Extreme Mechanics Letters 24 (2018) 1–5

Contents lists available at ScienceDirect

Extreme Mechanics Letters

journal homepage: www.elsevier.com/locate/eml

Accurate interatomic force field for molecular dynamics simulation by hybridizing classical and machine learning potentials

Peng Wang [a,b], Yecheng Shao [b,c], Hongtao Wang [b,c,*], Wei Yang [b,c,*]

[a] Materials Genome Institute, Shanghai University, Shanghai 200444, China
[b] Center for X-mechanics, Zhejiang University, Hangzhou 310027, China
[c] Institute of Applied Mechanics, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 10 July 2018
Received in revised form 14 August 2018
Accepted 15 August 2018
Available online 22 August 2018

ABSTRACT

Full atom simulations have demonstrated the scalability for billions of atoms, but still suffered from the transferability of semi-empirical interatomic potentials. We propose a dynamic multiscale molecular dynamics (MD) simulation method with both high accuracy and efficiency in interatomic force field calculation by hybridizing both classical and machine learning (ML) potentials. A dynamic procedure has been adopted by evaluating the centro-symmetry parameter of evolving microstructures during MD simulations and accordingly modifying the highly distorted regions depicted by ML potentials. Atomic force field calculation in near-perfect or perfect lattices remain sticking to the fast EAM potential, which precisely captures the long range elastic interactions. A handshaking region is introduced in order to enforce the continuity in atomic interactions. The MD simulations using a dynamic multiscale scheme can achieve the *ab-initio* accuracy without raising considerable computational cost. The foundation of this approach deeply roots in the facts that the ML method has comparable accuracy to *ab-initio* MD simulations and possesses the same order of computation complexity to the classical MD ($O(N)$). The proposed multiscale method, attaining both high accuracy and efficiency simultaneously, will pave the way for solving many material science problems involving complex interactions of both long range elastic forces and short range chemical bonding/debonding processes.

Full atom simulations have demonstrated the scalability for billions of atoms based on semi-empirical interatomic potentials, capable of unveiling complexity of material microstructures with unprecedented details [1,2]. The foundation of this approach, however, has been undermined by the transferability of interatomic potentials. For example, the embedded atom methods (EAM) [3,4], widely adopted in modelling deformation behaviours of metals, are optimized by fitting a finite set of material constants, such as the lattice constants, elastic moduli, surface energy, stable and unstable stacking fault energies, sublimation energy and vacancy formation energy. Most parameters are related to a selected equilibrium atomic arrangements [5,6], while many interesting phenomena involve structures far from equilibrium, such as the core of dislocations [7–9]. A sequential and hierarchy multiscale scheme has been proposed by incorporating *ab-initio* molecular dynamics (MD) simulation in some regions requiring higher accuracy. Several hybrid classical and quantum-mechanical MD simulation methods [10–12] have been promoted in the last decade. Though first principles calculations are only used when necessary (*e.g.* atoms close to a crack tip or a dislocation core), such methods still require high computational demands due to the large differences in efficiencies between classical ($O(N)$) and first principles ($O(N^3)$) calculations. The highly expensive computational cost puts stringent constraints on both the system size and the time duration of the *ab-initio* depicted region, the latter is characterized by a scale of a few hundred atoms for tens of picoseconds. Subtle balance between high accuracy and low cost in a single system has not been successfully achieved up to date. Recent years, there have been an upsurge in the applications of machine learning (ML) methods to several physical and material problems [13–15]. This data-driven paradigm has shown promising advantages over the classical methods, especially in accuracy. The ML potentials could "remember/interpolate", rather than recalculate, the previously "learned" first principles results with relatively low computational cost ($O(N)$) and high accuracy. The mean absolute error (MAE) between predicted forces in ML potential and the first principles results are generally less than 0.05 eV/Å [13], demonstrating great advantage in accuracy over semi-empirical interatomic potentials. Moreover, chemical reactions can be faithfully reproduced by MD simulation with ML potentials, which is rarely modelled by semi-empirical potentials [16]. On the other hand, the efficiency of ML potentials is still lower than that of classical potentials [13], preventing it from wide applications in large-scale MD simulations. In

* Corresponding authors.
E-mail addresses: htw@zju.edu.cn (H. Wang), yangw@zju.edu.cn (W. Yang).

https://doi.org/10.1016/j.eml.2018.08.002
2352-4316/

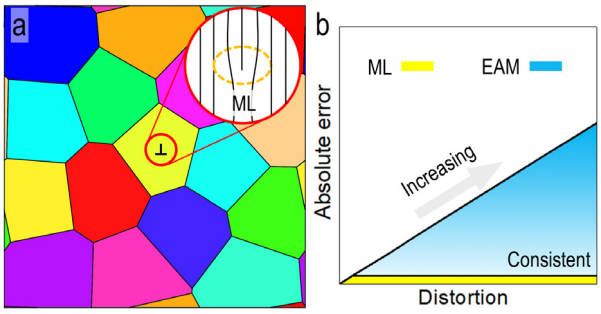

Fig. 1. (a) A schematic representation of the multiscale molecular dynamics architecture for a polycrystalline aggregate. Inset highlights a dislocation core with high distortion, identified by centro-symmetry parameter. (b) The relation between the error in force calculation and the distortion in atomic environment using ML and EAM potentials.

this work, we demonstrate that both high accuracy and efficiency can be achieved simultaneously in calculating interatomic force fields for MD simulations by hybridizing classical and machine learning potentials.

Fig. 1 shows a schematic representation of a multiscale molecular dynamics architecture for a polycrystalline aggregate. The ML potential is employed to depict atomic interactions in highly distorted regions. The accuracy is generally determined by the variation and the size of data sets derived from the first principles calculations [13–16], *i.e.* the coverage of the phase space sampling of a given material system. It is insensitive to specific distortion type across various atomic arrangements, rendering ML potentials excellent transferability within the phase space spanned by the data sets. The characteristics of bonding and debonding, atomic rearrangement and even chemical specifics can be well described by such models [17] with close accuracy to *ab-initio* simulations, while avoiding the high computational cost and complexity in modelling the electronic degrees of freedom. Atomic force field calculation in near-perfect or perfect lattices remain sticking to the fast EAM potential [2], which precisely captures the long range elastic interactions. Nevertheless, the error in atom force field calculation gradually increases with the disorder in atomic environments, as schematically shown in Fig. 1(b). Therefore, the boundary of the two regions is determined by matching the accuracy of two different methods. In this way, the whole system is modelled with one known accuracy given by the ML potential. As will be shown later (such as by Fig. 3(c)) that the MAE under the EAM potential increases monotonically with the average centro-symmetry (CS) parameter [18], a quantitative measure of the disorder in local atomic arrangements. Accordingly, the threshold value can be determined by relating the accuracy of the EAM potential to the CS parameter. A handshaking region, instead of an abrupt dividing surface, is introduced in order to enforce the continuity for atomic interactions. Two levels of representations are seamlessly coupled by a rule of force mixing:

$$F_{\text{hybrid}} = H(w-r)\frac{w-r}{w}F_{\text{ML}} + \frac{\min(r,\,w)}{w}F_{\text{EAM}} \tag{1}$$

where r is the shortest distance between the selected atom and the neighbouring ML atoms, w the thickness of the transition region, $H(r)$ the Heaviside function, and F_{ML} and F_{EAM} the forces calculated by ML and EAM potentials, respectively. To achieve higher efficiency while keeping the same accuracy, a dynamic procedure can be conveniently adopted by evaluating the CS parameter of evolving microstructures during MD simulations and accordingly modifying the highly distorted regions depicted by ML potentials. The above dynamics multiscale scheme has been implemented in

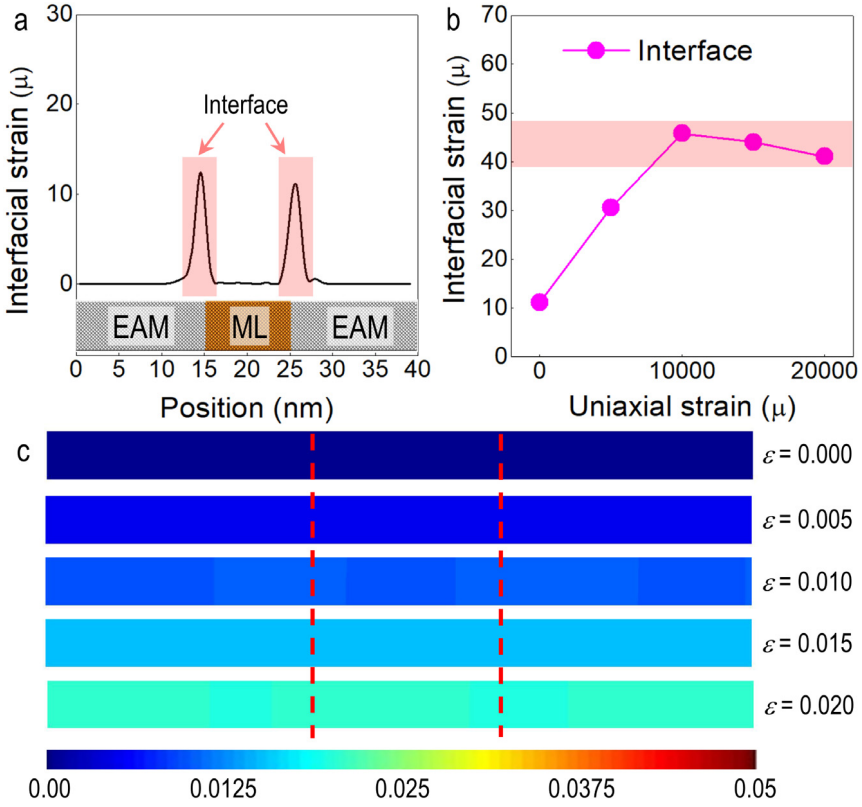

Fig. 2. (a) Strain variation along a nanowire after full relaxation using the multiscale MD simulations. The arrows indicate the locations of the numerical interfaces. (b) The variation of interfacial strain during uniaxial tension. (c) Strain distribution under different uniaxial strains of 0.000, 0.005, 0.010, 0.015 and 0.020, respectively. The middle of the handshaking region is marked by red dashed lines.

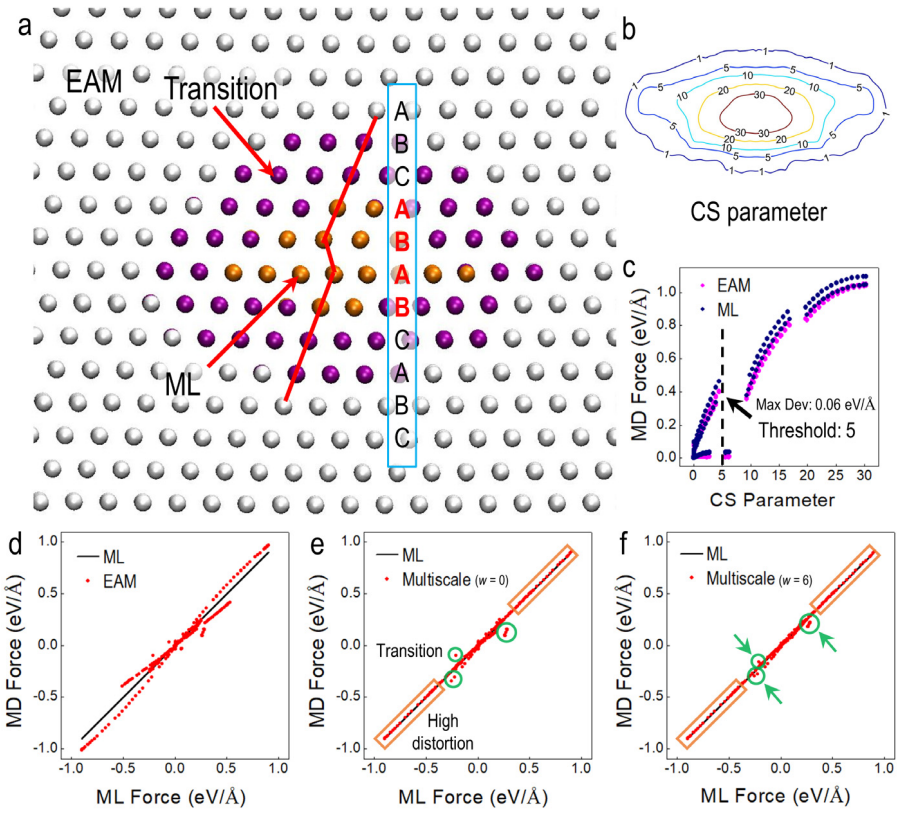

Fig. 3. (a) Atomic configuration of an edge dislocation with a Burger's vector $b = 1/2[\bar{1}10]$after fully relaxed by the multiscale MD simulations. Atomic forces are calculated using ML, EAM and mixed rules in different regions. (b) The contour of the CS parameter around the dislocation defines the boundaries of different regions in (a). (c) The scattering plot of atomic force versus the CS parameter for both ML and EAM potentials. Atom force prediction is calculated by (d) EAM model, (e) multiscale model ($w = 0$) and (f) multiscale ($w = 6$) model, taking the ML method as a reference. The dashed rectangles indicate the atom forces from the dislocation core regions.

the open source MD code of Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [19] as a single multiscale pair style. The source code and examples are provided as supplemental files.

An appropriate choice of compatible semi-empirical and ML potentials is critical for the multiscale MD method. In this way, a perfect numerical interface can be obtained when two neighbouring regions are described with different potentials. In this work, we choose the EAM potential, developed by Winey et al. [20], which well fits the lattice constant and the elastic moduli of aluminum. As a result, the calculated phonon dispersion curves show good agreement with the experimental data. The force-field ML potential, developed by Ramprasad et al. [13,21], is selected for the large diverse *ab-initio* training data set and is shown to have the high accuracy in atomic force prediction (0.05 eV/Å). In ML potentials, the atomic environment is numerically represented by the "atomic fingerprint" which is invariant to the basic atomic transformation operations. The uth component of the atomic force for atom i is predicted by comparing atom's fingerprints with a set of reference cases: [13,21]

$$F_i^u = \sum_t^{N_t} \alpha_t \cdot \exp\left[-\frac{(d_{i,t}^u)^2}{2l^2}\right] \tag{2}$$

where t labels each reference atomic structure, $d_{i,t}^u$ is the Euclidean distance between configuration i and t in fingerprint space. α_t and l are coefficients determined in training process.

The system is a "sandwiched" nanowire subjected to uniaxial tension strain. The middle part is calculated by the ML potential while two ends are modelled using the EAM potential (Fig. 2(a)). In the fully relaxed configuration, the interface coherency is well retained with small interfacial strain (∼1e−5) due to the difference in fitting parameters between two independently developed potentials. The interfacial strain increases under tension and reaches a plateau of ∼5e−5 around tension strain about 1% (Fig. 2(b–c)). Since the magnitude of interfacial strain is several orders less than the tension strain, it is expected that such artificial numerical interfaces may bare little influence to the simulation results.

The core of a straight dislocation in aluminum is further investigated to exemplify the multiscale scheme. Fig. 3(a) shows the fully relaxed configuration of an edge dislocation, which is identified by the local change of the {111} stacking sequence. The dislocation core region is delineated by the contours of the quantitative atomic distortion according to a sequence of the CS parameter (Fig. 3(b)). The core is highly distorted and localized due to the high stacking fault energy. Smooth transition to perfect lattice is observed outside the centre region. As marked in Fig. 3(a), the system is partitioned into three domains modelled by ML, EAM and force mixing methods, respectively. Fig. 3(c) shows the scattering plots of the atom force versus the CS parameter. The deviation between the two results from the ML and EAM methods enlarges monotonically with distortion in local atomic arrangements. Taken 0.05 eV/Å as the upper limit of acceptable force calculation error, the threshold CS parameter is uniquely determined and so is the ML-modelled core region. It is noted that the interactions between metallic atom and the 6th nearest neighbour are generally negligible. The hand-shaking region thickness is chosen to be 6 Å in order to balance both efficiency and accuracy. For comparison, the same system has also been investigated by a single ML or EAM method. All results are plotted in Fig. 3(d–f), taking atom forces from the ML method as a reference for the excellent agreement with first principles calculations (MAE < 0.05 eV/Å) [13]. The dashed rectangles indicate the atom forces of the dislocation core regions (marked in Fig. 3(a)). Limited by the transferability, the EAM potential fails to accurately

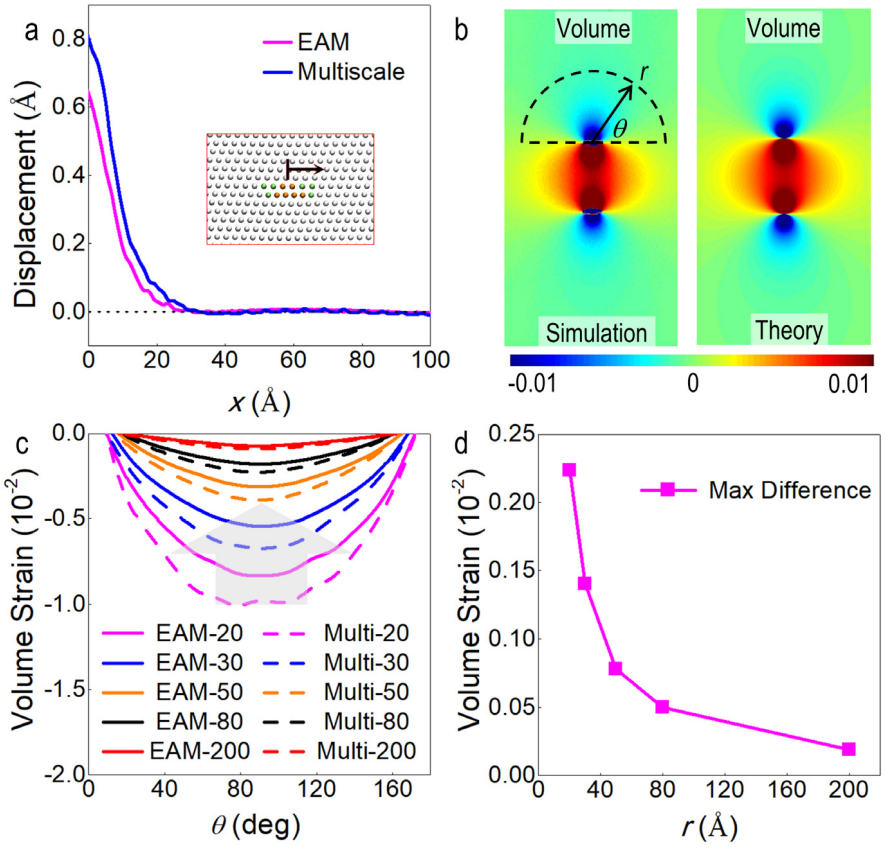

Fig. 4. (a) The relative atomic displacements from the centre of the dislocation core to the perfect lattice region. (b) The volume strain distribution of a dislocation dipole. Left: Simulation result with multiscale method. Right: Theoretical result based on solid mechanics. (c) The variation of volume strain as a function of the polar coordinates defined in (b). (d) The difference in the volume strain decreases slowly with the distance from the dislocation core.

Table 1
Computational cost for 10 ps.

Methods	Simulation time (s)	Normalization
EAM	1781.7	1
Machine Learning	137395.6	77.1
Multiscale	3759.4	2.11

model the highly distorted region (Fig. 3(d)). As expected, such deviation can be erased by using the proposed multiscale scheme (Fig. 3(e–f)). The errors in atom force calculation come mainly from the EAM component of the handshaking region. Clearly, smooth transition and improved accuracy can be achieved with a large thickness (w) by considering Eq. (1). An estimation of the relative computational cost is listed in Table 1. Given the same simulation setup, the requirements of both accuracy and efficiency have been achieved at the same time in MD simulations using the proposed multiscale scheme. It is noted that this balance can hardly be attained by the hybrid classical ($O(N)$) and quantum-mechanical ($O(N^3)$) MD simulation methods due to the intrinsic difference in the computational complexity. In contrast, both ML and EAM MD methods have the same computational demands ($O(N)$). But the efficiency of force field calculation using ML potentials is still much lower than that of EAM potentials. In general, the system size for ML method is in the order of tens of thousands of atoms. Therefore, hybridizing classical and machine learning potentials not only breaks through the transferability limitation of the semi-empirical potentials, but also retains the scalability. Moreover, the bonding and debonding process can be modelled by the ML method as well. This will greatly extend the domains of classical MD methods.

The accuracy of force calculation has substantial effect on the atomic configuration of defects. Compared in Fig. 4(a) is the relative atomic displacements from the centre of the dislocation core to the perfect lattice region, as indicated in the inset. The discrepancy is gradually enlarged near the dislocation core. The maximum error is about 30% between two different schemes. For the long range nature of the elastic deformation, the stress field outside the core region can be different as well. This subtle difference can be revealed by a dislocation dipole as shown in Fig. 4(b). The volume strain is measured in the polar coordinate system centred at one dislocation core. The strain distributions are similar for the two systems modelled by the multiscale and EAM methods, respectively. But the magnitude has a discrepancy as large as 20% outside the core region (Fig. 4(c)), which decreases slowly with the distance from the dislocation core (Fig. 4(d)). The above observation signifies the critical role of ML potential in accurately modelling the highly distorted region.

In this work, we have implemented a multiscale pair in the LAMMPS package by hybridizing ML and EAM potentials. The requirements in accuracy and efficiency have been attained simultaneously. The foundation for the proposed multiscale scheme roots in the facts that the ML method has demonstrated comparable accuracy to *ab-initio* MD simulations and same order of the computation complexity to the classical MD ($O(N)$). We note that successful employment of this method requires appropriate choice of compatible semi-empirical and ML potentials, which is critical to achieve a perfect numerical interface and minimize the artificial influence to the defect configuration. As atomic simulations have demonstrated the scalability for billions of atoms, the capability can be greatly extended by unveiling atomic details with *ab-initio* accuracy without increasing too much computational cost. Since

the requirement of both large scale and high accuracy can be met at the same time, this opens a door to many material science problems, *e.g.* phase nucleation and forest hardening, which involve complex interactions of both long range elastic forces and short range chemical bonding/debonding.

Acknowledgements

H. Wang acknowledges the financial support from the Natural Science Foundation of China (Grant No. 11725210) and the Fundamental Research Funds for the Central Universities, China (Grant No. 2018XZZX001-05).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eml.2018.08.002.

References

[1] D.L. McDowell, A perspective on trends in multiscale plasticity, Int. J. Plast. 26 (9) (2010) 1280–1309.
[2] P. Wang, S. Xu, J. Liu, X. Li, Y. Wei, H. Wang, H. Gao, W. Yang, Atomistic simulation for deforming complex alloys with application toward TWIP steel and associated physical insights, J. Mech. Phys. Solids 98 (2017) 290–308.
[3] M.S. Daw, M.I. Baskes, Semiempirical, Quantum mechanical calculation of hydrogen embrittlement in metals, Phys. Rev. Lett. 50 (17) (1983) 1285–1288.
[4] M.S. Daw, M.I. Baskes, Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals, Phys. Rev. B 29 (12) (1984) 6443–6453.
[5] P. Wang, Y. Wu, J. Liu, H. Wang, Impacts of atomic scale lattice distortion on dislocation activity in high-entropy alloys, Extreme Mech. Lett. 17 (2017) 38–42.
[6] P. Wang, H. Wang, Meta-Atom molecular dynamics for studying material property dependent deformation mechanisms of alloys, J. Appl. Mech-Trans. ASME 84 (11) (2017).
[7] W. Rui, W. Shaofeng, W. Xiaozhi, Edge dislocation core structures in FCC metals determined from ab initio calculations combined with the improved Peierls–Nabarro equation, Phys. Scr. 83 (4) (2011) 045604.
[8] S.L. Frederiksen, K.W. Jacobsen, Density functional theory studies of screw dislocation core structures in bcc metals, Phil. Mag. 83 (3) (2003) 365–375.
[9] D. Chen, Y. Kulkarni, Entropic interaction between fluctuating twin boundaries, J. Mech. Phys. Solids 84 (2015) 59–71.
[10] G. Csanyi, T. Albaret, M.C. Payne, A. De Vita, Learn on the fly: a hybrid classical and quantum-mechanical molecular dynamics simulation, Phys. Rev. Lett. 93 (17) (2004) 175503.
[11] R.C. Walker, M.F. Crowley, D.A. Case, The implementation of a fast and accurate QM/MM potential method in Amber, J. Comput. Chem. 29 (7) (2008) 1019–1031.
[12] G. Monard, M. Loos, V. Thery, K. Baka, J.L. Rivail, Hybrid classical quantum force field for modeling very large molecules, Int. J. Quantum Chem. 58 (2) (1996) 153–159.
[13] V. Botu, R. Batra, J. Chapman, R. Ramprasad, Machine learning force fields: construction, validation, and outlook, J. Phys. Chem. C 121 (1) (2017) 511–522.
[14] M. Caccin, Z. Li, J.R. Kermode, A. De Vita, A framework for machine-learning-augmented multiscale atomistic simulations on parallel supercomputers, Int. J. Quantum Chem. 115 (16) (2015) 1129–1139.
[15] N. Artrith, A. Urban, An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO_2, Comput. Mater. Sci. 114 (2016) 135–150.
[16] J. Behler, Representing potential energy surfaces by high-dimensional neural network potentials, J. Phys.: Condens. Matter 26 (18) (2014) 183001.
[17] V. Botu, R. Batra, J. Chapman, R. Ramprasad, Machine learning force fields: construction, validation, and outlook, J. Phys. Chem. C 121 (1) (2017) 511–522.
[18] C.L. Kelchner, S.J. Plimpton, J.C. Hamilton, Dislocation nucleation and defect structure during surface indentation, Phys. Rev. B 58 (17) (1998) 11085–11088.
[19] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. 117 (1) (1995) 1–19.
[20] J.M. Winey, K. Alison, Y.M. Gupta, A thermodynamic approach to determine accurate potentials for molecular dynamics simulations: thermoelastic response of aluminum, Model. Simul. Mater. Sci. Eng. 17 (5) (2009) 055004.
[21] T.D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, R. Ramprasad, A universal strategy for the creation of machine learning-based atomistic force fields, NPJ Comput. Mater. 3 (1) (2017) 37.