Computational Materials Science 273 (2026) 114900

Contents lists available at ScienceDirect

Computational Materials Science

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

Full length article

Software interoperability in electronic structure methods for atomistic materials simulations

Jesús Carrete [a], Pablo García-Fernández [b], Myrta Grüning [c], Xu He [d], Mike N. Pionteck [e], Nakib H. Protik [f], Milica Todorović [g], Atsushi Togo [i,h], Xing Wang [j], Roberta Poloni [k,*], Miguel Pruneda [l,m,*], Antonio Cammarata [n,*]

[a] Instituto de Nanociencia y Materiales de Aragón (INMA), CSIC-Universidad de Zaragoza, Zaragoza, 50009, Spain
[b] Facultad de Ciencias, Universidad de Cantabria, Avenida de los Castros s/n, Santander, 39005, Cantabria, Spain
[c] School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, United Kingdom
[d] Faculté des Sciences, Département de physique, Physique théorique des matériaux, Quantum Materials (Q-MAT), University of Liège, Liège, Belgium
[e] Institut für Theoretische Physik and Center for Materials Research (LaMa), Justus-Liebig-Universität Gießen, Gießen, 35392, Germany
[f] Institut für Physik and Center for the Science of Materials Berlin (CSMB), Humboldt-Universität zu Berlin, Berlin, 12489, Germany
[g] Department of Mechanical and Materials Engineering, University of Turku, Turku, 20014, Finland
[h] National Institute for Materials Science, Tsukuba, Japan
[i] Center for Elements Strategy Initiative for Structural Materials, Kyoto University, Kyoto, Japan
[j] Paul Scherrer Institute (PSI), Forschungsstrasse 111, Villigen PSI, 5232, Switzerland
[k] Univ. Grenoble Alpes, CNRS, Grenoble-INP, SIMaP, Grenoble, 38000, France
[l] Nanomaterials and Nanotechnology Research Center (CINN-CSIC), Universidad de Oviedo (UO), Principado de Asturias (PA), El Entrego, 33940, Spain
[m] Catalan Institute of Nanoscience and Nanotechnology (ICN2), BIST & CSIC, UAB Campus, Bellaterra, 08193, Spain
[n] Department of Control Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, Prague, 16627, Czech Republic

ARTICLE INFO

Keywords:
Material discovery
Ab initio
Automation
Machine-learning
Database
High-throughput

ABSTRACT

Atomistic materials simulations increasingly rely on complex, multi-stage workflows that combine electronic structure first principles methods, effective models, machine-learning techniques, and high-performance computing, making software interoperability a central requirement rather than a technical convenience. This perspective highlights how modular, interoperable software ecosystems overcome the limitations of monolithic electronic structure codes, enabling automation, reproducibility, and scalability across diverse applications such as lattice dynamics, electronic, magnetic, optical, and transport properties, as well as data-driven materials optimization. We discuss how standardized workflows, robust data provenance, and interoperable infrastructures are essential for high-throughput studies, AI-ready databases, and the transition toward autonomous materials discovery. This is the result of the debates of the Symposium F "Advanced interoperability in atomistic simulations of materials", which took place within the 2025 European Materials Research Society fall meeting.

1. Introduction

Atomistic simulations have become a central pillar of modern materials science, enabling predictive investigations of structural, electronic, vibrational, magnetic, optical, and transport properties across a wide range of materials systems. As these simulations increasingly target complex phenomena — such as anharmonic lattice dynamics, coupled electron–phonon and spin interactions, nonlinear optical responses, and large-scale phase-space sampling — the associated computational workflows have grown in both scale and sophistication. Today's state-of-the-art studies rarely rely on a single simulation code; instead, they depend on tightly orchestrated combinations of first principles engines, effective Hamiltonians, machine-learning models, workflow managers, and post-processing tools, often deployed on large-scale high-performance computing infrastructures. In this context, software interoperability is no longer optional but has become a fundamental requirement for progress, providing the structured backbone that emerging AI agents increasingly exploit for autonomous workflow orchestration.

Multi-agent AI systems represent an exciting frontier for atomistic simulations, with recent advances demonstrating autonomous orchestration of complex workflows, from structure generation and force-field selection to simulation execution and analysis, in domains such as

* Corresponding authors.
E-mail addresses: roberta.poloni@grenoble-inp.fr (R. Poloni), miguel.pruneda@icn2.cat (M. Pruneda), cammaant@fel.cvut.cz (A. Cammarata).

https://doi.org/10.1016/j.commatsci.2026.114900
Received 27 March 2026; Received in revised form 29 May 2026; Accepted 28 June 2026
Available online 9 July 2026

molecular dynamics and materials screening [1,2]. While these agentic frameworks have matured rapidly in force-field–based subfields, where standardized tasks and cheaper computations enable end-to-end automation, their application to electronic-structure methods remains earlier-stage [3,4], largely focused on workflow setup, error recovery, and hybrid integration with DFT rather than fully replacing human-designed protocols. This Perspective argues that interoperability should be grounded in shared, machine-readable schemas, provenance layers, and tool interfaces, which provide the essential foundation for workflows that remain accessible to researchers. Such a foundation is especially important as new theories, methods, and computed properties continue to evolve rapidly and still require active human oversight, rather than fully automated environments in which agents negotiate interfaces at runtime.

This has been the framework of the Symposium F "Advanced interoperability in atomistic simulations of materials", which took place within the 2025 European Materials Research Society fall meeting. The outcomes of the Symposium's debates are examined in this perspective, where we discuss the software ecosystem in atomistic simulations of materials as a unifying concept that enables modularity, automation, reproducibility, and scalability. We highlight how the traditional limitations of monolithic simulation codes can be overcome by interoperable frameworks that allow specialized tools to exchange data seamlessly and to be embedded within automated workflows.

Through representative examples drawn from across the electronic structure simulation landscape, this Perspective illustrates how interoperability supports advanced simulations at multiple levels. First, the role of interoperable workflow infrastructures in enabling complex multi-method simulations is discussed (Section 2). Section 3 focuses on electronic structure workflows, ranging from real-time simulations of nonlinear optical responses (Section 3.1), to parametrization of effective spin Hamiltonians for magnetic simulations (Section 3.2), and second-principles effective Hamiltonians derived from first principles (Section 3.3). Section 4 addresses thermal transport and lattice-dynamical properties, including coupled electron–phonon Boltzmann transport (Section 4.1) and recent machine-learning-accelerated approaches (Section 4.2). Section 5 discusses active learning strategies for data-efficient materials discovery and their integration within interoperable computational ecosystems. Finally, Section 6 reviews advances in molecular dynamics and phase-space sampling enabled by multiscale and adaptive frameworks.

A central theme of this work is the infrastructure required to support such interoperability, including standardized and extensible data formats, robust provenance tracking, inter-code communication layers, and workflow engines capable of managing complex, adaptive execution logic. We discuss how Python-based interfaces and modular software components facilitate extensibility and community-driven development, while high-throughput workflow frameworks and materials databases provide the backbone for systematic data generation and sharing. Particular attention is given to the role of interoperability in enabling FAIR (Findable, Accessible, Interoperable, Reusable), AI-ready data, where semantic consistency and traceability are essential for trustworthy machine learning and autonomous discovery pipelines.

By bringing together diverse methodological advances under a common connectivity perspective, this paper argues that the future of atomistic materials simulation lies in integrated software ecosystems rather than isolated codes. Such ecosystems not only accelerate the incorporation of new theoretical developments but also enable large scale, reproducible, and AI-driven exploration of materials. These developments provide key enabling infrastructure for future closed-loop discovery workflows, but their present impact is primarily in strengthening human-supervised, multi-code simulation pipelines.

2. Managing the multi-method realm of materials simulations

The increasing complexity of atomistic simulations, which often require chaining structure relaxation, sampling, and post-processing, has necessitated the shift from manual execution to automated workflows. These workflows are the engine behind high-throughput screening [5], granting researchers the ability to calculate material properties across vast chemical spaces with speed and consistency. To support this scale, the community has developed a rich landscape of workflow frameworks, including `AiiDA` [6,7], `Jobflow` [8], and `pyiron` [9]. Crucially, the massive amount of structured data generated by these high-throughput calculations serves as the foundation for modern databases [10–12], which are the prerequisite for the training of machine-learning (ML) models and AI-driven discovery. However, the rapid evolution of these distinct software ecosystems has created a fragmentation challenge: workflows and data generated in one framework are often incompatible with others. To sustain the momentum of AI-driven materials science, interoperability across four complementary layers is needed: the scientific protocol, the data provenance layer, the execution engine, and the user application layer (Fig. 1).

Layer 1: Interoperability of scientific protocols

Historically, comparing results across different Density Functional Theory (DFT) codes required laborious manual tuning of numerical parameters. To resolve this, the community has developed `Common Workflows` [13], which provide uniform interfaces for computing material properties (e.g., equations of state) regardless of the underlying quantum engine. By standardizing inputs and outputs, these interfaces treat the simulation code — whether `Quantum ESPRESSO` [14], `SIESTA` [15], or `CP2K` [16] — as a swappable component. This interoperability was the cornerstone of the recent massive verification of DFT precision [17], where identical protocols were executed across 11 different codes. This effort proved that diverse implementations can yield consistent results when orchestrated within a common workflow framework. Prioritizing these "abstract" workflow definitions over code-specific scripts ensures true scientific reproducibility and cross-code validation [18].

Layer 2: AI-ready data provenance and semantics

While standardizing execution is challenging, ensuring that the output of these simulations is machine readable is an immediate and high-impact priority. To fuel Artificial Intelligence (AI) models with reliable outcomes, data must be traceable, reproducible, and FAIR (Findable, Accessible, Interoperable, Reusable) [19]; otherwise, even sophisticated models propagate hidden errors. Provenance systems currently excel at linear multi-code execution but face challenges in highly iterative scenarios requiring adaptive path tracking. The `AiiDA` infrastructure addresses the traceability requirement by automatically storing the full data provenance in the database. This captures every input, output, and logical step of a calculation, ensuring that any data point used in machine learning can be audited and reproduced. Simultaneously, syntactic interoperability has been addressed by the `OPTIMADE API` [20,21], which allows uniform querying across databases like the `Materials Project` [10], `OQMD` [11], and `Materials Cloud` [12]. However, neither structured provenance nor syntactic uniformity is sufficient for full AI-readiness if semantics are missing. Currently, adding physical meaning often requires laborious *post hoc* curation. Workflow engines must desirably evolve to support ontology-aware semantics at runtime. Although several ontology- and metadata-centered infrastructures now wrap electronic-structure calculations, we are not aware of any widely used production DFT code that natively exposes a mature, runtime ontology-aware semantic layer; in practice, semantic information is still usually provided by external workflow systems, knowledge-graph layers, or domain experts

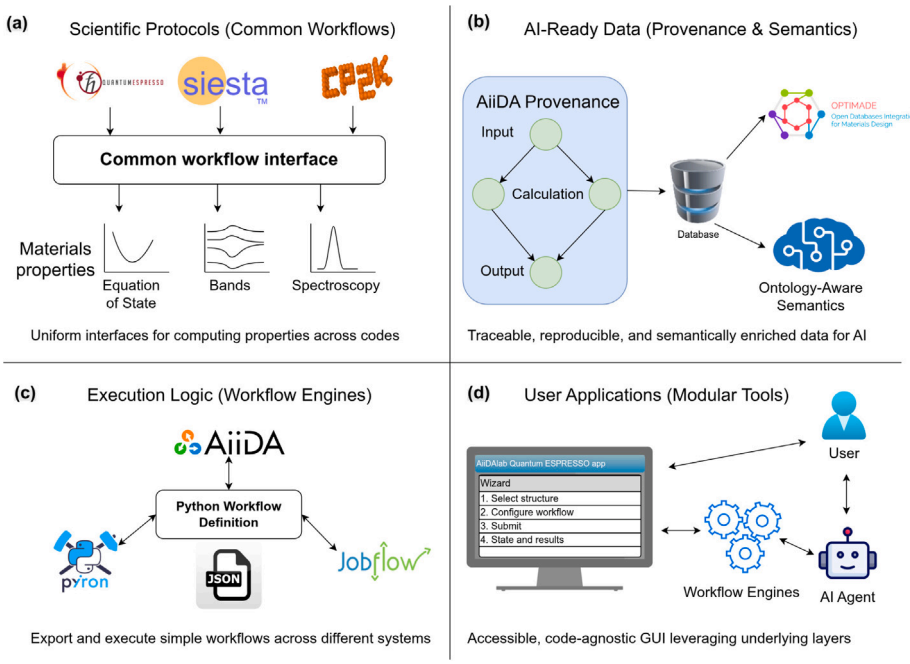

Fig. 1. The hierarchy of software interoperability. (a) Scientific protocols are standardized via `Common Workflows`. (b) Data is made AI-ready through automated provenance tracking and semantic integration. (c) Execution logic is shared via exchange formats like the `Python Workflow Definition` (PWD). (d) Users interact with these layers through modular applications like the `AiiDAlab Quantum ESPRESSO` app.

rather than by the solver itself. By embedding definitions from formal ontologies directly into the workflow logic, engines can generate rich, semantically granular data as it is created. This ensures that a floating-point value is not just a number, but is instantly tagged — for example, as a "potential energy" in "eV" relative to a specific ontology schema — allowing AI agents to autonomously interpret data without human intervention. Efforts toward shared ontologies are already underway in the `OPTIMADE` consortium (see the evolving OPTIMADE specification), through the MADICES conference series (https://madices.github.io/), or the European Materials Modelling Council (see https://github.com/emmo-repo/EMMO).

Layer 3: Interoperability of workflow engines

One aspirational "ideal" is the interoperability of the workflow engines themselves. As noted, the ecosystem is rich with systems like `AiiDA` [6,7], `Jobflow` [8], `FireWorks` [22], `Atomate` [23], and `pyiron` [9]. Currently, these systems operate as silos. Recent efforts, such as the `Python Workflow Definition` (PWD) [24], attempt to address this by defining a format to specify software dependencies, node definitions, and graph topology independent of the workflow management system. This allows simple Directed Acyclic Graph (DAG) workflows to be exported from one system (e.g., `AiiDA`) and executed by another (e.g., `pyiron`). However, achieving this for realistic, dynamic research applications — which often require complex flow control, error handling, and while-loops — remains a significant technical challenge. While valuable for simple demonstrations, widespread engine interoperability likely requires further maturation of these intermediate standards.

Layer 4: Interoperability of applications (tools)

Finally, these technical and scientific layers should be made accessible to non-specialists through interoperable graphical user interfaces (GUIs), whether installed locally or accessed via the cloud. The recent development of the `AiiDAlab Quantum ESPRESSO` app [25,26] demonstrates the efficacy of wrapping robust `AiiDA` workflows in a user-friendly "wizard" interface. By organizing the app around modular panels for inputs, processing steps, and results, various property plugins — ranging from electronic structure to spectroscopies — can share a unified interface and workflow logic. The future of GUIs lies in decoupling the frontend from the backend. By building "code-agnostic" interfaces that leverage the `Common Workflows` (Layer 1) and Provenance (Layer 2) described above, it is possible to create tools where a user selects a material and a property, and an assistant — potentially an LLM-driven agent — suggests a suitable code and workflow, lowering the barrier to advanced simulation methods. It is possible to imagine a research ecosystem where interoperability is built-in by design: a researcher should be able to define a simulation using a standard `Common Workflow`, execute it on any Workflow Engine via intermediate definitions like PWD, analyze it through a code-agnostic App, and automatically persist the results in an AI-ready database enriched with semantic ontologies. This vertical stack of interoperability is the key to transitioning from isolated calculations to autonomous materials discovery, enabling AI agents to seamlessly navigate the full cycle of design, simulation, and analysis.

3. Integrating electronic structure methods across scales

3.1. A toolkit for optical properties

Nonlinear optical spectroscopy is a powerful tool to probe (perturbative regime) or even control (non-perturbative regime) material properties. First principles simulations of the nonlinear optical response of materials can help the design and interpretation of nonlinear spectroscopy experiments. With this goal, the nonlinear optical run-level of the `yambo` code [27,28] implements the time-evolution of Bloch electrons in a homogeneous time-dependent electric field [29]. The workflow is shown in Fig. 2. The main ingredients in the equation of motion for the Bloch electrons $\{|v_{\mathbf{k}}, m\rangle\}$ are the external electric field $\mathcal{E}$ and the effective Hamiltonian $\hat{H}_{\mathbf{k}}$. The effective Hamiltonian is built from the Kohn–Sham Hamiltonian obtained from prior electronic structure calculations. Beyond the independent-particle approximation (IPA), quasiparticle corrections and the so-called collision matrix elements [30] — needed to calculate the self-energy at each time-step —

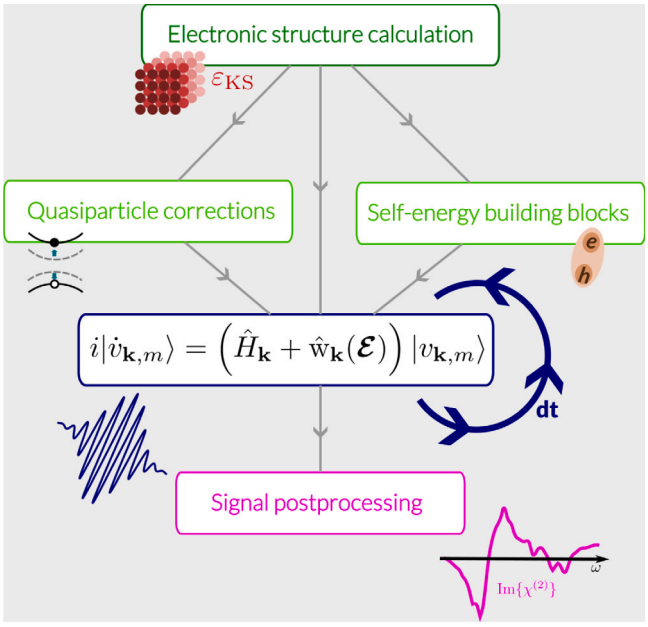

Fig. 2. Interoperability. In this workflow each box represents a separate run, executed by different codes/code run-levels. In the first step, the electronic structure is calculated. This is the input for the real-time simulation (blue box) and — in case of calculations beyond the independent particles — for the calculation of quasiparticle correction and the building-blocks for the self-energy operator, that are also input of the real-time simulation. The output from the real-time simulation is postprocessed to obtain, e.g. nonlinear spectral properties.

are added to the effective Hamiltonian. The integration of the equation of motion outputs the macroscopic polarization (and current), which are then post-processed to obtain the nonlinear responses. Obtaining the nonlinear optical response for a given material requires interoperability of different software: the nonlinear optical run-level integrates the equation of motion, the electronic structure is obtained from the plane-wave density functional code `Quantum ESPRESSO` [14], the quasiparticle corrections and collision matrix elements are computed by other run-levels of the `yambo` code [31], and `yambopy` is used for the postprocessing [32] The latter is a suite of Python scripts that includes access to the `yambo` databases and the analysis of the nonlinear response. The Python programming language makes it easy for users to extend the analysis tools depending on the experiments and properties they wish to address. As an example, in a recent work [33], nonlinear optical spectroscopy based on frequency-mixing has been addressed, and specifically the sum and difference frequency generation (SFG/DFG) (see top panels of Fig. 3). In SFG/DFG processes, two input fields at ω_1 and ω_2 generate an output at $\omega = \omega_1 \pm \omega_2$ providing direct access to the second-order response function $\chi^{(2)}(\omega;\omega_1,\omega_2)$. Before this work, only the analysis of a single monochromatic field was available, and nonlinear spectroscopy techniques treated were restricted to a single probe. By adding a few Python functions to `yambopy`, a new analysis tool was created and shared with other users that allows one to simulate frequency-mixing nonlinear optical spectroscopy. Specifically, robust signal-processing techniques has been employed, such as singular value decomposition (SVD) pseudoinversion and least-squares [33] rather than the discrete Fourier transform, which is impractical because commensurate frequencies lead to extremely long fundamental periods. Since the rest of the workflow could be used as it was (it was already possible to specify more than one monochromatic field), it is possible to study SFG/DFG incorporating *crucial* quasiparticle corrections and excitonic interactions within the so-called time-dependent adiabatic GW (TD-aGW) [30,34]. Results are shown in the bottom panels of Fig. 3 for monolayer *h*-BN. This work [33] — which also includes simulations on SFG/DFG of MoS_2 monolayer and terahertz-field-induced second-harmonic generation for a bilayer of *h*-BN — demonstrates how real-time simulations with TD-aGW capture excitonic effects in nonlinear frequency-mixing. Overall, thanks to the integration of different codes and run-levels, and the flexibility offered by the Python-based postprocessing, it has been possible to address several nonlinear properties beyond the original second- and third-harmonic generation and besides the one in the example. For instance, two-photon absorption [35] and, more recently, the shift-current [36] and the electric-field-induced second harmonic generation, including the ionic contribution [37]. Finally, being the workflow based on density functional theory, the nonlinear optical properties of any material can, in principle, be calculated. Several future directions appear promising: First, expanding the interoperability of the real-time framework is a natural next step. The structure of the workflow already allows different levels of theory to be plugged in (see Fig. 2), and extending the interoperability. For instance, the integration of additional run-levels, such as electron–phonon coupling or alternative screened interaction models would enable to treat multiple quantum mechanical effects within a single real-time workflow. A related improvement is the adoption of a fully Python-based interface for defining external electric fields, which would make the framework more accessible. Second, generalizing the present implementation to sequences of multiple pulses would gain access to a wide class of nonlinear probes, such as ultrafast pump–probe and time-resolved nonlinear spectroscopy. Third, model reduction and data-driven techniques [38] may dramatically accelerate TD-aGW propagation, enabling large-scale materials screening. Using model order reduction, low-rank approximations, or machine-learning surrogate models for the exchange kernel or the dynamical polarization could make high-resolution frequency grids feasible. Finally, the workflow can be integrated with materials informatics to generate systematic databases of nonlinear optical responses in 2D materials. For example, the present workflow could be made compatible with workflow managers such as `AiiDA` [39], which can manage high-throughput execution and store nonlinear response functions in standardized formats. Beyond the specific application to nonlinear spectroscopy, the underlying framework could be used for general signal analysis. Methods such as Fourier transform, SVD pseudoinvers, or least-squares fitting to analyze the nonlinear polarization do not rely on any assumption about the physical origin of the signal. With minor adaptations of the I/O interface, the same analysis tools can be applied to any multi-component time-domain signal, e.g., for spin and magnetization dynamics, charge or current oscillations, and other time-dependent collective modes obtained from real-time or molecular dynamics simulations.

This example maps naturally onto the interoperability stack described in Section 2, combining a well-defined protocol, data exchange between `Quantum ESPRESSO` and `yambo`, and Python-based postprocessing through `yambopy`. It also illustrates the broader challenge discussed above: while technical data exchange is well established, explicit semantic annotation of the exchanged quantities remains an important direction for future development.

3.2. Modeling magnetism at the nanoscale

The Heisenberg spin model is a cornerstone of research in magnetism, essential for simulating diverse phenomena ranging from ground-state ordering and thermodynamics to ultrafast spin dynamics across extended length and time scales. The generalized Heisenberg Hamiltonian captures the essential magnetic interactions, including isotropic exchange, anisotropic exchange, the Dzyaloshinskii-Moriya interaction (DMI), and single-ion anisotropy (SIA):

$$E = -\sum_i K_i(\vec{S}_i \cdot \vec{e}_i)^2 - \sum_{i\neq j}\left[J_{ij}^{iso}\vec{S}_i \cdot \vec{S}_j + \vec{S}_i\mathbf{J}_{ij}^{ani}\vec{S}_j + \vec{D}_{ij} \cdot (\vec{S}_i \times \vec{S}_j)\right]$$

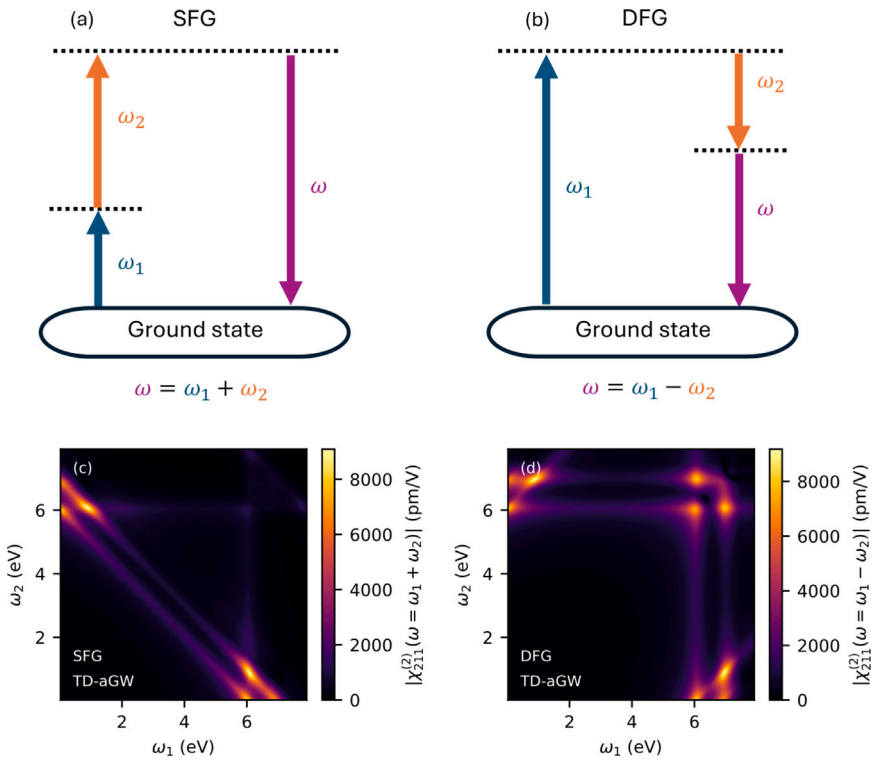

Fig. 3. Application. A schematic overview of (a) sum frequency generation (SFG) and (b) difference frequency generation (DFG). Calculated (c) SFG and (d) DFG spectra for *h*-BN at the TD-aGW level. These heatmaps have been generated using a frequency grid of $\omega_1 \times \omega_2 = 96 \times 96$ points. For each frequency pair, a real-time simulation was run, and the output signal processed.

Here, $\vec{S}_i$ denotes the unit spin vector at site i. The Hamiltonian accounts for the single-ion anisotropy K_i along the easy axis $\vec{e}_i$, as well as pairwise interactions: the isotropic exchange J_{ij}^{iso}, the symmetric anisotropic exchange tensor $\mathbf{J}_{ij}^{ani}$, and the antisymmetric Dzyaloshinskii-Moriya interaction vector $\vec{D}_{ij}$.

While the form of this Hamiltonian appears simple, accurately determining these interaction parameters for real materials is a critical and challenging task. The precise computation from first principles is essential for predictive simulations. There are several widely adopted approaches to compute them from DFT:

1. **Total energy mapping**: Calculating total energies of various collinear and non-collinear magnetic configurations (often using supercells) and fitting the parameters to the energy differences [40].
2. **Generalized Bloch Theorem (GBT)**: Using spin-spiral boundary conditions to calculate the energy of spin spirals and extracting parameters from the energy dispersion [41].
3. **Density Functional Perturbation Theory (DFPT)**: Computing the magnetic response functions by perturbing the density functional theory (DFT) electronic structure with respect to magnetic fields [42–46].
4. **Magnetic Force Theorem (MFT)**: Treating local spin rotations as perturbations to the electronic structure of a reference state (usually the ground state), in the framework of Green's function method, allowing for the calculation of all parameters from a single DFT calculation without supercells [47].

TB2J [48] is an implementation of the MFT within the Green's function formalism for computing magnetic interaction parameters. The central idea is to map the energy variation due to small rigid spin rotations in the DFT electronic structure to the classical Heisenberg spin model. According to the MFT, the change in total energy due to such a perturbation can be approximated by the change in the sum of single-particle energies, assuming the charge density remains fixed. This approach avoids the need for computationally expensive total energy calculations of multiple magnetic configurations and the associated convergence challenges. Instead, the exchange interactions are computed directly from the Green's functions of the magnetic ground state.

By treating the local spin rotation as a perturbation, the full set of magnetic interaction parameters, including isotropic exchange, anisotropic exchange, and Dzyaloshinskii–Moriya interaction (DMI), can be derived. The latter two quantities require a full-relativistic treatment in the DFT calculation [49]. To represent the local spin rotation, the method requires the Kohn–Sham Hamiltonian to be represented in an atomic-centered local basis set. TB2J supports Hamiltonians in the form of Wannier functions [50] or numerical atomic orbitals [51], and recently, gaussian orbitals. This flexibility allows the method to be applied to many DFT codes. A major advantage of this Green's function approach is its efficiency: it allows for the computation of magnetic interactions between atoms up to long distances from a single unit cell SCF calculation, without the need to construct large supercells.

TB2J is designed as a modular component of a larger ecosystem, ensuring interoperability and ease of integration into high-throughput workflows (Fig. 4). Central to this design is the `HamiltonIO` package, which establishes a standard interface for parsing electron Hamiltonians. By abstracting the details of specific file formats, `HamiltonIO` facilitates seamless interoperability between TB2J (or other code requiring electron Hamiltonians) and a wide range of DFT codes. This unified framework simplifies the addition of support for new DFT codes and promotes code reuse. Currently, `HamiltonIO` supports codes using the `Wannier90` interface to the many DFT codes, for example `ABINIT` [52], `QuantumESPRESSO` [53], `VASP` [54], `GPAW` [55], local atomic-orbital bases (`SIESTA` [15,56] via `sisl` [57], `OpenMX` [58], `ABACUS` [59]), and Gaussian-type orbitals (`Crystal` [60]). Additionally, it includes electron–phonon coupling parameters from `EPW` [61] for computing spin-lattice coupling.

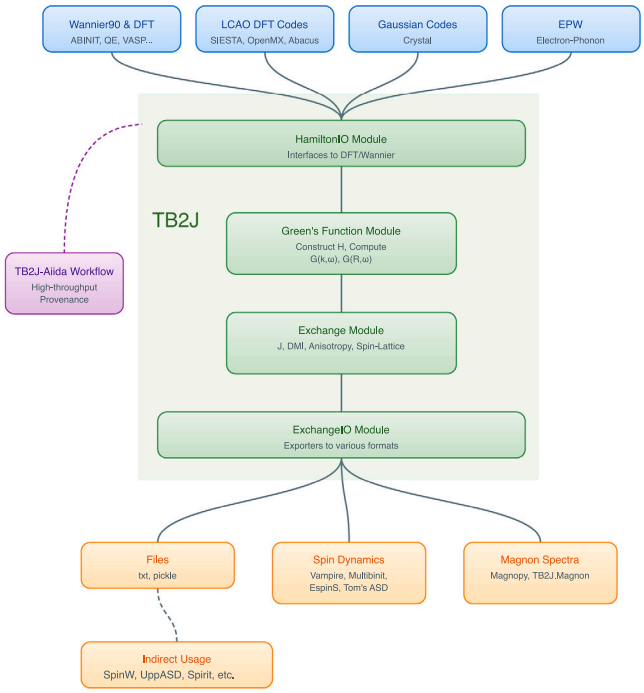

Fig. 4. The schematic architecture of TB2J and its integration in the ecosystem of DFT and spin simulation codes.

For output, TB2J generates exchange interaction parameters in formats compatible with several spin dynamics and Monte Carlo codes, such as `Vampire` [62], `MULTIBINIT` [52], `EspinS` [63], and `UppASD` [64]. This capability facilitates seamless integration into existing simulation pipelines, automating the otherwise tedious and error-prone task of format conversion. Beyond direct output, the computed parameters can be used for further analysis, such as calculating magnon spectra using linear spin-wave theory (e.g., via `SpinW` [65], `Magnopy` [66], or the internal `TB2J Magnon` module [67]). The package also includes a post-processing module for tasks like ligand spin correction and parameter symmetrization based on lattice symmetry. However, the diversity of file formats across various simulation codes remains a challenge for maintenance and interoperability. To address this, TB2J supports community efforts [https://github.com/spinham] to establish a standard spin Hamiltonian format, which would serve as a universal intermediate layer between electronic structure and spin simulation codes.

There is an `AiiDA` [39] workflow [67], which can be used to automatize the TB2J calculations and integrated to the existing `AiiDA`-DFT workflows. More recently, an agent skill [68,69] has been included in the TB2J package [48,70], exposing the main TB2J functionalities as agent tools. These tools can (i) automatically run upstream DFT calculations with parameters tailored for TB2J input, currently using `SIESTA`; (ii) identify existing DFT or `Wannier90` outputs, gather the required information, and generate TB2J input files automatically; (iii) run TB2J to construct the Heisenberg model and compute magnon band structures; (iv) detect possible pitfalls, such as non-converged Wannier functions; and (v) parse the output and generate a human-readable markdown report of the results, including links to the underlying theory, methodology, and implementation details. The same tools can be used interactively through step-by-step human commands or embedded in fully automated procedures. An agentic workflow based on these skills has also been developed within the Google Agent Development Kit (ADK) framework [71], and a Model Context Protocol (MCP) interface [72] has been built to expose TB2J capabilities to broader agent ecosystems.

The development of TB2J is largely community-driven, with many features added to address specific research needs. It relies on and integrates with existing open-source tools, such as the Atomic Simulation Environment (`ASE`) [73] and sisl [57]. The overlap between TB2J contributors and developers of major DFT codes (e.g. `ABINIT`, `SIESTA`, `ABACUS`, `Quantum ESPRESSO`, `EPW`) fosters a beneficial feedback loop, stimulating the development of new features.

The TB2J ecosystem provides a complementary example of interoperability centered on model construction. Here, the scientific protocol consists of mapping first-principles electronic structures onto spin Hamiltonians; the data layer requires robust exchange of Hamiltonians, Wannier or local-orbital representations, magnetic interaction parameters, and diagnostic metadata; the execution layer is supported by HamiltonIO, AiiDA workflows, agent workflows, MCP interfaces, and connections to electronic-structure and spin-dynamics codes; and the application layer includes both conventional user workflows and emerging agentic interfaces for human-guided or fully automated model generation.

3.3. Bridging scales: from first to second principles

One of the main theoretical tools in solid state physics are effective models that condense phenomena in relatively simple expressions that, typically, depend on a few parameters that can be obtained from fitting to the experiment. The most widespread of these models is, perhaps, the Ising Hamiltonian for magnetic systems but effective Hamiltonians are used to describe all kind of observables from electronic bands [74] to superconductivity [75] or ferroelectricity [76]. These techniques, like the related phase-field models [77], can be exploited with modern computers to simulate large structures like magnetic or polar skyrmions with typical sizes of tens or hundreds of nanometers. In this sense these semiempirical methods are complementary to first principles techniques, like DFT, that are predictive, in the sense that they do not require experimental input beyond the fundamental constants, but are limited to relatively small simulations cells, with a typical scale of 1 nm. When first principles calculations are used to provide the value of the parameters of effective Hamiltonians, they become predictive methods that may named as *second-principles*. However, effective Hamiltonians are usually designed to simulate a single property (magnetism, bands, phonons, structure among others) and, as a consequence, are very limited in scope when compared with full ab initio methods like DFT. Moreover, sometimes it is difficult to obtain model parameters in a consistent way [78,79] which, in part, can be due to neglecting degrees of freedom in the model that are important in the physical problem. For example, the Ising model does not include any geometry dependency but exchange constants are strongly dependent on structure. A solution to this problem involves creating a model Hamiltonian that is based on DFT and tries to simplify this method. There exist several approaches that try to achieve this goal like tight-binding DFT [80] or second-principles DFT (SPDFT) [81] that will be described here.

The basic concept behind SPDFT is the idea that many properties in a solid do not depend on the reaction of all the electrons of the system but only a few of them. Under this premise, one can separate the total density of the system into two contributions, the reference electronic density, n_0, and the deformation density δn:

$$n(\mathbf{r}) = n_0(\mathbf{r}) + \delta n(\mathbf{r}). \tag{1}$$

For example, the reference density in an insulator can be taken as the ground state of the system but when a dopant is introduced the extra charge and the distortion it introduces in the electron distribution of the insulator would be the deformation density. Using the generic expression for the DFT energy and using the deformation density as a perturbation, one can expand the exchange–correlation functional to second order [81] to obtain,

$$E \approx E^{(0)} + E^{(1)} + E^{(2)} + \cdots \tag{2}$$

where $E^{(0)}$ only depends on the reference density, while $E^{(1)}$ depends on it linearly, etc. The reference term, $E^{(0)}$, describes the energy surface of the system when the electron density is n_0 (that can change with the geometry). Thus, in an insulator, $E^{(0)}$ normally corresponds with the ground energy surface. In order to evaluate this term quickly an expansion of the ground energy surface around a particular geometry is employed [82], called *reference atomic geometry*, in terms of harmonic and anharmonic contributions, written as products of atomic displacements, that represents it very precisely. Thus, $E^{(0)}$ is the main contribution to the structural degrees of freedom and describes the lattice.

In order to represent the $E^{(1)}$ and $E^{(2)}$ terms, that involve changes in the electronic density a Wannier Functions (WF) [83] basis is used in SPDFT to represent the Kohn–Sham wavefunctions. The main variables in SPDFT are the one-electron reduced density matrices d_{ab}, $\mathrm{d}^{(0)}_{ab}$ and D_{ab} that are used, respectively, to represent n, n_0 and δn in the WF representation. Using this basis it can be seen that $E^{(1)}$ is the sum of the one-electron energies of the system, equivalent to the total energy of a tight-binding model, and $E^{(2)}$ includes the electron–electron interactions collecting the classical Hartree energy and also the exchange and correlation [81]. In this manner SPDFT includes structural and electronic degrees of freedom and can be used to describe a large variety of properties in solids, at difference with usual effective Hamiltonians. However, the definition of the model is only half of the problem, the other half being the determination of the parameters in the model. In SPDFT there are (i) structural parameters like force constants and anharmonic interactions, (ii) tight-binding-like bands, (iii) electron-lattice interactions that describe the change of the bands with the geometry and (iii) screened four-center electron–electron integrals [81]. Obtaining these parameters requires the interplay of the second-principles code with DFT programs in order to fully determine the model. Parameters in second-principles are obtained either from calculation or fitting of ab initio data [81,82,84,85]. Initial approaches involved fitting the models to user-defined datasets of first principles simulations [81]. This leads to models that are difficult to compare with each other, given that different users will create datasets that focus on different aspects of the studied system. Thus, much work has been devoted to refine the fitting process and the selection of the dataset used to create the models. An alternative to fitting is calculating the different parameters. This is usually a much more controlled procedure [85] and allows to create models that are much more comparable with one another. However, while some parameters, like force constants, tight-binding and electron-lattice (electron–phonon) couplings can be obtained with relative ease from first principles calculations, others, like anharmonic interactions or electron–electron integrals, are much more difficult to obtain. From the lattice side much work has been done with the `MULTIBINIT` code [52]. This is a lattice-only second-principles code that can use the DFT code `ABINIT` to semiautomatically create the harmonic and anharmonic couplings necessary to describe $E^{(0)}$. From the electron side the code `MODELMAKER` [85] interfaces with `SIESTA` [15] in order to create electron Hamiltonians with the help of `Wannier90` [86]. This code is able to create an electronic second-principles model by calculating the reference tight-binding Hamiltonian and the first- and second-order electron-lattice terms and fitting the electron–electron integrals. In order to define the quality of the model the user is required to input a few cutoffs that determine, for example, the range of the tight-binding interactions, the electron-lattice or electron–electron couplings.

SPDFT have found applications in the simulation of ferroelectric materials and their interfaces. They provide with a robust framework to describe the phase transitions in these materials [82]. The ability to carry out simulations with large simulation cells allows simulating polar topological textures like skyrmions or vortices [87] and study how electric fields, strain and temperature influences them. From the electron side SPDFT has been used to simulate magnetic states in transition metal lattices like NiO [81] or K_2AgF_4 [88]. They have also been used to calculate optical properties, including those of metals [89].

While at present not many applications of electron SPDFT models exist, the recent works dedicated to improve model generation [85] and to carry out time-dependent simulations [89] carry a great promise to deal with difficult transport problems like polaron hopping. Moreover, spectra calculated with real-time time-dependent SPDFT showed excitonic features associated with strong electron–hole correlations that show the flexibility and ability of second-principles to capture non-trivial physics. Thus, this class of methods shows the importance of continuous development of infrastructure capable of extracting information from first principles codes using various programs, like `SIESTA` or `ABINIT`, and postprocessing tools like `Wannier90` to create versatile, physically-based effective Hamiltonians.

Second-principles approaches further highlight that interoperability is not limited to file conversion, but also concerns the preservation of physical meaning across scales. The protocol layer is defined by the construction of effective Hamiltonians from first-principles data; the data layer involves the transfer of force constants, tight-binding parameters, electron-lattice couplings, and screened interactions; the execution layer relies on interfaces among DFT codes, and the application layer enables large-scale simulations using models whose parameters were obtained from ab initio calculations.

4. Interoperable workflows for thermal transport and lattice dynamics

Thermal-transport and lattice-dynamical workflows provide another view of the same four-layer structure. At the protocol level, first-principles calculations, force-constant extraction, electron–phonon and phonon-phonon interactions, and transport solvers must be combined consistently. At the data layer, the central challenge is to preserve the meaning and provenance of quantities such as interatomic force constants, scattering matrices, group velocities, and transport coefficients. At the execution and application layers, interoperable workflows make it possible to connect DFT/DFPT engines, Wannierization tools, Boltzmann solvers, phonon packages, and machine-learning potentials in automated and reusable pipelines.

4.1. Coupling electron and phonon

Over the past decades, there have been tremendous advances in DFT [91,92] and DFPT [93], both at the theory development and implementation fronts. This happened alongside an ever increasing availability of high performance computational tools. As such, accurate, parameter-free computations of the electron and phonon band structures have become relatively straightforward. DFPT, paired with the maximally localized Wannierization method [94], also allows computing the electron–phonon interactions [95] on wave vector meshes fine enough to solve the momentum resolved electronic Boltzmann transport equation (BTE), giving access to the charge carrier mobility and the thermopower. The DFT/DFPT framework also allows computing the phonon-phonon interactions, enabling a parameters-free solution of the phonon BTE, giving access to the lattice thermal conductivity. In this section, the focus is on the current state of the art of the parameters-free BTEs.

The DFT/DFPT+BTE workflow described above has been extremely popular in the transport physics community. At present, practitioners in the field can choose from a handsome selection of software packages that provide a solver for the BTE, either for electrons, or for phonons, or both. Examples include `epw` [61,95], `perturbo` [96], `elphbolt` [97], `Phoebe` [98], etc. for the electron BTE; `ShengBTE` [99], `AlmaBTE` [100], `Phono3py` [101], `FourPhonon` [102], `elphbolt` [97], `Phoebe` [98], etc. for the phonon BTE; and `elphbolt` [97] for the coupled electron–phonon BTEs.

The vast majority of the codes mentioned above make something known as the Bloch assumption [103] which states that the phonon

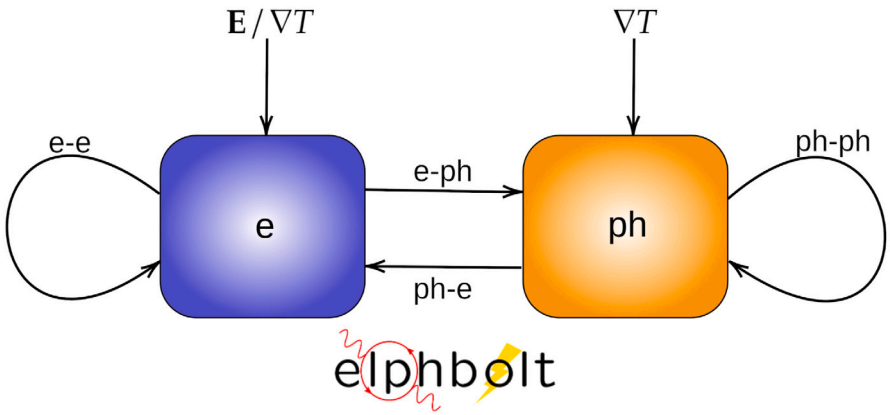

Fig. 5. Schematic showing the momentum loop between the electron (e) and phonon (ph) subsystems, as captured by the coupled e-ph BTEs. The external fields (electric field $\boldsymbol{E}$ and temperature gradient ∇T) can pump momentum into the system of particles. Via e-ph and the reverse ph-e interactions processes, the two subsystems can exchange momenta. There can be momentum redistribution within each subsystem via ph-ph and e-e processes; other possible scattering mechanisms are not shown for brevity. In the latest published version of elphbolt the e-e channel is not included. The code, however, has been recently extended to include this channel [90].

(electron) subsystem can be taken to remain in equilibrium when solving the electron (phonon) BTE. As has been pointed out by Peierls [104], this is an asymmetric treatment of the non-equilibrium state of electrons and phonons in a material where they interact. In reality, when electrons are driven out of equilibrium, they drag phonons out of equilibrium too, and vice versa, as illustrated in the momentum flow diagram in Fig. 5. This electron–phonon drag effect [105] has been shown to be extremely important for an accurate prediction of electron and phonon transport coefficients, especially the thermopower. While this has been observed early on in pioneering experiments [106–108], it is only recently that it was possible to compute these effects within a parameters-free formalism [109–112].

The `elphbolt` code, in particular, has, since its public release, enabled the transport physics community to carry out the dragful electron–phonon BTEs calculations. Over the past few years, this code has been used to make several predictions of strong drag effects in various materials including bulk Si [97], Si nanowires [113], diamond C [114], BAs [115], wGaN and wAlN [116], silicene [117], θ-TaN and MoS_2 [117,118].

The methodology described so far ignores the Coulomb interactions among the charge carriers. The prevailing wisdom is that electron–electron interactions are mostly of the momentum conserving (Normal) type and, as such, play a minor role in limiting the transport properties, which are mostly driven by momentum dissipative interactions (Umklapp). This picture is an oversimplification of the situation. While Normal processes are not directly dissipative, they do redistribute the energy and momentum among the states and, as such, indirectly affect transport properties. The most famous example of this is the hydrodynamic regime, where it is the preponderance of the non-dissipative processes that leads to the collective transport behavior. Very recently, the `elphbolt` code has been extended to include the Coulomb interactions [90]. In this work, the authors applied this tool to study the charge carrier transport properties of doped graphene. They found that the inclusion of Coulomb on top of the usual electron–phonon interactions was crucial to get decent agreement to recent measurements of the resistivity [119]. They also found that the extreme Coulomb drag effect between electrons and holes can lead to negative conductivity in one of the subsystems, a phenomenon that was very recently observed [119]. Lastly, they found that the Coulomb drag effect leads to the emergence of hydrodynamics in both the electron and hole subsystems at low temperatures. Moreover, and most importantly, they found that at low temperatures and low doping levels, the electron and the hole subsystems approach a shared hydrodynamic — so-called *bifluid* — state. None of these phenomena could be captured by the standard electron BTE theory that ignores the Coulomb interactions.

Now let us shift focus to the Boltzmann transport theory development side. Most of the codes described above ignore the interband quantum tunneling effects; the only exceptions are the `Phoebe` and the `phono3py` codes, and all of them ignore the fact that the interaction self-energies are non-band-diagonal quantities. The extension of the BTE to include the quantum tunneling effects have been derived in [120,121] for the phonon case and for the electronic case [122]. In a recent work [123], a far more general electronic BTE theory has been proposed which takes into account the non-diagonality of the self-energies and also the quantum tunneling effects. The new equation derived here — named the matrix BTE — further contains quantum corrections to the kinetic side of the transport equation. These are called kinetic corrections as they describe the effects of the interactions on the action of the external field on the particles. The standard electron BTE theory, and also the extended one in Ref. [122], ignores these terms. Furthermore, the matrix BTE, which was derived completely from first principles, has a fundamentally different off-diagonal part of the collision integral compared to the counterpart in Ref. [122] which uses a type of relaxation time approximation. The matrix BTE, however, is significantly more complex compared to the standard one due to the presence of several additional terms and the general matrix structure. However, with novel, order of magnitude cheaper methods of computing the electron–phonon matrix elements coming online [124], it is expected to see an implementation of this new theory in the near future.

4.2. *Accelerating equilibrium property sampling with machine-learning tools*

Phonons play a crucial role in various material properties, including thermal conductivity, phase transitions, thermoelectricity, and superconductivity [126–130]. Accurate prediction of phonon properties is therefore essential for understanding and designing materials [131]. Over the past few decades, first principles calculations based on DFT [91,92] have served as powerful engines for accurately predicting phonon properties. The first-principles characterization of the energy landscape of a crystalline solid around its equilibrium configuration as a truncated power series opens the door to predictive estimates of many relevant properties. The ingredients of those series are the nth derivatives of the potential energy evaluated at equilibrium, or *order-n interatomic force constants* (IFCs). First principles lattice-dynamical expansions usually rely on DFT to obtain those force constants. Low-order IFCs can be calculated employing DFPT [43,52,93,132–138], but the most widely used and versatile method to obtain all manner of IFCs is based on finite-difference formulas involving forces on

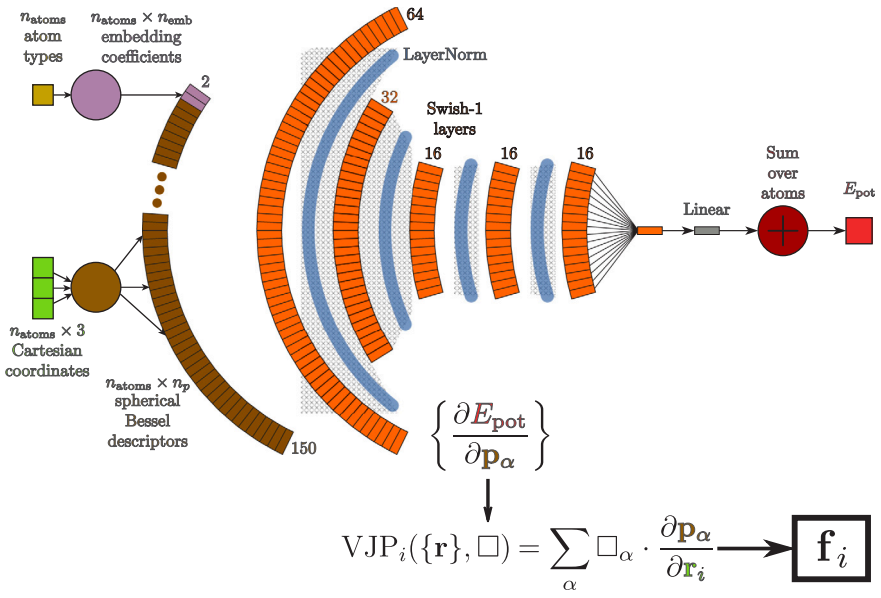

Fig. 6. Schematic representation of NeuralIL, an automatically differentiable MLIP, illustrating how the potential energy is calculated and how forward-mode AD (denoted as VJP in the figure) is applied to calculate the forces.
Source: Adapted from Ref. [125], licensed under CC-BY 4.0.

atoms [101,139–148]. However, the number of IFCs to be considered in order to obtain a good description of a crystal scales geometrically with their maximum order and, even if the symmetry of the system is optimally harnessed, so does the number of DFT runs required. This creates a significant bottleneck for this kind of workflow. The most frequent among these expansions is the harmonic approximation, where the series is truncated after the second order to calculate the phonon spectrum of the system. Third-order force constants additionally enable the calculation of the intrinsic lifetimes of those particles as limited by third-order processes; used as inputs to a solver of the Peierls-Boltzmann transport equation [99,100,141,147,149,150], these ingredients have provided accurate ab-initio predictions of the lattice thermal conductivity of many materials. Even for relatively simple cubic and hexagonal crystals, however, the workflow can easily involve thousands of DFT calculations for supercells [151]. Several extensions beyond the basic weakly-anharmonic-crystal model can make this already high cost skyrocket:

- Systems with point defects: these include doped semiconductors, ubiquitous in technological applications. Using methods based on Green's functions, the thermal conductivity of specific samples can be predicted accurately [152] and the physics of elastic phonon scattering by defects studied in great detail [153]. Unfortunately, this demands knowledge of the second-order IFCs of low-symmetry defect-laden supercells, often involving several charge states.
- Interfaces: parallel and perpendicular reconstructions, as well as interdiffusion, reduce the symmetry in this more drastic kind of crystallographic defect, whose modeling therefore requires calculations for large blocks of atoms [154].
- Higher-order anharmonicity: in situations where four-phonon processes make a significant contribution to scattering, their amplitudes must be calculated from the fourth-order IFCs [155]. This could, in principle, be extended to even higher orders, but the cost quickly becomes prohibitive.
- Temperature-dependent calculations: crystalline phases whose stability and vibrational spectrum depend on temperature require a self-consistent workflow for determining a set of effective IFCs, with a corresponding multiplication of the computational demands [156].

Fortunately, the recent explosion of machine learning has provided several strategies to accelerate lattice-dynamical calculations. For example, compressive sensing for force constant calculations [142,157] has enabled high-throughput lattice thermal conductivity calculations for approximately 6000 crystal structures by exploiting the sparsity of interatomic force constant tensors [158]. One related class of approaches consists in regarding the truncated power series as a generalized linear model whose coefficients are the IFCs. Thus, instead of finite-difference formulas based on configurations with a few atoms displaced, a global regression approach can be used, taking configurations with much more global displacements as inputs, and employing feature selection techniques to reduce the computational effort. The user-friendly and well documented package `hiPhive` [159] implements a wide range of these techniques.

However, the most revolutionary and flexible solution are machine-learning interatomic potentials (MLIPs; see Fig. 6), advanced regression models that, after being trained on a reasonably sized set of DFT calculations, can calculate the potential energy of arbitrary atomic configurations for compatible systems [160]. The progress of universal MLIPs has been remarkable, and they are now ready for application to harmonic phonon [161] and thermal conductivity calculations [160,162], although they are not always as accurate as conventional first principles approaches. Recent universal machine-learning interatomic potentials, including `MACE-MP-0` [163], `CHGNet` [164], and `M3GNet` [165], are rapidly expanding the scope of transferable atomistic simulations. While their potential applications to materials discovery have not yet been fully explored, further improvements in accuracy and applicability can be anticipated. On the other hand, task-specific MLIPs trained on smaller datasets for particular materials or material systems also prove useful. Since pre-trained models are typically trained on datasets generated by DFT calculations with specific choices of parameters such as exchange–correlation functionals, they are inherently not perfect for all applications.

Modern MLIPs [167] tend to be implemented on top of latest-generation machine-learning frameworks such as `PyTorch` [168] or `JAX` [169]. Besides other advantages like high efficiency and GPU acceleration, this makes those MLIPs automatically differentiable [125]. Not to be confused with numerical or symbolic differentiation, automatic differentiation (AD) is a family of algorithms using sophisticated versions of the chain rule to obtain differential operators based on code

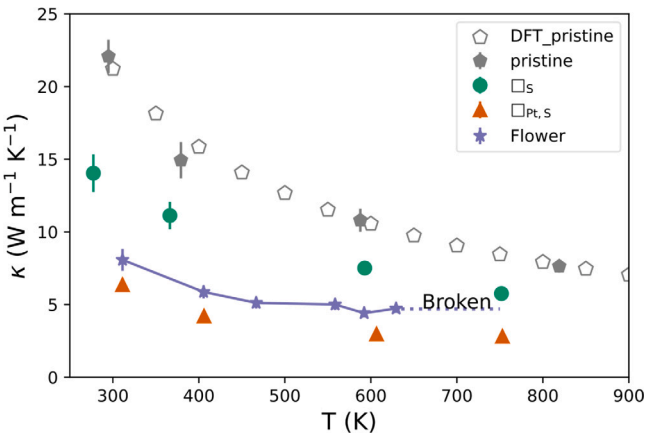

Fig. 7. Temperature dependence of the lattice thermal conductivity of a PtSTe bilayer with different kinds of defects at a concentration of 2.87×10^{13} cm^{-2}. All the ingredients for these calculations were obtained from a trained MLIP. Also included is a comparison with direct DFT calculations for the pristine system.
Source: Reproduced from Ref. [166], licensed under CC-BY 4.0.

built from primitive operations whose derivatives are known [170]. Reverse-mode automatic differentiation can calculate all the elements of a gradient down to machine precision with a cost comparable to that of the function being evaluated. Hence, MLIPs also yield forces efficiently, greatly accelerating all workflows for IFC calculations based on DFT (see Fig. 7 for an example of calculations for defect-laden crystals). However, AD can take lattice dynamics much further.

- Second-order IFCs can be calculated column-by-column by composing forward-mode AD on top of reverse-mode AD.
- IFCs can be bypassed entirely if the truncated power series they parameterized is evaluated directly using recursive passes of forward-mode AD, bringing previously unreachable orders of the expansion within reach [171].
- Since differential operators can be evaluated on bases other than the canonical basis, the dynamical matrix can be obtained in reciprocal space without passing through the IFCs, enabling an approach to phonon calculations that resembles that of DFPT.
- The dynamical matrix thus obtained is also a differentiable object, which gives access to group velocities (by differentiating with respect to the reciprocal-space point) and Grüneisen parameters (by differentiating with respect to volume).
- In the framework of Wigner transport [120], which improves upon the Boltzmann transport equation by including the contribution of coherences and is applicable to amorphous systems, the complete phonon group velocity operator, which determines the non-diagonal terms of the thermal flux, can also be calculated using AD.

On top of these direct contributions to lattice dynamics, MLIPs can also power alternative approaches whenever lattice-dynamical methods come short. For instance, for thermal conductivity calculations, the same potential can provide the ingredients to solve the Boltzmann transport equation for phonons and serve as the backend of a Green-Kubo molecular-dynamics run that include anharmonicity to all orders and scale better to larger systems [166,172–176]. Although MLIPs work well in combination with phonon calculations, their combined use remains insufficiently approachable to the broader community. Therefore, the development of user-friendly software implementations [101, 148] is of great importance. To make the prediction of phonon-related materials properties even more accessible to the scientific community, it is essential to minimize the need for users to delve into calculation details through reasonable automation and sensible default parameters. As mentioned before, achieving this goal requires the establishment of robust methodologies and workflows, along with their implementation in software. Optimizing the interplay among computational cost, numerical stability, and predictive accuracy turns out to be quite challenging in practice. Furthermore, long-term support for developed software packages, including comprehensive documentation, tutorials, easy installation channels, and application programming interfaces (APIs), is crucial to ensure their effective use not only by human users but also by AI agents. The future of lattice dynamics is therefore bright thanks to the integration between machine-learning tools and traditional workflows. Moreover, further blurring of the frontiers between the ab-initio calculations and the postprocessing part can be expected with the emergence of automatically differentiable DFT programs.

Computer simulations play an important role as scientific data generators for AI-driven materials discovery. High-throughput calculation frameworks such as `AiiDA` [39,177] and `Atomate2` [178] represent indispensable components for data generation in the scientific community. The generated data are shared through computational materials databases such as Materials Project [179] and NOMAD [180], where phonon calculation data can also be found [179–181]. All this information serve as valuable resources for materials discovery when combined with AI technologies.

The emergence of universal (or foundation) MLIPs further changes the role of interoperability in lattice-dynamical workflows. Such models replace DFT as faster input providers for phonon, thermal-transport, and molecular-dynamics calculations, thereby lowering the barrier to large-scale simulations. However, they also introduce new interoperability requirements, and challenge many of the assumptions in the simulation ecosystem. One major source of tension concerns provenance of data. The workflow must track not only structures, energies, forces, and derived force constants, but also the training data itself, as well as model version, architecture, uncertainty estimates, differentiation capabilities, validation domain, and known limitations of the MLIP. Data management becomes correspondingly more complex, because universal MLIPs are trained on aggregated heterogeneous databases produced with different codes, conventions, and standards, which makes semantic consistency across datasets increasingly important. At the same time, their reliability cannot be assessed through a single universal criterion, which motivates hybrid workflows that orchestrate DFT and MLIP calculations on demand. A second challenge concerns semantic ambiguity. In universal MLIP workflows, the same model may act as a force evaluator, a surrogate Hamiltonian, an uncertainty estimator, a data generator for active learning, or even a controller of the simulation loop itself. As a result, workflow graphs are no longer purely static and hierarchical, but become adaptive and partially agent-like, with runtime decisions depending on model confidence, data coverage, and task context. This also blurs the traditional separation between workflow execution and software deployment, since reproducibility now depends not only on the scientific code but also on model checkpoints, GPU-dependent runtime stacks, inference libraries, and versioned machine-learning infrastructure. More broadly, universal MLIPs raise a fundamentally philosophical issue: they shift part of the scientific method from an explicit theoretical construction to a statistical object distilled from previous calculations. In this setting, interoperability is no longer only about moving data between codes, but also about preserving epistemic meaning, reliability and provenance across models, databases, and workflows. Overall, systems like AiiDA, Atomate2, and FireWorks will not disappear, but will need to evolve into broader platforms that coordinate simulations, ML models, databases, and autonomous agents within a hybrid scientific workflow stack.

5. Active learning strategies for materials discovery

The discovery of functional materials is increasingly constrained by the sheer size and complexity of modern materials spaces, which

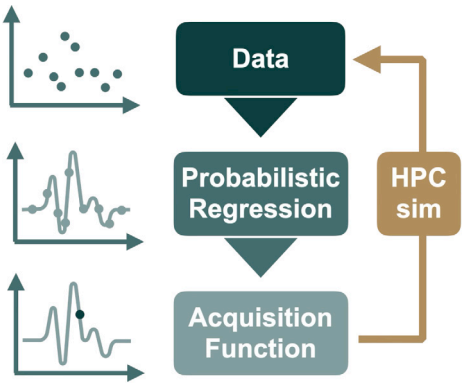

Fig. 8. Bayesian optimization algorithm, where data acquisition requires interoperability with simulations on HPC platforms.

far exceed what can be explored through trial-and-error or brute-force high-throughput screening. The combinatorial nature of materials design, combined with the high cost of evaluating many target properties, makes data-driven and machine-learning approaches essential for efficient materials discovery. In this framework, machine-learning models act as efficient surrogates for expensive structure–property relationships, enabling rapid predictions once trained on labeled data.

Over the past decade, ML has been successfully applied to a wide range of materials properties. However, a key limitation persists: most ML approaches implicitly rely on the availability of large, representative, and well-balanced datasets. In practice, materials datasets are often small, biased, and highly imbalanced, particularly when accurate calculations or experiments are required. As a result, model performance is frequently limited not by algorithmic sophistication, but by the quality and informativeness of the training data. Active learning addresses this limitation by explicitly coupling model training with adaptive data selection (Fig. 10). Rather than treating the dataset as fixed, active learning strategies iteratively identify which new data points should be labeled in order to maximally improve model performance. This paradigm is particularly well suited to materials science, where data are scarce and simulations or experiments are costly [182]. Existing active learning strategies in materials science range from model-free approaches based on diversity [183] or representativity [184] in feature space to model-based schemes that define an acquisition criterion based on the knowledge of the samples already labeled [185–188]. A prominent example of the latter is Bayesian optimization [189,190], which relies on a probabilistic surrogate model — most commonly a Gaussian process — to iteratively select new samples through acquisition functions that combine predicted performance and model uncertainty (see Fig. 10).

This probabilistic framework is implemented in the Bayesian Optimization Structure Search (BOSS) Python package for materials research [194]. BOSS constructs high-dimensional surrogate models that approximate a material's underlying energy or property landscape, allowing to identify optimal materials solutions. Surrogate models are progressively improved through strategic sampling that prioritize data points with high information gain. As indicated in Fig. 8, data sampling relies on algorithm interoperability with HPC simulations, and results in modest but informative datasets. This approach was applied in computational density functional theory studies of molecular surface adsorbates [191,195,196], thin film growth [197], solid–solid interfaces [198], molecular conformers [199], kinetic disorders [200] and magnetic interactions [192] in crystals (as illustrated in Fig. 9). In big-data machine learning, BO-like active learning strategies were deployed to prune large datasets and reduce computational needs. Iterative data labeling strategies that balance prediction uncertainty with structural diversity helped to halve big datasets without compromising AI model accuracy [201], and quickly identify materials with desired properties from a large pool of unlabeled data [202].

However data on-the-fly requires robust HPC workflows and software interoperability. Moreover, innovative algorithm development is needed to address modern challenges. Most optimization problems in chemistry involve satisfying multiple objectives so optimization procedures are often followed by Pareto front computations. Here, Pareto-targeting acquisition functions [203] could adjust the sampling toward identifying Pareto-optimal solutions directly. BO has been typically applied to optimize materials properties in continuous space, with discrete variables such as compound type excluded. Mixed discrete-continuous variable BO [204] would open the route toward more comprehensive optimal solutions, but here again specialized models and acquisition strategies are needed. Higher-dimensional BO searches could be conducted by downfolding the problem to a lower-dimensional feature space, which offers a more convenient search [205]. Specialized Gaussian process models and kernels could be adapted to account for materials symmetries and energy gradient information to reduce sampling needs. Lastly, Δ-learning and multi-task Gaussian processes have recently opened up the prospect of optimizations that combine simulation data computed with different atomistic simulations methods, with different accuracies and computational costs [206]. Multi-fidelity BO has much potential to expedite computational optimizations by majority sampling of inexpensive low fidelity data, accompanied by high-fidelity refinement.

While BO and related model-based active learning strategies generally outperform model-free approaches by exploiting target information, they often suffer from high computational cost, poor scaling with search dimensionality, and limited transferability across models and descriptors. This motivates the development of computationally efficient, model-agnostic active learning strategies that retain the ability to construct informative and well-balanced training sets while reducing labeling costs for a given input representation. In this context, recently developed regression tree-based active learning models offer a promising alternative, combining low computational cost with broad applicability across materials systems [193]. The method exploits the intrinsic partitioning of descriptor space provided by regression trees to guide the selection of informative training samples.

In practice, the number of samples selected from each leaf is determined by the variance of the target values and the fraction of unlabeled samples within the leaf. Because the leaves of a regression tree define regions of approximate homogeneity in the joint feature-target space, this acquisition strategy naturally balances target diversity with representativeness in feature space. As a result, it enables the selection of maximally informative samples—an essential aspect that is overlooked by many existing active learning approaches. The algorithm is available in the `scikit-activeml` [207] Python library. Experiments on numerous benchmark datasets demonstrated improvement over existing methods and showed its efficiency in learning a regression model from a very limited labeled dataset [193]. The effectiveness of this approach has been further demonstrated in the context of MOFs, where both electronic and adsorption properties present significant challenges for data-driven modeling [208]. Accurate predictions of MOF properties were achieved using substantially smaller training sets compared to conventional active learning approaches, such as Greedy sampling [183], Iterative Representativeness Diversity Maximization [184], Variance-Based Query by Committee [188], and Gaussian process [185]. This reduction in data requirements translates directly into savings in computational and experimental effort, while preserving — or even improving — predictive accuracy. Moreover, the tree-based partitions provide interpretable insight into structure–property relationships, an essential aspect for rational materials design.

Active learning makes the link between interoperability and autonomous materials discovery particularly explicit. In this context, the scientific protocol is no longer a fixed sequence of calculations, but an adaptive loop in which models, acquisition criteria, simulations,

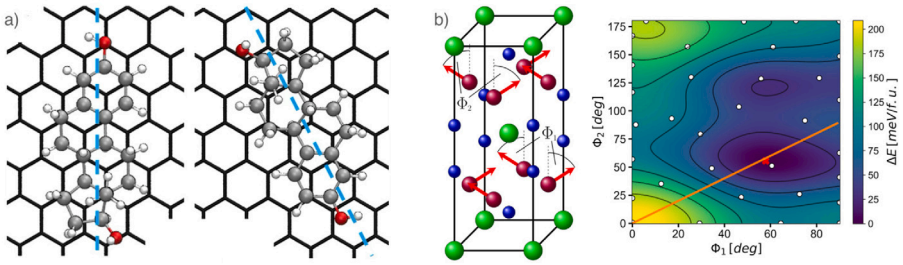

Fig. 9. Bayesian optimization for (a) atomistic structure search of hormone sensing on Gr and (b) studies of spin-spin interactions via magnetic energy landscapes. *Source:* (a) Adapted from Ref. [191] (b) Adapted from Ref. [192]. Licensed under CC-BY 4.0.

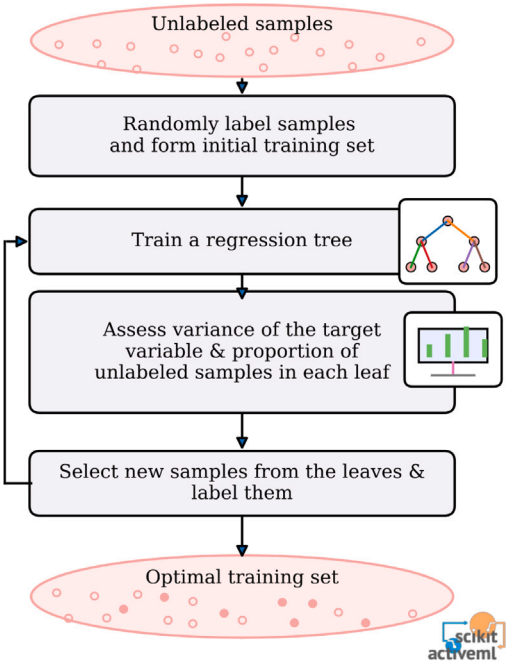

Fig. 10. Flowchart representing regression tree-based active learning [193], with the acquisition criterion described in detail.

and data updates interact. The data layer must therefore record not only successful calculations, but also model versions, selected and rejected candidates, uncertainty estimates, and failed or low-confidence branches. The execution layer must support iterative decision-making on HPC resources, while the application layer must expose these loops in a way that can be controlled either by human researchers or by autonomous agents.

6. Advanced molecular dynamics and phase-space sampling workflows

Molecular Dynamics (MD) simulations are among the most widely used computational approaches to investigate systems at atomic resolution, particularly when properties arising from configurations far from equilibrium are of interest [209]. These simulations powerfully exemplify the multifaceted nature of computational materials science, where diverse spatial and temporal scales — coupled with complex phenomena — demand specialized approximations, modular methodological implementations across distinct codes, and seamless interoperability to achieve computationally efficient simulations for a comprehensive understanding of the system under study.

Within the MD framework, the temporal evolution of a collection of atoms is governed by interatomic interactions encoded in a potential energy function, which depends explicitly on the atomic coordinates. These interactions can be supplied from classical *force fields* [209] or evaluated at each time step by quantum mechanical calculations as in *Ab Initio* Molecular Dynamics [210]. An intermediate strategy is offered by hybrid QM/MM methods [211], in which a subset of the system is treated quantum mechanically (QM) while the remainder is described using classical mechanics (MM). Unfortunately, these QM/MM methods require *a priori* identification of the QM region, and a reliable parameterized interatomic potential model for the MM region.

Recent advances allow adaptive descriptions of the potential energy surface, where the level of theoretical fidelity is no longer uniform across the simulation domain. *Adaptive-precision* schemes dynamically combine classical interatomic potentials with machine-learning-based force models, allowing forces on selected atoms to be evaluated at higher accuracy while the remainder of the system is treated using computationally inexpensive classical interactions; implementation examples include the use of atomic cluster expansion as machine learning potential fitted on DFT data as in Ref. [212]. Unlike conventional QM/MM schemes, this enables atom-resolved accuracy control without prior system-specific assumptions. All these schemes benefit from abstraction and modularization of the codes, as the MD algorithms can be easily decoupled from an external module that provides Energy and Forces (and is agnostic of the level of theory used, either parameterized semiempirical potentials, DFT calculations, or MLIPs).

On top of this simple structure, we can add new levels of methodological improvements, with modularity among different approximations. For example, the cost and rigidity of machine-learning force fields can be reduced by embedding active learning directly within Molecular Dynamics simulations. In active-learning-driven MD, the potential energy model is refined during the simulation itself whenever the explored configurations fall outside the domain of reliable interpolation, as quantified by uncertainty estimators. The `falcon` framework implements this idea by partitioning the training dataset into clusters of structurally similar configurations and training multiple specialized machine-learning models on DFT data instead of a single global one [213]. Another example is the extension of MD into regimes dominated by electronically excited states and non-adiabatic transitions. Machine-learning models capable of representing multiple electronic potential energy surfaces and their couplings have been developed to accelerate non-adiabatic molecular dynamics, particularly when combined with trajectory surface hopping algorithms [214,215]. These physics-informed models are trained on DFT or post-Hartree–Fock reference data which are obtained from ab initio packages; feeding of the ML model is based on *ad hoc* interfaces between the selected ML engine and the ab initio software. Such machine-learning-accelerated excited-state dynamics represents a significant departure from classical MD, as the nuclear motion is explicitly coupled to evolving electronic degrees of freedom [216].

Alongside these emerging techniques, several established non-standard molecular dynamics methodologies continue to play a central role in overcoming intrinsic limitations of conventional Newtonian simulations. These require modularity not on the Energy-and-Forces

provider, but on the solver of the dynamics evaluation itself. A couple of example illustrate how this can be used to reach larger system sizes, or longer simulation times (by better sampling of the dynamical space)

Long-range structural distortions govern phenomena such as rippling, thermal transport, diffusion processes, and nucleation [217–224], and studying their effect with conventional MD simulations requires large structural models that are computationally expensive. The recently proposed *Normal Dynamics* technique [225], allows to limit the computational cost by solving Newton's equations of motion in the reciprocal space. By exploiting the phonon mode formalism, dynamical trajectories are generated by sampling the reciprocal space instead of the direct space; this allows us to evolve only selected long-wave perturbations by a suitable choice of the reciprocal wave vectors, thus reducing the degrees of freedom to be integrated in time compared to standard Cartesian dynamics. The method is implemented in the `pindol` Fortran code [226], which is interfaced with the `phonopy` [227], `phono3py` [150] and `hiPhive` [159] packages for the preparation of the input files and postprocessing of the produced dynamical trajectories.

Moving toward even larger systems, coarse-grained molecular dynamics reduces the number of degrees of freedom by mapping groups of atoms onto effective interaction sites, thus smoothing the potential energy landscape and allowing simulations of mesoscale phenomena [228]. Adaptive resolution simulation methods extend coarse-graining by allowing particles to transition dynamically between atomistic and coarse-grained representations within a single simulation, ensuring local accuracy while retaining global efficiency [229–232].

Enhanced sampling techniques address the limited configurational exploration of standard MD by modifying the sampling process through bias potentials or ensemble transformations that accelerate rare events. Methods such as metadynamics, umbrella sampling, replica exchange molecular dynamics, and adaptive biasing force algorithms, also integrating machine-learning engines, enable the reconstruction of free-energy landscapes and the study of activated processes inaccessible on conventional MD timescales [233–236]. Another method is the path integral molecular dynamics, which incorporates nuclear quantum effects into molecular simulations by mapping each quantum particle onto a classical ring polymer representation derived from the Feynman path integral formalism [237–241]. Finally, dissipative particle dynamics represents a mesoscopic simulation paradigm in which coarse particles interact via conservative, dissipative, and stochastic forces that collectively reproduce correct hydrodynamic behavior [242–246].

All these methodologies illustrate a broad evolution of molecular dynamics and the sampling of the phase space toward adaptive, learning-based, and multiscale frameworks that transcend fixed Newtonian formulations and enable the simulation of increasingly complex physical phenomena. As molecular-dynamics workflows increasingly combine multiple levels of resolution, interoperability must extend beyond input–output exchange to include the evolving decision logic of the simulation. This reinforces the broader message of this viewpoint article: future interoperable ecosystems must support dynamic, provenance-aware, and machine-readable workflows suitable for both human users and autonomous agents.

7. Conclusion

We have shown that software interoperability has become central to atomistic materials simulations, as contemporary research increasingly relies on workflows that combine multiple codes, models, and data-processing stages. Interoperability enables, for example, lattice-dynamics calculations accelerated by machine-learning potentials, the construction of second-principles electronic and magnetic models, real-time simulations of nonlinear optical responses, and transport calculations based on coupled electron–phonon descriptions. Examples drawn from magnetism, molecular dynamics, and data-driven optimization further illustrate how modular software design allows parameters, models, and data to move seamlessly between electronic-structure calculations, effective Hamiltonians, and large-scale simulations.

At the infrastructure level, workflow frameworks, provenance tracking, and standardized interfaces emerge as essential enablers of automation, reproducibility, and high-throughput studies, as well as of the generation of data that can be reused in machine-learning and AI applications. Taken together, these developments point to a broader shift from isolated simulation codes toward interoperable software ecosystems. Progress will depend on clearer interfaces, shared workflow definitions, and more robust handling of data semantics and provenance, particularly as AI-driven and autonomous workflows become more prevalent.

Beyond the interoperability aspects discussed here, recent advances in agentic AI and multi-agent scientific workflows suggest new possibilities for more adaptive simulation and discovery pipelines. Nevertheless, the present Perspective focuses on the standards, provenance mechanisms, and workflow infrastructure that such systems would require in practice. In this sense, interoperability remains a necessary foundation for both human-supervised automation and future autonomous discovery, even if the latter is still emerging and unevenly developed across atomistic simulation subfields. As autonomous agents become more common users of simulation tools, Layer 4 will need to give greater importance to machine-actionable interfaces, such as structured error messages, self-describing APIs, and standardized output schemas. For electronic-structure codes, this means designing interfaces that can support not only human users, but also automated and agent-based workflows.

In this context, interoperability emerges not merely as desirable feature but an increasingly necessary technical requirement for scaling atomistic simulations, ensuring reproducibility, and enabling the next generation of materials discovery.

CRediT authorship contribution statement

Jesús Carrete: Writing – original draft. **Pablo García-Fernández:** Writing – original draft. **Myrta Grüning:** Writing – original draft. **Xu He:** Writing – original draft. **Mike N. Pionteck:** Writing – original draft. **Nakib H. Protik:** Writing – original draft. **Milica Todorović:** Writing – original draft. **Atsushi Togo:** Writing – original draft. **Xing Wang:** Writing – original draft. **Roberta Poloni:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Miguel Pruneda:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Antonio Cammarata:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

A.C. acknowledges support of the Czech Science Foundation (project No. 24-12643L), cofunding by the European Union under the project "Robotics and advanced industrial production" (reg. no. CZ.02.01.01/00/22_008/0004590), and the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90254).

M.P. acknowledges support from Grant PID2022-139776NB-C61 funded by the Spanish MCIN/AEI/10.13039/501100011033 and by the ERDF, A way of making Europe, Grant 2021SGR01519 from Generalitat de Catalunya, and by EU MaX CoE (Grant 101093374).

A.T. acknowledges support by JSPS KAKENHI Grant Numbers JP24K08021, JP24H00190, JP25H01246, and JP25H01252; some of the calculations in this study were performed on the Numerical Materials Simulator at NIMS.

We gratefully thank Giovanni Pizzi for insightful discussions on workflow interoperability and provenance semantics.

X.W. acknowledge financial support by the NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 205602).

N.H.P. acknowledges funding from the "Deutsche Forschungsgemeinschaft" (DFG, German Research Foundation) for an "Emmy Noether" research grant (Grant No. 534386252).

M.G. acknowledges funding from the UKRI Horizon Europe Guarantee funding scheme (EP/Y032659/1).

Data availability

No data was used for the research described in the article.

References

[1] A. Ghafarollahi, M.J. Buehler, Automating alloy design and discovery with physics-aware multimodal multiagent AI, Proc. Natl. Acad. Sci. 122 (4) (2025) e2414074122.

[2] F.Y. Wang, L. Marom, S. Pal, R.K. Luu, W. Lu, J.A. Berkovich, M.J. Buehler, Autonomous agents coordinating distributed discovery through emergent artifact exchange, 2026, arXiv preprint arXiv:2603.14312.

[3] I. Takahara, T. Mizoguchi, B. Liu, Accelerated inorganic materials design with generative AI agents, Cell Rep. Phys. Sci. 6 (12) (2025).

[4] T.D. Pham, A. Tanikanti, M. Keçeli, ChemGraph as an agentic framework for computational chemistry workflows, Commun. Chem. (2026).

[5] S. Curtarolo, G.L. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, Nat. Mater. 12 (3) (2013) 191–201, http://dx.doi.org/10.1038/nmat3568.

[6] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky, AiiDA: automated interactive infrastructure and database for computational science, Comput. Mater. Sci. 111 (2016) 218–230, http://dx.doi.org/10.1016/j.commatsci.2015.09.013.

[7] S.P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A.V. Yakutovich, C.W. Andersen, et al., AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance, Sci. Data 7 (1) (2020) 300, http://dx.doi.org/10.1038/s41597-020-00638-4.

[8] A.S. Rosen, M. Gallant, J. George, J. Riebesell, H. Sahasrabuddhe, J.-X. Shen, M. Wen, M.L. Evans, G. Petretto, D. Waroquiers, et al., Jobflow: Computational workflows made simple, J. Open Source Softw. 9 (93) (2024) 5995, http://dx.doi.org/10.21105/joss.05995.

[9] J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, J. Neugebauer, Pyiron: An integrated development environment for computational materials science, Comput. Mater. Sci. 163 (2019) 24–36, http://dx.doi.org/10.1016/j.commatsci.2018.07.043.

[10] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, et al., Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater. 1 (1) (2013) http://dx.doi.org/10.1063/1.4812323.

[11] S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies, Npj Comput. Mater. 1 (1) (2015) 1–15, http://dx.doi.org/10.1038/npjcompumats.2015.10.

[12] L. Talirz, S. Kumbhar, E. Passaro, A.V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S.P. Huber, S. Zoupanos, et al., Materials Cloud, a platform for open computational science, Sci. Data 7 (1) (2020) 299, http://dx.doi.org/10.1038/s41597-020-00637-5.

[13] S.P. Huber, E. Bosoni, M. Bercx, J. Bröder, A. Degomme, V. Dikan, K. Eimre, E. Flage-Larsen, A. Garcia, L. Genovese, et al., Common workflows for computing material properties using different quantum engines, Npj Comput. Mater. 7 (1) (2021) 136, http://dx.doi.org/10.1038/s41524-021-00594-6.

[14] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. De Gironcoli, P. Delugas, F. Ferrari Ruffino, et al., Quantum ESPRESSO toward the exascale, J. Chem. Phys. 152 (15) (2020) http://dx.doi.org/10.1063/5.0005082.

[15] A. García, N. Papior, A. Akhtar, E. Artacho, V. Blum, E. Bosoni, P. Brandimarte, M. Brandbyge, J.I. Cerdá, F. Corsetti, et al., Siesta: Recent developments and applications, J. Chem. Phys. 152 (20) (2020) http://dx.doi.org/10.1063/5.0005077.

[16] T.D. Kühne, M. Iannuzzi, M. Del Ben, V.V. Rybkin, P. Seewald, F. Stein, T. Laino, R.Z. Khaliullin, O. Schütt, F. Schiffmann, et al., CP2K: An electronic structure and molecular dynamics software package-Quickstep: Efficient and accurate electronic structure calculations, J. Chem. Phys. 152 (19) (2020) http://dx.doi.org/10.1063/5.0007045.

[17] E. Bosoni, L. Beal, M. Bercx, P. Blaha, S. Blügel, J. Bröder, M. Callsen, S. Cottenier, A. Degomme, V. Dikan, et al., How to verify the precision of density-functional-theory implementations via reproducible and universal workflows, Nat. Rev. Phys. 6 (1) (2024) 45–58, http://dx.doi.org/10.1038/s42254-023-00655-3.

[18] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I.E. Castelli, S.J. Clark, A. Dal Corso, et al., Reproducibility in density functional theory calculations of solids, Science 351 (6280) (2016) aad3000, http://dx.doi.org/10.1126/science.aad3000.

[19] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, et al., The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 1–9, http://dx.doi.org/10.1038/sdata.2016.18.

[20] C.W. Andersen, R. Armiento, E. Blokhin, G.J. Conduit, S. Dwaraknath, M.L. Evans, Á. Fekete, A. Gopakumar, S. Gražulis, A. Merkys, et al., OPTIMADE, an API for exchanging materials data, Sci. Data 8 (1) (2021) 217, http://dx.doi.org/10.1038/s41597-021-00974-z.

[21] M.L. Evans, J. Bergsma, A. Merkys, C.W. Andersen, O.B. Andersson, D. Beltrán, E. Blokhin, T.M. Boland, R.C. Balderas, K. Choudhary, et al., Developments and applications of the OPTIMADE API for materials discovery, design, and data exchange, Digit. Discov. 3 (8) (2024) 1509–1533, http://dx.doi.org/10.1039/D4DD00039K.

[22] A. Jain, S.P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, et al., FireWorks: A dynamic workflow system designed for high-throughput applications, Concurr. Comput.: Pr. Exp. 27 (17) (2015) 5037–5059, http://dx.doi.org/10.1002/cpe.3505.

[23] K. Mathew, J.H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I.-h. Chu, T. Smidt, B. Bocklund, M. Horton, et al., Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows, Comput. Mater. Sci. 139 (2017) 140–152, http://dx.doi.org/10.1016/j.commatsci.2017.07.030.

[24] J. Janssen, J. George, J. Geiger, M. Bercx, X. Wang, C. Ertural, J. Schaarschmidt, A.M. Ganose, G. Pizzi, T. Hickel, J. Neugebauer, A python workflow definition for computational materials design, Digit. Discov. 4 (2025) 3149–3161, http://dx.doi.org/10.1039/D5DD00231A.

[25] X. Wang, E. Bainglass, M. Bonacci, A. Ortega-Guerrero, L. Bastonero, M. Bercx, P. Bonfà, R. De Renzi, D. Du, P.N. Gillespie, et al., Making atomistic materials calculations accessible with the AiiDAlab quantum ESPRESSO app, 2025, http://dx.doi.org/10.48550/arXiv.2507.19670, arXiv preprint arXiv:2507.19670.

[26] A.V. Yakutovich, K. Eimre, O. Schütt, L. Talirz, C.S. Adorf, C.W. Andersen, E. Ditler, D. Du, D. Passerone, B. Smit, et al., AiiDAlab–an ecosystem for developing, executing, and sharing scientific workflows, Comput. Mater. Sci. 188 (2021) 110165, http://dx.doi.org/10.1016/j.commatsci.2020.110165.

[27] D. Sangalli, A. Ferretti, H. Miranda, C. Attaccalite, I. Marri, E. Cannuccia, P. Melo, M. Marsili, F. Paleari, A. Marrazzo, G. Prandini, P. Bonfà, M.O. Atambo, F. Affinito, M. Palummo, A. Molina-Sánchez, C. Hogan, M. Grüning, D. Varsano, A. Marini, Many-body perturbation theory calculations using the yambo code, J. Phys.: Condens. Matter. 31 (32) (2019) 325902, http://dx.doi.org/10.1088/1361-648X/ab15d0.

[28] D. Sangalli, R. Reho, I.M. Alliati, J. Cervantes-Villanueva, A. Esquembre Kučukalić, T. Geirsson, D.A. Leon Valido, B. Mellado Pinto, P. Milev, M. Nalabothula, F. Paleari, A. Romani, L. Bellentani, J.M. Castelo Ares, M. D'Alessandro, S. Laricchia, E.M. Montagna, S. Orlandini, J.E. Roman, F. Affinito, A. Molina-Sánchez, C. Attaccalite, M. Grüning, Lumen, 2026, http://dx.doi.org/10.5281/zenodo.20121198.

[29] C. Attaccalite, M. Grüning, Nonlinear optics from an ab initio approach by means of the dynamical Berry phase: Application to second- and third-harmonic generation in semiconductors, Phys. Rev. B 88 (2013) 235113, http://dx.doi.org/10.1103/PhysRevB.88.235113, URL https://link.aps.org/doi/10.1103/PhysRevB.88.235113.

[30] C. Attaccalite, M. Grüning, A. Marini, Real-time approach to the optical properties of solids and nanostructures: Time-dependent Bethe-Salpeter equation, Phys. Rev. B 84 (2011) 245110, http://dx.doi.org/10.1103/PhysRevB.84.245110, URL https://link.aps.org/doi/10.1103/PhysRevB.84.245110.

[31] A. Marini, C. Hogan, M. Grüning, D. Varsano, Yambo: An ab initio tool for excited state calculations, Comput. Phys. Comm. 180 (8) (2009) 1392–1403, http://dx.doi.org/10.1016/j.cpc.2009.02.003, URL https://www.sciencedirect.com/science/article/pii/S0010465509000472.

[32] F. Paleari, A. Molina-Sánchez, M. Nalabothula, R. Reho, M. Bonacci, J.M. Castelo, J. Cervantes-Villanueva, M. Pionteck, M. Silvetti, C. Attaccalite, H. Pereira Coutada Miranda, Yambopy, 2025, http://dx.doi.org/10.5281/zenodo.15012962.

[33] M.N. Pionteck, M. Grüning, S. Sanna, C. Attaccalite, A real-time approach to frequency-mixing spectroscopies: Application to sum and difference frequency generation in two-dimensional crystals, SciPost Phys. 19 (2025) 129, http://

dx.doi.org/10.21468/SciPostPhys.19.5.129, URL https://scipost.org/10.21468/SciPostPhys.19.5.129.

[34] Y.-H. Chan, D.Y. Qiu, F.H. da Jornada, S.G. Louie, Giant exciton-enhanced shift currents and direct current conduction with subbandgap photo excitations produced by many-electron interactions, Proc. Natl. Acad. Sci. 118 (25) (2021) e1906938118, http://dx.doi.org/10.1073/pnas.1906938118, arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1906938118. URL https://www.pnas.org/doi/abs/10.1073/pnas.1906938118.

[35] C. Attaccalite, M. Grüning, H. Amara, S. Latil, F. Ducastelle, Two-photon absorption in two-dimensional materials: The case of hexagonal boron nitride, Phys. Rev. B 98 (2018) 165126, http://dx.doi.org/10.1103/PhysRevB.98.165126, URL https://link.aps.org/doi/10.1103/PhysRevB.98.165126.

[36] Y. Mao, J. Zhou, M. Grüning, C. Attaccalite, Shift current in two-dimensional Janus transition-metal Dichalcogenides: The role of excitons, Phys. Rev. Mater. 9 (12) (2025) http://dx.doi.org/10.1103/qth8-6spb.

[37] S. Grillo, E. Cannuccia, M. Palummo, O. Pulci, C. Attaccalite, Tunable second harmonic generation in 2D materials: Comparison of different strategies, SciPost Phys. Core 7 (2024) 081, http://dx.doi.org/10.21468/SciPostPhysCore.7.4.081, URL https://scipost.org/10.21468/SciPostPhysCore.7.4.081.

[38] B. Hou, J. Wu, V. Chang Lee, J. Guo, L.Y. Liu, D.Y. Qiu, Data-driven low-rank approximation for the electron-hole kernel and acceleration of time-dependent GW calculations, Npj Comput. Mater. 11 (1) (2025) 204, http://dx.doi.org/10.1038/s41524-025-01680-9.

[39] S.P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A.V. Yakutovich, C.W. Andersen, F.F. Ramirez, C.S. Adorf, F. Gargiulo, S. Kumbhar, E. Passaro, C. Johnston, A. Merkys, A. Cepellotti, N. Mounet, N. Marzari, B. Kozinsky, G. Pizzi, AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance, Sci. Data 7 (1) (2020) 300, http://dx.doi.org/10.1038/s41597-020-00638-4.

[40] H. Xiang, C. Lee, H.-J. Koo, X. Gong, M.-H. Whangbo, Magnetic properties and energy-mapping analysis, Dalton Trans. 42 (4) (2013) 823–853, http://dx.doi.org/10.1039/C2DT31662E.

[41] L.M. Sandratskii, Energy band structure calculations for crystals with spiral magnetic structure, Phys. Status Solidi (B) 136 (1) (1986) 167–180, http://dx.doi.org/10.1002/pssb.2221360119.

[42] S.Y. Savrasov, Linear response calculations of spin fluctuations, Phys. Rev. Lett. 81 (12) (1998) 2570, http://dx.doi.org/10.1103/PhysRevLett.81.2570.

[43] A.H. Romero, D.C. Allan, B. Amadon, G. Antonius, T. Applencourt, L. Baguet, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, F. Bruneval, G. Brunin, D. Caliste, M. Côté, J. Denier, C. Dreyer, P. Ghosez, M. Giantomassi, Y. Gillet, O. Gingras, D.R. Hamann, G. Hautier, F. Jollet, G. Jomard, A. Martin, H.P.C. Miranda, F. Naccarato, G. Petretto, N.A. Pike, V. Planes, S. Prokhorenko, T. Rangel, F. Ricci, G.-M. Rignanese, M. Royo, M. Stengel, M. Torrent, M.J. van Setten, B.V. Troeye, M.J. Verstraete, J. Wiktor, J.W. Zwanziger, X. Gonze, ABINIT: Overview, and focus on selected capabilities, J. Chem. Phys. 152 (2020) 124102, http://dx.doi.org/10.1063/1.5144261.

[44] K. Cao, H. Lambert, P.G. Radaelli, F. Giustino, Ab initio calculation of spin fluctuation spectra using time-dependent density functional perturbation theory, plane waves, and pseudopotentials, Phys. Rev. B 97 (2) (2018) 024420, http://dx.doi.org/10.1103/PhysRevB.97.024420.

[45] F. Ricci, S. Prokhorenko, M. Torrent, M.J. Verstraete, E. Bousquet, Density functional perturbation theory within noncollinear magnetism, Phys. Rev. B 99 (18) (2019) 184404, http://dx.doi.org/10.1103/PhysRevB.99.184404.

[46] J.J. Phillips, J.E. Peralta, Towards the blackbox computation of magnetic exchange coupling parameters in polynuclear transition-metal complexes: Theory, implementation, and application, J. Chem. Phys. 138 (17) (2013) 174115, http://dx.doi.org/10.1063/1.4802776.

[47] A.I. Liechtenstein, M. Katsnelson, V. Antropov, V. Gubanov, Local spin density functional approach to the theory of exchange interactions in ferromagnetic metals and alloys, J. Magn. Magn. Mater. 67 (1) (1987) 65–74, http://dx.doi.org/10.1016/0304-8853(87)90721-9.

[48] X. He, N. Helbig, M.J. Verstraete, E. Bousquet, TB2J: A python package for computing magnetic interaction parameters, Comput. Phys. Comm. 264 (2021) 107938, http://dx.doi.org/10.1016/j.cpc.2021.107938.

[49] G. Martínez-Carracedo, L. Oroszlány, A. García-Fuente, B. Nyári, L. Udvardi, L. Szunyogh, J. Ferrer, Relativistic magnetic interactions from nonorthogonal basis sets, Phys. Rev. B 108 (2023) 214418, http://dx.doi.org/10.1103/PhysRevB.108.214418, URL https://link.aps.org/doi/10.1103/PhysRevB.108.214418.

[50] D.M. Korotin, V.V. Mazurenko, V.I. Anisimov, S.V. Streltsov, Calculation of exchange constants of the Heisenberg model in plane-wave-based methods using the Green's function approach, Phys. Rev. B 91 (2015) 224405, http://dx.doi.org/10.1103/PhysRevB.91.224405, URL https://link.aps.org/doi/10.1103/PhysRevB.91.224405.

[51] L. Oroszlány, J. Ferrer, A. Deák, L. Udvardi, L. Szunyogh, Exchange interactions from a nonorthogonal basis set: From bulk ferromagnets to the magnetism in low-dimensional graphene systems, Phys. Rev. B 99 (22) (2019) 224412, http://dx.doi.org/10.1103/PhysRevB.99.224412.

[52] X. Gonze, B. Amadon, G. Antonius, F. Arnardi, L. Baguet, J.-M. Beuken, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, N. Brouwer, F. Bruneval, G. Brunin, T. Cavignac, J.-B. Charraud, W. Chen, M. Côté, S. Cottenier, J. Denier, G. Geneste, P. Ghosez, M. Giantomassi, Y. Gillet, O. Gingras, D.R. Hamann, G. Hautier, X. He, N. Helbig, N. Holzwarth, Y. Jia, F. Jollet, W. Lafargue-Dit-Hauret, K. Lejaeghere, M.A.L. Marques, A. Martin, C. Martins, H.P.C. Miranda, F. Naccarato, K. Persson, G. Petretto, V. Planes, Y. Pouillon, S. Prokhorenko, F. Ricci, G.-M. Rignanese, A.H. Romero, M.M. Schmitt, M. Torrent, M.J. van Setten, B.V. Troeye, M.J. Verstraete, G. Zérah, J.W. Zwanziger, The abinit project: Impact, environment and recent developments, Comput. Phys. Comm. 248 (2020) 107042, http://dx.doi.org/10.1016/j.cpc.2019.107042.

[53] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M.B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, et al., Advanced capabilities for materials modelling with Quantum ESPRESSO, J. Phys.: Condens. Matter. 29 (46) (2017) 465901, http://dx.doi.org/10.1088/1361-648X/aa8f79.

[54] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (16) (1996) 11169, http://dx.doi.org/10.1103/PhysRevB.54.11169.

[55] J. Enkovaara, C. Rostgaard, J.J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H.A. Hansen, et al., Electronic structure calculations with GPAW: A real-space implementation of the projectoraugmented-wave method, J. Phys.: Condens. Matter. 22 (25) (2010) 253202, http://dx.doi.org/10.1088/0953-8984/22/25/253202.

[56] J.M. Soler, E. Artacho, J.D. Gale, A. García, J. Junquera, P. Ordejón, D. Sánchez-Portal, The SIESTA method for ab initio order-Nmaterials simulation, J. Phys.: Condens. Matter. 14 (11) (2002) 2745, http://dx.doi.org/10.1088/0953-8984/14/11/302.

[57] N.R. Papior, B. Jonas L, T. Frederiksen, S. Sanz Wuhl, Zerothi/sisl: v0. 9.8, 2020, Zenodo.

[58] T. Ozaki, H. Kino, Efficient projector expansion for the ab initio LCAO method, Phys. Rev. B 72 (2005) 045121, http://dx.doi.org/10.1103/PhysRevB.72.045121, URL https://link.aps.org/doi/10.1103/PhysRevB.72.045121.

[59] P. Li, X. Liu, M. Chen, P. Lin, X. Ren, L. Lin, C. Yang, L. He, Large-scale ab initio simulations based on systematically improvable atomic basis, Comput. Mater. Sci. 112 (2016) 503–517, http://dx.doi.org/10.1016/j.commatsci.2015.07.004.

[60] A. Erba, J.K. Desmarais, S. Casassa, B. Civalleri, L. Donà, I.J. Bush, B. Searle, L. Maschio, L. Edith-Daga, A. Cossard, et al., CRYSTAL23: A program for computational solid state physics and chemistry, J. Chem. Theory Comput. 19 (20) (2022) 6891–6932, http://dx.doi.org/10.1021/acs.jctc.2c00958.

[61] H. Lee, S. Poncé, K. Bushick, S. Hajinazar, J. Lafuente-Bartolome, J. Leveillee, C. Lian, J.-M. Lihm, F. Macheda, H. Mori, H. Paudyal, W.H. Sio, S. Tiwari, M. Zacharias, X. Zhang, N. Bonini, E. Kioupakis, E.R. Margine, F. Giustino, Electron–phonon physics from first principles using the EPW code, Npj Comput. Mater. 9 (1) (2023) 156, http://dx.doi.org/10.1038/s41524-023-01107-3.

[62] R.F. Evans, W.J. Fan, P. Chureemart, T.A. Ostler, M.O. Ellis, R.W. Chantrell, Atomistic spin model simulations of magnetic nanomaterials, J. Phys.: Condens. Matter. 26 (10) (2014) 103202, http://dx.doi.org/10.1088/0953-8984/26/10/103202.

[63] N. Rezaei, M. Alaei, H. Akbarzadeh, ESpinS: A program for classical Monte-Carlo simulations of spin systems, Comput. Mater. Sci. 202 (2022) 110947, http://dx.doi.org/10.1016/j.commatsci.2021.110947.

[64] M. Hortamani, L. Sandratskii, P. Kratzer, I. Mertig, M. Scheffler, Exchange interactions and critical temperature of bulk and thin films of MnSi: A density functional theory study, Phys. Rev. B—Condensed Matter Mater. Phys. 78 (10) (2008) 104402, http://dx.doi.org/10.1103/PhysRevB.78.104402.

[65] S. Toth, B. Lake, Linear spin wave theory for single-Q incommensurate magnetic structures, J. Phys.: Condens. Matter. 27 (16) (2015) 166002, http://dx.doi.org/10.1088/0953-8984/27/16/166002.

[66] A. Ivanov, V. Uzdin, H. Jónsson, Fast and robust algorithm for energy minimization of spin systems applied in an analysis of high temperature spin configurations in terms of skyrmion density, Comput. Phys. Comm. 260 (2021) 107749, http://dx.doi.org/10.1016/j.cpc.2020.107749.

[67] A. Tellez-Mora, X. He, E. Bousquet, L. Wirtz, A.H. Romero, Systematic determination of a material's magnetic ground state from first principles, Npj Comput. Mater. 10 (1) (2024) 20, http://dx.doi.org/10.1038/s41524-024-01202-z.

[68] Anthropic, Introducing agent skills, 2025, https://claude.com/blog/skills. (Accessed 25 May 2026).

[69] Anthropic, contributors, Agent skills: A standardized way to give AI agents new capabilities and expertise, 2025, Open specification. (Accessed 25 May 2026), https://agentskills.io.

[70] X. He, contributors, TB2J: A python package for computing magnetic interaction parameters, 2026, https://github.com/mailhexu/TB2J. (Accessed 25 May 2026).

[71] Google, Agent Development Kit (ADK): An open-source, code-first toolkit for building, evaluating, and deploying sophisticated AI agents, 2025, https://github.com/google/adk-python. (Accessed 25 May 2026).

[72] D. Soria Parra, J. Spahr-Summers, Model Context Protocol (MCP): Specification and documentation, 2024, https://github.com/modelcontextprotocol/modelcontextprotocol. (Accessed 25 May 2026).

[73] A.H. Larsen, J.J. Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, C. Hargus, et al., The atomic simulation environment — A Python library for working with atoms, J. Phys.: Condens. Matter. 29 (27) (2017) 273002, http://dx.doi.org/10.1088/1361-648X/aa680e.

[74] J. Patterson, B. Bailey, Solid-State Physics, Springer, Berlin, Germany, 2016.

[75] M. Cyrot, Ginzburg-Landau theory for superconductors, Rep. Progr. Phys. 36 (2) (1973) 103, http://dx.doi.org/10.1088/0034-4885/36/2/001.

[76] W. Zhong, D. Vanderbilt, K.M. Rabe, First-principles theory of ferroelectric phase transitions for perovskites: The case of $BaTiO_3$, Phys. Rev. B 52 (1995) 6301–6312, http://dx.doi.org/10.1103/PhysRevB.52.6301, URL https://link.aps.org/doi/10.1103/PhysRevB.52.6301.

[77] L.-Q. Chen, Phase-field models for microstructure evolution, Annu. Rev. Mater. Res. 32 (1) (2002) 113–140, http://dx.doi.org/10.1146/annurev.matsci.32.112001.132041.

[78] I. Thomas, S. Clark, T. Lancaster, Exchange constants in molecule-based magnets derived from density functional methods, Phys. Rev. B 96 (9) (2017) 094403, http://dx.doi.org/10.1103/PhysRevB.96.094403.

[79] J.G. Aldea, D.L. Esteras, S. Roche, J.H. Garcia, Challenges and strategies for first-principles simulations of two-dimensional magnetic phenomena, Nanoscale 17 (43) (2025) 24955–24989, http://dx.doi.org/10.1039/D4NR05503A.

[80] W.M.C. Foulkes, R. Haydock, Tight-binding models and density-functional theory, Phys. Rev. B 39 (17) (1989) 12520, http://dx.doi.org/10.1103/PhysRevB.39.12520.

[81] P. Garcia-Fernandez, J.C. Wojdeł, J. Íñiguez, J. Junquera, Second-principles method for materials simulations including electron and lattice degrees of freedom, Phys. Rev. B 93 (19) (2016) 195137, http://dx.doi.org/10.1103/PhysRevB.93.195137.

[82] J.C. Wojdeł, P. Hermet, M.P. Ljungberg, P. Ghosez, J. Íñiguez, First-principles model potentials for lattice-dynamical studies: general methodology and example of application to ferroic perovskite oxides, J. Phys.: Condens. Matter. 25 (30) (2013) 305401, http://dx.doi.org/10.1088/0953-8984/25/30/305401.

[83] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza, D. Vanderbilt, Maximally localized wannier functions: Theory and applications, Rev. Modern Phys. 84 (4) (2012) 1419–1475, http://dx.doi.org/10.1103/RevModPhys.84.1419.

[84] C. Escorihuela-Sayalero, J.C. Wojdeł, J. Íñiguez, Efficient systematic scheme to construct second-principles lattice dynamical models, Phys. Rev. B 95 (9) (2017) 094115, http://dx.doi.org/10.1103/PhysRevB.95.094115.

[85] N. Carral-Sainz, T. Fernández-Ruiz, J. Íñiguez González, J. Junquera, P. García-Fernández, Systematic generation of electron models for second-principles density functional theory methods, Phys. Rev. B 112 (15) (2025) 155146, http://dx.doi.org/10.1103/jv9g-fymp.

[86] A.A. Mostofi, J.R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, N. Marzari, An updated version of wannier90: A tool for obtaining maximally-localised Wannier functions, Comput. Phys. Comm. 185 (8) (2014) 2309–2310, http://dx.doi.org/10.1016/j.cpc.2014.05.003.

[87] S. Das, Y.L. Tang, Z. Hong, M.A.P. Gonçalves, M.R. McCarter, C. Klewe, K.X. Nguyen, F. Gómez-Ortiz, P. Shafer, E. Arenholz, V.A. Stoica, S.-L. Hsu, B. Wang, C. Ophus, J.F. Liu, C.T. Nelson, S. Saremi, B. Prasad, A.B. Mei, D.G. Schlom, J. Íñiguez, P. García-Fernández, D.A. Muller, L.Q. Chen, J. Junquera, L.W. Martin, R. Ramesh, Observation of room-temperature polar skyrmions, Nature 568 (7752) (2019) 368–372, http://dx.doi.org/10.1038/s41586-019-1092-8.

[88] I. Sanchez-Movellan, M. Moreno, J.A. Aramburu, P. Garcia-Fernandez, Strain-induced ferromagnetic to antiferromagnetic crossover in d^9-ion (Cu^{2+} and Ag^{2+})-layered perovskites, J. Phys. Chem. C 127 (17) (2023) 8332–8341, http://dx.doi.org/10.1021/acs.jpcc.3c01166.

[89] T. Fernández-Ruiz, J. Íñiguez González, J. Junquera, P. García-Fernández, Extension of second-principles density functional theory into the time domain, Phys. Rev. B 112 (23) (2025) 235109, http://dx.doi.org/10.1103/pc8l-xkq3.

[90] D. Paul, E. Trukhan, N.H. Protik, Ab initio study of coulomb drag driven electron-hole bifluidity in doped graphene, Phys. Rev. B (2026) http://dx.doi.org/10.1103/q4mx-9tmn, https://link.aps.org/doi/10.1103/q4mx-9tmn.

[91] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864–B871, http://dx.doi.org/10.1103/PhysRev.136.B864, URL https://link.aps.org/doi/10.1103/PhysRev.136.B864.

[92] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138, http://dx.doi.org/10.1103/PhysRev.140.A1133, URL https://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[93] S. Baroni, P. Giannozzi, A. Testa, Green's-function approach to linear response in solids, Phys. Rev. Lett. 58 (1987) 1861–1864, http://dx.doi.org/10.1103/PhysRevLett.58.1861, URL https://link.aps.org/doi/10.1103/PhysRevLett.58.1861.

[94] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza, D. Vanderbilt, Maximally localized wannier functions: Theory and applications, Rev. Modern Phys. 84 (2012) 1419–1475, http://dx.doi.org/10.1103/RevModPhys.84.1419, URL https://link.aps.org/doi/10.1103/RevModPhys.84.1419.

[95] J. Noffsinger, F. Giustino, B.D. Malone, C.-H. Park, S.G. Louie, M.L. Cohen, EPW: A program for calculating the electron–phonon coupling using maximally localized Wannier functions, Comput. Phys. Comm. 181 (12) (2010) 2140–2148, http://dx.doi.org/10.1016/j.cpc.2010.08.027.

[96] J.-J. Zhou, J. Park, I.-T. Lu, I. Maliyov, X. Tong, M. Bernardi, Perturbo: A software package for ab initio electron–phonon interactions, charge transport and ultrafast dynamics, Comput. Phys. Comm. 264 (2021) 107970, http://dx.doi.org/10.1016/j.cpc.2021.107970.

[97] N.H. Protik, C. Li, M. Pruneda, D. Broido, P. Ordejón, The elphbolt ab initio solver for the coupled electron-phonon Boltzmann transport equations, Npj Comput. Mater. 8 (1) (2022) 28, http://dx.doi.org/10.1038/s41524-022-00710-0.

[98] A. Cepellotti, J. Coulter, A. Johansson, N.S. Fedorova, B. Kozinsky, Phoebe: A high-performance framework for solving phonon and electron Boltzmann transport equations, J. Phys.: Mater. 5 (3) (2022) 035003, http://dx.doi.org/10.1088/2515-7639/ac86f6.

[99] W. Li, J. Carrete, N. A. Katcho, N. Mingo, ShengBTE: A solver of the Boltzmann transport equation for phonons, Comput. Phys. Comm. 185 (6) (2014) 1747–1758, http://dx.doi.org/10.1016/j.cpc.2014.02.015.

[100] J. Carrete, B. Vermeersch, A. Katre, A. van Roekeghem, T. Wang, G.K. Madsen, N. Mingo, almaBTE : A solver of the space-time dependent Boltzmann transport equation for phonons in structured materials, Comput. Phys. Comm. 220 (2017) 351–362, http://dx.doi.org/10.1016/j.cpc.2017.06.023.

[101] A. Togo, L. Chaput, T. Tadano, I. Tanaka, Implementation strategies in phonopy and phono3py, J. Phys.: Condens. Matter. 35 (35) (2023) 353001, http://dx.doi.org/10.1088/1361-648X/acd831.

[102] Z. Han, X. Yang, W. Li, T. Feng, X. Ruan, FourPhonon: An extension module to shengbte for computing four-phonon scattering rates and thermal conductivity, Comput. Phys. Comm. 270 (2022) 108179, http://dx.doi.org/10.1016/j.cpc.2021.108179.

[103] F. Bloch, Zum elektrischen Widerstandsgesetz bei tiefen Temperaturen, Z. Phys. 59 (3) (1930) 208–214, http://dx.doi.org/10.1007/BF01341426.

[104] R. Peierls, Zur Theorie der elektrischen und thermischen Leitfähigkeit von Metallen, Ann. Phys., Lpz. 396 (2) (1930) 121–148, http://dx.doi.org/10.1002/andp.19303960202, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19303960202, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19303960202.

[105] Y. Gurevich, O. Mashkevich, The electron-phonon drag and transport phenomena in semiconductors, Phys. Rep. 181 (6) (1989) 327–394, http://dx.doi.org/10.1016/0370-1573(89)90011-2.

[106] H.P.R. Frederikse, Thermoelectric power of germanium below room temperature, Phys. Rev. 92 (1953) 248–252, http://dx.doi.org/10.1103/PhysRev.92.248, URL https://link.aps.org/doi/10.1103/PhysRev.92.248.

[107] T.H. Geballe, G.W. Hull, Seebeck effect in Germanium, Phys. Rev. 94 (1954) 1134–1140, http://dx.doi.org/10.1103/PhysRev.94.1134, URL https://link.aps.org/doi/10.1103/PhysRev.94.1134.

[108] T.H. Geballe, G.W. Hull, Seebeck effect in silicon, Phys. Rev. 98 (1955) 940–947, http://dx.doi.org/10.1103/PhysRev.98.940, URL https://link.aps.org/doi/10.1103/PhysRev.98.940.

[109] J. Zhou, B. Liao, B. Qiu, S. Huberman, K. Esfarjani, M.S. Dresselhaus, G. Chen, Ab initio optimization of phonon drag effect for lower-temperature thermoelectric energy conversion, Proc. Natl. Acad. Sci. 112 (48) (2015) 14777–14782, http://dx.doi.org/10.1073/pnas.1512328112.

[110] M. Fiorentini, N. Bonini, Thermoelectric coefficients of *n*-doped silicon from first principles via the solution of the Boltzmann transport equation, Phys. Rev. B 94 (2016) 085204, http://dx.doi.org/10.1103/PhysRevB.94.085204, URL https://link.aps.org/doi/10.1103/PhysRevB.94.085204.

[111] F. Macheda, N. Bonini, Magnetotransport phenomena in *p*-doped diamond from first principles, Phys. Rev. B 98 (2018) 201201, http://dx.doi.org/10.1103/PhysRevB.98.201201, URL https://link.aps.org/doi/10.1103/PhysRevB.98.201201.

[112] N.H. Protik, B. Kozinsky, Electron-phonon drag enhancement of transport properties from a fully coupled ab initio Boltzmann formalism, Phys. Rev. B 102 (2020) 245202, http://dx.doi.org/10.1103/PhysRevB.102.245202, URL https://link.aps.org/doi/10.1103/PhysRevB.102.245202.

[113] M. Raya-Moreno, R. Rurali, X. Cartoixà, Degradation of the ZT thermoelectric figure of merit in silicon when nanostructuring: From bulk to nanowires, Int. J. Heat Mass Transfer 225 (2024) 125385, http://dx.doi.org/10.1016/j.ijheatmasstransfer.2024.125385, URL https://www.sciencedirect.com/science/article/pii/S0017931024002163.

[114] C. Li, N. Protik, P. Ordejón, D. Broido, Colossal phonon drag enhanced thermopower in lightly doped diamond, Mater. Today Phys. 27 (2022) 100740, http://dx.doi.org/10.1016/j.mtphys.2022.100740, URL https://www.sciencedirect.com/science/article/pii/S2542529322001389.

[115] C. Li, N.H. Protik, N.K. Ravichandran, D. Broido, High-frequency phonons drive large phonon-drag thermopower in semiconductors at high carrier density, Phys. Rev. B 107 (2023) L081202, http://dx.doi.org/10.1103/PhysRevB.107.L081202, URL https://link.aps.org/doi/10.1103/PhysRevB.107.L081202.

[116] Y. Quan, Y. Chen, B. Liao, Significant phonon drag effect in wide band gap GaN and AlN, Phys. Rev. B 107 (2023) 245202, http://dx.doi.org/10.1103/PhysRevB.107.245202, URL https://link.aps.org/doi/10.1103/PhysRevB.107.245202.

[117] Y. Quan, B. Liao, Electron drag effect on thermal conductivity in two-dimensional semiconductors, Nano Lett. 24 (26) (2024) 8143–8150, http://dx.doi.org/10.1021/acs.nanolett.4c02097.

[118] Y. Quan, B. Liao, Coupled electron-phonon hydrodynamics in two-dimensional semiconductors, Phys. Rev. Lett. 134 (2025) 226301, http://dx.doi.org/10.1103/PhysRevLett.134.226301, URL https://link.aps.org/doi/10.1103/PhysRevLett.134.226301.

[119] L.A. Ponomarenko, A. Principi, A.D. Niblett, W. Wang, R.V. Gorbachev, P. Kumaravadivel, A.I. Berdyugin, A.V. Ermakov, S. Slizovskiy, K. Watanabe, T. Taniguchi, Q. Ge, V.I. Fal'ko, L. Eaves, M.T. Greenaway, A.K. Geim, Extreme electron–hole drag and negative mobility in the Dirac plasma of graphene, Nat. Commun. 15 (1) (2024) 9869, http://dx.doi.org/10.1038/s41467-024-54198-x.

[120] M. Simoncelli, N. Marzari, F. Mauri, Unified theory of thermal transport in crystals and glasses, Nat. Phys. 15 (8) (2019) 809–813, http://dx.doi.org/10.1038/s41567-019-0520-x.

[121] M. Simoncelli, N. Marzari, F. Mauri, Wigner formulation of thermal transport in solids, Phys. Rev. X 12 (2022) 041011, http://dx.doi.org/10.1103/PhysRevX.12.041011, URL https://link.aps.org/doi/10.1103/PhysRevX.12.041011.

[122] A. Cepellotti, B. Kozinsky, Interband tunneling effects on materials transport properties using the first principles Wigner distribution, Mater. Today Phys. 19 (2021) 100412, http://dx.doi.org/10.1016/j.mtphys.2021.100412.

[123] E. Trukhan, N.H. Protik, Interband and kinetic corrections to the electronic Boltzmann transport equation, 2025, arXiv:2512.04835, URL https://arxiv.org/abs/2512.04835.

[124] S. Peng, D. Pinkston, J. Yao, S. Kliavinek, I. Maliyov, M. Bernardi, Efficient GPU parallelization of electronic transport and nonequilibrium dynamics from electron-phonon interactions in the perturbo code, 2025, arXiv:2511.03683, URL https://arxiv.org/abs/2511.03683.

[125] H. Montes-Campos, J. Carrete, S. Bichelmaier, L.M. Varela, G.K.H. Madsen, A differentiable neural-network force field for ionic liquids, J. Chem. Inf. Model. 62 (1) (2022) 88–101, http://dx.doi.org/10.1021/acs.jcim.1c01380.

[126] C. Kittel, Introduction to Solid State Physics, vol. 7, John Wiley & Sons, Inc. 1996.

[127] N.W. Ashcroft, N.D. Mermin, Solid State Physics, Thomson Learning, 1976.

[128] J.M. Ziman, Electrons and Phonons: the Theory of Transport Phenomena in Solids, Oxford University Press, 1960, http://dx.doi.org/10.1093/acprof:oso/9780198507796.001.0001.

[129] G.P. Srivastava, Physics of Phonons, CRC Press, 1990.

[130] R.E. Peierls, Quantum Theory of Solids, Oxford University Press, 2001.

[131] M.T. Dove, Introduction to Lattice Dynamics, Cambridge University Press, Cambridge, 1993.

[132] X. Gonze, J.-P. Vigneron, Density-functional approach to nonlinear-response coefficients of solids, Phys. Rev. B 39 (1989) 13120–13128, http://dx.doi.org/10.1103/PhysRevB.39.13120.

[133] P. Giannozzi, S. de Gironcoli, P. Pavone, S. Baroni, Ab initio calculation of phonon dispersions in semiconductors, Phys. Rev. B 43 (1991) 7231–7242, http://dx.doi.org/10.1103/PhysRevB.43.7231.

[134] S.Y. Savrasov, Linear-response theory and lattice dynamics: A muffin-tin-orbital approach, Phys. Rev. B 54 (1996) 16470–16486, http://dx.doi.org/10.1103/PhysRevB.54.16470, URL https://link.aps.org/doi/10.1103/PhysRevB.54.16470.

[135] X. Gonze, C. Lee, Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory, Phys. Rev. B 55 (1997) 10355–10368, http://dx.doi.org/10.1103/PhysRevB.55.10355.

[136] S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory, Rev. Modern Phys. 73 (2001) 515–562, http://dx.doi.org/10.1103/RevModPhys.73.515, URL https://link.aps.org/doi/10.1103/RevModPhys.73.515.

[137] L. Paulatto, F. Mauri, M. Lazzeri, Anharmonic properties from a generalized third-order ab initio approach: Theory and applications to graphite and graphene, Phys. Rev. B 87 (2013) 214303, http://dx.doi.org/10.1103/PhysRevB.87.214303, URL https://link.aps.org/doi/10.1103/PhysRevB.87.214303.

[138] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, D. Allan, First-principles computation of material properties: The ABINIT software project, Comput. Mater. Sci. 25 (3) (2002) 478–492, http://dx.doi.org/10.1016/s0927-0256(02)00325-7.

[139] K. Parlinski, Z.Q. Li, Y. Kawazoe, First-principles determination of the soft mode in cubic ZrO_2, Phys. Rev. Lett. 78 (1997) 4063–4066, http://dx.doi.org/10.1103/PhysRevLett.78.4063.

[140] D. Alfè, PHON: A program to calculate phonons using the small displacement method, Comput. Phys. Comm. 180 (12) (2009) 2622–2633, http://dx.doi.org/10.1016/j.cpc.2009.03.010.

[141] T. Tadano, Y. Gohda, S. Tsuneyuki, Anharmonic force constants extracted from first-principles molecular dynamics: Applications to heat transfer simulations, J. Phys.: Condens. Matter. 26 (22) (2014) 225402, http://dx.doi.org/10.1088/0953-8984/26/22/225402.

[142] T. Tadano, S. Tsuneyuki, Self-consistent phonon calculations of lattice dynamical properties in cubic $SrTiO_3$ with first-principles anharmonic force constants, Phys. Rev. B 92 (2015) 054301, http://dx.doi.org/10.1103/PhysRevB.92.054301.

[143] Y. Oba, T. Tadano, R. Akashi, S. Tsuneyuki, First-principles study of phonon anharmonicity and negative thermal expansion in ScF_3, Phys. Rev. Mater. 3 (2019) 033601, http://dx.doi.org/10.1103/PhysRevMaterials.3.033601.

[144] Y. Wang, L.-Q. Chen, Z.-K. Liu, YPHON: A package for calculating phonons of polar materials, Comput. Phys. Comm. 185 (11) (2014) 2950–2968, http://dx.doi.org/10.1016/j.cpc.2014.06.023.

[145] O. Hellman, I.A. Abrikosov, S.I. Simak, Lattice dynamics of anharmonic solids from first principles, Phys. Rev. B 84 (2011) 180301, http://dx.doi.org/10.1103/PhysRevB.84.180301.

[146] O. Hellman, I.A. Abrikosov, Temperature-dependent effective third-order interatomic force constants from first principles, Phys. Rev. B 88 (2013) 144301, http://dx.doi.org/10.1103/PhysRevB.88.144301.

[147] O. Hellman, P. Steneteg, I.A. Abrikosov, S.I. Simak, Temperature dependent effective potential method for accurate free energy calculations of solids, Phys. Rev. B 87 (2013) 104111, http://dx.doi.org/10.1103/PhysRevB.87.104111.

[148] A. Togo, First-principles phonon calculations with phonopy and phono3py, J. Phys. Soc. Japan 92 (1) (2023) 012001, http://dx.doi.org/10.7566/JPSJ.92.012001.

[149] A. Chernatynskiy, S.R. Phillpot, Phonon transport simulator (PhonTS), Comput. Phys. Comm. 192 (2015) 196–204, http://dx.doi.org/10.1016/j.cpc.2015.01.008.

[150] A. Togo, L. Chaput, I. Tanaka, Distributions of phonon lifetimes in Brillouin zones, Phys. Rev. B 91 (2015) 094306, http://dx.doi.org/10.1103/PhysRevB.91.094306, URL https://link.aps.org/doi/10.1103/PhysRevB.91.094306.

[151] N.H. Protik, A. Katre, L. Lindsay, J. Carrete, N. Mingo, D. Broido, Phonon thermal transport in 2H, 4H and 6H silicon carbide from first principles, Mater. Today Phys. 1 (2017) 31–38, http://dx.doi.org/10.1016/j.mtphys.2017.05.004, URL https://www.sciencedirect.com/science/article/pii/S2542529317300743.

[152] A. Katre, J. Carrete, N. Mingo, Unraveling the dominant phonon scattering mechanism in the thermoelectric compound ZrNiSn, J. Mater. Chem. A 4 (2016) 15940–15944, http://dx.doi.org/10.1039/C6TA05868J.

[153] B. Dongre, J. Carrete, A. Katre, N. Mingo, G.K.H. Madsen, Resonant phonon scattering in semiconductors, J. Mater. Chem. C 6 (2018) 4691–4697, http://dx.doi.org/10.1039/C8TC00820E.

[154] Z.-Y. Ong, Tutorial: Concepts and numerical techniques for modeling individual phonon transmission at interfaces, J. Appl. Phys. 124 (15) (2018) 151101, http://dx.doi.org/10.1063/1.5048234, arXiv:https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/1.5048234/19911791/151101_1_1.5048234.pdf.

[155] A. Kundu, X. Yang, J. Ma, T. Feng, J. Carrete, X. Ruan, G.K.H. Madsen, W. Li, Ultrahigh thermal conductivity of θ-phase tantalum nitride, Phys. Rev. Lett. 126 (2021) 115901, http://dx.doi.org/10.1103/PhysRevLett.126.115901, URL https://link.aps.org/doi/10.1103/PhysRevLett.126.115901.

[156] S. Bichelmaier, J. Carrete, M. Nelhiebel, G.K.H. Madsen, Accurate first-principles treatment of the high-temperature cubic phase of hafnia, Phys. Status Solidi (RRL) – Rapid Res. Lett. 16 (10) (2022) 2100642, http://dx.doi.org/10.1002/pssr.202100642, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/pssr.202100642, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pssr.202100642.

[157] F. Zhou, W. Nielson, Y. Xia, V. Ozoliņš, Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations, Phys. Rev. Lett. 113 (2014) 185501, http://dx.doi.org/10.1103/PhysRevLett.113.185501.

[158] M. Ohnishi, T. Deng, P. Torres, Z. Xu, T. Tadano, H. Zhang, W. Nong, M. Hanai, Z. Tian, M. Hu, X. Ruan, R. Yoshida, T. Suzumura, L. Lindsay, A.J.H. McGaughey, T. Luo, K. Hippalgaonkar, J. Shiomi, Database and deep-learning scalability of anharmonic phonon properties by automated brute-force first-principles calculations, 2025, arXiv:2504.21245, URL https://arxiv.org/abs/2504.21245.

[159] F. Eriksson, E. Fransson, P. Erhart, The hiphive package for the extraction of high-order force constants by machine learning, Adv. Theory Simul. 2 (5) (2019) 1800184, http://dx.doi.org/10.1002/adts.201800184.

[160] J. Riebesell, R.E.A. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A.A. Lee, A. Jain, K.A. Persson, A framework to evaluate machine learning crystal stability predictions, Nat. Mach. Intell. 7 (6) (2025) 836–847, http://dx.doi.org/10.1038/s42256-025-01055-1.

[161] A. Loew, D. Sun, H.-C. Wang, S. Botti, M.A.L. Marques, Universal machine learning interatomic potentials are ready for phonons, Npj Comput. Mater. 11 (1) (2025) 178, http://dx.doi.org/10.1038/s41524-025-01650-1.

[162] B. Póta, P. Ahlawat, G. Csányi, M. Simoncelli, Thermal conductivity predictions with foundation atomistic models, 2025, arXiv:2408.00755, URL https://arxiv.org/abs/2408.00755.

[163] I. Batatia, et al., A foundation model for atomistic materials chemistry, J. Chem. Phys. 163 (2025) 184110, http://dx.doi.org/10.1063/5.0310984.

[164] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C.J. Bartel, G. Ceder, Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling, Nat. Mach. Intell. 5 (2023) 1031–1041, http://dx.doi.org/10.1038/s42256-023-00716-3.

[165] C. Chen, S.P. Ong, A universal graph deep learning interatomic potential for the periodic table, Nat. Comput. Sci. 2 (2022) 718–728, http://dx.doi.org/10.1038/s43588-022-00349-3.

[166] L. Pan, J. Carrete, Z. Wang, G.K.H. Madsen, Phonon transport in defect-laden bilayer janus PtSTe studied using neural-network force fields, J. Phys. Chem. C 128 (26) (2024) 11024–11032, http://dx.doi.org/10.1021/acs.jpcc.4c02454.
[167] I. Batatia, D.P. Kovács, G.N.C. Simm, C. Ortner, G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, 2023, arXiv:2206.07697, URL https://arxiv.org/abs/2206.07697.
[168] PyTorch Team, Pytorch, 2025, https://pytorch.org. (Accessed 24 December 2025).
[169] JAX Team, JAX, 2024, https://www.jax.dev. (Accessed 24 December 2025).
[170] A. Griewank, A. Walther, Evaluating Derivatives, second ed. Society for Industrial and Applied Mathematics, 2008, http://dx.doi.org/10.1137/1.9780898717761, URL https://epubs.siam.org/doi/abs/10.1137/1.9780898717761.
[171] S. Bichelmaier, J. Carrete, G.K.H. Madsen, Evaluating the efficiency of power-series expansions as model potentials for finite-temperature atomistic calculations, Int. J. Quantum Chem. 123 (11) (2023) e27095, http://dx.doi.org/10.1002/qua.27095, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.27095.
[172] B. Mortazavi, E.V. Podryabinkin, I.S. Novikov, T. Rabczuk, X. Zhuang, A.V. Shapeev, Accelerating first-principles estimation of thermal conductivity by machine-learning interatomic potentials: A MTP/ShengBTE solution, Comput. Phys. Comm. 258 (2021) 107583, http://dx.doi.org/10.1016/j.cpc.2020.107583, URL https://www.sciencedirect.com/science/article/pii/S0010465520302800.
[173] T. Kocabaş, M. Keçeli, Á. Vázquez-Mayagoitia, C. Sevik, Gaussian approximation potentials for accurate thermal properties of two-dimensional materials, Nanoscale 15 (2023) 8772–8780, http://dx.doi.org/10.1039/D3NR00399J.
[174] L. Reicht, L. Legenstein, S. Wieser, E. Zojer, Designing accurate moment tensor potentials for phonon-related properties of crystalline polymers, Molecules 29 (16) (2024) http://dx.doi.org/10.3390/molecules29163724, URL https://www.mdpi.com/1420-3049/29/16/3724.
[175] S.K. Jana, T. Thanasarnsurapong, A. Boonchun, Machine learning-guided discovery of multifunctional Nb2CX2 MXenes (X= S, Se, Te): Insights into mechanical properties and thermal conductivity, J. Mater. Chem. C 13 (2025) 8631–8640, http://dx.doi.org/10.1039/D4TC05430J.
[176] Y. Dou, K. Shimizu, J. Carrete, H. Fujioka, S. Watanabe, Machine-learning potential for phonon transport in AlN with defects in multiple charge states, Phys. Rev. Mater. 9 (2025) 034601, http://dx.doi.org/10.1103/PhysRevMaterials.9.034601, URL https://link.aps.org/doi/10.1103/PhysRevMaterials.9.034601.
[177] M. Uhrin, S.P. Huber, J. Yu, N. Marzari, G. Pizzi, Workflows in AiiDA: Engineering a high-throughput, event-based engine for robust and modular computational workflows, Comput. Mater. Sci. 187 (2021) 110086, http://dx.doi.org/10.1016/j.commatsci.2020.110086.
[178] A.M. Ganose, H. Sahasrabuddhe, M. Asta, K. Beck, T. Biswas, A. Bonkowski, J. Bustamante, X. Chen, Y. Chiang, D.C. Chrzan, J. Clary, O.A. Cohen, C. Ertural, M.C. Gallant, J. George, S. Gerits, R.E.A. Goodall, R.D. Guha, G. Hautier, M. Horton, T.J. Inizan, A.D. Kaplan, R.S. Kingsbury, M.C. Kuner, B. Li, X. Linn, M.J. McDermott, R.S. Mohanakrishnan, A.A. Naik, J.B. Neaton, S.M. Parmar, K.A. Persson, G. Petretto, T.A.R. Purcell, F. Ricci, B. Rich, J. Riebesell, G.-M. Rignanese, A.S. Rosen, M. Scheffler, J. Schmidt, J.-X. Shen, A. Sobolev, R. Sundararaman, C. Tezak, V. Trinquet, J.B. Varley, D. Vigil-Fowler, D. Wang, D. Waroquiers, M. Wen, H. Yang, H. Zheng, J. Zheng, Z. Zhu, A. Jain, Atomate2: Modular workflows for materials science, Digit. Discov. (2025) http://dx.doi.org/10.1039/D5DD00019J.
[179] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, APL Mater. 1 (1) (2013) 011002, http://dx.doi.org/10.1063/1.4812323.
[180] M. Scheidgen, L. Himanen, A.N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J.A. Márquez, S. Brockhauser, S. Brückner, L.M. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J.F. Rudzinski, N. Daelman, J.M. Pizarro, M. Kuban, C. Salazar, P. Ondračka, H.-J. Bungartz, C. Draxl, NOMAD: A distributed web-based platform for managing materials science research data, J. Open Source Softw. 8 (90) (2023) 5388, http://dx.doi.org/10.21105/joss.05388.
[181] A. Togo, MDR phonon calculation database, 2025, https://mdr.nims.go.jp/collections/d7aab932-8512-4b9a-b93d-b61f6e5e7019.
[182] A. Nandy, C. Duan, H.J. Kulik, Audacity of huge: Overcoming challenges of data scarcity and data quality for machine learning in computational materials discovery, Curr. Opin. Chem. Eng. 36 (100778) (2022) 100778, http://dx.doi.org/10.1016/j.coche.2021.100778.
[183] D. Wu, C.-T. Lin, J. Huang, Active learning for regression using greedy sampling, Inform. Sci. 474 (2019) 90–105, http://dx.doi.org/10.1016/j.ins.2018.09.060.
[184] Z. Liu, X. Jiang, H. Luo, W. Fang, J. Liu, D. Wu, Pool-based unsupervised active learning for regression using iterative representativeness-diversity maximization (iRDM), Pattern Recognit. Lett. 142 (2021) 11–19, http://dx.doi.org/10.1016/j.patrec.2020.11.019, URL https://www.sciencedirect.com/science/article/pii/S0167865520304323.
[185] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
[186] W. Cai, Y. Zhang, J. Zhou, Maximizing expected model change for active learning in regression, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 51–60.
[187] J. Goetz, A. Tewari, P. Zimmerman, Active learning for non-parametric regression using purely random trees, in: Advances in Neural Information Processing Systems, vol. 31, 2018.
[188] R. Burbidge, J.J. Rowland, R.D. King, Active learning for regression based on query by committee, in: H. Yin, P. Tino, E. Corchado, W. Byrne, X. Yao (Eds.), Intelligent Data Engineering and Automated Learning, IDEAL 2007, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 209–218.
[189] B. Shahriari, et al., Taking the human out of the loop: A review of Bayesian optimization, Proc. IEEE 104 (1) (2016) 148–175, http://dx.doi.org/10.1109/JPROC.2015.2494218.
[190] P.I. Frazier, A tutorial on Bayesian optimization, 2018, http://dx.doi.org/10.48550/arXiv.1807.02811, arXiv:1807.02811.
[191] S. Sippola, M. Todorović, E. Peltola, First-principles structure search study of 17-β-estradiol adsorption on graphene, ACS Omega 9 (32) (2024) 34684–34691, http://dx.doi.org/10.1021/acsomega.4c03485.
[192] J. Baumsteiger, L. Celiberti, P. Rinke, M. Todorović, C. Franchini, Exploring noncollinear magnetic energy landscapes with Bayesian optimization, Digit. Discov. 4 (2025) 1639–1650, http://dx.doi.org/10.1039/D4DD00402G.
[193] A. Jose, J.P.A. de Mendonça, E. Devijver, N. Jakse, V. Monbet, R. Poloni, Regression tree-based active learning, Data Min. Knowl. Discov. 38 (2) (2024) 420–460, http://dx.doi.org/10.1007/s10618-023-00951-7.
[194] M. Todorović, M.U. Gutmann, J. Corander, P. Rinke, Bayesian inference of atomistic structure in functional materials, Npj Comput. Mater. 5 (1) (2019) http://dx.doi.org/10.1038/s41524-019-0175-2.
[195] J. Järvi, P. Rinke, M. Todorović, Detecting stable adsorbates of (1S)-camphor on Cu(111) with Bayesian optimization, Beilstein J. Nanotechnology 11 (2020) 1577–1589, http://dx.doi.org/10.3762/bjnano.11.140.
[196] J. Järvi, B. Alldritt, O. Krejčí, M. Todorović, P. Liljeroth, P. Rinke, Integrating Bayesian inference with scanning probe experiments for robust identification of surface adsorbate configurations, Adv. Funct. Mater. 31 (32) (2021) 2010853, http://dx.doi.org/10.1002/adfm.202010853, arXiv:https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adfm.202010853, URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/adfm.202010853.
[197] A.T. Egger, L. Hörmann, A. Jeindl, M. Scherbela, V. Obersteiner, M. Todorović, P. Rinke, O.T. Hofmann, Charge transfer into organic thin films: A deeper insight through machine-learning-assisted structure search, Adv. Sci. 7 (15) (2020) 2000992, http://dx.doi.org/10.1002/advs.202000992, arXiv:https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/advs.202000992, URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/advs.202000992.
[198] A. Fangnon, M. Dvorak, V. Havu, M. Todorović, J. Li, P. Rinke, Protective coating interfaces for perovskite solar cell materials: A first-principles study, ACS Appl. Mater. Interfaces 14 (10) (2022) 12758–12765, http://dx.doi.org/10.1021/acsami.1c21785.
[199] L. Fang, E. Makkonen, M. Todorović, P. Rinke, X. Chen, Efficient amino acid conformer search with Bayesian optimization, J. Chem. Theory Comput. 17 (3) (2021) 1955–1966, http://dx.doi.org/10.1021/acs.jctc.0c00648.
[200] J. Li, F. Pan, G.-X. Zhang, Z. Liu, H. Dong, D. Wang, Z. Jiang, W. Ren, Z.-G. Ye, M. Todorović, P. Rinke, Structural disorder by octahedral tilting in inorganic halide perovskites: New insight with Bayesian optimization, Small Struct. 5 (11) (2024) 2400268, http://dx.doi.org/10.1002/sstr.202400268, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/sstr.202400268, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sstr.202400268.
[201] H. Homm, J. Laakso, P. Rinke, Efficient dataset generation for machine learning halide perovskite alloys, Phys. Rev. Mater. 9 (2025) 053802, http://dx.doi.org/10.1103/PhysRevMaterials.9.053802, URL https://link.aps.org/doi/10.1103/PhysRevMaterials.9.053802.
[202] K. Ghosh, M. Todorović, A. Vehtari, P. Rinke, Active learning of molecular data for task-specific objectives, J. Chem. Phys. 162 (1) (2025) 014103, http://dx.doi.org/10.1063/5.0229834.
[203] S. Ament, S. Daulton, D. Eriksson, M. Balandat, E. Bakshy, Unexpected improvements to expected improvement for Bayesian optimization, 2025, arXiv:2310.20708, URL https://arxiv.org/abs/2310.20708.
[204] Y. Zhang, M. Vaara, A. Alesafar, D.B. Nguyen, P. Silva, L. Koskelo, J. Ristolainen, M. Stosiek, J. Löfgren, J. Vapaavuori, P. Rinke, Data-efficient optimization of thermally-activated polymer actuators through machine learning, Mater. Des. 253 (2025) 113908, http://dx.doi.org/10.1016/j.matdes.2025.113908, URL https://www.sciencedirect.com/science/article/pii/S0264127525003284.
[205] R. Moriconi, M.P. Deisenroth, K.S. Sesh Kumar, High-dimensional Bayesian optimization using low-dimensional feature spaces, Mach. Learn. 109 (9) (2020) 1925–1943, http://dx.doi.org/10.1007/s10994-020-05899-z.
[206] H.C. Herbol, M. Poloczek, P. Clancy, Cost-effective materials discovery: Bayesian optimization across multiple information sources, Mater. Horiz. 7 (2020) 2113–2123, http://dx.doi.org/10.1039/D0MH00062K.
[207] Online documentation page, 2026, https://github.com/scikit-activeml/scikit-activeml. (Accessed 29 January 2026).

[208] A. Jose, E. Devijver, N. Jakse, R. Poloni, Informative training data for efficient property prediction in metal–organic frameworks by active learning, J. Am. Chem. Soc. 146 (9) (2024) 6134–6144, http://dx.doi.org/10.1021/jacs.3c13687, PMID: 38404041.

[209] D. Frenkel, B. Smit, Understanding Molecular Simulation: from Algorithms to Applications, Academic Press, San Diego, California, USA, 1996.

[210] D. Marx, J. Hutter, Ab Initio Molecular Dynamics, Cambridge University Press, Cambridge, England, 2009.

[211] A. Warshel, M. Levitt, Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. 103 (2) (1976) 227–249, http://dx.doi.org/10.1016/0022-2836(76)90311-9.

[212] D. Immel, R. Drautz, G. Sutmann, Adaptive-precision potentials for large-scale atomistic simulations, J. Chem. Phys. 162 (2025) 114119, http://dx.doi.org/10.1063/5.0245877.

[213] N. Felis, W. Dononelli, FALCON: Fast active learning for machine learning potentials in atomistic and ab initio molecular dynamics simulations, Npj Comput. Mater. (2025) http://dx.doi.org/10.1038/s41524-025-01897-8.

[214] J. Westermayr, P. Marquetand, Machine learning for electronically excited states of molecules, Chem. Rev. 121 (2019) 9873–9926, http://dx.doi.org/10.1021/acs.chemrev.0c00749.

[215] P.O. Dral, Quantum chemistry in the age of machine learning, J. Phys. Chem. Lett. 11 (2020) 2336–2347, http://dx.doi.org/10.1021/acs.jpclett.9b03664.

[216] J.C. Tully, Molecular dynamics with electronic transitions, J. Chem. Phys. 93 (1990) 1061–1071, http://dx.doi.org/10.1063/1.459170.

[217] Y. Hu, J. Chen, B. Wang, On the intrinsic ripples and negative thermal expansion of graphene, Carbon 95 (2015) 239–249, http://dx.doi.org/10.1016/j.carbon.2015.08.022.

[218] U. Monteverde, J. Pal, M.A. Migliorato, M. Missous, U. Bangert, R. Zan, R. Kashtiban, D. Powell, Under pressure: Control of strain, phonons and bandgap opening in rippled graphene, Carbon 91 (2015) 266–274, http://dx.doi.org/10.1016/j.carbon.2015.04.044.

[219] B. Smith, L. Lindsay, J. Kim, E. Ou, R. Huang, L. Shi, Phonon interaction with ripples and defects in thin layered molybdenum disulfide, Appl. Phys. Lett. 114 (22) (2019) 221902, http://dx.doi.org/10.1063/1.5099103.

[220] X. Gu, Y. Wei, X. Yin, B. Li, R. Yang, Colloquium: Phononic thermal properties of two-dimensional materials, Rev. Modern Phys. 90 (2018) 041002, http://dx.doi.org/10.1103/RevModPhys.90.041002.

[221] F. Grasselli, Investigating finite-size effects in molecular dynamics simulations of ion diffusion, heat transport, and thermal motion in superionic materials, J. Chem. Phys. 156 (13) (2022) 134705, http://dx.doi.org/10.1063/5.0087382.

[222] A. Mahata, M.A. Zaeem, Size effect in molecular dynamics simulation of nucleation process during solidification of pure metals: Investigating modified embedded atom method interatomic potentials, Modelling Simul. Mater. Sci. Eng. 27 (8) (2019) 085015, http://dx.doi.org/10.1088/1361-651X/ab4b36.

[223] S.H. Jamali, L. Wolff, T.M. Becker, A. Bardow, T.J.H. Vlugt, O.A. Moultos, Finite-size effects of binary mutual diffusion coefficients from molecular dynamics, J. Chem. Theory Comput. 14 (5) (2018) 2667–2677, http://dx.doi.org/10.1021/acs.jctc.8b00170.

[224] S.-F. Tsay, C.F. Liu, System-size effects in the molecular-dynamics simulation of metallic crystallization, Phys. Lett. A 192 (5) (1994) 374–378, http://dx.doi.org/10.1016/0375-9601(94)90222-4.

[225] A. Cammarata, M. Dašić, P. Nicolini, Integrating Newton's equations of motion in the reciprocal space, J. Chem. Phys. 161 (8) (2024) 084111, http://dx.doi.org/10.1063/5.0224108.

[226] PINDOL, 2026, The code is available free of charge at https://github.com/acammarat/pindol, Last retrieved 2026-01-21.

[227] A. Togo, I. Tanaka, First principles phonon calculations in materials science, Scr. Mater. 108 (2015) 1–5, http://dx.doi.org/10.1016/j.scriptamat.2015.07.021.

[228] W.G. Noid, Perspective: Coarse-grained models for biomolecular systems, J. Chem. Phys. 139 (2013) 090901, http://dx.doi.org/10.1063/1.4818908.

[229] M. Praprotnik, L. Delle Site, K. Kremer, Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly, J. Chem. Phys. 123 (22) (2005) 224106, http://dx.doi.org/10.1063/1.2132286.

[230] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N.E. Charron, G. de Fabritiis, F. Noé, C. Clementi, Machine learning of coarse-grained molecular dynamics force fields, ACS Cent. Sci. 5 (5) (2019) 755–767, http://dx.doi.org/10.1021/acscentsci.8b00913.

[231] M. Li, L. Wang, Z. Zheng, Coarse-grained machine learning potential for mesoscale multilayered graphene, Npj Comput. Mater. 11 (1) (2025) 374, http://dx.doi.org/10.1038/s41524-025-01849-2.

[232] A.B. Poma, A. Hinostroza Caldas, L.F. Cofas-Vargas, M.S. Jones, A.L. Ferguson, L. Medrano Sandonas, Recent advances in machine learning and coarse-grained potentials for biomolecular simulations, Biophys. J. 125 (2) (2026) 327–343, http://dx.doi.org/10.1016/j.bpj.2025.06.019.

[233] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. 99 (2002) 12562–12566, http://dx.doi.org/10.1073/pnas.202427399.

[234] A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method, Phys. Rev. Lett. 100 (2008) 020603, http://dx.doi.org/10.1103/PhysRevLett.100.020603.

[235] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1) (1999) 141–151, http://dx.doi.org/10.1016/S0009-2614(99)01123-9.

[236] L. Bonati, Y.-Y. Zhang, M. Parrinello, Neural networks-based variationally enhanced sampling, Proc. Natl. Acad. Sci. 116 (36) (2019) 17641–17647, http://dx.doi.org/10.1073/pnas.1907975116.

[237] D. Chandler, P.G. Wolynes, Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids, J. Chem. Phys. 74 (7) (1981) 4078–4095, http://dx.doi.org/10.1063/1.441588.

[238] M. Ceriotti, T.E. Markland, Efficient methods and practical guidelines for simulating isotope effects, J. Chem. Phys. 138 (1) (2013) 014112, http://dx.doi.org/10.1063/1.4772676.

[239] C. Li, G.A. Voth, Using machine learning to greatly accelerate path integral ab initio molecular dynamics, J. Chem. Theory Comput. 18 (2) (2022) 599–604, http://dx.doi.org/10.1021/acs.jctc.1c01085.

[240] A. Solovykh, N. Rybin, I. Novikov, A. Shapeev, Path-integral molecular dynamics with actively-trained and universal machine learning force fields, Comput. Phys. Comm. 319 (2026) 109902, http://dx.doi.org/10.1016/j.cpc.2025.109902.

[241] M.S. Chen, A. Robledo, C. Schäfer, K.Y. Han, C. Clementi, M.E. Tuckerman, Machine learning-accelerated path integral molecular dynamics simulations of reactive organic electrolytes, J. Chem. Phys. 163 (14) (2025) 144110, http://dx.doi.org/10.1063/5.0288380.

[242] P.J. Hoogerbrugge, J.M.V.A. Koelman, Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics, Europhys. Lett. 19 (3) (1992) 155, http://dx.doi.org/10.1209/0295-5075/19/3/001.

[243] P. Español, P. Warren, Statistical mechanics of dissipative particle dynamics, Europhys. Lett. 30 (4) (1995) 191, http://dx.doi.org/10.1209/0295-5075/30/4/001.

[244] E. Lavagnini, J.L. Cook, P.B. Warren, C.A. Hunter, Translation of chemical structure into dissipative particle dynamics parameters for simulation of surfactant self-assembly, J. Phys. Chem. B 125 (15) (2021) 3942–3952, http://dx.doi.org/10.1021/acs.jpcb.1c00480.

[245] S. Zhang, T. Ye, B. Jing, H. Yin, D. Pan, Non-Markovian intelligent dissipative particle dynamics integrated with machine learning for enhancing coarse-grained simulations, J. Comput. Phys. 542 (2025) 114356, http://dx.doi.org/10.1016/j.jcp.2025.114356.

[246] F. Sepehr, S.J. Paddison, Dissipative particle dynamics interaction parameters from ab initio calculations, Chem. Phys. Lett. 645 (2016) 20–26, http://dx.doi.org/10.1016/j.cplett.2015.12.032.