scientific reports

OPEN

A machine-learned spin-lattice potential for dynamic simulations of defective magnetic iron

Jacob B. J. Chapman [✉] & Pui-Wai Ma

A machine-learned spin-lattice interatomic potential (MSLP) for magnetic iron is developed and applied to mesoscopic scale defects. It is achieved by augmenting a spin-lattice Hamiltonian with a neural network term trained to descriptors representing a mix of local atomic configuration and magnetic environments. It reproduces the cohesive energy of BCC and FCC phases with various magnetic states. It predicts the formation energy and complex magnetic structure of point defects in quantitative agreement with density functional theory (DFT) including the reversal and quenching of magnetic moments near the core of defects. The Curie temperature is calculated through spin-lattice dynamics showing good computational stability at high temperature. The potential is applied to study magnetic fluctuations near sizable dislocation loops. The MSLP transcends current treatments using DFT and molecular dynamics, and surpasses other spin-lattice potentials that only treat near-perfect crystal cases.

The success of density functional theory (DFT)[1,2] has drastically advanced the scientific and technological aspects of materials development due to its unprecedented predictive power at a modest computational cost. However, the order $O(n^3)$ scalability of DFT calculations, where n is the number of electrons, has severely limited the simulation box size and time scale. Machine-learned potentials have demonstrated their ability to perform scalable atomic scale simulations with DFT accuracy using only a fraction of its computational requirements[3]. Since the seminal work of Behler and Parrinello[4], who introduced the concept of invariant descriptors to represent local chemical environment, a range of machine-learned potentials based on kernel methods[5,6] and network networks[7–10] have been developed and applied to investigate real physical problems.

Spin-polarized and non-collinear magnetism are well established extensions of DFT for magnetic materials but their results are valid only for the electronic ground state. Attempts to mimic magnetic excitation by coupling spin dynamics to constrained non-collinear calculations have been made[11,12]. However, the limitations of the DFT method on the simulation box size has yet to be overcome. In addition the effects of magnetic excitation and their interaction with atomic trajectories are irreconcilable within the framework of classical molecular dynamics (MD)[13].

Nevertheless, magnetic effects cannot be ignored in many situations. In magnetic iron, the bcc-fcc and fcc-bcc phase transitions at 1185K and 1667K, respectively, are due to the competing phonon and magnon free energies[14–18]. The softening of tetragonal shear modulus C' near the Curie temperature T_C[19,20] and stability of anomalous ⟨110⟩ dumbbell self-interstitial atom (SIA) configurations[21–23] are also believed to be magnetically driven. Itinerant ferromagnetism, in the form of increased magnitudes of magnetic moment, have been linked to the stability of grain boundaries and intergranular cohesion[24].

Spin-lattice dynamics[25] was developed to treat both spin (magnetic) and lattice degrees of freedom within a unified framework. Spin-lattice dynamics is a general framework similar to molecular dynamics and applicable to any arbitrary atomic scale Hamiltonian. The latest development on the Langevin spin equation of motion[26] allows simultaneous treatment of both the rotational (direction) and longitudinal fluctuations (magnitude) of magnetic moments. In most other studies the magnitudes of magnetic moments are assumed to be fixed[27–29] or have been performed on a fixed lattice[30,31]. Whilst spin-lattice dynamics has been used to investigate a variety of microscopic dynamic effects in iron[14,25,27–29,32,33], there is still not a spin-lattice potential capable of simultaneously modelling mechanical deformations, magnetic fluctuations and defect properties[13].

The difficulty of developing spin-lattice potentials are two-fold. First, a spin-lattice potential has double the degrees of freedoms (6N) of a conventional MD potential (3N), where N is the number of atoms. A substantial amount of extra data is required for potential fitting for each extra degree of freedom, drastically expending the

		MSLP			DFT (VASP)			DFT (OpenMX)		
		a_0 (Å)	$|\mathbf{M}|$ (μ_B)	ΔE (eV/atom)	a_0 (Å)	$|\mathbf{M}|$ (μ_B)	ΔE (eV/atom)	a_0 (Å)	$|\mathbf{M}|$ (μ_B)	ΔE (eV/atom)
BCC	FM	2.817	2.16		2.831	2.19		2.842	2.25	
	SL-AFM	2.824	1.54	0.36	2.800	1.34	0.46			
	NM	2.753	0.00	0.42	2.764	0.00	0.47	2.766	0.00	0.56
FCC	DL-AFM	3.470	2.08	0.08	3.466	2.04	0.08	3.476	2.38	0.10
	SL-AFM	3.494	0.96	0.16	3.494	1.30	0.12	3.435	2.00	0.13
	FM	3.47	1.03	0.15	3.50	1.00	0.16	3.648	2.63	0.12
	NM	3.428	0.00	0.18	3.456	0.00	0.16	3.462	0.00	0.25

Table 1. The equilibrium lattice constant a_0, the magnitude of spontaneous magnetic moment $|\mathbf{M}|$, and the relative energy difference with respect to the BCC ground state ΔE calculated using our machine-learned spin-lattice potential (MSLP) for iron at non-magnetic (NM), ferromagnetic (FM), single layer antiferromagnetic (SL-AFM), and double layer antiferromagnetic (DL-AFM) states in BCC and FCC structures. DFT calculations using VASP and OpenMX are shown for comparison. Details are in Supplementary Materials.

representable phase space. Recent data-driven techniques can aid in parameter optimisation for such cases[33]. Second, potentials that adopt the Heisenberg or Heisenberg-Landau functional form in various studies[23] are shown to be too restrictive to near-perfect crystal cases. A good functional form that is applicable to both perfect and defective configurations is yet to be derived.

Machine-learned potentials for spin-lattice dynamics that go beyond the need of a well defined functional form could be a viable solution[10,34]. While the number of machine-learned potentials for iron has rapidly increased over the past decade[3,35–37], applications including explicit spin degrees of freedoms are very limited. Recently, Nikolov et al.[33] produced a machine-learned spectral neighbor analysis potential. Since they kept using the Heisenberg functional form, the potential does not consider the change of the magnitudes of magnetic moments due to thermal excitation or the change of local atomic environment. Novikov et al.[38] developed a moment tensor spin-lattice potential that includes longitudinal fluctuation, but they limited their approach to collinear configurations near perfect crystal structures. Domina et al.[34] extended the spectral-neighbour representation to be applicable to non-unit vector fields such as spin. Whilst no dynamics was performed, their approach shows an excellent ability to predict the energies of non-collinear states relative to a prototype model of iron for configurations with small atomic displacements from the perfect BCC lattice.

In this work, an alternative approach is taken. We built on a conventional spin-lattice model known to work well near equilibrium conditions for BCC iron. Then, an additional neural network term is trained to reduce the error of the conventional spin-lattice model near equilibrium and to learn the missing physics as the environment deforms. Such an approach is distinct from the method by Nikolov[33] which trained a SNAP potential and converted into a spin-lattice model by adding a pairwise Heisenberg potential. However, bilinear exchange interactions between different phases of iron are incompatible[14] and non-Heisenberg exchange interactions become important in defect states[23,39,40]. In the presented work, the magnetic interactions are trained to provide complex interactions beyond conventional functional forms.

We show that our newly developed machine-learned spin-lattice potential (MSLP) is capable of describing the complex magnetic states at highly deformed as well as near-perfect configurations. Our MSLP for iron has good quantitative agreement with DFT data and good computational stability at high temperature simulations. The calculated T_C is also in good agreement with experimental value. We also apply the MSLP to study the magnetic effect of mesoscopic scale dislocation loops in iron at finite temperature, which cannot be achieved using DFT and MD, or using other available machine-learned spin-lattice potentials.

Results

Magnetic states in BCC and FCC structures.

We investigated an essential feature being necessary for a MSLP for iron, which is the relative stability of various magnetic states in BCC and FCC structures. We initialized the ferromagnetic (FM), single-layer antiferromagnetic (SL-AFM), and non-magnetic (NM) states in both BCC and FCC structures, and additionally the double-layer antiferromagnetic (DL-AFM) state in FCC. We relaxed the simulation box and magnetic moments using conjugate gradient method, but with a small mixing step, to ensure the relaxation would stop at a local minimum. Table 1 summarizes our results. It shows the equilibrium lattice constant a_0, the magnitude of spontaneous magnetic moment $|\mathbf{M}|$, and the relative energy difference with respect to the BCC ground state ΔE. DFT data calculated using both VASP[41–44] and OpenMX[45] packages are shown for comparison.

FM BCC is the most stable state. There is small underestimation of the a_0 (-0.5%) and $|\mathbf{M}|$ (-1.4%) compared to VASP data. The DL-AFM is the lowest energy collinear state in FCC, which is 80 meV/atom higher than the FM BCC phase. The energy of other magnetic states are also in quantitative agreement with DFT data. The NM FCC was shown to have free energy lower than NM BCC at all temperatures[14]. Magnetism stabilizes the BCC structure[14–18]. Our MSLP reproduces this phenomenon.

In BCC iron, the formation of spontaneous magnetic moment reduces the energy by 0.42 eV/atom. By varying the magnitude of magnetic moment, we can plot the Landau-functional-like energy well (Supplementary Materials). The position and depth of the minimum for FM state is well reproduced resulting in accurate
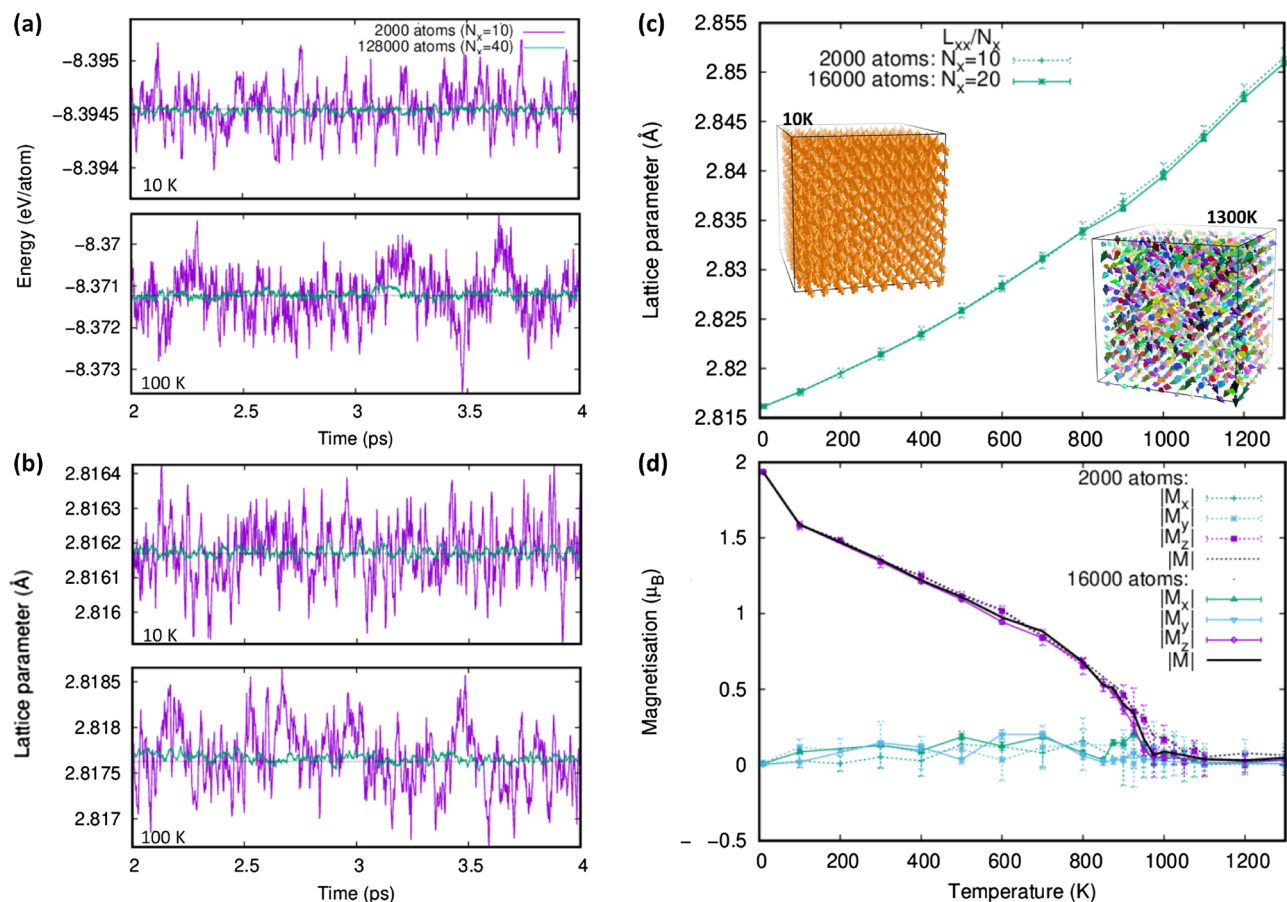
Figure 1. Dynamic stability and properties of the machine-learned spin-lattice dynamics potential (MSLP) for iron. Comparison of instantaneous (**a**) total energies per atom and (**b**) lattice parameter per unit cell between 2000 atom ($10 \times 10 \times 10$) and 128,000 atom ($40 \times 40 \times 40$) simulation cells at 10 K and 100 K. (**c**) Lattice constants and (**d**) magnetization ($|\mathbf{M}| = |\sum_i \mathbf{M}_i|/N$) calculated using our MSLP for iron using a simulation boxes containing 2000 and 16,000 atoms. Standard deviation of data are shown as error bars. Details are in Supplementary materials. Subfigures in (**c**) show snapshots of the ferromagnetic arrangement of magnetic moments at 10 K and the paramagnetic state at 1300 K.

properties of the FM BCC phase. However, a small discrepancy on the profile of the curve compared to DFT data can be observed for small magnetic moments. We note our MSLP predicts a different order of stability of the magnetic states in FCC relative to VASP data, where a low spin FM state has slightly lower energy than the SL-AFM. On the other hand, DFT data from OpenMX predicts the same order of stability as our potential. This highlights the complexity of the potential energy surface of iron where the relative stability of magnetic states is in the order of 0.01 eV.

Our MSLP produced various magnetic states quantitatively as good as the moment tensor spin-lattice potential developed recently by Novikov et al.[38], which is valid only near-perfect crystal collinear regime.

Whilst some perfect HCP configurations were included in the database to help smooth the trained potential energy surface, we would not expect the current parameterisation to perform well for HCP structures since the data was limited and weakly weighted. More details on the comparison of energies, forces, stresses and effective magnetic fields with respect to DFT data are in Supplementary Materials.

Finite temperature properties: lattice constant and Curie temperature. The main purpose of developing a MSLP is to perform dynamic simulations at finite temperature and observe the time evolution of a system. We implemented our MSLP into SPILADY[46]. It allows spin-lattice dynamics to be performed with longitudinal fluctuations of magnetic moments[25,26,47] which is a unique feature of the code and a fundamental concept built into the MSLP.

The initial calculations prove dynamic stability. In spin-lattice dynamics it is important that the potential energy surface is smooth and continuous because both atomic forces and effective magnetic fields are derivatives of the Hamiltonian. A small abnormality may generate unexpected artefacts such as large forces or magnetic fields that destroy the system. Figure 1a shows the total energies of 2,000 and 128,000 atom FM BCC Fe spin-lattice dynamics simulations in NPT ensembles. The magnitude of energy fluctuation is inversely proportional to the number of particles. The average energy of both size runs are equivalent with no evidence of drift. Figure 1b shows the lattice parameters of the same calculations confirming the consistency of the potential with simulation

Defect	Site	MSLP	VASP-PAW	OpenMX[23]	VASP-USSP[51]
⟨110⟩$_{DB}$	Core	−0.28	−0.21	−0.30	−0.18
	Compressive	1.70	1.66	1.87	1.52
	Tensile	2.31	2.37	2.45	2.30
Vacancy	1NN	2.23	2.43	2.53	2.70
	2NN	2.08	2.08	2.13	2.41
	3NN	2.10	2.21	2.24	
	Bulk-like	2.11	2.19	2.22	2.52

Table 2. Magnetic moments in the vicinity of a ⟨110⟩ dumbbell self-interstitial atom and vacancy configurations. The core, compressive and tensile sites refer to the positions defined in the inset of Fig. 2. They are all in unit of Bohr magneton (μ_B). Bulk-like refers to atoms far from the defect core.

size. Scalability is important since simulations of the order 10^5 and larger are beyond the current capability of DFT studies of metallic systems.

We examined the change of lattice constants and T_C of our MSLP for BCC iron. We created cubic simulation boxes containing 2,000 and 16,000 atoms. Figure 1c shows the lattice constant, which is calculated from the time average of the linear dimension of a varying simulation box with pressure set to zero. The lattice constant monotonically increases with a smooth slope as temperature increases. It is generally underestimated but comparable to other MD potentials[48]. The standard deviation, which is shown as the error bar, remains small even at high temperature showing good stability of our potential.

T_C is an unique indicator of a spin-lattice potential. BCC iron undergoes ferromagnetic to paramagnetic phase transition at 1043K[49]. Figure 1d shows the calculated magnetization $|\mathbf{M}| = |\sum_i \mathbf{M}_i|/N$. Calculations were performed in a smaller step of 25K near the T_C. The calculated T_C is around 900K, which is in reasonable agreement with experiment.

We also tried to quench a system from 1000 K, which is in the paramagnetic regime for the parameterisation, to temperatures below T_C. A ferromagnetic monodomain is recovered. However, since spin-orbit coupling has not been considered in the model, there is no magnetic easy axis, so the magnetisation may align along any orientation.

Upon thermalisation an unexpected and unphysical drop in the magnetisation is observed ($T < 100$ K). This artifact of the current parameterisation arises due to the poor reproduction in the curvature of the magnetisation energy (SM-Fig 4) when defect states were included in the training data. As a further consequence, the reduction in the magnitude of the excited magnetic moments reduces their interaction strength, reducing correlation at high temperatures and can partially account for the underestimation of the Curie temperature. Tailoring the Curie temperature during the fitting process is limited as it is not a directly trainable quantity and relates to the curvature of the exchange interactions. Improvements could be made by training future models using more complete feature representations such as spectral neighbours for vector fields derived recently by Domina et al.[34]. One may also observe our model produces the classical profile of the magnetisation curve despite careful training to accurate quantum mechanical data from DFT calculations. This is due to the mapping of the fluctuation-dissipation terms to the Gibbs distribution in the generalised Langevin spin dynamics approach used to evolve our equations of motion[26,47]. Qunatisation of the vibrational thermodynamics could be considered by using the quantum fluctuation-dissipation relations which incorporate Bose-Einstein statistics[50].

Point defects: self-interstitial atom and vacancy. DFT calculations show that in highly distorted lattice structures, complex magnetic configurations can be observed. Existing spin-lattice potentials[14,25,33,38] remain incapable of capturing such phenomena even for point defects. The magnitudes of magnetic moments near the defect core can be suppressed or even in reverse alignment with respect to the bulk[51]. Models that only adopt the Heisenberg Hamiltonian cannot produce physically correct point defect migration as the model does not allow change of magnitudes of magnetic moments according to the change of local electronic structure[23]. Additional Landau terms in the Hamiltonian which are functions of local environment may be a solution, but correct treatment of itinerant properties remain unresolved[23].

In BCC iron, the most stable SIA configuration is a ⟨110⟩ dumbbell configuration[21,22]. Using our MSLP, we performed annealing simulations with initial configurations including either a ⟨110⟩ or ⟨111⟩ dumbbell, in a cell with 2001 atoms. The cells were initially thermalised to 10K and gradually decreased to 0K for 5ps. Both SIA configurations relaxed to maintain/form a ⟨110⟩ dumbbell. We can understand this through nudged elastic band DFT calculations which show no intermediary energy barriers across the migration pathway between the ⟨111⟩ and ⟨110⟩ SIA configurations (see Supplementary Materials). A ⟨111⟩ SIA configuration will inevitably relax to a ⟨110⟩ dumbbell when small perturbations exist. A snapshot of a spin-lattice dynamics simulation of the ⟨110⟩ configuration at 10K is shown in Fig. 2a. The magnetic moments were plotted with unit magnitude for ease of viewing. Their magnitudes are represented by colour.

We examined the magnetic configuration in the core of a ⟨110⟩ dumbbell. The magnetic moments in and surrounding the core are listed in Table 2. It shows very good agreement in comparison to DFT calculations: VASP with PAW pseudopotentials (current work), VASP with ultrasoft pseudopotentials (USPP)[51], and OpenMX[23]. The magnitude of the magnetic moments within the defect core are larger than the VASP-PAW data that the potential was trained to, but is similar to those produced by OpenMX. Likewise, for the tensile site the magnetic moments
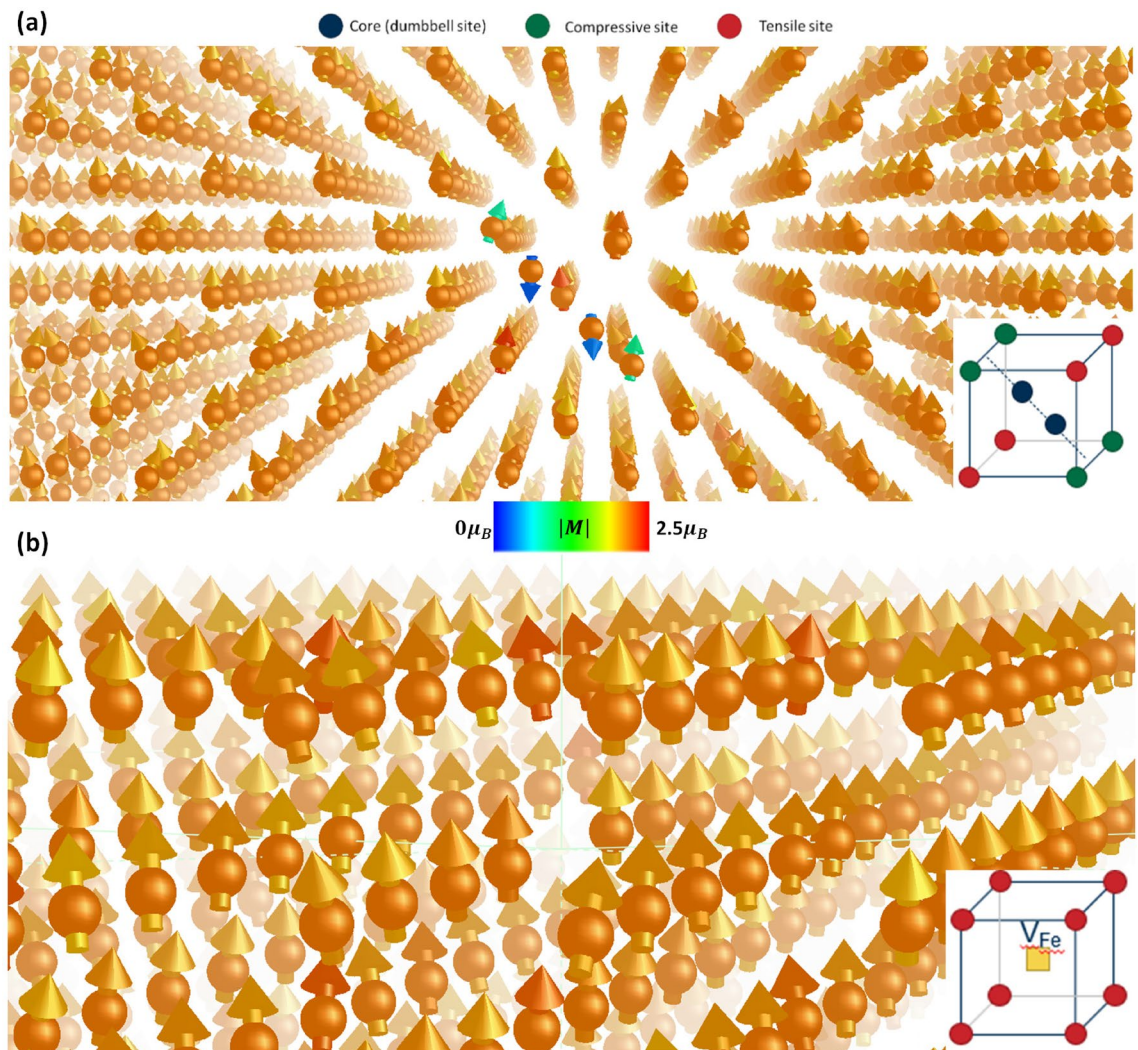
4

Figure 2. A snapshot of (**a**) a ⟨110⟩ dumbbell self-interstitial atom configuration and (**b**) a mono-vacancy in iron at 10 K. Magnitudes of magnetic moments are indicated according to the colour bar. The insets shows the schematic configurations. For the SIA, the schematic indicates the core (blue), compressive (green) and tensile (red) sites.

predicted by the MSLP are smaller than VASP-PAW data but are comparable to VASP-USSP data. Generally, the magnetic moments are reproduced in quantitative agreement with DFT calculations. In the core of the interstitial defect, magnitudes of magnetic moments are approximately 1/10th of bulk and in anti-alignment to the bulk. Enhanced magnitudes can be observed on the tensile sites and slightly reduced magnitudes on the compressed sites. In additional to the most stable configuration, our MSLP reproduced the correct order of stability of SIA, i.e. the formation energy of ⟨110⟩ < tetrahedral < ⟨111⟩ < ⟨100⟩ < octhahedral (see Supplementary Materials).

Another point defect that we explored is the mono-vacancy (V_{Fe}). Annealing simulations were performed using a 1999 atoms cell containing a single vacancy. Table 2 shows the calculated values of magnetic moments in the vicinity of the defect site. Figure 2b shows a snapshot of the system near the vacancy during dynamics. DFT calculations indicate that the magnetic moments directly adjacent to a vacancy are larger. This occurs due to the increased volume to which their moments can relax. The magnitude of the magnetic moments in the 1st nearest neighbour (NN) sites are approximately 11%, 14% and 6.7% larger for the VASP-PAW, OpenMX and VASP-USSP calculations. The increase is only 5.6% using the MSLP. Conversely, the magnetic moments of the 2nd NN to the vacancy have reduced magnitudes. Our potential predicts a reduced magnetic moment relative to bulk in line with DFT calculations, but the proportion is diminished. By the 3rd NN sites, the magnetic moments are *bulk-like* in all cases. Our MSLP predicted the correct trend of the changes, but generally gives a smaller value.

Extended defects: prismatic dislocation loops.

We applied our MSLP to sizable systems that cannot be addressed by DFT. We constructed two simulation cells which are pre-relaxed using the Malerba 2010 Fe potential[52] through the conjugate gradient implementation in LAMMPS. In the first cell we created a square SIA loop with **b** = a_0[001] consisting of 265 atoms in a box containing 128,265 atoms. In the second, a circular SIA loop with **b** = $\frac{a_0}{2}$[111] was constructed with 261 atoms in a box containing 139,287 atoms. The relaxed prismatic

dislocation loops identified using the dislocation extraction algorithm (DXA) are shown in Fig. 3(i)a for the $\langle 100 \rangle$ loop and Fig. 3(ii)a for the $\frac{1}{2}\langle 111 \rangle$ loop.

We chose these dislocation loops as representative examples because both kinds of loop can be experimentally observed in α-iron. Iron is known to be anomalous, forming $\langle 100 \rangle$-type prismatic edge dislocations at temperatures above 550°C[53,54] despite the isotropic elasticity favoring dislocation loops with smaller Burgers vectors such as $\frac{1}{2}\langle 111 \rangle$. Analytic linear elasticity solution suggests the softening of C', which is a magnetic effect, accounts for the observation of square $\langle 100 \rangle$ loops at high temperature[55,56].

The MSLP offers analysis of magnetic excitation in the vicinity of these extended defects for the first time. As such, we performed spin-lattice dynamics calculations in NPT ensembles at both 10K and 800K using the MSLP, evaluating the local stress and magnetic configurations of both loop types. Data for the $\langle 100 \rangle$ prismatic loop at 10 K are shown in Fig. 3(i) whereas the $\frac{1}{2}\langle 111 \rangle$ loop results are in Fig. 3(ii). For both loop types we present positive and negative isosurfaces of the stress field introduced by the defects. Specifically, we evaluated $\text{Tr}(\sigma_{ij}^k)$, where σ_{ij}^k is the Virial stress tensor of atom k computed using our MSLP.

Yellow/blue isosurfaces show the compressed/tensile regions where atoms contribute $\text{Tr}(\sigma_{ij}^k)=\pm 0.017$ GPa to the stress for the $\langle 100 \rangle$ loop and $\text{Tr}(\sigma_{ij}^k)=\pm 0.024$ GPa for the $\frac{1}{2}\langle 111 \rangle$ loop. (c) and (d) show the contour maps of $\text{Tr}(\sigma_{ij}^k)$ and magnetic moment magnitudes on a (100) or ($\bar{1}2\bar{1}$) plane intersecting the centre of the dislocation loops. (e) shows the magnetic moment vectors of atoms near the loop superimposed on the contour map. To visually compare the stress and magnetic configuration, we present an overlay of the stress isosurface with a snapshot of the non-collinear magnetic moments. The magnetic moments are coloured according to their magnitude, with red/orange hues representing oversized moments, blue/dark-green hues for small and green moments for bulk-like magnitudes ($|\mathbf{M}| \approx 2.0\mu_B$). Vector fields of the time averaged moments with the same colour scheme are presented in (g) highlighting the magnetic moments in the core region of the dislocation. A strong correlation is evident between the local stress and the magnetic moments. Regions under tension, which have a comparatively larger volumes per atom, cause magnetic moments to increase relative to their bulk value. On the other hand, regions which are compressed result in reduced magnitudes of the magnetic moments. This is most evident in the core of the dislocation line.

Data pertaining to the $\langle 100 \rangle$ and $\frac{1}{2}\langle 111 \rangle$ prismatic loops at 800 K are presented in Fig. 4i and ii, respectively. Despite the simulations operating at high temperature on structures far from the training data, the simulations remain stable and well behaved. To smooth variations due to thermodynamic perturbations, the isosurfaces presented in (a) represent $\text{Tr}(\langle \sigma_{ij}^k(t) \rangle)$, where $\langle \sigma_{ij}^k(t) \rangle$ is the time averaged Virial stress over the simulation. (b) and (c) present contour maps of the time averaged stress ($\text{Tr}(\langle \sigma_{ij}^k(t) \rangle)$) and magnitudes of the magnetic moments on a (100) or ($\bar{1}2\bar{1}$) plane cross section through the dislocation loops. Snapshots of the non-collinear magnetic moments in and near the dislocation loops during dynamics at 800 K are shown in (d). The magnetic moments near the defects become highly disordered relative to the bulk-like atoms. Importantly, the time averaged magnetic moments shown in (e) for the tensile region of the dislocation are already acting paramagnetic despite the sample being below the Curie temperature.

Discussion

A new machine-learned spin-lattice potential (MSLP) for iron that can simultaneously simulate the mechanical and magnetic responses at finite temperature for both near-perfect and highly distorted configurations is developed. It is achieved through combining the knowledge of a conventional spin-lattice potential and a neural network term implemented using both local atomic and magnetic descriptors. Each magnetic moment is a three dimensional vector, where both the direction and magnitude depend on the local atomic environment and can be perturbed by thermal excitation.

Our MSLP shows near DFT accuracy on perfect crystals and point defect configurations. It produces quantitatively accurate predictions of various magnetic states in both BCC and FCC phases. The complex magnetic configurations in the vicinity of the core of vacancy and self-interstitial atom configurations, including the magnetic moment reversal and quenching at the core, were correctly reproduced. The order of stability of SIA configurations is compatible with DFT, where the $\langle 110 \rangle$ dumbbell is most stable.

Spin-lattice dynamics is performed to calculate the Curie temperature, which is in good agreement with experiment[49]. We apply our potential to study the magnetism of mesoscopic scale dislocation loops at finite temperature. Non-collinear magnetic moments about prismatic dislocation loops are investigated for the first time. We show moment magnitudes are suppressed in regions of compressive stress and are enhanced in regions of tensile stress. This transcends the capability of DFT and MD methods, as well as currently available MSLPs for iron. These simulations show good numerical stability at high temperature. Whilst the current MSLP is tailored to iron, the framework is flexible and can be applied to a large class of magnetic materials and alloys.

Methods

Hamiltonian of machine-learned spin-lattice potential. In many other developments of machine-learned potentials[4,9,33,38], the potential energy is defined as the output of a machine-learned machinery without any presumptions. The difference in the energy landscape can be up to the order of several eV, whilst requiring the accuracy and precision to be within at least 10^{-3} eV. Smoothness of the energy landscape is also a requirement of machine-learned potentials because atomic forces are calculated as the derivative of the potential energy. It necessitates a broad coverage of training data, especially near extrema. One can optimize their machine-learned potential by supplying sufficient data to cover important parts of the phase space[38] or generate massive amount of data in brute-force to cover the whole phase space[33].
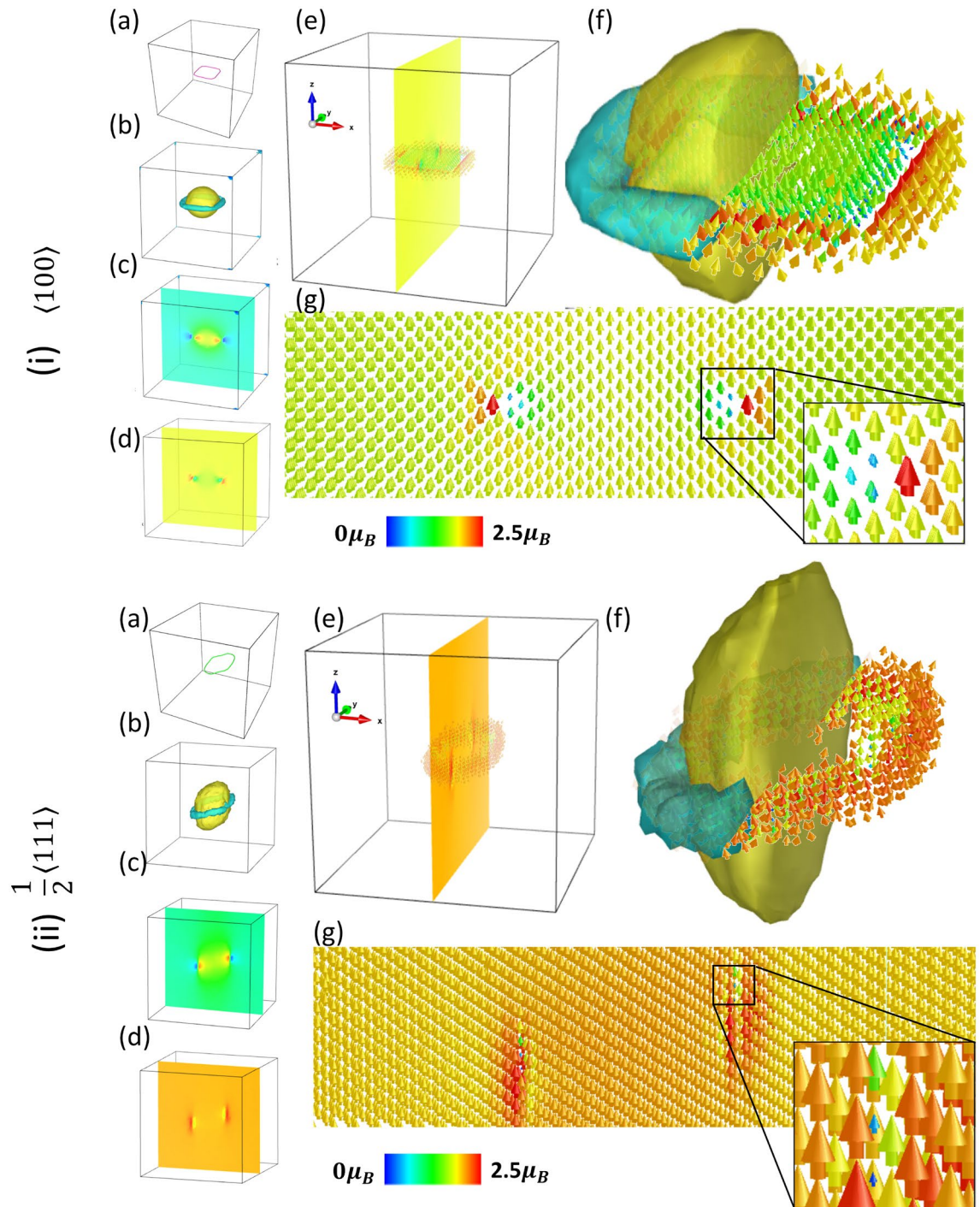
Figure 3. Spin-lattice dynamics simulations performed at 10 K for the (i) ⟨100⟩ and (ii) $\frac{1}{2}$⟨111⟩ interstitial dislocation loops. (**a**) Dislocation identified using the dislocation extraction algorithm (DXA). (**b**) The compressive and tensile stresses caused by the dislocations are shown via the isosurface of $\text{Tr}(\sigma_{ij}^k)$, where σ_{ij}^k is the Virial stress tensor of atom k. (**c**) 2D contour plot of $\text{Tr}(\sigma_{ij}^k)$ through the (010) or ($\bar{1}2\bar{1}$) plane bisecting the centre of loop. (**d**) 2D contour plot of the magnetic moment magnitudes on the same plane as (**d**). (**e**) Contour plot of magnetic moment magnitudes overlayed with a sample of the instantaneous magnetic moment vectors for atoms near the loop. (**f**) Overlay of magnetic moments near the loop with the stress tensor isosurface showing relation between compressive (blue) and tensile (yellow) regions with large (red/orange) and small (blue/green) magnetic moments (see colourbar). (**g**) Vector field of the magnetic moments near the dislocation loops highlighting the core of the dislocations with suppressed magnetic moments and the enhanced moments directly adjacent.
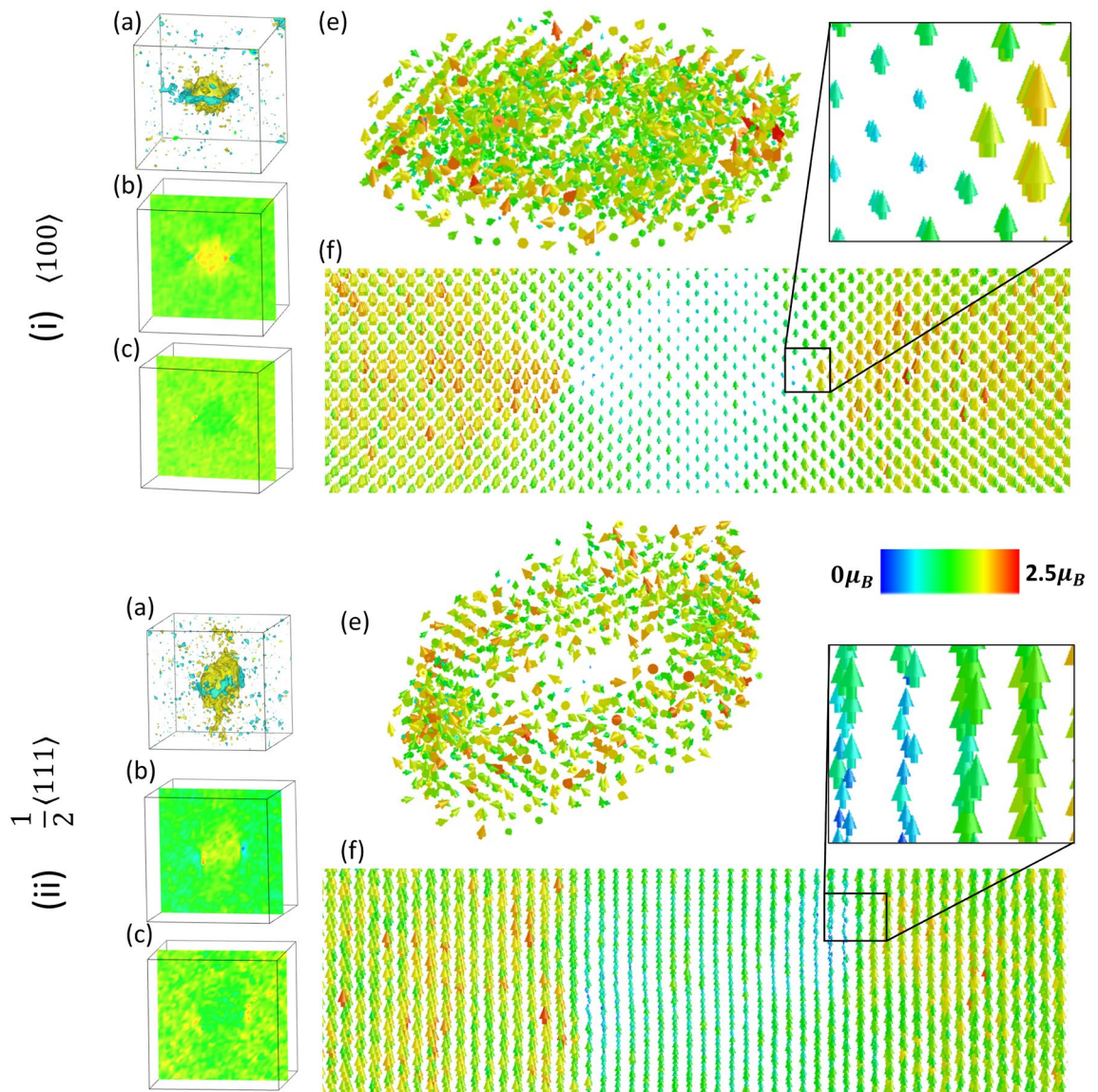
Figure 4. Spin-lattice dynamics simulations performed at 800 K for the (i) $\langle 100 \rangle$ and (ii) $\frac{1}{2}\langle 111 \rangle$ interstitial dislocation loops. (**a**) The compressive and tensile stresses caused by the dislocations are shown via the isosurface of $\mathrm{Tr}(\langle \sigma_{ij}^{k}(t) \rangle)$, where $\langle \sigma_{ij}^{k} \rangle$ is the time averaged Virial stress tensor of atom k. (**b**) 2D contour plot of $\mathrm{Tr}(\langle \sigma_{ij}^{k}(t) \rangle)$ through the $(0\bar{1}0)$ or $(\bar{1}2\bar{1})$ plane bisecting the centre of loop. (**c**) 2D contour plot of the magnetic moment magnitudes on the same plane as (**b**). (**d**) Sample of the instantaneous magnetic moments during dynamics of atoms in and near the dislocation loop. (**e**) Vector field representing the averaged magnetic moments near the loop.

On the other hand, if one can supply a *mean function* before performing the learning process, the machine-learned machinery can then be used as a correction term. A properly chosen mean function can significantly reduce the amount of training data[57]. We follow this logic and define our spin-lattice Hamiltonian as follows:

$$\mathcal{H}(\mathcal{R}, \mathcal{P}, \mathcal{M}) = \sum_i \frac{\mathbf{p}_i^2}{2m} + V(\mathcal{R}, \mathcal{M}), \tag{1}$$

where

$$V(\mathcal{R}, \mathcal{M}) = V^{\mathrm{NM}}(\mathcal{R}) + V^{\mathrm{HL}}(\mathcal{R}, \mathcal{M}) + V^{\mathrm{NN}}(\mathcal{R}, \mathcal{M}). \tag{2}$$

The Hamiltonian depends on the momenta $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N\}$, atomic positions $\mathcal{R} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N\}$ and magnetic moments $\mathcal{M} = \{\mathbf{M}_1, \mathbf{M}_2, ..., \mathbf{M}_N\}$. The potential energy V contains three terms. The non-magnetic term V^{NM}, which adopts the embedded atom method (EAM) functional form, takes care of the non-magnetic contributions. The Heisenberg-Landau (HL) term V^{HL} takes care part of the magnetic contribution. The neural

network (NN) term V^{NN} takes care of contributions missed in V^{NM} and V^{HL}. We may think of its application as a correction term. The functional form of the first two terms follow conventional spin-lattice potentials[14,26] that performs well near perfect crystal, but not in highly deformed configurations.

The NN term is defined as a sum of local contributions:

$$V^{\text{NN}} = V_0^{\text{NN}} \sum_i \mathcal{N}(\mathbf{x}_i^0; \{\mathbf{W}, \mathbf{b}\}), \tag{3}$$

where V_0^{NN} is a fitting parameter to correct the energy unit. The neural network $\mathcal{N}$ is trained by adjusting the weight $\mathbf{W}$ and bias $\mathbf{b}$ parameters. The translation, rotation and permutation invariant atom centred symmetry (ACS) descriptors $\mathbf{x}_i^0$ of atom i are extended to depend on both the local atomic and magnetic environments. This follows the usual assumption that the local environment is sufficient to determine the atomic energy[6], where a cutoff distance r_{cut} is adopted.

Details of each term in the Hamiltonian are defined below. The potential energy surface V of our MSLP was trained to a large amount of DFT data consisting of energies, forces, stresses and effective fields from 4,601 non-polarised, 5,827 collinear and 2,245 non-collinear configurations. A more detailed audit of the training data is available in the Supplementary Materials (Section 1). The choice of Loss function, training data, training procedure, and final parameters are provided in the Supplementary Materials. In short, we developed a non-magnetic potential V^{NM}, followed by fitting parameters in V^{HL} and V^{NN}. Finally, we optimized the whole potential V. One may consider V^{NM} and V^{HL} are used to construct a temporary mean function. On the contrary, each function in the NM and HL terms may be considered as a descriptor. As such, we can treat the V as a special kind of machine-learned machinery that flexibly combines the known and unknown physics when trained to good quality data.

The non-magnetic term. The Hamiltonian being adopted contains several terms, the non-magnetic term is chosen to have the same functional form of the embedded atom method (EAM)[58,59]:

$$V^{\text{NM}}(\mathscr{R}) = \sum_i F(\rho_i) + \frac{1}{2} \sum_{i,j} V_{ij}(r_{ij}). \tag{4}$$

$F(\rho_i)$ is a many-body term depending on the effective electron density ρ_i. V_{ij} is a pairwise potential depending on $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, which is the distance between atom i and j. The many-body term follows the functional form proposed by Mendelev[60] and Ackland[61]:

$$F[\rho_i] = -\sqrt{\rho_i} + \phi \rho_i^2, \tag{5}$$

where ϕ is a fitting parameter. The effective electron density ρ_i is defined as a sum of the square of a pairwise function t_{ij} which has the physical meaning corresponding to the hopping integral in tight binding model[14]:

$$\rho_i = \sum_{j \in \{k | r_{ik} < r_{\text{cut}}\}} t_{ij}^2, \tag{6}$$

and

$$t_{ij}(r_{ij}) = \sum_n^{N_t} t_n (r_n^t - r_{ij})^3 \Theta(r_n^t - r_{ij}), \tag{7}$$

where t_n are fitting parameters, r_n^t are knot points, and Θ is Heaviside function. The pairwise potential V_{ij} is split into three parts:

$$V_{ij} = \begin{cases} V^{ZBL}(r_{ij}), & \text{if } r_{ij} \leq r_1, \\ V^{\text{it}}(r_{ij}), & \text{if } r_1 < r_{ij} < r_2, \\ \sum_n^{N_V} V_n (r_n^V - r_{ij})^3 \Theta(r_n^V - r_{ij}), & \text{if } r_{ij} \geq r_2, \end{cases} \tag{8}$$

where $r_1 = 1.3$Å and $r_2 = 1.8$Å. The short-range part is ZBL potential[62]. The middle-range part is a 5$^{\text{th}}$ order polynomial which ensures the function V_{ij} being continuous up to second derivatives at r_1 and r_2. The longer-range part is a cubic spline, where V_n are fitting parameters and r_n^V are knot points. Numerical values of fitting parameters are mentioned in Supplementary Materials.

The Heisenberg-Landau term. The Heisenberg-Landau term is a sum of a Heisenberg term V^{H} and a Landau term V^{L}, such that

$$V^{\text{HL}}(\mathscr{R}, \mathscr{M}) = V^{\text{H}}(\mathscr{R}, \mathscr{M}) + V^{\text{L}}(\mathscr{R}, \mathscr{M}). \tag{9}$$

Conventional Heisenberg Hamiltonian assumes localised electron model with fixed magnitudes of magnetic moments or spins[25,27,33,63]. However, even for perfect crystalline configurations it has been observed that the adiabatic magnetic exchange-energy hypersurface parameterized by the bilinear Heisenberg Hamiltonian is insufficient[64–67]. An accurate representation necessitates longitudinal fluctuations to be considered[26,66,68], due to

the itinerant nature of electrons. First, we write the Heisenberg Hamiltonian in a form that allows the change of magnitude[23,69]:

$$V^{\mathrm{H}} = -\frac{1}{2} \sum_{ij} J_{ij}(r_{ij}) \mathbf{M}_i \cdot \mathbf{M}_j. \tag{10}$$

The pairwise exchange coupling parameter J_{ij} can be calculated through DFT according to the magnetic force theorem (MFT)[70]. We adopt a 5th order polynomial here that fits well to perfect BCC cases[14]:

$$J(r_{ij}) = J_0 \left(1 - \frac{r_{ij}}{r_{\mathrm{cut}}}\right)^5 \Theta(r_{\mathrm{cut}} - r_{ij}). \tag{11}$$

Second, the Heisenberg term can be improved by introducing higher order terms that describe longitudinal fluctuations[23,26,68,71]. By using a Landau expansion, we introduce self-energy terms which create an energy well for a finite magnetic moment such that a spontaneous moment is formed and whose length can be variably controlled. We write the Landau term:

$$V^{\mathrm{L}} = \sum_i \left(A(\rho_i)\mathbf{M}_i^2 + B(\rho_i)\mathbf{M}_i^4 + C\mathbf{M}_i^6\right). \tag{12}$$

One can find more details regarding the philosophy of the Landau coefficients and how one may extract them directly from DFT calculations in Ref.[14,23]. Here, we simply treat them as fitting parameters. We assume an underlying quadratic polynomial functional form for both A and B coefficients, parameterised with respect to ρ_i used in the EAM potential (Eqn. 6):

$$A(\rho_i) = a_0 + a_1\rho_i + a_2\rho_i^2 \tag{13}$$

$$B(\rho_i) = b_0 + b_1\rho_i + b_2\rho_i^2 \tag{14}$$

The coefficient for the 6th order term is independent of the local environment and serves to prevent a divergence in the Landau energy well. It has been shown that such functional form is sufficient for strained on-lattice configurations, but is insufficient when lattice distortions are introduced[23]. Further, DFT calculations have shown the magnitude of Landau coefficients in the core of defects can change by several orders of magnitude due to the suppression of the magnetic moments[23]. As such, these terms provide an initial approximation. The neural network term serves as a necessary adjustment allowing the potential to move away from near-perfect crystal.

It is necessary to note that the inclusion of self-interactions terms is not solely sufficient to include longitudinal fluctuations in a spin-lattice dynamics model when using Langevin dynamics. One cannot use the stochastic Landau-Lifshitz equations to evolve the system, as is typically employed in most available codes, since the magnitudes of magnetic moments are assumed constant. Instead, the fluctuation-dissipation relations must be re-derived from the Fokker-Planck equation to correctly account for the longitudinal fluctuations. A generalised Langevin spin dynamics approach and an expression for the dynamic spin temperature have been derived in Ref.[26] valid for any spin Hamiltonian with longitudinal fluctuations.

The neural network term. Conventional spin-lattice potentials are insufficient to reproduce the relative stability of BCC and FCC phases. The magnetic force theorem reveals the exchange coupling parameter J_{ij} has completely different functional form for each crystal structures[14,23]. Previous work[14] defined two different set of J_{ij} and Landau coefficients for each phase, such that the phase must be labelled *a priori*, allowing free energy differences between the BCC and FCC phase to be extracted. It means such approach cannot be applied to arbitrary systems. A possible alternative is to calculate the J_{ij} and Landau coefficients by DFT on the spot, but it is not feasible for large-scale atomic scale simulation. Besides, calculation of atomic Landau coefficient requires knowing the atomic energy, which is not a well defined quantity in most DFT implementations.

To overcome the limitation imposed by the functional form, aiming at simulating arbitrary crystal structures, we apply machine learning techniques to develop a new potential. We choose artificial neural network (ANN) which abstractions between layers ensure magnetic interactions go beyond the bilinear form of the Heisenberg potential and local fluctuation of the Landau potential.

Behler and Parinello[72] and others[4,73] successfully applied the ANN for atomic simulation based on feed-forward multilayer perceptrons. It composes of multiple layers of Threshold Logic Units (TLUs). They are fully connected between adjacent layers. It is a feed forward ANN in which data provided to the input nodes are transmitted through one or more hidden layers until producing an output signal at the final layer. Unlike other architectures such as recurrent NN, cyclical connections between layers are not used. When two or more hidden layers are used, it is often referred to as a deep-ANN (DNN)[74]. We simply call it NN in this work.

In some works of machine-learned potentials for MD[3,35–37] a single machine-learned machinery, such as Gaussian process or NN, is used to predict the total energy, or more precisely the energy of an atom depending on the local atomic environment. Instead, we use NN to predict the contribution that cannot be captured by the non-magnetic term and Heisenberg-Landau term. In addition to the V^{NM} and V^{HL} term, the potential energy contains a NN term:

10

$$V^{\text{NN}} = V_0^{\text{NN}} \sum_i \mathcal{N}(\mathbf{x}_i^0; \{\mathbf{W}, \mathbf{b}\}), \tag{15}$$

where V_0^{NN} is a fitting parameter to match the scale and unit of the NN contribution to the MSLP Hamiltonian. $\mathbf{x}_i^0 \in \mathbb{R}^{N_0}$ is a vector of descriptors with N_0 elements being supplied to the input layer of NN. Descriptors are functions depending on the atomic positions and magnetic moments within a cutoff distance r_{cut} from atom i, representing the local atomic environment. The NN with n layers is a mapping:

$$\mathcal{N}(\mathbf{x}^0; \{\mathbf{W}, \mathbf{b}\}) = \mathscr{P}^n \circ \mathscr{P}^{n-1} \circ \mathscr{P}^{n-2} \circ \ldots \circ \mathscr{P}^0(\mathbf{x}^0), \tag{16}$$

where the operator $\circ$ represents the composition of functions. $\mathscr{P}^k$ is the mathematical description of a perceptron at layer k. It acts as a mapping from layer $k - 1$ to the adjacent layer indexed k ($0 \leq k \leq n$) which includes the composition of a linear transformation, followed by a non-linear transformation using a component-wise activation function $\mathbf{f}_a^k$:

$$\mathbf{x}^k = \mathscr{P}^k(\mathbf{x}^{k-1}) = \mathbf{f}_a^k(\mathbf{z}^k), \tag{17}$$

where

$$\mathbf{z}^k = \mathbf{W}^k \mathbf{x}^{k-1} + \mathbf{b}^k. \tag{18}$$

$\mathbf{x}^k \in \mathbb{R}^{N_k}$ is a vector representation of the input signals from each of the N_k nodes (neurons) in the kth layer of the NN. The weight matrix $\mathbf{W}^k \in \mathbb{R}^{N_k \times N_{k-1}}$ controls the strength of the signal transferred from each node in the $k - 1^{\text{th}}$ layer to each node in the kth layer. $\mathbf{b}^k \in \mathbb{R}^{N_k}$ is a bias vector. The vector $\mathbf{z}^k$ is an intermediate quantity referred to as the weighted input.

The activation function acts to abstractify the signals from the kth layer by adding non-linearity (since a linear combination of linear operations can itself be transformed into a single linear operation). It performs a component-wise operation on the weighted input $\mathbf{z}^k$ produced from the linear transformation of $\mathbf{x}^{k-1}$, such that

$$\mathbf{f}_a^k(\mathbf{z}^k) = \left(f_a(z_0^k), f_a(z_1^k), \ldots, f_a(z_{N_{k-1}}^k) \right). \tag{19}$$

We have chosen to use an unconventional unbounded activation function defined as:

$$f_a(z) = \frac{z}{1 + |z|} + az, \tag{20}$$

for all TLUs, except the output layer. The functional form of the activation function was chosen since the profile of $z/(1 + |z|)$ is qualitatively similar to $\tanh(z)$ for small z (approximately linear), but it is computationally cheaper than the hyperbolic tangent. Linear twisting is included to help prevent saturation for large values of z, which would result in a vanishing gradient of the Loss function, as originally proposed for the tanh function[75]. The mapping of the final layer performs a linear transformation only, that is $x^n = f_a^n(z^n) = z^n$ and produces a scalar output. Therefore, for n^{th} layer the weight and bias terms have dimension $\mathbb{R}^{1 \times N_{n-1}}$ and $\mathbb{R}^1$, respectively.

The power of ANNs are due to their universality. For instance, a two-layer feed forward ANN with non-linear activation functions have been shown to be an universal function approximator. As such, for a given continuous function there exists a neural network which can approximate it on a compact set of $\mathbb{R}^N$ arbitrarily well[76]. Furthermore, the universal approximator theorem has been shown to hold true for unbounded non-linear activation functions[77]. If linear activation functions were to be chosen, a DNN with any number of hidden layers may be represented as a single linear transformation and therefore cannot be a universal approximator.

Local atomic and magnetic descriptors.
In our MSLP, descriptors are functions representing the $6N$ coordinate and spin space, where N is the number of atoms. The purpose of the NN term is to map descriptors to part of the atomic energy. Since atomic energy is a scalar, descriptors should be translational, rotational and permutational invariant. We defined four sets of descriptors. The first set depends only on atomic positions. The other three depend on both atomic positions and magnetic moments.

Our atomic descriptors are based on the $G^{(2)}$ radial basis functions within the Atom Centred Symmetry class of effective coordinates[4,72]. We drop the (2) superscript for brevity and refer to the descriptor as G2 in-text. It has been successfully used for a variety of materials including water[78,79], aluminium and its alloys[80], germanium telluride[81] as well as carbon allotropes[82]. It is written as:

$$G_{i,h} \triangleq G_{i,h}^{(2)} = \sum_j G_{ij,h}, \tag{21}$$

where

$$G_{ij,h} = w_i w_j f_c(r_{ij}) \exp\left(-\eta(r_{ij} - R_s)^2\right), \tag{22}$$

and h is a compound index representing a unique triplet of hyperparameters $\{R_c, R_s, \eta\}$. The G2 descriptors are R_s centred Gausssians spread according to η. w_i and w_j are weights which characterise different atomic species and are not uniquely defined. Often one maps a unique integer to each element type. Here we define them as the atomic mass (for iron $w = 55.847$). Use of a species weight is advantageous as it enables descriptors to be

defined which do not scale with the number of species. That is, the length of the input descriptor vector does not change with the number of chemical species.

The smoothness criterion is satisfied by employing an envelope function which, as well as its first derivative, decays smoothly to zero at the cutoff radius:

$$f_c(r_{ij}) = \begin{cases} \frac{1}{2}\left(\cos\left(\frac{\pi r_{ij}}{R_c}\right) + 1\right), & \text{if } r_{ij} \leq R_c, \\ 0, & \text{if } r_{ij} > R_c. \end{cases} \tag{23}$$

In this work we fix R_c making it equals to the cutoff distance of the pair potential r_{cut}. It reduces the number of hyperparameters to 2. We used 9 equally spaced G2 descriptors with $R_s = 2.0 + x(R_c - 2.0)/8$ where $x = 0, 1, 2, ..., 9$. We set $\eta = 5$ Å^{-2} to provide a small overlap between the Gaussian basis functions. Often a large number of G2 descriptors (5-200) are used varying η from 10^{-2} to 1 Å^{-2} with $R_s = 0$ Å[3]. We opted to fix the Gaussian width and varying their centering, in order to reduce the correlation between the data encoded by each descriptor. Reducing the correlation can also be achieved using more advanced orthogonal descriptors such as SOAP[6] at the expense of greater computation time per descriptor. Since we consider $6N$ degrees of freedoms, we chose G2 descriptor as it is computationally less demanding.

The design of the magnetic descriptors is based on the G2 function. Inspired by the Heisenberg and Landau functional forms, we write three further sets of descriptors. The first set of magnetic descriptors is written as:

$$G_{i,h}^H = \sum_j G_{ij,h}^H, \tag{24}$$

where the 2-body contributions are defined as:

$$G_{ij,h}^H \triangleq G_{ij,h}\mathbf{M}_i \cdot \mathbf{M}_j. \tag{25}$$

The scalar product of the magnetic moments ensures the invariant properties are maintained. Smoothness is guaranteed by the G2 prefactor which contains the envelope function f_c. By reusing the G2 in the magnetic descriptors we reduce the computational cost of the descriptor calculations which must be performed for every atom, at every timestep if dynamics is to be performed. Each G2 in the Heisenberg-like descriptors may be considered to be surrogate exchange parameters with different dependencies on the local environment as set by the chosen hyperparameters.

Similarly, we defined descriptors inspired by the Landau term up to the 4$^{\text{th}}$ order. A 6$^{\text{th}}$ order term is provided in the classical Landau expression to prevent divergences. As with the Heisenberg-like term, the Landau-like descriptors are built from a sum of two body contributions:

$$G_{ij,h}^A \triangleq G_{ij,h}\mathbf{M}_i^2, \tag{26}$$

$$G_{ij,h}^B \triangleq G_{ij,h}\mathbf{M}_i^4. \tag{27}$$

Each hidden layer of NN provides successively higher order representations of the exchange interactions beyond the original bilinear, quadratic and quartic input interactions.

We use nine G2 radial basis functions. The input of the NN has a dimension of $\mathbb{R}^{N_0}$, where $N_0 = 4 \times 9 = 36$. That is, nine structural descriptors $\{G^{(2)}\}$, nine Heisenberg-like descriptors $\{G^H\}$, nine Landau-A-like descriptors $\{G^A\}$ and nine Landau-B-like descriptors $\{G^B\}$. Every descriptor has analytical derivatives with respect to both changes in position and magnetic moment (see Supplementary Materials). Whilst we opted to use two-body G2 as the basis of our magnetic descriptors, the principle is extendable to N-body descriptor representations.

Fitting procedure. Once the database has been constructed, the model parameters can be trained by minimising the Loss function (see Supplementary Materials). Each component in the model Hamiltonian is motivated by different physical properties. To reflect this our fitting workflow consisted of four distinct stages to allow each term to learn their respective physical behaviours.

1. First, we introduce the underlying behaviour of metallic bonds for BCC, FCC and HCP iron in the absence of magnetic interactions by fitting the parameters of the non-magnetic potential $V^{\text{NM}}(\mathbf{p}^{\text{NM}})$ to the configurations in the non-magnetic database. The non-magnetic parameters are the subset $\mathbf{p}^{\text{NM}} = (\{V^t\}, \{r^t\}, \{t^N\}, \{r^t\}, \phi)$. We maintain the parameters of the ZBL potential. The coefficients of the interpolation potential V^{it} are not fit but are analytically derived to maintain continuity.
2. Next, the characteristic behaviour of *band splitting* (i.e. the spontaneous formation of a magnetic moment) and their itinerant magnetic interactions are added by fitting the Heisenberg-Landau parameters $\mathbf{p}^{\text{HL}}$ to bulk-like BCC configurations in the magnetic database. During this process $\mathbf{p}^{\text{NM}}$ are held constant such that the total energy considered by the Loss function is $\mathscr{U} \rightarrow V^{\text{NM}} + V^{\text{MC}}(\mathbf{p}^{\text{HL}})$. The Heisenberg-Landau parameters are $\mathbf{p}^{\text{HL}} = (J_0, \{a\}, \{b\}, c)$.
4. Magnetic interactions beyond the parametric constraints of the Heisenberg-Landau formalism are produced by training the NN weights and biases $\mathbf{p}^{\text{NN}} = (\{\mathbf{W}\}, \{\mathbf{b}\}, V_0)$ to all desired observables in the magnetic database. This also introduces the magnetic behaviour of the FCC phase into the Hamiltonian. During this stage

the total energy is given by the full model $\mathscr{U} \rightarrow V^{NM} + V^{HL} + V^{NN}(\mathbf{p}^{NN})$, where the parameters $\mathbf{p}^{NM}$ and $\mathbf{p}^{HL}$ are fixed.

5   Finally, we make minor adjustments to the parameter space by enabling all variables $\mathbf{p} = (\mathbf{p}^{NM}, \mathbf{p}^{HL}, \mathbf{p}^{NN})$ to be simultaneously adjusted with total energy: $\mathscr{U} \rightarrow V^{NM}(\mathbf{p}^{NM}) + V^{HL}(\mathbf{p}^{HL}) + V^{NN}(\mathbf{p}^{NN})$. In this stage the maximum step size of the minimization algorithm is reduced.

Once the Loss function has been minimised with respect to the training database its generalisation and stability may be validated through dynamic simulations. This is usually performed using MD. However, our model Hamiltonian has coupled spin and lattice degrees of freedoms and is designed to incorporate itinerant behaviour. Consequently, in order to evolve with time and temperature we may use spin-lattice dynamics to treat atomic and magnetic interactions on equal footing which already incorporated both transverse and longitudinal magnetic fluctuations.

Data availibility

The data that supports the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The spin-lattice dynamics SPILADY code is available under the Apache Licence version 2.0. One can download it from https://ccfe.ukaea.uk/resources/spilady/. Version 1.0.1 is suitable for single element simulation of molecular dynamics, spin dynamics, spin-lattice dynamics and spin-lattice-electron dynamics[46]. It is capable of considering the longitudinal fluctuations of magnetic moments. The modified SPILADY code developed for this work, which includes the MSLP potential, is available upon request from the corresponding author and will be included in the next version release. The potential parameters used in the generation of results presented in this work are available in the Supplementary Data (mslp.in). Further details are available in the Supplementary Material.

References

1. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B* **136**, 864 (1964).
2. Kohn, W. & Sham, L. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* **140**, 1133 (1965).
3. Goryaeva, A. M., Maillet, J.-B. & Marinica, M.-C. Towards better efficiency of interatomic linear machine learning potentials. *Comput. Mater. Sci.* **166**, 200 (2019).
4. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural net-work potentials. *J. Chem. Phys.* **134**, 074106 (2011).
5. Bartók, A., Payne, M. C., Kondor, K. & Csányi,. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
6. Bartók, A., Kondor, K. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
7. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
8. Cubuk, E. *et al.* Identifying structural flow defects in disordered solids using machine learning methods. *Phys. Rev. Lett.* **114**, 108001 (2015).
9. Cooper, A. M., Kästner, J., Urban, A. & Artrith, N. Efficient training of ann potentials by including atomic forces via taylor expansion and application to water and a transition-metal oxide. *npj Comput. Mater.* **6**, 54 (2020).
10. Eckhoff, M. & Behler, J. High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions. *NPJ Comput. Mater.* **7**, 170 (2021).
11. Antropov, V. P., Katsnelson, M. I., Harmon, B. N., Schilfgaarde, M. V. & Kusnezov, D. Spin dynamics in magnets: Equation of motion and finite temperature effects. *Phys. Rev. B* **54**, 1019 (1996).
12. Hellsvik, J. *et al.* General method for atomistic spin-lattice dynamics with first-principles accuracy. *Phys. Rev. B* **99**, 104302 (2019).
13. Ma, P. W. & Dudarev, S. L. *Atomistic Spin-Lattice Dynamics* 1017–1035 (Springer International Publishing, 2020).
14. Ma, P.-W., Dudarev, S. L. & Wrobel, J. S. Dynamic simulation of structural phase transitions in magnetic iron. *Phys. Rev. B* **96**, 094418 (2017).
15. Körmann, F., Dick, A., Grabowski, B., Hickel, T. & Neugebauer, J. Atomic forces at finite magnetic temperatures: Phonons in paramagnetic iron. *Phys. Rev. B* **85**, 125104 (2012).
16. Körmann, F. *et al.* Temperature dependent magnon-phonon coupling in bcc fe from theory and experiment. *Phys. Rev. Lett.* **113**, 165503 (2014).
17. Hasegawa, H. & Pettifor, D. G. Microscopic theory of the temperature-pressure phase diagram of iron. *Phys. Rev. Lett.* **50**, 130 (1983).
18. Lavrentiev, M. Y., Nguyen-Manh, D. & Dudarev, S. L. Magnetic cluster expansion model for bcc-fcc transitions in Fe and Fe–Cr alloys. *Phys. Rev. B* **81**, 184202 (2010).
19. Hasegawa, H., Finnis, M. W. & Pettifor, D. G. A calculation of elastic constants of ferromagnetic iron at finite temperatures. *J. Phys. F: Met. Phys.* **15**, 19 (1985).
20. Dever, D. J. Temperature dependence of the elastic constants in α-iron single crystals: Relationship to spin order and diffusion anomalies. *J. Appl. Phys.* **43**, 3293 (1972).
21. Nguyen-Manh, D., Horsfield, A. P. & Dudarev, S. L. Self-interstitial atom defects in bcc transition metals: Group-specific trends. *Phys. Rev. B* **73**, 020101(R) (2006).
22. Derlet, P. M., Nguyen-Manh, D. & Dudarev, S. L. Multiscale modelling of crowdion and vacancy defects in body-centred-cubic transition metals. *Phys. Rev. B* **76**, 054107 (2007).
23. Chapman, J. B. J., Ma, P. W. & Dudarev, S. L. Effect of non-heisenberg magnetic interactions on defects in ferromagnetic iron. *Phys. Rev. B* **102**, 224106 (2020).
24. Yesilleten, D., Nastar, M., Arias, T. A., Paxton, A. T. & Yip, S. Stabilizing role of itinerant ferromagnetism in intergranular cohesion in iron. *Phys. Rev. Lett.* **81**, 2998 (1998).
25. Ma, P.-W., Woo, C. H. & Dudarev, S. L. Large-scale simulation of the spin-lattice dynamics in ferromagnetic iron. *Phys. Rev. B* **78**, 024434 (2008).
26. Ma, P.-W. & Dudarev, S. L. Longitudinal magnetic fluctuations in langevin spin dynamics. *Phys. Rev. B* **86**, 054416 (2012).

27. Tranchida, J., Plimpton, S. J., Thilbaudeau, P. & Thompson, A. P. Massively parallel symplectic algorithm for coupled magnetic spin dynamics and molecular dynamics. *J. Comput. Phys.* **372**, 406 (2018).
28. Mudrick, M., Eisenbach, M., Perera, D., Stocks, G. M. & Landau, D. P. Combined molecular and spin dynamics simulation of bcc iron with lattice vacancies. *J. Phys: Conf. Ser.* **921**, 012007 (2017).
29. Evans, R. F. L. *et al.* Atomistic spin model simulations of magnetic nanomaterials. *J. Phys.: Condens. Matter* **26**, 103202 (2014).
30. Chapman, J. B. J., Ma, P.-W. & Dudarev, S. L. Dynamics of magnetism in FeCr alloys with Cr clustering. *Phys. Rev. B* **99**, 184413 (2019).
31. Malerba, L. *et al.* Multiscale modelling for fusion and fission materials: The m4f project. *Nucl. Mater. Energy* **29**, 101051 (2021).
32. Wen, H., Ma, P. W. & Woo, C. Spin-lattice dynamics study of vacancy formation and migration in ferromagnetic bcc iron. *J. Nucl. Mater.* **440**, 428 (2013).
33. Nikolov, S. *et al.* Data-driven magneto-elastic predictions with scalable classical spin-lattice dynamics. *npj Comput. Mater.* **7**, 153 (2021).
34. Domina, M., Cobelli, M. & Sanvito, S. Spectral neighbor representation for vector fields: Machine learning potentials including spin. *Phys. Rev. B* **105**, 214439 (2022).
35. Dragoni, D., Daff, T. D., Csányi, G. & Marzari, N. Achieving dft accuracy with a machine-learning interatomic potential: Thermo-mechanics and defects in bcc ferromagnetic iron. *Phys. Rev. Mater.* **2**, 013808 (2018).
36. Goryaeva, A. M. *et al.* Efficient and transferable machine learning potentials for the simulation of crystal defects in bcc fe and w. *Phys. Rev. Mater.* **5**, 103803 (2021).
37. Wang, Y. *et al.* Machine-learning interatomic potential for radiation damage effects in bcc-iron. *Comput. Mater. Sci.* **202**, 110960 (2022).
38. Novikov, I., Grabowski, B., Körmann, F. & Shapeev, A. Magnetic moment tensor potentials for collinear spin-polarized materials reproduce different magnetic states of bcc fe. *npj Comput. Mater.* **8**, 13 (2022).
39. Kvashnin, Y. O. *et al.* Microscopic origin of heisenberg and non-heisenberg exchange interactions in ferromagnetic bcc Fe. *Phys. Rev. Lett.* **116**, 217202 (2016).
40. Szilva, A. *et al.* Theory of noncollinear interactions beyond Heisenberg exchange: Applications to bcc Fe. *Phys. Rev. B* **96**, 144413 (2017).
41. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558(R) (1993).
42. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251 (1994).
43. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
44. Kresse, G. & Furthmüller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
45. Ozaki, T. *et al.* http://www.openmx-square.org/ (2003).
46. Ma, P.-W., Dudarev, S. L. & Woo, C. H. Spilady: A parallel cpu and gpu code for spinlattice magnetic molecular dynamics simulations. *Comput. Phys. Commun.* **207**, 350 (2016).
47. Ma, P.-W. & Dudarev, S. Langevin spin dynamics. *Phys. Rev. B* **83**, 134418 (2011).
48. Proville, L., Rodney, D. & Marinica, M.-C. Quantum effect on thermally activated glide of dislocations. *Nat. Mater.* **11**, 845 (2012).
49. Lavrentiev, M. Y. *et al.* Magnetic cluster expansion simulation and experimental study of high temperature magnetic properties of Fe-Cr alloys. *J. Phys.: Condens. Matter* **24**, 326001 (2012).
50. Woo, C. H., Wen, H., Semenov, A. A., Dudarev, S. L. & Ma, P. W. Quantum heat bath for spin-lattice dynamics. *Phys. Rev. B* **91**, 104306 (2015).
51. Olsson, P., Domain, C. & Wallenius, J. Ab initio study of Cr interactions with point defects in bcc Fe. *Phys. Rev. B* **75**, 014110 (2007).
52. Malerba, L. *et al.* Comparison of empirical interatomic potentials for iron applied to radiation damage studies. *J. Nucl. Mater.* **406**, 19–38 (2010).
53. Masters, B. C. Dislocation loops in irradiated iron. *Nature* **200**, 254 (1963).
54. Little, E. A. & Eyre, B. L. The geometry of dislocation loops generated in α-iron by 1 mev electron irradiation at 550°c. *J. Microsc.* **97**, 107 (1973).
55. Dudarev, S. L., Bullough, R. & Derlet, P. M. Effect of the $\alpha - \gamma$ phase transition on the stability of dislocation loops in bcc iron. *Phys. Rev. Lett.* **100**, 135503 (2008).
56. Dudarev, S. L., Derlet, P. M. & Bullough, R. The magnetic origin of anomalous high-temperature stability of dislocation loops in iron and iron-based alloys. *J. Nucl. Mater.* **386**, 45 (2009).
57. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning* (MIT Press, 2006).
58. Daw, M. S., Foiles, S. M. & Baskes, M. I. The embedded-atom method: A review of theory and applications. *Mater. Sci. Reports* **9**, 251 (1993).
59. Daw, M. S. & Baskes, M. I. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B* **29**, 6443 (1984).
60. Mendelev, M. I., Srolovitz, D. J., Ackland, G. J., Sun, D. Y. & Asta, M. Development of new interatomic potentials appropriate for crystalline and liquid iron. *Philosophical Mag.* **83**, 3977 (2003).
61. Ackland, G. J., Mendelev, M. I., Srolovitz, D., Han, S. & Barashev, A. V. Development of an interatomic potential for phosphorus impurities in α-iron. *J. Phys.: Condens. Matter* **16**, S2629 (2004).
62. Biersack, J. P. & Ziegler, J. F. Refined universal potentials in atomic collisions. *J. Nucl. Instrum. Methods* **143**, 93 (1982).
63. Perera, D. *et al.* Phonon-magnon interactions in body centered cubic iron: A combined molecular and spin dynamics study. *J. Appl. Phys.* **115**, 17D124 (2014).
64. Drautz, R. & Fähnle, M. Parametrization of the magnetic energy at the atomic level. *Phys. Rev. B* **72**, 212405 (2005).
65. Okatov, S. V., Gornostyrev, Y. N., Lichtenstein, A. I. & Katsnelson, M. I. Magnetoelastic coupling in γ-iron investigated within an ab initio spin spiral approach. *Phys. Rev. B* **84**, 214422 (2011).
66. Singer, R., Dietermann, F. & Fähnle, M. Spin interactions in bcc and fcc Fe beyond the Heisenberg model. *Phys. Rev. Lett.* **107**, 017204 (2011).
67. Singer, R., Dietermann, F. & Fähnle, M. Erratum: Spin interactions in bcc and fcc fe beyond the Heisenberg model. *Phys. Rev. Lett.* **107**, 119901(E) (2011).
68. Ruban, A. V., Khmelevskyi, S., Mohn, P. & Johansson, B. Temperature-induced longitudinal spin fluctuations in Fe and Ni. *Phys. Rev. B* **75**, 054402 (2007).
69. Wang, H., Ma, P.-W. & Woo, C. H. Exchange interaction for spin-lattice coupling in bcc iron. *Phys. Rev. B* **82**, 144304 (2010).
70. Lichtenstein, A. I., Katnelson, M. I. & Gubanov, V. A. Exchange interactions and spin-wave stiffness in ferromagnetic materials. *J. Phys. F: Met. Phys.* **14**, L125 (1984).
71. Rosengaard, N. M. & Johansson, B. Finite-temperature study of itinerant ferromagnetism in fe, co, and ni. *Phys. Rev. B* **55**, 14975 (1997).
72. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).

73. Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys.: Condens. Matter* **26**, 183001 (2014).
74. Wang, H., Zhang, L., Han, J. & Weinan, E. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **228**, 178 (2018).
75. Montavon, G., Orr, G. B. & Müller, K.-R. *Neural Networks: Tricks of the Trade* (Springer, 2012).
76. Goodfellow, I. Bengio, Y. & Courville, A. Deep Learning. http://www.deeplearningbook.org (publisherMIT Press, 2016).
77. Sonoda, S. & Murata, N. Neural network with unbounded activation functions is universal approximator. *Appl. Comput. Harmon. Anal.* **43**, 233 (2017).
78. Kondati Natarajan, S., Morawietz, T. & Behler, J. Representing the potential energy surface of protanated water clusters by high-dimensional neural network potentials. *Phys. Chem. Chem. Phys.* **17**, 8356 (2015).
79. Morawietz, T. & Behler, J. A density functional theory-based neural network potential for water clusters including van-der-waals corrections. *J. Phys. Chem. A* **117**, 7356 (2013).
80. Kobayashi, R., Giofré, D., Junge, T., Ceriotti, M. & Curtin, W. A. Neural network potentials for Al–Mg–Si alloys. *Phys. Rev. Mater.* **1**, 053604 (2017).
81. Sosso, G. C., Miceli, G., Caravati, S. & Behler, J. Neural network interatomic potential for the phase change material gete. *Phys. Rev. B* **85**, 174103 (2012).
82. Hhaliullin, R. Z., Eshet, H., Kühne, T., Behler, J. & Parrinello, M. Graphite-diamond phase coexistence study employing a neural network mapping of the ab initio potential energy surface. *Phys. Rev. B* **81**, 100103 (2010).

Acknowledgements

Author contributions

P.W.M. conceived the original M.S.L.D. method with contributions from J.B.J.C. J.B.J.C. and P.W.M. developed the M.S.L.D. training software. J.B.J.C. performed the potential fitting. J.B.J.C. implemented the M.S.L.P. into the SPILADY code originally written by P.W.M. Both J.B.J.C. and P.W.M. analysed the results, wrote and reviewed the manuscript.

Competing interest

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-25682-5.

Correspondence and requests for materials should be addressed to J.B.J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.