TUDelft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Restrepo, S. E., Mohandas, N. K., Sluiter, M. H. F., & Paxton, A. T. (2025). Applicability of universal machine learning interatomic potentials to the simulation of steels. *Modelling and Simulation in Materials Science and Engineering, 33*(3), Article 035003. https://doi.org/10.1088/1361-651X/adb483

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

PURPOSE-LED
PUBLISHING™

PAPER • OPEN ACCESS

Applicability of universal machine learning interatomic potentials to the simulation of steels

View the article online for updates and enhancements.

You may also like

Applicability of universal machine learning interatomic potentials to the simulation of steels

Sebastián Echeverri Restrepo[1,2] ⓘ**, Naveen K Mohandas**[3] ⓘ**, Marcel H F Sluiter**[3,*] ⓘ **and Anthony T Paxton**[2,4] ⓘ

[1] SKF Research & Technology Development (RTD), SKF B.V., Meidoornkade 14, 3992 AE Houten, The Netherlands
[2] Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom
[3] Department of Materials Science and Engineering, Delft University of Technology, Mekelweg 2, Delft 2628 CD, The Netherlands
[4] Department of Materials, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

E-mail: M.H.F.Sluiter@tudelft.nl

CrossMark

Abstract
Bearing steels are complex materials composed of an iron matrix and a well defined and precise amount of several alloying elements. In order to improve sustainability and circularity, there is a tendency to increase the utilisation of scrap material for their production. The variability of the composition of scrap material has a direct impact on the properties of the final steels: There is less control on their composition due to the possible presence of larger amounts of tramp and alloying elements. One way to study the effect of tramp elements is by using universal machine learning interatomic potentials. These types of potential render the investigation of multi-element systems possible. They permit the study of interactions between iron atoms in the matrix and multiple concurrent tramp and alloying elements, a feature that is currently not available in classical potentials. In this work, we present a benchmark of four state-of-the-art universal machine learning interatomic potentials (Crystal Hamiltonian

---

* Author to whom any correspondence should be addressed.

Graph Neural Network (Deng *et al* 2023 *Nat. Mach. Intell.* **5** 1031–41) (*v0.2.0* and *v0.3.0*), Materials 3-body Graph Network (Chen and Ping Ong 2022 *Nat. Comput. Sci.* **2** 718–28), Multiple Atomic Cluster Expansion (Batatia *et al* 2022 *Advances in Neural Information Processing Systems* vol 35 pp 11423–36)) and SevenNet (Park *et al* 2024 *J. Chem. Theory Comput.* **20** 4857–68), and study their applicability to the simulation of systems relevant to steels. For pure Fe, all potentials accurately predict the equilibrium lattice parameter, but the accuracy varies for other properties. For most solute–solute and solute–vacancy interactions all interatomic potentials tend to capture the general trends though there is a disparity in the predicted magnitudes. While currently 'off-the-shelf' universal machine learning interatomic potentials fail to predict some key properties, some of them show significant potential to serve as starting point for further training and refinement.

Keywords: interatomic potential, machine learning, steel modelling, solute atoms, diffusion, point defects

1. Introduction

The steel industry is transitioning towards a more sustainable and circular economy, emphasising increased scrap utilisation and inclusion of lower-quality scrap in steel production. This shift poses significant challenges for controlling the composition of final steel products and requires the development of steels that are more robust to compositional variability.

To understand the impact of various elements on steel processability and final properties, atomic-scale simulations offer invaluable insights. These simulations allow for the investigation of fundamental phenomena that govern macroscopic material behaviour. While *ab-initio* methods like density functional theory (DFT) are suitable for studying phenomena in small systems, larger-scale phenomena, such as dislocations and their interactions, require more extensive simulations. Recent advancements in universal machine learning potentials have emerged as promising candidates for large-scale atomic simulations.

Universal machine learning interatomic potentials (MLIPs) are interatomic potentials designed to reproduce the interactions of systems containing atoms from any combination of elements of the periodic table. Trained on extensive sets of DFT data, these potentials have demonstrated remarkable accuracy in reproducing energies and forces of test datasets, achieving average energy errors below $0.05\,\mathrm{eV\,atom^{-1}}$ [1]. Although not always the case, many universal MLIPs are now open-source, making them accessible for broader use and testing.

Applications of universal potentials extend to materials discovery, serving as preliminary filters to assess the stability of candidate materials before performing specific DFT calculations and experimental verifications. Their performance in reproducing *ab-initio* results for equations of state, atomic and lattice relaxations, formation energies, and vibrational properties has been previously studied [2]. Their performance has also been recently reviewed for free surfaces [3].

However, the suitability of these potentials for simulating systems containing elements and processes relevant to steels remains underexplored. This study aims to address this gap by investigating the applicability of universal MLIPs for steel-related simulations.

2. Methods

In order to test the suitability of universal MLIPs for the simulation of systems relevant to steels, we prepared a number of test configurations, calculated some of their properties, and compared the results with DFT values found in the literature. In this section, we describe the details of the simulations performed.

The universal MLIPs were chosen based on their open source availability and their ranking on the Matbench Discovery leaderboard [1]. We consider four state-of-the-art pre-trained universal MLIPs, namely Crystal Hamiltonian Graph Neural Network [4] (CHGNet), Materials 3-body Graph Network [5] (M3GNet), Multiple Atomic Cluster Expansion [6] (MACE) and Scalable EquiVariance Enabled Neural Network [7] (SevenNet).

CHGNet [4] is a graph neural network-based machine-learning interatomic potential designed to model the universal potential energy surface. The architecture of the CHGNet model takes as input a crystal structure with unknown atomic charge and predicts energy, force, stress, and magnetic moments (although it is important to note that it is trained on, and can only predict, the absolute values of the magnetic moments). It employs atom and bond graphs to represent pairwise bond and angle information, which are processed through basis expansions and embedding layers to create features for atoms, bonds, and angles. These features are updated through interaction blocks, and properties are predicted at output layers using fully connected layers. The interaction blocks allow atoms, bonds, and angles to share and update information, while the atom convolution layer calculates neighbouring atom and bond information through weighted message passing. One of the main differences between CHGNet and other universal potentials is the incorporation of magnetic moments, which carry information about both local ionic environments and charge distribution. For CHGNet, we consider separately two published versions (*v0.2.0* and *v0.3.0*).

M3GNet [5] is a universal MLIP based on graph neural networks with three-body interactions. This model combines the many-body features of traditional interatomic potentials with flexible graph-based material representations. The three-body angular interactions are expanded by a complete and orthogonal spherical Bessel function and spherical harmonics basis set. The model architecture begins with a position-included graph, followed by featurisation, the main block, and a readout module that outputs energy, force, and stress. Featurisation includes a graph featuriser and a many-body computation module. The graph featuriser embeds atomic numbers into a continuous feature space and expands pair-bond distances into a basis set. The many-body computation module calculates interaction indices and angles. The main block has two steps: many-body to bond, which calculates new bond information, and standard graph convolution, which updates bond, atom, and state information. In the readout stage, atom information is passed to a gated MLIP to obtain atomic energy, which sums to the total energy, with its derivatives providing force and stress outputs. In the present work, we use version *M3GNet-MP-2021.2.8-PES*.

The MACE architecture integrates equivariant message passing with many-body messages. It retains components of equivariant graph neural networks, such as element embedding with tensor decomposition and higher-order equivariant messages via tensor product operations. It uses 4-body order equivariant features in each layer, requiring two layers of message passing, and non-linear activations in the radial basis and final readout layer [6]. In the present work, we use their pretrained foundational model, MACE-MP-0 (Large, *MACE_MPtrj_2022.9.model*) [8].

SevenNet [7] is a package that implements an efficient parallelisation scheme compatible with graph neural network interatomic potentials. The neural network architecture used is identical to that of NequIP [9], it is programmed using PyTorch [10] and it leverages the

e3nn [11] library to ensure equivariance in mathematical operations. Here we use the pre-trained SevenNet-o model, version *11July2024*.

Calculations were performed using the Atomic Simulation Environment (ASE) [12], where the potentials were loaded as calculator objects. All relaxations were carried out using the ABC-FIRE minimisation method [13–15], with the convergence criterion that the force on all individual atoms should be less than $1\,\mathrm{meV\,\AA^{-1}}$ [16] while for the calculation of the stiffness tensor elements a stricter criterion was used, namely $0.1\,\mathrm{meV\,\AA^{-1}}$ [17].

2.1. Pure iron

The performance of the selected universal MLIPs was initially assessed in iron (Fe) systems, as Fe is the main constituent element in steels. This evaluation provides a baseline understanding of their ability to accurately describe Fe–Fe interactions, a crucial aspect of simulating steel alloy properties. We focus on the following properties: lattice parameter, vacancy formation energy, surface energies, and elastic properties, and, for the CHGNet potentials only, some local magnetic moment properties.

The lattice parameter (a_0^{Fe}) is calculated on cubic supercells with an underlying body-centred cubic (BCC) lattice of various sizes, ranging from $(1 \times 1 \times 1)$ to $(5 \times 5 \times 5)$ unit cells to check the robustness of the potentials. The supercells were structurally relaxed, allowing the atomic positions and the box size to change while constraining the box to remain orthogonal. The relaxed system of size $(4 \times 4 \times 4)$ unit cells was taken as input for the calculation of the vacancy formation energy (E_f^v), using the following equation:

$$(E_f^v) = E\left[\mathrm{Fe}_{n-1}\square\right] - \frac{n-1}{n}E\left[\mathrm{Fe}_n\right] \tag{1}$$

where $E[\mathrm{Fe}_{n-1}\square]$ is the total energy of a system containing $(n-1)$ Fe atoms in a BCC lattice and one vacancy, and $E[\mathrm{Fe}_n]$ is the energy of a system containing a perfect BCC Fe crystal with n atoms.

The relaxed energies of the (111), (110) and (100) surfaces on BCC Fe are also evaluated. The surfaces are generated in orthogonal supercells using the equilibrium lattice parameter previously calculated for each potential. The dimensions of the supercells are $(4 \times 4 \times 20)$ repetitive units in the initial unrelaxed cell. The surface normal directions are [111], [110] and [100] along the z direction of the simulation box. A 10 Å vacuum is inserted in the z direction to avoid interactions between opposing surfaces, due to periodicity. The energy of the system is minimised by allowing the movement of the atoms while keeping the dimensions and angles of the simulation box fixed. The energy of a surface defined by the indices ijk (E_s^{ijk}) is calculated as follows:

$$E_s^{ijk} = \frac{E\left[\mathrm{Fe}_n^{ijk}\right] - E\left[\mathrm{Fe}_n\right]}{2A^{ijk}} \tag{2}$$

where $E[\mathrm{Fe}_n^{ijk}]$ is the total energy of a system containing free surfaces with normal $[ijk]$; A^{ijk} is the area of the free surface.

For the two versions of CHGNet, we also extracted the magnetic moment of an atom located at the top layer of the surface.

The components (C_{ij}) of the stiffness tensor $(\boldsymbol{C})$ were calculated using the stress–strain method [18, 19]. It is based on Hooke's law: $\boldsymbol{\sigma} = \boldsymbol{C}\varepsilon$, where $\boldsymbol{C}$ is obtained from the computed stress tensor $\boldsymbol{\sigma}$ that results from the imposed Green–Lagrange strain tensor ε. The method involves four steps: (1) structurally relaxing the system, (2) applying six strain tensors with

four magnitudes each (-0.01, -0.005, 0.005 and 0.01), (3) relaxing ionic positions while maintaining the applied strain and computing the resulting stresses, and (4) calculating the stiffness tensor components by fitting Hooke's law via least squares.

The cubic elastic stability criteria are calculated from the components of the stiffness tensor: the bulk modulus $B = (C_{11} + 2C_{12})/3$, and the shear moduli C_{44} and $C' = (C_{11} - C_{12})/2$.

2.2. Carbon in iron

Carbon (C) is one of the most important interstitial alloying elements in steels, playing a crucial role in determining their properties and behaviour. The amount of C in steel can vary widely, and its effects on steel properties are complex and multifaceted, depending on factors such as the overall composition of the steel, processing history, and thermal and mechanical treatments applied. Carbon is also one of the most frequent foreign interstitial atoms in the Fe matrix, along with nitrogen atoms. Even a very small amount of C in interstitial positions can drastically influence steel properties through strong interactions with lattice defects. The interaction of C with dislocations and/or substitutional atoms significantly affects the yield stress and subsequent mechanical properties of the material.

We study the behaviour of C within BCC Fe by calculating the binding energy of various configurations where C atoms, and/or vacancies, are present in a BCC Fe lattice. The configurations correspond to those previously studied in [20–22], namely configurations with: one vacancy and one C atom, two vacancies and one C atom, two C atoms, and one vacancy and a C dimer, see figure 1. We put special emphasis on the configurations with one vacancy and a C dimer, since DFT calculations have shown that vacancies in both low- and high-carbon steels bind strongly with a carbon dimer [22, 23]. The binding energies, E_b, are calculated as follows,

$$E_b = E\left[\text{Fe}_{n-p}\Box_p\right] + q\left(E\left[\text{Fe}_n\text{C}\right]\right) - q\left(E\left[\text{Fe}_n\right]\right) - E\left[\text{Fe}_{n-p}\Box_p\text{C}_q\right], \qquad (3)$$

where $E[\text{Fe}_{n-p}\Box_p]$ is the total energy of a system containing $(n-p)$ Fe atoms and p vacancies, $E[\text{Fe}_n\text{C}]$ the energy of a system containing n Fe atoms and one interstitial C atom, $E[\text{Fe}_n]$ the energy of a system containing a perfect Fe crystal with n atoms, and $E[\text{Fe}_{n-p}\Box_p\text{C}_q]$ is the total energy of the system of interest, containing $(n-p)$ Fe atoms, p vacancies and q C atoms. Here and elsewhere, positive values of E_b indicate attractive interactions. In addition the energy of migration of C between octahedral interstitial positions in BCC Fe is calculated. For the systems containing a carbon dimer and a vacancy (configurations j, k, l and m), we also calculate relative energies between configurations using the following notation: $\Delta E_{r-s} = E_b^r - E_b^s$ is the energy difference between configurations r and s.

All calculations containing C and/or vacancies are performed using as a base a BCC Fe system of size $(4 \times 4 \times 4)a_0^{\text{Fe}}$, where the C atoms are inserted as interstitials, and/or where Fe atoms are removed to generate vacancies. The total energies are calculated by relaxing the atomic positions and box dimensions while constraining the box to remain orthogonal.

2.3. Solutes and vacancies

Substitutional solutes play an important role in determining the microstructure and properties of steels. These solutes interact with the iron lattice, with each other [24], with interstitial elements such as carbon [25, 26], and with lattice defects [27], thereby influencing the behaviour of the material. The interactions between substitutional solutes are known to enhance solid solution strengthening [28–31]. Solutes can exhibit short-range order also, where they tend
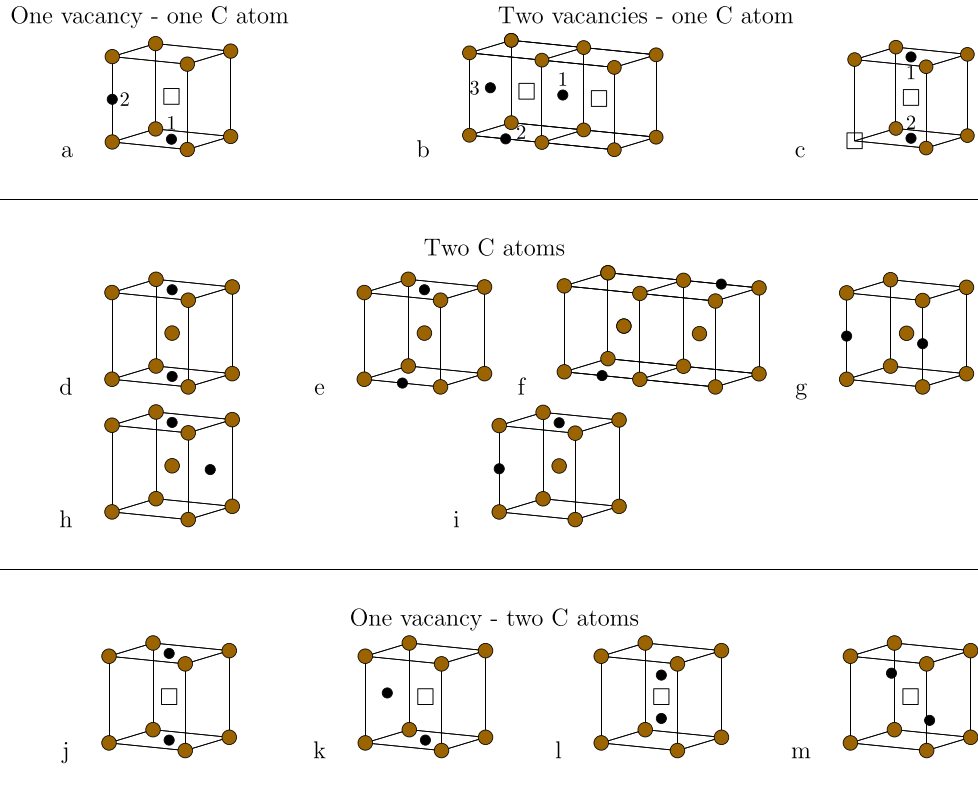
Figure 1. Configurations containing C and/or vacancies within a BCC Fe lattice used for the assessment of the potentials.

to cluster or segregate in specific arrangements within the lattice, affecting the mechanical properties [32–34]. Electronic interactions between substitutional solutes and the iron lattice can influence the electronic structure and properties, such as corrosion resistance [35, 36]. Furthermore, lattice distortion caused by substitutional solutes affects strength, ductility, and toughness [37–39]. The thermodynamic stability of various phases in steel, such as austenite, ferrite, and martensite, and the microstructure is influenced by substitutional solutes [40–42].

Interactions between vacancies and solute atoms in steels play a role in diffusion-mediated processes, such as ordering and decomposition, which occur in alloys during heat treatment and irradiation. By affecting the formation of precipitates, grain growth, and microstructure, vacancy-solute interactions can alter the strength, hardness, ductility, and corrosion resistance of steel. In addition, they can influence the movement and interaction of dislocations, which additionally has an effect on the fatigue behaviour. A thorough understanding of vacancy-solute interactions is crucial for optimising steel properties and behaviour, and for developing new steel alloys with improved performance and reliability.
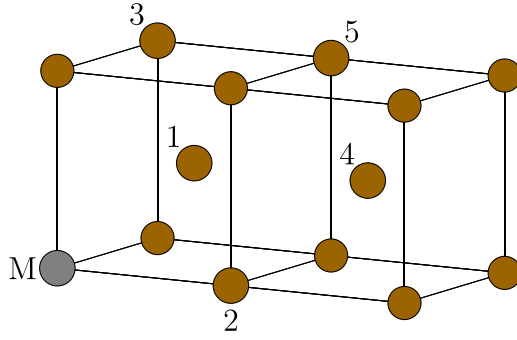
Figure 2. The positions of the first to fifth nearest neighbours of the solute M in the body-centred cubic (BCC) lattice. Site numbering as in [24] and [43].

2.3.1. Solute–solute interactions. We test the performance of the various potentials by calculating the binding energy of pairs of substitutional solute atoms in BCC Fe. We consider five distinct configurations, namely the first five nearest neighbour separations, see figure 2.

The solute–solute interactions, E_b^{ss}, are quantified using the following equation,

$$E_b^{ss} = 2E\left[Fe_{n-1}M\right] - E\left[Fe_n\right] - E\left[Fe_{n-2}M_2\right], \tag{4}$$

where $E[Fe_{n-1}M]$ is the total energy of a system containing $(n-1)$ Fe atoms and one substitutional atom of element M; $E[Fe_n]$ is the energy of a system containing (n) Fe atoms, and $E[Fe_{n-2}M_2]$ is the energy of a system containing $(n-2)$ Fe atoms and two substitutional M atoms.

The total energies $E[Fe_{n-1}M]$, $E[Fe_n]$, and $E[Fe_{n-2}M_2]$ are calculated in cubic supercells of size $(4 \times 4 \times 4)a_0^{Fe}$ (for a value of $n = 128$ atoms), where a_0^{Fe} is the lattice parameter of BCC Fe. As before, the supercells are structurally relaxed, allowing the atomic positions and the box volume and shape to change while constraining the box to remain orthogonal.

2.3.2. Solute–vacancy interactions. Analogous to the previous case of solute–solute interactions within BCC Fe, the solute–vacancy interactions are calculated for the first five neighbor distances, see figure 2. The solute–vacancy binding energy, E_b^{sv}, is defined as,

$$E_b^{sv} = E\left[Fe_{n-1}\square\right] + E\left[Fe_{n-1}M\right] - E\left[Fe_n\right] - E\left[Fe_{n-2}M\square\right] \tag{5}$$

where $E[Fe_n]$, $E[Fe_{n-1}\square]$, and $E[Fe_{n-1}M]$ are defined as before, and $E[Fe_{n-2}M\square]$ is the energy of system containing $(n-2)$ Fe atoms, one M atom, and one vacancy. The supercell size and the relaxation procedure are identical to those used for the calculation of the solute–solute interactions.

2.3.3. Solute migration. The migration energy barrier of a M atom in a Fe BCC matrix (E_m^M) is calculated using the string method [44–47] as implemented in the ASE software. A $(4 \times 4 \times 4)a_0^{Fe}$ supercell is used as before. The migration path is shown in figure 3. The generated initial and final configurations are then structurally relaxed, allowing relaxation of the atomic positions only. The string method is applied using seven images, generated using a linear interpolation between the initial and final images.
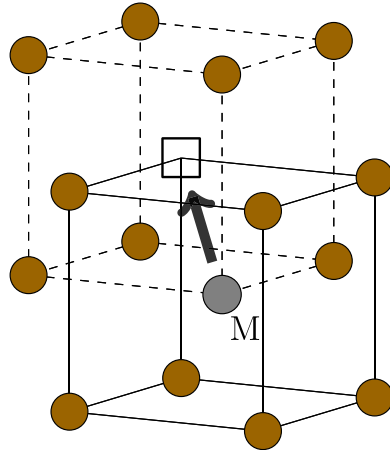
Figure 3. Route for a substitutional atom of element M (in grey) to migrate towards a vacancy (depicted as a white square) in a BCC lattice of Fe along a $\langle 111 \rangle$ direction.

3. Results

3.1. Pure iron

Results of the calculations performed in systems containing exclusively Fe atoms are presented in table 1, together with some literature values from DFT calculations for reference. All the universal MLIP potentials are able to reproduce the lattice parameter (a_0^{Fe}) of BCC Fe to within 1% of the Materials Project target value [48]. The performance of the potentials for the reproduction of the vacancy formation energy ((E_f^v)) is not good. If we take as a reference a DFT calculated value of 1.95 eV [49], the closest values are given by MACE and SevenNet-o, with values of 1.38 eV and 1.29 eV, which translate into errors of 29% and 34%, respectively. The other potentials show larger errors, namely 57% (M3GNet), 58% CHGNet version *v0.3.0* (abbreviated as CHG3), and 62% CHGNet version *v0.2.0* (abbreviated as CHG2).

The migration energy is well reproduced by CHG2 and M3GNet, with values of 0.65 eV and 0.66 eV, respectively. These two values agree with a DFT reference 0.65 eV [49]. CHG3 presents the poorest performance, underestimating the migration energy with a value of 0.30 eV, less than half of the reference value.

Regarding the energy of the (100), (110) and (111) surfaces, all the potentials are able to capture the trend present in DFT calculations: $E_s^{(111)} > E_s^{(100)} > E_s^{(110)}$. The potential with the lowest error is MACE, which underestimates by up to 17%. SevenNet-o, CHG3 and CHG2 produce higher errors, which go up to 50%. For M3GNet, we notice that the surfaces (110) and (111) are not stable: for (111), the surface atoms reconstruct, and for (110) a semi amorphous region is formed close to the free surface.

The magnetic moments of bulk and the (111), (100) and (110) surfaces calculated with CHG2 and CHG3 give low errors, the largest ones being in the order of 10 CHG2 gives rather similar values for the three surfaces, while CHG3 gives similar values for $\mu^{(111)}$ and $\mu^{(110)}$. On a positive note, both potentials coincide with the DFT predictions that the atoms in the (100) surface have the highest magnetic moments.

The elastic properties are compared with the values from materials project database [48]. CHG2 has the least mean absolute average error for the elastic properties with a value of 20%

Table 1. Calculated and DFT reference values of the lattice parameter (a_0^{Fe}), vacancy formation energy (E_f^v), migration energy (E_{m}^{Fe}), surface energies ($E_{\text{s}}^{(100)}$, $E_{\text{s}}^{(110)}$, $E_{\text{s}}^{(111)}$), bulk (μ^{bulk}) and surface magnetic moments ($\mu^{(111)}$, $\mu^{(110)}$, $\mu^{(100)}$), and elastic coefficients (B, C', C_{44}). The best predictions are highlighted in bold for each property (except for the lattice parameter and the elastic coefficients). Fields labelled with "NC" indicate no convergence after 10^5 iterations.

		MLIPs					DFT									
		CHG2	CHG3	M3GNet	MACE	SevenNet-o	MP[a]	[18]	[50]	[51]	[52]	[49]	[53]	[54]	[43]	[55]
a_0^{Fe}	[Å]	2.849	2.847	2.839	2.859	2.846	2.86	2.84	2.863	2.836	2.869		2.855	2.84	2.831	
(E_f^v)	[eV]	0.73	0.82	0.83	**1.38**	1.29						1.95	1.94		2.18	
E_{m}^{Fe}	[eV]	**0.65**	0.3	**0.66**	0.58	0.47						0.65			0.7	
$E_{\text{s}}^{(100)}$	[Jm^{-2}]	1.35	1.54	1.08	**2.39**	1.85	2.49	2.5	2.47	2.49	2.29					
$E_{\text{s}}^{(110)}$	[Jm^{-2}]	1.09	1.16	NC	**2.03**	1.25	2.44	2.45	2.37		2.27					
$E_{\text{s}}^{(111)}$	[Jm^{-2}]	1.59	1.72	NC	**2.49**	2.04	2.73	2.73	2.58		2.52					
μ^{bulk}	[μ_B]	**2.35**	2.44				2.64	2.18	2.2			2.32	2.31	2.21		
$\mu^{(111)}$	[μ_B]	2.5	**2.54**						2.81					2.84		
$\mu^{(110)}$	[μ_B]	2.5	**2.52**						2.59					2.6		
$\mu^{(100)}$	[μ_B]	2.56	**2.7**						2.95					2.94		
B	[GPa]	182	111	158	139	97	208	182					177			211
C'	[GPa]	55	−1	55	29	19	49.5	49					51			51
C_{44}	[GPa]	55	96	63	120	110	89	97					101			125

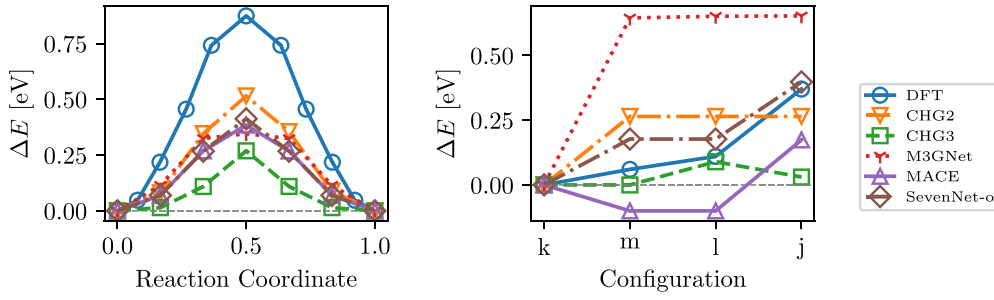[a] MP—Materials Project database [56].

Figure 4. Comparison of C behaviour according to the considered universal potentials and a DFT [25, 60]. (Left) Minimum energy paths for the migration of a carbon atom between two octahedral sites in BCC Fe. (Right) ΔE_{r-k} where r refers to configurations *k, m, l* and *j*. DFT data from [22]. The configurations are shown in figure 1.

followed by M3GNet with a value of 21%. However, the bulk modulus B is underestimated by all models, the worst prediction being SevenNet-o with a value of 97 GPa, about half of the reference value. The results for C' are comparable with DFT with an error of 10% for CHG2 and M3GNet. Noteworthy of CHG3 is that $C' < 0$, which violates one of the mechanical stability criteria [19, 57, 58]. Thereby making the application of CHG3 impractical for Fe-rich systems. C_{44} is underestimated by CHG2 (38%) and by M3GNet (29%) while MACE and SevenNet-o overestimate by similar percentages (35% and 24%, respectively).

3.2. Carbon in iron

The two interstitial positions in BCC Fe hold special significance for carbon: the octahedral position is its lowest energy position while the tetrahedral position is the saddle point energy configuration for a carbon atom along the diffusion path between octahedral positions. The minimum energy path for a carbon atom between two octahedral sites within a BCC Fe lattice is presented in figure 4(left). We obtain migration energy values of 0.52, 0.27, 0.35, 0.39 and 0.41 eV for CHG2, CHG3, M3GNet, MACE and SevenNet-o, respectively. All potentials underestimate the migration energy predicted by DFT (0.83 [25], 0.865 [59], 0.875 eV [60]). It is remarkable that the best (CHG2) and worst (CHG3) results are given by potentials of the same type that, moreover, were trained with the same data.

In figure 4(right) we show the relative energies of the j, k, l and m configurations (see figure 1) relative to the lowest energy state, configuration k (according to DFT [22]). In the figure, configurations are ordered from more favourable at left to less favourable at right, lower energies representing more stable structures.

The potential that best reproduces the trend of the DFT energies ($\Delta E_{k-k} < \Delta E_{m-k} < \Delta E_{l-k} < \Delta E_{j-k}$) is SevenNet-o, although configuration m is metastable and transforms to configuration l during relaxation. CHG2 partially follows the trend, but configurations *m, l* and *j* converge to a single value, somewhere close to configuration l. M3GNet presents a similar behaviour to CHG2, but the energy differences are larger with respect to DFT. The CHG3 and especially the MACE MLIPs completely fail to reproduce the DFT tendencies. Table 2 shows the binding energies of configurations a to m presented in figure 1.

When one vacancy and one carbon atom are present, all the potentials correctly favour having the C atom at the nearest neighbour (a_1) instead of at the second neighbour position (a_2) -except M3GNet that does not reach force convergence. For the case of two vacancies at second-nearest neighbour positions and one carbon atom (configurations b), only MACE

Table 2. Calculated and DFT reference binding energies (E_b) for configurations a to i introduced in figure 1. The best predictions are highlighted in bold for each property. The units are eV. Fields labelled with "NC" indicate no convergence after 10^5 iterations.

	DFT [20]	CHG2	CHG3	M3GNet	MACE	SevenNet-o
$E_b^{a_1}$	0.47	0.30	0.09	NC	**0.31**	0.29
$E_b^{a_2}$	−0.01	−0.17	−0.16	−0.47	**−0.15**	−0.16
$E_b^{b_1}$	0.68	0.02	0.05	−0.57	**0.59**	0.43
$E_b^{b_2}$	0.47	−0.16	−0.09	**0.15**	0.11	0.13
$E_b^{b_3}$	0.34	0.19	0.09	−0.02	**0.36**	0.43
$E_b^{c_1}$	0.58	0.29	0.11	NC	**0.38**	**0.38**
$E_b^{c_2}$	0.12	−0.23	−0.32	−0.48	0.01	**0.02**
E_b^{d}	0.16	0.04	0.01	0.06	**0.07**	0.06
E_b^{e}	0.13	0.05	−0.01	**0.13**	0.01	0.01
E_b^{f}	−0.09	**−0.06**	−0.15	−0.02	−0.13	−0.18
E_b^{g}	−0.09	0.05	**−0.01**	0.12	0.05	0.03
E_b^{h}	−0.65	−0.63	−0.63	−0.23	**−0.66**	−0.54
E_b^{i}	−1.67	**−1.18**	−0.01	−0.81	−1.14	−0.91
E_b^{j}	1.07	**0.37**	0.15	0.25	0.26	0.36
E_b^{k}	1.50	0.63	0.18	**0.90**	0.43	0.75
E_b^{l}		0.37	0.09	0.25	0.53	0.58
E_b^{m}		0.37	0.18	0.25	0.53	0.58

correctly predicts configuration b_3 to have the highest binding energy. None of the potentials is able to capture the DFT trend $E_b^{b_1} > E_b^{b_2} > E_b^{b_3}$. When the two vacancies are nearest neighbours (configurations c), all the MLIPs favour c_1. CHG2, CHG3 and M3GNet incorrectly produce repulsive interactions for c_2.

For the final cases (d to i), where two carbon atoms are present in bulk Fe, MACE, SevenNet-o and CHG3 correctly predict configuration d to be the most favourable one. However, none of the potentials are able to reproduce the trend of the binding energies as predicted by DFT. Considering the configurations a through i, i.e. excluding the poorly described one vacancy—two C atom configurations, MACE most often agrees best with DFT and has the smallest average absolute difference with the DFT values as well as the smallest average difference with the DFT values squared. When all configurations, a through m, are considered SevenNet-o becomes comparable to MACE. However, for all MLIPs performance for configurations *j* through m is very poor.

3.3. Solutes and vacancies

3.3.1. Solute–solute interactions.   Calculations of the binding energies, as defined in equation (4), for the solute–solute interactions are presented in figure 5. All the relaxations reach energy convergence to within a tolerance of 1×10^{-6} eV with less than 10^5 iterations, except for two cases with M3GNet, namely for the elements W and Hf. We note that although convergence in the total energy of the system is reached with a tolerance of 1×10^{-6} eV, the convergence criterion that the force on all individual atoms should be less than 1×10^{-3} eV Å^{-1} forces is never reached.
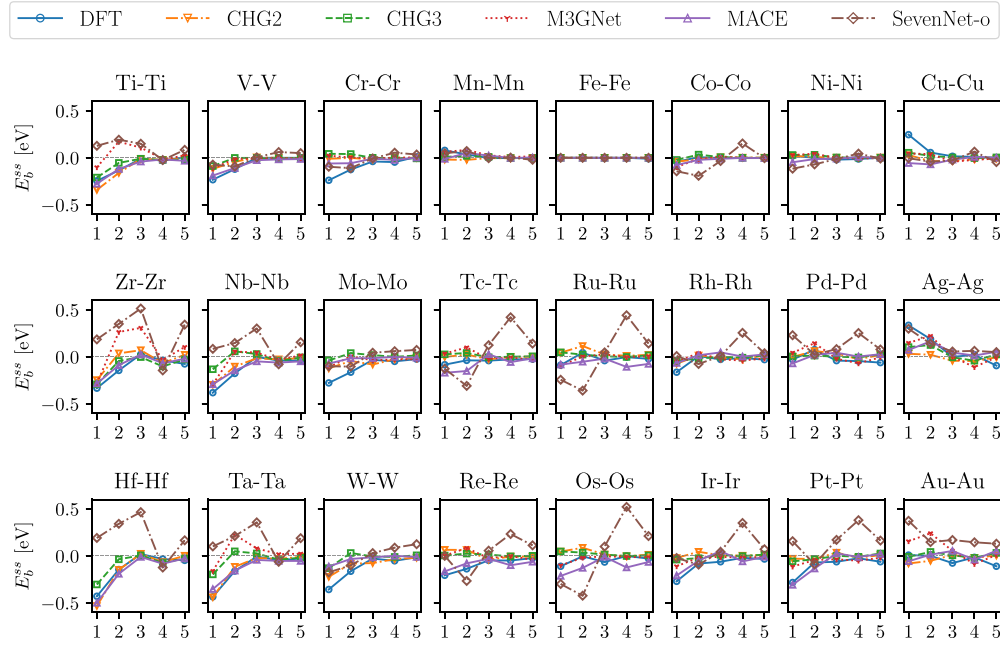
Figure 5. Solute–solute interactions for transition metals in BCC Fe. The binding energies for the first to fifth-nearest neighbours are presented, as calculated with various universal potentials. The DFT reference values are taken from [24].

Table 3. Calculated and DFT binding energies for first nearest neighbour interaction for some of the solutes. The best predictions are highlighted in bold.

Solute	DFT [24]	CHG2	CHG3	M3GNet	MACE	SevenNet-o
Co–Co	−0.043	**−0.040**	−0.022	−0.097	−0.077	−0.143
Cr–Cr	−0.239	−0.010	0.040	0.023	−0.060	**−0.095**
Cu–Cu	0.246	**0.301**	0.054	0.046	−0.056	−0.004
Mn–Mn	0.079	−0.019	0.009	**0.062**	−0.011	0.047
Mo–Mo	−0.277	**−0.122**	−0.038	−0.111	−0.070	−0.100
Nb–Nb	−0.381	**−0.296**	−0.131	−0.291	−0.291	0.085
Ni–Ni	0.020	**0.020**	0.030	0.029	−0.047	−0.115
Ti–Ti	−0.246	−0.341	−0.210	−0.100	**−0.276**	0.127
V–V	−0.233	−0.118	−0.087	−0.111	**−0.189**	−0.072

The first-nearest neighbour interaction of some common alloying elements in steel are given in table 3 for a quantitative comparison. CHG2 most often agrees best with DFT and it has the smallest average absolute difference with the DFT values as well as the smallest average difference with the DFT values squared. CHG2 is the most reliable in capturing attractive and repulsive tendencies, except for Cr–Cr and Mn–Mn, while for Cr–Cr, Cu–Cu, Zr–Zr, Ag–Ag and the platinum group it does not capture properly the distance dependence. This is not the case for other potentials, with MACE being the worst as it predicts all interactions to be repulsive. SevenNet-O shows the largest mean absolute error while CHG3 and M3GNet perform poorly too. It is also noteworthy that although CHG3 is an 'improved' version of CHG2, it performs worse.
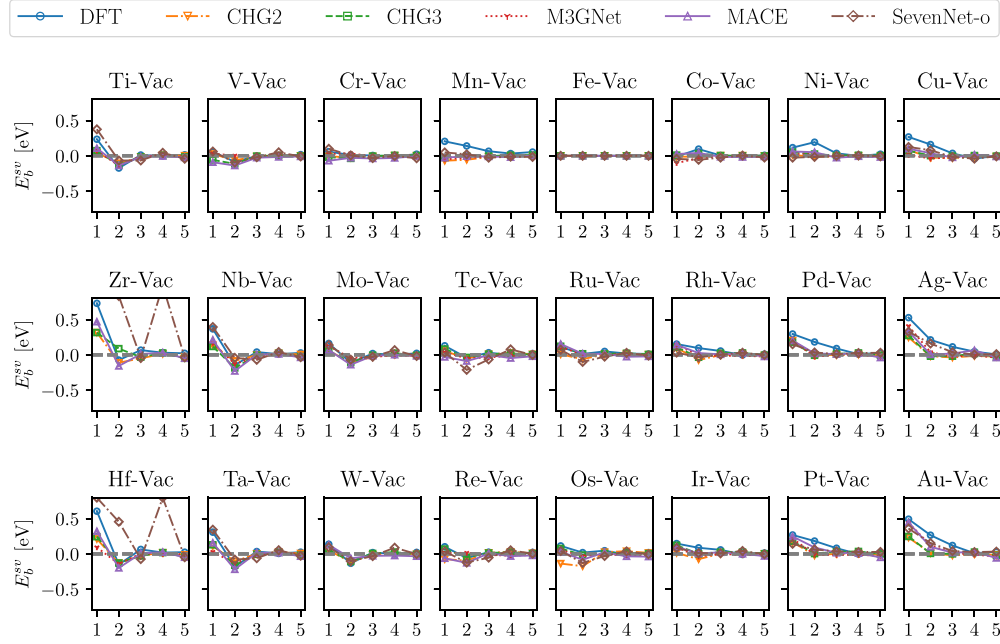
Figure 6. Solute–vacancy interactions for transition metals in BCC Fe. The binding energies for the first to fifth-nearest neighbours are presented, as calculated with various universal potentials. The DFT reference values are taken from [24].

Generally, the potentials struggle to capture the stronger first and second-nearest neighbour interactions, while they are able to capture the decay of the interactions as they get closer to the fifth-nearest neighbour.

Some specific elements that are worth mentioning are those from group 4 of the periodic table, namely Ti, Zr, and Hf. CHG2, CHG3 and MACE properly describe the strong interaction at the first-nearest neighbour and the rapid weakening with distance. A similar predictive behaviour is seen for the elements of group 5 (V, Nb, and Ta) but to a lesser extent. Cu and Ag also stand out, as these are the only elements that present strong attractive interactions according to DFT. None of the potentials are able to predict the attractive Cu–Cu and Ag–Ag interactions at first-nearest neighbour distances.

3.3.2. Solute–vacancy interactions. Results of calculations of the binding energies between solutes and vacancies (E_b^{sv}) (see equation (5)) at various nearest neighbour separations (first to fifth) are presented in figure 6. In general, the potentials accurately capture a common trend across all elements: an attractive solute–vacancy interaction that weakens with increasing separation—although some elements present a slightly repulsive interaction at the second-nearest neighbour distance.

In terms of mean absolute average deviations from DFT [24], for the first-nearest neighbour interactions involving 3d transition metals and the for steel important alloying elements Nb and Mo, SevenNet-o presents the best performance, with a value of 0.08 eV with the other MLIPs giving all about twice higher values. SevenNet-o also most often gives the best agreement with DFT (L0 norm), and the least average difference with DFT [24] squared (L2 norm) for the first-neighbour configurations.
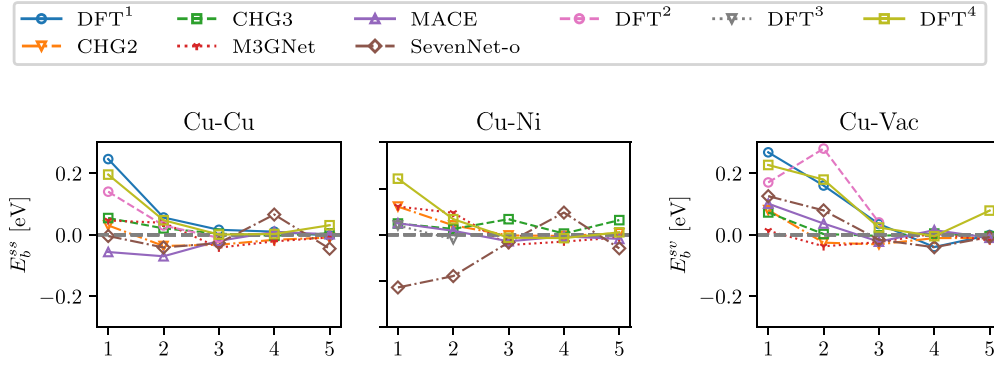
Figure 7. (Left and centre) Solute–solute interactions for Cu–Cu and Cu–Ni in BCC Fe. (Right) Solute–vacancy interactions for Cu in BCC Fe. The binding energies from the first to the fifth-nearest neighbour positions are presented. Apart from the universal potentials, additional DFT results from literature are shown: [1] [24], [2] [61], [3] [62], [4] [34].

Notable outliers are the predictions of SevenNet-o for the systems containing Zr and Hf, where the binding values at the first, second, and fourth-nearest neighbour separations are more than 0.5 eV away from the reference value, and the general decaying trend is also not reproduced. SevenNet-o is not unique in failing for Zf and Hf, M3GNet does not reach structural relaxation convergence within 10^5 iterations for these systems.

Similar to the solute–solute case, the potentials cannot accurately model the attractive interactions predicted by DFT for Cu. This is also the case for Mn. All MLIP potentials systematically underestimate the attraction between late transition metals and a vacancy, especially for the 3d series and less so for the 5d series. For group 10 (Ni, Pd, and Pt), the second-nearest neighbour interactions are underestimated by all the potentials.

In figure 7, we present in more detail the binding energies for the case of Cu–Cu, Cu-vacancy, and the additional case of Cu–Ni interactions with additional DFT results. We note that there is significant variation in the DFT results. For the first-nearest neighbour interactions, the DFT results [24, 61] differ by 0.11 eV for Cu–Cu, and by 0.10 eV for Cu-vacancy. The trend of the binding energies from the first to the third-nearest neighbour interactions is not uniform either; for Cu-Ni there is no consensus between the DFT results about the attractive or repulsive nature of the second-nearest neighbour interactions [34, 62]. However, all DFT results [24, 34, 61, 62] indicate a rapid vanishing of the interactions with distance.

3.3.3. Solute migration. In figure 8 we show the migration energy for single atoms of substitutional elements within a BCC Fe lattice as calculated with various universal interatomic potentials. We also include reference DFT values for comparison [43]. The various universal interatomic potentials roughly capture the behaviour displayed by DFT: for each of the considered periods (4, 5, 6) there is an increasing trend starting with the elements of group 4 (Ti, Zr, Hf), peaking at group 9 (Co, Rh, Ir), and decreasing towards group 12 (Cu, Ag, Au) – the only exception being Mn. The calculation of Ir with M3GNet, failed to converge.

On average, CHG3 performs worst, with a mean absolute error of 0.34 eV and with a systematic underestimation for all the elements considered. SevenNet-o erroneously predicts migration energies close to zero for the elements Ti, Zr, Nb, Hf, and Ta. This is in part due to the strong attractive solute–vacancy interaction that SevenNet-o predicts for such elements.
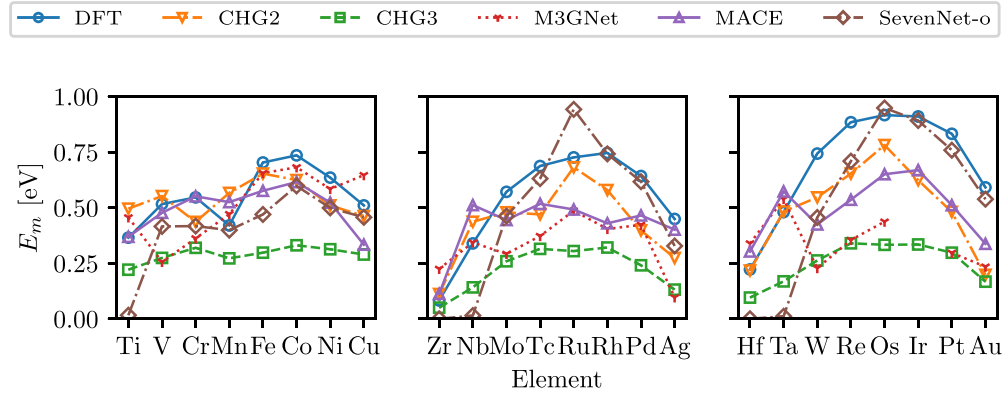
Figure 8. Migration energy for single atoms of substitutional elements within a BCC Fe lattice (see figure 3), as calculated with various universal interatomic potentials. The DFT reference values are taken from [43].

We also note that their equilibrium position is slightly shifted towards the vacancy, making the transition path for vacancy migration shorter.

The best average performance is obtained by CHG2, with a mean absolute error of 0.14 eV.

4. Discussion

Universal MLIPs offer significant potential provided that at least three conditions are met: (i) accurate reproduction of DFT results, at (ii) much greater computational speed, with (iii) better scaling with system size. Here, mostly the accuracy aspects will be discussed with only minor attention to speed and scaling. It is clear from the data we have presented here that the currently offered universal potentials are not good enough to use off the shelf. As table 4 indicates, all MLIPs fail badly for several properties with high relevance to modelling steel. In fact, some properties such as the vacancy formation energy and the interaction of multiple C atoms with a single vacancy are not well described by any of the five MLIPs. Especially troublesome is the failure of CHG3 for the mechanical stability of the BCC Fe crystal structure as evidenced by the negative value of C' in table 1 and the failure of M3GNet to converge for surfaces, see table 1, and for migration energy barriers for C and some substitutional solids, as well as for some configurations involving C, see table 2. The remaining three MLIPS, CHG2, MACE, and SevenNet-o, are quite similar in overall performance, but differ in their details.

The elastic constants, of great importance in modelling dislocations, are best described by CHG2, with MACE as a second choice. For simulations involving free surfaces MACE is to be favoured over CHG2.

For properties involving binding energies it is necessary that the model captures the type of interaction (attraction or repulsion), and that it properly identifies the most stable configuration. In simulations where C plays an important role MACE and CHG2 are of similar performance, with SevenNet-o less favoured because it performs less well on elasticity, surfaces and migration in pure Fe. When substitutional solutes are concerned, solute–solute interactions, important to describe clustering, precipitation and ordering, appear best described by CHG2, while MACE and SevenNet-o are a little less successful. SevenNet-o outperforms CHG2 and

Table 4. Performance of universal MLIPs on various materials properties. Vacancies are indicated with □ and substitutional solute atoms with *M*. Interaction is abbreviated as int.act. For energy properties the symbols are assigned according to absolute numerical differences with the DFT target values (norm 1 loss): + differs less than 0.1 eV, ○ differs between 0.1 and 0.25 eV, − differs 0.25–0.5 eV, −− differs more than 0.5 eV. For surface energies (elastic properties) the same numerical boundaries but in units of Jm^{-2} (100 GPa). Pathological errors, such a failure to reach energy converge after 10^5 structural optimization iterations, or mechanical instability of the BCC Fe crystal, are indicated by a $*$ superscript.

	CHG2	CHG3	M3GNet	MACE	SevenNet-o
Pure Fe					
a_0^{Fe}	+	+	+	+	+
(E_f^v)	−−	−−	−−	−−	−−
E_m^{Fe}	+	−	+	+	○
$E_s^{(hkl)}$	−−	−−	−−*	○	−−
C_{ijkl}	○	−−*	−	−	−−
C migration in Fe	−	−−	−−*	−	−
C-□ int.act.	○	−	−	○	○
C-□-□ int.act.	−	−	−−*	○	○
C–C int.act.	○	−	−	○	○
C–C–□ int.act.	−−	−−	−−	−−	−−
M–M int.act.	+	○	○	○	○
M-□ int.act.	○	○	○	○	+
M migration in Fe	○	−	○*	○	○

MACE on solute–vacancy interactions which are crucial for properly describing the kinetics of substitutional solutes. All three MLIPs, CHG2, MACE, and SevenNet-o, perform similarly for solute migration energy barriers. It must be mentioned that none of these three MLIPs performs particularly well on solute energetics.

Overall, among the compared potentials CHG2 appears to be qualitatively reliable except for the iron–carbon system. It should be noted that in kinetic simulations involving thermally activated processes, quantities like the vacancy formation energy and migration energies appear in exponentials. Therefore, even relatively small errors in MLIPs can give rise to large discrepancies between MLIP predictions and DFT or reality. The carbon-vacancy binding is of great importance [23] because it indicates how strong carbon segregates towards excess volume, such as at grain boundaries and at dislocations. CHG2, MACE, and SevenNet-o all perform very comparable when it comes to a single C interacting with a single vacancy, with a typical error of 150 meV. In a real sense, the vacancy diffusion in pure iron is of little relevance since even with carbon concentrations at ppm levels it is expected that vacancies are decorated with two or three carbon atoms. Therefore it is vital that models for steels reproduce faithfully the binding energies and their ordering as found in [23], and as we see here this is not the case for the universal potentials we have tested. When two C are interacting with a vacancy the errors are much larger, typically 0.5 to 1 eV, with the second C atom giving no, or almost no energy gain in marked contrast to the DFT results, see table 2. DFT calculations have shown that carbon-vacancy complexes with even more C atoms occur at much higher concentrations than vacancies without any bonded carbon, see e.g. figure 3 in [23].

In the very recent motivation in steel metallurgy to study tramp elements in remelted steels as well as the roles that substitutional elements play in properties including strength, toughness, hardenability and so on, it is certainly very attractive to have access to universal interatomic potentials that are advertised as covering essentially the whole periodic table. The current results show that MLIP performance is strongly solute dependent. Although CHG2 overall performs best for solute–solute interactions, Ti–Ti and V–V were best described by MACE, while Cr–Cr was best described by SevenNet-o. For Cu–Cu none of the MLIPs performed well. For solute–vacancy interactions SevenNet-o performed considerably better than CHG2 and MACE. Nevertheless, for some cases SevenNet-o failed in dramatic fashion: Mn, Co, Ni, Zr, Hf were poorly described. The other MLIPs performed not much better for these solute–vacancy interactions.

The performance variability occurs for all universal MLIPs. We surmise that this originates from incomplete training datasets. Although there is a large number of configurations in the training datasets, they do not cover the configuration space well. Defected crystal structures, which are essential for any realistic materials modelling, are under-represented as the current datasets[5] consist of equilibrium configurations or configurations close to equilibrium [64]. This makes most property predictions, other than the equilibrium lattice parameter, an extrapolation for the MLIPs, which are known to perform poorly when extrapolating. This is especially true when calculating the surface energy or the vacancy formation energy where the atomic environments differ much from the training datasets. To improve the MLIPs, relevant defected configurations must be included during the initial training or must be used for subsequent fine-tuning. Furthermore, the loss functions and metrics that are used for training and rating universal MLIPs are not representative of specific materials, or of realistic materials modelling centred around defect behaviour. A case in point may be modelling a large cell with about a hundred atoms and a single vacancy. The error in the energy averaged per atom may be very small, say 10 meV, but 'averaged' over this single vacancy it will be two orders of magnitude larger, at which point it approaches the value of the vacancy formation energy itself.

The variability in DFT results of ostensibly the same quantity is a matter of concern. This is exemplified in variations in elastic constants, see table 1, and also in the Cu binding curves, see figure 7. While it is customary to assert in the materials science community that there exists a well-defined DFT value for a given property, there are at least three reasons why DFT results differ across multiple publications. The first may be called 'user error', in the sense that convergence with respect to basis set, **k**-point mesh and supercell size may not have been achieved. The second is a consequence of the variety of exchange and correlation functionals: the standard local-spin-density functional forms [65, 66], are not used these days as they underestimate the lattice constant of iron and overestimate the bulk modulus; on the other hand there is a variety of gradient-corrected functionals which yield the correct lattice constant but overestimate the magnetic moment. Thirdly, there is the choice of electron-ion potential: these vary from all-electron descriptions to various flavours of pseudopotentials to the augmentation methods employed in the projector augmented-wave approach [67].

As a consequence, training data sets need to be composed using a uniform methodology and using uniform computational settings.

---

[5] Even the latest databases, such as OMAT24 [63], mention specifically that the focus is on perfectly periodic structures without defects.

5. Conclusions

Universal MLIPs have been used to calculate a range of properties relevant to modelling of steel. The properties were limited to those for which DFT values have been published. Accuracy of (defect) energies only was considered, computational speed, memory requirement, scaling with system size, and even efficiency of structural optimization, i.e. number of required ionic iterations, were not considered. It might be worthwhile to mention that force convergence was much poorer than what was expected on the basis of energy convergence. Globally, the MLIPs performed impressively on energetic properties. Qualitatively at least, MLIPs appear remarkably able to reproduce trends and to predict some properties of pure iron, such as, say, surface energies. Especially CHG2 stood out, but also MACE and SevenNet-o often gave surprisingly sensible results. However, as the current MLIPs are trained on perfect, defect-free crystal structures, they generally perform much worse on defect properties as that the published loss values would suggest. modelling realistic materials usually involves defects and often thermally activated processes mediated by defects as well. Therefore, current MLIPs are not yet capable to accurately represent realistic processes in specific materials such as steel.

The extreme subtleties arising from the presence of both interstitials and numerous substitutional elements mean that for now better models, maybe with a narrower scope, are needed.

We have also exposed in the this study the surprising conclusion that even with modern DFT codes, their results taken from various sources can be as widely differing as the universal machine learning potentials, hence values to compare must be carefully chosen. This fact should not be ignored.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

Appendix. Comparison between CHG2 and CHG3

Three key distinctions exist between CHG2 and CHG3. Firstly, the data partitioning for training, validation, and testing differs. CHG2 employs an 80-10-10 split, whereas CHG3 utilizes a 90-5-5 split. Secondly, variations are observed in the model parameters (table A1). Finally, discrepancies in the training parameters are also evident (table A2).

Table A1. Model parameter comparison between CHG2 and CHG3.

Model Parameter	CHG2	CHG3
Number of radial components	9	31
Number of angular basis to use	9	31
Normalization layer to use on the gated MLP	None	layer
Readout MLP hidden dimensions	[64, 64]	[64, 64, 64]
Cutoff radius (Å) in creating the graph adjacency list	5	6
Cutoff strength used in graph smooth cutoff function	5	8
Bias in the output MLP Linear layer	True	False

Table A2. Training parameter comparison between CHG2 and CHG3.

Training Parameter	CHG2	CHG3
Decay Fraction	1×10^{-2}	0.5×10^{-2}
Training epochs	20	30
Initial learning rate	1e-3	5e-3

ORCID iDs

Sebastián Echeverri Restrepo https://orcid.org/0000-0001-8692-1879
Naveen K Mohandas https://orcid.org/0000-0002-6854-9767
Marcel H F Sluiter https://orcid.org/0000-0002-6514-4318
Anthony T Paxton https://orcid.org/0000-0003-4182-8210

References

[1] Riebesell J, Goodall R E A, Benner P, Chiang Y, Deng B, Lee A A, Jain A and Persson K A 2023 Matbench Discovery – a framework to evaluate machine learning crystal stability predictions (arXiv:2308.14920)
[2] Yu H, Giantomassi M, Materzanini G, Wang J and Rignanese G-M 2024 Systematic assessment of various universal machine-learning interatomic potentials (arXiv:2403.05729)
[3] Focassio B, Freitas L P M and Schleder G R 2024 Performance assessment of universal machine learning interatomic potentials: challenges and directions for materials' surfaces *ACS Appl. Mater. Interfaces* (https://doi.org/10.1021/acsami.4c03815)
[4] Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel C J and Ceder G 2023 CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling *Nat. Mach. Intell.* **5** 1031–41
[5] Chen C and Ping Ong S 2022 A universal graph deep learning interatomic potential for the periodic table *Nat. Comput. Sci.* **2** 718–28
[6] Batatia I, Kovacs D P, Simm G, Ortner C and Csanyi G 2022 MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields *Advances in Neural Information Processing Systems* vol 35, ed S Koyejo, S Mohamed, A Agarwal, D Belgrave, K Cho and A Oh (Curran Associates, Inc) pp 11423–36 (available at: https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf)
[7] Park Y, Kim J, Hwang S and Han S 2024 Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations *J. Chem. Theory Comput.* **20** 4857–68
[8] Batatia I *et al* 2023 A foundation model for atomistic materials chemistry (arXiv:2401.00096)
[9] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B 2022 E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13** 2453

[10] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library *Advances in Neural Processing Systems* **vol 32** pp 8026–37 (available at: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf)

[11] Geiger M and Smidt T 2022 e3nn: euclidean neural networks (arXiv:2207.09453)

[12] Larsen A H *et al* 2017 The atomic simulation environment-a Python library for working with atoms *J. Phys.: Condens. Matter* **29** 273002

[13] Bitzek E, Koskinen P, Gähler F, Moseler M and Gumbsch P 2006 Structural relaxation made simple *Phys. Rev. Lett.* **97** 1–4

[14] Guénolé J, Nöhring W G, Vaid A, Houllé F'eric, Xie Z, Prakash A and Bitzek E 2020 Assessment and optimization of the fast inertial relaxation engine (fire) for energy minimization in atomistic simulations and its implementation in LAMMPS *Comput. Mater. Sci.* **175** 109584

[15] Echeverri Restrepo S and Andric P 2023 ABC-FIRE: accelerated bias-corrected fast inertial relaxation engine *Comput. Mater. Sci.* **218** 111978

[16] The force convergence criterion value ΔF can be related to the energy convergence ΔE through volume V, bulk modulus B, and pressure P: $2\Delta E = BV(\Delta V/V)^2 \rightarrow \Delta V/V = \sqrt{2\Delta E/(BV)}$; $\Delta F = |AP|$ with $P = -B\Delta V/V$ and $A = V^{(2/3)} \rightarrow \Delta F = V^{(1/6)}\sqrt{2B\Delta E}$, where A is the cross-sectional area per atom. Now using: $\Delta E = 10^{-6}$ eV/atom, $V = 16$ Å^3/atom, $B = 150$ GPa $\approx$ 1 eV/Å$^3 \rightarrow \Delta F \approx 2$ meV/Å

[17] Echeverri Restrepo S, Sluiter M H F and Thijsse B J 2013 Atomistic relaxation of systems containing plasticity elements *Comput. Mater. Sci.* **73** 154–60

[18] de Jong M *et al* 2015 Charting the complete elastic properties of inorganic crystalline compounds *Sci. Data* **2** 150009

[19] Echeverri Restrepo S 2022 Density functional theory characterisation of cementite (Fe$_3$C) with substitutional molybdenum (Mo) atoms *Physica* B **631** 413669

[20] Becquart C S, Raulot J M, Bencteux G, Domain C, Perez M, Garruchet S and Nguyen H 2007 Atomistic modeling of an Fe system with a small concentration of C *Comput. Mater. Sci.* **40** 119–29

[21] Lau T T, Först C J, Lin Xi, Gale J D, Yip S and Van Vliet K J 2007 Many-body potential for point defect clusters in Fe-C alloys *Phys. Rev. Lett.* **98** 98–101

[22] Paxton A T and Elsässer C 2013 Analysis of a carbon dimer bound to a vacancy in iron using density functional theory and a tight binding model *Phys. Rev.* B **87** 224110

[23] Först C J, Slycke J, Van Vliet K J and Yip S 2006 Point defect concentrations in metastable Fe-C alloys *Phys. Rev. Lett.* **96** 175501

[24] Olsson P, Klaver T P C and Domain C 2010 *Ab initio* study of solute transition-metal interactions with point defects in bcc Fe *Phys. Rev.* B **81** 054102

[25] Simonovic D, Ande C K, Duff A I, Syahputra F and Sluiter M H F 2010 Diffusion of carbon in bcc Fe in the presence of Si *Phys. Rev.* B **81** 054116

[26] Echeverri Restrepo S, Andric P and Paxton A T 2022 Effect of crystal defects in iron on carbon diffusivity: analytical model married to atomistics *Phys. Rev. Mater.* **6** L100801

[27] Gorbatov O I, Hosseinzadeh Delandar A, Gornostyrev Y N, Ruban A V and Korzhavyi P A 2016 First-principles study of interactions between substitutional solutes in bcc iron *J. Nucl. Mater.* **475** 140–8

[28] Takeuchi S 1969 Solid-solution strengthening in single crystals of iron alloys *J. Phys. Soc. Japan* **27** 929–40

[29] Schmauder S and Kohler C 2011 Atomistic simulations of solid solution strengthening of α-iron *Comput. Mater. Sci.* **50** 1238–43

[30] Zeng D, Lu L, Gong Y, Zhang Y and Zhang J 2017 Influence of solid solution strengthening on spalling behavior of railway wheel steel *Wear* **372–373** 158–68

[31] Andric P, Echeverri Restrepo S and Maresca F 2023 Predicting dislocation density in martensite $ab − initio$ *Acta Mater.* **243** 118500

[32] Auger P, Pareige P, Akamatsu M and Van Duysen J-C 1994 Microstructural characterization of atom clusters in irradiated pressure vessel steels and model alloys *J. Nucl. Mater.* **211** 194–201

[33] Hyde J M, Sha G, Marquis E A, Morley A, Wilford K B and Williams T J 2011 A comparison of the structure of solute clusters formed during thermal ageing and irradiation *Ultramicroscopy* **111** 664–71

[34] Whiting T M, Burr P A, King D J M M and Wenman M R 2019 Understanding the importance of the energetics of Mn, Ni, Cu, Si and vacancy triplet clusters in bcc Fe *J. Appl. Phys.* **126** 115901

[35] Gavriljuk V G, Shanina B D and Berns H 2008 *Ab initio* development of a high-strength corrosion-resistant austenitic steel *Acta Mater.* **56** 5071–82

[36] Bhadeshia H K D H 2012 Steels for bearings *Prog. Mater. Sci.* **57** 268–435

[37] Scheuer U and Lengeler B 1991 Lattice distortion of solute atoms in metals studied by x-ray-absorption fine structure *Phys. Rev.* B **44** 9883–94

[38] Zhang H, Johansson B, Ahuja R and Vitos L 2012 First-principles study of solid-solution hardening in steel alloys *Comput. Mater. Sci.* **55** 269–72

[39] Fellinger M R, Hector L G and Trinkle D R 2018 Effect of solutes on the lattice parameters and elastic stiffness coefficients of body-centered tetragonal Fe *Comput. Mater. Sci.* **152** 308–23

[40] Purdy G R and Brechet Y J M 1995 A solute drag treatment of the effects of alloying elements on the rate of the proeutectoid ferrite transformation in steels *Acta Metall. Mater.* **43** 3763–74

[41] Rawers. J C 2008 Alloying effects on the microstructure and phase stability of Fe-Cr-Mn steels *J. Mater. Sci.* **43** 3618–24

[42] Yan P and Bhadeshia H K D H 2015 Austenite-ferrite transformation in enhanced niobium, low carbon steel *Mater. Sci. Technol.* **31** 1066–76

[43] Messina L, Nastar M, Sandberg N and Olsson P 2016 Systematic electronic-structure investigation of substitutional impurity diffusion and flux coupling in bcc iron *Phys. Rev.* B **93** 184302

[44] Weinan E, Ren W and Vanden-Eijnden E 2002 String method for the study of rare events *Phys. Rev.* B **66** 052301

[45] Weinan E, Ren W and Vanden-Eijnden E 2007 Simplified and improved string method for computing the minimum energy paths in barrier-crossing events *J. Chem. Phys.* **126** 164103

[46] Cameron M, Kohn R V and Vanden-Eijnden E 2011 The string method as a dynamical system *J. Nonlinear Sci.* **21** 193–230

[47] Makri S, Ortner C and Kermode J R 2019 A preconditioning scheme for minimum energy path finding methods *J. Chem. Phys.* **150** 094109

[48] Jain A *et al* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Materials* **1** 011002

[49] Domain C and Becquart C S 2001 *Ab initio* calculations of defects in Fe and dilute Fe-Cu alloys *Phys. Rev.* B **65** 024103

[50] Błoński P and Kiejna A 2007 Structural, electronic and magnetic properties of bcc iron surfaces *Surf. Sci.* **601** 123–33

[51] Echeverri Restrepo S, Di Stefano D, Mrovec M and Paxton A T 2020 Density functional theory calculations of iron - vanadium carbide interfaces and the effect of hydrogen *Int. J. Hydrog. Energy* **45** 2382–9

[52] Spencer M J S, Hung A, Snook I K and Yarovsky I 2002 Density functional theory study of the relaxation and energy of iron surfaces *Surf. Sci.* **513** 389–98

[53] Ackland G J, Mendelev M I, Srolovitz D J, Han S and Barashev A V 2004 Development of an interatomic potential for phosphorus impurities in -iron *J. Phys.: Condens. Matter* **16** S2629–42

[54] Fischer G and Gerber. I C 2014 Magnetic properties of bcc iron surfaces and the influence of the chemical environment: electronic structure calculations *J. Phys.: Condens. Matter.* **27** 015004

[55] Lin Y, Chong X, Ding Y, Zhou Y, Gan M, Xu L, Wei S and Feng J 2021 First-principles calculations of thermal and electrical transport properties of bcc and fcc Dilute Fe-X (X = Al, Co, Cr, Mn, Mo, Nb, Ni, Ti, V and W) binary alloys *Metals* **11** 1988

[56] Jain A *et al* 2013 Commentary: The materials project: a materials genome approach to accelerating materials innovation: Data retrieved from the materials project for Fe (mp-13) from database version v2023.11.1.

[57] Born M and Huang K 1954 *Dynamical theory of crystal lattices International Series of Monographs on Physics* (Clarendon)

[58] Shang S L, Saengdeejing A, Mei Z G, Kim D E, Zhang H, Ganeshan S, Wang Y and Liu Z K 2010 First-principles calculations of pure elements: equations of state and elastic stiffness constants *Comput. Mater. Sci.* **48** 813–26

[59] Souissi M, Chen Y, Sluiter M H F and Numakura H 2016 *Ab initio* characterization of B, C, N and O in bcc iron: solution and migration energies and elastic strain fields *Comput. Mater. Sci.* **124** 249–58

[60] Maugis P and Kandaskalov D 2020 Revisiting the pressure effect on carbon migration in iron *Mater. Lett.* **270** 127725

[61] Becquart C S and Domain C 2003 *Ab initio* contribution to the study of complexes formed during dilute FeCu alloys radiation *Nucl. Instrum. Methods Phys. Res* B **202** 44–50

[62] Vincent E, Becquart C S and Domain C 2007 Atomic kinetic Monte Carlo model based on *Ab initio* data: simulation of microstructural evolution under irradiation of dilute Fe-CuNiMnSi alloys *Nucl. Instrum. Methods Phys. Res.* B **255** 78–84

[63] Barroso-Luque L, Shuaibi M, Fu X, Wood B M, Dzamba M, Gao M, Rizvi A, Zitnick C L and Ulissi Z W 2024 Open materials 2024 (OMat24) inorganic materials dataset and models (arXiv:2410.12771)

[64] Deng B, Choi Y, Zhong P, Riebesell J, Anand S, Li Z, Jun K, Persson K A and Ceder G 2024 Overcoming systematic softening in universal machine learning interatomic potentials by fine-tuning (arXiv:2405.07105)

[65] Ceperley D M and Alder B J 1980 Ground state of the electron gas by a stochastic method *Phys. Rev. Lett.* **45** 566–9

[66] von Barth U and Hedin L 1972 A local exchange-correlation potential for the spin polarized case. i *J. Phys. C: Solid State Phys.* **5** 1629–42

[67] Blöchl P E 1994 Projector augmented-wave method *Phys. Rev.* B **50** 17953–79