Review | Received 29 March 2026; Revised 19 August 2026; Accepted 11 September 2026; Published 28 September 2026
https://doi.org/10.55092/aimat20260012

Empowering material defect research with machine learning

Feifei Ren[1],†, Shijie Hu[1],†, Guanghui Zhang[1], Zenghua Cai[2] and Dan Han[1],*

[1] Key Laboratory of Automobile Materials of MOE, School of Materials Science and Engineering, Jilin University, Changchun, China

[2] Jiangsu Key Laboratory of Micro and Nano Heat Fluid Flow Technology and Energy Application, School of Physical Science and Technology, Suzhou University of Science and Technology, Suzhou, China

† These authors contributed equally to this work.

* Correspondence author; E-mail: hand@jlu.edu.cn.

Highlights:

- Machine-learning interatomic potentials (MLIPs) extend defect simulations across scales.
- Defect-specific data determine the reliability of machine-learning models.
- Active learning targets rare and uncertain defect configurations.
- ML-assisted atomistic models enable long-timescale kinetic simulations.
- Validation criteria should match the defect class and target property.

Abstract: Material defects govern the performance limits, degradation pathways, and service lifetimes of functional and structural materials, yet their predictive modeling remains difficult because local electronic reconstruction, long-range interactions, rare kinetic events, and finite-temperature effects are coupled across scales. First-principles calculations provide reliable descriptions of defect cores and charge states but are limited by system size, time scale, and sampling cost, whereas empirical molecular dynamics (MD) can access larger systems but often lacks transferability in reconstructed or chemically complex defect environments. This review examines machine learning (ML) for defect modeling when model choice, training data, and validation criteria are matched to the physical problem. We discuss machine-learning interatomic potentials (MLIPs), sparse Gaussian-process and committee-machine models, active-learning workflows, high-throughput defect screening, and ML-assisted multiscale simulations that connect atomistic calculations with longer-timescale modeling. Particular attention is given to the distinct requirements of point defects, dislocations, surfaces, grain boundaries, and interfaces, including configurational sampling, long-range interactions, charge response, and uncertainty control. Representative applications in semiconductors, halide perovskites, solid electrolytes, battery cathodes, irradiation damage, and catalytic nanoparticles show where ML expands the feasible scientific questions and where first-principles benchmarks or experimental validation remain essential. Reliable ML-assisted defect modeling ultimately depends on validating the physical quantities that control the relevant defect mechanism, rather than relying on aggregate model accuracy alone.

Keywords: material defects; machine-learning interatomic potentials; graph neural networks; active learning; multiscale modeling; high-throughput computing

1. Introduction

In modern materials science, defects are increasingly treated as controllable features of material structure rather than simply as imperfections. Their concentration, configuration, and distribution can be modified through compositional design, non-equilibrium processing, and external stimuli to modulate functional and structural responses [1,2]. Predictive defect engineering requires quantitative descriptions of defect formation thermodynamics, electronic-structure responses, migration pathways, and long-term evolution across relevant spatial and temporal scales [1,3,4]. These requirements become particularly demanding in structurally soft halide perovskites, chemically complex battery cathodes, solid electrolytes, and high-entropy alloys, where defect landscapes contain multiple metastable configurations and coupled electronic, ionic, and structural responses.

Predictive defect modeling is constrained by a trade-off between accuracy and the accessible spatial and temporal scales [5–7]. Local bond breaking, charge transfer, and electronic reconstruction within defect cores often require first-principles descriptions based on density functional theory (DFT) [1,8,9]. In contrast, phenomena involving long-range elastic or electrostatic fields, defect–defect interactions, dislocation motion, interfacial restructuring, and microstructural evolution often extend beyond the spatial and temporal scales routinely accessible to DFT [10–13]. In standard diagonalization-based implementations, the computational cost of DFT scales approximately as $O(N^3)$, limiting routine calculations to relatively small systems and short or static sampling windows [14–17]. Such calculations may fail to capture rare migration events, finite-temperature entropic effects, anharmonic lattice fluctuations, and metastable defect configurations that become relevant under operating conditions. Empirical MD can access larger systems and longer trajectories, but conventional interatomic potentials may lose accuracy when defect evolution involves bond rearrangement, charge redistribution, or chemically diverse local environments.

Automated defect workflows, including PyCDT, the Defect and Structure Prediction Package (DASP), doped, and AiiDA-based tools, have improved the standardization, reproducibility, and throughput of defect calculations [18–20]. Nevertheless, these workflows still depend on costly DFT calculations, which limits exhaustive exploration of large defect configurational spaces [21–23]. ML can reduce this bottleneck by extending the use of first-principles data to defect configurations and time scales that are impractical to sample directly with DFT. Machine-learning interatomic potentials (MLIPs) can sample defect relaxations, migration pathways, and finite-temperature dynamics at much lower cost than direct DFT [24–28]. Sparse Gaussian-process models provide data-efficient predictions with uncertainty estimates, while committee-machine strategies improve scalability in chemically heterogeneous environments. Equivariant graph neural networks (GNNs) can describe strongly distorted and chemically diverse local atomic environments. Active-learning workflows then determine when additional first-principles labels are required [29–32].

In this review, we focus on how ML models should be selected and validated for specific defect problems, while retaining first-principles and mechanistic descriptions where they are required. We organize the discussion around three practical considerations. We first compare the theoretical and computational requirements of point, line, and interfacial defects. We then analyze defect-oriented MLIPs, sparse Gaussian-process and committee-machine models, active-learning workflows, high-throughput defect informatics, and multiscale couplings from ML-driven MD to kinetic Monte Carlo (kMC), phase-field, and continuum models. Particular attention is given to cases in which ML enables defect configurations, time scales, or system sizes that are difficult to access directly with first-principles calculations, including finite-temperature point-defect thermodynamics, defect-mediated ion migration in solid electrolytes, dopant-defect interactions in battery cathodes, halide-perovskite defect and ion-migration problems, and alloy nanoparticle degradation. Throughout the review, we emphasize how the choice and validation of ML models depend on the underlying defect mechanism and the quantity of interest.

2. Development of defect theory and methods

2.1. Definition, classification, and task-specific modeling requirements

In an ideal crystal, atoms occupy periodically repeated lattice positions. Real materials, however, inevitably contain local departures from this periodic arrangement, generally referred to as defects [33,34]. Despite their localized nature, defects can have macroscopic consequences because they perturb the local electronic and structural environment and may alter transport or kinetic processes [35,36].

As shown in Figure 1a, crystal defects are commonly classified by spatial dimension into zero-dimensional (0D) point defects, one-dimensional (1D) line defects, and two-dimensional (2D) planar or interfacial defects. Point defects such as vacancies, interstitials, and antisites perturb the local bonding and electronic environment, thereby influencing defect stability and transport [37–39]. Dislocations, the most common line defects, are central to plastic deformation. Their mechanical effects depend on the structure and mobility of the dislocation core as well as its interactions with solutes and other defects [40]. Planar and interfacial defects, such as surfaces and grain or phase boundaries, create locally broken coordination and chemical environments that can promote segregation, reconstruction, and interfacial transport [41,42]. As illustrated in Figure 1b, defect engineering exploits these structural irregularities as controllable variables. The practical objective is therefore not only to identify defects, but also to predict how their population, configuration, and evolution alter material properties [43–47]. This places a central requirement on defect modeling: the computational method must resolve the physics controlling the target defect process at the relevant length and time scales [48].

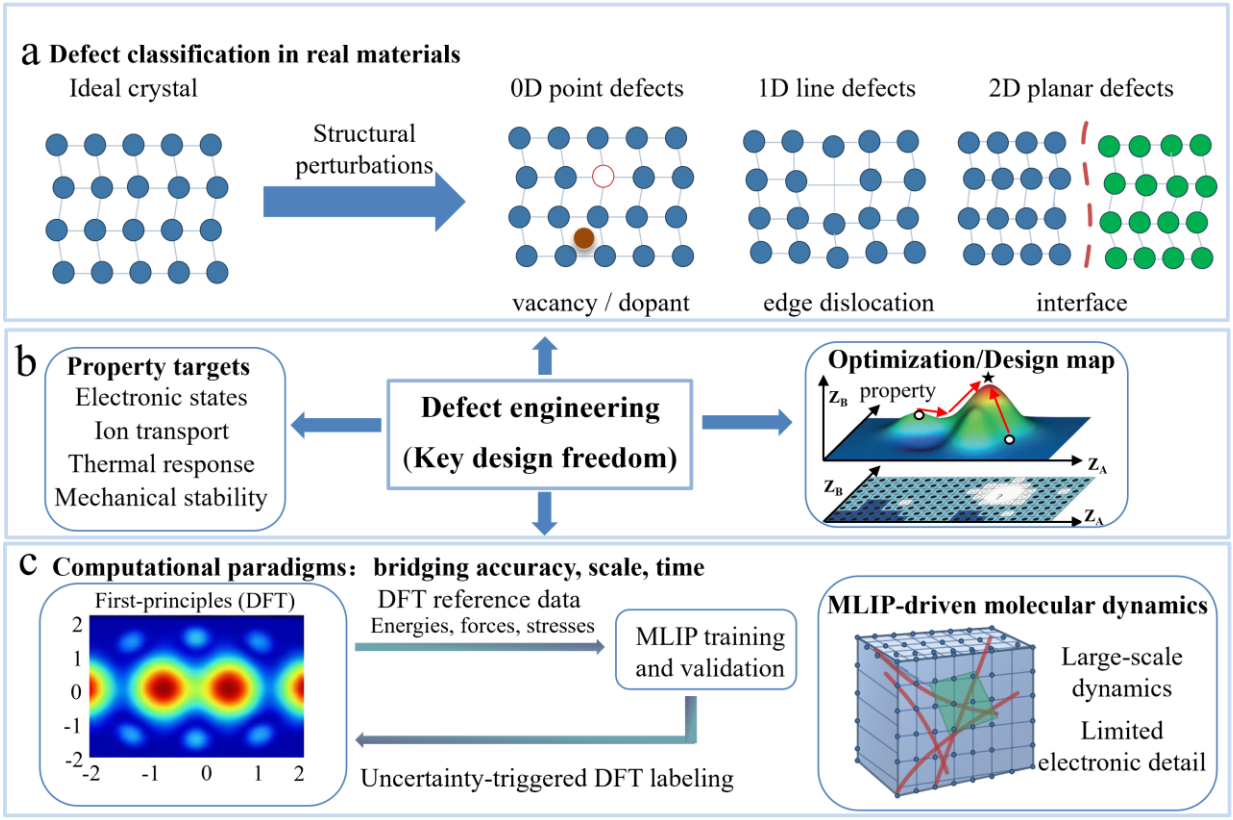

Figure 1. Defect classes, property targets, and ML-enabled modeling routes. **(a)** Representative point, line, and planar defects, with the line defect illustrated by a cross-sectional edge dislocation featuring an extra half-plane of atoms terminating within the crystal lattice; **(b)** Defect engineering for property tuning and materials design; **(c)** Uncertainty-aware MLIP workflow connecting DFT data with large-scale molecular dynamics.

The modeling requirements vary strongly with defect class. For point defects, the relevant requirement depends on the target observable: formation and charge-transition energies demand accurate local electronic structure, whereas diffusion problems additionally require reliable migration barriers and finite-temperature configurations [49–52]. For dislocations, the difficulty shifts from local chemistry alone to the simultaneous description of reconstructed cores and long-range elastic fields, often under applied stress. Interfaces pose a different challenge because broken symmetry permits reconstruction, segregation, and local changes in composition or charge, requiring configurational sampling beyond a single idealized structure [53–55]. Disordered defect networks further require finite-temperature sampling and event-level validation. Accordingly, the suitability of an ML strategy should be judged by whether it captures the physics and observables required for the target defect process, rather than by accessible simulation scale alone (Table 1).

Table 1. Defect classes, principal modeling challenges, and recommended ML routes [39,40,52,54,55].

Defect class	Representative systems	Main challenge	Recommended ML route
point defects	vacancies; interstitials; antisites	charge correction; local reconstruction; metastability	defect-specific MLIP training; active sampling; NEB screening
line defects	dislocations; disconnections; solute-dislocation complexes	long-range elasticity; core reconstruction; stress response	large-cell MLIP MD; accelerated dynamics; ML-informed DDD
planar and interfacial defects	surfaces; stacking faults; grain boundaries; heterointerfaces	broken symmetry; segregation; reconstruction; stoichiometry	interface-focused training; MLIP relaxation; structure search
disordered defect networks	amorphous phases; irradiation cascades; ionic and soft-lattice disorder	complex energy landscapes; bond rearrangement; finite-temperature transport	On-the-fly learning; sparse or committee models; MLIP MD–kMC coupling

A persistent difficulty in defect modeling is the mismatch between the spatial and temporal scales required to describe defect evolution and the cost of electronic-structure calculations. As shown in Figure 1c, first-principles calculations provide the electronic-structure accuracy needed to resolve local defect energetics, but their computational cost restricts the cell sizes and trajectory lengths accessible for defect−defect interactions and thermally activated evolution [1,56]. Empirical potentials extend atomistic simulations to much larger systems, yet their fixed functional forms may become unreliable when defect processes involve changes in local bonding, coordination, or chemical environment [54,57]. MLIPs offer an intermediate route by reproducing first-principles energy and force data at substantially lower computational cost, thereby enabling larger atomistic systems and longer trajectories within the configurational domain represented by the training data [5,7,15,16].

2.2. Limitations of traditional computational methods and software ecosystems

In defect research, first-principles methods remain the main benchmark for describing defect electronic structure, local defect levels, charge states, and thermodynamic stability [58]. However, the computational cost of conventional Kohn-Sham DFT scales approximately as $O(N^3)$ with system size in standard diagonalization-based implementations. The actual cost becomes particularly high for large supercells and calculations involving hybrid functionals or charged-defect corrections. In practice, defect calculations therefore require trade-offs among supercell size, charge-state sampling, k-point convergence, and exchange-correlation accuracy. Semilocal functionals such as the Local Density Approximation (LDA) and Generalized Gradient Approximation (GGA) further suffer from self-interaction errors and band-gap underestimation, which can distort charge localization and defect-level positions [59,60]. Hybrid functionals and many-body perturbation methods such as GW provide a more accurate description of defect electronic structures [61–63]. Their high computational cost, however, makes systematic sampling of charge states, migration pathways, and metastable configurations difficult, especially across diverse chemical environments.

Empirical-potential-based MD is commonly used for defect processes that involve collective atomic motion over relatively long timescales [64]. Typical examples include dislocation motion and interactions, the diffusion and aggregation of radiation-induced defects, and grain-boundary migration or reconstruction. Compared with direct first-principles simulations, MD can access substantially larger

systems and longer timescales, which makes it suitable for studying defect kinetics and microstructural evolution (Figure 2a) [65]. Its reliability, however, depends strongly on the functional form and parameterization of the interatomic potential. Empirical potentials often struggle when defect processes involve changes in chemical bonding or local electronic structure, including charge redistribution and changes in valence states [66]. Because these electronic effects can strongly influence defect-core structures, migration barriers, and relative stability, empirical potentials may lose accuracy outside the environments included in their parameterization [67,68]. This limitation becomes particularly important for chemically complex systems, interacting defects, and extreme conditions.

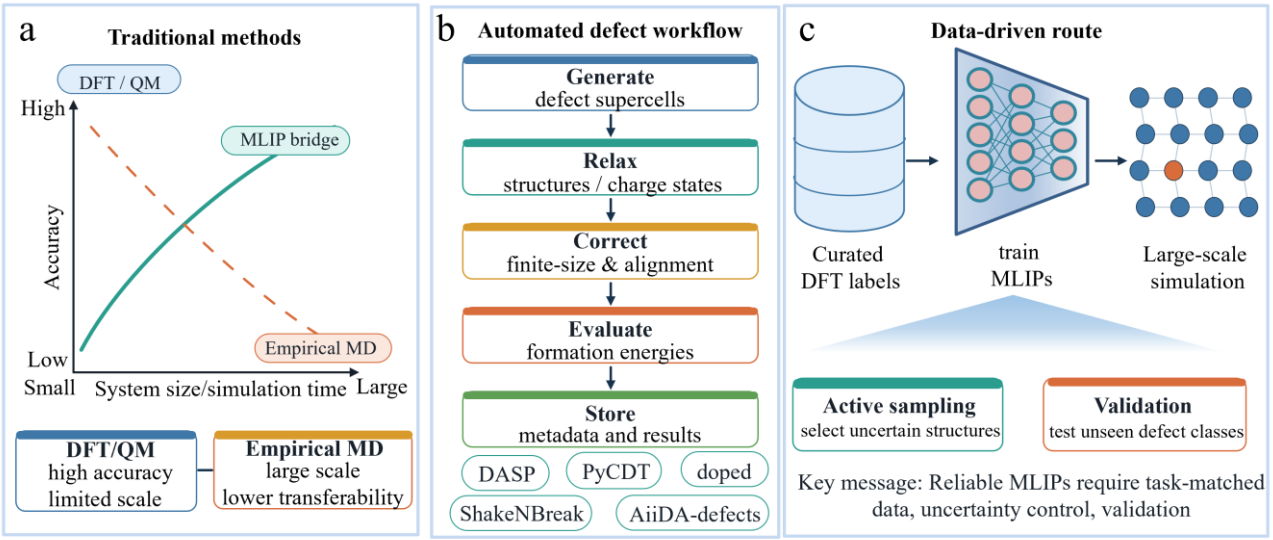

Figure 2. Computational paradigms and automated workflows for defect modeling. **(a)** Accuracy and accessible-scale trade-offs among first-principles calculations, empirical molecular dynamics, and MLIP-based simulations; **(b)** Automated defect-generation, relaxation, correction, analysis, and data-storage workflows; **(c)** Data-driven modeling based on curated DFT labels, MLIP training, large-scale atomistic simulation, active sampling, and validation.

Automated defect workflows have improved the reproducibility and throughput of first-principles defect calculations. As shown in Figure 2b, representative platforms include DASP, doped, AiiDA defects, PyCDT, pydefect, and ShakeNBreak, which support tasks such as defect generation, charge state enumeration, chemical potential analysis, finite-size correction, formation energy evaluation, and low-energy configuration search [69–72]. These tools reduce manual intervention and provide more systematic reference data for subsequent data-driven modeling [73]. Nevertheless, these workflows still require repeated first-principles calculations, and the underlying computational bottleneck therefore remains [74]. The problem becomes particularly severe when exploring large configurational spaces, interacting defects, reconstructed interfaces, or chemically complex systems. In addition, predefined defect-construction rules and manually designed search strategies can bias the explored configuration space, making it difficult to identify reconstructions or metastable structures that were not anticipated in advance [75]. MLIPs can extend first-principles accuracy to much broader configurational sampling, while uncertainty-aware workflows help identify configurations that require additional quantum-mechanical calculations (Figure 2c). Quantum-mechanical benchmarks remain essential, but MLIPs can substantially reduce how often such calculations are needed during structural relaxation, pathway

searches, and finite-temperature sampling [71,76–78]. These challenges have driven increasing interest in ML-based approaches that retain near-first-principles accuracy while extending the accessible configurational and spatiotemporal scales.

3. Defect-data construction and thermodynamic descriptors

3.1. Construction of defect structures: enumeration, perturbation and search

The initial set of defect structures defines which stable minima, metastable states, and migration-related configurations are accessible to subsequent thermodynamic and kinetic analyses [18,71]. Defects often induce local structural rearrangement, electronic redistribution, and long-range lattice distortion. Starting too close to the ideal lattice may prevent relaxation toward the physically relevant geometry, thereby influencing calculated formation energies, defect levels, and migration barriers [79,80]. In ML-enabled defect modeling, the sampled structures also define the configurational space represented in the training data, including defect cores, distorted coordination environments, migration-related regions, and reconstructed interfaces.

A common starting point is the enumeration of symmetry-inequivalent sites. Symmetry-based enumeration works well for simple, high-symmetry crystals, where vacancies, interstitials, antisites, and substitutions can be generated from well-defined crystallographic positions [1,72]. Ideal enumeration remains largely restricted to configurations implied by the parent lattice. In low symmetry, chemically complex, or structurally soft materials, direct relaxation from ideal starting coordinates can converge to only one local minimum and miss lower energy reconstructions [1,70,72]. Enumeration is therefore often combined with symmetry-breaking perturbations, local random displacements, or chemically informed distortions around the defect center [71,81]. Relaxation from these perturbed structures broadens the sampled region of the local potential-energy surface and can reveal alternative stable and metastable defect geometries [71,82].

When disorder, interfaces, multiple sublattices, or high defect concentrations are involved, the number of possible configurations increases rapidly and exhaustive enumeration becomes impractical [21,83]. Simulated annealing, Monte Carlo search, evolutionary algorithms, and adaptive optimization methods can explore broader configurational spaces and identify low-energy or representative metastable structures [14,84,85]. The reliability of these searches depends on whether the underlying energy model remains transferable in distorted defect environments. Low-cost enumeration and perturbation can first generate candidate structures, which are then relaxed and screened using MLIPs or other surrogate models. Selected low-energy configurations, together with structures that are poorly represented by the surrogate model, can subsequently be reevaluated using first-principles calculations.

As illustrated in Figure 3a, automated and ML-assisted approaches are increasingly being used to explore defect configurations beyond manually constructed ideal structures. On-the-fly sparse Gaussian process (SGP) optimization offers one way to perform this adaptive exploration [86]. During the search, the model can be updated as new configurations are encountered, while uncertainty estimates help identify poorly represented structures that require additional first-principles calculations. Such adaptive sampling is useful when relevant minima cannot be anticipated from the initial structure, as in reconstructed vacancies, interstitial clusters, ion-migration bottlenecks, surfaces, and grain-boundary defect complexes. ML-assisted search primarily improves sampling efficiency but does not remove the

need for independent validation. Low-energy structures separated by small energy differences, as well as migration-related saddle points, should still be confirmed by DFT or benchmarked against a separately validated high-accuracy MLIP.

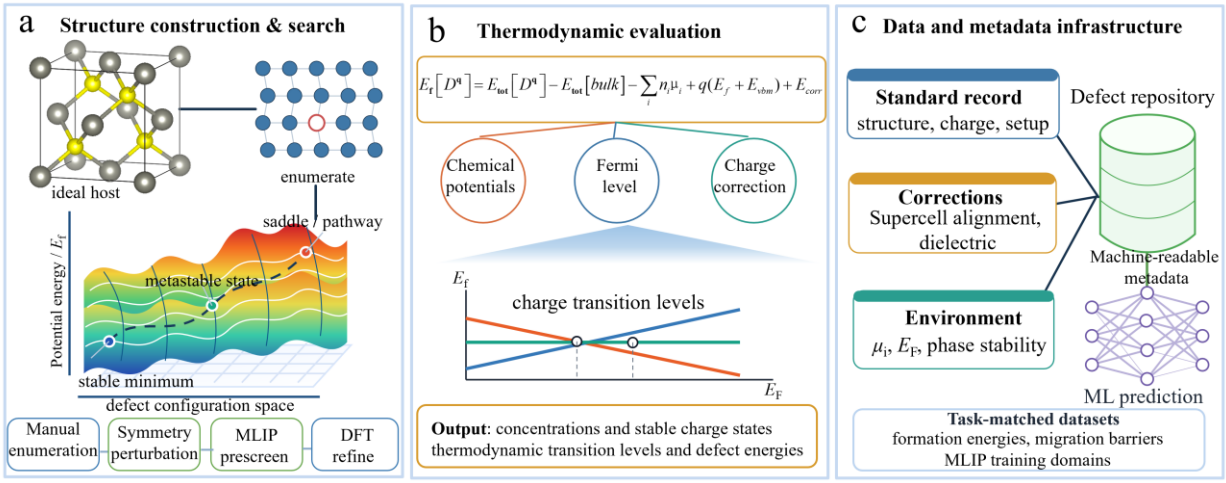

Figure 3. Theoretical frameworks and data infrastructure for defect science. **(a)** Evolution from manual enumeration to ML-assisted global search for stable and metastable defect structures; **(b)** Thermodynamic evaluation based on formation energies and transition levels, including necessary charge corrections; **(c)** Automated high-throughput workflows generating standardized defect data and metadata for data-driven modeling.

3.2. Defect formation energies and charge-transition levels

Defect formation energies determine the thermodynamic stability and equilibrium concentrations of defects. As shown in Figure 3b, the formation energy is obtained from the energy difference between the defect and perfect supercells, with the atomic chemical potentials constrained by the thermodynamic stability range of the host phase [58]. For charged defects, correction schemes such as the Freysoldt–Neugebauer–Van de Walle (FNV) method are generally required to mitigate spurious periodic image interactions, specifically the slowly decaying Coulomb potential (q/ϵ_r), thereby recovering the physical behavior of an isolated defect in the dilute limit [1,80,87,88]. Using $n_i > 0$ for atoms added to the supercell and $q > 0$ for electrons removed, the formation energy of a defect D in charge state q is written as:

$$E_f[D^q] = E_{\text{tot}}[D^q] - E_{\text{tot}}[bulk] - \sum_i n_i\,\mu_i + q(E_{\text{F}} + E_{vbm}) + E_{corr} \quad (1)$$

here, $E_{\text{tot}}[\text{D}^q]$ and $E_{\text{tot}}[\text{bulk}]$ are the total energies of the defective and pristine supercells, respectively. n_i denotes the number of atoms of species i added to or removed from the supercell, and μ_i is the corresponding atomic chemical potential. E_{F} denotes the Fermi-level position measured relative to the valence-band maximum, with $E_{\text{F}} = 0$ at the VBM, while E_{VBM} provides the corresponding band-edge reference in the total-energy expression. E_{corr} accounts for finite-size electrostatic and potential-alignment corrections. The chemical potentials are constrained by the stability of the host phase against elemental and competing phases. Band-edge and potential alignment must also be treated consistently across different charge states [87,88].

Plotting ($E_f[D^q]$) as a function of E_f yields thermodynamic charge-transition levels, corresponding to the Fermi-level positions at which two charge states have equal formation energies. These levels are closely related to electron or hole trapping and can be compared with experimental measurements such as photoluminescence and deep-level transient spectroscopy. High-throughput first-principles calculations and data-driven models now enable the screening of defect formation energies, charge-state stability, and defect levels across broad chemical and structural spaces [89–92]. Their predictive accuracy remains sensitive to the treatment of reference energies, chemical-potential ranges, charge corrections, and band-edge alignment.

The conventional formation-energy formalism is usually evaluated using relaxed 0 K structures and static total energies. Finite-temperature machine-learning force-field (MLFF) studies indicate that metastable defect geometries, orientational degeneracy, vibrational and spin entropies, and anharmonicity can contribute appreciably to defect free energies. For CdTe, MLFF-based thermodynamic integration showed that room-temperature metastable configurations can be populated and that thermal effects can change predicted interstitial concentrations by approximately two orders of magnitude [93]. The distinction between static formation energies and finite-temperature formation free energies is therefore important when comparing defect data, particularly for photovoltaic absorbers and ionic conductors.

3.3. Data, annotation, and metadata for defect machine learning

For defect-related ML, model performance depends not only on dataset size but also on whether the training data adequately represent the relevant defect configurations, charge states, and local chemical environments. In practice, many defect calculations are performed to address specific materials or physical mechanisms rather than to construct transferable ML datasets, resulting in substantial differences in supercell construction, charge-state treatment, convergence criteria, and post-processing protocols across DFT datasets [94]. These inconsistencies become more problematic when data are combined across different chemical-potential conditions, charge states, and supercell sizes. In addition, properties such as migration barriers and local density of states (LDOS) require configurations that may not be adequately sampled in datasets originally generated for defect formation-energy calculations.

High-throughput defect workflows can reduce these inconsistencies by applying common generation, relaxation, and analysis procedures across large sets of defect configurations, as illustrated in Figure 3c. Automating defect generation, structural relaxation, phase-stability analysis, charge-state screening and energy evaluation allows these calculations to be performed under a more consistent computational protocol, reducing operator-dependent variation across different chemical-potential conditions and Fermi-level ranges. Platforms such as PyCDT and pymatgen-based workflows, together with Materials Project-related infrastructure, have helped standardized key steps ranging from defect generation and supercell construction to formation-energy analysis [95]. For ML applications, such protocol consistency is particularly important because systematic differences arising from supercell size, charge corrections, or convergence settings may otherwise be learned as apparent defect-property trends rather than recognized as computational artifacts.

Atomic coordinates and total energies alone are insufficient to interpret many defect properties, making the associated computational metadata an integral part of the dataset. Key metadata should include supercell size, charge state, chemical potentials, dielectric constant, band-edge reference and charge-correction scheme. The DFT settings and relaxation procedure should also be documented. For MLIP datasets, each structure should be labeled according to the environment it represents, such as a

bulk region, defect core, saddle region, reconstructed interface, or finite-temperature configuration. High-uncertainty structures should be marked separately. Machine-readable metadata make the dataset easier to reproduce and reuse. They also make it clearer which configurations are represented in the training data and where the model may be less reliable [96].

The key challenge is therefore to construct defect datasets that match the target task, rather than simply increasing the amount of available data. Formation-energy prediction requires consistent chemical potentials and charge corrections, whereas migration-barrier prediction requires transition-state and pathway configurations, and MLIP training requires forces and stresses from defect cores, interfaces, and finite-temperature trajectories. This task-dependent view connects the data infrastructure in Figure 3c with the defect-oriented data design and sampling strategies discussed in Section 4.4.

4. Machine-learning interatomic potentials for defect environments

MLIPs learn potential energy surfaces (PESs) from quantum-mechanical reference data and enable atomistic simulations at length and time scales beyond routine first-principles calculations [97]. Unlike traditional empirical potentials, MLIPs do not assume simple analytical function forms. Instead, they learn the relationship between atomic environments and energies, forces, and stresses from training data, providing a flexible description of many-body interactions. Defect environments often involve bonding, coordination, strain, and chemical states that differ substantially from those in pristine crystals [84,98]. At the same time, the reliability of an MLIP is determined by the chemical and configurational domain represented in the training set, making validation and uncertainty assessment essential for defect applications.

4.1. What ML changes in defect modeling

Figure 4 groups MLIP applications in defect studies into four areas: reconstruction and energetics, migration and transport, collective evolution, and degradation and lifetime. These simulations can sample more configurations and longer trajectories than direct first-principles calculations, provided that the relevant chemical and structural environments are represented in the training data and tested independently [21,99].

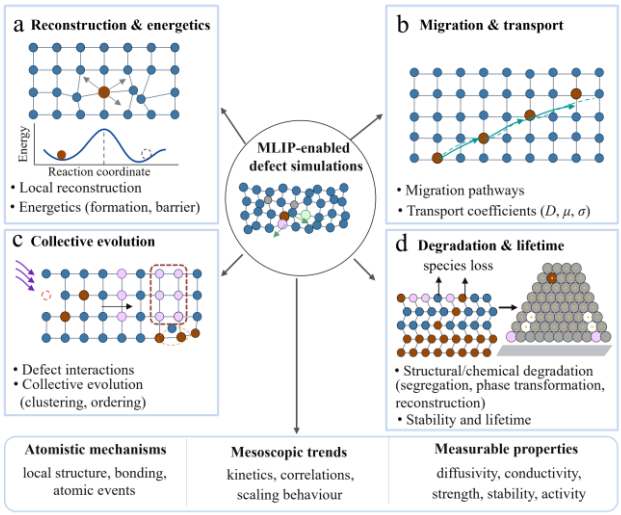

Figure 4. Task-oriented applications of MLIPs in defect simulations. **(a)** Local reconstruction and defect energetics; **(b)** Migration pathways and transport; **(c)** Defect interactions and collective evolution; **(d)** Structural and chemical degradation. The lower panel connects atomistic mechanisms with mesoscale behavior and observable material properties.

For point defects, MLIPs can accelerate structural relaxation, metastable-state searches, and migration-path sampling, with selected structures subsequently refined by DFT. For dislocations and other extended defects, MLIP-driven MD enables larger simulations of glide, climb, cross-slip, and interactions with solutes, vacancies, interstitials, and precipitates. At grain boundaries and heterointerfaces, relevant applications include reconstruction, segregation, and defect-mediated transport. MLIPs are also useful for irradiation and degradation simulations. They can track defect generation, recombination, clustering, and bond rearrangement over trajectories that are inaccessible to direct first-principles calculations [93,99,100]. Migration studies require adequate coverage of saddle-point configurations. Dislocation simulations instead need highly strained and reconstructed environments. Ionic and polar materials may also require explicit treatment of charge response and long-range electrostatics.

4.2. Model families and geometric deep learning

As illustrated in Figure 5, MLIP development has gradually shifted from hand-crafted descriptors toward geometric deep-learning models. Geometric models can represent distorted and low-symmetry environments more directly than fixed invariant descriptors. In the traditional modeling paradigm (Figure 5a), models relied primarily on manually constructed local descriptors, such as Smooth Overlap of Atomic Positions (SOAP) or atom-centered symmetry functions (ACSFs) to characterize atomic environments [101]. These descriptors transform local atomic coordinates into fixed-dimensional feature vectors. Rotational, translational, and permutational invariance is introduced through predefined mathematical forms. The resulting descriptors are then used as inputs to regression or neural-network models, such as GAP or BPNNs [7,15]. These descriptors provide translational, rotational, and permutational invariance at relatively low computational cost. Their accuracy depends on descriptor resolution, cutoff radius, basis completeness, and training-set coverage. These factors become particularly important for defect cores, saddle-point structures, and reconstructed interfaces.

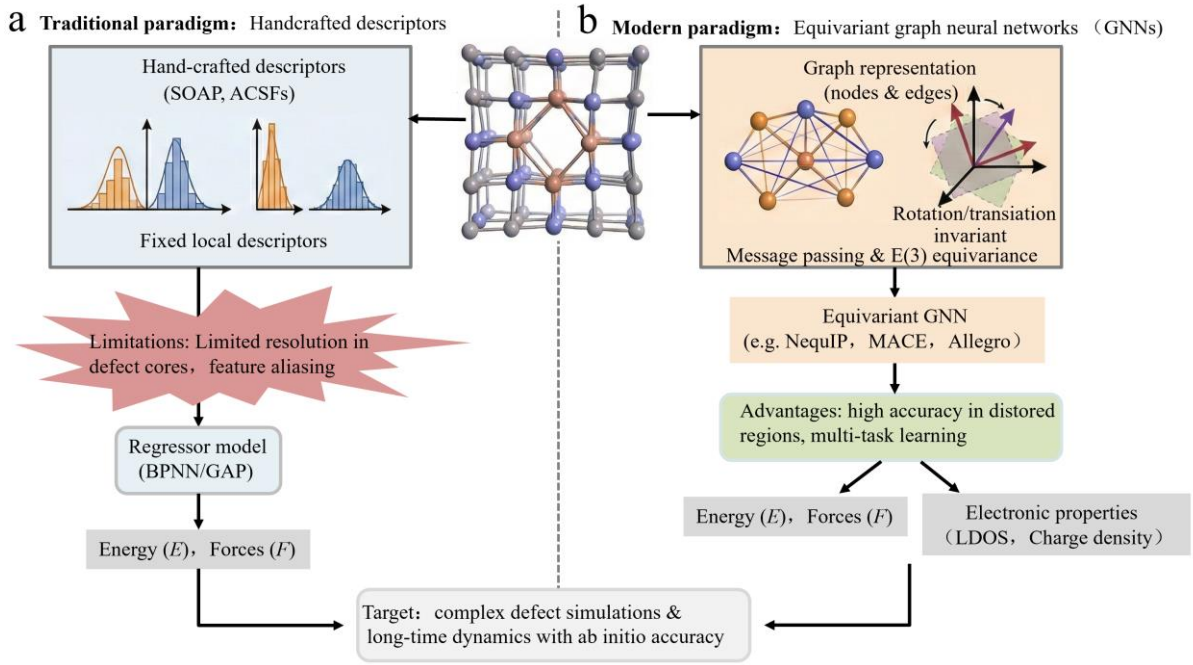

Figure 5. Evolution of structural representations used in MLIPs from **(a)** traditional hand-crafted descriptors to **(b)** modern geometric deep-learning paradigm.

In the geometric deep-learning paradigm shown in Figure 5b, atomic structures are represented as graphs and updated through symmetry-preserving message passing. Equivariant graph neural networks, such as NequIP, MACE, and Allegro, preserve the transformation behavior of scalar, vector, and higher-order tensor features under three-dimensional symmetry operations. This improves the description of distorted and low-symmetry defect environments, where local structural perturbations can strongly influence energies and forces. In practice, the selection of an equivariant MLIP should depend on the target defect problem, the required accuracy, the available training data, and the computational scale of the simulation.

Overall, model selection in defect research should balance accuracy, transferability, uncertainty control, and inference cost. The balance between accuracy and efficiency depends on the target problem. Point-defect reconstruction and migration barriers require high accuracy in distorted environments. Dislocation motion, irradiation cascades, and long-time diffusion place greater emphasis on inference efficiency. Chemically heterogeneous interfaces require broad training coverage, whereas charged and polar defects may require explicit treatment of long-range interactions and charge response. Table 2 summarizes these considerations for representative MLIP families.

Table 2. Qualitative comparison of representative MLIP families for defect modeling, including their main strengths, representative defect tasks, training and inference costs, and principal limitations or validation requirements [7,23–25,28].

MLIP family	Main strengths	Representative defect tasks	Qualitative cost profile[a]	Main cautions
GAP/sparse GP	data efficiency; native uncertainty	defect cores; saddle points; ion migration	medium/ low-medium	kernel scaling; uncertainty calibration
ACE/ MTP/ NEP	efficient many-body inference	dislocations; irradiation cascades; thermal transport	medium/low	basis/cutoff sensitivity; sparse-defect coverage
Deep Potential	scalable MD; mature active-learning	diffusion; interfaces; large-scale defect dynamics	high/low	long-range electrostatics; charge response
NequIP	data-efficient equivariant accuracy	point-defect reconstruction; migration barriers	medium/high	message-passing cost; large-cell scaling
MACE	higher-order equivariance; accuracy-cost balance	defect cores; migration pathways; finite-temperature dynamics	medium/medium	body-order choice; charged-defect validation
Allegro	local equivariant; parallel scalability	dislocation motion; irradiation cascades; long trajectories	high/low[b]	locality assumptions; long-range effects
CHGNet/M3GNet	broad pretrained chemical coverage	initial screening; relaxation; transfer learning	low–medium[c]/medium	domain shift; DFT validation
Sparse/active BCM	local experts; scalable uncertainty	interfaces; amorphous/ heterogeneous phases	medium/medium	expert partitioning; virial validation

[a] Cost levels are qualitative and refer to defect-oriented training and inference under comparable dataset sizes and computational conditions.

[b] The low inference cost of Allegro assumes an optimized parallel implementation.

[c] For pretrained models, the training cost refers to defect-specific fine-tuning and excludes the original pretraining cost.

No single MLIP family is optimal for all defect problems. Model choice should therefore follow the target mechanism and simulation scale. Atomic cluster expansion (ACE), moment tensor potential (MTP), and neuroevolution potential (NEP) provide efficient many-body representations and are suitable for large-scale dynamics. Deep Potential models combine scalable neural-network potentials with established active-learning workflows. Sparse Gaussian-process and committee-machine approaches are attractive when data efficiency and uncertainty estimation are important. Pretrained models such as CHGNet and M3GNet can also be used for initial screening or defect-specific transfer learning. Quantitative defect formation energies, migration barriers, charged-defect configurations, and highly distorted structures still require targeted first-principles validation or defect-specific fine-tuning. Some ML models also predict electronic or magnetic quantities, including LDOS and charge-related descriptors. These predictions should be validated against explicit electronic-structure calculations and should not be treated as equivalent to standard MLIP energy and force outputs [102,103].

4.3. Long-range interactions, charge response, and hybrid frameworks

The charged-defect formation energies and correction terms discussed in Section 3.2 are closely related to the treatment of long-range electrostatics in MLIP construction. Defects can generate slowly decaying electrostatic or elastic fields that extend well beyond the region of local structural distortion. Long-range electrostatic effects are particularly important in insulators and semiconductors. As shown in Figure 6a, charged point defects generate long-range electrostatic and polarization responses, whereas dislocations and crack tips produce elastic fields that extend beyond the locally distorted core region. Their magnitude and spatial decay depend on defect geometry, dielectric or elastic properties, and boundary conditions. Most conventional MLIPs describe atomic environments using finite cutoff radii. Without an explicit nonlocal component, such local representations may miss contributions arising outside the cutoff region, including long-range electrostatic, polarization, and elastic effects [97].

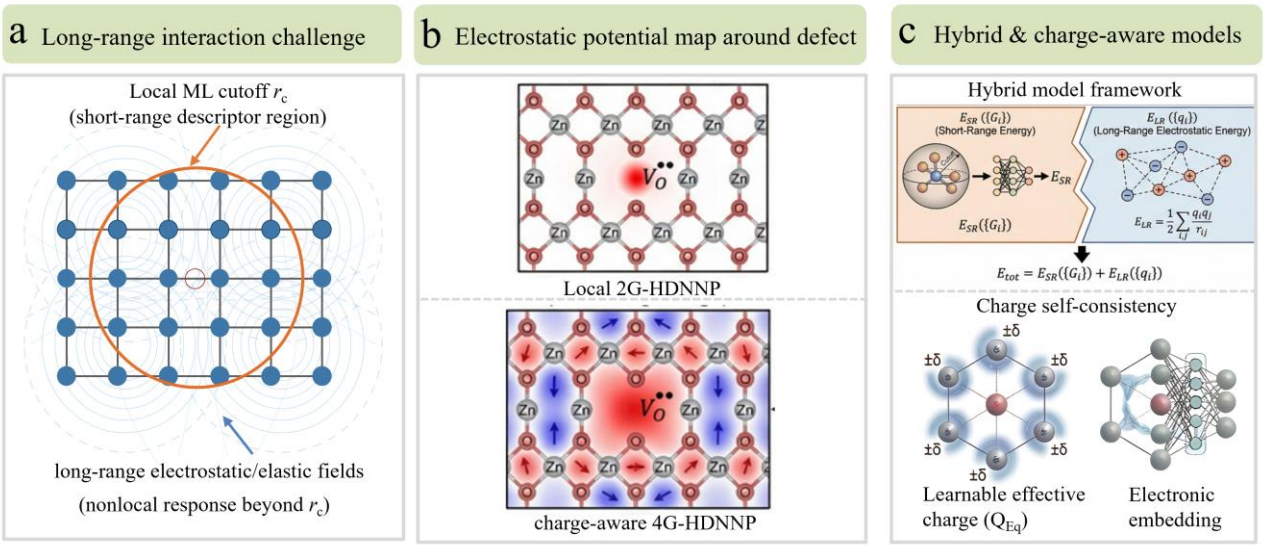

Figure 6. Treatment of long-range interactions in defect MLIPs. **(a)** Limitations of finite-cutoff descriptors; **(b)** Electrostatic potentials predicted by 2G- and 4G-HDNNPs around a charged oxygen vacancy; **(c)** Hybrid decomposition into short-range ML and explicit long-range contributions.

Figure 6b compares second- and fourth-generation high-dimensional neural network potentials for a charged oxygen vacancy. The local 2G-HDNNP cannot reproduce the variation of the electrostatic potential beyond its cutoff radius. In the 4G-HDNNP, environment-dependent charges are obtained through global charge equilibration. This treatment recovers the long-range potential gradient and the associated lattice-polarization response [104].

FNV enters at a different stage of the calculation. It is applied during the analysis of periodic charged-defect DFT calculations. It reduces the artificial interaction between repeated charged defects and aligns the electrostatic potential used to calculate defect formation energies [87,88]. It does not update charge redistribution or polarization as the atoms move along an MD trajectory. Charge-equilibration and polarizable MLIPs perform this update for each atomic configuration, and the resulting electrostatic contribution enters both the energy and the forces [104,105]. FNV-corrected formation energies can therefore serve as reference or validation targets, while dynamical training still requires a consistent treatment of periodic energies, forces, and long-range contributions.

Figure 6c shows one way to construct such a dynamical model. A short-range ML term describes bonding, repulsion, and local structural rearrangement, while explicit electrostatic, polarization, or elastic terms account for the nonlocal response. Depending on the system, these contributions may be evaluated using Ewald summation, multipole models, charge equilibration, or Green's-function-based elastic corrections [102,105]. The energy decomposition used during training must be consistent with the reference calculation. Otherwise, periodic-image artifacts may remain in the training targets, or the long-range electrostatic contribution may be counted twice. Static defect energetics and configuration-dependent polarization should therefore be validated separately [102,105,106].

The choice of electrostatic treatment depends on the material and defect process. Fixed-charge or strictly local models are often adequate for metals and covalent materials when screening is strong and charge transfer is limited. They may become insufficient in ionic insulators, polar oxides, halide perovskites, and charged-defect systems when electrostatic or polarization effects extend beyond the local cutoff. Variable-charge, charge-equilibration, polarizable, or hybrid short-range/long-range models are better suited to systems in which redox chemistry, dielectric screening, polarons, or charged vacancies strongly affect the defect energetics. Vacancy and interstitial migration in metals mainly requires accurate local forces and sufficient sampling of distorted saddle-point configurations. Migration of charged defects in semiconductors, ionic insulators, polar oxides, and halide perovskites additionally requires consistent treatment of dielectric screening, electrostatics, charge states, and possible polaronic distortions.

4.4. Defect-oriented data design, sparse sampling, and fidelity transfer

Compared with bulk crystals, defect systems require more selective training-data construction because the relevant configurations are sparse and many target properties depend on small relative energy differences. Defect cores and the associated distorted environments occupy only a small fraction of the total configurational space. If training data are dominated by bulk-like structures or near-harmonic bulk vibrations, the resulting model may perform well on average while remaining inaccurate in the defect regions that matter most [30,78,107]. Effective dataset construction therefore requires targeted sampling of vacancies, interstitials, defect clusters, dislocation cores, and grain-boundary reconstructions, together with adequate coverage of transition states when diffusion or reactions are considered.

Training-data imbalance can also be handled at the loss-function level. Gaussian density function (GDF) weighting provides one approach in which sparsity in feature space is used to assign larger effective weights to underrepresented configurations during neural-network-potential training [108]. For defect MLIPs, this idea is directly relevant because defect cores and transition states may be present in the training set but still be statistically overwhelmed by bulk-like vibrations. Density-aware weighting can therefore reduce the risk that rare defect and transition-state configurations are masked by the much larger number of bulk-like environments.

Sparse Gaussian-process regression (SGPR) is well suited to data-efficient defect potentials because it combines flexible regression with predictive uncertainty estimates. In defect science, physically important configurations are rarely distributed uniformly in configurational space. Defect cores, migration saddle points, and reconstructed interfaces are typically sparse but property-controlling regions. The model is trained using a compact set of representative atomic environments. During active sampling, configurations with high predicted uncertainty are selected for additional first-principles labeling [109]. Compared with a single global neural-network potential, sparse GP models are particularly useful when the reference dataset is small, when active sampling is central to the workflow, or when uncertainty estimates are needed to guide further data acquisition.

Sparse Bayesian committee-machine (BCM) strategies partition a chemically diverse configurational space among local expert models and combine their predictions. This architecture is suitable for trajectories containing distinct environments, including bulk regions, defect cores, surfaces, grain boundaries, amorphous structures, and high-pressure configurations. It limits the growth of individual kernel models, while disagreement among experts provides a practical uncertainty measure [110]. For finite-temperature and NPT simulations, stresses, pressures, and virials must be validated together with energies and forces. Active sparse BCM potentials support on-the-fly learning under constant-pressure conditions and have been demonstrated for $Li_{10}Ge(PS_6)_2$ and other condensed-phase systems [111].

Multi-fidelity learning can reduce the amount of expensive high-level reference data required for defect MLIPs. Large-scale structural and force data can be generated using lower-cost electronic-structure methods, while selected high-fidelity calculations can be used to calibrate key defect energetics and electronic quantities [32,112]. For example, hybrid-functional calculations can correct semilocal density-functional errors in band gaps, charge localization, and defect formation energies. Many-body perturbation-theory corrections, such as GW-level information, are more naturally used to refine band-edge alignment, defect levels, or charge-transition levels rather than to provide routine large-scale force labels for PES training. Residual-learning or delta-learning strategies can then learn the correction between low- and high-fidelity targets, which is especially useful when semilocal DFT errors would otherwise propagate into charged-defect energetics or electronic transition levels.

Loss design and physical regularization are also important when defect configurations are sparsely represented in the training set. Dynamic weighting can balance energy, force, stress, and defect-specific configurations during training, while physical constraints such as energy-force consistency, stress-tensor symmetry, elastic constants, and phonon stability can suppress nonphysical extrapolation in data-sparse regions [113,114]. Specific active-learning decision criteria, such as uncertainty metrics and relabeling triggers, are discussed separately in Section 5.3.

4.5. Representative defect applications across material classes

Machine-learning interatomic potentials have been used to connect atomistic defect processes with transport and degradation behavior. In the solid electrolyte $Li_7P_3S_{11}$, a sparse Gaussian-process potential achieved a force RMSE of 0.14 eV $Å^{-1}$ with $R^2 = 0.944$ on an independent test set. For the 672-atom supercell considered in that study, sparse-GP-driven molecular dynamics was reported to be approximately 10^4 times faster than the corresponding ab initio molecular dynamics simulation. The calculated room-temperature (RT) ionic conductivity of the α phase was 3.5×10^{-2} S cm^{-1}, compared with previously reported values of 4.5×10^{-2} S cm^{-1} and 5.7×10^{-2} S cm^{-1}. The RT conductivity of β-phase is 2.3×10^{-3} S cm^{-1}. The calculated Li-diffusion activation energies were 0.20 and 0.30 eV for the α and β phases, respectively [115]. Transferable ML potentials have also been explored across related Li-containing solid electrolytes [116]. In lithium-ion cathodes, combined ML and DFT calculations have been used to reveal how Al doping suppresses capacity and voltage fading by regulating dopant–defect interactions, oxygen-related distortions, transition-metal migration, and Li diffusion [117].

Metal halide perovskites present a more challenging defect problem because ion migration is coupled to defect charge states, strain, local polarization, and phase stability [118]. ML potentials extend ab initio molecular dynamics to larger cells and longer trajectories, enabling the investigation of phase stability, ion migration, surface and grain-boundary defects, and interfacial degradation. For $CsPbI_3$, ML force-field migration barriers generally agreed with DFT climbing-image nudged elastic band calculations to within approximately 0.1 eV, although the deviation reached 0.22 eV for the neutral iodide vacancy. Independent validation yielded $R^2 > 0.94$ and a force MAE below 54.83 meV $Å^{-1}$. At least five independent 2 ns MLMD trajectories were subsequently performed at each temperature, yielding charge-state-dependent diffusion activation energies of 0.16–0.30 eV [119]. The larger barrier error for the neutral iodide vacancy, despite the low overall force MAE, shows that aggregate force accuracy does not guarantee uniform barrier accuracy across defect charge states and migration pathways. Defect-specific migration barriers should therefore be validated independently against first-principles calculations.

MLIPs also extend defect studies beyond static 0 K energetics. MLIP-based free-energy calculations can incorporate vibrational, anharmonic, and metastable-configuration contributions to finite-temperature defect thermodynamics [93]. In metals, alloys and catalytic nanoparticles, MLIP-derived energetics can be coupled with kinetic Monte Carlo or related mesoscale methods to connect atomic migration and segregation events with long-time structural evolution. For example, MLIP-assisted off-lattice kMC has been used to investigate electrochemical degradation in Pt_3Co nanoparticles [120].

Across these material classes, MLIP reliability depends on whether the model preserves the defect mechanism governing the target property. Validation should therefore be hierarchical: energy and force errors assess local potential-energy-surface fidelity, defect-specific barrier errors determine the reliability of rare-event kinetics, and observables such as diffusion coefficients, ionic conductivity, voltage stability, and degradation times evaluate the resulting multiscale predictions. A single aggregate error is insufficient to establish the reliability of an ML-enabled defect workflow.

5. Defect configuration search, active learning, and data-driven screening

5.1. Defect configuration space and MLIP-assisted global search

The defect configuration space becomes extremely large once multiple sublattices, disorder, or interfaces are involved. In multicomponent alloys, solid solutions, and interfacial structures, a defect may occupy several inequivalent sites and relax into multiple reconstructed geometries or clusters. Manual construction of candidate structures can capture obvious motifs, but it rarely samples the full low-energy landscape. For this reason, global optimization methods such as simulated annealing, basin hopping, and evolutionary algorithms are increasingly used to explore defect configurations under stoichiometric and geometric constraints [14,84,121].

For defect structure search, the configurational problem can be divided into two coupled levels. The first level is site combination, which determines the lattice sites occupied by vacancies, interstitials, substitutional impurities, or defect complexes. The second level is local reconstruction, which determines how atoms relax around a given occupancy pattern and whether new low-symmetry or metastable configurations appear. As shown in Figure 7a, MLIP-assisted search can use an outer loop to enumerate site occupancies and an inner loop to perturb, relax, and rank the resulting structures. MLIPs can evaluate large numbers of candidate structures at much lower cost than direct first-principles relaxation. Representative low-energy structures can then be refined by DFT [10,122,123].

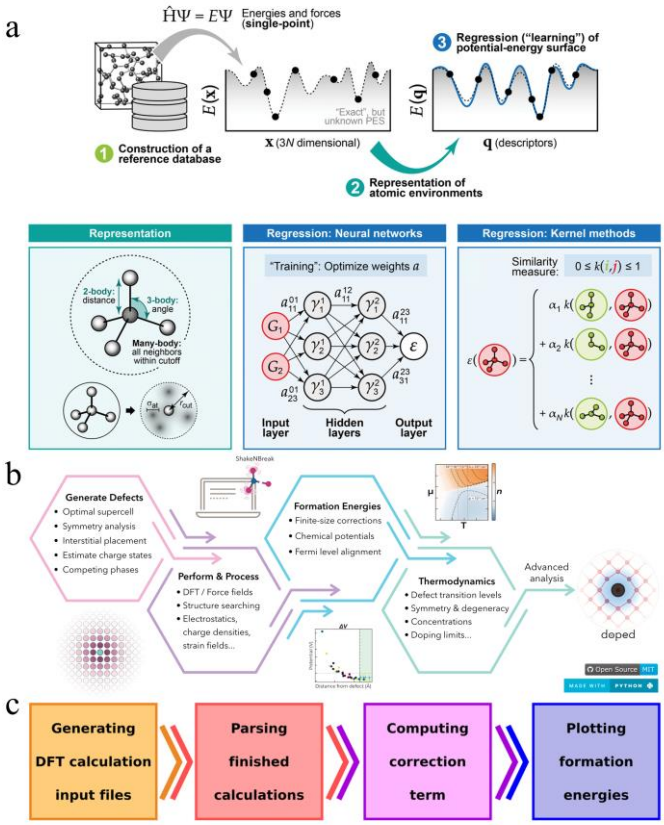

Figure 7. Representative workflows for defect structure search and high-throughput screening. **(a)** MLIP-assisted exploration of defect PESs. Reprinted with permission [16]. Copyright 2019 Wiley-VCH GmbH; **(b)** Workflow of the doped package for defect generation and analysis. Reproduced with permission [70]. Licensed under CC BY 4.0; **(c)** Workflow of PyCDT for high-throughput defect calculations. Reproduced with permission [18]. Licensed under CC BY 4.0.

On-the-fly SGP potentials can accelerate defect-structure optimization [86]. During exploration, the potential is updated when new local environments are encountered. This is useful for defect searches because vacancies, interstitials, dopants, surfaces, and interfaces may relax into low-symmetry structures that are poorly represented in the initial training set. For more complex defects, including grain boundaries, phase boundaries, and amorphous environments, global optimization usually needs additional physical constraints to avoid non-physical structures. Typical constraints include stoichiometry preservation, bounds on coordination environments, and soft penalties that suppress unrealistic collapse during relaxation [124]. General structure-prediction methods such as USPEX and CALYPSO have also been applied to defect problems, with MLIPs used for rapid prescreening and DFT for final reoptimization in some studies [125,126]. Such workflows broaden the range of defect configurations that can be explored, but low-energy structures relevant to defect thermodynamics or kinetics still require first-principles validation.

5.2. Automated defect calculation and ML-enhanced screening workflows

High-throughput defect screening relies on standardized workflows that generate comparable defect structures, apply consistent thermodynamic corrections, and retain the metadata required for validation. A typical defect workflow includes supercell construction, symmetry analysis, defect generation, structural relaxation, charge-state enumeration, formation-energy evaluation, finite-size correction, band-edge alignment, and evaluation of charge-transition levels. Software including PyCDT, pymatgen, AiiDA, atomate, and doped support different parts of this workflow, ranging from defect generation and correction schemes to workflow orchestration and data management [18,95,127,128]. As illustrated in Figure 7b,c, workflow automation improves both computational throughput and reproducibility. The latter depends on retaining calculation metadata, including supercell size, dielectric tensor, band-gap correction scheme, chemical-potential condition, charge-correction method, and band-edge alignment strategy.

ML can reduce the dependence of high-throughput defect screening on direct DFT calculations in three main ways. MLIPs can first be used to accelerate structure relaxation and migration-path searches, allowing a larger number of defect configurations to be explored at substantially lower cost [129]. Such acceleration is particularly useful for sampling defect sites, defect complexes, competing relaxed configurations, and transition states, where the number of candidate structures can become prohibitively large. By enabling broader configurational sampling, MLIPs can also increase the likelihood of identifying low-energy structures before selected configurations are refined with first-principles calculations.

A complementary strategy is to use supervised models based on GNNs or handcrafted descriptors as surrogates for defect energetics. Depending on the available representations and training data, these models can predict defect formation energies and migration barriers and, in some cases, estimate charge-transition levels from compositional and structural information. Once trained, these models can rank large libraries of candidate materials or defect configurations before DFT refinement [22,130]. Their applicability is more limited for charged defects, strongly reconstructed local environments, and systems that are poorly represented in the training data. DFT calculations are therefore still required for the subset of candidates that determines the final defect energetics or physical interpretation.

More adaptive workflows couple ML predictions with uncertainty estimates to determine the required level of calculation. Depending on the predicted uncertainty, a configuration may be accepted

at the surrogate-model level, further relaxed with an MLIP, or selected for DFT calculation and subsequent model refinement [131,132]. Such uncertainty-guided selection is relevant to defect problems because saddle points, charged defect geometries, reconstructed defect cores, and ion-migration bottlenecks may lie outside the configurational space adequately represented by the training data. The resulting workflow concentrates first-principles calculations on configurations for which ML predictions are less reliable or for which higher accuracy is required.

For high-throughput defect screening, ML acceleration should be integrated with a traceable workflow that retains the quantities required for defect thermodynamics, including formation energies, charge-transition levels, migration barriers, and correction terms. Standardized defect generation, ML-assisted candidate ranking, uncertainty-guided DFT refinement, and systematic metadata recording together provide a practical framework for automated defect calculations.

5.3. Active learning decision rules for defect PES refinement

In modeling complex defect PESs, it is generally impractical to generate a training set that covers all relevant configurations from the outset. Active learning provides a practical way to address this problem. As shown in Figure 8a, active learning iterates between model training and data sampling. New configurations are selected for first-principles labeling according to the current model uncertainty, reducing the number of DFT calculations required to reach a given level of accuracy [78,133–135]. The main challenge is to determine which configurations provide sufficient information gain to justify additional DFT calculations.

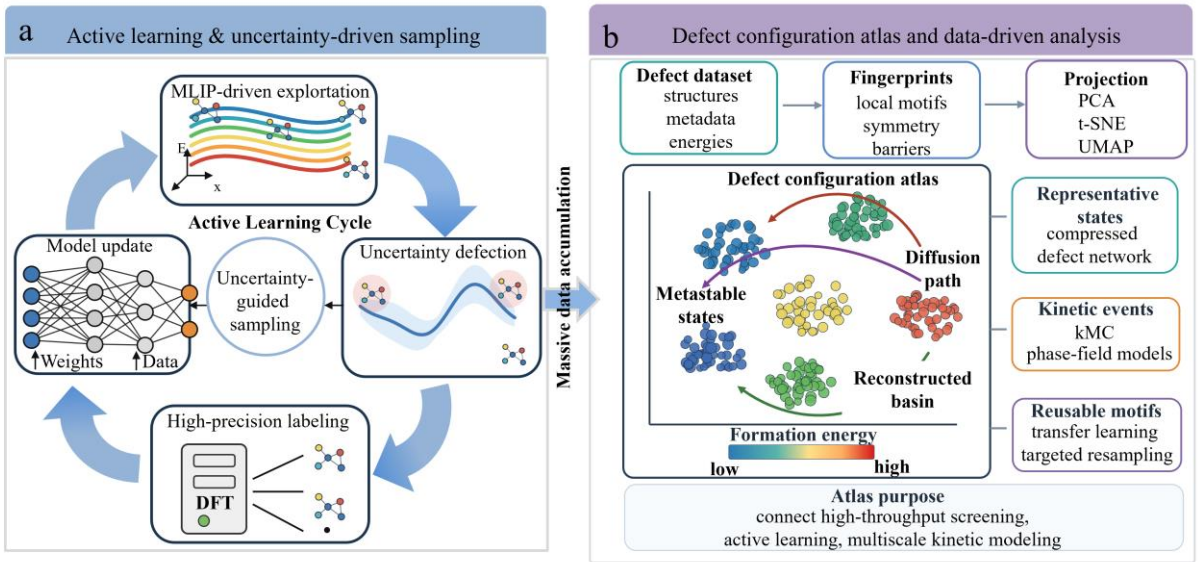

Figure 8. Data driven exploration of the defect configuration space. **(a)** Active learning and uncertainty driven sampling; **(b)** Construction and analysis of defect configuration atlases.

Uncertainty estimation is an important component of active learning. Gaussian processes (GPs) and Bayesian neural networks (BNNs) provide uncertainty estimates through their predictive distributions [26]. For conventional neural-network potentials, uncertainty is commonly approximated using model ensembles, Monte Carlo dropout, force disagreement, energy disagreement, or descriptor space distance [136,137]. In active sparse BCM workflows, disagreement among local expert models also be used as an uncertainty indicator when simulations encounter heterogeneous environments such as defect

cores, surfaces, grain boundaries, amorphous regions, or configurations sampled under finite pressure [111]. Because defect calculations often depend on small energy differences and accurately resolved transition states, both energy and force uncertainties should be considered, particularly for saddle-point structures and strongly distorted local environments [138].

A typical active-learning cycle consists of configuration sampling, uncertainty evaluation, first-principles labeling, model retraining, and validation. Candidate configurations may be generated by molecular dynamics, global structure searches, nudged elastic band calculations, defect enumeration, or finite-temperature sampling, depending on the process of interest. The uncertainty metric should be calibrated on an independent DFT dataset by comparing predicted uncertainty with actual errors in forces, energies, stresses, or migration barriers. After retraining, the model should be tested on defect types, migration pathways, and local chemical environments that were not represented in the training set.

A practical two-threshold scheme can be used for active-learning decisions. Let u denote the uncertainty indicator, with lower and upper thresholds u_{low} and u_{high}, respectively. The lower threshold can be calibrated on a held-out DFT dataset so that configurations exceeding a specified reference-error tolerance are identified with sufficiently high sensitivity. Simulations may continue when $u < u_{\mathrm{low}}$. Configurations with $u_{\mathrm{low}} \leq u < u_{\mathrm{high}}$ are added to the DFT-labeling queue. When $u \geq u_{\mathrm{high}}$, the configuration lies well outside the reliable region of the current model and should be labeled before further exploration. Because the uncertainty distribution changes as the training set grows, the thresholds should be reassessed after substantial model updates and, where necessary, for different phases, compositions, charge states, temperatures, stress states, and defect environments.

No single threshold applies to all materials. In DP-GEN, multiple models are trained on the same dataset with different parameter initializations, and the model deviation is defined using the largest atomic-force standard deviation across the ensemble. Reported candidate windows are 0.05–0.15 eV $\mathrm{Å}^{-1}$ for Al and Al–Mg and 0.03–0.13 eV $\mathrm{Å}^{-1}$ for Mg [134]. These values are specific to the corresponding models, datasets, and sampling conditions and should not be treated as universal criteria. Migration studies may require additional checks on saddle-point forces and the relative energies of competing pathways. For defect screening, the stability of formation-energy rankings should also be examined, whereas simulations under constant pressure or mechanical loading require validation of stresses and virials. The uncertainty criterion should therefore be calibrated according to the quantity of interest and the acceptable error of the target calculation [139–141].

5.4. Defect configuration atlases for kinetics and transfer learning

Large-scale searches and high-throughput defect calculations can generate many configurations that differ only in local reconstruction, charge state, coordination environment, or migration geometry. As illustrated in Figure 8b, a defect configuration atlas provides a framework for organizing these structures according to their structural and energetic similarities. Structural fingerprints are first calculated for individual defect configurations. Clustering and dimensionality-reduction methods such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP) can then be used to project high-dimensional structural descriptors onto a low-dimensional representation [142,143]. Formation energies, local symmetries, charge states, migration barriers, electronic descriptors, and model uncertainties may be mapped onto the same representation. Clusters of low-energy structures may identify stable or metastable defect

families, whereas configurations between clusters may correspond to reconstruction pathways or migration intermediates.

For kinetic modeling, configurations belonging to the same free-energy basin can be grouped into a single state. Transition pathways between these states, together with the corresponding barriers or free-energy differences, can then be used to determine event rates. The resulting network contains fewer states than the original atomistic dataset and can be used as an event catalog for kinetic Monte Carlo simulations [5,144]. Phase-field modeling requires an additional coarse-graining step because individual atomic configurations are no longer resolved. Information from the atlas and related atomistic calculations may nevertheless be used to estimate free-energy landscapes, kinetic coefficients, and transitions between competing defect populations.

The same structural representation can also be used to assess transfer to related materials. A local defect motif in a related compound or interface may fall within a region of configuration space already represented in the training set. In such cases, the existing model may provide a useful starting point for fine-tuning with a limited amount of additional data. In contrast, isolated configurations or structures associated with large prediction uncertainties indicate poorly sampled regions and require additional first-principles calculations [145,146]. Once labeled, these configurations can be incorporated into the training set and used for subsequent model retraining. Updating the atlas during this process also provides a way to track the expansion of the sampled defect configuration space.

6. Defect dynamics and multiscale simulation

6.1. MLIP-driven molecular dynamics and accelerated dynamics

Many macroscopic properties of materials depend on how defects evolve under service conditions, including irradiation, thermal exposure, and mechanical loading. At the atomistic level, MD remains a primary tool for studying such processes, but conventional force fields often lack the accuracy needed when bond rearrangement or defect core chemistry is important. MLIPs have extended the spatial and temporal range of MD while retaining improved accuracy. This has enabled simulations of irradiation cascades, vacancy-interstitial recombination, defect clustering, and related processes in systems much larger than those accessible to direct DFT [147].

For long trajectories, however, the reliability of MLIP-based MD must be checked explicitly. Common tests include monitoring energy drift in microcanonical simulations, comparing migration mechanisms across integration settings, and verifying diffusion coefficients or structural statistics against reference calculations or experiment when possible. MLIPs can also be combined with path integral MD to describe light element diffusion when nuclear quantum effects are relevant [148,149]. A small average force error does not necessarily imply accurate defect kinetics. Errors can become more important near saddle points, highly distorted coordination environments, and rare local rearrangements.

Rare event dynamics remain a separate challenge because many defect processes occur on timescales beyond standard MD [150]. Accelerated methods can extend the accessible timescale. Examples include temperature-accelerated dynamics, hyperdynamics, parallel replica dynamics, and metadynamics [151]. MLIPs reduce the cost of repeated energy and force evaluations during transition-pathway searches. However, barrier heights and representative transition states should still

be validated against DFT or a high-accuracy reference model when they control rate ordering or long time kinetic predictions.

Beyond dynamical processes, MLIPs have extended the calculation of finite-temperature defect thermodynamics, which is critical for predicting defect concentrations under real operating conditions. Static 0 K formation energies are rarely sufficient. Defect free energies may contain vibrational, electronic, magnetic, and configurational contributions. Finite-temperature sampling can also populate metastable configurations that are not represented by a single relaxed 0 K structure. A recent study on CdTe showed that finite-temperature entropic contributions can alter predicted defect concentrations by orders of magnitude compared to static DFT calculations [93]. We therefore distinguish three levels of defect thermodynamic modeling. The first uses static DFT formation energies and is well suited to rapid high-throughput screening. The second incorporates harmonic or quasi-harmonic corrections to improve accuracy at moderate temperatures. The third relies on MLIP-enabled anharmonic thermodynamic integration or free-energy sampling when temperature-dependent disorder is central to the material function.

6.2. From MLIPs to kinetic Monte Carlo rate models

KMC provides a route from atomistic barrier information to long-timescale defect evolution. In lattice kMC, the system evolves through a set of discrete events. Each event is assigned a rate from transition-state theory or a related kinetic model. KMC can reach timescales far beyond those of MD. Its accuracy, however, depends on whether the relevant events are included in the event library. Building that library remains one of the main practical bottlenecks in applying kMC to complex defect problems.

Barrier errors are amplified when they are converted into kinetic rates. Within transition-state theory, the rate of event i is given by Equation (2):

$$k_i = \nu_i \exp\left(-\frac{E_{\mathrm{b},i}}{k_{\mathrm{B}}\mathrm{T}}\right), \tag{2}$$

Where ν_i is the rate prefactor and $E_{\mathrm{b},i}$ is the migration barrier. If the ML-predicted barrier differs from the reference value, the barrier error is defined by Equation (3):

$$\Delta E_{b,i} = E_{b,i}^{\mathrm{ML}} - E_{b,i}^{\mathrm{ref}}, \tag{3}$$

The corresponding rate ratio is given by Equation (4):

$$\frac{k_i^{\mathrm{ML}}}{k_i^{\mathrm{ref}}} = exp\left(-\frac{\Delta E_{b,i}}{k_B T}\right). \tag{4}$$

For a barrier error of 0.05 eV, the predicted rate changes by a factor of approximately 6.9 at 300 K, 2.6 at 600 K, and 1.8 at 1000 K, assuming that the rate prefactor remains unchanged. Underestimating the barrier accelerates the event, while overestimating it slows the event. Because kMC uses these rates to select events and advance the simulation time, barrier errors can reorder competing pathways and shift the predicted defect populations and timescales.

Barrier uncertainty can be propagated directly into kMC. One approach is to sample barriers from calibrated distributions and construct a separate rate catalog for each sample. The resulting simulations provide distributions of defect concentrations, diffusion coefficients, segregation profiles, and degradation

times. Sensitivity analysis can identify the events that contribute most strongly to the final uncertainty. These events are natural targets for additional DFT transition-state calculations. If the ordering of competing events changes across the sampled catalogs, several mechanisms should be retained in the interpretation. Section 6.5 summarizes the corresponding reporting requirements.

ML can reduce the cost of constructing and updating the event library in two ways. First, MLIPs can provide efficient energy and force evaluators for transition-state searches. MLIPs can be combined with automated transition-state methods such as CI-NEB and the dimer method [107,152]. This combination reduces the cost of locating saddle points and evaluating migration barriers in large atomic systems. Active learning can further prioritize high-uncertainty configurations near the saddle region for DFT refinement. As illustrated in Figure 9a, uncertainty estimates determine whether an event is evaluated directly by the ML model or submitted for additional DFT validation [78,153]. A second approach is to predict migration barriers or event rates directly from local structural descriptors. This approach enables on-the-fly rate evaluation without explicitly enumerating and storing every possible event. Conventional lattice kMC becomes less suitable when the atomic topology changes substantially. Examples include dislocation climb and vacancy-cluster growth. Off-lattice and graph-based kMC methods provide greater flexibility for such processes [154].

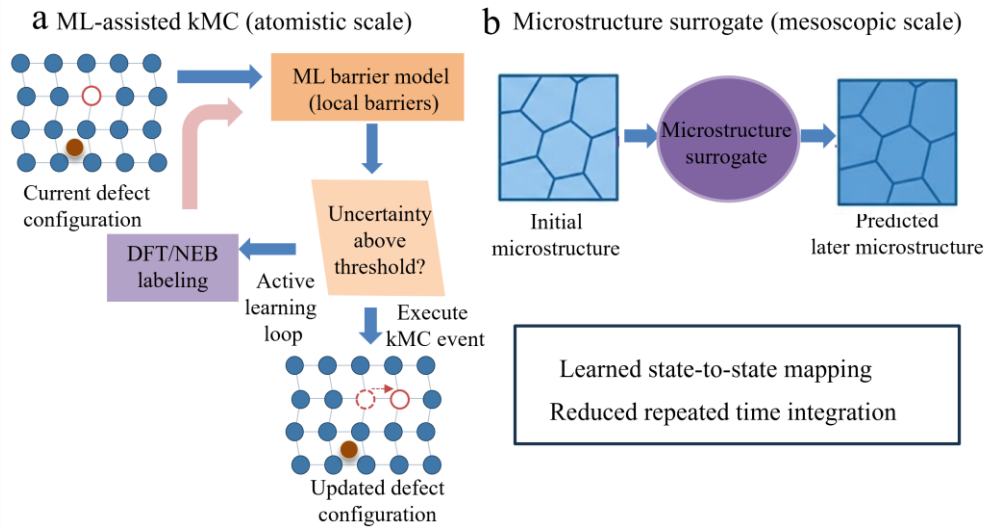

Figure 9. ML-assisted strategies for long-timescale defect kinetics. **(a)** Uncertainty-aware barrier prediction, first-principles labeling, and model updating for kMC simulations; **(b)** Surrogate prediction of mesoscale microstructure evolution within a validated parameter and time range.

Jung *et al.* coupled an ACSF-based neural-network potential with off-lattice kMC to investigate degradation of Pt_3Co nanoparticles [120]. The training dataset contained 4576 DFT configurations, corresponding to 514,449 local atomic environments. The validation-set RMSEs were 4.46 meV atom^{-1}, 0.13 eV Å^{-1}, and 6.72 kbar for energy, force, and stress, respectively. For surface-vacancy formation energies and vacancy-migration barriers, the RMSEs were 0.136 and 0.227 eV, with corresponding R^2 values of 0.938 and 0.847. The off-lattice kMC simulations considered nanoparticles with diameters of 2.52, 3.28, and 4.04 nm and followed surface-atom dissolution and vacancy migration for simulated physical times of up to 50 h. Notably, the migration-barrier RMSE was much larger than the energy RMSE, illustrating why kinetic benchmarks should be reported separately from general potential accuracy.

6.3. Mesoscopic scale: dislocation dynamics and phase-field models

At the mesoscopic scale, Discrete Dislocation Dynamics (DDD) and phase-field models are important tools for linking atomic scale defect behavior with microstructural evolution. Traditional DDD simulations typically treat dislocations as line defects evolving within a continuous elastic background [155,156], while phase-field models describe microstructure evolution through continuous order parameters and free energy functionals. The predictive accuracy of both methods depends strongly on input parameters such as dislocation mobility laws, defect interaction energies, elastic constants, interfacial energies, diffusion coefficients, and kinetic mobilities. These parameters have traditionally been obtained from experiments or a limited number of first-principles calculations. Such datasets may be insufficient for complex multicomponent alloys and multiphase systems.

MLIP-driven MD can sample dislocation motion over selected stresses, temperatures, compositions, and chemical environments [157,158]. The resulting data can be used to construct mobility laws or regression models [159]. These models relate dislocation velocity to stress, temperature, local composition, and nearby defects. This approach is useful when dislocation motion depends strongly on local chemistry or defect-core reconstruction. Under these conditions, simple empirical mobility laws may not transfer well between systems.

MLIPs can also support phase-field parameterization. High-throughput atomistic simulations can sample energy and stress responses across different compositions, phases, and defect concentrations. These quantities can be used to parameterize phase-field models. Surrogate models can also be trained on phase-field trajectories (Figure 9b) [160]. They may predict later microstructures from an initial state or advance the system over selected time intervals. Evaluation should go beyond image similarity. Morphological statistics, conserved quantities, boundary conditions, and kinetic trends provide additional tests over the intended parameter and time range.

Uncertainty propagation remains important when atomistic quantities are transferred to mesoscale models. Errors in MLIP-predicted defect energies, migration barriers, elastic constants, and interfacial energies can affect dislocation mobility laws, phase-boundary motion, and microstructure evolution. MLIP-informed DDD and phase-field studies should therefore report the training domain, reference accuracy, parameter uncertainty, and validation range.

6.4. Neural operators and physics-informed continuum modeling

At the macroscopic scale, the mechanical, thermal, and electrical transport behavior of defect-containing materials are commonly described by partial differential equations (PDEs) [161]. Neural operators and physics-informed neural networks (PINNs) are increasingly used to accelerate PDE calculations and inverse parameter identification [162–164]. The two approaches, however, use different learning strategies. Neural operators learn mappings between functions rather than pointwise input-output relations. PINNs take a different approach by incorporating governing equations and physical constraints into the training objective [163,165].

In modeling defect-containing materials, MLIPs can be combined with Neural Operators or PINNs to build multiscale data-driven PDE solvers. MLIPs and atomistic simulations can quantify how defects modify elastic constants, diffusion coefficients, and interface mobilities (Figure 10a), These defect-dependent properties can then be transferred to continuum models as spatially varying material

parameters (Figure 10b). Neural Operators can learn the response of the continuum system to these heterogeneous inputs, allowing rapid prediction of structural response and degradation for prescribed defect distributions and boundary conditions (Figure 10c) [166].

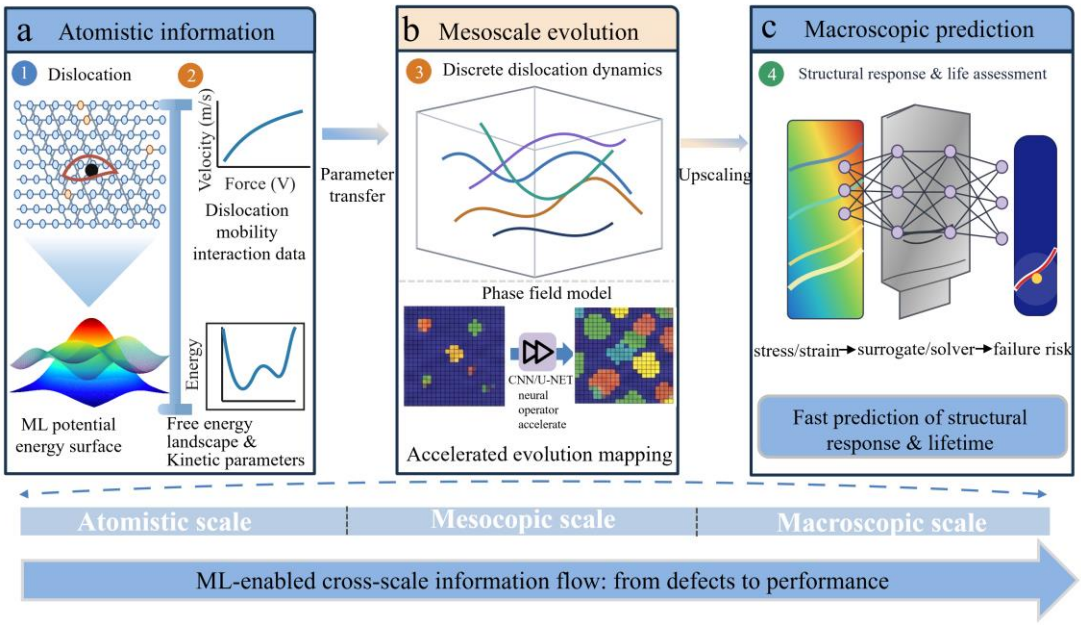

Figure 10. ML-enabled multiscale information flow. **(a)** Atomistic defect and kinetic information; **(b)** Mesoscale evolution through dislocation and phase-field models; **(c)** Macroscopic prediction of structural response and lifetime.

This strategy is attractive for rapid screening, but several challenges remain, including robust treatment of boundary conditions, strict enforcement of conservation laws, and accurate representation of singular fields near crack tips or dislocation cores [167,168]. Uncertainty in MLIP-derived inputs, including diffusion coefficients, elastic constants, and defect mobilities, should also be propagated to continuum-scale predictions. Neural operators and PINNs are most useful as components of a validated multiscale workflow. They can accelerate repeated calculations, but they do not remove the need for atomistic benchmarks or physically consistent continuum models.

6.5. Comparative applicability and quantitative reporting across scales

These methods differ mainly in accessible length and time scales, the amount of physical detail retained, and computational cost [169]. MLIP-based MD provides direct atomic-scale information on processes such as defect formation, migration, and recombination. KMC can extend simulations to experimentally relevant timescales, provided that the dominant events and their rates are adequately represented. However, its reliability depends critically on the comprehensiveness of the event library. DDD and phase-field methods span larger spatial and temporal scales by introducing coarse grained descriptions, making them particularly suitable for studying dislocation network evolution and microstructural coarsening. Neural operators and PINNs operate at the continuum-field level and can accelerate repeated calculations over large parameter spaces.

Coarse-graining inevitably removes part of the lower-scale information. The accuracy of multiscale predictions therefore depends on how the remaining information is transferred between models. Atomistic quantities transferred to higher-level models need independent benchmarks [170,171]. Their

uncertainty should also be reported when it can affect the higher-scale prediction. Conversely, continuum or mesoscale predictions, such as stress strain response or defect density evolution, should be validated against experimental observations whenever possible [172]. Multiscale predictions are more convincing when atomistic parameterization is combined with validation at higher scales [173,174]. Accuracy cannot be assumed to transfer unchanged from one model to the next.

Quantitative reporting conventions. For consistent comparison, let y_i^{ML} and y_i^{ref} denote the predicted and reference values, respectively, for M evaluated quantities. The mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R2) are defined by Equations (5)–(7), respectively:

$$\mathrm{MAE}=\frac{1}{M}\sum_{i=1}^{M}\left|y_i^{ML}-y_i^{ref}\right|, \tag{5}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{M}\sum_{i=1}^{M}(y_i^{ML}-y_i^{ref})^2}, \tag{6}$$

$$R^2=1-\frac{\sum_{i=1}^{M}(y_i^{ML}-y_i^{ref})^2}{\sum_{i=1}^{M}(y_i^{ref}-\bar{y}^{ref})^2}. \tag{7}$$

here, y_i may represent energies, cartesian force components, stress components, or other scalar targets. In MLIP benchmarking, each reported error should be accompanied by the target quantity, dataset split, normalization convention, unit, and reference level [171].

For migration pathway j, the barrier is defined as the energy difference between the transition state and the initial state [137,151], as given by Equation (8):

$$E_{b,j}=E_{TS,j}-E_{IS,j}, \tag{8}$$

where $E_{TS,j}$ and $E_{IS,j}$ are the energies of the transition and initial states, respectively. The signed barrier error is defined by Equation (9):

$$\Delta E_{b,j}=E_{b,j}^{ML}-E_{b,j}^{ref}, \tag{9}$$

whereas barrier MAE or RMSE is evaluated over the complete set of benchmark pathways. The term barrier error should therefore be accompanied by the specific metric used.

Training-set size should distinguish the number of atomic configurations from the number of local atomic environments or training points. The total number of local environments is given by Equation (10):

$$N_{env}=\sum_{S=1}^{N_{conf}}N_s, \tag{10}$$

where N_S is the number of atoms in configuration S. Accessible timescale denotes simulated physical time rather than computational wall time. For MD, $t_{MD}=N_{step}\Delta t$, whereas the kMC timescale is accumulated from stochastic event-time increments. System size should be specified by the number of atoms and a physical dimension, such as the simulation-cell dimensions or nanoparticle diameter.

The most relevant metrics depend on the type of multiscale coupling. For MLIP-kMC workflows, the most important quantities are the migration barriers and event rates. Reports should therefore specify the DFT reference level, barrier errors, event-catalog coverage, system size, simulated physical time,

and uncertainty in the kinetic rates. For MLIP-informed DDD and phase-field models, attention should shift to the accuracy of transferred parameters. Relevant quantities include elastic constants, defect interaction energies, interfacial energies, diffusion coefficients, and mobility parameters. Their validated composition, temperature, and defect-concentration ranges should also be stated. For continuum and neural-operator models, uncertainty in MLIP-derived parameters should be propagated to predicted stress, transport properties, and degradation behavior. Recent reviews and device-scale studies have considered accuracy, transferability, computational cost, and application range together when assessing MLIP performance [175,176] These benchmarks allow readers to judge whether atomistic predictions are accurate enough for use at the next scale. This is particularly important for migration barriers, diffusion coefficients, dislocation mobility relations, and interfacial energies. Without such quantitative information, links between atomistic predictions and device-scale performance remain difficult to assess and validate.

7. Challenges, validation, and frontier trends

ML approaches can accelerate defect-structure search, MLIP development, high-throughput screening, and large-scale atomistic simulations, but average test errors alone are insufficient to establish model reliability. Defect applications often involve rare configurations, small energy differences, long-range interactions, finite-temperature effects, and extrapolation to new chemistries or structures. This section discusses task-specific validation, physical consistency, data quality, and emerging workflows that retain human oversight.

7.1. Model validation, extrapolability, and benchmarking

Figure 11 reframes validation as a closed workflow rather than a single accuracy score. Defect datasets are typically highly imbalanced, as common local environments are well represented whereas rare configurations, unusual charge states, high-temperature structures, strained states, and chemically complex interfaces remain sparsely sampled. Under these conditions, models evaluated using conventional random splits may appear accurate while still failing for unseen defect classes or thermodynamic conditions outside the training domain [158,177]. More stringent validation should use out-of-distribution test sets that hold out entire defect families, chemistries, charge states, temperature ranges, or stress regimes relevant to the target task [178,179].

Defect-specific benchmarks should cover quantities such as formation energies, charge-transition levels, migration barriers, transport coefficients, and damage evolution [180,181]. Validation should establish whether the model is reliable within its intended domain, requires additional training data, or is unsuitable for the target task. This assessment should combine numerical errors with physical checks, including energy-force consistency, stress symmetry, charge conservation, long-range electrostatic response, and finite-temperature behavior [24,182].

Table 3 translates the workflow in Figure 11 into a reporting checklist for ML-assisted defect studies. Not every item applies equally to all defect classes. For example, charge-state and electrostatic corrections are essential for charged point defects, whereas stress and virial validation is more relevant to mechanical, dislocation, constant-pressure, and pressure-dependent simulations.

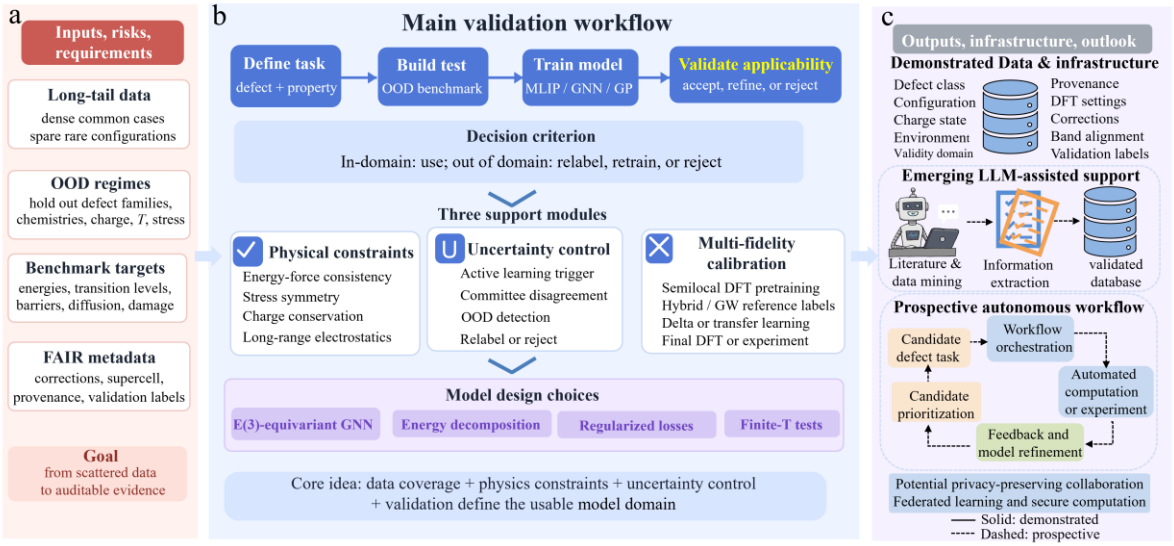

Figure 11. Validation workflow and emerging directions for ML-enabled defect research. **(a)** Input risks, benchmark targets, and metadata requirements for defect-oriented model development; **(b)** Task-specific validation within a defined applicability domain, including physical constraints, uncertainty control, and multi-fidelity calibration; **(c)** Established data and metadata infrastructure together with emerging LLM-assisted, autonomous, and privacy-preserving workflow. Solid elements denote demonstrated components, while dashed elements denote emerging directions that require human oversight and independent physical validation.

Table 3. Minimum reporting checklist for ML-assisted defect studies.

Checklist item	Minimum reporting or validation requirement
Defect task	defect class; host chemistry; target property; temperature; stress; chemical-potential conditions
Charge-state treatment	charge state(s); Fermi-level range; charge localization; polaron treatment
Reference electronic-structure method	DFT code; exchange-correlation functional; pseudopotential or basis; cutoff and k-point sampling; spin or SOC treatment; convergence criteria
Finite-size and electrostatic corrections	supercell size; charge-correction method; dielectric response; potential or band-edge alignment; chemical-potential convention
Training-domain coverage	bulk structures; defect cores; metastable states; saddle-point regions; interfaces; finite-temperature; strain and high-uncertainty configurations relevant to the task
Uncertainty control	uncertainty metric; calibration dataset; query or relabeling threshold; accept/retrain/reject criterion
Task-specific validation	energies; forces; structures; formation energies; transition levels; migration barriers; diffusion coefficients; kinetic observables
Stress/virial and physical consistency	stress; virial; pressure validation; energy-force consistency; symmetry and conservation tests; finite-temperature stability
Transferability assessment	held-out defect classes; chemistries; charge states; temperatures; stress states; in-domain and out-of-domain performance
Reproducibility and provenance	numbers of configurations and local environments; dataset split; model and software versions; data provenance; independent DFT or experimental benchmarks

NOTE: Each item should be identified as applied, not applicable, or not tested. Accuracy targets and validation depth should reflect the target defect mechanism and the downstream use of the ML predictions in thermodynamic, kinetic, or multiscale analyses.

7.2. Frontier trends: physical constraints, data quality, and multi-fidelity strategies

Physical constraints become particularly important when the target property depends on physics that is poorly represented in the training data. Depending on the material and defect mechanism, relevant constraints may include equivariance, energy–force consistency, stress symmetry, charge conservation, and explicit treatments of long-range electrostatics or dielectric response. For charged defects, polarons, redox-active materials, and polar semiconductors, validation should also examine charge localization, electrostatic response, and defect-induced polarization, rather than only short-range energies and forces [104,182,183]. These constraints can reduce unphysical predictions but do not replace defect-specific validation [184].

Data fidelity remains a major limitation. Large semilocal DFT datasets are useful for pretraining and broad chemical coverage, but defect energetics often depend on band gaps, charge localization, band-edge alignment, and charge-transition levels. These quantities may require hybrid functionals or many-body corrections for reliable calibration. Multi-fidelity approaches can combine large, low-cost datasets with smaller high-fidelity reference sets through transfer learning, multitask learning, or delta learning [112,185]. The transferability of models trained mainly on lower-fidelity data should be assessed after calibration against a limited high-fidelity reference set. For point defects, this should be tested by holding out entire defect types, charge states, transition levels, and chemical potential regimes, rather than only testing bulk energies and forces.

Finite-temperature effects should also be treated as a validation target rather than a minor correction. Static zero-temperature formation energies may miss vibrational, configurational, anharmonic, and metastable state contributions to defect free energies. MLIP-based thermodynamic integration and finite-temperature sampling can capture these contributions, provided that the underlying potential is accurate for thermally accessible defect configurations.

7.3. Data, software ecosystems, and emerging autonomous workflows

Broader application of ML in defect research depends on reliable data and software infrastructure. Defect databases should follow Findable, Accessible, Interoperable, and Reusable (FAIR) principles while recording key methodological information, including supercell settings, charge-correction schemes, computational parameters, validation results, and data provenance [186–189]. Software platforms should also support workflow management, metadata checking, and standardized interfaces between electronic-structure calculations and machine-learning models. These components form the established infrastructure shown in Figure 11a.

Several parts of ML-assisted defect workflows are already automated. Existing software can automate defect-structure generation, charge-state enumeration, finite-size corrections, and calculation record management. More recently, first-principles calculations have been combined with MLIPs and machine-learning Hamiltonians for structural relaxation, defect formation-energy prediction, and configurational search. For example, machine-learning Hamiltonians have been used for structural relaxation and oxygen-vacancy formation-energy calculations in amorphous SiO_2 [190]. While MLIPs have been used to accelerate defect-structure exploration in amorphous silica [191]. These studies demonstrate that several key steps in defect calculations can already be assisted by ML and incorporated into the validation-centered workflow outlined in Figure 11b. However, current implementations mainly involve individual computational modules rather than fully autonomous, end-to-end defect-discovery workflows.

The dashed outline in Figure 11c denotes less mature LLM-assisted workflows. Current uses include literature and data retrieval, information extraction, script generation, workflow documentation, and error-log analysis [192–195]. These tools can reduce routine work, but their outputs still require human verification. Defect-model selection, mechanistic interpretation, uncertainty assessment, and first-principles validation should therefore remain subject to expert review and independent physical calculations.

A fully autonomous defect-discovery platform would integrate candidate-structure generation, high-throughput calculations, active learning, uncertainty-based decision making, and, where available, automated synthesis or characterization within a closed-loop workflow. However, end-to-end implementations remain uncommon in defect research. Reliable operation will require robust interfaces between modules, calibrated uncertainty thresholds, complete data provenance, human oversight, and independent physical validation. Federated learning may support cross-institutional collaboration without centralized data sharing, although its application to defect research remains limited [196].

8. Conclusion

ML can broaden the range of tractable defect problems by improving configurational sampling, accelerating atomistic calculations, and transferring information across different length and time scales. In particular, MLIPs, active-learning strategies and off-lattice kinetic models can extend the accessible system sizes and simulation times within their validated domains. However, sparse training data, limited transferability, and inappropriate physical assumptions can still lead to substantial errors, so ML predictions should remain tied to first-principles calculations and mechanistic analysis. Reliable use therefore requires direct tests of the physics relevant to the target problem, including long-range electrostatic interactions in polar materials, finite-temperature anharmonic effects, and out-of-distribution configurations along defect-transition pathways. ML has already been applied to defect-structure exploration, migration analysis, finite-temperature sampling, and selected multiscale simulations. Fully autonomous multi-fidelity workflows and self-driving defect-discovery platforms, however, are still at an early stage. Further progress requires calibrated uncertainty estimates, traceable data provenance, human oversight, and independent validation against first-principles calculations or experiments. The most reliable workflows are therefore likely to combine physical insight with data-driven models and explicit uncertainty assessment.

Data availability statement

No supplementary or additional data were generated in this study.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) tools only for language polishing, readability improvement, and preliminary visual ideation. All scientific content, literature interpretation, figure design decisions, and final conclusions were checked and approved by the authors. AI-generated images are not used as a substitute for author-created scientific figures in the final revised graphical materials.

Acknowledgments

D.H. acknowledges the support by the National Natural Science Foundation of China (Grant No. 12404268), and the Fundamental Research Funds for the Central Universities, JLU.

Authors' contribution

Conceptualization, D.H.; investigation, F.R. and S.H.; data curation, F.R. and S.H.; writing—original draft preparation, F.R. and S.H.; writing—review and editing, G.Z., Z.C. and D.H.; project administration and funding acquisition, D.H. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, *et al*. First-principles calculations for point defects in solids. *Rev. Mod. Phys.* 2014, 86(1):253–305.

[2] Mosquera-Lois I, Huang Y, Lohan H, Ye J, Walsh A, *et al*. Multifaceted nature of defect tolerance in halide perovskites and emerging semiconductors. *Nat. Rev. Chem.* 2025, 9:287–304.

[3] Barah OO, Natukunda F, Bori I, Ukagwu KJ. Mechanisms and modelling of diffusion in solids: a multiscale framework with industrial case studies and AI enhancements. *Discover Sustain.* 2025, 6:804.

[4] Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, *et al*. The high-throughput highway to computational materials design. *Nat. Mater.* 2013, 12(3):191–201.

[5] Zhang L, Han J, Wang H, Car R, E W. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* 2018, 120(14):143001.

[6] Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, *et al*. Machine learning force fields. *Chem. Rev.* 2021, 121(16):10142–10186.

[7] Bartók AP, Payne MC, Kondor R, Csányi G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* 2010, 104(13):136403.

[8] Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature* 2018, 559(7715):547–555.

[9] Chen H, Ortner C, Wang Y. QM/MM methods for crystalline defects. Part 3: machine-learned MM models. *Multiscale Model. Simul.* 2022, 20(4):1490–1518.

[10] Freitas R, Cao Y. Machine-learning potentials for crystal defects. *MRS Commun.* 2022, 12(5):510–520.

[11] Unruh D, Meidanshahi RV, Goodnick SM, Csányi G, Zimányi GT. Gaussian approximation potential for amorphous Si: H. *Phys. Rev. Mater.* 2022, 6(6):065603.

[12] Fan Y. Atomistic modeling of microstructural defect evolution in alloys under irradiation: a comprehensive review. *Appl. Sci.* 2025, 15(16):9110.

[13] Elder KR, Grant M. Modeling elastic and plastic deformations in nonequilibrium processing using phase field crystals. *Phys. Rev. E* 2004, 70(5):051605.

[14] Pickard CJ, Needs RJ. Ab initio random structure searching. *J. Phys.: Condens. Matter* 2011, 23(5):053201.
[15] Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* 2007, 98(14):146401.
[16] Deringer VL, Caro MA, Csanyi G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* 2019, 31(46):e1902765.
[17] Chen L. Phase-field models for microstructure evolution. *Annu. Rev. Mater. Res.* 2002, 32(1):113–140.
[18] Broberg D, Medasani B, Zimmermann NER, Yu G, Canning A, *et al*. PyCDT: a Python toolkit for modeling point defects in semiconductors and insulators. *Comput. Phys. Commun.* 2018, 226:165–179.
[19] Huang M, Zheng Z, Dai Z, Guo X, Wang S, *et al*. DASP: defect and dopant ab-initio simulation package. *J. Semicond.* 2022, 43(4):042101.
[20] Huber SP, Zoupanos S, Uhrin M, Talirz L, Kahle L, *et al*. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* 2020, 7(1):300.
[21] Mosquera-Lois I, Kavanagh SR, Ganose AM, Walsh A. Machine-learning structural reconstructions for accelerated point defect calculations. *npj Comput. Mater.* 2024, 10(1):121.
[22] Lu S, Zhou Q, Ouyang Y, Guo Y, Li Q, *et al*. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* 2018, 9(1):3405.
[23] Chen C, Ong SP. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* 2022, 2(11):718–728.
[24] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, *et al*. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* 2022, 13(1):2453.
[25] Musaelian A, Batzner S, Johansson A, Sun L, Owen CJ, *et al*. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* 2023, 14(1):579.
[26] Bartók AP, Kermode J, Bernstein N, Csányi G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* 2018, 8(4):041048.
[27] Jiang C, Marianetti CA, Khafizov M, Hurley DH. Machine learning potential assisted exploration of complex defect potential energy surfaces. *npj Comput. Mater.* 2024, 10(1):21.
[28] Jia W, Wang H, Chen M, Lu D, Lin L, *et al*. Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC20)*, Atlanta, USA, November 9–19, 2020. pp. 1–14.
[29] Frolov T, Olmsted DL, Asta M, Mishin Y. Structural phase transformations in metallic grain boundaries. *Nat. Commun.* 2013, 4:1899.
[30] Bonati L, Parrinello M. Silicon liquid structure and crystal nucleation from Ab initio deep metadynamics. *Phys. Rev. Lett.* 2018, 121(26):265701.
[31] Byggmästar J, Hamedani A, Nordlund K, Djurabekova F. Machine-learning interatomic potential for radiation damage and defects in tungsten. *Phys. Rev. B* 2019, 100(14):144105.
[32] Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* 2017, 3(1):54.
[33] Nieminen RM. Issues in first-principles calculations for defects in semiconductors and oxides. *Model. Simul. Mater. Sci. Eng.* 2009, 17(8):084001.

[34] Wang F, Bai S, Tress W, Hagfeldt A, Gao F. Defects engineering for high-performance perovskite solar cells. *npj Flex. Electron.* 2018, 2(1):22.

[35] Dubajic M, Neilson JR, Klarbring J, Liang X, Bird SA, *et al*. Dynamic nanodomains dictate macroscopic properties in lead halide perovskites. *Nat. Nanotechnol.* 2025, 20(6):755–763.

[36] Zhang H, Lv R. Defect engineering of two-dimensional materials for efficient electrocatalysis. *J. Materiomics* 2018, 4(2):95–107.

[37] Liu D, Luo D, Iqbal AN, Orr KWP, Doherty TAS, *et al*. Strain analysis and engineering in halide perovskite photovoltaics. *Nat. Mater.* 2021, 20(10):1337–1346.

[38] He M, Ma Y, Wu S, Tan H, Dong W, *et al*. Amorphous grain boundary engineering for scalable flexible perovskite photovoltaics with improved stability. *Nat. Energy* 2026, 11(2):262–273.

[39] Loh L, Ning S, Kieczka D, Chen Y, Yang J, *et al*. Electron ptychography for atom-by-atom quantification of 1D defect complexes in monolayer MoS_2. *ACS Nano* 2025, 19(6):6195–6208.

[40] Zhang Y, Liu J, Xu Y, Xie C, Wang S, *et al*. Design and regulation of defective electrocatalysts. *Chem. Soc. Rev.* 2024, 53(21):10620–10659.

[41] Liu C, Yang Y, Chen H, Spanopoulos I, Bati ASR, *et al*. Two-dimensional perovskitoids enhance stability in perovskite solar cells. *Nature* 2024, 633:359–364.

[42] Grancini G, Nazeeruddin MK. Dimensional tailoring of hybrid perovskites for photovoltaics. *Nat. Rev. Mater.* 2019, 4(1):4–22.

[43] Wang Z, Gao H, Wu D, Meng J, Deng J, *et al*. Defects and defect passivation in perovskite solar cells. *Molecules* 2024, 29(9):2104.

[44] Park JS, Kim S, Xie Z, Walsh A. Point defect engineering in thin-film solar cells. *Nat. Rev. Mater.* 2018, 3(7):194–210.

[45] Yu J, Tao J, Yuan F, Zhang X. The dislocation dynamics study and machine learning approach of the critical stress for dislocation shearing and bypassing precipitates. *Acta Mech. Solida Sin.* 2025, 39(6):798–806.

[46] Yin W, Shi T, Yan Y. Unusual defect physics in $CH_3NH_3PbI_3$ perovskite solar cell absorber. *Appl. Phys. Lett.* 2014, 104(6):063903.

[47] Parrott ES, Milot RL, Stergiopoulos T, Snaith HJ, Johnston MB, *et al*. Effect of structural phase transition on charge-carrier lifetimes and defects in $CH_3NH_3PbI_3$ perovskite. *J. Phys. Chem. Lett.* 2016, 7(7):1321–1326.

[48] Jain A, Ong SP, Hautier G, Chen W, Richards WD, *et al*. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* 2013, 1(1):011002.

[49] Zhang S, Wei S, Zunger A. A phenomenological model for systematization and prediction of doping limits in II-VI and I-III-VI_2 compounds. *J. Appl. Phys.* 1998, 83(6):3192–3196.

[50] Cai X, Luo J, Li S, Wei S, Deng H. Overcoming the doping limit in semiconductors via illumination. *Phys. Rev. B* 2022, 106(21):214102.

[51] Curtarolo S, Setyawan W, Hart GLW, Jahnatek M, Chepulskii RV, *et al*. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* 2012, 58:218–226.

[52] Goyal A, Gorai P, Peng H, Lany S, Stevanović V. A computational framework for automation of point defect calculations. *Comput. Mater. Sci.* 2017, 130:1–9.

[53] Hu J, Chen Z. Atomistic study of shock Hugoniot in columnar nanocrystalline copper. *Comput. Mater. Sci.* 2021, 197:110635.

[54] Mishin Y, Asta M, Li J. Atomistic modeling of interfaces and their impact on microstructure and properties. *Acta Mater.* 2010, 58(4):1117–1151.
[55] Bulatov VV, Hsiung LL, Tang M, Arsenlis A, Bartelt MC, *et al*. Dislocation multi-junctions and strain hardening. *Nature* 2006, 440(7088):1174–1178.
[56] Zhang Z, Li X, Wang H, Wang J, Cui Y, *et al*. Mechanism of dislocation network formation and strengthening in additive manufactured alloys. *Acta Mater.* 2024, 266:119665.
[57] Kulichenko M, Nebgen B, Lubbers N, Smith JS, Barros K, *et al*. Data generation for machine learning interatomic potentials and beyond. *Chem. Rev.* 2024, 124(24):13681–13714.
[58] Dreyer CE, Van de Walle CG. First-principles theory of point defects in semiconductors. *Annu. Rev. Mater. Res.* 2024, 54:147–175.
[59] Perdew JP. Density functional theory and the band gap problem. *Int. J. Quantum Chem.* 1985, 28(S19):497–523.
[60] Lany S, Zunger A. Assessment of correction methods for the band-gap problem and for finite-size effects in supercell defect calculations: case studies for ZnO and GaAs. *Phys. Rev. B* 2008, 78(23):235104.
[61] Zhu Y, Du Y, Liu J, Wang H, Mukei T, *et al*. A generalizable machine learning framework for predicting quasiparticle band structures. *Nat. Comput. Sci.* 2024, 4:566–576.
[62] Hedin L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys. Rev.* 1965, 139(3A): A796–A823.
[63] Komsa HP, Rantala TT, Pasquarello A. Finite-size supercell correction schemes for charged defect calculations. *Phys. Rev. B* 2012, 86(4):045112.
[64] Pilania G, Goldsmith BR, Yoon M, Dongare AM. Recent advances in computational materials design: methods, applications, algorithms, and informatics. *J. Mater. Sci.* 2022, 57(23):10471–10474.
[65] Li Z, Wang X, Zhang L, Wan J, Zhang S. Logic of the layout of the solid electrolyte interphase. *Nat. Nanotechnol.* 2024, 19:89–97.
[66] Senftle TP, Hong S, Islam MM, Kylasa SB, Zheng Y, *et al*. The ReaxFF reactive force-field: development, applications and future directions. *npj Comput. Mater.* 2016, 2(1):15011.
[67] Mishin Y. Atomistic modeling of the γ and γ′-phases of the Ni-Al system. *Acta Mater.* 2004, 52(6):1451–1467.
[68] Brenner DW. The art and science of an analytic potential. *Phys. Status Solidi B* 2000, 217(1):23–40.
[69] Yan S, Wei J, Wang S, Huang M, Wu Y, *et al*. Defect physics of the quasi-two-dimensional photovoltaic semiconductor GeSe. *Chin. Phys. B* 2022, 31(11):116103.
[70] Kavanagh SR, Squires AG, Nicolson A, Mosquera-Lois I, Ganose AM, *et al*. doped: python toolkit for robust and repeatable charged defect supercell calculations. *J. Open Source Softw.* 2024, 9(96):6433.
[71] Mosquera-Lois I, Kavanagh SR, Walsh A, Scanlon DO. Identifying the ground state structures of point defects in solids. *npj Comput. Mater.* 2023, 9:25.
[72] Mosquera-Lois I, Kavanagh SR, Walsh A, Scanlon DO. ShakeNBreak: navigating the defect configurational landscape. *J. Open Source Softw.* 2022, 7(80):4817.
[73] Choudhary K, Garrity KF, Sharma V, Biacchi AJ, Walker ARH, *et al*. High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric and dielectric responses. *npj Comput. Mater.* 2020, 6(1):64.

[74] Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, *et al*. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* 2015, 1(1):15010.

[75] Olawade DB, Fapohunda O, Usman SO, Akintayo A, Ige AO, *et al*. Artificial intelligence in computational and materials chemistry: prospects and limitations. *Chem. Afr.* 2025, 8(7):2707–2721.

[76] Drautz R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* 2019, 99(1):014104.

[77] Shuaibi M, Sivakumar S, Chen RQ, Ulissi ZW. Enabling robust offline active learning for machine learning potentials using simple physics-based priors. *Mach. Learn.: Sci. Technol.* 2021, 2(2):025007.

[78] Podryabinkin EV, Shapeev AV. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* 2017, 140:171–180.

[79] Lany S, Zunger A. Polaronic hole localization and multiple hole binding of acceptors in oxide wide-gap semiconductors. *Phys. Rev. B* 2009, 80(8):085202.

[80] Zhang S, Northrup JE. Chemical potential dependence of defect formation energies in GaAs: application to Ga self-diffusion. *Phys. Rev. Lett.* 1991, 67(17):2339–2342.

[81] Li H, Wang Z, Zou N, Ye M, Xu R, *et al*. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* 2022, 2(6):367–377.

[82] Lany S, Zunger A. Anion vacancies as a source of persistent photoconductivity in II-VI and chalcopyrite semiconductors. *Phys. Rev. B* 2005, 72(3):035215.

[83] Hatton P, Uberuaga BP, Perez D. Hop-Decorate: an automated atomistic workflow for generating defect transport data in chemically complex materials. *Comput. Mater. Sci.* 2025, 259:114165.

[84] Vasylenko A, Asher BM, Collins CM, Gaultois MW, *et al*. Inferring energy-composition relationships with Bayesian optimization enhances exploration of inorganic materials. *J. Chem. Phys.* 2024, 160(5):054110.

[85] Wang Y, Lv J, Zhu L, Ma Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B* 2010, 82(9):094116.

[86] Hajibabaei A, Umer M, Anand R, Ha M, Kim KS. Fast atomic structure optimization with on-the-fly sparse Gaussian process potentials. *J. Phys.: Condens. Matter* 2022, 34(34):344007.

[87] Freysoldt C, Neugebauer J, Van de Walle CG. Fully ab initio finite-size corrections for charged-defect supercell calculations. *Phys. Rev. Lett.* 2009, 102(1):016402.

[88] Kumagai Y, Oba F. Electrostatics-based finite-size corrections for first-principles point defect calculations. *Phys. Rev. B* 2014, 89(19):195205.

[89] Tao S, Schmidt I, Brocks G, Jiang J, Tranca I, *et al*. Absolute energy level positions in tin- and lead-based halide perovskites. *Nat. Commun.* 2019, 10(1):2560.

[90] Chandrasekaran A, Kamal D, Batra R, Kim C, Chen L, *et al*. Solving the electronic structure problem with machine learning. *npj Comput. Mater.* 2019, 5(1):22.

[91] Meggiolaro D, De Angelis F. First-principles modeling of defects in lead halide perovskites: best practices and open issues. *ACS Energy Lett.* 2018, 3(9):2206–2222.

[92] Bartel CJ. Review of computational approaches to predict the thermodynamic stability of inorganic solids. *J. Mater. Sci.* 2022, 57(23):10475–10498.

[93] Mosquera-Lois I, Klarbring J, Walsh A. Point defect formation at finite temperatures with machine learning force fields. *Chem. Sci.* 2025, 16(20):8878–8888.
[94] Lim S, Kim TW, Park T, Heo YS, Yang S, *et al*. Large-scale analysis of defects in atomically thin semiconductors using hyperspectral line imaging. *Small* 2024, 20(42):e2400737.
[95] Ong SP, Richards WD, Jain A, Hautier G, Kocher M, *et al*. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* 2013, 68:314–319.
[96] Draxl C, Scheffler M. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* 2019, 2(3):036001.
[97] Artrith N, Behler J. High-dimensional neural network potentials for metal surfaces: a prototype study for copper. *Phys. Rev. B* 2012, 85(4):045439.
[98] Mishin Y. Machine-learning interatomic potentials for materials science. *Acta Mater.* 2021, 214:116980.
[99] Barros de Moraes EA, D'Elia M, Zayernouri M. Machine learning of nonlocal micro-structural defect evolutions in crystalline materials. *Comput. Methods Appl. Mech. Eng.* 2023, 403:115743.
[100] Johnson L, Malone W, Rizk J, Chen R, Gibson T, *et al*. Machine learning method to determine concentrations of structural defects in irradiated materials. *Comput. Mater. Sci.* 2024, 242:113079.
[101] Bartók AP, Kondor R, Csányi G. On representing chemical environments. *Phys. Rev. B* 2013, 87(18):184115.
[102] Grisafi A, Ceriotti M. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.* 2019, 151(20):204105.
[103] Li H, Tang Z, Gong X, Zou N, Duan W, *et al*. Deep-learning electronic-structure calculation of magnetic superstructures. *Nat. Comput. Sci.* 2023, 3(4):321–327.
[104] Ko TW, Finkler JA, Goedecker S, Behler J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* 2021, 12(1):398.
[105] Zhang L, Wang H, Muniz MC, Panagiotopoulos AZ, Car R, *et al*. A deep potential model with long-range electrostatic interactions. *J. Chem. Phys.* 2022, 156(12):124107.
[106] Unke OT, Chmiela S, Gastegger M, Schütt KT, Sauceda HE, *et al*. SpookyNet: learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* 2021, 12(1):7273.
[107] Peterson AA. Acceleration of saddle-point searches with machine learning. *J. Chem. Phys.* 2016, 145(7):074106.
[108] Jeong W, Lee K, Yoo D, Lee D, Han S. Toward reliable and transferable machine learning potentials: uniform training by overcoming sampling bias. *J. Phys. Chem. C* 2018, 122(39):22790–22795.
[109] Willow SY, Hajibabaei A, Ha M, Yang DC, Myung C, *et al*. Sparse Gaussian process based machine learning first principles potentials for materials simulations: application to batteries, solar cells, catalysts, and macromolecular systems. *Chem. Phys. Rev.* 2024, 5(4):041307.
[110] Willow SY, Kim S, Yang DC, Ha M, Hajibabaei A, *et al*. A sparse Bayesian committee machine potential for oxygen-containing organic compounds. *Chem. Phys. Rev.* 2025, 6(2):021401.
[111] Willow SY, Kim DG, Sundheep R, Hajibabaei A, Kim KS, *et al*. Active sparse Bayesian committee machine potential for isothermal-isobaric molecular dynamics simulations. *Phys. Chem. Chem. Phys.* 2024, 26(33):22073–22082.

[112] Smith JS, Nebgen BT, Zubatyuk R, Lubbers N, Devereux C, *et al*. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* 2019, 10(1):2903.

[113] Batatia I, Batzner S, Kovács DP, Musaelian A, Simm GNC, *et al*. The design space of E(3)-equivariant atom-centred interatomic potentials. *Nat. Mach. Intell.* 2025, 7:56–67.

[114] Vondrák M, Reuter K, Margraf JT. Pushing charge equilibration-based machine learning potentials to their limits. *npj Comput. Mater.* 2025, 11(1):288.

[115] Hajibabaei A, Myung CW, Kim KS. Sparse Gaussian process potentials: application to lithium diffusivity in superionic conducting solid electrolytes. *Phys. Rev. B* 2021, 103(21):214102.

[116] Hajibabaei A, Kim KS. Universal machine learning interatomic potentials: surveying solid electrolytes. *J. Phys. Chem. Lett.* 2021, 12(33):8115–8120.

[117] Ha M, Hajibabaei A, Kim DY, Singh AN, Yun J, *et al*. Al-doping driven suppression of capacity and voltage fadings in 4d-element containing Li-ion-battery cathode materials: machine learning and density functional theory. *Adv. Energy Mater.* 2022, 12(30):2201497.

[118] Bian T, Zhu W, Lei Q, Yin J. Machine learning potentials for inorganic and hybrid lead halide perovskites: from phase stability to defects and interfaces. *ACS Appl. Mater. Interfaces* 2026, 18(12):17219–17238.

[119] Tyagi V, Pols M, Brocks G, Tao S. Tracing ion migration in halide perovskites with machine learned force fields. *J. Phys. Chem. Lett.* 2025, 16(20):5153–5159.

[120] Jung J, Ju S, Kim PH, Hong D, Jeong W, *et al*. Electrochemical degradation of Pt_3Co nanoparticles investigated by off-lattice kinetic Monte Carlo simulations with machine-learned potentials. *ACS Catal.* 2023, 13(24):16078–16087.

[121] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science* 1983, 220(4598):671–680.

[122] Bernstein N, Csányi G, Deringer VL. De novo exploration and self-guided learning of potential-energy surfaces. *npj Comput. Mater.* 2019, 5(1):99.

[123] Hessmann SS, Schütt KT, Gebauer NW, Gastegger M, Oguchi T, *et al*. Accelerating crystal structure search through active learning with neural networks for rapid relaxations. *npj Comput. Mater.* 2025, 11(1):44.

[124] Deringer VL, Bernstein N, Csanyi G, Ben Mahmoud C, Csányi M, *et al*. Origins of structural and electronic transitions in disordered silicon. *Nature* 2021, 589(7840):59–64.

[125] Podryabinkin EV, Tikhonov EV, Shapeev AV, Oganov AR. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B* 2019, 99(6):064114.

[126] Gao B, Gao P, Lu S, Lv J, Wang Y, *et al*. Interface structure prediction via CALYPSO method. *Sci. Bull.* 2019, 64(5):301–309.

[127] Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* 2016, 111:218–230.

[128] Mathew K, Montoya JH, Faghaninia A, Dwarakanath S, Aykol M, *et al*. Atomate: a high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* 2017, 139:140–152.

[129] Byggmästar J, Nordlund K, Djurabekova F. Gaussian approximation potentials for body-centered-cubic transition metals. *Phys. Rev. Mater.* 2020, 4(9):093802.
[130] Diao X, Diao Y, Tang Y, Zhao G, Gu Q, *et al*. High-throughput screening of stable and efficient double inorganic halide perovskite materials by DFT. *Sci. Rep.* 2022, 12(1):12633.
[131] Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput. Mater.* 2019, 5(1):21.
[132] Bassman Oftelie L, Rajak P, Kalia RK, Nakano A, Sha F, *et al*. Active learning for accelerated design of layered materials. *npj Comput. Mater.* 2018, 4(1):74.
[133] Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* 2019, 5(1):83.
[134] Zhang L, Lin D, Wang H, Car R, E W. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* 2019, 3(2):023804.
[135] Zhang Y, Wang H, Chen W, Zeng J, Zhang L, *et al*. DP-GEN: a concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput. Phys. Commun.* 2020, 253:107206.
[136] Bilbrey JA, Firoz JS, Lee MS, Choudhury S. Uncertainty quantification for neural network potential foundation models. *npj Comput. Mater.* 2025, 11(1):109.
[137] Chen LY, Li YP. Uncertainty quantification with graph neural networks for efficient molecular design. *Nat. Commun.* 2025, 16(1):3262.
[138] Imbalzano G, Zhuang Y, Kapil V, Rossi K, Engel EA, *et al*. Uncertainty estimation for molecular dynamics and sampling. *J. Chem. Phys.* 2021, 154(7):074102.
[139] Ocampo D, Posso D, Namakian R, Gao W. Adaptive loss weighting for machine learning interatomic potentials. *Comput. Mater. Sci.* 2024, 244:113155.
[140] Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, *et al*. Scaling deep learning for materials discovery. *Nature* 2023, 624(7990):80–85.
[141] Vandermause J, Torrisi SB, Batzner S, Xie Y, Sun L, *et al*. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Comput. Mater.* 2020, 6(1):20.
[142] Smith JS, Zubatyuk R, Nebgen B, Lubbers N, Barros K, *et al*. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* 2020, 7(1):134.
[143] Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* 2018, 1(9):696–703.
[144] Li Z, Kermode JR, De Vita A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* 2015, 114(9):096405.
[145] Cheng M, Fu C, Okabe R, Chotrattanapituk A, Boonkird A, *et al*. Artificial intelligence-driven approaches for materials design and discovery. *Nat. Mater.* 2026, 25(2):174–190.
[146]Deringer VL, Csányi G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* 2017, 95(9):094203.
[147] Wen T, Wang R, Zhu L, Zhang L, Wang H, *et al*. Specialising neural network potentials for accurate properties and application to the mechanical response of titanium. *npj Comput. Mater.* 2021, 7(1):206.
[148] Kapil V, Rossi M, Marsalek O, Petraglia R, Litman Y, *et al*. i-PI 2.0: a universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* 2019, 236:214–223.

[149] Cheng B, Engel EA, Behler J, Dellago C, Ceriotti M. Ab initio thermodynamics of liquid and solid water. *Proc. Natl. Acad. Sci. U.S.A.* 2019, 116(4):1110–1115.
[150] Tewary VK. Extending the time scale in molecular dynamics simulations: propagation of ripples in graphene. *Phys. Rev. B* 2009, 80(16):161409.
[151] Sørensen MR, Voter AF. Temperature-accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.* 2000, 112(21):9599–9606.
[152] Koistinen OP, Dagbjartsdottir FB, Ásgeirsson V, Vehtari A, Jónsson H. Nudged elastic band calculations accelerated with Gaussian process regression. *J. Chem. Phys.* 2017, 147(15):152720.
[153] Jinnouchi R, Miwa K, Karsai F, Kresse G, Asahi R. On-the-fly active learning of interatomic potentials for large-scale atomistic simulations. *J. Phys. Chem. Lett.* 2020, 11(17):6946–6955.
[154] Ravipati S, Savva GD, Christidi IA, Guichard R, Nielsen J, *et al*. Coupling the time-warp algorithm with the graph-theoretical kinetic Monte Carlo framework for distributed simulations of heterogeneous catalysts. *Comput. Phys. Commun.* 2022, 270:108148.
[155] Kubin LP, Canova G, Condat M, Devincre B, Pontikis V, *et al*. Dislocation microstructures and plastic flow: a 3D simulation. *Solid State Phenom.* 1992, 23–24:455–472.
[156] Arsenlis A, Cai W, Tang M, Rhee MS, Oppelstrup T, *et al*. Enabling strain hardening simulations with dislocation dynamics. *Model. Simul. Mater. Sci. Eng.* 2007, 15(6):553–595.
[157] Szlachta WJ, Bartók AP, Csányi G. Accuracy and transferability of Gaussian approximation potential models for tungsten. *Phys. Rev. B* 2014, 90(10):104108.
[158] Dragoni D, Daff TD, Csányi G, Marzari N. Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron. *Phys. Rev. Mater.* 2018, 2(1):013808.
[159] Bertin N, Zhou F. Accelerating discrete dislocation dynamics simulations with graph neural networks. *J. Comput. Phys.* 2023, 487:112180.
[160] Gaikwad S, Kasilingam T, Ahmad O, Mukherjee R, Bhowmick S. Deep learning-driven prediction of microstructure evolution via latent space interpolation. *Phys. Rev. Mater.* 2025, 9(10):103804.
[161] Abueidda DW, Koric S, Sobh NA, Sehitoglu H. Deep learning for plasticity and thermo-viscoplasticity. *Int. J. Plast.* 2021, 136:102852.
[162] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 2019, 378:686–707.
[163] Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* 2021, 3(3):218–229.
[164] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, *et al*. Physics-informed machine learning. *Nat. Rev. Phys.* 2021, 3(6):422–440.
[165] Wang S, Wang H, Perdikaris P. Learning the solution operator of parametric partial differential equations with physics-informed DeepONet. *Sci. Adv.* 2021, 7(40):eabi8605.
[166] Ma L, Liu H, Williams M, Levine L, Ramazani A. Integrating phase field modeling and machine learning to develop process-microstructure relationships in laser powder bed fusion of IN718. *Metallogr. Microstruct. Anal.* 2024, 13(5):983–995.

[167] Lian H, Zhao P, Zhang M, Wang P, Li Y. Physics informed neural networks for phase field fracture modeling enhanced by length-scale decoupling degradation functions. *Front. Phys.* 2023, 11:1258661.
[168] Kharazmi E, Zhang Z, Karniadakis GEM. hp-VPINNs: variational physics-informed neural networks with domain decomposition. *Comput. Methods Appl. Mech. Eng.* 2021, 374(1):113547.
[169] Capone M, Romanelli M, Castaldo D, Parolin G, Bello A, *et al*. A vision for the future of multiscale modeling. *ACS Phys. Chem. Au.* 2024, 4(3):202–225.
[170] Perez D, Subramanyam APA, Maliyov I, Swinburne TD. Uncertainty quantification for misspecified machine learned interatomic potentials. *npj Comput. Mater.* 2025, 11(1):263.
[171] Zuo Y, Chen C, Li X, Deng Z, Chen Y, *et al*. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* 2020, 124(4):731–745.
[172] Mahadule D, Demiral M, Mulki H, Khatirkar RK. Experiments and crystal plasticity finite element simulations of texture development during cold rolling in a Ti-15V-3Cr-3Sn-3Al alloy. *Crystals* 2023, 13(1):137.
[173] Stach E, DeCost B, Kusne AG, Hattrick-Simpers J, Brown KA, *et al*. Autonomous experimentation systems for materials development: a community perspective. *Matter* 2021, 4(9):2702–2726.
[174] Vergara D, del Bosque A, Fernández-Arias P. A decade of digital twins in materials science and engineering. *Comput. Mater. Contin.* 2025, 85(1):41–64.
[175] Wang G, Wang C, Zhang X, Li Z, Zhou J, *et al*. Machine learning interatomic potential: bridge the gap between small-scale models and realistic device-scale simulations. *iScience* 2024, 27(5):109673.
[176] Wang G, Sun Y, Zhou J, Sun Z. PotentialMind: graph convolutional machine learning potential for Sb-Te binary compounds of multiple stoichiometries. *J. Phys. Chem. C* 2023, 127(51):24724–24733.
[177] Riebesell J, Goodall RE, Benner P, Chiang Y, Deng B, *et al*. A framework to evaluate machine learning crystal stability predictions. *Nat. Mach. Intell.* 2025, 7(6):836–847.
[178] Li K, Rubungo AN, Lei X, Persaud D, Choudhary K, *et al*. Probing out-of-distribution generalization in machine learning for materials. *Commun. Mater.* 2025, 6(1):9.
[179] Omee SS, Fu N, Dong R, Hu M, Hu J. Structure-based out-of-distribution (OOD) materials property prediction: a benchmark study. *npj Comput. Mater.* 2024, 10:144.
[180] Bertoldo F, Ali S, Manti S, Thygesen KS. Quantum point defects in 2D materials-the QPOD database. *npj Comput. Mater.* 2022, 8(1):56.
[181] Yang Z, Liu X, Zhang X, Huang P, Novoselov KS, *et al*. Modeling crystal defects using defect informed neural networks. *npj Comput. Mater.* 2025, 11(1):229.
[182] Deng B, Zhong P, Jun K, Riebesell J, Han K, *et al*. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* 2023, 5(9):1031–1041.
[183] Unke OT, Meuwly M. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* 2019, 15(6):3678–3693.
[184] Willard J, Jia X, Xu S, Steinbach M, Kumar V. Integrating scientific knowledge with machine learning. *ACM Comput. Surv.* 2022, 55(4):1–37.
[185] Rahman MH, Gollapalli P, Manganaris P, Yadav SK, Pilania G, *et al*. Accelerating defect predictions in semiconductors using graph neural networks. *APL Mach. Learn.* 2024, 2(1):016122.

[186] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, *et al*. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 2016, 3:160018.

[187] Tanifuji M, Matsuda A, Yoshikawa H. Materials data platform—a FAIR system for data-driven materials science. In *Proceedings of the 2019 8th International Congress on Advanced Applied Informatics (IIAI–AAI)*, Toyama, Japan, July 7–12, 2019, pp. 1021–1022.

[188] Andersen CW, Armiento R, Blokhin E, Conduit GJ, Dwaraknath S, *et al*. OPTIMADE, an API for exchanging materials data. *Sci. Data* 2021, 8(1):217.

[189] Himanen L, Geurts A, Foster AS, Rinke P. Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* 2019, 6(21):1900808.

[190] Dai Z, Yang Z, Ni M, Huang M, Xiang H, *et al*. Machine learning Hamiltonian enables scalable and accurate defect calculations: the case of oxygen vacancies in amorphous SiO_2. *npj Comput. Mater.* 2026.

[191] Li X, Guo X, Huang M, Gong X, Chen S. Machine learning interatomic potentials accelerate defect exploration in amorphous silica. *Phys. Rev. Mater.* 2026, 10:015003.

[192] Polak MP, Morgan D. Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nat. Commun.* 2024, 15(1):1569.

[193] Dagdelen J, Dunn A, Lee S, Walker N, Rosen AS, *et al*. Structured information extraction from scientific text with large language models. *Nat. Commun.* 2024, 15(1):1418.

[194] Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. *Nature* 2023, 624(7992):570–578.

[195] Jablonka KM, Ai Q, Al-Feghali A, Badhwar S, Bocarsly JD, *et al*. 14 examples of how LLMs can transform materials science and chemistry: a reflection on a large language model hackathon. *Digital Discovery* 2023, 2(5):1233–1250.

[196] Hanser T. Federated learning for molecular discovery. *Curr. Opin. Struct. Biol.* 2023, 79:102545.