

# Семинар НУГ «Машинное обучение в атомистическом моделировании»

**Тема:** Уменьшение размера обучающей выборки с помощью малоранговой аппроксимации Moment Tensor Potential без потери точности

**Докладчик:** Иван Новиков, к.ф.-м.н., доцент факультета компьютерных наук НИУ ВШЭ

Москва, 2026

# Moment Tensor Potential

Потенциальная энергия  $E^{\text{MTP}}$  является суммой вкладов  $V(\mathbf{n}_i)$  атомных окружений  $\mathbf{n}_i$ ,  $i = \overline{1, n}$ :

$$E^{\text{MTP}} = \sum_{i=1}^n V(\mathbf{n}_i) = \sum_{i=1}^n \sum_{\alpha_i} \xi_{\alpha} B_{\alpha}(\mathbf{n}_i) \quad (*)$$

$$\mathbf{n}_i = \{r_{ij}, z_i, z_j : j = 1, \dots, N_{nb}^i\}$$

$B_{\alpha}$  - некоторые свёртки Moment Tensor Descriptors:

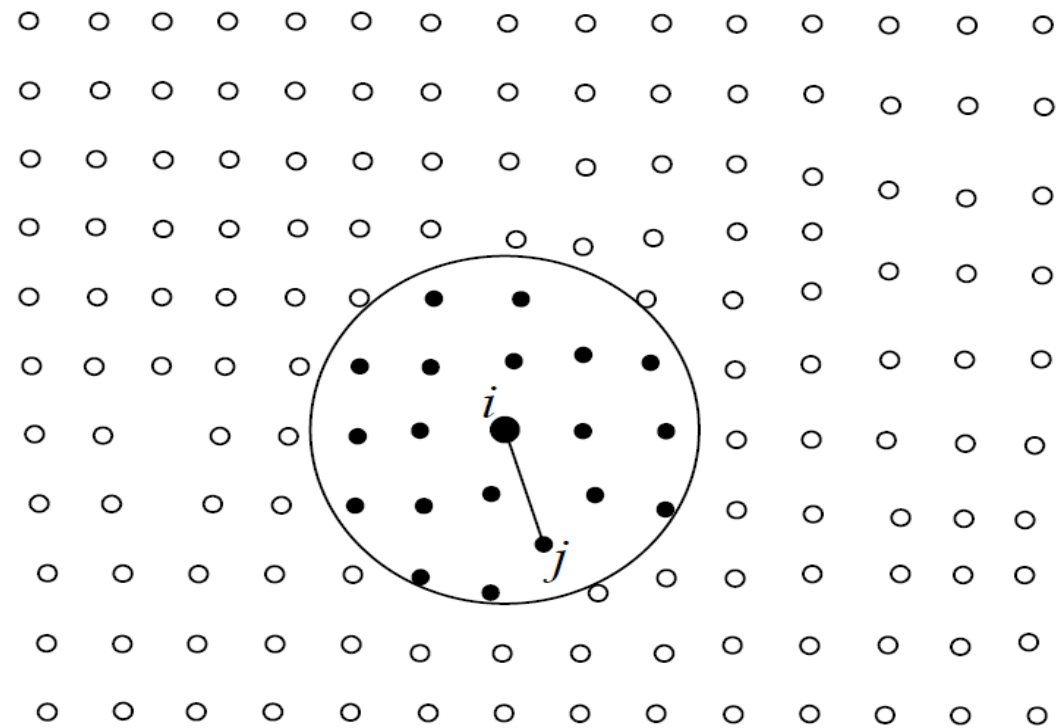
$$M_{\mu, \nu}(\mathbf{n}_i) = \sum_j f_{\mu}(|r_{ij}|, z_i, z_j) \underbrace{\vec{r}_{ij} \otimes \dots \otimes \vec{r}_{ij}}_{\nu \text{ times}}$$

$$f_{\mu}(|r_{ij}|, z_i, z_j) = \sum_{\beta} c_{\mu, z_i, z_j}^{(\beta)} Q_{\beta}(|r_{ij}|) (R_{cut} - |r_{ij}|)^2$$

Введем:  $\text{lev } M_{\mu, \nu} = 2 + 4\mu + \nu$  и  $\text{lev } \prod_{p=1}^P M_{\mu_p, \nu_p} = \sum_{p=1}^P (2 + 4\mu_p + \nu_p)$ .

Для определения формы МТР мы выбираем некоторый  $\text{lev}_{\max}$  и включаем в (\*) любую  $B_{\alpha}$  с  $\text{lev } B_{\alpha} \leq \text{lev}_{\max}$ . Далее обозначаем  $\boldsymbol{\theta} = \{\boldsymbol{\xi}, \mathbf{c}\}$  и  $E^{\text{MTP}} = E^{\text{MTP}}(\boldsymbol{\theta}; \mathbf{R}, \mathbf{Z}, \mathbf{L}) = E^{\text{MTP}}(\boldsymbol{\theta})$ .

Число радиальных параметров:  $N_T^2 N_{\mu} N_Q$



# Tensor-factorized MTP (TFMTP)

Разложение тензорного произведения:

$$c_{z_i, z_j, \mu}^{(\beta)} = \sum_{\alpha_1=1}^{r_1} \sum_{\alpha_2=1}^{r_2} \sum_{\alpha_3=1}^{r_3} G_{1, z_i, \alpha_1}^{(1)} G_{\alpha_1, z_j, \alpha_2}^{(2)} G_{\alpha_2, \mu, \alpha_3}^{(3)} G_{\alpha_3, \beta, 1}^{(4)}$$

Число радиальных параметров:  $N_T r_1 + r_1 N_T r_2 + r_2 N_\mu r_3 + r_3 N_\beta$

Остальная архитектура остается такой же как и у MTP!

В случае с TFMTP число параметров не растет квадратично с числом типов атомов!

## Обучение МТР

Пусть  $K$  – число конфигураций в обучающей выборке с рассчитанными энергиями, силами и напряжениями. Обучение заключается в нахождении параметров  $\theta$ , минимизирующих целевую функцию:

$$\sum_{k=1}^K \left[ w_e \left( E_k^{\text{QM}} - E_k^{\text{MTP}}(\theta) \right)^2 + w_f \sum_{i=1}^n \left| f_{i,k}^{\text{QM}} - f_{i,k}^{\text{MTP}}(\theta) \right|^2 + w_s \sum_{i=1}^6 \left( \sigma_{i,k}^{\text{QM}} - \sigma_{i,k}^{\text{MTP}}(\theta) \right)^2 \right] \rightarrow \min,$$

где  $w_e$ ,  $w_f$  и  $w_s$  - неотрицательные веса. В большинстве случаев  $w_e = 1$ ,  $w_f = 0.01$  и  $w_s = 0.001$ .

Алгоритм оптимизации МТР и TFMTR не меняется!

## Активное обучение с помощью алгоритма MaxVol: активное множество

Обозначим энергию МТР за  $E = E(\bar{\boldsymbol{\theta}}, \text{cfg})$ . Под активным множеством мы понимаем подмножество из  $m$  конфигураций в обучающей выборке, которое дает максимальный детерминант (объем) у матрицы:

$$A = \begin{bmatrix} \frac{\partial E}{\partial \theta_1}(\bar{\boldsymbol{\theta}}, \text{cfg}_1) & \cdots & \frac{\partial E}{\partial \theta_m}(\bar{\boldsymbol{\theta}}, \text{cfg}_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial \theta_1}(\bar{\boldsymbol{\theta}}, \text{cfg}_m) & \cdots & \frac{\partial E}{\partial \theta_m}(\bar{\boldsymbol{\theta}}, \text{cfg}_m) \end{bmatrix}$$

Если  $K > m$ , то для нахождения активного множества мы используем алгоритм MaxVol. Если  $K \leq m$ , мы включаем некоторые (или все)  $K$  в активное множество (его также определяем с помощью MaxVol) и заполняем остальные строки единицами на диагонали и нулями вне диагонали.

## Степень экстраполяции

Предположим, что проходит какое-либо атомистическое моделирование. Для того, чтобы определить, рассматривать ли текущую конфигурацию  $\text{cfg}^*$  как кандидата на включение в обучающую выборку, мы вводим так называемую степень экстраполяции:

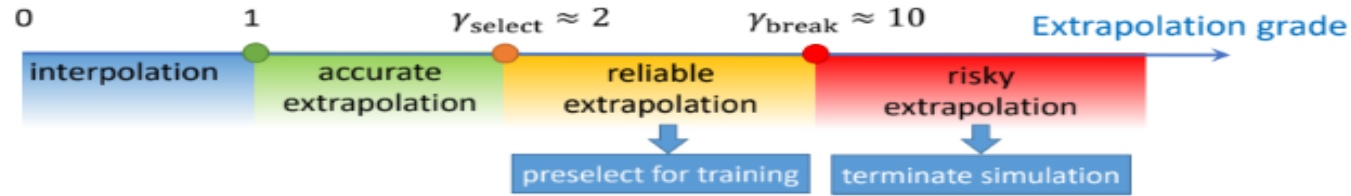
$$\gamma(\text{cfg}^*) = \max_{1 \leq j \leq m} (|c_j|), \text{ где}$$

$$\mathbf{c} = \left( \frac{\partial E}{\partial \theta_1}(\bar{\boldsymbol{\theta}}, \text{cfg}^*) \dots \frac{\partial E}{\partial \theta_m}(\bar{\boldsymbol{\theta}}, \text{cfg}^*) \right) A^{-1}.$$

Степень экстраполяции – это множитель, на который будет увеличен  $\det(A)$  в случае, если конфигурация  $\text{cfg}^*$  будет добавлена в активное множество, т.е., если  $\gamma(\text{cfg}^*) > 1$ , то  $\det(A)$  может быть увеличен. Мы говорим, что МТР экстраполирует, если  $\gamma(\text{cfg}^*) > 1$  и интерполирует в противном случае.

## Двухпороговая схема активного обучения

Для построения так называемого предотобранного множества конфигураций, мы задаем два порога:  $\gamma_{\text{select}} \approx 2$  and  $\gamma_{\text{break}} \approx 10$ .



Если  $\gamma_{\text{select}} \leq \gamma(\text{cfg}^*) \leq \gamma_{\text{break}}$ , то мы добавляем конфигурацию в предотобранное множество и продолжаем моделирование. Если  $\gamma(\text{cfg}^*) > \gamma_{\text{break}}$ , то мы прерываем моделирование, строим матрицу  $P \times m$ , где  $P$  - число предотобранных конфигураций:

$$B = \begin{bmatrix} \frac{\partial E}{\partial \theta_1}(\bar{\theta}, \text{cfg}_1) & \cdots & \frac{\partial E}{\partial \theta_m}(\bar{\theta}, \text{cfg}_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial \theta_1}(\bar{\theta}, \text{cfg}_P) & \cdots & \frac{\partial E}{\partial \theta_m}(\bar{\theta}, \text{cfg}_P) \end{bmatrix}$$

## Обновление активного множества и обучающей выборки

С помощью алгоритма MaxVol мы отбираем конфигурации, который делают максимальным детерминант матрицы  $A$  (or,  $\det(A)$ ). Т. е., мы выбираем  $S \leq m$  конфигураций из предотобранного множества, или, мы находим подматрицу максимального объема в матрице:

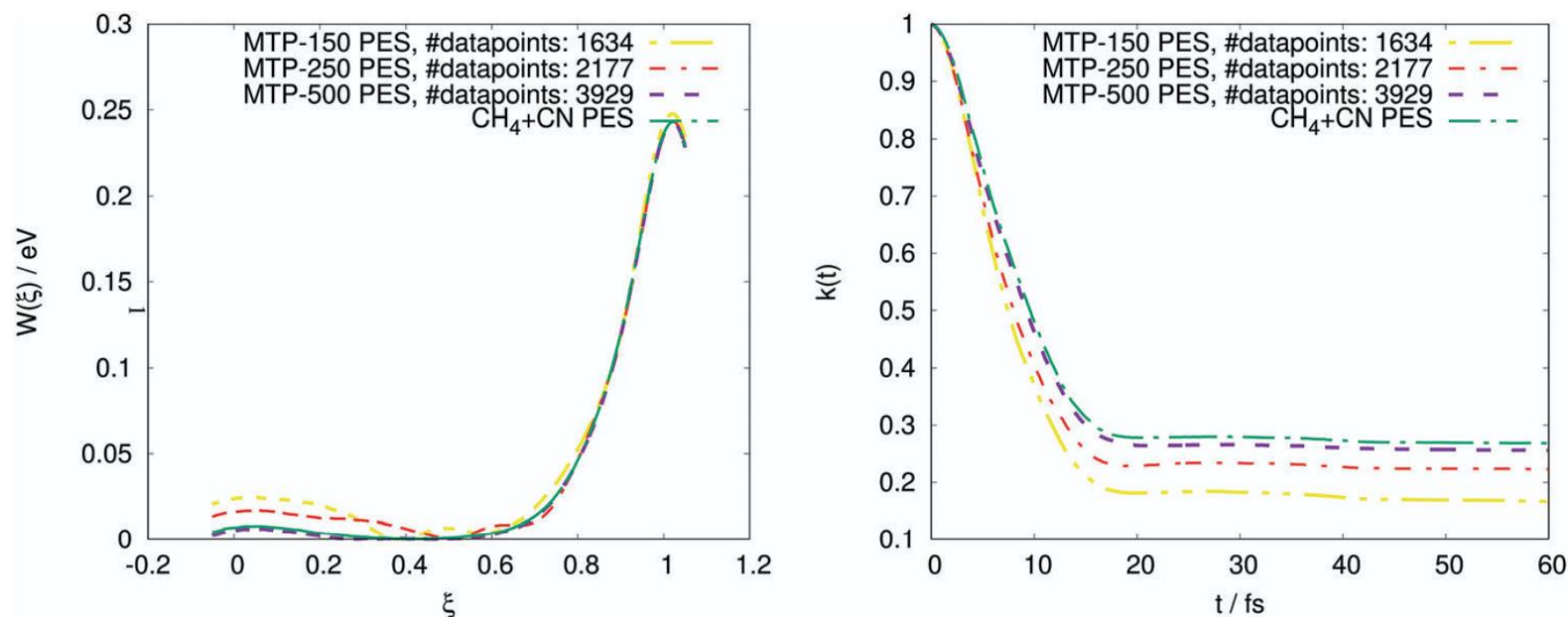
$$\tilde{B} = \begin{bmatrix} \frac{\partial E}{\partial \theta_1}(\bar{\theta}, \text{cfg}_1) & \cdots & \frac{\partial E}{\partial \theta_m}(\bar{\theta}, \text{cfg}_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial \theta_1}(\bar{\theta}, \text{cfg}_{P+m}) & \cdots & \frac{\partial E}{\partial \theta_m}(\bar{\theta}, \text{cfg}_{P+m}) \end{bmatrix}$$

где первые  $m$  строк из матрицы  $A$  на предыдущем шаге активного обучения. Далее мы проводим DFT вычисления для получения энергий, сил и напряжений этих конфигураций и добавляем их в обучающую выборку.

Мы также обновляем активное множество путем добавления (замены)  $S$  строк в матрице  $A$ . Наконец, мы дообучаем МТР на обновленной обучающей выборке.

# Основная идея презентации

Важным гиперпараметром алгоритма активного обучения является число параметров потенциала  $m$ . В работе [I. S. Novikov, Y. V. Suleimanov, and A. V. Shapeev, “Automated calculation of thermal rate coefficients using ring polymer molecular dynamics and machine-learning interatomic potentials with active learning,” Physical Chemistry Chemical Physics 20, 29503–29512 (2018)] было установлено, что число параметров и точность потенциала растет с числом его параметров.



Однако в той работе менялось не просто число параметров потенциала, но и уровень.

В настоящей работе мы фиксируем уровень потенциала и уменьшаем только радиальное число параметров, а далее активно обучаем оригинальный и «сжатый» МТР и сравниваем их точность и число требуемых данных.

## Результаты для системы FLiNaK (1)

Обучение ансамбля из пяти MTP и TFMTP на модели MatterSim во время молекулярно-динамического моделирования в NVT-ансамбле при 600 K с помощью алгоритма активного обучения основанного на MaxVol. Стартовые обучающие выборки для всех потенциалов включали 20 конфигураций, полученных с помощью МД моделирования на MatterSim.

| Model | # parameters | # selected configurations |
|-------|--------------|---------------------------|
| TFMTP | 360          | $927 \pm 20$              |
| MTP   | 568          | $1608 \pm 36$             |

Число параметров моделей и число отобранных конфигураций.

| Model | loss            | energy error<br>meV/atom | force error<br>meV/Å | stress error<br>eV |
|-------|-----------------|--------------------------|----------------------|--------------------|
| TFMTP | $2.01 \pm 0.15$ | $0.63 \pm 0.04$          | $26.8 \pm 0.8$       | $0.232 \pm 0.008$  |
| MTP   | $1.78 \pm 0.14$ | $0.58 \pm 0.06$          | $25.2 \pm 0.9$       | $0.227 \pm 0.008$  |

Ошибки по энергиям, силам и напряжениям на базе данных, полученной на MTP.

## Результаты для системы FLiNaK (2)

Обучение потенциалов MTP и TFMTP на данных DFT (код QuantumEspresso) во время молекулярно-динамического моделирования в NVT-ансамбле при 1200 K с помощью алгоритма активного обучения основанного на MaxVol. Стартовые обучающие выборки для всех потенциалов включали 20 конфигураций, полученных с помощью МД моделирования на Quantum Espresso.

| Model | # parameters | # selected configurations |
|-------|--------------|---------------------------|
| TFMTP | 400          | 844                       |
| MTP   | 608          | 1735                      |

Число параметров моделей и число отобранных конфигураций.

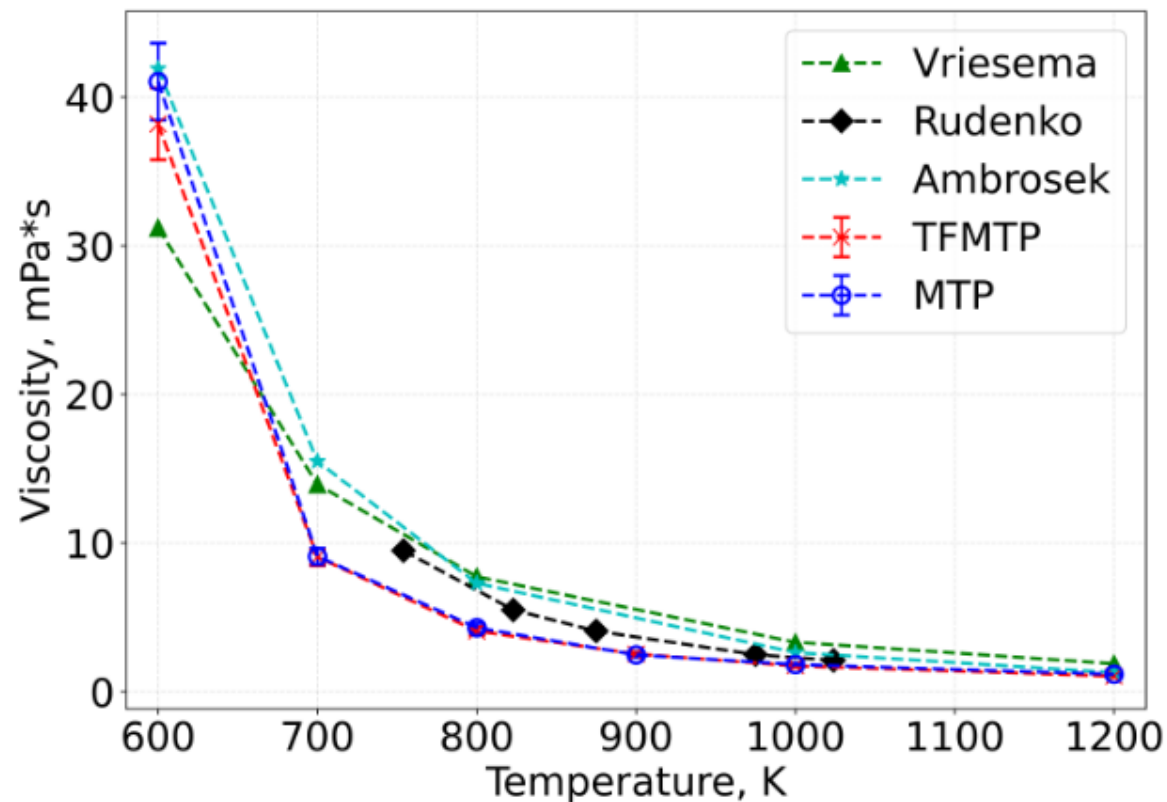
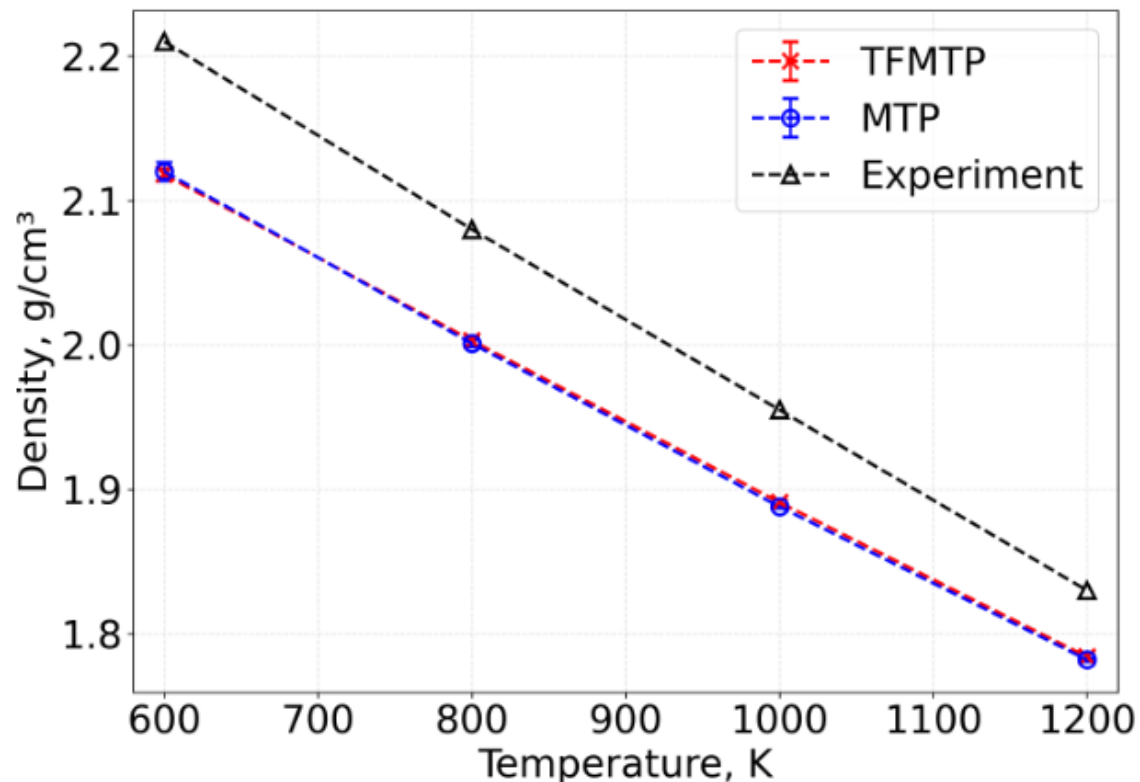
| Model | loss  | energy error<br>meV/atom | force error<br>meV/Å | stress error<br>eV |
|-------|-------|--------------------------|----------------------|--------------------|
| TFMTP | 0.321 | 1.04                     | 32.5                 | 0.324              |
| MTP   | 0.292 | 0.97                     | 31.0                 | 0.306              |

Ошибки по энергиям, силам и напряжениям на базе данных, полученной на MTP.

Вывод: имея в 1.5 раза меньше параметров, TFMTP требует в 2 раза меньше данных!

## Результаты для системы FLiNaK (3)

Расчет физико-химических свойств в диапазоне температур 600-1200 К.



Вывод: TFMTP и MTP дали идентичные показатели плотности и вязкости!

$$\eta = \frac{V}{k_B T} \int_0^{\infty} \langle P_{\alpha\beta}(t) P_{\alpha\beta}(0) \rangle dt$$

## Результаты для эквиатомного сплава MoNbTaWV (1)

Обучение потенциалов MTP и TFMTP на данных DFT (код Quantum Espresso) во время релаксации структур с помощью алгоритма активного обучения основанного на MaxVol. Начальные обучающие выборки включали 110 конфигураций: 10 отрелаксированных конфигураций и 100 конфигураций с возмущенными атомами.

| Model | # parameters | # selected configurations |
|-------|--------------|---------------------------|
| TFMTP | 858          | 696                       |
| MTP   | 1293         | 1541                      |

Число параметров моделей и число отобранных конфигураций.

| Model | loss  | energy error<br>meV/atom | force error<br>meV/Å | stress error<br>eV |
|-------|-------|--------------------------|----------------------|--------------------|
| TFMTP | 0.183 | 0.192                    | 25.9                 | 0.245              |
| MTP   | 0.190 | 0.203                    | 26.5                 | 0.238              |

Ошибки по энергиям, силам и напряжениям на базе данных, полученной на MTP.

Вывод: независимо от атомистической системы, имея в 1.5 раза меньше параметров, TFMTP требует в 2 раза меньше данных!

## Результаты для эквиатомного сплава MoNbTaWV (2)

### Расчет констант упругости и модуля всестороннего сжатия.

| Model | $C_{11}$ , GPa  | $C_{12}$ , GPa  | $C_{44}$ , GPa | Bulk modulus, GPa |
|-------|-----------------|-----------------|----------------|-------------------|
| TFMTP | $338.2 \pm 1.4$ | $160.4 \pm 1.3$ | $48.0 \pm 1.1$ | $219.7 \pm 1.0$   |
| MTP   | $332.4 \pm 1.2$ | $158.2 \pm 1.4$ | $48.9 \pm 1.0$ | $216.2 \pm 1.2$   |
| DFT   | $334.3 \pm 5.1$ | $159.8 \pm 2.7$ | $56.6 \pm 4.5$ | $217.9 \pm 1.3$   |

Константы упругости и модули всестороннего сжатия, рассчитанные с помощью MTP, TFMTP и DFT и осредненные по десяти 54-атомным твердым растворам MoNbTaWV.

| Model                | $C_{11}$ , GPa  | $C_{12}$ , GPa  | $C_{44}$ , GPa | Bulk modulus, GPa |
|----------------------|-----------------|-----------------|----------------|-------------------|
| TFMTP                | $335.7 \pm 0.3$ | $160.1 \pm 0.2$ | $48.4 \pm 0.1$ | $218.6 \pm 0.2$   |
| MTP                  | $331.1 \pm 0.5$ | $158.0 \pm 0.3$ | $49.5 \pm 0.1$ | $215.7 \pm 0.3$   |
| tabGAP <sup>42</sup> | $382.2 \pm 0.6$ | $124.5 \pm 0.3$ | $47.5 \pm 0.3$ | $210.4 \pm 0.3$   |

Константы упругости и модули всестороннего сжатия, рассчитанные с помощью MTP, TFMTP и tabGAP и осредненные по пятидесяти 2000-атомным твердым растворам MoNbTaWV.

[42] J. Byggmästar, K. Nordlund, and F. Djurabekova, “Modeling refractory high-entropy alloys with efficient machine-learned interatomic potentials: Defects and segregation,” Physical Review B 104, 104101 (2021).

Вывод: TFMTP и MTP дают близкие результаты как друг к другу, так и к другим моделям!

# Технические особенности обучения TFMTP

Потенциал TFMTP при обучении методом BFGS удается успешно обучить, стартуя не с любого начального приближения и «застревание» алгоритма происходит на этапе linesearch.

1.  $H_k \mathbf{p}_k = -\nabla f(\mathbf{x}_k)$ .

2.  $\alpha_k = \arg \min_{\alpha} f(\mathbf{x}_k + \alpha \mathbf{p}_k)$ , но на практике ищем, например, бинарным поиском пока не выполняются условия:

(i)  $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \mathbf{p}_k^T \nabla f(\mathbf{x}_k)$

(ii)  $\mathbf{p}_k^T \nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \geq c_2 \mathbf{p}_k^T \nabla f(\mathbf{x}_k)$

3.  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$ ,  $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$ ,  $H_{k+1} = H_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{H_k \mathbf{s}_k \mathbf{s}_k^T H_k}{\mathbf{s}_k^T H_k \mathbf{s}_k}$

Чтобы избежать этого «застревания»:

- из энергий, рассчитанных на DFT, вычиталось минимальное значение энергии, которое было в изначальной базе данных;
- начальное приближение выбиралось на интервале (0.1,0.8), используя равномерное распределение;
- веса минимизируемого функционала брались в 10 раз меньше стандартных:  $w_e = 0.1$ ,  $w_f = 0.001$ ,  $w_s = 0.0001$  для FLiNaK, а для MoNbTaWV вес  $w_s = 0.001$  остался стандартным.

## Заключение

TFMTP, имеющий в 1.5 раза меньше параметров, чем MTP, требует в 2 раза меньше данных для обучения однако дает схожую точность как в метрике ошибок по энергиям, силам и напряжениям, так и в рамках расчета физических свойств.

## Будущие планы

Разработать метод оптимизации TFMTP, позволяющий оптимизировать его без подбора скейлинга.

Спасибо за внимание!