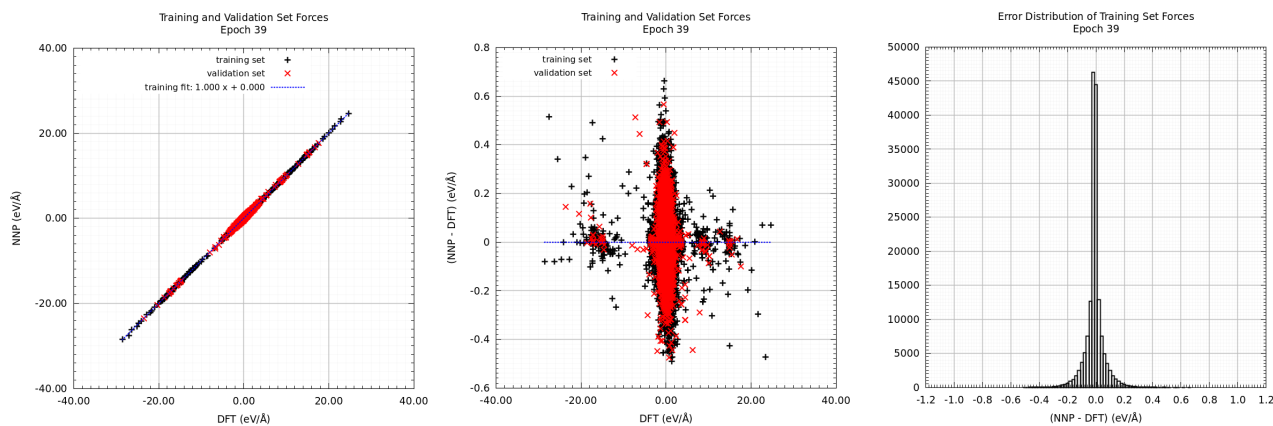


- Energies_Error_Distribution_ValidationSet.png содержит гистограмму ошибок прогнозируемых энергий машинно-обученного потенциала для проверочного набора:

Error Distribution of Validation Set Energies
Epoch 39

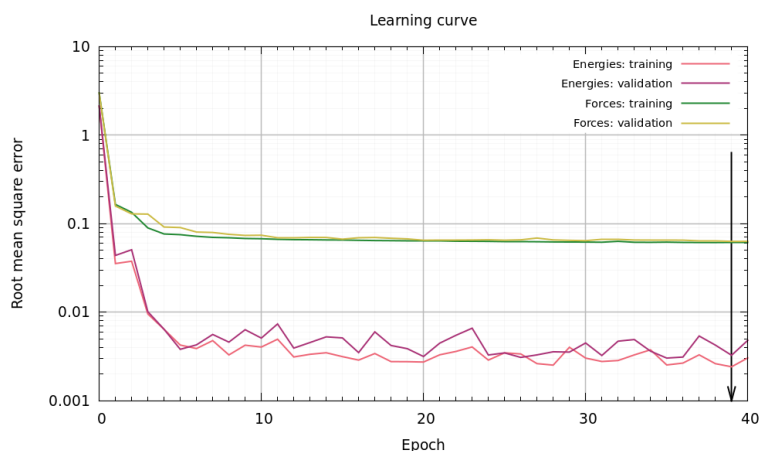


- Аналогичные зависимости в виде графиков доступны для сил:



Некоторые из перечисленных выше графиков могут отсутствовать, так как создаются в зависимости от конфигурации этапа **MLPG**. На графиках свойств, рассчитанных с помощью машинно-обученного потенциала, результаты показаны как функции свойств, рассчитанных с помощью ТФП. В идеале на графике должна быть прямая линия с наклоном 45° . Корреляция расчетов ТФП и машинно-обученного потенциала также отображена на графике в виде уравнения. Идеальное соответствие в этом уравнении дает наклон, равный единице, и прохождению через ноль. Графики разности ТФП и машинно-обученного потенциала показывают отклонение свойства, рассчитанного с помощью машинно-обученного потенциала, от соответствующих значений ТФП как функцию свойства ТФП. В идеальном случае эти отклонения равны нулю. Графики распределения ошибок содержат гистограммы ошибок. Эти распределения должны быть как можно более узкими и центрироваться в нуле.

При работе с нейронной сетью также генерируется график **Learning-curve.png**, который показывает среднеквадратичную ошибку как функцию эпохи для энергий и сил в обучающих и тестовых наборах. Оптимальная эпоха указана на графике ниже черной стрелкой.



3.2.4. Использование машинно-обученного потенциала, созданного с помощью программного модуля MLPG

Если генератор машинно-обучаемого потенциала (MLPG) достиг удовлетворительного согласия с обучающим набором, рассчитанным из первых принципов, его можно использовать в дальнейших расчетах с помощью MedeA LAMMPS. При каждом запуске генератора машинно-обучаемого потенциала (MLPG) создается файл с расширением .frc с параметрами машинно-обученного потенциала. Этот файл можно найти в папке соответствующего этапа блок-схемы. Его точное местоположение указано в файле Job.out. Сохраните этот файл на своем компьютере в *MD/data/Forcefields/custom* и прочитайте его с помощью *Forcefields -> Read...* При выполнении вычислений на внешнем сервере убедитесь, что этот файл силового поля присутствует на этом компьютере в том же месте, что и на вашем сервере.

Контрольные вопросы и дополнительные задания:

1. Какие свойства обучающего набора, рассчитанные из первых принципов, используются для генерации машинно-обученного потенциала?
2. Какие конфигурации системы могут быть включены в обучающий набор? Как конфигурации, включенные в обучающий набор, определяют применимость сгенерированного машинно-обученного потенциала?
3. Каковы требования к симметрии системы и параметрам расчетов ТФП для обучающей выборки?
4. Основные параметры настройки потенциалов нейронной сети и спектрального анализа соседей.
5. Как производится оценка качества сгенерированных машинно-обученных потенциалов?

4. Практические задания

4.1. Использование готового машинно-обученного потенциала на примере предсказания свойств никеля (Ni) в коде LAMMPS

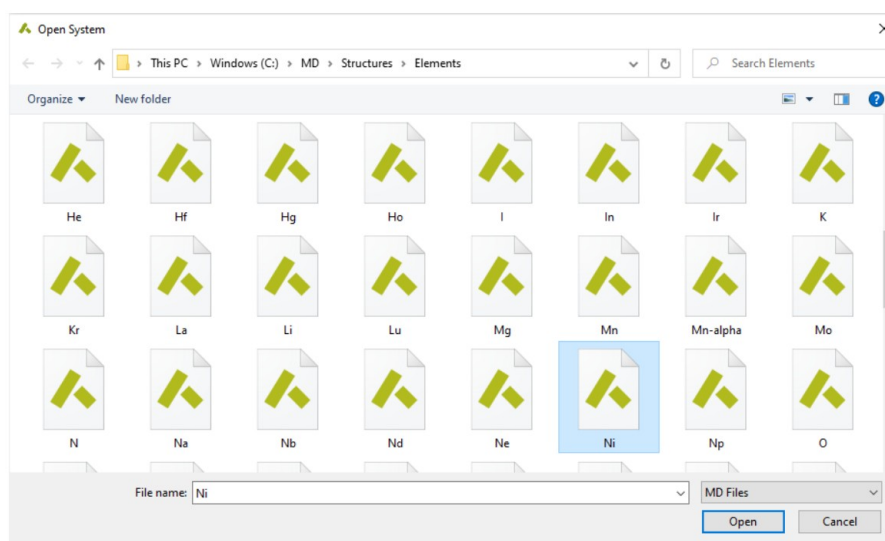
Цель: Вычислить свойства металлической системы Ni, используя готовый машинно-обученный потенциал из библиотеки.

Задачи:

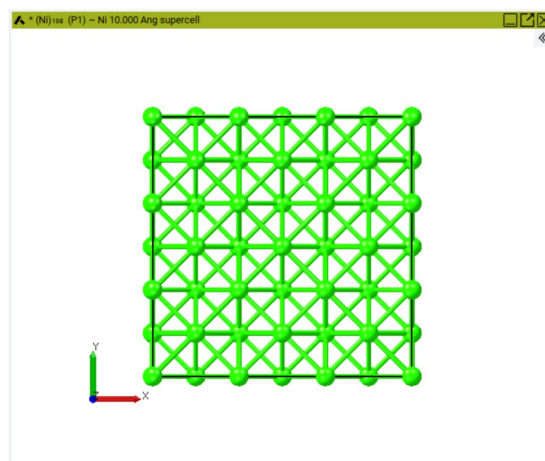
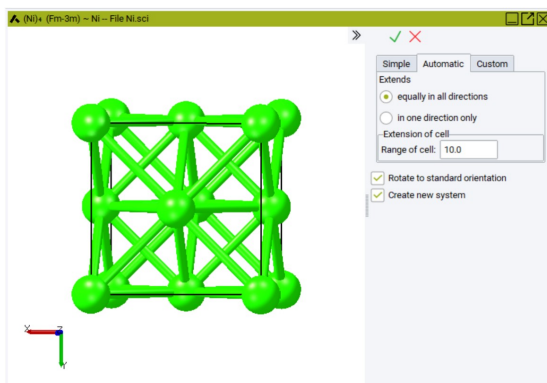
1. Построить суперъячейку Ni с помощью **MedeA Supercell Builder**.
2. Использовать машинно-обученный потенциал для выполнения моделирования молекулярной динамики в NPT ансамбле с помощью программного модуля MedeA LAMMPS.
3. Найти равновесную плотность при заданной температуре.
4. Сравнить полученные результаты с экспериментальными данными, сделать выводы об эффективности машинно-обученного потенциала для предсказания свойств металлов на примере никеля.

Порядок выполнения работы:

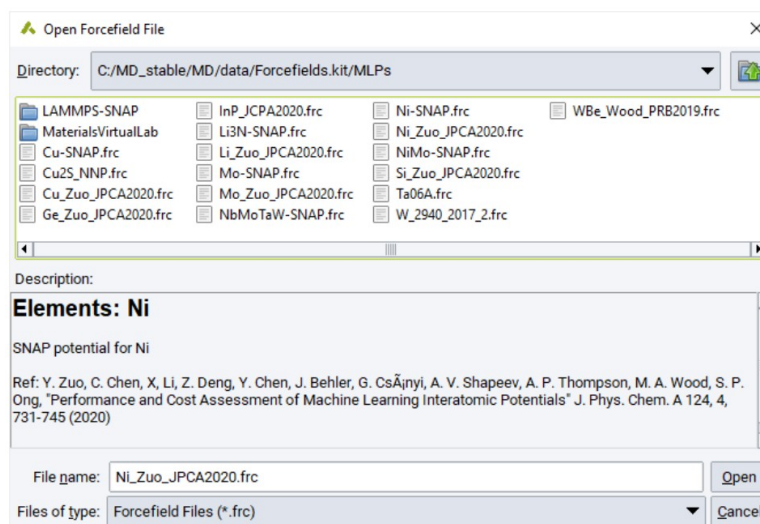
1. Постройте гранецентрированную кубическую (ГЦК) суперъячейку Ni, размером около $10\text{\AA} \times 10\text{\AA} \times 10\text{\AA}$. Откройте File » Open structure from disk, появится диалоговое окно Open System. Перейдите к папке MD/Structures/Elements и выберите Ni.sci. Нажмите Open.



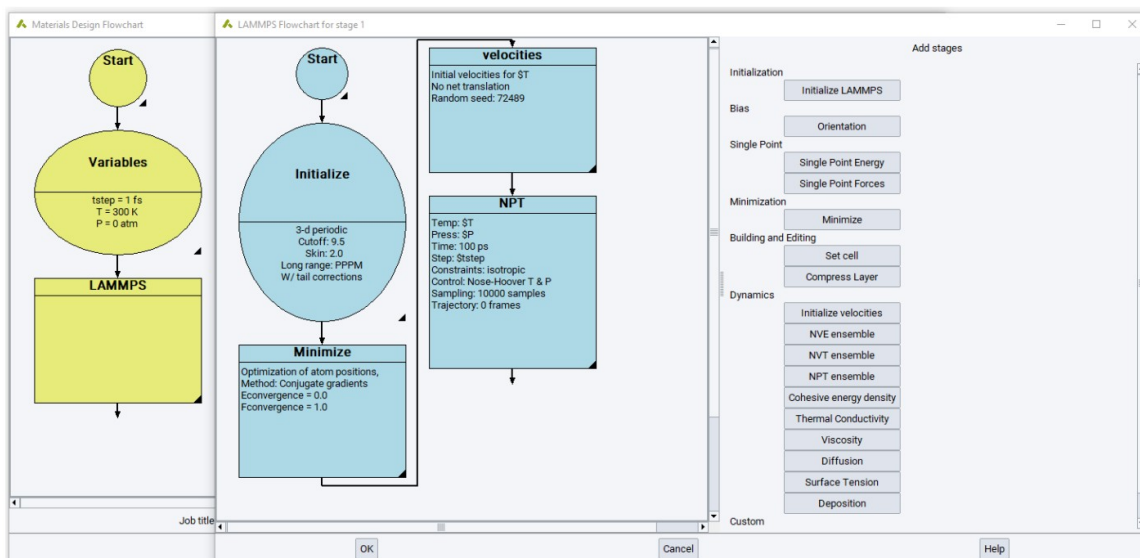
Откройте боковую панель Supercell Builder с помощью Builders » Supercell Builder и перейдите на вкладку Automatic, сохранив настройки по умолчанию и проследив, чтобы диапазон ячейки был 10\AA .



2. Активируйте машинно-обученный потенциал, нажав Forcefields » Read... и перейдя к MD/data/Forcefields.kit/MLPs. Выберите файл Ni Zuo JPCA2020.frc, содержащий параметры MLP для Ni, и откройте его.



3. В последующих симуляциях используйте это силовое поле (машинно-обученный потенциал в данном случае). После активации силового поля в виде машинно-обученного потенциала его можно использовать для расчета широкого спектра свойств, аналогично использованию любых других силовых полей. Напомним, что данный потенциал будет работать только в рамках тренировочного набора, использованного для его генерации. Для примера предлагается провести простое моделирование методом молекулярной динамики (MD) в NPT ансамбле при температуре и давлении окружающей среды.
4. Составьте алгоритм (блок-схему с параметрами) моделирования с помощью интерфейса Medea flowchart. Откройте Jobs » New Job... для активации интерфейса блок-схемы. В правом верхнем разделе, Reading and writing flowcharts, нажмите кнопку Open Library... Перейдите MD/data/ Flowcharts.kit/LAMMPS и откройте Simple molecular dynamics.flow.



5. Загрузится простая блок-схема для расчетов методом молекулярной динамики, которая выполняет моделирование в NPT ансамбле при температуре T и давлении P , заданным в блоке Variables. Временной шаг моделирования определяется с помощью переменной $tstep$. Дважды щелкните на блок Variables, если хотите его изменить. Протокол моделирования LAMMPS определяется в блоке LAMMPS. Также можно дважды щелкнуть на этот блок для его изменения.
6. Дважды щелкните на блок Variables и измените давление на 1 atm . После ввода названия, например, Ni MLP tutorial в Job title, нажмите Run, чтобы отправить задачу на сервер для выполнения.
7. После отправки расчета на сервер временные файлы ввода/вывода можно проверить на JobServer, где также хранятся результаты моделирования.
8. Откройте файл Job.out на сервере. Файл содержит две части, первая из которых представляет собой сводку этапов и параметров расчета.

```

#-----
#           This calculation has 2 stages
#-----

Stage 1: Variables
  tstep = 1 fs
  T = 300 K
  P = 1 atm

Stage 2: This LAMMPS calculation has 4 stages:

  Stage 2.1: Initialize LAMMPS for a SNAP type of forcefield 'Ni_Zuo_JPCA2020' (from Forcefields.kit::MLPs/
  ↔Ni_Zuo_JPCA2020.frc)
    Current system formula: Ni108 (Ni)
    Use of an SNAP forcefield results in the specialized nonbond term handling in LAMMPS:
    Reported "van der Waals" energy (Evdw) includes all nonbond, non-Coulombic energy terms.

  Stage 2.2: Minimize using conjugate gradients and linesearch fast
    Relative energy convergence: 0.0
    Force convergence: 1.0
    Maximum atomic motion: 0.05
    Maximum number of iterations: 1000
    Maximum number of force evaluations: 10000

  Stage 2.3: Set the velocities for 300 K

  Stage 2.4: NPT integration for 100 ps with a timestep of 1 fs
    P is 1 atm with cell constrained isotropically
    T is 300 K
  
```

Можно проверить переменные, системную формулу, критерии минимизации и параметры NPT ансамбля. Когда расчет завершен, появляется вторая часть файла – Job.out, которая содержит результаты моделирования и плотность системы на этапе NPT:

```

Stage 2.4: NPT integration for 100 ps with a timestep of 1 fs
P is 1 atm with cell constrained isotropically
T is 300 K

```

| Property | Value | +/- Uncertainty | Units | After Steps | % Run |
|----------|----------|-----------------|--------|-------------|-------|
| t: | 100000.0 | | fs | | |
| T: | 300.6 | +/- 1.9 | K | 0 | 0.0% |
| P: | -2.1 | +/- 7.5 | atm | 0 | 0.0% |
| V: | 1190.601 | +/- 0.09 | Ang^3 | 0 | 0.0% |
| rho: | 8.84103 | +/- 0.00067 | g/mL | 0 | 0.0% |
| Etotal: | -59428.9 | +/- 5 | kJ/mol | 0 | 0.0% |
| a: | 10.59875 | +/- 0.00027 | Ang | 0 | 0.0% |
| b: | 10.59875 | +/- 0.00027 | Ang | 0 | 0.0% |
| c: | 10.59875 | +/- 0.00027 | Ang | 0 | 0.0% |
| alpha: | 90.00022 | +/- 0 | degree | 0 | 0.0% |
| beta: | 90.00022 | +/- 0 | degree | 0 | 0.0% |
| gamma: | 90.00022 | +/- 0 | degree | 0 | 0.0% |
| Epot: | -59830 | +/- 2.5 | kJ/mol | 0 | 0.0% |
| Ekin: | 401.1 | +/- 2.5 | kJ/mol | 0 | 0.0% |
| Evdw: | -59830 | +/- 2.5 | kJ/mol | 0 | 0.0% |
| Ecoult: | 0 | +/- 0 | kJ/mol | 0 | 0.0% |
| Sxx: | -380 | +/- 290 | atm | 0 | 0.0% |
| Syy: | 360 | +/- 230 | atm | 0 | 0.0% |
| Szz: | -60 | +/- 290 | atm | 1000 | 10.0% |
| Syz: | -200 | +/- 500 | atm | 0 | 0.0% |
| Sxz: | -10 | +/- 130 | atm | 0 | 0.0% |
| Sxy: | -260 | +/- 220 | atm | 0 | 0.0% |

```

LAMMPS stage successfully completed on 2 core(s) on Wed 07 July 2021 at 19:11:08 CEST after 3902 s (1:05:02)

```

Температура T составляет 300.6 ± 1.9 K.

Давление $P = -2.1 \pm 7.5$ atm. Степень флуктуации типична для такой небольшой системы. Плотность $\rho = 8.84103 \pm 0.00067$ g/mL, что согласуется с экспериментальным значением 8.902 g/mL.

Размеры ячейки a , b и c уравновешены плотностью.

Следует напомнить, что когда моделирование происходит за пределами режима, охватываемого обучающими набором, машинно-обученные потенциалы работают намного хуже, чем классические силовые поля. Важно убедиться, что все необходимые конфигурации, к которым применяется машинно-обученный потенциал, были включены во время обучения. Другими словами, если Вы планируете моделировать систему в каких-то экстремальных условиях, то в обучающей выборке должны быть соответствующие конфигурации. Если это не так, следует создать новый машинно-обученный потенциал с измененным обучающим набором с использованием программного модуля MedeA MLPG.

Контрольные вопросы и дополнительные задания:

- Для никеля (Ni) проведите расчеты с другими имеющимися в распоряжении потенциалами, например EAM, MEAM, PCFF+.

 - Сравните точность и время расчетов с помощью потенциала машинного обучения по сравнению с другими потенциалами, экспериментальными данными (из литературы) и ТФП (произведя расчет для элементарной ячейки).
 - Вычислите с помощью различных потенциалов механические свойства (модуль упругости, модуль Юнга и другие на Ваше усмотрение), постоянные решетки и плотность. Сделайте таблицу, в которой будет произведено сравнение полученных данных с экспериментом и ТФП. Вычислите относительные погрешности.
 - Сделайте выводы о точности и времени расчетов.
 - Постройте необходимые диаграммы для визуального сравнения использованных потенциалов.

- д) **Дополнительно:** Выполните расчеты при различных температурах. Подумайте, как можно найти точку плавления. Выполните необходимые вычисления и сравните с экспериментальными данными.
2. Для меди (Cu) выполните моделирование методом молекулярной динамики в NPT ансамбле с готовым машиннообученным потенциалом, сгенерированным в подходе SNAP. Файл Cu-SNAP.frc находится в следующей директории:
MD/data/Forcefields.kit/MLPs.
- а) Выполните расчеты с помощью других силовых полей, имеющихся в директориях.
б) Сравните полученные данные с экспериментальными (из литературы) и ТФП (из расчета для одной элементарной ячейки).
3. **Дополнительно:** аналогичные расчеты выполните для платины (Pt), сравните результаты, полученные с помощью машиннообученного потенциала, с имеющимися в распоряжении потенциалами MEAM и EAM. Файлы силового поля находятся в следующей директории: MD/data/ Forcefields.kit/metallic. Вычислите упругие константы, параметры решетки и проведите NPT моделирование.

PS. Инструкцию, как делать расчеты механических свойств с помощью модуля LAMMPS, можно посмотреть на <https://www.youtube.com/watch?v=uzFnHASMS60> или по инструкции в папке: PCS-SERVER/SharePCS/0TutorialsMedeA/ElasticProp, где также есть инструкция по определению упругих свойств с помощью модуля VASP. Не забудьте, что при работе с первопринципными методами необходимо использовать минимально возможную ячейку для сокращения времени расчетов.

В результате выполнения Упражнения должна получиться подобная таблица для каждого из исследованных металлов:

| | | | | | | | | | |
|----|--------------------------------------|------------|--------|------------|-----------|----------------|------------------------|--------|---------|
| 1 | Ni | | | | | | | | |
| 2 | Оптимизация | | | | | lattice params | | | |
| 3 | | job number | server | timing,sec | a | b | c | Volume | density |
| 4 | Ni MLP | | | | | | | | |
| 5 | Ni MEAM | | | | | | | | |
| 6 | Ni EAM | | | | | | | | |
| 7 | Ni PCFF+ | | | | | | | | |
| 8 | Ni DFT | | | | | | | | |
| 9 | | | | | | | | | |
| 10 | | | | | | | | | |
| 11 | | | | | | | | | |
| 12 | | | | | | | | | |
| 13 | Механические свойства Ni | | Bulk | Shear, Gpa | Young,Gpa | Melting temp | и другие на усмотрение | | |
| 14 | ML | | | | | | | | |
| 15 | MEAM | | | | | | | | |
| 16 | EAM | | | | | | | | |
| 17 | DFT | | | | | | | | |
| 18 | experiment | | | | | | | | |
| 19 | (PS. Привести ссылку на эксперимент) | | | | | | | | |
| 20 | Построить диаграмму! | | | | | | | | |

4.2. Создание обучающего набора и генерация машинно-обученного потенциала для титана (Ti). Исследование фазовых переходов в титане

Фазовая стабильность элементарных металлов, металлических соединений и сплавов представляет безусловно большой интерес. Знание стабильных фаз в условиях окружающей среды и их превращений в новые фазы, например, при изменении температуры и/или давления жизненно важно во многих приложениях. Конкретные составы и обработка также могут иметь большое значение. Например, INVAR, – сплав, состоящий из 64 % Fe и 36 % Ni, обладает почти нулевым тепловым расширением, и понимание физической основы такого явления открывает большие возможности для повышения эффективности, инноваций и улучшения технологических процессов. Знание свойств металлов и сплавов значительно улучшилось благодаря использованию методов компьютерного моделирования, например, на основе теории функционала плотности. Такие методы объясняют, например, фазовую стабильность переходных металлов и существование хорошо известных гексагональной, объемно-центрированной, кубической, шестигуговой плотно упакованной фазы. В последнее время интерес обратился к сложным металлическим системам, таким как высокоэнтропийные сплавы, требующим более масштабных системных расчетов.

Несмотря на невероятный рост вычислительной мощности, наблюдаемый в последние десятилетия, вычисления ТФП больших систем по-прежнему требуют больших вычислительных ресурсов. В частности, моделирование *ab initio* молекулярной динамики для больших систем, содержащих 100000 или более атомов, при длительном времени моделирования остается недоступным. Как следствие, явления, происходящие при повышенных температурах, такие как температурно-индуцированные фазовые переходы, не могут быть хорошо изучены. Это серьезно снижает применимость расчетов *ab initio* именно тогда, когда необходимы точные результаты, ограничивая их актуальность для промышленных приложений.

В течение многих лет потребность в крупномасштабном моделировании молекулярной динамики удовлетворялась за счет использования классических силовых полей. Специально для металлов потенциал Embedded Atom Model (EAM) Zhou et al. оказался очень полезным в большом количестве приложений. Однако, будучи вдохновленными физическими моделями с точки зрения связей и атомных плотностей, применимость классических силовых полей, например, для изучения фазовых переходов и химических реакций, ограничена. Кроме того, определение параметров силового поля для новых материалов требует большого опыта и тщательной подгонки.

В последнее десятилетие ограничения классических силовых полей инициировали разработку машинно-обученных потенциалов. Они используют набор общих параметризуемых математических функций для описания зависимости энергии, сил и напряжения от атомной структуры, о чем говорилось в первых главах учебного пособия. Эти параметры определяются из расчетов ТФП, выполненных для набора структур (обучающий набор) с использованием методов машинного обучения. На сегодняшний день используется несколько подходов, включая методы нейронной сети и регрессии.

Цель: Изучить фазовый переход между альфа и бета фазой титана.

Задачи:

1. Составить обучающий набор и выполнить расчеты ТФП для каждой конфигурации.
2. Собрать результаты расчетов ТФП в структурный лист обучающего набора.

3. Использовать полученные *ab-initio* MD траектории для обучения машинно-обученного потенциала.
4. Выполнить расчеты методом классической молекулярной динамики с помощью полученного машинно-обученного потенциала для апробации.

Порядок выполнения работы:

1. **Составление тренировочного набора** – первый этап для создания потенциала. Нас будет интересовать фазовый переход между альфа и бета фазой Ti. Для описания этого перехода необходимо в тренировочный набор включить: α, β и ω (шестиугольная) фазы титана и ГЦК. **Для каждой из 4 фаз необходимо выполнить структурную оптимизацию методом ТФП.** Параметры расчета должны быть одинаковыми для всех, например, 300 eV или 350 eV, 0.2 1/Angstrom, accurate, real-space projection, Gaussian smearing with 0.05 eV, 10^{-6} SCF convergence, PBE functional.

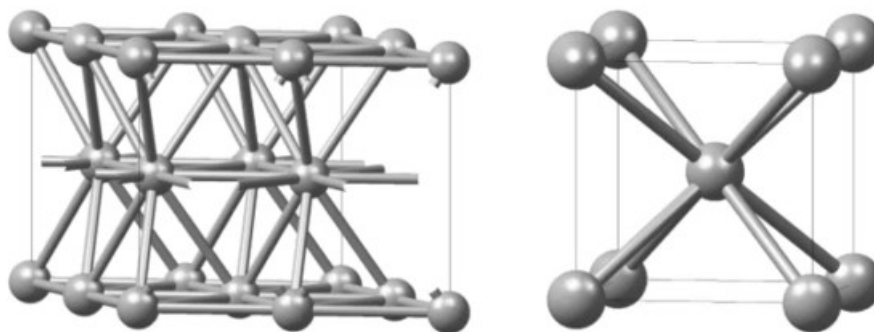


Рис. 7. Суперъячейки $3 \times 3 \times 2$ α -фазы ГПУ и $3 \times 3 \times 3$ β -фаза ОЦК.

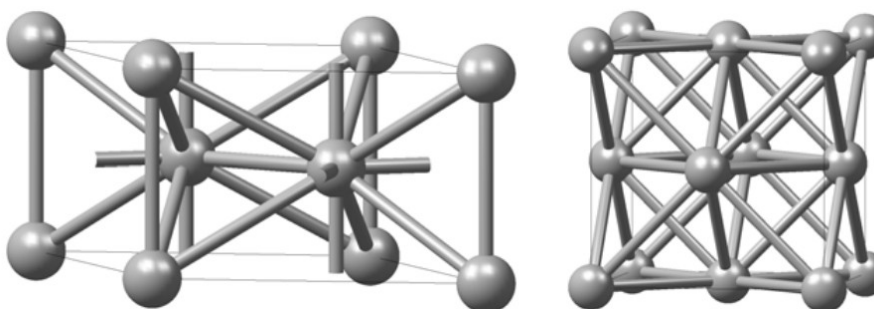


Рис. 8. Суперъячейки $2 \times 2 \times 3$ ω -фазы ГП и $2 \times 2 \times 2$ ГЦК.

- Для каждой фазы создайте конфигурации (структуры), соответствующие разным структурным деформациям, а именно: изотропные (12) и одноосные (12*3) деформации с напряжением $\pm 0.5\%$, $\pm 1\%$, $\pm 1.5\%$, $\pm 2\%$, $\pm 4\%$ и $\pm 6\%$, а также угловые деформации (3 угла * 3) 0.5° , 1.5° и 2° . Получится множество структур, которые будут составлять обучающий набор (в нашем случае $4 \cdot (12 + 12 \cdot 3 + 9) = 228$ для 4х фаз).
- Постройте суперъячейки, состоящие, например, из $3 \times 3 \times 3$ элементарных. Проверьте, сколько атомов получилось для каждой фазы. Должно быть 32-54 атома в зависимости от фазы.
- Для каждой из суперъячеек необходимо выполнить NVT моделирование методом молекулярной динамики для получения большего числа конфигураций. Параметры для расчета VASP-MD: 120 фсек с шагом 4 фсек в диапазоне 300-1700 К (это даст 30 конфигураций для каждой из фаз и для каждой деформации).

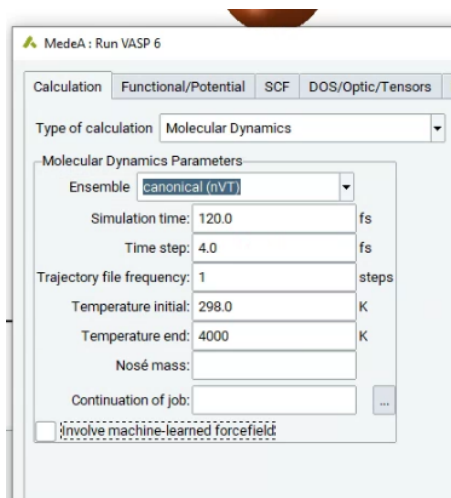
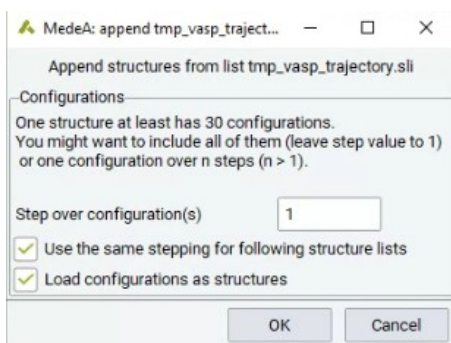


Рис. 9. Окно для расчета методом VASP-MD

- Сколько в итоге получится конфигураций? Соберите все конфигурации в один лист (structure list file). Для этого откройте: File -> Structure list Editor. В появившемся окне File -> Open Structure List from job и добавьте таким образом в структурный файл все ваши конфигурации из выполненных расчетов.

При добавлении нового списка каждый раз будет выходить окошко, в котором можно выбрать, какие из структур расчета включить в список. Также включите следующие параметры для гарантированного согласования обучающего набора:



В итоге, должен получиться файл, содержащий все созданные конфигурации. Этот один файл содержит всю необходимую для обучения информацию, например, энергии, силы на всех атомах и действующие напряжения.

MedeA: structure list editor

File Add structure(s) Display Properties QT: QSAR Toolbox

SQLite structure list file (5808128 bytes): /home/volker/MD/Structures/Demo/Ti/TiHs1-2021-10-10.sll
Containing 1005 structure(s)

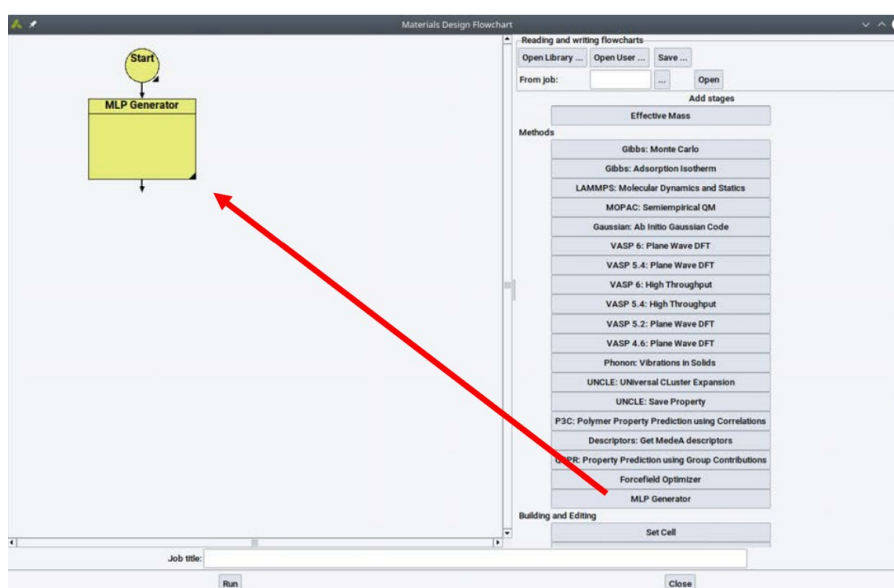
Display structures from: 1 to: 200 Apply

| Order | Name | Structural Formula | # atoms | # configurations | Symmetry | Cell parameters | Epot | |
|-------|-------------------------------|--------------------|---------|------------------|----------|-----------------------------------|------------|-------------|
| 1 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.30442 90 90 120 | -282.12713 | {{-0 -0}} |
| 2 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.67059 8.67059 9.16486 90 90 120 | -281.64894 | {{-0 -0}} |
| 3 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.67059 8.80263 9.30442 90 90 120 | -282.02039 | {{0.0238}} |
| 4 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.67059 9.30442 90 90 120 | -282.02039 | {{-0.0202}} |
| 5 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.16486 90 90 120 | -282.056 | {{0 0}} |
| 6 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.7146 8.7146 9.21138 90 90 120 | -281.91211 | {{-0 -0}} |
| 7 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.7146 8.80263 9.30442 90 90 120 | -282.07732 | {{0.0169}} |
| 8 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.7146 9.30442 90 90 120 | -282.07732 | {{-0.0121}} |
| 9 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.21138 90 90 120 | -282.09622 | {{-0 -0}} |
| 10 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.75862 8.75862 9.2579 90 90 120 | -282.07021 | {{0 -0}} |
| 11 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.75862 8.80263 9.30442 90 90 120 | -282.11285 | {{0.0072}} |
| 12 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.75862 9.30442 90 90 120 | -282.11285 | {{-0.0072}} |
| 13 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.2579 90 90 120 | -282.11898 | {{-0 -0}} |
| 14 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.84664 8.84664 9.35094 90 90 120 | -282.08559 | {{-0 -0}} |
| 15 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.84664 8.80263 9.30442 90 90 120 | -282.12049 | {{-0.0082}} |
| 16 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.84664 9.30442 90 90 120 | -282.12049 | {{0.0064}} |
| 17 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.35094 90 90 120 | -282.11801 | {{0 -0}} |
| 18 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.89066 8.89066 9.39747 90 90 120 | -281.95006 | {{-0 0}} |
| 19 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.89066 8.80263 9.30442 90 90 120 | -282.09365 | {{-0.0152}} |
| 20 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.89066 9.30442 90 90 120 | -282.09365 | {{0.0149}} |
| 21 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.39747 90 90 120 | -282.09074 | {{0 -0}} |
| 22 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.93467 8.93467 9.44399 90 90 120 | -281.72348 | {{-0 -0}} |
| 23 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.93467 8.80263 9.30442 90 90 120 | -282.04652 | {{-0.0221}} |
| 24 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.93467 9.30442 90 90 120 | -282.04652 | {{0.0227}} |
| 25 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.44399 90 90 120 | -282.0446 | {{-0 -0}} |
| 26 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.30442 89 90 120 | -282.09186 | {{-0 -9.2}} |
| 27 | Ti-P63mmc-CCMR75535-min_3x3x2 | Ti36 | 36 | 1 | P1 | 8.80263 8.80263 9.30442 90 89 120 | -282.09182 | {{-0.0002}} |

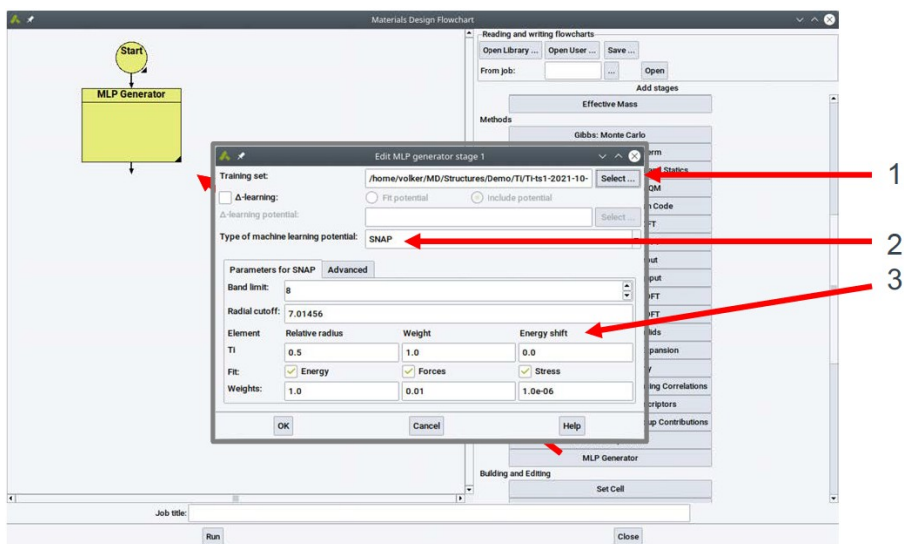
Close

Если по каким-то причинам у вас не получилось собрать список структур, то готовый скруктурный лист можно открыть из папки PCS-SERVER -> SharePCS -> 0TutorialsMedeA -> TutorialsCompDesign -> ML -> Tutorial-Generating-Ti-NNP с помощью File -> Fitting Data Manager.

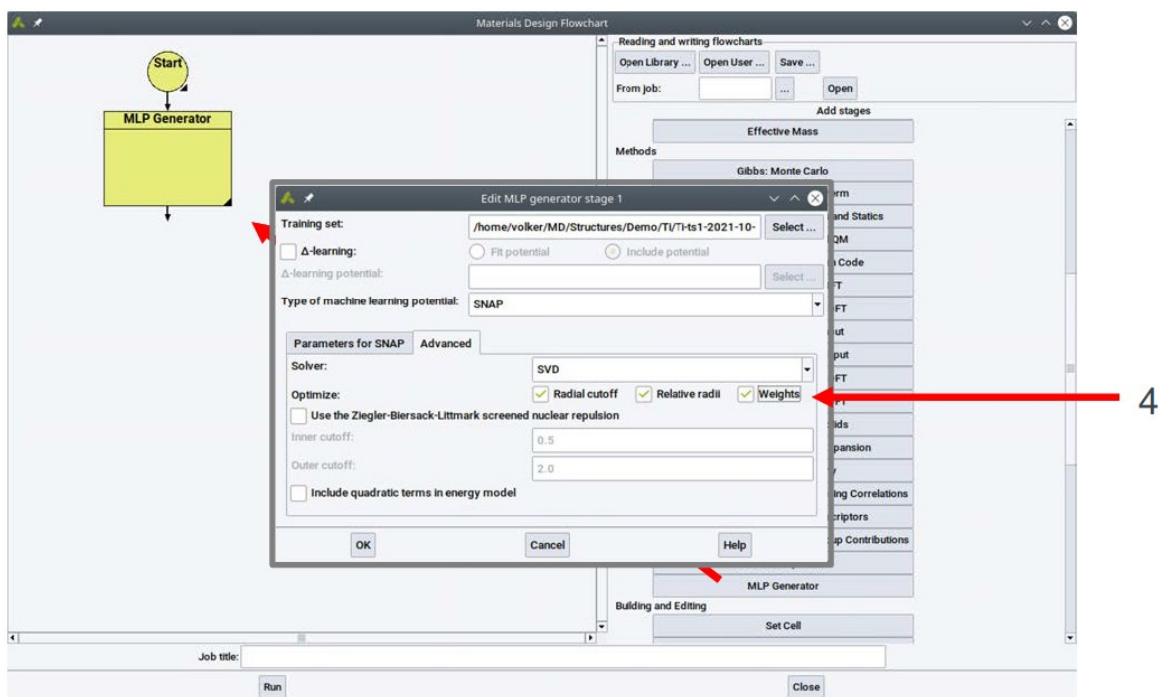
6. **Перейдите к созданию потенциала.** Откройте интерфейс MedeA flowchart: Jobs -> New Job... и добавьте MLP Generator в пустую блок-схему.



Откройте только что добавленный блок, появится окно, в котором можно задать параметры генерации потенциала:



1 – обучающий набор, 2 – тип потенциала, 3 – основные параметры. На данном этапе необходимо указать обучающий набор и тип потенциала. Используйте потенциал **SNAP**. 4 – гиперпараметры, которые могут быть указаны дополнительно. Можно выполнить точную настройку и указать дополнительные параметры, такие как **Relative radius**, **Weigh** и **Radial cutoff**. Устанавливаем флажок на всех.



7. Запустите расчет. После завершения откройте файл с результатами, в котором содержится информация о параметрах и ссылка на сгенерированный файл с расширением **frc**. Это и есть файл сгенерированного силового поля, который может быть использован для последующих расчетов.

Stage 1: MLP Generator

```

Input for FitSNAP is in "Ti-tsl-2021-10-10.in":
Band limit: 10
Radial cutoff: 7.01456 Ang
Elements:
+-----+
| Rel. radius  Weight  Energy shift |
+-----+
| Ti      1.0000  0.5000  0.000000 |
+-----+

Properties used in the fit: energy forces stress
with weights:           1.0 0.05 1.0e-05

The Ziegler-Biersack-Littmark screened nuclear repulsion with inner/outer radius of 0.5/2.0 Ang is being used.

The following parameters are optimized during the MLP generation run:
Parameter      Initial  Optimized
Radial cutoff  7.01456 6.02630
Relative radius for Ti  0.50000 0.59345
Weight for Ti   1.00000 1.00917
    
```

5
6
7

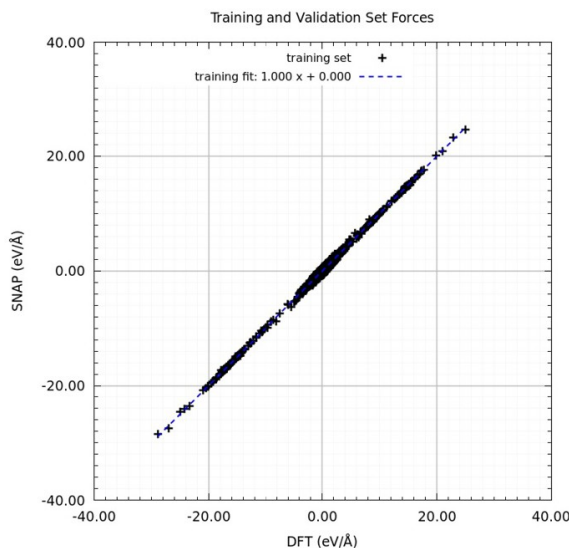
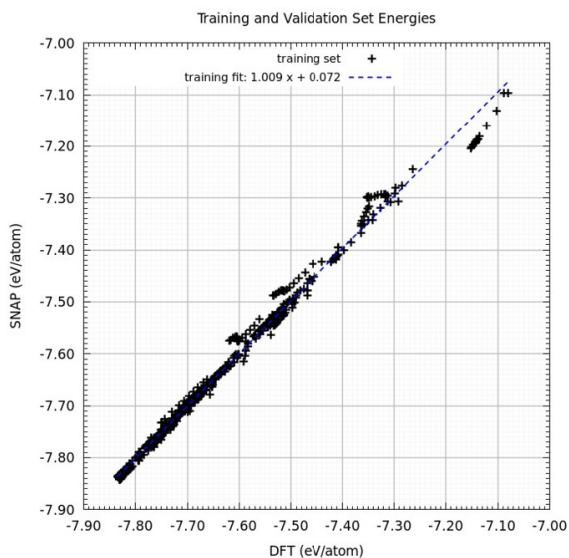
```

Fitting statistics
+-----+
| Group  Weighting  Subsystem  Count  Mean absolute error  Root mean absolute error  Root mean squared error  Relative root mean squared error  R squared |
+-----+
| all Unweighted Combined 127401 59.9335 0.0739 487.8104 0.0586 0.9966 |
| all Unweighted Energy 1005 0.0133 0.1452 0.0178 0.1341 0.9820 |
| all Unweighted Forces 120366 0.0583 0.3595 0.1091 0.2428 0.9411 |
| all Unweighted Stress 6030 1265.1000 0.0738 2242.2232 0.0586 0.9966 |
| all Weighted Combined 127401 0.0035 0.0453 0.0074 0.0108 0.9999 |
| all Weighted Energy 1005 0.0133 0.1452 0.0178 0.1341 0.9820 |
| all Weighted Forces 120366 0.0029 0.3595 0.0055 0.2428 0.9411 |
| all Weighted Stress 6030 0.0127 0.0738 0.0224 0.0586 0.9966 |
+-----+
Units: energy: eV/atom, force: eV/Angstrom, pressure: bar

Generated SNAP has been converted to "/data/medea/ContractJobs/dir97000/97338/Stage_1/Ti-tsl-2021-10-10.frc".
    
```

5 – оптимизированные параметры, 6 – относительная среднеквадратичная ошибка, 7 – путь к сгенерированному **frc** файлу.

8. Проанализируйте полученные результаты. Это можно сделать с помощью оценки ошибки расчетов по сравнению с результатами ТФП моделирования. В случае энергий и сил получается прямая линия с наклоном, близким к единице, что свидетельствует о хорошем соответствии полученных в рамках машинно-обучаемого потенциала результатов с вычислениями ТФП.



Сохраните сгенерированный потенциал. Для этого в списке выходных файлов найдите тот, у которого расширение **.frc**, нажав правой кнопкой мыши, сохранить как, и выберите путь **C->MD->data->** и выберите любую удобную папку. Важно, чтобы потенциал «лежал» на всех компьютерах в одной папке для проведения дальнейших расчетов.

9. Далее перейдем к валидации полученного потенциала.

Рассчитайте равновесные и упругие свойства для этих фаз [61–65]. Постройте гистограммы, где представлено сравнение упругих констант C_{11} , C_{13} , C_{44} , объемного, сдвигового модулей упругости, модуля Юнга, параметров решеток, плотности, подсчитанных вашим потенциалом с ТФП, экспериментальными данными, рассчитанными полуэмпирическими потенциалами. Также можно произвести генерацию потенциала с помощью нейронной сети (для этого в параметрах MLP вместо SNAP надо выбрать NN). Сравните полученные результаты с представленными диаграммами.

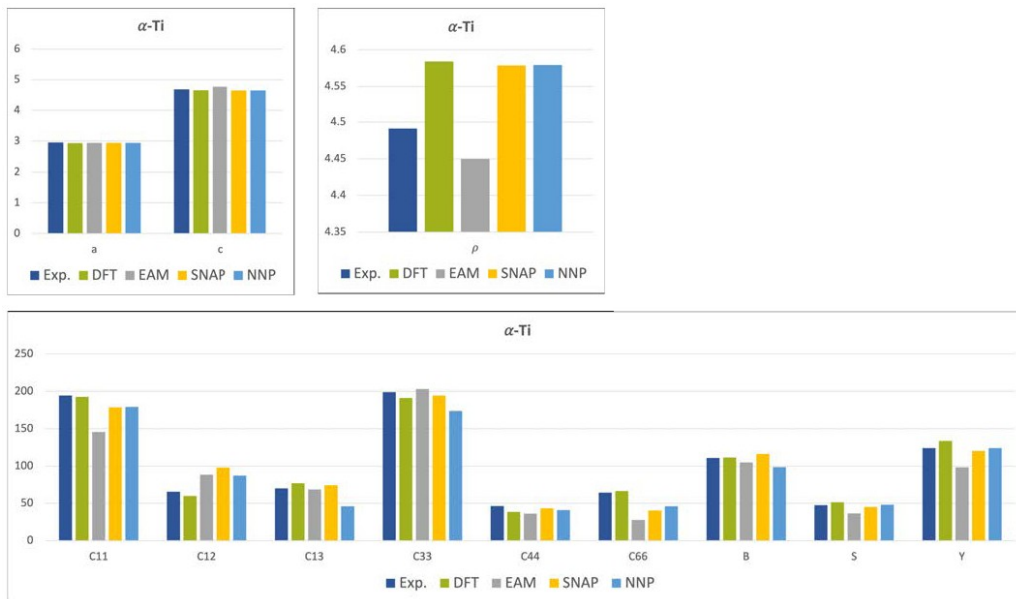


Рис. 10. Сравнение с экспериментальными данными параметров решетки, плотности и упругих констант α -Ti, полученных с помощью ТФП и с помощью машинно-обученного потенциала. Параметры решетки a и c указаны в \AA , плотность ρ в г/см^3 и упругие константы C_{ij} в ГПа. Можно видеть хорошее согласование данных

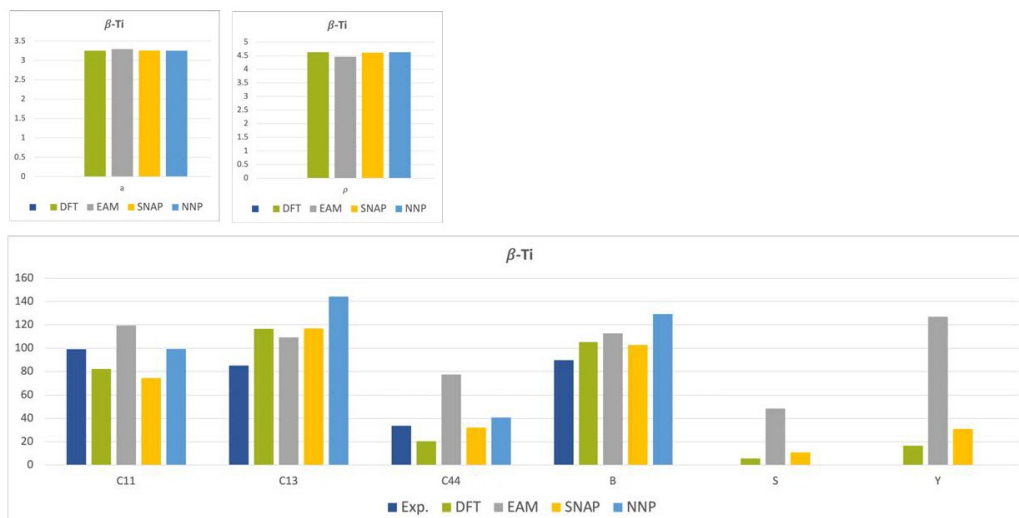


Рис. 11. Сравнение с экспериментальными данными параметров решетки, плотности и упругих констант β -Ti, полученных с помощью ТФП и с помощью машинно-обученного потенциала

Сделайте выводы о качестве предсказательных способностей различных методов. В чем могут быть причины расхождений?

10. **Перейдем к исследованию фазовых переходов в титане (Ti) с помощью полученного потенциала [58–60].** В то время как предыдущие результаты могут быть получены как из расчетов *ab initio*, так и из расчетов с помощью силовых полей и, таким образом, могут быть использованы для проверки машинно-обученного потенциала, исследование $\alpha - \beta$ фазового перехода выходит за рамки возможностей методов ТФП и, следовательно, машинно-обученные потенциалы могут полностью реализовать свои возможности, а именно их высокую эффективность в сочетании с точностью, унаследованной от ТФП.

Чтобы проиллюстрировать возможности машинно-обученного потенциала, $\alpha - \beta$ фазовый переход был исследован путем моделирования молекулярной динамики при различных температурах от 300 до 2200 К с использованием программного модуля MedeA LAMMPS с использованием сгенерированного машинно-обученного потенциала SNAP-MLP. При каждой температуре суперъячеечные структуры уравнивались в течение 20 пс с временным шагом 1 фс. Вы можете попробовать самостоятельно построить блок-схему для этого расчета или открыть готовую из папки PCS-SERVER-> SharePCS -> 0TutorialsMedeA -> Методичка-ML-задания -> упражнение2. В частности, при нагреве и охлаждении моделирование начиналось при 200 и до 2200 К с суперъячейками ГПУ и ОЦК структур, состоящих из 108 ($3 \times 3 \times 3$ суперъячейка) и 128 ($4 \times 4 \times 4$ суперъячейка) атомов соответственно. Постройте график зависимости энергии уравновешенных структур при каждой температуре, как показано на Рис. 12

На рисунке можно видеть, что при температуре около 1300 К происходит изменение структуры Ti (скачок энергии), что указывает на наличие скрытой теплоты. Сравните с тем значением, которое получилось у вас. При температурах выше 1300 К титан находится в β -фазе, а при 1900 К переходит в жидкую фазу. Эти данные соответствуют экспериментальными данными о температурах фазового перехода и плавления, которые составляют 1156 К и 1946 К соответственно.

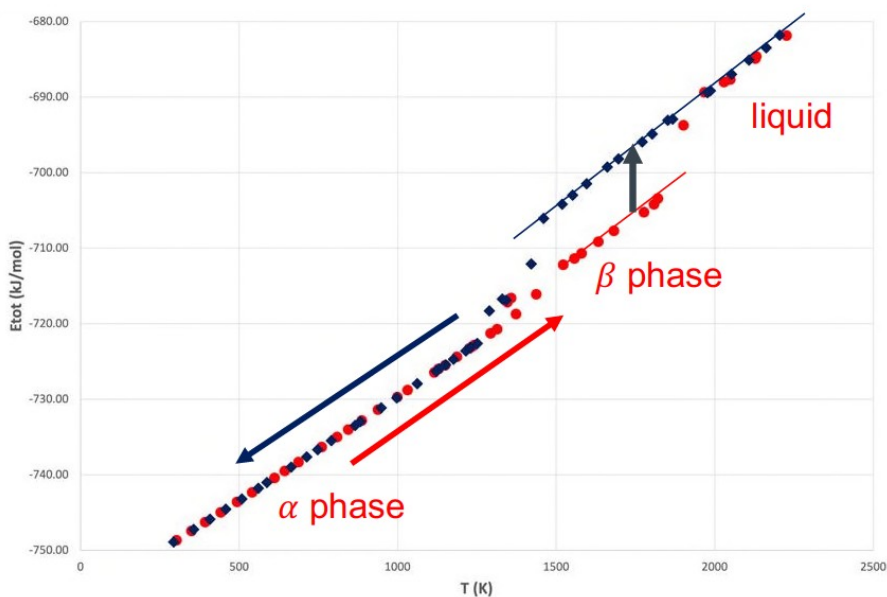


Рис. 12. Исследование фазовых переходов Ti

Дополнительное задание: Похожий график зависимости энергии от температуры можно получить с помощью фононного расчета. Для этого откройте окно для построения блок-схем, откройте LAMMPS Phonon flowchart из библиотеки Open Library, выберите блок-схему MD/data/Flowcharts.kit/Phonon и выберите LAMMPS.flow. Выбор параметров можно оставить по умолчанию. Расчет выполняется достаточно быстро. Для разных фаз на одном поле постройте зависимости $E(T)$. Точка пересечения соответствует фазовому переходу. Сравните с результатом, полученным с помощью LAMMPS на предыдущем рисунке.

- Также можно проанализировать **парные корреляционные функции** для каждой фазы (Рис. 13). Для этого откройте каждую из структур и выполните расчет методом молекулярной динамики в ансамбле NPT для суперъячеек, состоящих из нескольких элементарных. Блок-схему можно открыть готовую из библиотеки Open Library-> Lammps->Simple molecular dynamics, выбрав температуру выше температуры плавления, например 2000 К.

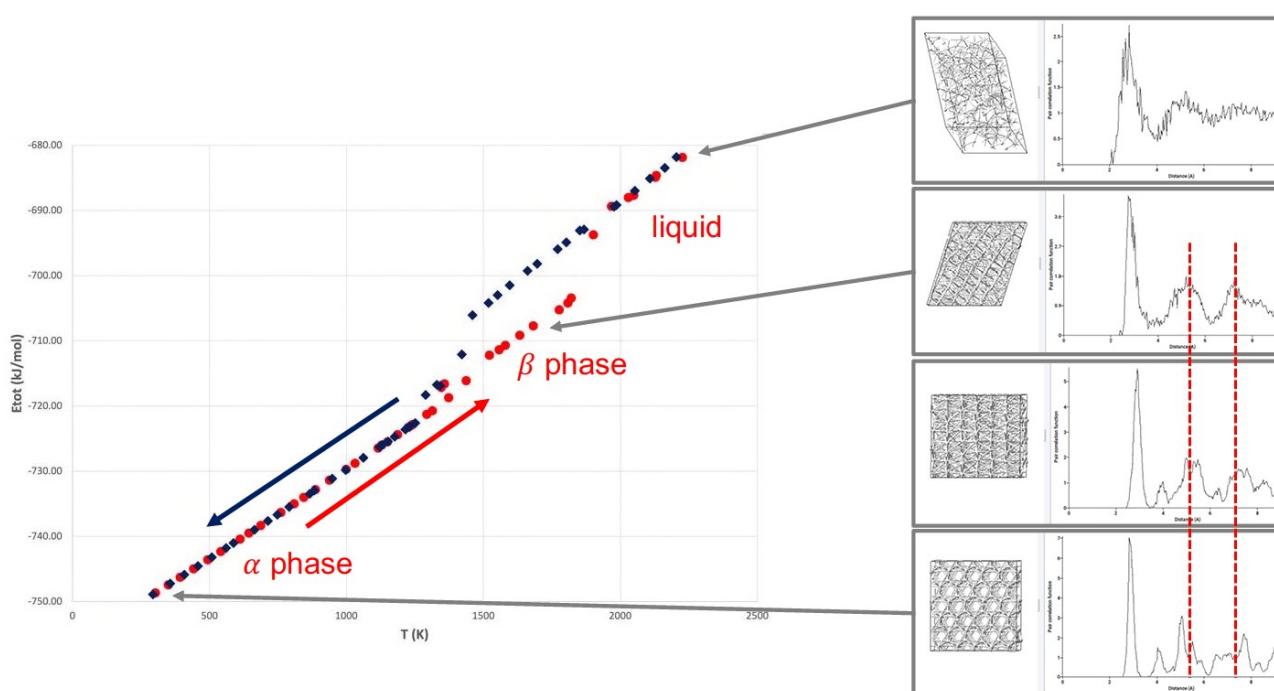


Рис. 13. Парные корреляционные функции, соответствующие различным фазам Ti

Чтобы открыть окно для наблюдения парных корреляционных функций, откройте конечную конфигурацию после проведения расчета, Analysis -> Geometric analysis -> Pair Correlation. Сравните полученную зависимость с Рис. 11. Действительно, можно увидеть, что для самой высокотемпературной фазы отсутствует дальний порядок, что указывает на жидкую фазу. Напротив, для α и β фаз наблюдаем пики в дальнем порядке.

Контрольные вопросы и дополнительные задания:

- Сколько структур получилось в вашем обучающем наборе? Подумайте, насколько полученный набор является полным, какие еще конфигурации можно включить в набор?
- По каким результатам можно судить о точности полученного потенциала? Насколько точным получился ваш?

3. Представьте результаты расчетов зависимости энергии от температуры для исследуемых фаз, определите температуру фазового перехода, сравните ее с экспериментальными данными.
4. Вычислите парные корреляционные функции для различных фаз. Что они показывают? Соответствуют ли полученные зависимости представленным в описании?
5. Представьте результаты механических свойств, а также сравнение с экспериментальными данными. Сделайте выводы о точности.

4.3. Создание обучающих наборов и машинно-обучаемых потенциалов на примере поверхности Nb

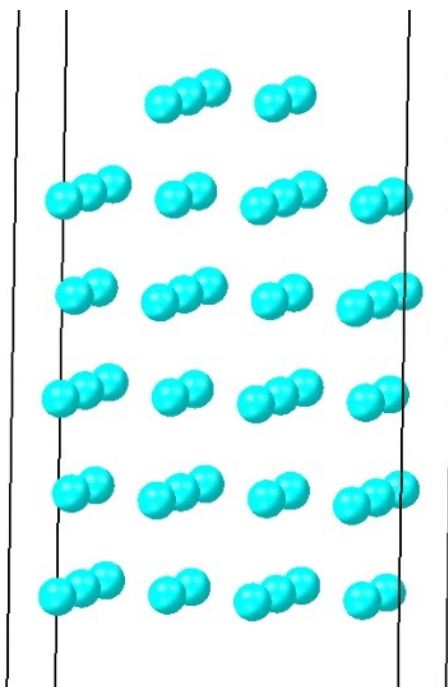
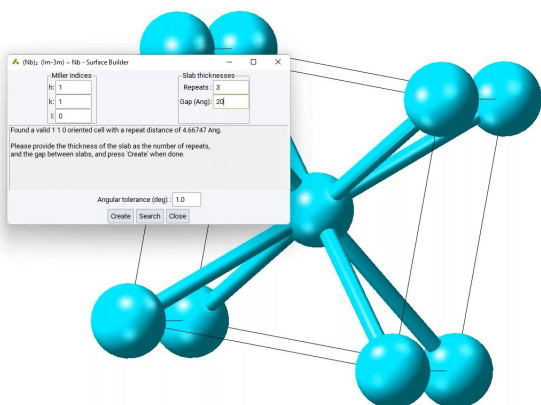
Цель: Создана машинно-обученного потенциала для моделирования воздействия нанокластеров ниобия на его поверхность.

Задачи:

1. Построить поверхность ниобия (Nb).
2. Использовать **VASP** для моделирования *ab-initio* молекулярной динамики.
3. Получить *ab-initio* MD траектории и использовать их для обучения потенциала спектрального анализа соседей **SNAP MLP**.
4. Применить машинно-обученный потенциал для дальнейшего моделирования свойств.

Порядок выполнения работы:

1. Постройте поверхность Nb, состоящую из 44 атомов. Для этого перейдите **MD/data/Structure** и найдите *Nb.sci*. После открытия перейдите во вкладку *Builders -> Build surface...* Установите индексы Миллера $h : 1, k : 1, l : 0$, **Repeats** на 3 и **Gap** 20. Нажмите создать и, не меняя никаких параметров, нажмите **ОК**. Получите поверхность Nb.
2. Создайте суперъячейку $2 \times 2 \times 1$. Для этого снова откройте *Builders -> Build surface...* и во вкладке **Simple** установите **Nuber of cell** $a = 2, b = 2, c = 1$, нажмите **ОК**. Удалите 2 строки атомов Nb на одной стороне поверхности для понижения симметрии. Для этого выделите крайние атомы и удалите их.



3. Запустите *ab-initio* MD в модуле **VASP**. Рекомендуемые параметры моделирования:

K-mesh in Brillouin Zone for SCF

| Constraint | Mesh points | Spacing (1/Å) |
|------------|-------------|---------------|
| x: | 3 | 0.317 |
| y: | 3 | 0.224 |
| z: | 1 | 0.185 |

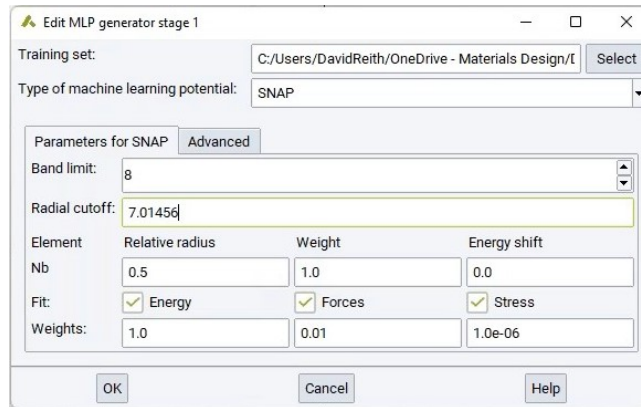
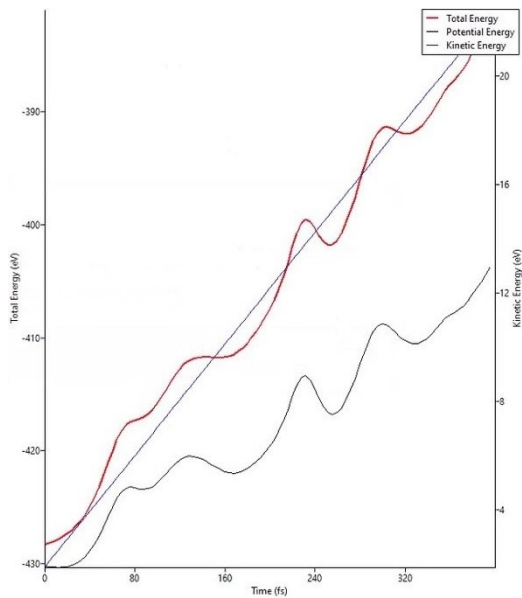
- Тип расчета (Type of calculation): Molecular Dynamics
- Ансамбль (Ensemble): temperature scaling (nVE)
- Продолжительность моделирования (Simulation time): 400 fs
- Начальная температура (Temperature initial): 300 K
- Конечная температура (Temperature final): 4000 K
- Энергия отсечки плоской волны (Plane wave cutoff): 350 eV
- Проекция (Projection): Real space
- Вкладка SCF: Spacing of k-points: 0.4 1/Å, тип размытия (Type of Smearing): Gaussian

4. Запустите расчеты. После завершения получите зависимость полной, кинетической и потенциальной энергий от времени. Это и будет основой обучающего набора, который будет содержать 100 конфигураций системы.

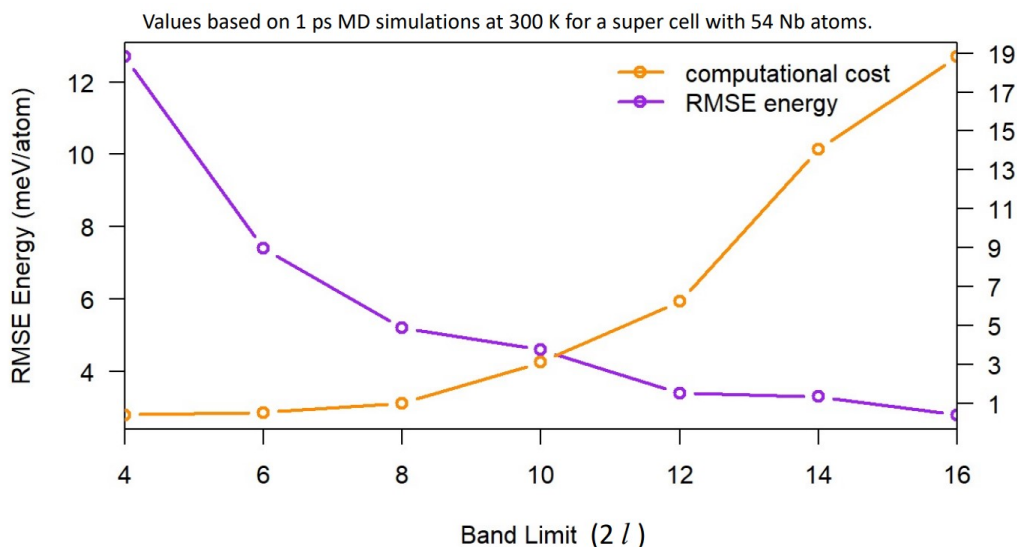
Затем откройте File -> New structure list file и создайте новый список. Добавьте траекторию через Add structure(s) -> Append list/trajectory from job.

Далее нажмите **ОК** для согласования обучающего набора.

5. Настройте параметры для генерации машинно-обучаемого потенциала. Установите тип потенциала SNAP (в дополнение после можно выполнить генерацию и с помощью нейронной сети). Установите угловую точность потенциала, она должна быть кратна двум ($2l$, где l – степень сферических гармоник).



Необходимо отметить, что с ростом l возрастают и вычислительные затраты, что можно видеть на графике ниже. Вычислительные затраты увеличиваются в геометрической прогрессии, а среднеквадратичное отклонение уменьшается экспоненциально. Таким образом, оптимальным выбором будет диапазон от 8 до 12. Выберите для начала $l = 4$.



Результат для $2l = 8$:

Stage 1: MLP Generator

Input for FitsNAP is in "pure_Nb_UGM.in":

Band limit: 8
Radial cutoff: 7.01456 Ang
Elements:

| | Rel. radius | Weight | Energy shift |
|----|-------------|--------|--------------|
| Nb | 1.0000 | 0.5000 | 0.000000 |

Properties used in the fit: energy forces stress
with weights: 1.0 0.01 1.0e-06

The Ziegler-Biersack-Littmark screened nuclear repulsion is not used.

Fitting statistics

| Group | Weighting | Subsystem | Count | Mean absolute error | Root mean absolute error | Root mean squared error | Relative root mean squared error | R squared |
|----------------|-----------|-----------|-----------|---------------------|--------------------------|-------------------------|----------------------------------|-----------|
| all Unweighted | Combined | 13900 | 249.3306 | 0.7052 | 1503.9435 | 0.6337 | 0.5984 | |
| all Unweighted | Energy | 100 | 0.0167 | 0.0659 | 0.0182 | 0.0599 | 0.9964 | |
| all Unweighted | Forces | 13200 | 0.3624 | 0.4448 | 0.4883 | 0.4360 | 0.8099 | |
| all Unweighted | Stress | 600 | 5768.1841 | 0.8496 | 7238.7448 | 0.8514 | 0.2752 | |
| all Weighted | Combined | 13900 | 0.0038 | 0.0513 | 0.0052 | 0.0067 | 1.0000 | |
| all Weighted | Energy | 100 | 0.0167 | 0.0659 | 0.0182 | 0.0599 | 0.9964 | |
| all Weighted | Forces | 13200 | 0.0036 | 0.4448 | 0.0049 | 0.4360 | 0.8099 | |
| all Weighted | Stress | 600 | 0.0058 | 0.8496 | 0.0072 | 0.8514 | 0.2752 | |

Units: energy: eV/atom, force: eV/Angstrom, pressure: bar

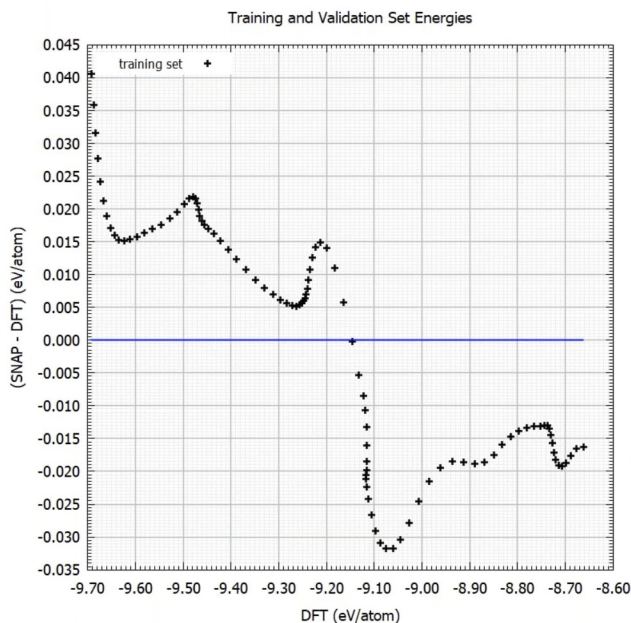
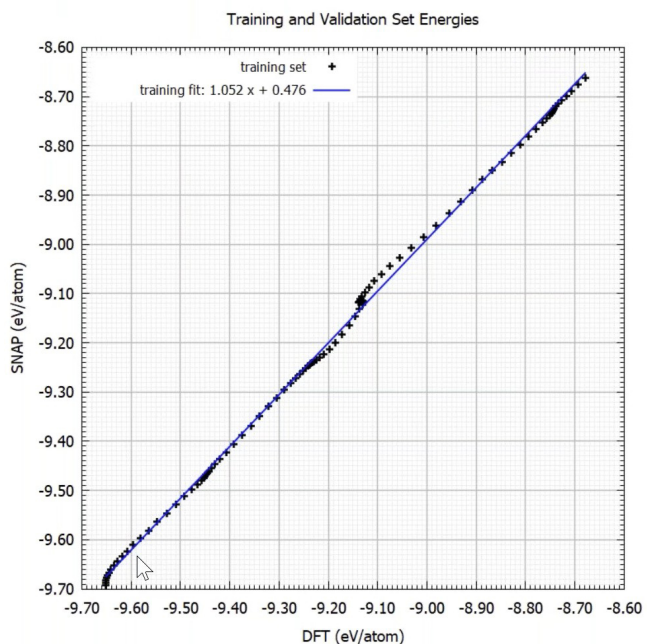
Generated SNAP has been converted to "C:/MD/Job/Dir0/39/Stage_1/pure_Nb_UGM.frc".

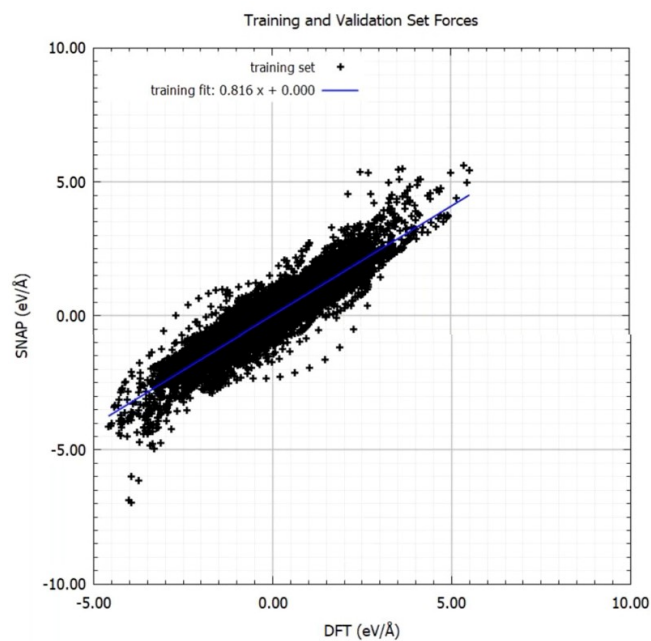
Publications using potentials generated with the SNAP option should quote the following work:

A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, *J. Comp. Phys.* 285, 316-330 (2015)
(<https://doi.org/10.1016/j.jcp.2014.12.018>)

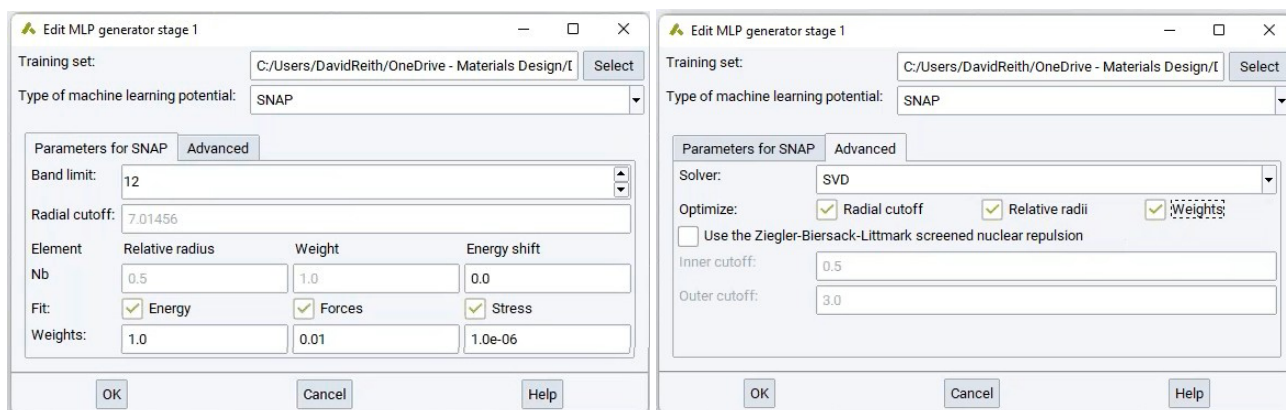
M. A. Wood and A. P. Thompson, Extending the accuracy of the SNAP interatomic potential form, *J. Chem. Phys.* 148, 241721 (2018)
(<https://doi.org/10.1063/1.5017641>)

Можно видеть, что ошибка для энергий относительно высока. Для силы она имеет приемлемое значение.





На графиках выше видно, что для энергии наблюдается хорошая согласованность результатов с ТФП расчетами, но имеются систематические колебания. Касательная на графике сил имеет отклонение (порядка 5° относительно линии). Поэтому необходимо улучшить результат, оптимизировав параметры. Установите $2l = 12$, затем во вкладке **Advanced** отметьте флажком все пункты строки **Optimize**. Далее запустите расчет.



Результат для $2l = 12$

Stage 1: MLP Generator

```
Input for FitsNAP is in "pure-Nb.in":
Band limit: 12
Radial cutoff: 7.01456 Ang
Elements:
+-----+
| Rel. radius Weight Energy shift |
+-----+
| Nb 1.0000 0.5000 0.000000 |
+-----+

Properties used in the fit: energy forces stress
with weights: 1.0 0.01 1.0e-06

The Ziegler-Biersack-Littmark screened nuclear repulsion is not used.

Fitting statistics
+-----+
| Group Weighting Subsystem Count Mean abso- Root mean ab- Root mean Relative root mean R
| | | | | lute error | solute error | squared error | squared |
+-----+
| all Unweighted Combined 13900 58.9803 0.1668 355.5365 0.1498 0.9776 |
| all Unweighted Energy 100 0.0028 0.0201 0.0034 0.0197 0.9996 |
| all Unweighted Forces 13200 0.1691 0.2076 0.2313 0.2065 0.9574 |
| all Unweighted Stress 600 1362.6567 0.2007 1711.2594 0.2013 0.9595 |
| all Weighted Combined 13900 0.0017 0.0221 0.0023 0.0029 1.0000 |
| all Weighted Energy 100 0.0028 0.0201 0.0034 0.0197 0.9996 |
| all Weighted Forces 13200 0.0017 0.2076 0.0023 0.2065 0.9574 |
| all Weighted Stress 600 0.0014 0.2007 0.0017 0.2013 0.9595 |
+-----+

Units: energy: eV/atom, force: eV/Angstrom, pressure: bar

The following parameters are optimized during the MLP generator run:
Parameter Initial Optimized
Radial cutoff 7.01456 6.23623
Relative radius for Nb 0.50000 0.68287
Weight for Nb 1.00000 0.97502

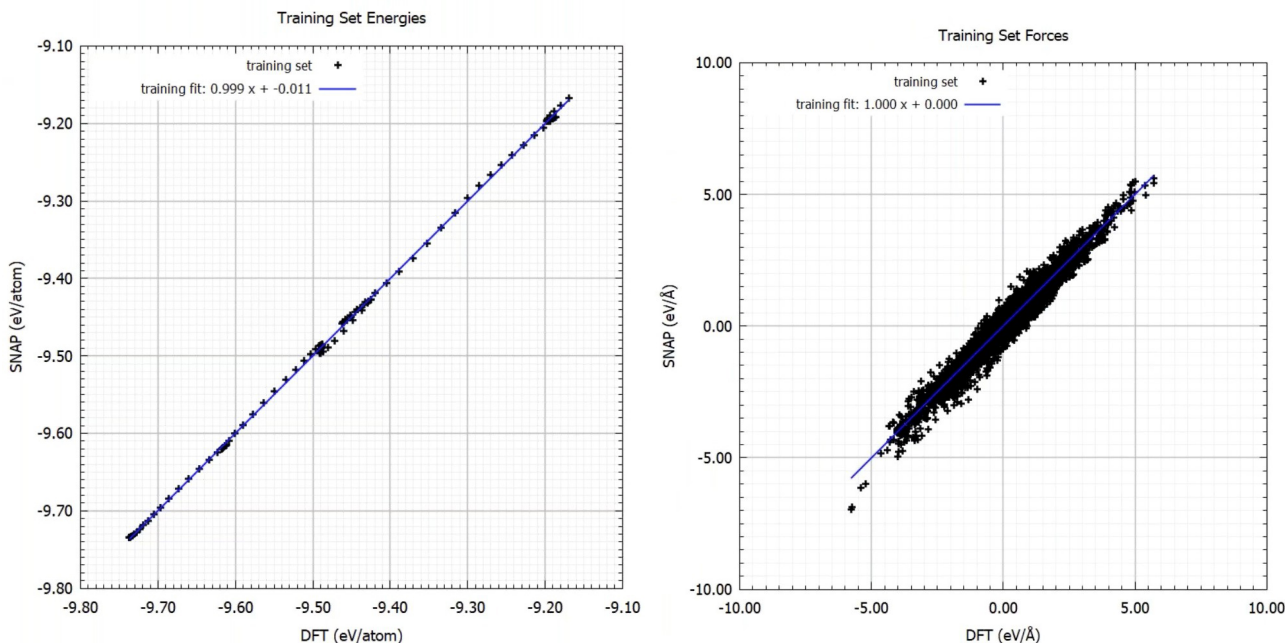
Generated SNAP has been converted to "C:/MD_nightly/Jobs/dir0/60/Stage_1/pure-Nb.frc".

Publications using potentials generated with the SNAP option should quote the following work:
A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated
generation of quantum-accurate interatomic potentials, J. Comp. Phys. 285, 316-330 (2015)
(https://doi.org/10.1016/j.jcp.2014.12.018)

M. A. Wood and A. P. Thompson, Extending the accuracy of the SNAP interatomic potential form, J. Chem. Phys. 148, 241721 (2018)
(https://doi.org/10.1063/1.5017641)
```

Среднеквадратичная ошибка энергии и силы уменьшена в несколько раз. На графиках все еще есть небольшие отклонения, но соответствие стало лучше. Наклон прямой близок к 45° , что говорит об однозначном соответствии значениям ТФП. В итоге получается хорошее силовое поле, которое можно использовать для дальнейших расчетов. Для удобства сохраните файл с расширением **frc** в папку **Forcefields** или в любую другую.

Примечание: При использовании силового поля файл с расширением **frc** нужно сохранить на компьютере, где производятся вычисления, и на сервере. В противном случае файл не будет найден.



6. Перейдите к выполнению основной задачи. Для этого включите силовое поле, зайдя во вкладку **Forcefields** -> **Read** и найдя папку, в которую сохранили файл. Обратите внимание на автоматически сгенерированное описание, например, количество структур в обучающем наборе 100; среднеквадратичная ошибка энергии и сил – $0.0034 \frac{eV}{atom}$ и $0.2313 \frac{eV}{Ang}$. Нажмите **Open** и силовое поле активируется.

Open Forcefield File

Directory: C:/MD/data/Forcefields/MLP

| | | | |
|--------------------------------------------|-------------------------------|---------------------------------------|-----------------------|
| H20-training-set-trajectories.frc | Nb_and_oxygen-bandlimit12.frc | pure-Nb-bandlimit14.frc | Ti-ts1-2021-10-13.000 |
| H20-training-set-trajectories_-_no_ice.frc | pure-Nb-bandlimit4.frc | pure-Nb-bandlimit16.frc | Ti-ts1-2021-10-13.000 |
| Nb_and_oxygen-bandlimit8.frc | pure-Nb-bandlimit6.frc | pure-Nb-UGM.frc | TiO2-MLPG-trainingli |
| Nb_and_oxygen-bandlimit10.frc | pure-Nb-bandlimit8.frc | Ti-ts1-2021-10-10.000079-dev58197.frc | UGM-TiO2-full-trainin |
| Nb_and_oxygen-bandlimit12-ZBL-r1.0.frc | pure-Nb-bandlimit10.frc | Ti-ts1-2021-10-10.000079.frc | |
| Nb_and_oxygen-bandlimit12-ZBL.frc | pure-Nb-bandlimit12.frc | Ti-ts1-2021-10-10.000191-dev58328.frc | |

Description:

Elements: Nb

Automatically generated Spectral Neighbor Analysis Potential (SNAP) forcefield file from job 60 based on the structure list pure-Nb.

The 1 structures in the structure list are partitioned into training set and validation set.
Number of structures in training set: 100

Errors:

| Term | RMSE |
|---------------------|----------------|
| Training set energy | 0.0034 eV/atom |
| Training set forces | 0.2313 eV/Ang |
| Training set stress | 1711.2594 bar |

Publications using the generated forcefield should quote the following work:

A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker,
Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials,
J. Comp. Phys. 285, 316-330 (2015)

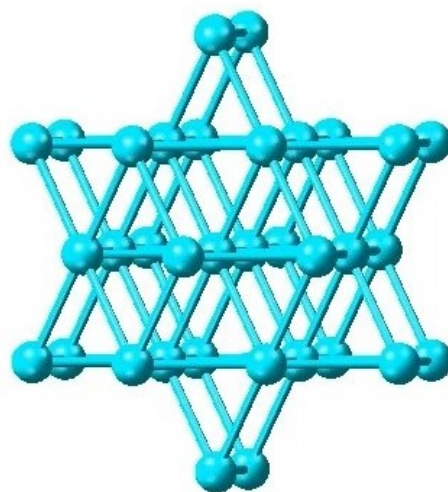
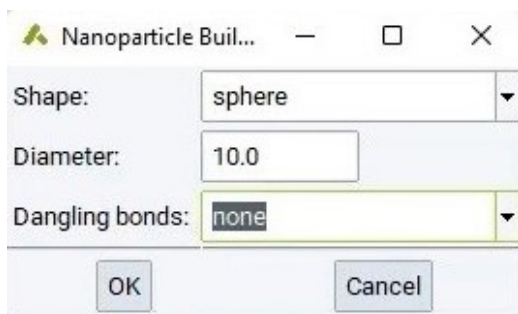
File name: pure-Nb-UGM.frc Open

Files of type: Forcefield Files (*.frc) Cancel

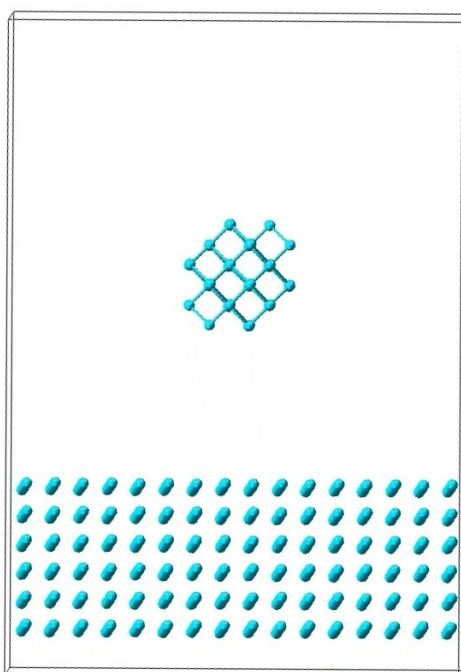
Во вкладке **Choose** можно проверить наличие установленного силового поля.



Создайте ячейку для моделирования. Она будет состоять из ранее созданной поверхности и наночастицы Nb. Создайте наночастицу. Для этого снова откройте файл

Nb.sci и перейдите во вкладку *Builders -> Nanoparticles....* Частица будет иметь форму сферы и радиус 10Å.



На первом рисунке представлены параметры частицы, а на втором – полученная наночастица. Теперь с помощью правой кнопки мыши можно скопировать структуру и вставить в окно с поверхностью. Должна получиться следующая картина:



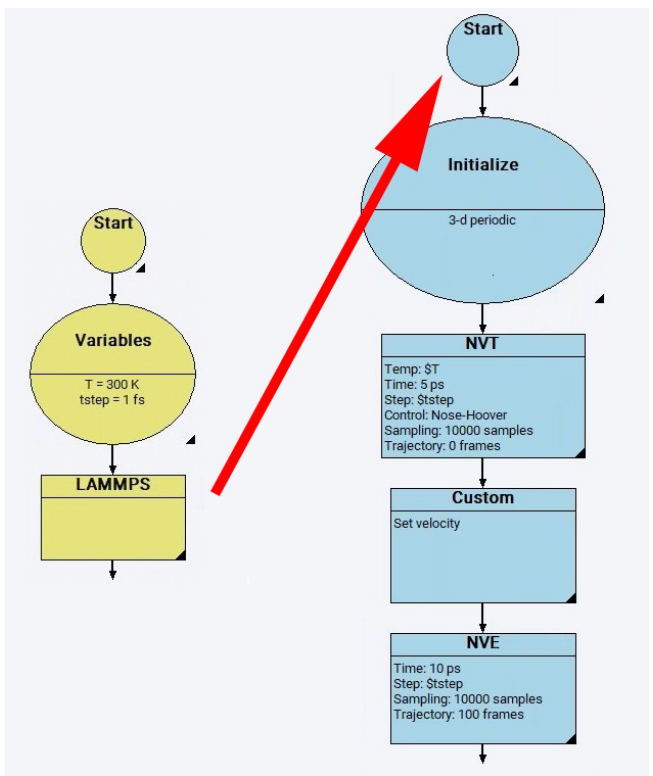
Далее необходимо смоделировать воздействие наночастицы Nb на поверхность, следовательно кинетическая энергия частицы будет передана поверхности и она начнет двигаться вниз. Чтобы не допустить этого, необходимо заморозить атомы на границе ячейки. Для этого выбираем на верхней панели  и с помощью  выбираем атомы на границе и замораживаем их.

The screenshot shows the Medea software interface. On the left, a 3D model of a nanosystem is displayed, consisting of a central cluster of atoms and a larger, flat structure below it. On the right, an 'Atoms' table lists the coordinates and properties of the atoms in the system. The table has the following columns: Atomic Number, Wyckoff Position, Wyckoff Equation, X, Y, Z, Freeze, Occupancy, and Spin. The data is as follows:

| Atomic Number | Wyckoff Position | Wyckoff Equation | X | Y | Z | Freeze | Occupancy | Spin |
|---------------|------------------|------------------|--------------|-----------|----------|--------|-----------|------|
| 41 | 1a | xyz | 2.71707e-009 | 0.0898191 | 0.239045 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.0898191 | 0.109398 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.0273191 | 0.239045 | xyz | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.0273191 | 0.195829 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.0898191 | 0.152614 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.0898191 | 0.195829 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.0273191 | 0.152614 | xyz | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.0273191 | 0.28226 | xyz | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.0898191 | 0.066183 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.0898191 | 0.28226 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.0273191 | 0.066183 | xyz | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.152319 | 0.109398 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.214819 | 0.239045 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.214819 | 0.109398 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.152319 | 0.239045 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.152319 | 0.195829 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.214819 | 0.152614 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.214819 | 0.195829 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.152319 | 0.152614 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.152319 | 0.28226 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.214819 | 0.066183 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.214819 | 0.28226 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.152319 | 0.066183 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.277319 | 0.109398 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.339819 | 0.239045 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.339819 | 0.109398 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.277319 | 0.239045 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.277319 | 0.195829 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.339819 | 0.066183 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.339819 | 0.28226 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.277319 | 0.066183 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.402319 | 0.109398 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.464819 | 0.239045 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.464819 | 0.109398 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.402319 | 0.239045 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.402319 | 0.195829 | --- | 1 | 0 |
| 41 | 1a | xyz | 2.71707e-009 | 0.464819 | 0.152614 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.464819 | 0.195829 | --- | 1 | 0 |
| 41 | 1a | xyz | 0.0625 | 0.402319 | 0.152614 | --- | 1 | 0 |

Для удобства выделите наночастицу другим цветом. Нажмите правой кнопкой мыши на структуру, выберите Subsets -> Select atoms in subset -> Particle. Потом на панели найдите View -> Options... и в открытом окне выберите Change subset color... Откройте и выберите красный цвет.

После подготовки модели настройте протокол моделирования с помощью интерфейса Medea flowchart.



Используйте параметры моделирования, показанные на рисунке выше. Установите температуру, временной шаг и прочие параметры. Скорость частицы установите $40 \text{ \AA}/\text{ps}$. Для этого в блоке Custom пропишите `velocity subset_particle set 0 0 -40`. Запустите расчет.

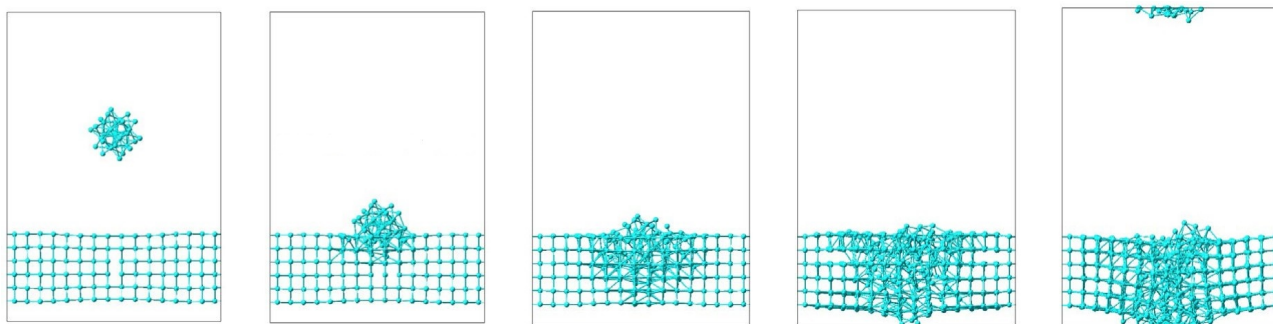
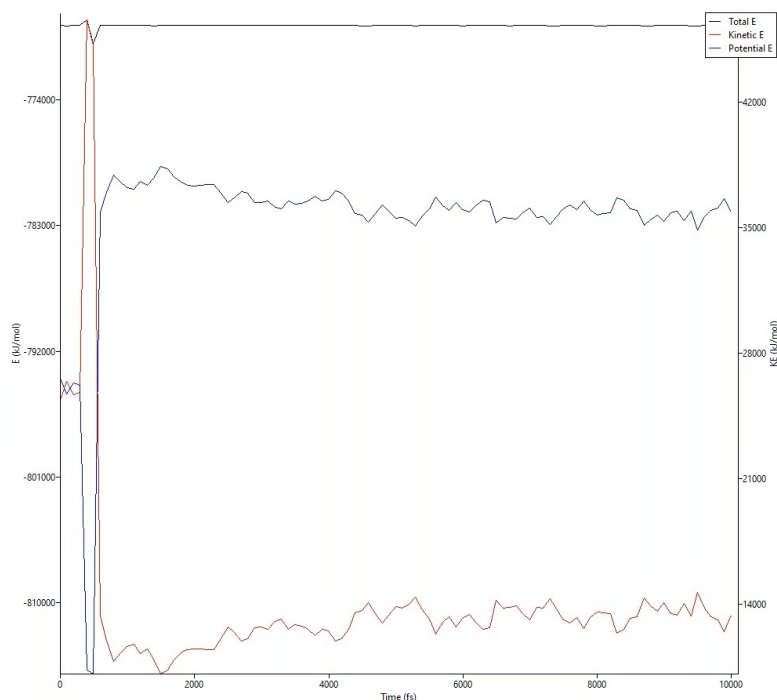


График с траекториями представлен ниже.



Видно, что происходит передача импульса наночастицей поверхности ниобия, которая начинает колебаться.

В итоге, с помощью машинного обучения можно пронаблюдать процесс осаждения наночастицы Nb на поверхность Nb (110). Из траектории, содержащей 44 атома, задача была решена для 809 атомов.

Контрольные вопросы и дополнительные задания:

1. Можно ли использовать потенциал нейронной сети (NNP) для решения данной задачи? При наличии времени и компьютерных ресурсов можно выполнить сравнение результатов, полученных разными методами.
2. Попробуйте оценить выигрыш во времени, получаемый с помощью машинного обучения. Сравните временные затраты на потенциалы SNAP и NNP.
3. Подумайте, как изменятся результаты моделирования при увеличении размера наночастицы, при изменении скорости осаждения. Проверьте предположения с помощью расчетов.

4.4.1. Дополнительные задания: моделирование десорбции воды на поверхности ниобия (Nb)

Цель: Расширьте обучающий набор таким образом, чтобы можно было учитывать взаимодействие с кислородом (десорбция кислорода на поверхность Nb).

Порядок выполнения работы:

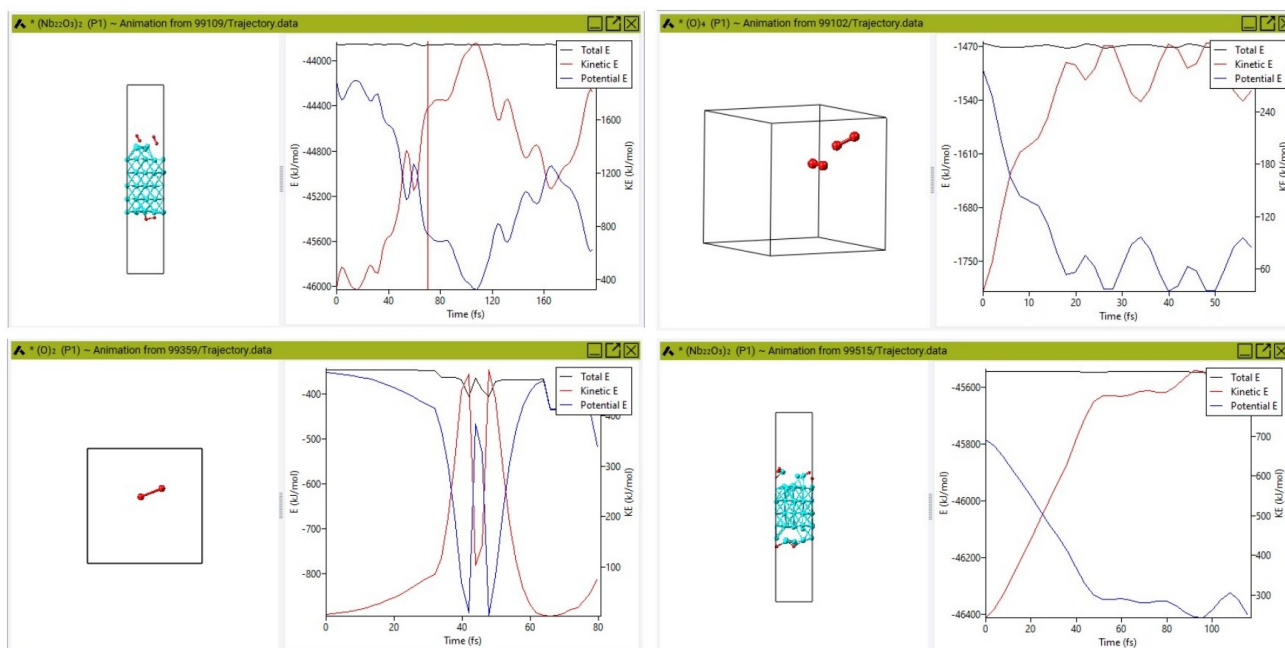
1. Параметры для **VASP** (соответствующим образом отрегулируйте временной шаг и продолжительность моделирования):

- Тип расчета (Type of calculation): Molecular Dynamics.
- Ансамбль (Ensemble): temperature scaling (nVE)
- Энергия отсечки плоской волны (Plane wave cutoff): 350 eV.
- Проекция (Projection): Real space
- Вкладка Функционал/Потенциал (Functional/Potential): Specific Potentials per Element O.
- Вкладка SCF: Spacing of k-points: 0.4 1/Ang, флажок на Use odd size grids, Type of smearing: Gaussian.

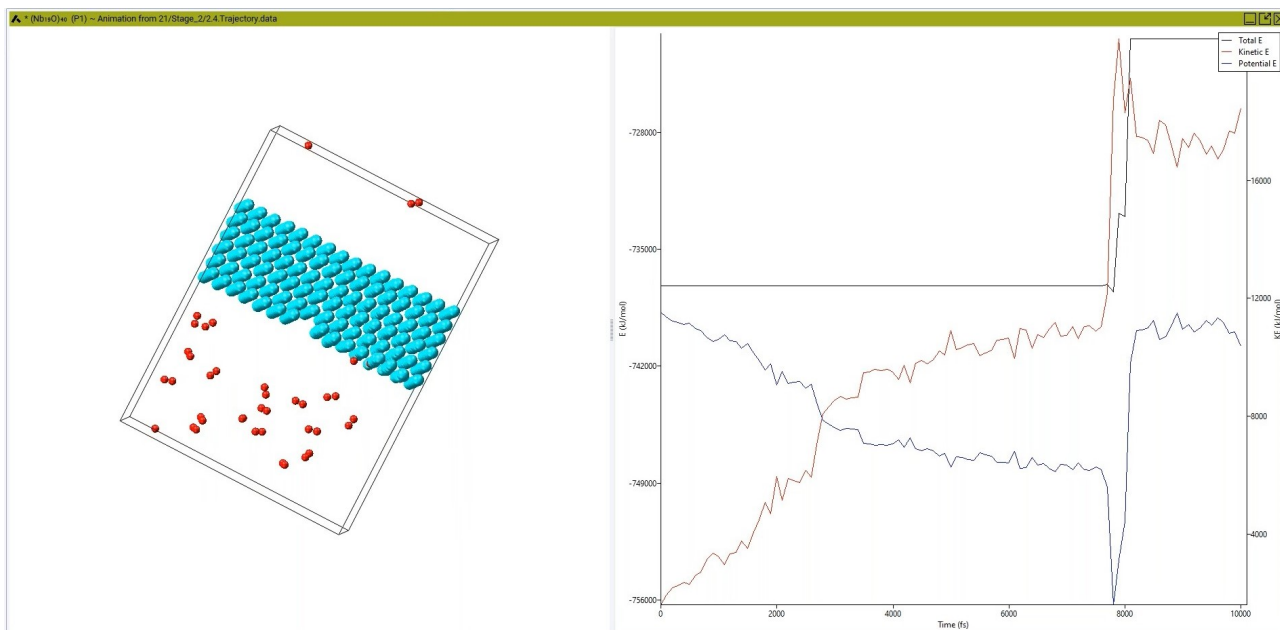
2. Параметры для SNAP MLP:

- Band limit: 12.
- Флажок на всех трех параметрах в строке Optimize (Radial cutoff, Relative radii, Weights).

Получатся следующие итоговые траектории:



Далее необходимо смоделировать взаимодействие O_2 с поверхностью Nb (110):



3. Создать машинно-обученный потенциал для воды.

Указания: Использовать три структуры с разной плотностью системы ρ , равной 900, 1000 и 1100 kg/m^3 .

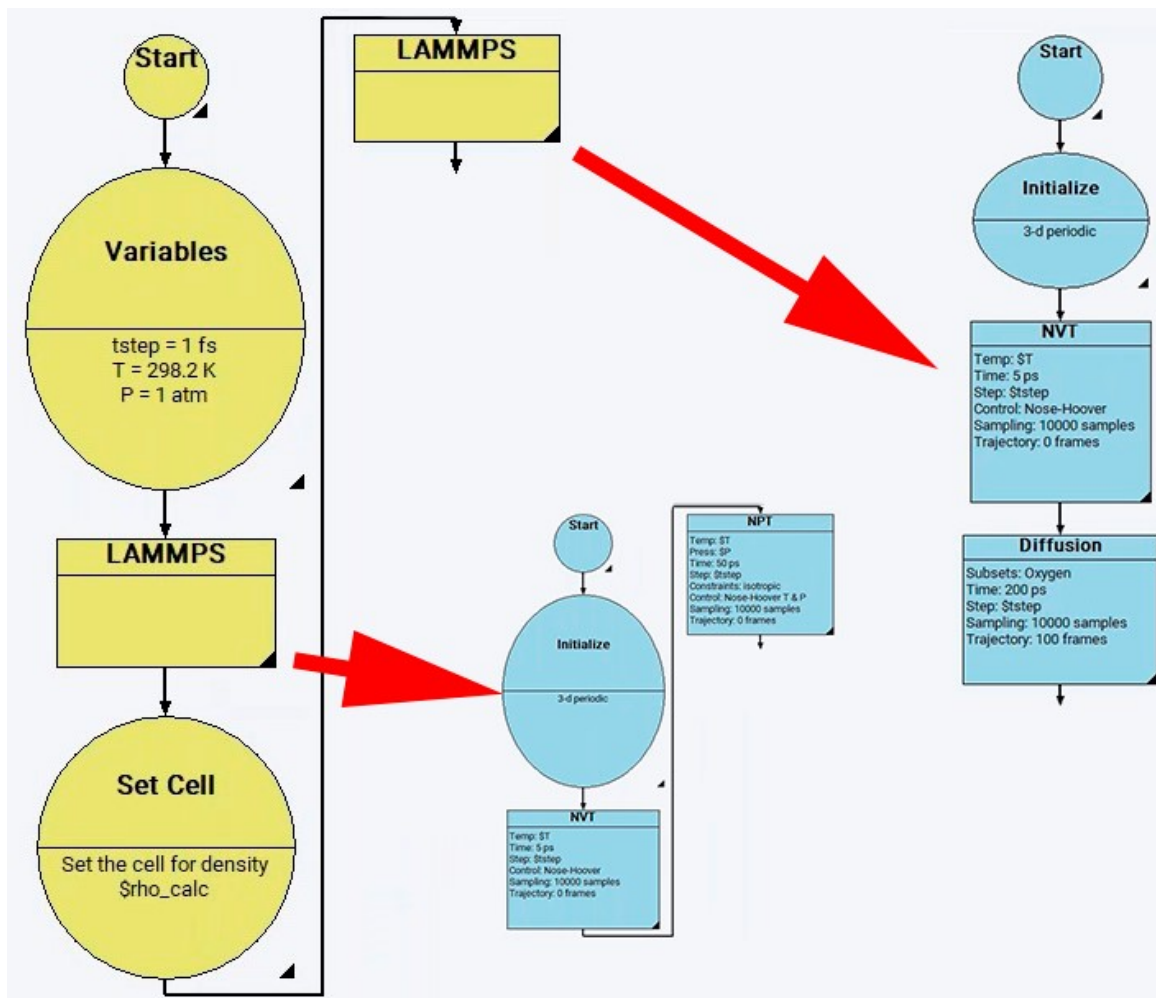
Параметры для **VASP** (соответствующим образом отрегулируйте временной шаг и продолжительность моделирования):

- Тип расчета (Type of calculation): Molecular Dynamics
- Ансамбль (Ensemble): micro canonical (nVE)
- Начальная температура (Temperature initial): 298 K.
- Коррекцией для ван-дер-ваальсового взаимодействия (Van der Waals): DFT-D3 zero-damping
- Точность (Precision): Standard 500.
- Вкладка SCF: Spacing of k-points: 0.21/Ang, флажок на Use odd size grids, Type of smearing: Gaussian.

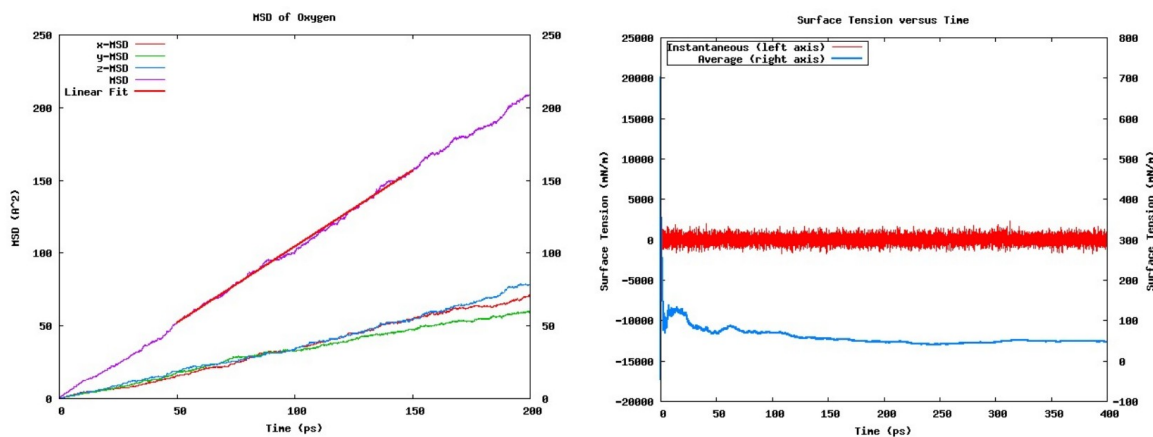
Параметры для SNAP MLP:

- Band limit: 12.
- Флажок на всех трех параметрах в строке Optimize (Radial cutoff, Relative radii, Weights).

Необходимо рассчитать среднеквадратичное смещение (диффузию). Рекомендуется использовать следующую блок-схему с 400 (примерно) молекулами воды в ячейке:



Должны получиться следующие графики:



Найдите коэффициент диффузии D и энэргию γ . Сравните результат с экспериментальными значением $2.229 \cdot 10^{-5} \text{ см}^2/\text{с}$ и 71.99 мН/м . Определите, как хорошо результат описывают другие модели и SNAP, используя представленные ниже значения коэффициента для некоторых моделей.

| Модели | SPC | SPC/Еб | TIP3P | TIP3P-FB | TIP4P | TIP4P/2005 | TIP5P |
|------------------------------------|------|--------|-------|----------|-------|------------|-------|
| $D(10^{-5} \text{ см}^2/\text{с})$ | 4.52 | 2.10 | 6.14 | 2.20 | 3.86 | 2.31 | 2.93 |
| $\gamma(\text{мН/м})$ | 49.5 | 60.4 | 45.9 | 59.8 | 51.4 | 62.6 | 47.9 |

4.4.2. Дополнительные задания: создание обучающего набора и генерирование машинно-обученного потенциала для нитрида алюминия AlN

AlN обладает высокой теплопроводностью и является хорошим электрическим изолятором. Подложки из AlN используются при изготовлении силовых модулей (MOSFET, IGBT), светодиодных корпусов, подложек лазерных диодов, кристаллов, в качестве теплоотводящих подложек в силовой электронике и термоэлектрических модулей.

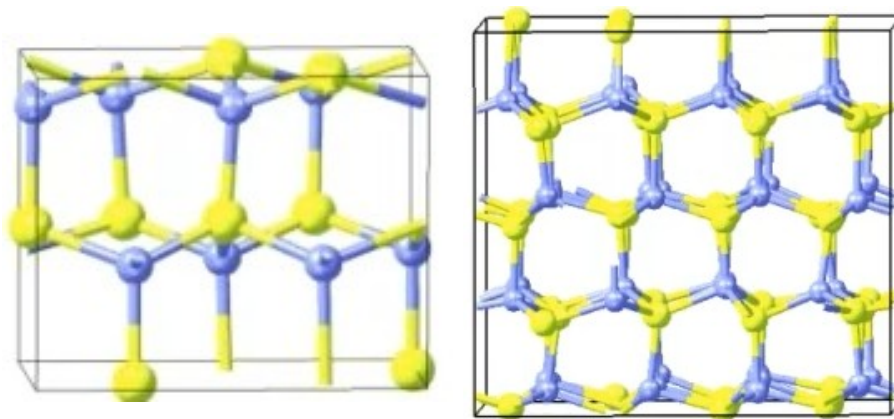
Цель: Создание машинно-обученного потенциала для нитрида алюминия AlN.

Задачи:

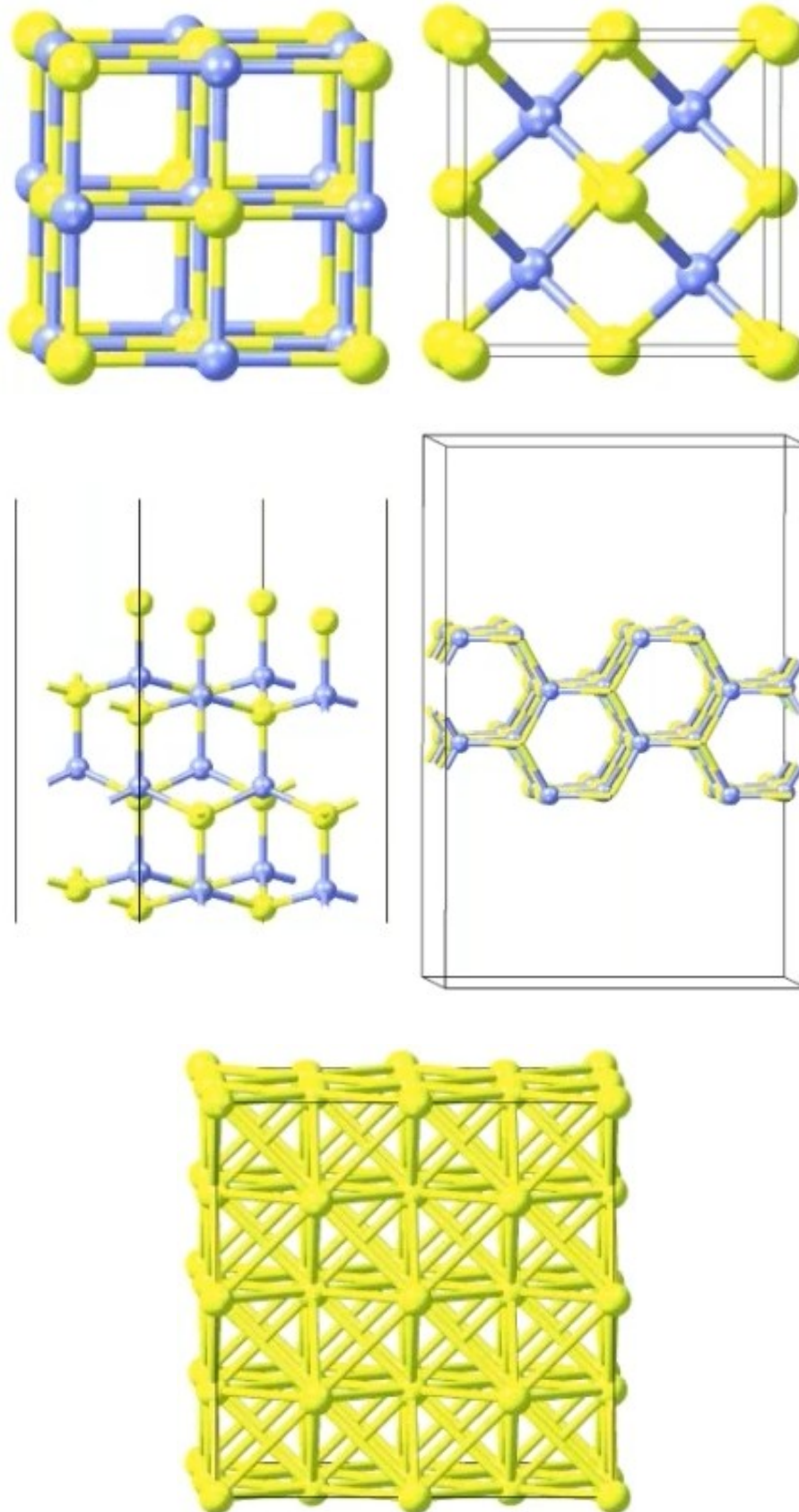
1. Составить обучающий набор и выполнить расчеты ТФП для каждой конфигурации.
2. Собрать результаты ТФП расчетов в структурный лист обучающего набора.
3. Использовать полученные *ab-initio* MD траектории для обучения MLP.
4. Выполнить расчеты методом классической молекулярной динамики с помощью полученного MLP.

Порядок выполнения работы:

1. Для создания обучающего набора постройте суперъячейки вюрцита AlN с 16 и 93 атомами.
2. Проведите *ab-initio* MD моделирование в NPT ансамбле при температурах $T = 150K$, $T = 300K$, $T = 600K$, $T = 1500K$.

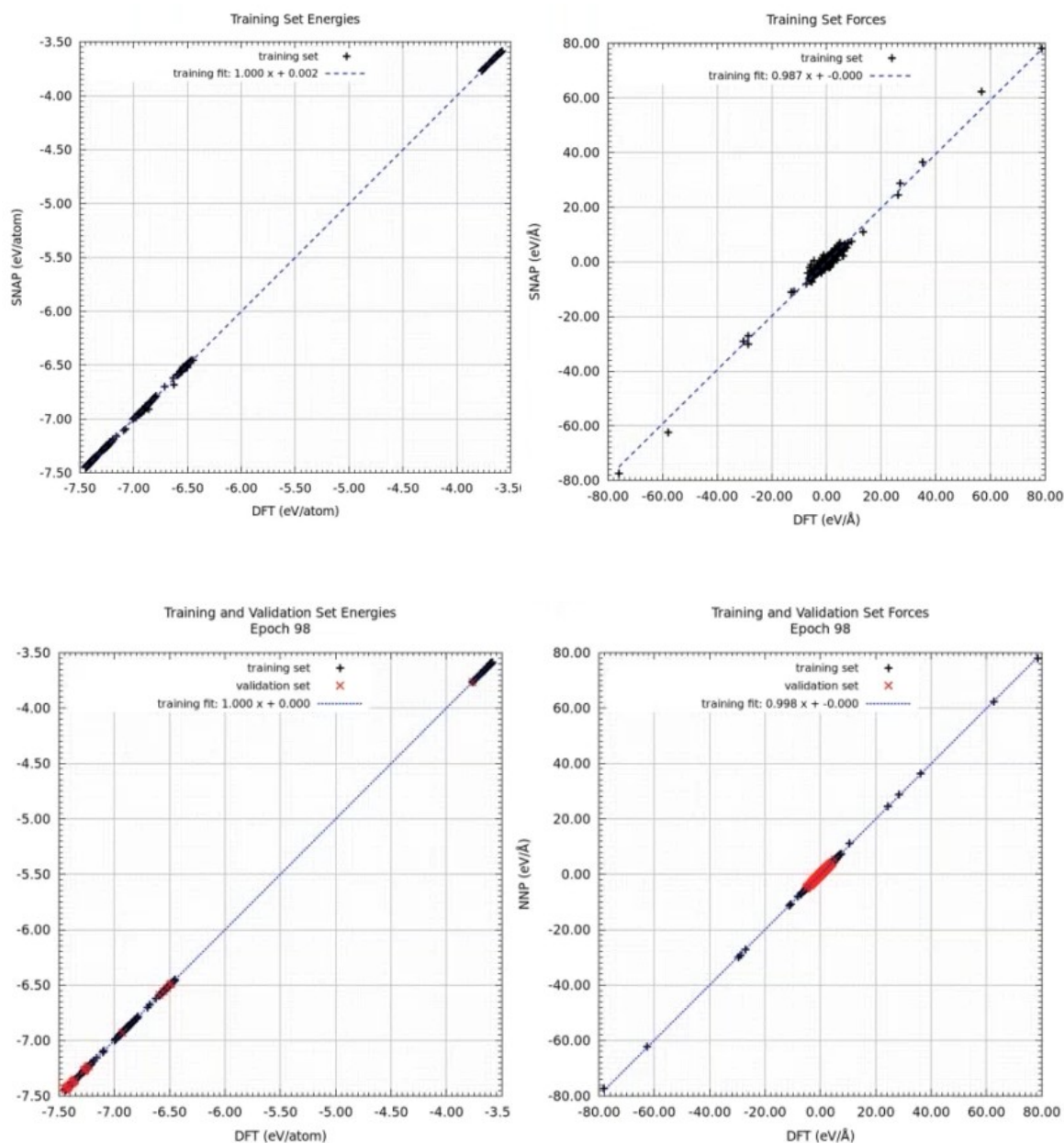


3. Примените одноосные, изотропные и угловые деформации для структур AlN: вюрцита, каменной соли (rocksalt) и цинковой обманки (zinc blende).
4. Выполните *ab-initio* MD моделирование в NVT ансамбле поверхности вюрцита AlN (001) и (110).
5. И, наконец, проведите *ab-initio* MD моделирование для чистой ячейки алюминия ГЦК из 32 атомов в NPT ансамбле.



В итоге получится 431 структура в обучающем наборе. Далее создайте структурный лист и создайте SNAP и NNP потенциал на основе рекомендаций из предыдущих упражнений.

6. Для оценки полученных результатов ориентируйтесь на графики, показанные ниже.



7. Определите, какой из потенциалов лучше описывает фононный спектр. Сравните упругие константы, полученные с помощью потенциалов, с экспериментальными значениями^{1 2}. Найдите коэффициент теплопроводности при трех разных температурах и также сравните с экспериментальными значениями.

Получится 298 структур. Параметры для **VASP**:

- Обменно-корреляционный член DFT функционала (DFT exchange-correlation): GGA-PBE с коррекцией для ван-дер-ваальсового взаимодействия Grimme D3.
- Энергия отсечки плоской волны (Plane wave cutoff): 510 eV.
- Интервал k-сетки (k spacing): 0.178 1/Å.

Далее создайте SNAP и NNP потенциалы. Найдите параметры решетки для всех структур и сравните с экспериментальными данными.

¹C.M. Lin et al., Appl. Phys. 109, 033514 (2011)

²McNeil et al., J. Am. Ceram. Soc. 76, 1132 (1993)

Литература

- [1] MedeA version 3.8; MedeA is a registered trademark of Materials Design, Inc., San Diego, USA. – URL: <https://www.materialsdesign.com/> (дата обращения: 03.07.2024)
- [2] *Mohr S.* Linear scaling DFT calculations for large tungsten systems using an optimized local basis / S. Mohr, M. Eixarch, M. Amsler, et al. // *Nuclear Materials and Energy.* – 2018. – V. 15. – P. 64–70.
- [3] *Shao Y.* Advances in methods and algorithms in a modern quantum chemistry program package / Y. Shao, L.F. Molnar, Y. Jung et al. // *Physical Chemistry Chemical Physics.* – 2006. – V. 8. – № 27. – P. 3172–3191.
- [4] *Needs R.J.* Continuum variational and diffusion quantum Monte Carlo calculations / R.J. Needs, M.D. Towler, N.D. Drummond et al. // *Journal of Physics: Condensed Matter.* – 2009. – V. 22. – № 2. – 023201.
- [5] *Coulomb C.A.* Mémoires sur l'électricité et la magnétisme / C.A. Coulomb. – Chez Bachelier, libraire, 1789.
- [6] *Lennard-Jones J.E.* Cohesion / J.E. Lennard-Jones // *Proceedings of the Physical Society.* – 1931. – V. 43. – № 5. – P. 461–482.
- [7] *Buckingham R.A.* The classical equation of state of gaseous helium, neon and argon / R.A. Buckingham // *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences.* – 1938. – V. 168. – № 933. – P. 264–283.
- [8] *Stillinger F.H.* Computer simulation of local order in condensed phases of silicon / F.H. Stillinger, T.A. Weber // *Physical review B.* – 1985. – V. 31. – № 8. – P. 5262–5271.
- [9] *Daw, M.S.* Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals / M.S. Daw, M.I. Baskes // *Physical Review B.* – 1984. – V. 29. – № 12. – P. 6443–6453.
- [10] *Brenner D. W.* Quantum-Based Analytic Interatomic Forces and Materials Simulation / D. W. Brenner, O. A. Shenderova, D. A. Areshkin // *Reviews in Computational Chemistry.* – 1998. – P. 207–239.
- [11] *Brenner D.W.* Relationship between the embedded-atom method and Tersoff potentials / D.W. Brenner // *Physical review letters.* – 1989. – V. 63. – № 9. – P. 1022–1022.
- [12] *Tersoff J.* New empirical approach for the structure and energy of covalent systems / J. Tersoff // *Physical review B.* – 1988. – V. 37. – № 12. – P. 6991–7000.
- [13] *Baskes M.I.* Modified embedded-atom potentials for cubic materials and impurities / M.I. Baskes // *Physical review B.* – 1992. – V. 46. – № 5. – P. 2727–2742.

- [14] *Liang T.* Classical atomistic simulations of surfaces and heterogeneous interfaces with the charge-optimized many body (COMB) potentials / T. Liang, T. Shan, R. Cheng et al. // *Materials Science and Engineering: R: Reports.* – 2013. – V. 74. – № 9. – P. 255–279.
- [15] *Van Duin A.C.* ReaxFF: a reactive force field for hydrocarbons / A. C. Van Duin, S. Dasgupta, F. Lorant et al. // *The Journal of Physical Chemistry A.* – 2001. – V. 105. – № 41. – P. 9396–9409.
- [16] *Becker C.A.* Considerations for choosing and using force fields and interatomic potentials in materials science and engineering / C.A. Becker, F. Tavazza, Z.T. Trautt et al. // *Current Opinion in Solid State and Materials Science.* – 2013. – V. 17. – № 6. – P. 277–283.
- [17] *Hernandez A.* Fast, accurate, and transferable many-body interatomic potentials by symbolic regression / A. Hernandez, A. Balasubramanian, F. Yuan et al. // *NPJ Computational Materials.* – 2019. – V. 5. – № 1. – P. 1–11.
- [18] *Hynninen T.* Limitations of reactive atomistic potentials in describing defect structures in oxides / T. Hynninen, T. Musso, A.S. Foster // *Modelling and Simulation in Materials Science and Engineering.* – 2016. – V. 24. – № 3. – 035022.
- [19] *Chan H.* Machine learning classical interatomic potentials for molecular dynamics from first-principles training data / H. Chan, B. Narayanan, M.J. Cherukara et al. // *The Journal of Physical Chemistry C.* – 2019. – V. 123. – № 12. – P. 6941–6957.
- [20] *Ercolessi F.* Interatomic potentials from first-principles calculations: the force-matching method / F. Ercolessi, J.B. Adams // *Europhysics Letters.* – 1994. – V. 26. – № 8. – P. 583–588.
- [21] *Blank T.B.* Neural network models of potential energy surfaces / T.B. Blank, S.D. Brown, A.W. Calhoun et al. // *The Journal of chemical physics.* – 1995. – V. 103. – № 10. – P. 4129–4137.
- [22] *Plimpton S.J.* Computational aspects of many-body potentials / S.J. Plimpton, A.P. Thompson // *MRS bulletin.* – 2012. – V. 37. – № 5. – P. 513–521.
- [23] *Mueller T.* Machine learning in materials science: Recent progress and emerging applications / T. Mueller, A.G. Kusne, R. Ramprasad // *Reviews in computational chemistry.* – 2016. – V. 29. – P. 186–273.
- [24] *Shapeev A.V.* Moment tensor potentials: A class of systematically improvable interatomic potentials / A.V. Shapeev // *Multiscale Modeling Simulation.* – 2016. – V. 14. – № 3. – P. 1153–1173.
- [25] *Rumelhart D.E.* Learning representations by back-propagating errors / D.E. Rumelhart, G.E. Hinton, R.J. Williams // *Nature.* – 1986. – V. 323. – № 6088. – P. 533–536.
- [26] *Widrow B.* 30 years of adaptive neural networks: perceptron, madaline, and backpropagation / B. Widrow, M.A. Lehr // *Proceedings of the IEEE.* – 1990. – V. 78. – № 9. – P. 1415–1442.
- [27] *Zuo Y.* Performance and cost assessment of machine learning interatomic potentials / Y. Zuo, C. Chen, X. Li et al. // *The Journal of Physical Chemistry A.* – 2020. – V. 124. – № 4. – P. 731–745.

- [28] *Podryabinkin E. V.* Active learning of linearly parametrized interatomic potentials / E.V. Podryabinkin, A. V. Shapeev // Computational Materials Science. – 2017. – V. 140. – P. 171–180.
- [29] *Behler J.* Generalized neural-network representation of high-dimensional potential-energy surfaces / J. Behler, M. Parrinello // Physical review letters. – 2007. – V. 98. – № 14. – 146401.
- [30] *Behler J.* Representing potential energy surfaces by high-dimensional neural network potentials / J. Behler // Journal of Physics: Condensed Matter. – 2014. – V. 26. – № 18. – 183001.
- [31] *Chmiela S.* Machine learning of accurate energy-conserving molecular force fields / S. Chmiela, A. Tkatchenko, H. E. Sauceda et al. // Science advances. – 2017. – V. 3. – № 5. – e1603015.
- [32] *Unke O. T.* PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges / O. T. Unke, M. Meuwly // Journal of chemical theory and computation. – 2019. – V. 15. – № 6. – P. 3678–3693.
- [33] *Dragoni D.* Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron / D. Dragoni, T.D. Daff, G. Csányi et al. // Physical Review Materials. – 2018. – V. 2. – № 1. – 013808.
- [34] *Yao K.* The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics / K. Yao, J.E. Herr, D.W. Toth et al. // Chemical science. – 2018. – V. 9. – № 8. – P. 2261–2269.
- [35] *Artrith N.* High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide / N. Artrith, T. Morawietz, J. Behler // Physical Review B. – 2011. – V. 83. – № 15. – 153101.
- [36] *Huang X.* Ab initio potential energy and dipole moment surfaces for H_5O_2^+ / X. Huang, B.J. Braams, J.M. Bowman // The Journal of chemical physics. – 2005. – V. 122. – № 4. – 044308.
- [37] *Bartók A.P.* On representing chemical environments / A.P. Bartók, R. Kondor, G. Csányi // Physical Review B. – 2013. – V. 87. – № 18. – 184115.
- [38] *Rupp M.* Fast and accurate modeling of molecular atomization energies with machine learning / M. Rupp, A. Tkatchenko, K.R. Müller et al. // Physical review letters. – 2012. – V. 108. – № 5. – 058301.
- [39] *Huan T.D.* A universal strategy for the creation of machine learning-based atomistic force fields / T.D. Huan, R. Batra, J. Chapman et al. // NPJ Computational Materials. – 2017. – V. 3. – № 1. – P. 1–8.
- [40] *Oganov A. R.* How to quantify energy landscapes of solids / A. R. Oganov, M. Valle // The Journal of chemical physics. – 2009. – V. 130. – № 10. – 104504.
- [41] *Jacobsen T.L.* On-the-fly machine learning of atomic potential in density functional theory structure optimization / T.L. Jacobsen, M.S. Jørgensen, B. Hammer // Physical review letters. – 2018. – V. 120. – № 2. – 026102.

- [42] *Seko A.* Group-theoretical high-order rotational invariants for structural representations: Application to linearized machine learning interatomic potential / A. Seko, A. Togo, I. Tanaka // *Physical Review B.* – 2019. – V. 99. – № 21. – 214108.
- [43] *Batra R.* General atomic neighborhood fingerprint for machine learning-based methods / R. Batra, H.D. Tran, C. Kim et al. // *The Journal of Physical Chemistry C.* – 2019. – V. 123. – № 25. – P. 15859-15866.
- [44] *Bartók A. P.* Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons / A.P. Bartók, M.C. Payne, R. Kondor et al. // *Physical review letters.* – 2010. – V. 104. – № 13. – 136403.
- [45] *Thompson A.P.* Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials / A P. Thompson, L.P. Swiler, C.R. Trott et al. // *Journal of Computational Physics.* – 2015. V. 285. – P. 316–330.
- [46] *Wood M.A.* Extending the accuracy of the SNAP interatomic potential form / M.A. Wood, A.P. Thompson // *The Journal of chemical physics.* – 2018. – V. 148. – № 24. – 241721.
- [47] *Gilmer J.* Neural message passing for quantum chemistry: Proceedings of the 34th International Conference on Machine Learning (06–11 Aug, 2017) / J. Gilmer, S.S. Schoenholz, P.F. Riley et al. // – PMLR. – 2017. – P. – 1263–1272.
- [48] *Schütt K.T.* Quantum-chemical insights from deep tensor neural networks / K.T. Schütt, F. Arbabzadah, S. Chmiela et al. // *Nature communications.* – 2017. – V. 8. – № 1. – P. 1–8.
- [49] *Schütt K. T.* Schnet — a deep learning architecture for molecules and materials / K.T. Schütt, H.E. Sauceda, P.J. Kindermans et al. // *The Journal of Chemical Physics.* – 2018. – V. 148. – № 24. – 241722.
- [50] *Lorenz S.* Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks / S. Lorenz, A. Groß, M. Scheffler // *Chemical Physics Letters.* – 2004. – V. 395. – № 4-6. – P. 210–215.
- [51] *Behler J.* Atom-centered symmetry functions for constructing high-dimensional neural network potentials / J. Behler // *The Journal of chemical physics.* – 2011. V. – 134. – № 7. – 074106.
- [52] *Behler J.* First principles neural network potentials for reactive simulations of large molecular and condensed systems / J. Behler // *Angewandte Chemie International Edition.* – 2017. – V. 56. – № 42. – P. 12828–12840.
- [53] *Thompson A.* SNAP: Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials for LAMMPS, talk given at the LAMMPS Users' Workshop and Symposium / A. Thompson // 2015.
- [54] *Powell M.J.* The NEWUOA software for unconstrained optimization without derivatives / M.J. Powell // In *Large-scale nonlinear optimization* / edited by G. Pillo, M. Roma: Springer, Boston, MA. – P. 255–297.
- [55] *Singraber A.* Parallel multistream training of high-dimensional neural network potentials / A. Singraber, T. Morawietz, J. Behler et al. // *Journal of chemical theory and computation.* – 2019. – V. 15. – № 5. – P. 3075–3092.

- [56] *Imbalzano G.* Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials / G. Imbalzano, A. Anelli, D. Giofré et al. // The Journal of chemical physics. – 2018. – V. 148. – № 24. – 241730.
- [57] *Gastegger M.* wACSF—Weighted atom-centered symmetry functions as descriptors in machine learning potentials / M. Gastegger, L. Schwiedrzik, M. Bittermann et al. // The Journal of chemical physics. – 2018. – V. 148. – № 24. – 241709.
- [58] *McClure J.L.* Measurement of the heat of fusion of titanium and a titanium alloy (90Ti-6Al-4V) by a microsecond-resolution transient technique / J.L. McClure, A. Cezairliyan // International journal of thermophysics. – 1992. – V. 13. – № 1. – P. 75–81.
- [59] *Kaschnitz E.* Heat capacity of titanium in the temperature range 1500 to 1900 K measured by a millisecond pulse-heating technique / E. Kaschnitz, P. Reiter // Journal of thermal analysis and calorimetry. – 2001. – V. – 64. – № 1. – P. 351–356.
- [60] *Kaschnitz E.* Enthalpy and temperature of the titanium alpha-beta phase transformation / E. Kaschnitz, P. Reiter // International journal of thermophysics. – 2002. – V. – 23. – № 5. – P. 1339–1345.
- [61] *Cheng C.* Development and application of EAM potentials for Ti, Al and Nb with enhanced planar fault energy of Ti / C. Cheng, Y. Ma, Q. Bao et al. // Computational Materials Science. – 2020. – V. 173. – 109432.
- [62] *Dickel D.* Neural network aided development of a semi-empirical interatomic potential for titanium / D. Dickel, D.K. Francis, C.D. Barrett // Computational Materials Science. – 2020. – V. 171. – 109157.
- [63] *Hao P.D.* Anisotropic elastic and thermodynamic properties of the HCP-Titanium and the FCC-Titanium structure under different pressures / P.D. Hao, P. Chen, L. Deng et al. // Journal of Materials Research and Technology. – 2020. – V. 9. – № 3. – P. 3488–3501.
- [64] *Ledbetter H.* Elastic constants of body-centered-cubic titanium monocrystals / H. Ledbetter, H. Ogi, S. Kai et al. // Journal of applied physics. – 2004. – V. – 95. – № 9. – P. 4642–4644.
- [65] *Fisher E.* The Science, Technology and Application of Titanium / E. Fisher, D. Dever // edited by R. I. Jaffee, N. E. Promisel, 1973. – P. 2739.