# Application of machine-learning for construction of bias potential: a case study of add-atom hyperdynamics and straight screw dislocation migration.

Ivan S. Novikov

February 27, 2021

**Abstract**

In this report we describe a bias potential for add-atom global hyperdynamics on the basis of machine-learning (ML) interatomic potential (namely, Moment Tensor Potential, MTP). We compare the results obtained using the ML-bias potential with the ones obtained with conventional bond-boost bias potential. We also discuss possibilities for construction of a ML-bias potential for acceleration a migration of straight screw dislocation.

## 1 Global hyperdynamics and bond-boost bias potential

Hyperdynamics (HD) method was proposed in [8]. This method allows to extend the time scale $t^{\mathrm{MD}}$ of a molecular dynamics (MD) simulation and to observe infrequent events (like migrations of atoms, screw dislocations) with higher frequency.

Consider a system located in a basin $A$ of the potential energy function $V$. In addition we introduce a bias potential $\Delta V^b$, which is non-negative around the local basin $A$ and is equal to zero near the transition state region. We also choose such the bias potential which does not block any escape paths, i.e., any subminima within the state $A^b$ have escape times substantially shorter than the escape time $\tau^A_{\mathrm{esc}}$ for the state $A$. In [8] it was shown, that the average time to escape from the state $A$

$$\tau^A_{\mathrm{esc}} = \frac{1}{n_{\mathrm{esc}}} \sum_{l=1}^{n_{\mathrm{tot}}} \Delta t^{\mathrm{MD}} e^{\frac{\Delta V^b_l}{k_B T}}, \tag{1}$$

where $n_{\mathrm{esc}}$ is the number of escape attempts, $l$ is the current number of MD step, $\Delta t^{\mathrm{MD}}$ is the MD time step, $n_{\mathrm{tot}}$ is the total number of MD steps. At long time scale (infinitely

long trajectory) the boost time $t^b = \tau^A_{\rm esc} n_{\rm esc}$ and, thus

$$t^b = \sum_{l=1}^{n_{\rm tot}} \Delta t^{\rm MD} e^{\frac{\Delta V_l^b}{k_B T}}. \tag{2}$$

With $t^{\rm MD} = \Delta t^{\rm MD} n_{\rm tot}$ we conclude, that $t^b = t^{\rm MD}$ where the bias potential is equal to zero (as for normal MD).

One of the possible bias potentials for global HD is the bond-boost potential proposed in [4]. The bond of atoms $i$ and $j$ is defined if the equilibrium distance $r^0_{ij}$ between these two atoms is smaller than specific constant value, which is defined for each specific system. The current strain of the bond $ij$ on the $l$-th MD step

$$\varepsilon^l_{ij} = \frac{r^l_{ij} - r^0_{ij}}{r^0_{ij}}, \tag{3}$$

where $r^l_{ij}$ is current (non-equilibrium) distance between the atoms. The bias energy $\Delta V^b_{l,ij}$ of any bond is defined as

$$\Delta V^b_{l,ij} = \begin{cases} V_{\rm max}\left(1 - \left(\dfrac{\varepsilon^l_{ij}}{q}\right)^2\right) & |\varepsilon^l_{ij}| < q \\ 0 & |\varepsilon^l_{ij}| \geq q. \end{cases} \tag{4}$$

Here the non-negative factor $V_{\rm max}$ and the threshold $q$ for strain (or, maximal permissible strain) are the adjustable parameters. Let the maximum strain on the $l$-th step be $\varepsilon^l_{\rm max} = \max\limits_{ij} |\varepsilon^l_{ij}|$. Then, the total potential energy of the single bond with the strain $\varepsilon^l_{\rm max}$ on the current step is equal to sum of the potential function $V_l$ and a bias function $\Delta V^b_{l,\rm max}$. We note that $\Delta V^b_{l,\rm max} = 0$ if $\varepsilon^l_{\rm max} \geq q$. The bias force acting on the bond with the maximum strain

$$f^b_{l,\rm max} = -\frac{\partial \Delta V^b_{l,\rm max}}{\partial \varepsilon^l_{\rm max}} = \begin{cases} \dfrac{2V_{\rm max}\varepsilon^l_{\rm max}}{q^2} & \varepsilon^l_{\rm max} < q \\ 0 & \varepsilon^l_{\rm max} \geq q. \end{cases} \tag{5}$$

The force can be decomposed into an equal and opposite force acting only on the atoms which form the bond with $\varepsilon^l_{\rm max}$.

It was demonstrated in [4] that the bond-boost potential works good even for local HD, in which the bias potential applied to a group of bonds (as opposite to global HD, described above). Therefore, in order to investigate a possibility to use machine-learning bias (ML-bias) potential for boosting we decided to start with add-atom global HD and consider the bond-boost potential (4) as the reference one.

# 2 Moment Tensor Potential

Moment Tensor Potential (MTP) is a machine-learning interatomic potential which was proposed in [3, 7]. MTP is local, i.e., the energy $E^{\mathrm{MTP}}$ is the sum of contributions $V^{\mathrm{MTP}}(\mathfrak{n}_i)$ of atomic neighborhoods $\mathfrak{n}_i$ for $N$ atoms

$$E^{\mathrm{MTP}} = \sum_{i=1}^{N} V^{\mathrm{MTP}}(\mathfrak{n}_i). \tag{6}$$

Each neighborhood is a tuple $\mathfrak{n}_i = (\{r_{i1}, z_i, z_1\}, \ldots, \{r_{ij}, z_i, z_j\}, \ldots, \{r_{iN_{\mathrm{neigh.}}}, z_i, z_{N_{\mathrm{neigh.}}}\})$, where $r_{ij}$ are relative atomic positions, $z_i$, $z_j$ are the types of central and neighboring atoms, $N_{\mathrm{neigh.}}$ is the number of atoms in neighborhood. We also denote the maximum number of atomic types occured in all the neighborhoods by $N_{\mathrm{types}}$. Each contribution $V^{\mathrm{MTP}}(\mathfrak{n}_i)$ in the potential energy $E^{\mathrm{MTP}}$ expands through a set of basis functions

$$V^{\mathrm{MTP}}(\mathfrak{n}_i) = \sum_{\alpha=1}^{N_{\mathrm{lin.}}} \xi_\alpha B_\alpha(\mathfrak{n}_i), \tag{7}$$

where $B_\alpha$ are the MTP basis functions, $\xi_\alpha$ are the linear parameters to be found, and $N_{\mathrm{lin.}}$ is the number of these parameters. To define the functional form of the MTP basis functions and the number $N_{\mathrm{lin.}}$ we introduce the so-called moment tensor descriptors

$$M_{\mu,\nu}(\mathfrak{n}_i) = \sum_{j=1}^{N_{\mathrm{neigh.}}} f_\mu(|r_{ij}|, z_i, z_j) \underbrace{r_{ij} \otimes \ldots \otimes r_{ij}}_{\nu \text{ times}}. \tag{8}$$

The descriptor consists of the angular part $r_{ij} \otimes \ldots \otimes r_{ij}$ (the symbol "$\otimes$" denotes the outer product of vectors and, thus, the angluar part is the tensor of $\nu$-th order) and the radial part $f_\mu(|r_{ij}|, z_i, z_j)$ of the following form

$$f_\mu(|r_{ij}|, z_i, z_j) = \sum_{\beta=0}^{N_{\mathrm{polyn.}}-1} c^{(\beta)}_{\mu,z_i,z_j} T_\beta(|r_{ij}|)(R_{\mathrm{cut}} - |r_{ij}|)^2. \tag{9}$$

Here $\mu = 0, \ldots, N_{\mathrm{rad.}} - 1$ is the number of the radial function $f_\mu$ (the method to define a concrete number $N_{\mathrm{rad.}}$ of radial functions needed to construct all the MTP basis functions is detailed below), $c^{(\beta)}_{\mu,z_i,z_j}$ are the radial parameters to be found, $T_\beta(|r_{ij}|)$ are Chebyshev polynomials of the order $\beta$, $N_{\mathrm{polyn.}}$ is the number of Chebyshev polynomials (the highest order is $N_{\mathrm{polyn.}} - 1$ because the sequence of the polynomials starts from the one of 0 order). The number of the radial MTP parameters $c^{(\beta)}_{\mu,z_i,z_j}$ is $N_{\mathrm{polyn.}} \times N_{\mathrm{rad.}} \times N^2_{\mathrm{types}}$. Finally, the term $(R_{\mathrm{cut}} - |r_{ij}|)^2$ is introduced to ensure smothness w.r.t. to the atoms leaving and entering the sphere with the cut-off radius $R_{\mathrm{cut}}$.

By definition, the MTP basis function $B_\alpha$ is a multiplication of one or more moment tensor descriptors, yielding a scalar, e.g.

$$M_{1,0},\ (M_{1,2}M_{0,1}) \cdot M_{0,1},\ M_{1,2} : M_{3,2}, \ldots,$$

where "$\cdot$" is an inner product of vectors, ":" is a Frobenius product of matrices. In order to construct the basis functions $B_\alpha$, and, thus, determine a particular functional form of MTP, we define the so-called level of moment tensor descriptor

$$\mathrm{lev}M_{\mu,\nu} = 2+4\mu+\nu, \tag{10}$$

for example, $\mathrm{lev}M_{0,0}=2$, $\mathrm{lev}M_{2,1}=11$, $\mathrm{lev}M_{0,3}=5$. We also define the level of moment tensor descriptor multiplication, and, in particular, the level of the MTP basis function

$$\mathrm{lev}B_\alpha = \mathrm{lev}\underbrace{\prod_{p=1}^{P} M_{\mu_p,\nu_p}}_{\text{scalar}} = \sum_{\mathrm{p}=1}^{\mathrm{P}} (2+4\mu_{\mathrm{p}}+\nu_{\mathrm{p}}). \tag{11}$$

A set of MTP basis functions and, thus, a particular functional form of MTP depends on the maximum level, $\mathrm{lev}_{\mathrm{max}}$, which we also call the level of MTP. We include in the set of the MTP basis functions only the ones with $\mathrm{lev}B_\alpha \leq \mathrm{lev}_{\mathrm{max}}$, e.g., the MTP of 8-th level includes nine basis functions

$$B_1 = M_{0,0};\ \mathrm{lev}M_{0,0}=2\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_2 = M_{1,0};\ \mathrm{lev}M_{1,0}=6\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_3 = M_{0,0}^2;\ \mathrm{lev}M_{0,0}^2=4\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_4 = M_{0,1}\cdot M_{0,1};\ \mathrm{lev}(M_{0,1}\cdot M_{0,1})=6\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_5 = M_{0,2} : M_{0,2};\ \mathrm{lev}(M_{0,2} : M_{0,2})=8\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_6 = M_{0,0}M_{1,0};\ \mathrm{lev}(M_{0,0}M_{1,0})=8\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_7 = M_{0,0}^3;\ \mathrm{lev}M_{0,0}^3=6\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_8 = M_{0,0}(M_{0,1}\cdot M_{0,1});\ \mathrm{lev}(M_{0,0}(M_{0,1}\cdot M_{0,1}))=8\leq \mathrm{lev}_{\mathrm{max}}=8,$$

$$B_9 = M_{0,0}^4;\ \mathrm{lev}M_{0,0}^4=8\leq \mathrm{lev}_{\mathrm{max}}=8.$$

Thus, the number of linear parameters $N_{\mathrm{lin.}}$ and the number $N_{\mathrm{rad.}}$ of radial functions depend on the level of MTP and are fixed for each level (e.g., if $\mathrm{lev}_{\mathrm{max}}=8$ then $N_{\mathrm{lin.}}=9$, $N_{\mathrm{rad.}}=2$).

Using the method described above, it is possible to calculate MTP energy. As the energy is a smooth function, then it is possible to find its derivatives analytically, e.g., MTP forces and MTP stresses. Nevertheless, as it was mentioned above, typically the MTP parameters are not known, or, non-optimal (e.g., they do not allow to predict energies, forces and/or stresses of any realistic physical system). For finding the optimal MTP parameters we need a training set, i.e. a set of atomic configurations with the known reference energies, (and, probably, forces and stresses) to which we approximate the MTP energies (and, probably, MTP forces and MTP stresses). We assume, that

the reference energies, forces, and stresses have physical meaning and correspond to any realistic physical system, and, thus, by using of them, we can optimize the MTP parameters, and the potential will predict physically relevant properties of a material. We emphasize, that in this section we only describe a general method to find optimal parameters, but do not discuss any details on construction of a training set (see the details in the sections below), because it depends on a particular problem we are interested in.

Let $K$ be a number of configurations in a training set, and $N^{(k)}$ is a number of atoms in $k$-th configuration. Denote a vector of MTP parameters to be found by $\boldsymbol{\theta} = (\xi_\alpha, c^{(\beta)}_{\mu,z_i,z_j})$. We find the optimal parameters $\bar{\theta}$ by solving the following optimization problem (minimization of the objective function)

$$\sum_{k=1}^{K}\Big[w_{\mathrm{e}}\left(E_k^{\mathrm{REF}}-E_k^{\mathrm{MTP}}(\boldsymbol{\theta})\right)^2+w_{\mathrm{f}}\sum_{i=1}^{N^{(k)}}\sum_{a=1}^{3}\left(f_{i,a,k}^{\mathrm{REF}}-f_{i,a,k}^{\mathrm{MTP}}(\boldsymbol{\theta})\right)^2 +w_{\mathrm{s}}\sum_{a,b=1}^{3}\left(\sigma_{ab,k}^{\mathrm{REF}}-\sigma_{ab,k}^{\mathrm{MTP}}(\boldsymbol{\theta})\right)^2\Big]\rightarrow\min. \tag{12}$$

We start from randomly initialized MTP parameters on the interval $(-10^{-7},10^{-7})$. It should be noticed, that the order of parameters on the interval does not matter, because before the optimization such a re-normalization of parameters occurs that the norm of the coefficient vector is equal to unity (we do not describe the details here). The optimal parameters $\bar{\theta}$ are found numerically, by using the iterative method for minimization of the nonlinear objective function, namely, Broyden-Fletcher-Goldfarb-Shanno algorithm. Thus, after the optimization, the parameters $\bar{\theta}$ are near the local minimum of the objective function. In this objective function, $E_k^{\mathrm{REF}}$, $f_{i,a,k}^{\mathrm{REF}}$, and $\sigma_{ab,k}^{\mathrm{REF}}$ are the reference energies, forces, and stresses, i.e. the ones to which we approximate the MTP energies $E_k^{\mathrm{MTP}}$, forces $f_{i,a,k}^{\mathrm{MTP}}$, and stresses $\sigma_{ab,k}^{\mathrm{MTP}}$, and, thus, optimize the MTP parameters $\boldsymbol{\theta}$. The factors $w_{\mathrm{e}}$, $w_{\mathrm{f}}$, and $w_{\mathrm{s}}$ are non-negative weights which express the importantance of energies, forces, and stresses w.r.t. each other. We refer to the minimization problem (12) as the fitting of MTP. Using a training set and MTP trained we calculate the so-called extrapolation grade per atom (described below) which we use to construct ML-bias potential.

# 3 Extrapolation grade per atom

We introduce a concept of extrapolation grade per atom for nonlinear MTP on the basis of the papers [6] (the grade per atom was introduced for linear MTP) and [3] (the extrapolation grade of configuration was proposed for nonlinear MTP). We start with the consideration of (12). Assume that we know the vector of optimal MTP parameters $\bar{\theta}$ and let the length of the vector be $m$. Then we can linearize each term in the objective

function, in particular, the term with the difference between energies

$$E_k^{\mathrm{REF}} - E_k^{\mathrm{MTP}}(\boldsymbol{\theta}) \approx E_k^{\mathrm{REF}} - \sum_{p=1}^{m} (\theta_p - \bar{\theta}_p) \frac{\partial E_k^{\mathrm{MTP}}(\bar{\boldsymbol{\theta}})}{\partial \theta_p}.$$

We can consider fitting of MTP as the solution of the overdertemined system

$$\sum_{p=1}^{m} \theta_p \frac{\partial E_k^{\mathrm{MTP}}(\bar{\boldsymbol{\theta}})}{\partial \theta_p} = E_k^{\mathrm{REF}} + \sum_{p=1}^{m} \bar{\theta}_p \frac{\partial E_k^{\mathrm{MTP}}(\bar{\boldsymbol{\theta}})}{\partial \theta_p}. \tag{13}$$

Substituting (6) in (13) and taking into account that $V^{\mathrm{MTP}}$ also depends on $\boldsymbol{\theta}$, we obtain

$$\sum_{i=1}^{N^{(k)}} \sum_{p=1}^{m} \theta_p \frac{\partial V^{\mathrm{MTP}}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^{(k)})}{\partial \theta_p} = \sum_{i=1}^{N^{(k)}} \left( V_{i,k}^{\mathrm{REF}} + \sum_{p=1}^{m} \bar{\theta}_p \frac{\partial V^{\mathrm{MTP}}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^{(k)})}{\partial \theta_p} \right), \; k = \overline{1,K}. \tag{14}$$

The matrix of the system (14) is

$$\mathsf{B} = \begin{pmatrix} \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_1^{(1)}) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_1^{(1)}) \\ \vdots & & \vdots \\ \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_{N^{(1)}}^{(1)}\right) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_{N^{(1)}}^{(1)}\right) \\ \vdots & & \vdots \\ \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_1^{(K)}) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_1^{(K)}) \\ \vdots & & \vdots \\ \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_{N^{(K)}}^{(K)}\right) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_{N^{(K)}}^{(K)}\right) \end{pmatrix},$$

where the first $N^{(1)}$ rows correspond to the first configuration (of $N^{(1)}$ atoms) in the training set, the last $N^{(K)}$ rows correspond to the last ($K$-th) configuration, and, thus, the size of the matrix is $\left(\sum_{k=1}^{K} N^{(k)}\right) \times m$.

Following [6] we find a square $m \times m$ submatrix $\mathsf{A}$ of matrix $\mathsf{B}$ using the maxvol algorithm [2] (i.e., we find the matrix with maximum volume, or, $|\det \mathsf{A}|$). In order to find grades per atom for a configuration with atomic neighborhoods $\mathfrak{n}_i^*$, $i = 1, \dots, N$, we compose a matrix

$$\mathsf{C} = \begin{pmatrix} \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_1^*) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_p}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_1^*) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_1^*) \\ \vdots & & \vdots & & \vdots \\ \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^*\right) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_p}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^*) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^*\right) \\ \vdots & & \vdots & & \vdots \\ \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_N^*\right) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_p}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_N^*) & \dots & \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_N^*\right) \end{pmatrix} \mathsf{A}^{-1},$$

where the derivatives $\frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^*\right)$ are calculated analytically. The grade per $i$-th atom (or, the local grade) is the maximum absolute value in the $i$-th row of the matrix $\mathsf{C}$, namely,

$$\gamma_i = \gamma(\mathfrak{n}_i^*) = \max_{1\le p\le m}(|C_{i,p}|), \text{ where}$$
$$C_{i,p} = \left(\frac{\partial V^{\mathrm{MTP}}}{\partial \theta_1}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^*\right), \ldots, \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_p}(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^*), \ldots, \frac{\partial V^{\mathrm{MTP}}}{\partial \theta_m}\left(\bar{\boldsymbol{\theta}}, \mathfrak{n}_i^*\right)\right)\mathsf{A}^{-1}, \tag{15}$$

and the extrapolation grade of configuration (or, the global grade) is the maximum absolute element of the matrix $\mathsf{C}$, i.e.,

$$\gamma = \max_{1\le i\le N,\ 1\le p\le m}(|C_{i,p}|), \tag{16}$$

thus, the global grade is indeed the maximum through local grades.

As it could be seen, the extrapolation grade depends on geometry of configuration (or, atomic neighborhoods). In hyperdynamics we typically deal with local basins (local minima) and transition state with different neighborhoods for atoms which participate in transition (and, thefore, in boosting). Due to this reason, the extrapolation grades are also different, e.g., the local grades per atoms in transition state could be greater than the ones in local minima. This is a motivation to use the global grade as a measure to stop boosting (i.e., if the global grade is greater than a maximal permissible grade, then we stop boosting, and, therefore, the maximal permissible grade is an analog of the maximal permissible strain which determines a moment to stop boosting in bond-boost potential). The concept of global grade per atom is applied for construction of a ML-bias potential used for global hyperdynamics (e.g., in the add-atom dynamics, see the next two sections), whereas the concept of local grade could be used for local hyperdynamics.

# 4 Machine-learning bias potential for global hyperdynamics

As it was discussed above, we construct a ML-bias potential for global hyperdynamics under the assumption that it depends on global grade $\gamma$: $\Delta V^{\mathrm{ML-b}} = \Delta V^{\mathrm{ML-b}}(\gamma)$. Thus, at every $l$-th MD step we calculate global grade $\gamma^l$ with (16) and choose the non-negative factor $V_{\max}$ and threshold $\gamma_{\mathrm{th}}$ (like the threshold $q$ in the bond-boost potential). To a single atom of a maximum grade we apply the ML-bias energy of the following form

$$\Delta V_{l,\max}^{\mathrm{ML-b}} = \begin{cases} V_{\max}\mathrm{f}\left(1-\dfrac{\gamma^l}{\gamma_{\mathrm{th}}}\right) & \gamma^l < \gamma_{\mathrm{th}} \\ 0 & \gamma^l \ge \gamma_{\mathrm{th}}, \end{cases} \tag{17}$$

where $\mathrm{f}\left(1-\dfrac{\gamma^l}{\gamma_{\text{th}}}\right)$ is any non-negative function which allows to avoid too large boosting time $t^b$ (i.e., the so-called "ballistic" effect, when the boosting energy is too large, and, therefore, the boosting time is also too high). As it was shown above, the extrapolation grade depends on atomic positions. That's why it is possible to calculate the ML-bias force. This force has the form

$$f^{\text{ML-b}}_{l,\max} = -\frac{\partial \Delta V^{\text{ML-b}}_{l,\max}}{\partial r^l_{\max}} = \begin{cases} -V_{\max}\dfrac{\partial \mathrm{f}\left(1-\dfrac{\gamma^l}{\gamma_{\text{th}}}\right)}{\partial \gamma^l}\dfrac{\partial \gamma^l}{\partial r^l_{\max}} & \gamma^l < \gamma_{\text{th}} \\ 0 & \gamma^l \geq \gamma_{\text{th}}, \end{cases} \tag{18}$$

where $r^l_{\max}$ is the coordinate of atom with maximum grade. As the depence of extrapolation grade on atomic positions is the composition function, i.e.,

$$\gamma^l = \gamma^l(V^{\text{MTP}}(\mathfrak{n}_1),\ldots,V^{\text{MTP}}(\mathfrak{n}_i),\ldots,V^{\text{MTP}}(\mathfrak{n}_N)),$$

then it is difficult to calculate the derivative $\dfrac{\partial \gamma^l}{\partial r^l_{\max}}$ analytically, that's why we calculated it numerically, using the central differences. We generate such a training set that the extrapolation grade is small enough for configurations close to local basin, and is greater for configurations close to transition state. The parameter $\gamma_{\text{th}}$ could be close to the grade in transition state, or greater, if during simulation we occur configurations with higher grade, than in transition state, but the configuration with this grade does not correspond to the transition state itself.

# 5 Machine-learning bias vs bond-boost bias: a case study of add-atom dynamics for Pt-fcc

For estimation of the ML-bias potential efficiency we compare it with the bond-boost bias potential by using of these two potentials in add-atom dynamics on the example of Pt-fcc. The example was taken from the `LAMMPS` open source software which is available at https://lammps.sandia.gov/doc/Install_git.html (see the folder `examples/hyper/`). As the potential energy function we use EAM potential for Pt-fcc generated from Voter potential (see the file `ptvoterlammps.eam` in the folder mentioned above) with the cut-off radius of 5.6 Å (which is close to the distance between any atom if fcc-lattice and its second nearest neighbor). We consider an orthogonal unit cell of 577 atoms with the lattice parameters $a = b = 23.52$ Å (along the axes $Ox$ and $Oy$), $c = 19.992$ Å (along the axis $Oz$), the crystallographic directions are $< 100 >$. The first 576 atoms are located inside the surface including 8 layers (72 atoms per layer), the distance between layers is 1.96 Å. We investigate a migration of the 577-th atom at $T = 700$ K from one

local basin (where the coordinate of this atom is $(13.72, 13.72, 15.09)$ Å, this atom is above the 8-th layer) to one of four possible local basins, where the 577-th moves to the 8-th layer and its coordinate will be close to the coordinate of one of the four nearest neighbors: $(13.72, 11.76, 13.72)$ Å, $(11.76, 13.72, 13.72)$ Å, $(15.68, 13.72, 13.72)$ Å, and $(13.72, 15.68, 13.72)$ Å, and one of the nearest neighbors leaves the surface and will be located above it (the so-called "add-atom exchange" mechanism). The distance between the additional atom and any nearest neighbor is appoximately equal to 2.4 Å (see the Fig. 1, left top, a view from above), thus, in order to migrate from one local basin to another it is necessary to do the path of $\approx 2.4$ Å for both atoms. In the case shown below the additional atom migrates to the position $(13.72, 11.76, 13.72)$ Å, and the new coordinate of this nearest neighbor is $(13.72, 9.8, 15.09)$ Å (see the Fig. 1).

We emphazsize, that in this simulation we do not consider the so-called "bridge-hoping" diffusion of 577-th atom. Moreover, this type of atom migration is not possible with the probability close to 1 for the Pt-fcc system described above at $T = 700$ K. To verify this, we ran 1000 pure MDs of 2 ns starting from different initial velocities. For 983 of 1000 cases we observed one of four "add-atom exchange" migrations with the close probability (it varies from 22 % to 26 %), and did not observe any migration for the rest 17 cases. Thus, we may state, that with a probability close to one we will observe "add-atom exchange" migrations with the transition state, close to the one shown in the Fig. 1. Our boosting should be turned-off near this transition state.

In order to use MTP for boosting we, first, should create a training set and fit the potential. We generated the following training set. We ran MD of 1.5 ns with the timestep of $\Delta t^{\mathrm{MD}} = 5$ fs (thus, we had 300000 steps) in NVT ensemble at $T = 700$ K. Then, in order to avoid the correlation between configurations, we considered each 10000-th configuration from MD simulation (i.e., 31 configurations) and added in the training set only the configurations with small deviations (thermal vibratons) of atoms from their starting positions. The maximum permissible deviation was chosen equal to 0.8 Å, i.e., approximately 30% of the path which is necessary to do for migration from one local basin to other. Finally, we obtained the training set including 29 configurations with small thermal vibrations of atoms (thus, we excluded 2 configurations in which the deviations were more than 0.8 Å). We consider only small deviations of atoms in the training set and do not include the configurations with big deviations, because our aim is to have smaller extrapolation grades near local basins (at around 1 or smaller, to have interpolation on these configurations), and higher ones (at around 2 or higher, to have extrapolation) near transition state. The big deviations could correspond to transition state, thus, adding the configurations with big deviations in the training set leads to small extrapolation grade (or, even to interpolation) near transition state and we will not be able to use the extrapolation grade as a measure to stop boosting.

We fit MTP with $\mathrm{lev}_{\max} = 8$ ($N_{\mathrm{lin.}} = 9$), $N_{\mathrm{rad.}} = 2$, $N_{\mathrm{polyn.}} = 8$, thus, MTP has 26 parameters. The cut-off radius $R_{\mathrm{cut.}} = 5.8$ Å (a bit greater than the distance between any atom in fcc-lattice and its second nearest neighbor). For optimization of MTP parameters (or, fitting of MTP) we solved the problem (12). The reference energies, forces, and

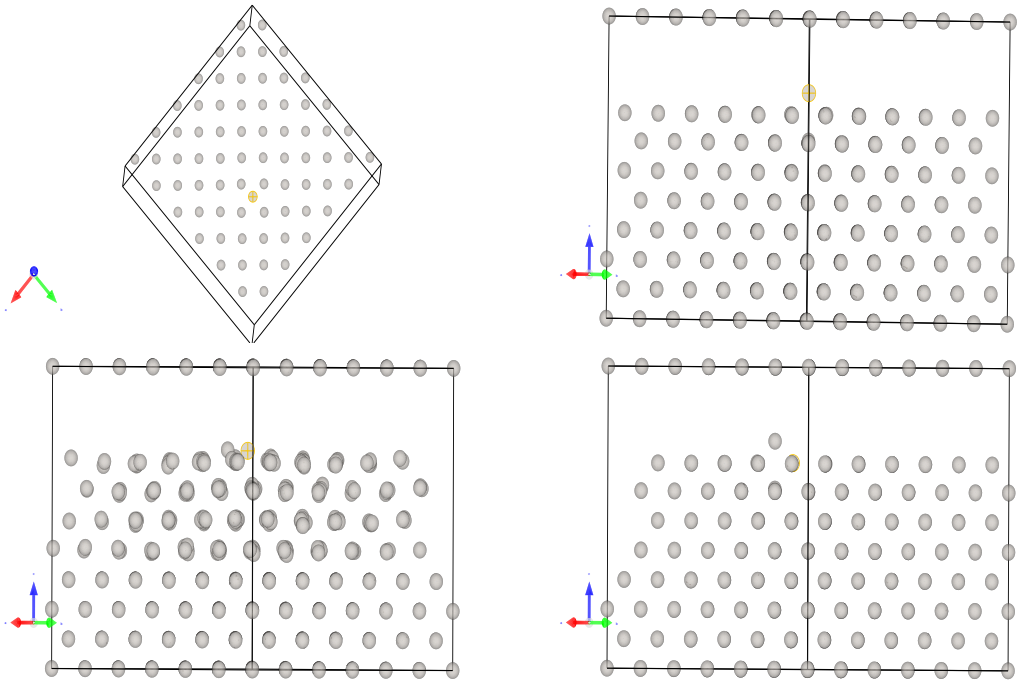

Figure 1: The left top figure illustrates the nearest neighbors of the additional atom (crossed). In the right top figure the first local basin is shown: the coordinate of the additional atom is $(13.72, 13.72, 15.09)$ Å, the coordinate of the nearest neighbor to be migrated is $(13.72, 11.76, 13.72)$ Å). In the left bottom figure we see the approximate transition state. The new coordinate of the additional atom is $(13.7945, 13.1917, 14.47)$ Å, thus, the displacement is approximately equal to 0.8 Å, the atom done at around 30% of the path to its new local basin. The new coordinate of the nearest neighbor is $(13.7675, 10.8092, 14.5261)$ Å, therefore, the displacement is close to 1.2 Å, the atom approximately done 50% of the path to its new local minimum. Finally, in the right bottom figure, the second local basin is shown: the coordinate of the additional atom is $(13.72, 11.76, 13.72)$ Å, now it is located in the 8-th layer, thus, we do not see the crossed atom in this figure, and the coordinate of the nearest neighbor is $(13.72, 9.8, 15.09)$ Å, it is migrated from the layer.

stresses in (12) (or, the energies, forces, and stresses in the training set) were calculated using EAM potential for Pt-fcc mentioned above, the factors before the differences between energies, forces, and stresses were $w_e = 1$, $w_f = 0.01$, $w_s = 0$. We chose $w_s = 0$, because it is not necessary to predict stresses (or, elastic constants).

Once we trained MTP (or, optimized the parameters of MTP), we can calculate the extrapolation grade and ML-bias potential. For the add-atom dynamics we consider the ML-bias energy for the atom with maximum extrapolation grade of the following form

$$\Delta V_{l,\max}^{\mathrm{ML-b}} = \begin{cases} V_{\max}\left(1 - \dfrac{\gamma^l}{\gamma_{\mathrm{th}}}\right)^p & \gamma^l < \gamma_{\mathrm{th}} = 10,\ p = 8 \\ 0 & \gamma^l \geq \gamma_{\mathrm{th}} = 10,\ p = 8, \end{cases} \tag{19}$$

with different non-negative weights $V_{\max}$ from the interval (0,1). We boost an atom with maximal extrapolation grade (there could be any of 577 atoms, either add-atom, or one of atoms inside layers) if the global extrapolation grade $\gamma^l$ of current configuration is greater than $\gamma_{\mathrm{th}}$, and do not boost any atom otherwise. For choosing the threshold $\gamma_{\mathrm{th}}$ we, first, ran 100 add-atom with bond-boost bias potential starting from different initial velocities, then, calculated the grades for all the configurations sampled. We found that the maximum grade among all the runs was at around 10. Such a high grade corresponds to configuration in which additional atom moved for at around 1 Å above the surface compared to its original position. Thus, $\gamma_{\mathrm{th}} = 10$. We also found that the grade of configuration in transition state is equal to 2-3, thus, our ML-bias potential should be close to zero if $\gamma^l = 3$ for turning-off boosting near the transition state. Finally, our ML-bias potential should not be "ballistic", i.e., a boosting energy should not be too large. In order to choose a power of the boost function which allows to satisfy all the above criterions, we considered the 2-nd, 4-th, 6-th, 8-th, 10-th, and 20-th powers of (19). The bias energies, and their derivatives w.r.t. to extrapolation grade $\gamma$ for $V_{\max} = 0.8$ are plotted in the Fig. 2.

We then ran 100 add-atom dynamics with the ML-bias potentials described above and calculated the average boosting time $t^b$ for all runs. As it will be shown below, typical number of steps, needed to detect an event (i.e., jump of the additional atom) with bond-boost potential for $V_{\max} = 0.8$ with the probability close to 1 is 100000 with the timestep $\Delta t^{\mathrm{MD}} = 5$ fs. We adjusted the critical time of boosting $t^b_{\mathrm{critical}} = 2$ ns, i.e., if the hyperdynamics time is four times greater than the pure MD time (0.5 ns), than, the ML-bias potential is considered as "ballistic" and could not be used for boosting. The MD time was exceeded for the ML-bias potentials with $p = 2$, $p = 4$, and $p = 6$ (i.e., $t^b > 2$ ns for 100000 steps), thus, the potentials with these powers are the "ballistic" ones. Also, these bias potentials are not suitable, because the bias energies are close to zero too far from the grade near transition state. All the potentials with the powers greater than 8 are not the "ballistic" ones. We chose the power $p = 8$, because the derivatives of bias energy w.r.t. to extrapolation grade are greater for this power, than for $p = 10$, and $p = 20$ if the extrapolation grades are greater than 1 (i.e., when we deal with extrapolation and close to transition state).

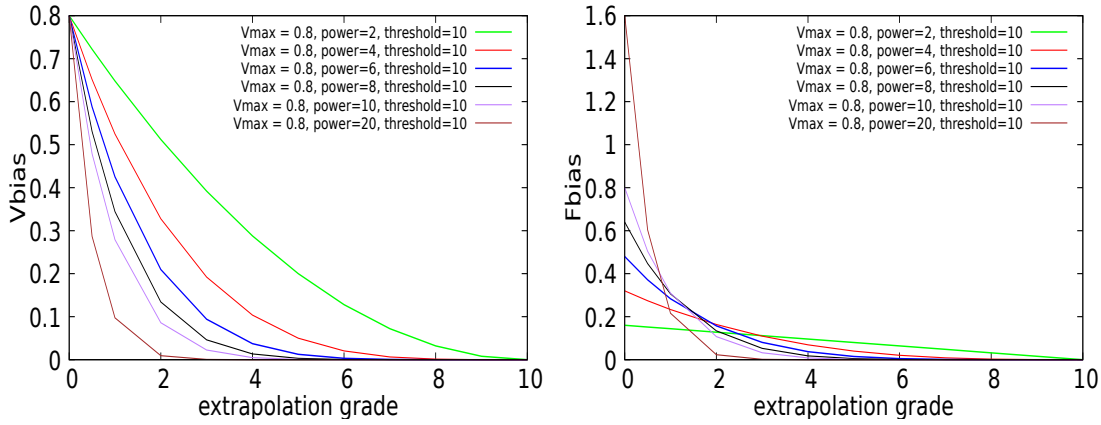


Figure 2: Energies of ML-bias potential (left) and their derivatives w.r.t. extrapolation grade (right).

For a comparison of ML-bias potential and bond-boost bias potential we ran add-atom dynamics at $T = 700K$ for different parameters $V_{\max} = 0.4, 0.6, 0.8$ and constructed the probability functions of distribution of the first jumps of atoms (i.e., we consider only the first jump of additional atom and one of its nearest neighbohrs and did not consider next jumps) on a certain time intervals. For this aim, depending on the value of $V_{\max}$, we ran from 1000 to 3000 simulations starting from different initial velocities. As it is expected, smaller the parameter $V_{\max}$, higher the probability to detect the first jumps of atoms at further time moments (i.e., more time intervals to detect jumps we have), thus, we need more simulations for the small parameters $V_{\max}$ for more accurate estimation of the probability function. The MD timestep of each simulation was $\Delta t^{\mathrm{MD}} = 5$ fs, the time of each simulation was 1.5 ns, each time interval (bin) width was 50 ps (i.e., we binned simulation time on the intervals (0, 50) ps, (50, 100) ps, ..., (1450, 1500) ps), the number of simulations was 100 times greater than the number of the bin for which the "latest" first jumps were detected (i.e., if the "latest" first jumps for a specified $V_{\max}$ occured on the 20-th interval, (950, 1000) ps, then the number of simulations was equal to 2000). In order to demonstrate that ML-bias potential accelerates pure MD (i.e., the probability to detect the first jumps in the very beginning of simulation interval is greater than for pure MD) we also ran 2000 MD simulations of 10 ns, the width of bin was 500 ps (greater than for hyperdynamics) because it is expected, that the probability to detect the first jumps with pure MD simulations will be smaller, than for the one with boosting.

The probability functions of distribution of the first jumps of atoms are shown in the Fig. 3, the same function for pure MD is shown in the Fig. 4.

The probability to detect the first jumps "faster" is greater for simulations with ML-

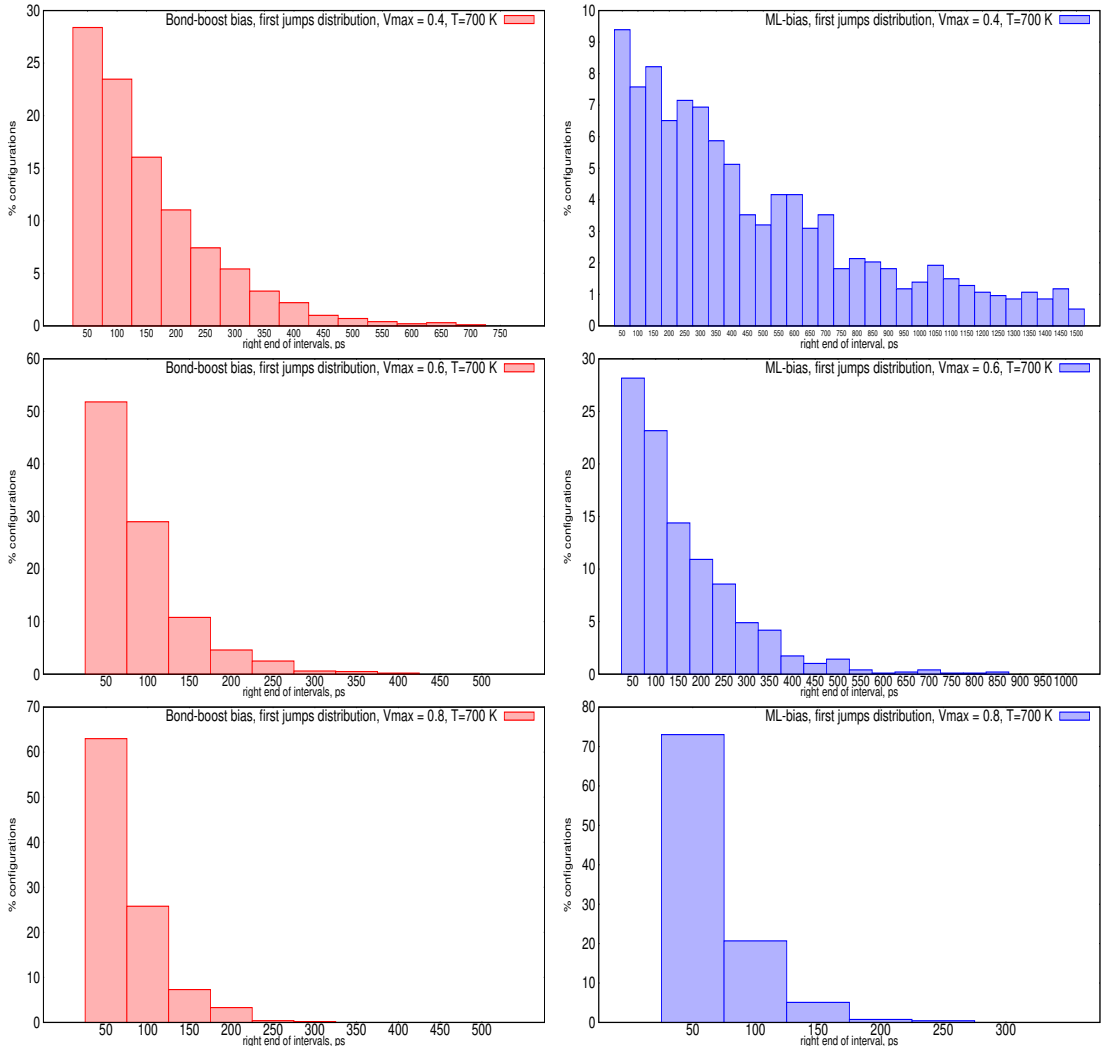


Figure 3: Probability functions of distribution of the first jumps of atoms for the bond-boost bias potential (left red histograms) and the ML-bias potential (right blue histograms) for $V_{\max} = 0.4, 0.6, 0.8$. For two small $V_{\max} = 0.4, 0.6$ the bond-boost bias potential gives better acceleration of pure MD, than the ML-bias potential, whereas for the greatest $V_{\max} = 0.8$ the ML-bias potential accelerates pure MD "faster".

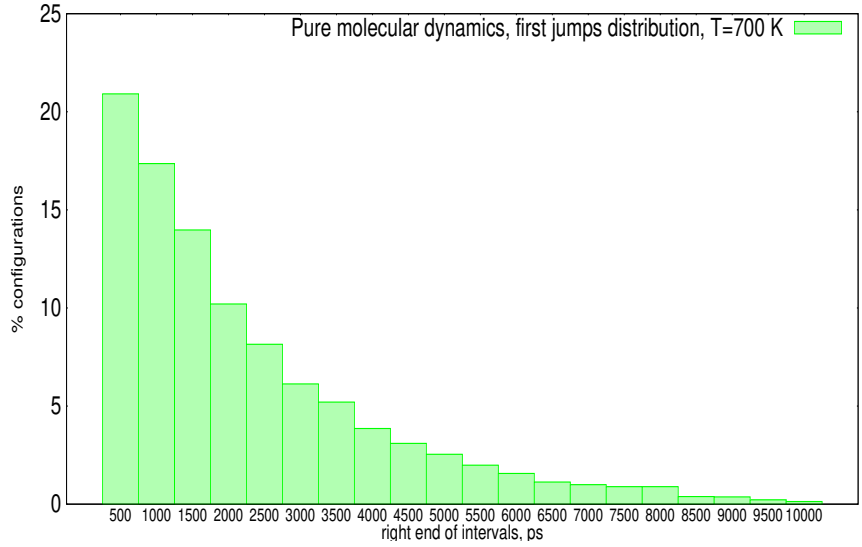


Figure 4: Probability function of distribution of the first jumps of atoms for pure MD (without any boosting).

bias potential, than without it (pure MD). Thus, both ML-bias potentails accelerate pure MD. As it was expected, greater the parameter $V_{\max}$, higher the frequency (probability) of the first jumps in the very beginning of add-atom dynamics. For two small $V_{\max} = 0.4, 0.6$ the bond-boost bias potential gives better acceleration of pure MD, than the ML-bias potential (i.e., the probability to detect the jumps "faster" is greater for the bond-boost bias potential than for the ML-bias potential), whereas for the greatest $V_{\max} = 0.8$ the ML-bias potential accelerates pure MD "faster". This effect may be related to either the choice of ML-bias energy, or to the training set. Nevertheless, for add-atom dynamics, we did not have the aim to create a ML-bias potential that accelerates pure MD better than bond-boost potential. Here we just demonstrated that the conception of extrapolation grade, in principle, could be used for construction of a bias potential.

# 6 Examination of MTP for prediction of Nb-bcc elastic constants

The second problem was investigation of a possibility to construct a ML-bias potential for acceleration of a migration of straight screw dislocation in Nb-bcc structure.

As the first step we investigated the accuracy of prediction of elastic constants with MTPs of different levels. We constructed the training sets on the basis of Nb-bcc configuration including 2 atoms in orthogonal box with the lattice parameters $a = b =$

| Nb-bcc, $(\pm 2.1; \pm 1.5; \pm 0.9; \pm 0.6; \pm 0.3)$% bulk deformation (10), shear deformation (extension/compression, 12), random atomic displacements (10), fully relaxed configuration (1), total # configurations: 33<br># atoms: 54; 128; 250; 432; 686; 1024<br># training sets: 54 × 33 configurations | | | |
|---|---|---|---|
| Shear deformation, %<br>Random atom displacement, Å | ± 0.6 | ± 1.3 | ± 2.0 |
| (-0.10, 0.10) | + | + | + |
| (-0.15, 0.15) | + | + | + |
| (-0.20, 0.20) | + | + | + |

Table 1: Nb-bcc training sets for fitting MTPs of different levels. All the results of training are demonstrated for 2 % shearing of lattice vectors, and for atomic random displacements from the interval (-0.2, 0.2) Å.

$c = 3.3$ Å (along all the three axes: $Ox$ is parallel to the direction [100], $Oy$ is parallel to the direction [010], $Oz$ is parallel to the direction [001]). We started with study the accuracy of MTPs fitting w.r.t. size of the unit cell, or, how many atoms in configurations do we need for accurate and reliable fitting of MTP. For this aim we created fully relaxed configurations with orthogonal box of size from $(9.9, 9.9, 9.9)$ Å to $(26.4, 26.4, 26.4)$ Å (thus, we increased the number of atoms from 54 to 1024). Each training set includes 33 configurations of the same size: 1 fully relaxed configuration, 10 extended/compressed configurations with bulk deformations, 12 extended/compressed configurations with different shear deformations ($xx$, $yy$, $zz$, $yz$, $xz$, $xy$), and 10 configurations with random displacements of each atom in the fully relaxed configuration. The quantities of bulk and shear deformations, as well as the intervals of atomic perturbations (from which the random values of atomic displacements were chosen) are shown in the Table. 1. All the results below are demonstrated for 2 % shearing (extension/compression) of lattice vectors, and for random displacement of each atom from the interval $(-0.2, 0.2)$ Å. Refence energies, forces, and stresses in the training sets were calculated with the Farkas EAM potential [1]. This potential predicts different properties of Nb-bcc with high accuracy, in particular, the elastic constants are close to the experimental ones.

We fit MTPs of three levels: 8 ($N_{\text{lin.}} = 9$), 12 ($N_{\text{lin.}} = 29$), and 16 ($N_{\text{lin.}} = 92$). For all the MTPs we choose $R_{\text{cut}} = 5$ Å, $N_{\text{polyn.}} = 8$, therefore, the total of parameters in these MTPs are 26, 54, 125, respectively. We denote these potentials by MTP-26, MTP-54, and MTP-125. The weights in the optimization problem (12) are $w_{\text{e}} = 1$, $w_{\text{f}} = 0.01$, and $w_{\text{s}} = 0.001$, because it was shown in [5], that the weights close to these allow to predict energies, forces, and stresses with high accuracy.

For each level of MTP we fit an ensemble (or, group) of five potentials in order to be able to estimate uncertainty of MTP prediction of different values (e.g., errors, elastic

constants, etc.) due to randomness of the fitting, i.e., we always start from a random set of MTP parameters for each fitting, and, thus, each MTP trained could be considered as a random quantity. We considered five potentials for each of the three ensembles, because, first, we manually checked that in each ensemble all five potentials converge to various local minimum, and, second, we compared the ensembles of five and ten potentials and verified, that the uncertainties of estimations (i.e., standard deviations of values predicted by each potential from the average value predicted by ensemble) were close to each other.

As we fit the ensembles of potentials, we calcute the average absolute energy, force, and stress fitting root-mean square errors (RMSEs)

$$\bar{E}_{\mathrm{RMSE}} = \frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\frac{E_k^{\mathrm{REF}}}{N^{(k)}} - \frac{E_k^{\mathrm{MTP}_j}}{N^{(k)}}\right)^2}, \tag{20}$$

$$\bar{F}_{\mathrm{RMSE}} = \frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\frac{\sum\limits_{i=1}^{N^{(k)}}\sum\limits_{a=1}^{3}\left(f_{i,a,k}^{\mathrm{REF}} - f_{i,a,k}^{\mathrm{MTP}_j}\right)^2}{3N^{(k)}}}, \tag{21}$$

$$\bar{S}_{\mathrm{RMSE}} = \frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\frac{\sum\limits_{a,b=1}^{3}\left(\sigma_{ab,k}^{\mathrm{REF}} - \sigma_{ab,k}^{\mathrm{MTP}_j}\right)^2}{9}}, \tag{22}$$

where $\mathrm{MTP}_j$ is the $j$-th MTP in the ensemble. The average absolute energy (20), force (21), and stress (22) fitting RMSEs for the three ensembles of MTPs and the standard deviations of MTPs' absolute errors are shown in the Figs. 5, 6, 7, respectively. From the figures we can conclude, that increasing of the MTP level decreases training errors. Mainly (except for the ensemble of MTPs with $\mathrm{lev}_{\max} = 8$ and several points for MTPs of higher levels), the standard deviations from average errors are reasonable, less than 10 %. It indicates, that the uncertainty of errors prediction is small enough, the average values of errors predicted are reliable.

The average absolute energy and force RMSEs do not signifficantly depend on size of configurations in the training sets, whereas the average absolute stress RMSEs decrease while increasing number of atoms in configurations. Finally, for the ensembles of MTP-54 and MTP-125 all the average errors for the configurations with 686 atoms and 1024 atoms are also close to each other (i.e., we observe the convergence w.r.t. number of atoms). Thus, it could be reasonable to consider configurations with 686 atoms in the training set.

For calculation of elastic constants it is necessary to find such a lattice constant at which cohesive energy of Nb-bcc structure reaches minimum. For all the three ensembles of potentials and for each potential we obtained very close dependence of cohesive

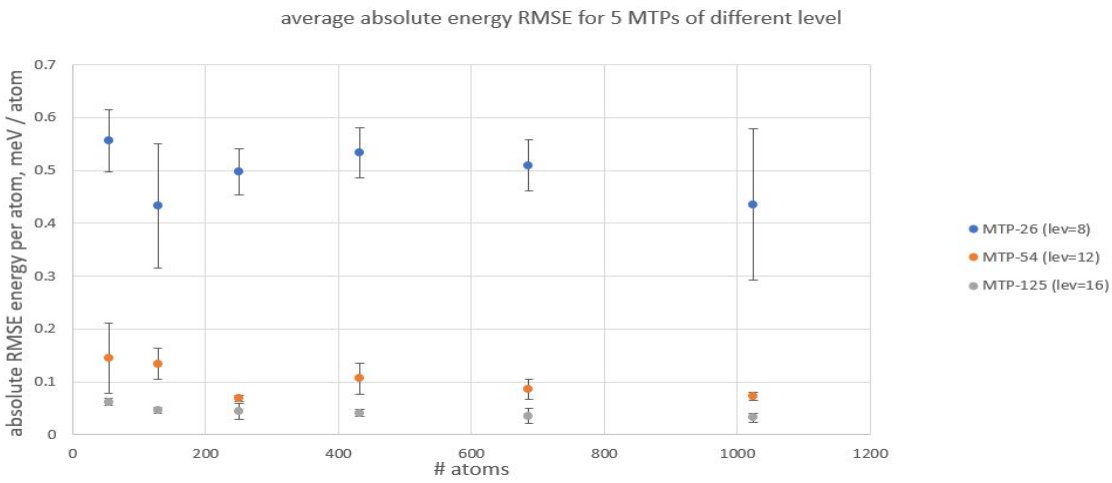


Figure 5: Average absolute energy fitting RMSEs and the standard deviations of MTPs' energy errors for different configuration sizes in the training sets. Higher the level of MTP, smaller the average energy RMSEs. The energy errors do not signifficantly depend on number of atoms in each configuration in the training set.

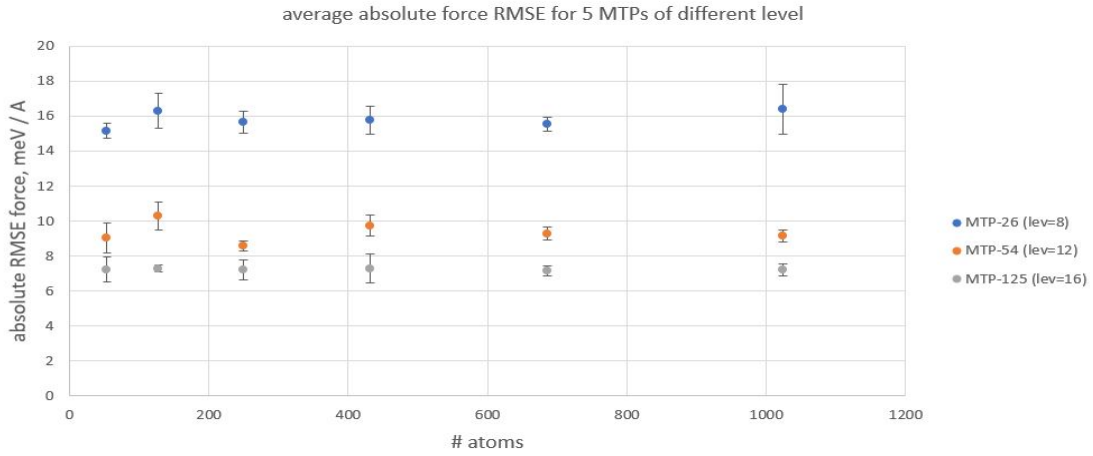


Figure 6: Average absolute force training RMSEs and the standard deviations of MTPs' force errors for different number of atoms in the training sets. Greater the number of parameters in MTP, smaller the errors. The force errors do not depend on sizes of configurations in the training set.

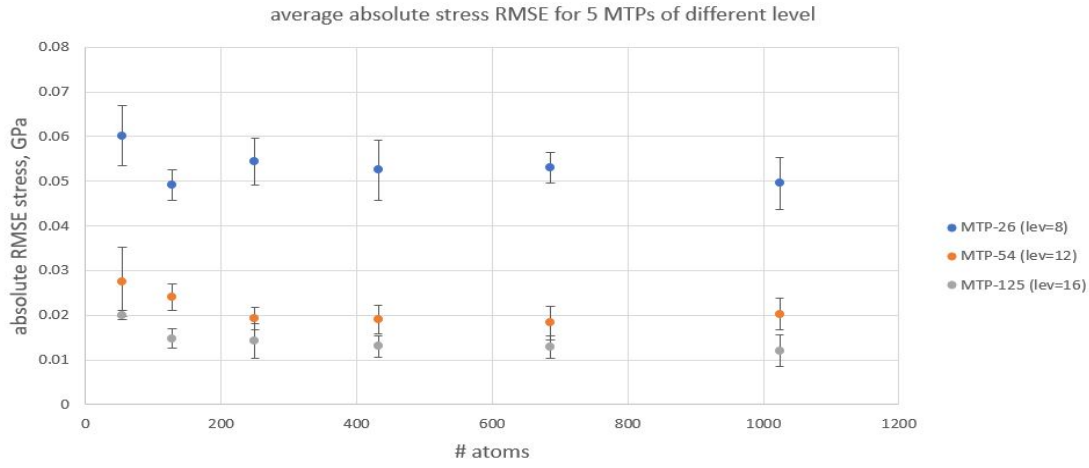


Figure 7: Average absolute stress fitting RMSEs and the standard deviations of MTPs' stress errors for different number of atoms in the training sets. Higher the level of MTP, smaller the average stress RMSEs. The stress errors decrease with increasing the number of atoms of each configuration in the training set.

energy on a lattice constant, the dependence is shown in the Fig. 8 for one of MTPs. The MTP curve is close to the reference (EAM) curve with the minimum at $a = 3.3$ Å.

Once we found the proper lattice constant of Nb-bcc structure, we calculate the elastic constants $C_{11}$, $C_{12}$, and $C_{44}$. The dependence of the average elastic constants for the three ensembles of MTPs on the size of configurations in the training sets, and the elastic constants calculated with the Farkas EAM are shown in the Fig. 9. Tthe maximum relative error in constants prediction is less than 25 %, but the constants $C_{11}$ and $C_{44}$ were better predicted with MTP-54 and MTP-125, than with MTP-26 (see the description of the Fig. 9 for details). Thus, it is better to use MTPs of high levels (12-th and 16-th) for prediction of Nb-bcc elastic properties and calculation of local extrapolation grades (on the basis of which a ML-bias potential should be constructed) of Nb configurations with screw dislocations. The maximim uncertainty (standard deviation) of the elastic constants estimation is less than 15 %. Finally, as for stress RMSEs, we observe the convergence of the elastic constants w.r.t. number of atoms for configurations with 686 and 1024 atoms. Due to this reason we used configurations with 686 atoms for the training of MTPs, on the basis of which we are trying to construct a ML-bias potential.

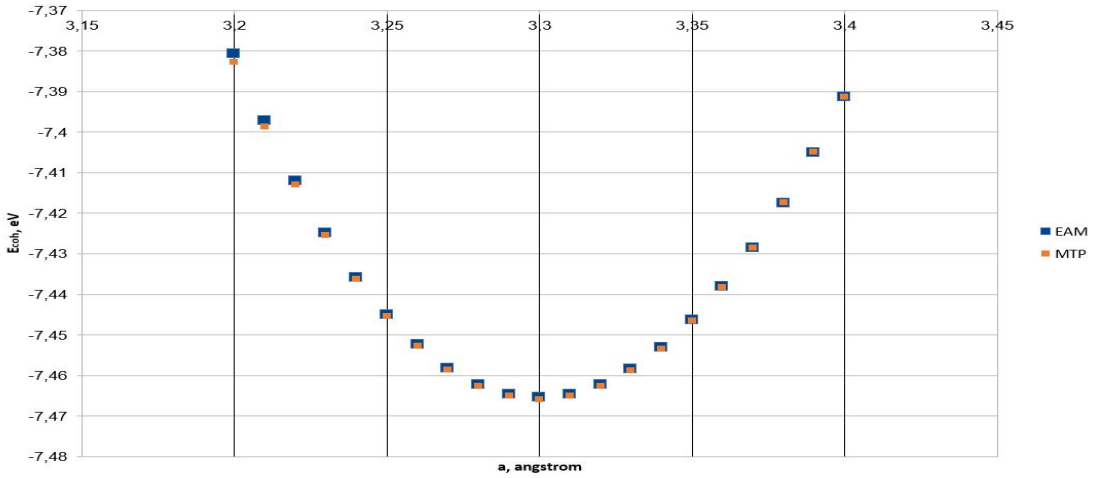


Figure 8: The dependence of cohesive energy on the lattice constant for Nb-bcc. Both the Farkas EAM potential and the MTP dependences are close to each other, cohesive energy reaches the minimum at $a = 3.3$ Å.

# 7 Extrapolation grades for structures with straight screw dislocations and kink-pairs

Before trying to construct a ML-bias potential for acceleration of a screw dislocation migration it is necessary to construct a training set which allows to differentiate local minima and a transition state, and find a threshold grade $\gamma_{\text{th}}$ at which a boosting of atoms is turned-off (or, to recognize a transition state with an extrapolation grade). For this aim we, first, discuss the process of screw dislocation migration. Screw dislocations are shown in the Fig. 10, the process of screw dislocation migration is shown in the Fig. 11. The configurations were obtained during MD simulations with the Farkas EAM potential.

As it could be seen from the figures, there are at least two types of atomic neighborhoods in any configuration with screw dislocations: neighborhoods including only bcc atoms, and neighborhoods including the atoms from the cores of dislocation. Thus, the grades of atoms from the cores of dislocations and in bcc-lattice are expected to be different. The local minimum is the configuration with straight screw dislocations (SDs). The transtition state is somewhere between the configuration with nucleation (an unstable kink-pair, KP) and the configuration with growing kink-pair (a stable one). It is necessary to stop boosting at the time frame "before" the configuration with a stable KP, thus, we should differentiate structures with unstable and stable KPs. In the language of extrapolation grades, a grade of a configuration with stable KP should be greater than a

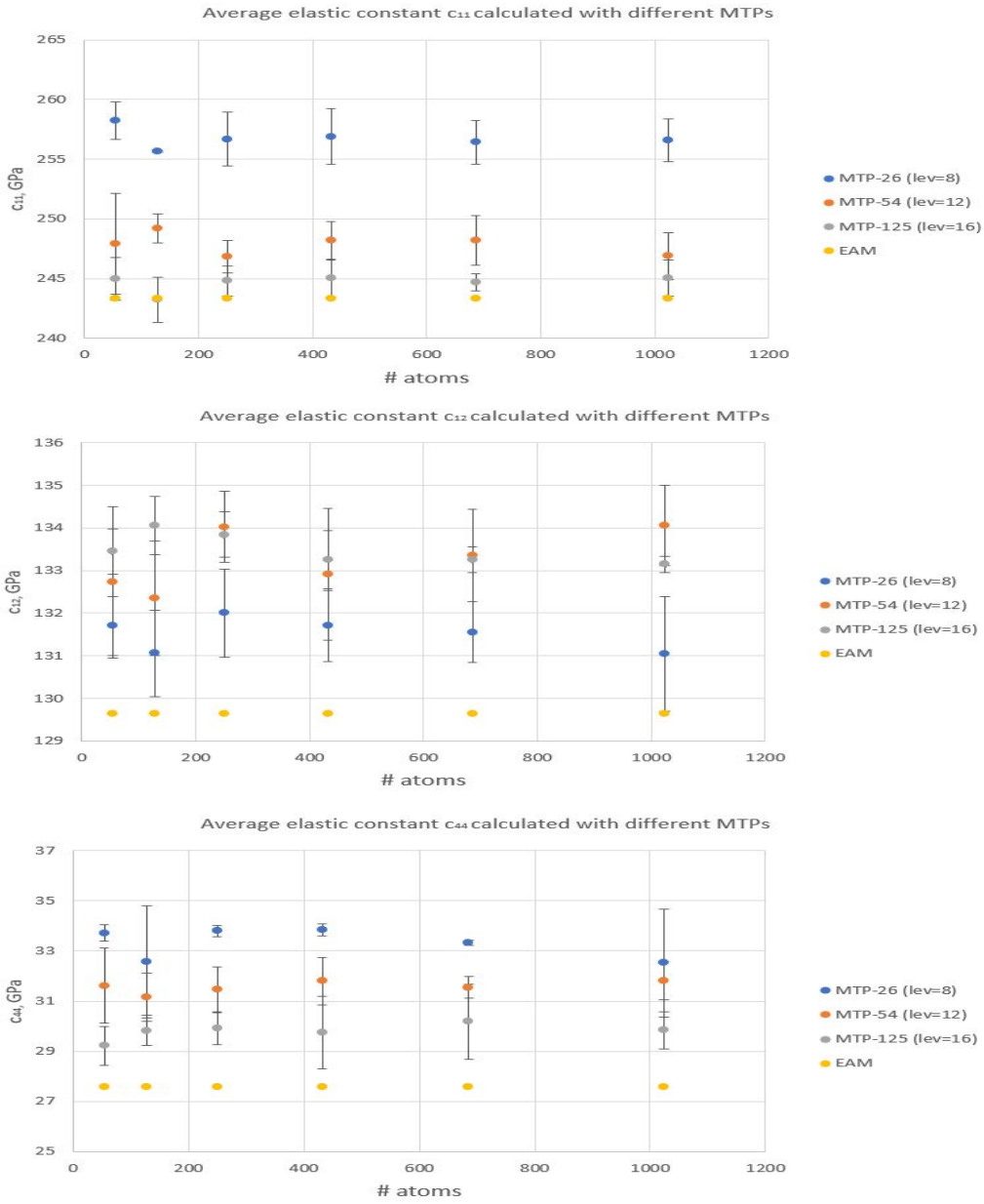


Figure 9: The dependence of the average elastic constants $C_{11}, C_{12}, C_{44}$ calculated with the three ensembles of MTPs on the size of configurations together with the constants predicted by the Farkas EAM potential. The $C_{12}$ constants obtained with MTPs of three levels are close to each other (the maximum error in the constant $C_{12}^{\mathrm{EAM}} \approx 129.5$ GPa estimation is at around 4 GPa (3 %)). The $C_{11}$, $C_{44}$ constants were better predicted with MTP-54 and MTP-125, than with MTP-26. The maximum error in the constant $C_{11}^{\mathrm{EAM}} \approx 243.4$ GPa prediction is $\approx 6$ GPa (2.5 %) for MTP-54 and MTP-125, and is approximately equal to 14 GPa (5.5 %) for MTP-26. For the constant $C_{44}^{\mathrm{EAM}} \approx 27.5$ GPa the maximum error is 3.5 GPa (12.5 %, for MTP-54 and MTP-125),and is 6.5 GPa (24 %, for MTP-26). The maximum uncertainty (standard deviation) of the elastic constants estimation is 5 GPa.

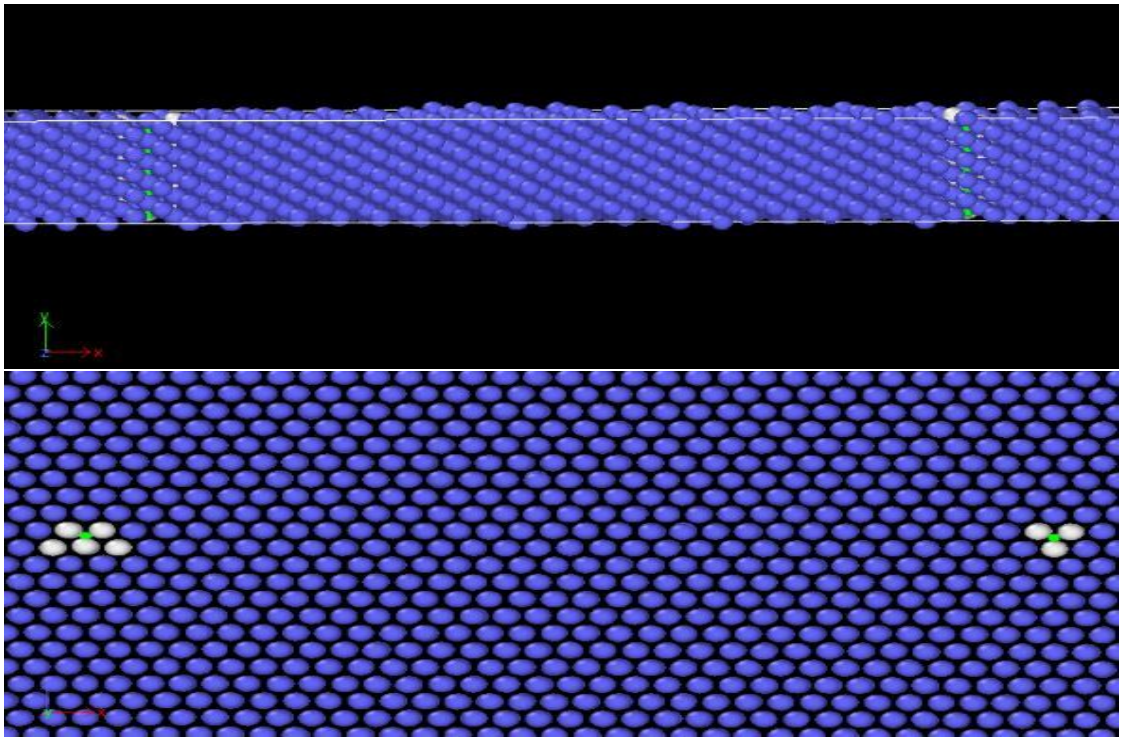

Figure 10: Screw dislocations in Nb-bcc structure: the top-view is perpendicular to the direction $[2\bar{1}1]$ (which is parallel to $Oz$), the bottom-view is perpendicular to the direction [111] (which is parallel to $Oy$), the direction $[0\bar{1}1]$ is perpendicular to $Ox$. The white atoms are the ones in the cores of screw dislocations (along green lines), the blue atoms are the bcc ones.

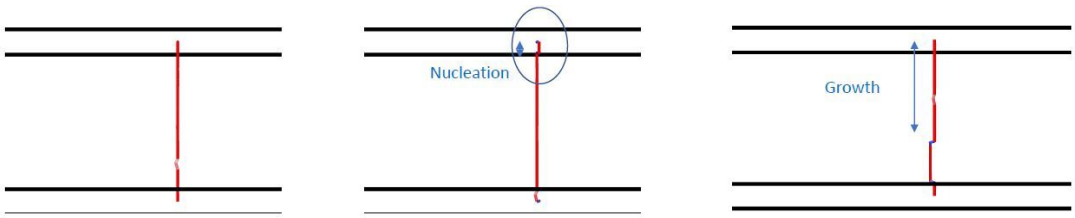


Figure 11: Process of screw dislocation migration under applied external shear stress. It starts from almost straight screw dislocation which corresponds to the local minimum (left), then the nucleation of the kink-pair takes place (middle), and, finally, the kink-pair is growing (right), this state is close to the transition state after which the whole dislocation jumps to the right side completely.

grade of a configuration with unstable KP, and a grade of a configuration with straight SDs. A threshold grade should be close to a grade of stable KP.

In order to investigate the possibility to satisfy all the above-mentioned conditions we created two different types of training sets: a training set with 33 configurations of 686 atoms (or, the training set without screw dislocations at $T = 0$ K described in the previous section), and a training set with a configuration including two straight screw dislocations with opposite Burger vectors (or, dipoles) at $T = 75$ K (see the Fig. 10). We used configurations with two dipoles because this allows to apply 3D periodic boundary conditions required for ML-bias potentials. The configuration with screw dislocations is of 14400 atoms. We emphasize, that many MTPs of 8-th, 12-th, and 16-th level with cut-off radia varying from 3.6 Å to 6.2 Å were fitted on the training sets mentioned above, but here we present only the results for MTPs-54 and MTPs-125, because the training errors and elastic constants were reasonable for these MTPs (see the previous section). We also note, that for each training set we fitted an ensemble of five MTPs with the same parameters which are not optimized (like, level, cut-off radius, etc.), but with a random vector of parameters to be optimized. After training we calculated the average errors (20), (21), (22) for the ensemble of MTPs and chose the one with the errors closest to the average ones (i.e., we choose the "representative" MTP from the ensemble and demonstrate all the results for this MTP). We compared extrapolation grades of the series of configurations obtained during a simulation illustrated in the Fig. 11.

First we discuss the results obtained with MTPs-54 ($\mathrm{lev}_{\max} = 12$) fitted on the training set including configuration with two sctraight screw dislocations. We considered MTPs with $R_{\mathrm{cut}} = 3.6$ Å and $R_{\mathrm{cut}} = 5$ Å. The extrapolation grades calculated with these MTPs for the configurations corresponding to the local minima and the transition state (we refer a set of these configurations to as a migration set 1, $\mathrm{MS}_1$) are shown in the Table 2. The series of these configurations obtained at $T = 75$ K during MD runs under shearing with the strain rate $10^8\ s^{-1}$. We compared the average grades of the migrating core of dislocation $\bar{\gamma}_{\mathrm{core}}$, maximum grades of the core of dislocation $\gamma_{\mathrm{core}}^{\max}$, and maximum grades of the bcc atoms $\gamma_{\mathrm{bcc}}^{\max}$. As it could be seen, the average grades are close to each other for both MTPs, but the MTP with $R_{\mathrm{cut}} = 3.6$ Å better differentiates the bcc and the core atoms (the difference between the maximum core and bcc grades is mainly greater). Therefore, it is better to use MTP with smaller cut-off radius for grade calculation. However, no one MTP among these two could differentiate the local minimum and the transition state, thus, a threshold grade $\gamma_{\mathrm{th}}$ could not be determined.

As the next step, we fitted MTP-125 ($\mathrm{lev}_{\max} = 16$) with $R_{\mathrm{cut}} = 3.6$ Å on the training sets (TS) with and without straight SDs. The grades of the configurations in the $\mathrm{MS}_1$ for two MTPs-125 fitted and for the MTP-54 with $R_{\mathrm{cut}} = 3.6$ Å fitted on the training set with straight SDs are shown in the Table 3. The MTPs-125 allow to recognize the transition state (i.e., the extrapolation grade is maximum near the transition state), a threshold grade $\gamma_{\mathrm{th}}$ could be determined and used as a measure for stop boosting.

In order to check whether MTPs of $\mathrm{lev}_{\max} = 16$ always allow to differentiate local

| Description | $\bar{\gamma}_{\text{core}}$ | $\gamma_{\text{core}}^{\max}$ | $\gamma_{\text{bcc}}^{\max}$ | $R_{\text{cut}}$, Å |
|---|---|---|---|---|
| 2 straight SDs | 0.80 | 5.0 | 1.5 | 3.6 |
| | 0.75 | 2.3 | 1.2 | 5.0 |
| KP nucleation | 0.92 | 3.2 | 1.6 | 3.6 |
| | 0.98 | 5.6 | 1.7 | 5.0 |
| KP growth | 0.71 | 2.2 | 1.5 | 3.6 |
| | 0.78 | 1.5 | 1.9 | 5.0 |
| Jump, 2 SDs | 0.84 | 4.9 | 1.3 | 3.6 |
| | 0.84 | 2.4 | 1.4 | 5.0 |

Table 2: Comparison of average grades $\bar{\gamma}_{\text{core}}$ and maximum grades $\gamma_{\text{core}}^{\max}$ of the migrating core of dislocation, and maximum grades of bcc atoms $\gamma_{\text{bcc}}^{\max}$ calculated using MTPs-54 with $R_{\text{cut}} = 3.6$ Å and $R_{\text{cut}} = 5$ Å. A set of the configurations for which the grades were calculated obtained at $T = 75$ K under shearing with the strain rate $10^8\ s^{-1}$. The set is called a migration set 1, $\text{MS}_1$. The MTP with smaller cut-off radius better differentiates the bcc and the core atoms (the difference between $\gamma_{\text{core}}^{\max}$ and $\gamma_{\text{bcc}}^{\max}$ is mainly greater). The average grades $\bar{\gamma}_{\text{core}}$ are close to each other for all the configurations in $\text{MS}_1$. The transition state (the KP growth) could not be reconsized by these potentials (the extrapolation grades do not reach maximum in the transition state).

minima and transition state, we created another migration set ($\text{MS}_2$) starting from different initial velocities in MD simulation. The extrapolation grades for the configurations in the $\text{MS}_2$ calculated with the MTPs-125 are represented in the Table 4. From the Table 4 it could be concluded that the MTPs-125 do not allow to recognize the transition state in the $\text{MS}_2$, and, therefore, the extrapolation grade could not be reliably used as a measure for stop boosting.

Taking into account the above limited results we conclude, that usage of the extrapolation grade as a measure for the differentiation of staraight SDs and KPs, and for stop boosting is difficult for the problem of screw dislocation migration. Therefore, a construction of a ML-bias potential on the basis of extrapolation grade would require future additional work. A probable reason is that the neighborhoods of straight SDs and KPs are similar and, thus, the extrapolation grades are also similar.

# 8 Outlook

In this section we provide suggestions for future work. For further development of ML-bias potential for add-atom hyperdynamics it could be useful to consider other temperatues (not only $T = 700$ K) and other positions of additional atom, and study the conditions under which simple hoping of add-atom or exchange mechanism dominate. Then, it could be meaningful to investigate whether different levels of MTPs improve

| Description | $\bar{\gamma}_{\text{core}}$ | $\gamma_{\text{core}}^{\text{max}}$ | $\gamma_{\text{bcc}}^{\text{max}}$ | $\text{lev}_{\text{max}}$, TS |
|---|---|---|---|---|
| 2 straight SDs | 0.75 | 5.0 | 1.5 | 12, TS with SDs |
| | 15.7 | 432 | 348 | 16, TS without SDs |
| | 12.8 | 26 | 19 | 16, TS with SDs |
| KP nucleation | 0.98 | 3.2 | 1.6 | 12, TS with SDs |
| | 54.0 | 1157 | 192 | 16, TS without SDs |
| | 13.3 | 36 | 15 | 16, TS with SDs |
| KP growth (1) | - | - | - | 12, TS with SDs |
| | 1318 | 90596 | 87315 | 16, TS without SDs |
| | 17 | 216 | 175 | 16, TS with SDs |
| KP growth (2) | 0.78 | 2.2 | 1.5 | 12, TS with SDs |
| | 20.1 | 343 | 231 | 16, TS without SDs |
| | 13.1 | 25 | 16 | 16, TS with SDs |
| Jump, 2 SDs | 0.78 | 4.9 | 1.3 | 12, TS with SDs |
| | 137 | 6023 | 41 | 16, TS without SDs |
| | 14 | 71 | 8 | 16, TS with SDs |

Table 3: Comparison of the grades for the configurations in the $\text{MS}_1$ calculated with the MTP-54 (TS with SDs) and with the MTPs-125 (TS with and without SDs). Transition state could be easily recognized with the MTPs-125 (KP growth (1)) and $\gamma_{\text{core}}^{\text{max}}$, it does not matter whether to include the configuration with SDs in TS, or not. A threshold grade $\gamma_{\text{th}}$ could be determined.

| Description | $\bar{\gamma}_{\text{core}}$ | $\gamma_{\text{core}}^{\max}$ | $\gamma_{\text{bcc}}^{\max}$ | $\text{lev}_{\max}$, TS |
|---|---|---|---|---|
| 2 straight SDs | 36 | 1711 | 1202 | 16, TS without SDs |
| | 13 | 44 | 30 | 16, TS with SDs |
| KP nucleation | 12 | 345 | 287 | 16, TS without SDs |
| | 12 | 23 | 19 | 16, TS with SDs |
| KP growth (1) | 39 | 1025 | 603 | 16, TS without SDs |
| | 12.5 | 39 | 24 | 16, TS with SDs |
| KP growth (2) | 26 | 572 | 297 | 16, TS without SDs |
| | 12 | 33 | 16 | 16, TS with SDs |
| KP growth (3) | 17 | 413 | 357 | 16, TS without SDs |
| | 11 | 28 | 17 | 16, TS with SDs |
| Jump, 2 SDs | 17 | 204 | 91 | 16, TS without SDs |
| | 11 | 20 | 12 | 16, TS with SDs |

Table 4: Comparison of the grades for the configurations in the $MS_2$ calculated with with the MTPs-125 (TS with and without SDs). Transition state could not be reliably recognized with the MTPs-125 (the grades of KP growth (1)-(3) are mainly smaller than the grades of 2 SDs). A threshold grade $\gamma_{\text{th}}$ could not be determined.

the results of boosting with ML-bias potential (for the moment the ML-bias potential for add-atom hyperdynamics was created on the basis of MTP with $\text{lev}_{\max} = 8$), and it is necessary to create other training sets and fit MTPs on them (e.g., with various random perturbations of atoms in initial non-perturbed configuration). Finally, it could be useful to pay more attention on selection of ML-bias energy $\Delta V_{l,\max}^{\text{ML}-\text{b}}$ (in this report we consider only the power function for the bias energy).

In the context of development of ML-bias potential for acceleration of a screw dislocation migration one can try other criterion (measure) for stop boosting than a threshold of extrapolation grade (e.g., maximum strain of the bond). Next, it could be meaningful to investigate a main direction of ML-bias forces which could be crucial for effective acceleration of a SD migration, and try to apply these forces to atoms in the core of SD. Finally, Nb configurations of other crystallographic directions could be analyzed.

# 9 Acknowledgements

The author thanks Prof. Dr. Blazej Grabowski for discussions and valuable advices in solving the problem of ML-bias potential construction for add-atom hyperdynamics. The author also thanks PD Dr. Nikolay Zotov for preparing the migrations sets and calculations of the extrapolation grades of configurations in these sets. The author acknowledges the financial support under the European Union's Horizon 2020 research and innovation programme (grant agreement No 639211).

# References


[1] Diana Farkas and Chris Jones. Interatomic potentials for ternary nb-ti-al alloys. Modelling and Simulation in Materials Science and Engineering, 4(1):23, 1996.

[2] SA Goreinov, IV Oseledets, DV Savostyanov, EE Tyrtyshnikov, and NL Zamarashkin. How to find a good submatrix. pages 247–256, 2010.

[3] Konstantin Gubaev, Evgeny V Podryabinkin, and Alexander V Shapeev. Machine learning of molecular properties: Locality and active learning. J. Chem. Phys., 148(24):241727, 2018.

[4] Soo Young Kim, Danny Perez, and Arthur F Voter. Local hyperdynamics. The Journal of chemical physics, 139(14):144110, 2013.

[5] II Novoselov, AV Yanilkin, AV Shapeev, and EV Podryabinkin. Moment tensor potentials as a promising tool to study diffusion processes. Computational Materials Science, 164:46–56, 2019.

[6] Evgeny V Podryabinkin and Alexander V Shapeev. Active learning of linearly parametrized interatomic potentials. Comput. Mater. Sci., 140:171–180, 2017.

[7] A.V. Shapeev. Moment tensor potentials: a class of systematically improvable interatomic potentials. Multiscale Model. Simul., 14(3):1153–1173, 2016.

[8] Arthur F Voter. A method for accelerating the molecular dynamics simulation of infrequent events. The Journal of chemical physics, 106(11):4665–4677, 1997.