

# Особенности выбора параметров для нейронной сети Бехлера—Парринелло на примере кристалла лития

А.Г. Потапов, А.В. Романов

Воронежский государственный университет

С момента публикации в 2011 году статьи Й. Бехлера о симметричных атомно-центрированных функциях для нейронных сетей было предпринято несколько успешных попыток разработать на их основе программы для поиска машинно-обучаемых межатомных потенциалов. Тем не менее, несмотря на очевидный прогресс в этой области, все еще существует проблема выбора параметров к функциям Бехлера—Парринелло для описания конкретных систем. В статье обсуждаются особенности подбора этих параметров на примере кристалла лития. Для тренировки нейросети был выполнен отжиг квантовой молекулярной динамикой с последующим обучением на полученных результатах.

*Ключевые слова:* нейросеть, межатомный потенциал, молекулярная динамика, функции Бехлера-Парринелло.

## 1. Введение

Являясь мощным инструментом теоретической химии, метод классической молекулярной динамики, тем не менее, имеет существенный недостаток. Для его использования необходимо знать межатомный потенциал, конструирование которого представляет известную сложность и требует больших временных затрат. С другой стороны, квантовая молекулярная динамика [1, 2] не имеет таких трудностей, однако необходимость больших вычислительных ресурсов делает этот метод малоприменимым для исследования больших структур.

Долгое время связующим звеном между этими двумя методиками было построение межатомных потенциалов по результатам квантово-химической модели. Однако с развитием аппаратных средств и программного обеспечения для нейронных сетей [3, 4] стало возможным «перекинуть мост» от квантовой молекулярной динамики к классической без значительных сложностей.

Важной вехой во внедрении нейронных сетей в молекулярную динамику стоит считать разработку специальных функций, описывающих ближнее окружение конкретного атома, например, полиномов Чебышева [5, 6] или функций Бехлера—Парринелло [7, 8]. Они позволили перейти от декартовой системы координат к описанию системы, допускающей трансляционную симметрию и, как следствие, неограниченное масштабирование. В последние несколько лет появился ряд программных пакетов, позволяющих обучать нейронные сети по результатам квантово-химической молекулярной динамики с использованием как центральных процессоров, так и связки процессора с графическими сопроцессорами. Одним из таких пакетов является нейронная сеть Fortnet [9].

Тем не менее, неразрешенными остаются сложности с подбором входных параметров для обучения нейронной сети. Выяснилось, что ожидаемый результат может быть получен за приемлемое время при наличии качественных исходных данных, однако эффективность обучения нейросети напрямую зависит от количества нейронов в скрытых слоях, вида и числа симметричных функций, их собственных параметров, а также от радиуса отсечки, сужающего диапазон рассматриваемых атомов в окружении. Недостаточное число нейронов приводит к невозможности точной аппроксимации даже при точно подобранных функциях, в то время как ширина и сдвиг центра гауссиан, угловое разрешение и другие дополнительные параметры, влияющие на эффективность их использования, также нуждаются в вычислении. Следует отметить, что в современных пакетных реализациях нейронных сетей, как правило, предложены собственные способы их расчета, однако полученные формулы на данный момент выводятся эмпирическим путем.

## 2. Теоретическая часть

Перед тем, как использовать функции Бехлера—Парринелло на реальных задачах, необходимо ознакомиться с их формулами и базовыми свойствами. Их проверка на простых тестовых примерах [10] показала наличие важных особенностей, связанных, в том числе, с порядком получаемой величины, которые нужно будет учитывать при дальнейшей работе.

### 2.1 Функции Бехлера—Парринелло

Ниже в (2)–(6) приведено математическое описание каждой из симметричных функций. Формула (1) содержит функцию отсечки, используемую для определения атомов, входящих в атомное окружение заданного атома.

$$f_c(R_{ij}) = \begin{cases} 0,5 \cdot \left[ \cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1 \right] & \text{для } R_{ij} \leq R_c \\ 0 & \text{для } R_{ij} > R_c \end{cases} \quad (1)$$

$$G_i^1 = \sum_j f_c(R_{ij}) \quad (2)$$

$$G_i^2 = \sum_j e^{-\eta(R_{ij}-R_s)^2} \cdot f_c(R_{ij}) \quad (3)$$

$$G_i^3 = \sum_j \cos(\kappa R_{ij}) \cdot f_c(R_{ij}) \quad (4)$$

$$G_i^4 = 2^{1-\zeta} \sum_{j,k \neq i}^{all} (1 + \lambda \cos \theta_{ijk})^\zeta e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}) \quad (5)$$

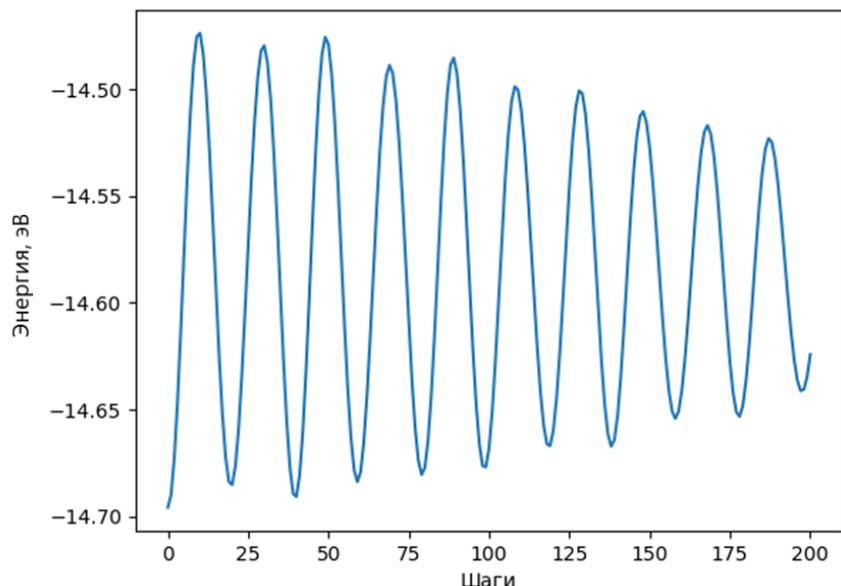
$$G_i^5 = 2^{1-\zeta} \sum_{j,k \neq i}^{all} (1 + \lambda \cos \theta_{ijk})^\zeta \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2)} \cdot f_c(R_{ij}) \cdot f_c(R_{ik}) \quad (6)$$

Здесь  $\eta$  – ширина гауссиан, регулирующая эффективное радиальное расширение функций симметрии;  $R_s$  – сдвиг центра гауссиан на определенный радиус;  $\kappa$  – параметр, определяющий период функции косинуса;  $\theta$  – угол между всеми комбинациями из трех атомов, содержащих заданный атом;  $\lambda$  – параметр, отвечающий за ориентацию графика  $G(\theta)$ ;  $\zeta$  – параметр, регулирующий угловое разрешение, то есть плавность изменения значения функции  $G$ .

Функции  $G^1$ ,  $G^2$ ,  $G^3$  относятся к группе радиальных и учитывают только взаимные меж-атомные расстояния. Функции  $G^4$ ,  $G^5$  являются угловыми и учитывают более сложные изменения атомных конфигураций. Поскольку функции внутри одной группы обладают довольно схожими свойствами, в сети Fortnet по умолчанию используются только две из них, относящихся к разным категориям:  $G^2$  и  $G^5$ .

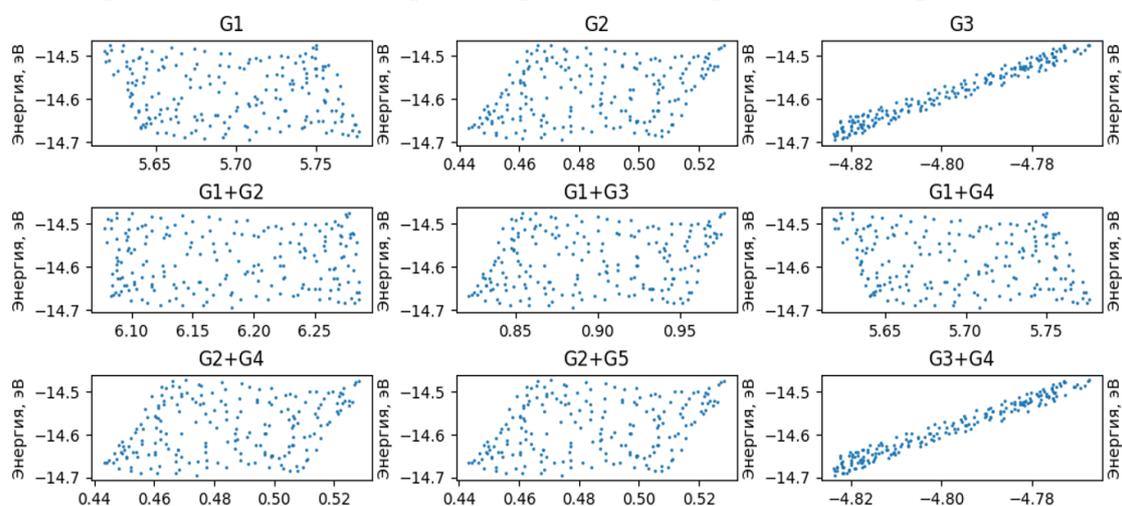
### 2.2 Необходимость нормализации

Для предварительного анализа свойств симметричных функций было осуществлено моделирование химической системы, состоящей из 4 атомов меди, расположенных в ГЦК-решетке, методом молекулярной динамики. Атомы были нагреты до 733 К. Предполагалось, что относительно высокое число степеней свободы в движущейся системе позволит наглядно продемонстрировать влияние не только радиальных, но и угловых функций на возможность аппроксимации ее энергии, график изменения которой с течением времени представлен на рис. 1.



**Рис. 1.** Колебания энергии в ходе эксперимента с  $\text{Cu}_4$

Для тестирования зависимости энергии от функций Бехлера—Парринелло были рассчитаны все возможные их линейные комбинации с единичными параметрами. Поскольку в данной системе невозможно выделить единственный неподвижный атом, относительно которого было бы удобнее всего рассматривать изменение всей конфигурации, было решено использовать сумму значений функций для каждого из 4 атомов. Некоторые из полученных графиков зависимости энергии системы от симметричных функций  $E(G)$  представлены на рис. 2.



**Рис. 2.** Графики  $E(G)$  для некоторых возможных линейных комбинаций  $G^1$ – $G^5$

На рисунке видно, что некоторые из полученных графиков почти полностью совпадают, из чего был сделан вывод о наличии внутри симметричных функций определенной иерархии порядка: модули значений некоторых из них существенно превышают модули значений других ( $|G^1| > |G^3| > |G^2| \gg |G^5| > |G^4|$ ), что приводит к несбалансированному их влиянию на конечный результат. Следовательно, необходимо использовать нормализацию при дальнейшей работе для компенсации данного эффекта.

Кроме того, наименьший разброс демонстрируют  $G^3$ -подобные графики, что может быть связано с наличием в ее структуре формулы косинуса, хорошо описывающей затухающие периодические колебания энергии.

### 3. Практическая часть

В ходе исследования было проведено моделирование кристалла лития с ячейкой на 27 атомов в рамках квантовой молекулярной динамики. С помощью калькулятора ASE[11] / GPAW [12–13] были рассчитаны траектории движения атомов моделируемых систем, а также получены их потенциальная и полная энергии в каждый момент времени. На основании полученных данных было проведено обучение нейросети Fortnet с различными гиперпараметрами, в числе которых количество радиальных и угловых симметричных функций, число нейронов в скрытых слоях и радиус отсечки. Кроме того, были сопоставлены полученные значения  $G$  и полная энергия  $E(G)$ .

#### 3.1 Величина датасета

В первом эксперименте было взято 27 атомов лития, был выполнен молекулярно-динамический отжиг с использованием динамики Langevin. В качестве начального распределения атомов по скоростям было выбрано распределение Максвелла-Больцмана при температуре в 300 К. Сходимость по температуре достигнута спустя 35000 шагов.

Полный объем полученных данных оказался не слишком удобным для первоначального тестирования влияния входных параметров модели Fortnet по причине высокой ресурсоемкости, в связи с чем были взяты лишь 372 точки расчета с шагом в 64 точки (для охвата значительной доли области анализа). Было использовано два скрытых слоя BPNN с 2 нейронами в каждом из них, 10 радиальных и 10 угловых симметричных функций. Радиус отсечки составил 20 Å (достаточно, чтобы захватить все 27 атомов). Обучение происходило в течение 5000 итераций. Сравнение исходной (DFT) и предсказанной (NN) энергий представлено на рис. 3.

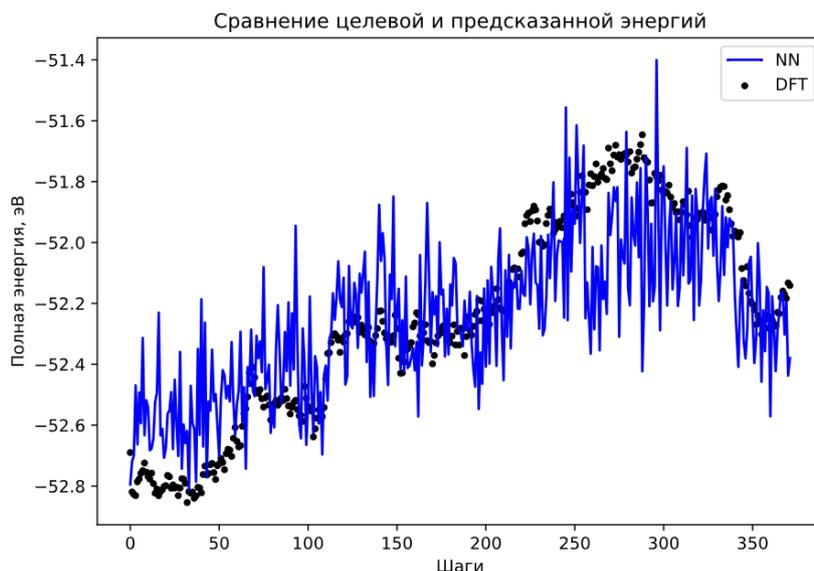
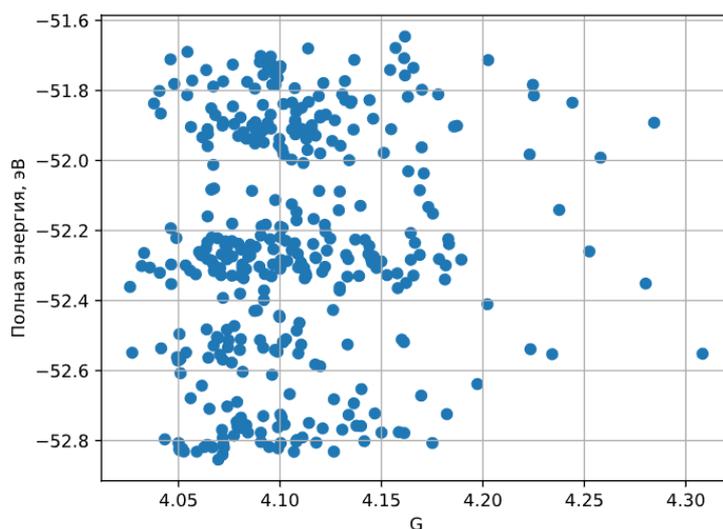


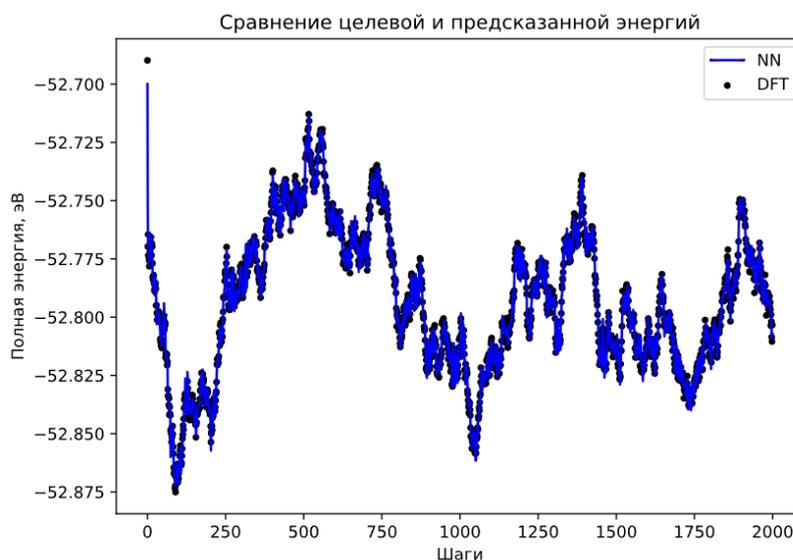
Рис. 3. Результаты работы сети Fortnet для 372 точек и 20 симметричных функций

Как видно, точность предсказания полученной модели крайне низка, она описывает лишь общую тенденцию изменения энергии и демонстрирует слишком большой процент выбросов. На рис. 4 представлен график изменения энергии в зависимости от суммы полученных значений функций Бехлера—Парринелло. Видно, что одним и тем же значениям  $G$  соответствует слишком широкий спектр значений  $E$ , что может свидетельствовать о том, что данный набор входных параметров не подходит для описания заданной системы.



**Рис. 4.** График  $E(G)$  для 372 точек и 20 симметричных функций

На втором этапе тестирования было взято большее количество шагов для предыдущего расчета, были внесены изменения в способ формирования входных данных: вместо охвата всего временного отрезка были взяты лишь первые 2000 значений подряд, что позволило работать с данными, скорость изменения которых приближена к реальной. Было существенно увеличено число нейронов в скрытых слоях: по 27 в каждом из двух слоев BPNN. Кроме того, результаты предыдущего эксперимента дали основание полагать, что число симметричных функций также нуждается в увеличении, поэтому для новой модели было взято 25 радиальных и 25 угловых функций. Для обучения оказалось достаточно 4283 итерации, после чего средняя ошибка MSE составила порядка  $5,95 \cdot 10^{-6}$ , что меньше целевого значения. Время обучения на 12 ядрах процессора Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz составило 12,2 минут. Итоговый график сравнения исходной и предсказанной энергий можно видеть на рис. 5.



**Рис. 5.** Результаты работы сети Fortnet для 2000 точек и 50 симметричных функций

На графике видно, что первая точка сильно выделяется из общего диапазона. Она связана с начальным приближением и не имеет значения для последующих расчетов. Кроме того, особое внимание стоит уделить графику  $E(G)$ , изображенному на рис. 6. Видно, что для одного и того же значения суммы значений всех 50 функций Бехлера—Парринелло относительно нулевого атома степень изменчивости энергии заметно меньше, чем на рис. 4. Тем не менее, она все еще гораздо больше, чем можно наблюдать на рис. 5, что было бы невозможно при использовании

сеть Fortnet линейной комбинации значений  $G$  для аппроксимации энергии. Следовательно, полученные 50 значений следует рассматривать как отдельные переменные, которые нельзя информативно изобразить на двумерном графике.

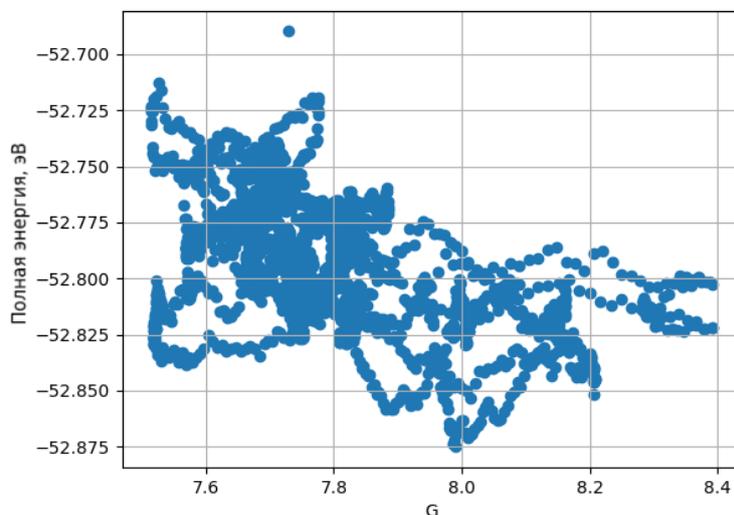


Рис. 6. График  $E(G)$  для 2000 точек и 50 симметричных функций

### 3.2 Влияние радиуса отсечки

Изменение задаваемого радиуса отсечки при сохранении величины датасета, числа нейронов в скрытых слоях и самих симметричных функций способно оказать видимое влияние на результат. С одной стороны, его уменьшение способно упростить формат входных данных и ускорить обучение (450 сек. для  $R_c = 8 \text{ \AA}$  против 715 сек. для  $R_c = 20 \text{ \AA}$  для 12 нейронов в каждом из 2 скрытых слоев). Однако при детальном рассмотрении можно заметить появление лишних всплесков на графике предсказанной энергии, что свидетельствует о недостаточной точности обучения (ср. рис. 5 и рис. 7).

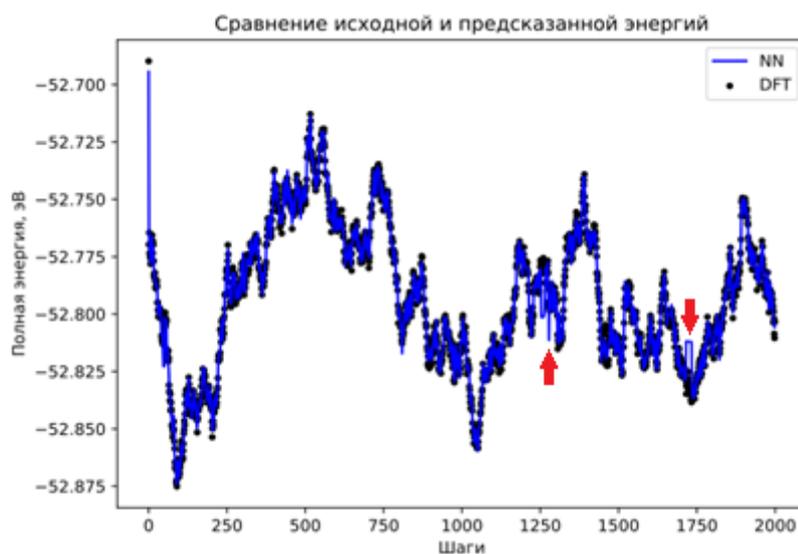


Рис. 7. Результаты работы сети Fortnet для 27 и 27 нейронов, 50 функций,  $R_c = 8 \text{ \AA}$

### 3.3 Вид симметричных функций

Ручное задание разнообразных сочетаний используемых симметричных функций при их общем неизменном числе, равном 24, и радиусе отсечки  $R_c = 8 \text{ \AA}$  позволило сделать некоторые выводы касательно их собственной эффективности. Были рассмотрены комбинации  $12G^1+12G^4$ ,

$12G^1+12G^5$ ,  $12G^2+12G^4$ ,  $12G^2+12G^5$ ,  $6G^1+6G^2+12G^4$ ,  $6G^1+6G^2+12G^5$ ,  $12G^1+6G^4+6G^5$ ,  $12G^2+6G^4+6G^5$ ,  $6G^1+6G^2+6G^4+6G^5$ . Наименее удачной стала комбинация  $12G^1+12G^4$  (см. рис. 8). Наиболее эффективной — комбинация  $12G^2+6G^4+6G^5$  (см. рис. 9).

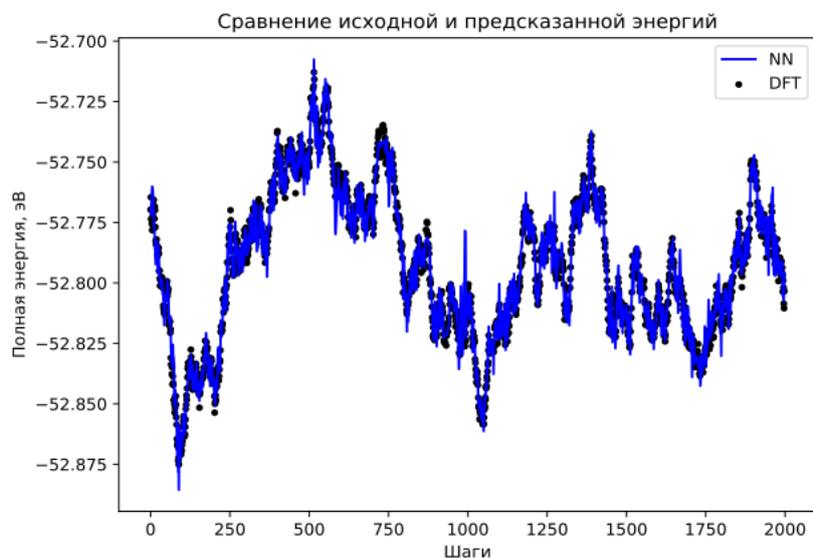


Рис. 8. Результаты работы сети Fortnet для  $12G^1+12G^4$ ,  $R_c = 8 \text{ \AA}$

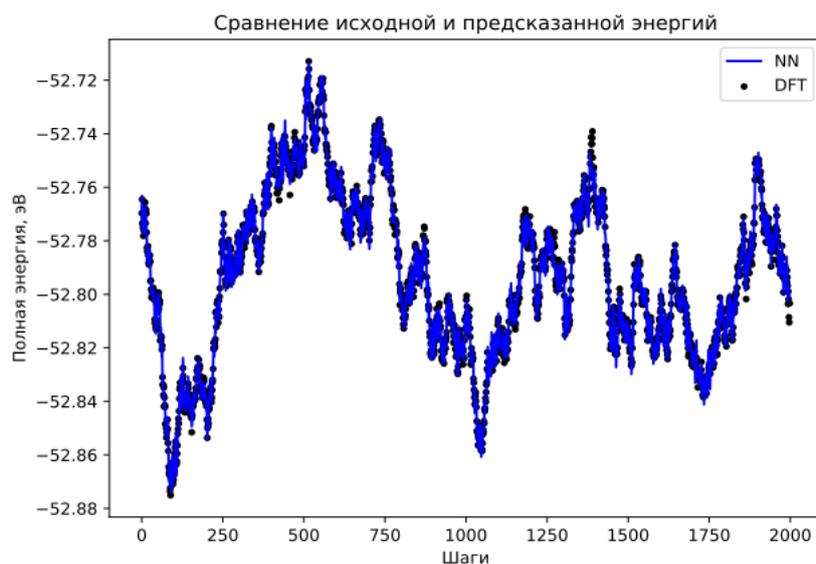


Рис. 9. Результаты работы сети Fortnet для  $12G^2+6G^4+6G^5$ ,  $R_c = 8 \text{ \AA}$

### 3.4 Число симметричных функций

Помимо описанных ранее моделей были опробованы другие варианты изменения гиперпараметров с целью выявления наиболее эффективных сочетаний. В частности, уменьшение числа симметричных функций с 50 до 12 привело, с одной стороны, к значительному замедлению обучения (потребовалось 16507 итераций и 32,9 минут), но, с другой стороны, несколько снизило процент шумовых осцилляций, что могло положительно сказаться на обобщающей способности модели. Результирующий график можно наблюдать на рис. 10.

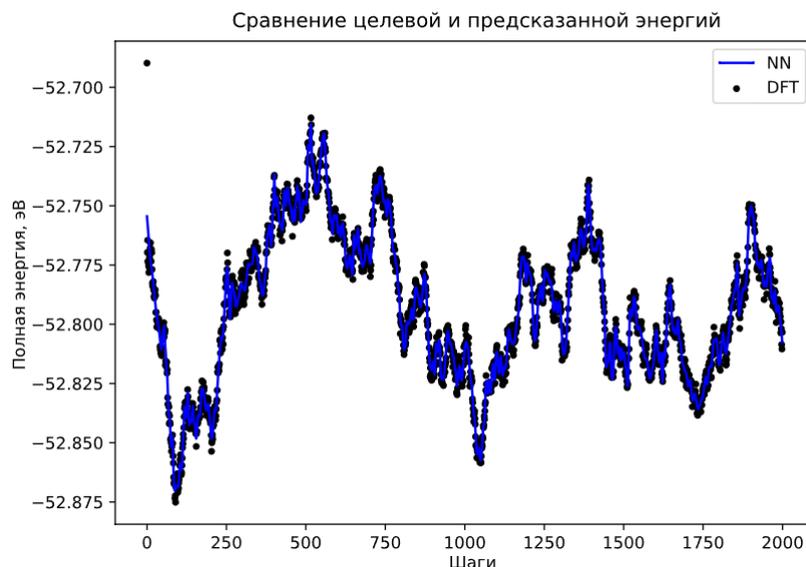


Рис. 10. Результаты работы сети Fortnet для 2000 точек и 12 симметричных функций

Кроме того, попытка уменьшить число нейронов до 13 в каждом из двух скрытых слоев привела к чуть менее точным результатам ( $MSE = 1,43 \cdot 10^{-5}$ ) при практически неизменной скорости обучения (около 11 минут).

#### 4. Заключение

На основе датасета для кристалла лития, полученного методом квантовой молекулярной динамики, показано, что с помощью нейронной сети возможно аппроксимировать значение полной энергии системы в каждый момент времени, основываясь только на данных о взаимных межатомных расстояниях и валентных углах. Переводя декартовы координаты атомов в базисный набор симметричных функций Бехлера—Парринелло, программа учитывает трансляционную и вращательную симметрию получающихся атомных конфигураций, что делает возможным использование результатов обучения нейросетевой модели в классической молекулярной динамике.

Тем не менее, существует набор параметров, оказывающих непосредственное влияние на степень совпадения исходной и предсказанной энергий. В частности, большое значение имеет выбор качественных и достаточно полных исходных данных. Чем более детальным и плавно меняющимся является массив расстояний, углов и энергий, тем более корректными получаются результирующие графики.

Еще одним важным фактором является число нейронов в скрытых слоях нейронной сети. Близкое к числу атомов число нейронов в проведенных экспериментах являлось минимально необходимым, поскольку с его уменьшением точность обучения начинала неизменно падать, а время обучения – расти.

Наконец, количество выбираемых радиальных и угловых симметричных функций напрямую зависит от самой рассматриваемой системы. Чем больше это число, тем более многомерным становится признаковое пространство, подаваемое на вход нейронной сети, задача которой заключается в установлении функциональной зависимости между ним и одномерным целевым вектором энергий. По этой причине становится крайне затруднительно (а при большом числе функций невозможно) оценить возможность нахождения такой связи, исходя из графического отображения зависимости  $E(G)$ . Увеличивается число параметров нейронной сети, а вместе с ним и вероятность переобучения.

В то же время работа с относительно небольшим числом симметричных функций может быть более эффективной в плане перспективного применения нейросети для оценки новых данных, но занимать больше времени, поскольку гибкость процесса обучения, связанная с быстрой и масштабной коррекцией весовых коэффициентов модели, снижается. Так или иначе,

получаемые модели способны выдать корректные результаты для самых разных входных параметров, причем лучше всего показало себя сочетание функций  $G^2+G^4+G^5$ .

В дальнейшем планируется расширить набор исследуемых систем, чтобы выявить новые закономерности в использовании различных параметров нейронной сети и функций Бехлер—Парринелло.

## Литература

1. Marx D., Hutter J. *Ab initio Molecular Dynamics: Basic Theory and Advanced Methods* // Cambridge University Press. 2009. Vol. 63, no. 3. 567 p. DOI: 10.1017/CBO9780511609633.
2. Kirchner B., di Dio P.J., Hutter J. *Real-World Predictions from Ab Initio Molecular Dynamics Simulations* // *Topics in Current Chemistry*. 2011. Vol. 307. P. 109–153. DOI: 10.1007/128\_2011\_195.
3. Kyuhyun L., Dongsun Y., Wonseok J. et al. *SIMPLE-NN: An efficient package for training and executing neural-network interatomic potentials* // *Computer Physics Communications*. 2019. Vol. 242. P. 95–103. DOI: 10.1016/j.cpc.2019.04.014.
4. Blank T.B., Brown S.D., Calhoun A.W. et al. *Neural network models of potential energy surfaces* // *The Journal of Chemical Physics*. 1995. Vol. 103, no. 10. P. 4129–4137. DOI: 10.1063/1.469597.
5. Lindsey R.K., Fried L.E., Goldman N. *CHIMES: A force matched potential with explicit three body interactions for molten carbon* // *Journal of Chemical Theory and Computation*. 2017. Vol. 13, no. 12. 8 p. DOI: 10.1021/acs.jctc.7b00867.
6. Lindsey R.K., Fried L.E., Goldman N. et al. *Active learning for robust, high-complexity reactive atomistic simulations* // *The Journal of Chemical Physics*. 2020. Vol. 153, no. 13. 16 p. DOI: 10.1063/5.0021965.
7. Behler J. *Constructing High-Dimensional Neural Network Potentials: A Tutorial Review* // *International Journal of Quantum Chemistry*. 2015. Vol. 115, no. 16. 19 p. DOI: 10.1002/qua.24890.
8. Behler J. *Atom-centered symmetry functions for constructing high-dimensional neural network potentials* // *The Journal of Chemical Physics*. 2011. Vol. 134, no. 3. 14 p. DOI: 10.1063/1.3553717.
9. van der Heide T., Kullgren J., Broqvist P. et al. *Fortnet, a software package for training Behler Parrinello neural networks* // *Computer Physics Communications*. 2023. Vol. 284. 13 p. DOI: 10.5281/zenodo.5969907.
10. Потапов А.Г., Романов А.В. Моделирование межатомного потенциала по результатам квантово-химического расчета с использованием симметричных функций Бехлера-Парринелло // *Труды молодых ученых факультета компьютерных наук ВГУ*. 2024. № 4. С. 215–222.
11. Larsen A.H., Mortensen J.J., Blomqvist J. et al. *The Atomic Simulation Environment – A Python library for working with atoms* // *J. Phys.: Condens.* 2017. Vol. 29, no. 27. 31 p. DOI: 10.1088/1361-648X/aa680e.
12. Mortensen J.J., Larsen A.H., Kuisma M. et al. *GPAW: An open Python package for electronic structure calculations featured* // *J. Chem. Phys.* 2024. Vol. 160, no. 9. 42 p. DOI: 10.1063/5.0182685.
13. Mortensen J.J., Hansen L.B., Jacobsen K.W. *Real-space grid implementation of the projector augmented wave method* // *Phys. Rev. B*. 2005. Vol. 71, no. 3. 11 p. DOI: 10.1103/PhysRevB.71.035109.