

Device-Scale Atomistic Simulations of Heat Transport in Advanced Field-Effect Transistors

Ke Xu,^{1,2,*} Gang Wang,¹ Ting Liang,² Yang Xiao,¹ Dongliang Ding,² Haichang Guo,² Xiang Gao,³ Lei Tong,⁴ Xi Wan,⁵ Gang Zhang,^{6,†} and Jianbin Xu^{2,‡}

¹*College of Physical Science and Technology, Bohai University, Jinzhou, P. R. China*

²*Department of Electronic Engineering and Materials Science and Technology Research Center, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, 999077, P. R. China*

³*CAS Key Laboratory of Mechanical Behavior and Design of Materials, Department of Modern Mechanics, University of Science and Technology of China, Hefei, Anhui 230027, China*

⁴*School of Integrated Circuits, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*

⁵*Engineering Research Center of IoT Technology Applications (Ministry of Education),*

School of Integrated Circuits, Jiangnan University, Wuxi, Jiangsu 214122, P.R. China

⁶*Yangtze Delta Region Academy in Jiaxing, Beijing Institute of Technology, JiaXing, 314019 China*

(Dated: November 26, 2025)

Self-heating in next-generation, high-power-density field-effect transistor limits performance and complicates fabrication, presenting significant challenges for accurate heat transport simulations. To address this, we introduce NEP-FET, a machine-learned framework for device-scale heat transport simulations of field-effect transistors. Built upon the neuroevolution potential, the model extends a subset of the OMat24 dataset through an active-learning workflow to generate a chemically diverse, interface-rich reference set. Coupled with the FETMOD structure generator module, NEP-FET can simulate realistic field-effect transistor geometries at sub-micrometer scales containing millions of atoms, and delivers atomistic predictions of temperature fields, per-atom heat flux, and thermal stress in device structures with high fidelity. This framework enables rapid estimation of device-level metrics, including heat-flux density and effective thermal conductivity. Our results reveal pronounced differences in temperature distribution between fin-type and gate-all-around transistor architectures. The framework closes a key gap in multiscale device modeling by combining near-quantum-mechanical accuracy with device-scale throughput, providing a systematic route to explore heat transport and thermo-mechanical coupling in advanced transistors.

I. INTRODUCTION

As device dimensions have been aggressively scaled in accordance with Moore’s law, transistor architectures have evolved from bipolar junction transistors to planar MOSFETs, followed by Finfield-effect transistor (FET) and gate-all-around (GAA) transistors [1, 2]. At nanometer scales, heat dissipation is governed almost entirely by conduction, rendering heat transport a primary bottleneck for further performance gains and, in advanced FET, a key driver of self-heating and reliability degradation [3, 4].

In this regime, Fourier’s law fails to capture the underlying physics of heat conduction [3, 5, 6], motivating microscopic approaches to heat transport. Ab-initio calculations can provide interatomic force constants, Hamiltonians, and vibrational eigenmodes, which, when combined with the Boltzmann transport equation (BTE), non-equilibrium Green’s function (NEGF) or molecular dynamics (MD) methods, yield detailed insight into phonon transport in low-dimensional and nanostructured materials [7–10]. However, NEGF methods scale poorly for large or complex systems and are less ef-

fective under strong anharmonicity or high-temperature conditions [11]. In comparison, BTE solvers provide higher computational efficiency and bridge macroscopic Fourier models with atomistic treatments [12, 13], though their computational cost remains substantial for realistic three-dimensional (3D) device geometries [14–16]. MD simulations, on the other hand, explicitly resolve atomic motion and naturally capture strong anharmonicity, enabling studies of diverse heat transport properties, ranging from size-dependent thermal conductivity [17, 18] to transport in low-dimensional materials [19, 20], and interfacial thermal conductance (ITC) [21–25]. However, the accuracy of MD depends critically on the interatomic potential, and MD simulations remain computationally demanding for large, heterogeneous, or multi-component systems. It is therefore difficult to identify a single method that can simultaneously capture quantum effects, anharmonic scattering, and realistic device geometries, motivating the development of multiscale frameworks that combine atomistic accuracy with continuum-scale efficiency [26].

To integrate these complementary strengths, multiscale thermal frameworks have been advanced. Hu et al. introduced GiftBTE [27], an open-source nongray phonon BTE solver that achieves high computational efficiency and supports fully 3D simulations of real materials and devices. Building on this foundation, they proposed a parameter-free BTE formulation based on first-

* kickhsu@gmail.com

† gangzhang2006@gmail.com

‡ jbxu@ee.cuhk.edu.hk

principles phonon properties [28], eliminating empirical parameters while maintaining quantitative agreement with experiments across diverse materials. By coupling first-principles nongray BTE with hydrodynamic heat equations (heat diffusion equations (HDE)), they realized multiscale thermal simulations for FET [8]. Gabourie et al. [7] proposed a physics-grounded, artificial intelligence (AI)-accelerated multiscale pipeline that links atomic-level material realism to chip-scale temperature prediction. Their workflow employs density functional theory (DFT)-level machine-learned potential (MLP) combined with MD to derive thermal conductivity, then uses Monte Carlo methods to obtain device-scale temperature maps, and finally applies AI models to centimeter-scale and device-level temperature measurements. This approach reduces ad hoc calibrations and corrects drift-diffusion biases in hotspot locations and peak temperatures. Together, these multiscale approaches highlight the central role of MD in quantitative FET-level thermal simulations.

Recently, MLP has emerged as a powerful solution to the accuracy-scale trade-off in MD. Foundation-style models such as MACE-MP-0 [30], CHGNet [31], M3GNet [32], DPA3 [33], and NEP89 [34] demonstrate broad chemical coverage, spanning multiple elements, phases, mixtures, and high-entropy alloys, while retaining near-quantum-mechanical accuracy. Combined with GPU-accelerated MD, these models now enable simulations with thousands to millions of atoms on commodity hardware. For instance, NEP can deliver high fidelity for systems with tens of millions of atoms on a single GPU, unlocking fully atomistic device-scale modeling [34].

Building on these advances, we first develop a specialized MLP tailored for transistor devices, achieving near-quantum-mechanical accuracy in multi-million-atom simulations on a single GeForce RTX 4090 card. Our workflow reaches sub-micrometer spatial scales, enabling fully atomistic modeling of individual transistors. By considering the full-frequency phonon excitations intrinsic to MD, we compare heat-flux driving schemes and temperature-response characteristics in realistic transistor geometries, revealing the coupling between device architecture, heat generation, and nanoscale transport. This capability enables direct, quantum-mechanical-level investigation of self-heating phenomena in nanoscale transistors, bridging atomistic physics with device-level thermal behavior.

II. RESULTS

NEP-FET framework: NEP-FET is a GPU-accelerated simulation workflow for FET built on Graphics Processing Units Molecular Dynamics (GPUMD) [35] and NEP [36]. The same paradigm has been successfully validated in other systems, such as liquid water with NEP-MB-pol [37]. The framework combines the DFT-level accuracy characteristic of NEP with high speed

and broad elemental coverage, enabling efficient simulations of complex materials and systems. To ensure broad coverage of device-relevant configurations for NEP-FET, training data were drawn from the structurally diverse OMat24. Since OMat24 contains limited explicit interfacial structures, additional interfacial data were incorporated, generated via AIMD and AL simulations (see Methods and Supplementary Materials S1, S2 for details). We use the enriched dataset to train NEP models tailored for FET systems (see Methods and Supplementary Materials S3 for details). Following model training, device geometries are constructed using our FET-MOD module, after which simulations are performed in GPUMD to evaluate the relevant material properties and device-level responses. To demonstrate the workflow, we consider a 3D silicon-based FET, adapted from Refs. [8, 38]. This example serves as a representative test case for benchmarking the efficiency and predictive capability of the NEP-FET framework. The overall pipeline is illustrated in Fig. 1a.

FET device models: We compare two distinct device architectures: FinFET and the emerging GAAFET, both of which are shown in Fig. 1b,c, with a particular focus on the cross-sectional view of the GAAFET. The device contains five representative interfaces: Si/SiO₂, Si/HfO₂, HfO₂/SiO₂, SiO₂/TiN, and HfO₂/TiN. The gate electrode is composed of titanium nitride (TiN), while the drain, source, and substrate are constructed from crystalline silicon (Si). shallow-trench isolation (STI) and other dielectric layers are SiO₂, with hafnium dioxide (HfO₂) used for the fin sidewall dielectric. For the FinFET, the source width W_1 is 8 nm, the fin height H_4 is 18 nm, the HfO₂ shell has a thickness of 1 nm, the gate width is 14 nm, and the gate-to-substrate distance is 2 nm. In the GAA device, each channel (fin) has a height of 4 nm and a pitch of 5 nm. The simulation domain has dimensions of $35 \times 22 \times 35$ nm³ and contains approximately 2.0×10^6 atoms. All models were built with our FETMOD module; additional sizes were generated by preserving these aspect ratios (see Methods and Supplementary Materials S4 for details).

NEP model: MLPs trained on first-principles data have become a key component of modern materials-by-design workflows, as they can be adapted to specific classes of materials while retaining broad transferability. Among the available MLP architectures, the NEP family achieves an attractive compromise between accuracy, speed, and accessible system size, and has been successfully applied to a wide variety of materials [34, 37]. In this work, we therefore employ NEP as our primary MLP framework.

The NEP formalism [36] represents the total energy as a sum of atomic contributions, $E = \sum_i U_i$, where each site energy U_i is obtained from a neural network that takes as input a set of local descriptors of the atomic environment. These descriptors are constructed from radial and angular basis functions built using Chebyshev and Legendre polynomials [36], and are confined to a finite cutoff radius, so that the computational effort grows lin-

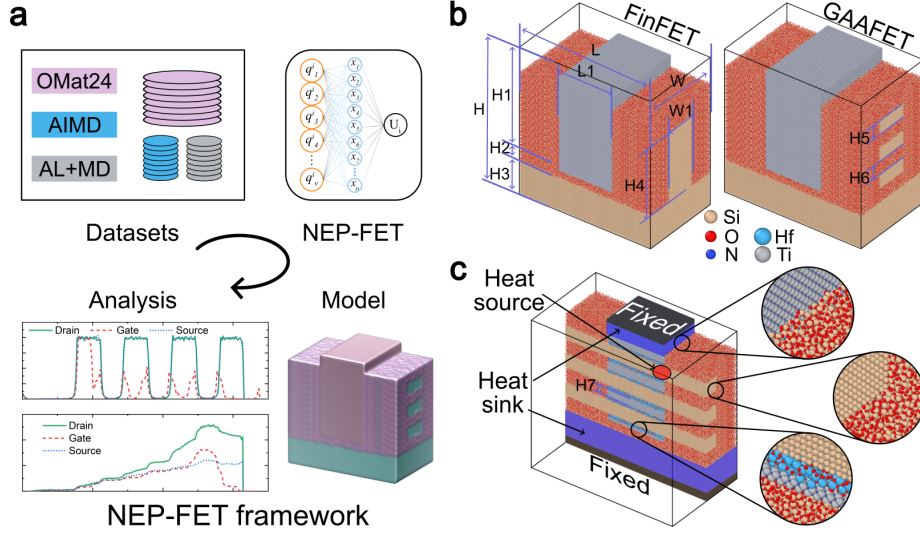


FIG. 1. **Overview of the NEP-FET framework and representative FET geometries.** (a) The NEP-FET framework is used for the simulation of device-level systems. The training set consists mainly of three components: (i) structures from OMat24 [29], (ii) configurations generated by AIMD, and (iii) configurations obtained via AL. These training sets are then processed and analyzed before being integrated into NEP for training. A realistic all-atom FET model is then generated using the FETMOD module. Finally, device-level property calculations are performed using GPU-accelerated MD software. (b) shows two typical FET structures established using FETMOD module: FinFET and GAAFET, with typical device dimensions marked. Different colored regions are used to group the atomic models to accommodate subsequent MD-related calculations. (c) A magnified view shows the detailed features of the device structure.

early with the number of atoms. The components of the descriptor vector serve as the input neurons of the network, and a single output neuron returns the scalar site energy U_i for the central atom i . The network parameters are optimized using a regularized training procedure based on the separable natural evolution strategy [41]. Forces and virial stresses are computed from the trained potential by taking analytical derivatives of the total energy with respect to atomic positions and strain. Chemical species are encoded through the expansion coefficients associated with a set of radial basis functions: for each ordered pair of species, the model introduces an independent set of N_{ec} expansion coefficients that are optimized during training, and each species is further assigned its own set of N_{nn} trainable neural network weights and biases. Detailed NEP parameters are provided in the Methods section and Supporting Information.

Dataset construction and coverage: Starting from OMat24, which contains periodic bulk structures for many semiconductors [29]. Then, we applied farthest-point sampling to obtain a compact subset. Because OMat24 contains only periodic bulk structures and thus few explicit interfaces, we complemented this dataset with additional device-relevant environments: six classes of hetero-junctions and five types of surface structures. Representative members of these families were sampled by AIMD and active-learning MD, additional interface and surface configurations after DFT labeling. Fig. 2a illustrates the overall dataset coverage: we perform principal component analysis (PCA) on the local-environment descriptors to visualize the distribution of atomic en-

vironments across structures. Point colors indicate the mean atomic energy for each configuration, demonstrating both diversity and continuity of local motifs. Typical device interfaces represented in the dataset correspond to those in Fig. 1b.

Model accuracy and validation: The trained NEP model for the FET systems (training protocol and hyperparameters in Methods and Supplementary Materials S1) achieves MAE relative to DFT references of 25.1 meV atom⁻¹ (energies), 287.5 meV Å⁻¹ (forces), and 0.687 GPa (stresses) (Fig. 2b). Fig. 2d shows the parity plot of the forces from DFT and NEP, while the corresponding parity plots for the energy, virials, and stress are provided in the Supplementary Materials S6. These metrics are well within the accuracy range required for capturing anharmonic lattice dynamics and interfacial phonon scattering, key ingredients for predicting heat transport in disordered dielectric environments. To further validate the model's transferability beyond the training set, we benchmarked the NEP-FET model on the ITC between HfO₂ and SiO₂, a technologically relevant oxide-oxide interface (see Methods for details). As shown in Fig. 2c, the predicted thermal conductance agrees closely with previous work [39], demonstrating that the model not only reproduces static energy and force properties but also faithfully captures anharmonic interfacial heat transport. These results confirm that the NEP-FET model provides a reliable and computationally efficient surrogate for large-scale simulations of dielectric interfaces in advanced transistor architectures.

Computational performance: NEP-FET attains

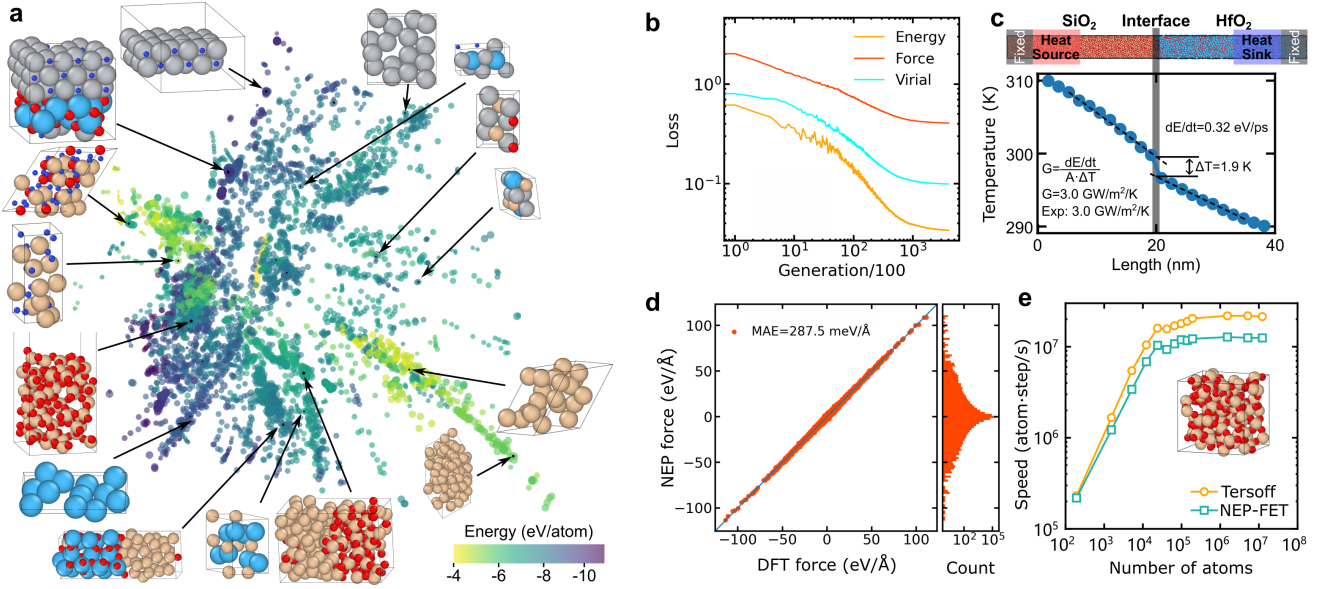


FIG. 2. Dataset composition for NEP-FET and training accuracy. (a) Distribution of the NEP-FET datasets in the reduced descriptor space spanned by the first two principal components. The color intensity indicates the average per-atom energy (in units of eV/atom) in the NEP-FET dataset, while the dot size represents the number of atoms contained in each individual training configuration. (b) Evolution of the training loss for the NEP-FET model. (c) shows the prediction of the interfacial heat transport properties of SiO₂ and HfO₂ materials by the NEP-FET model, demonstrating that the model can accurately capture the interfacial heat transport properties of the system [39]. (d) Final force-prediction accuracy of the force prediction (MAE = 287.5 meV Å⁻¹). (e) shows the benchmark results of computational efficiency of NEP-FET and traditional empirical potential model (Tersoff-SiO₂ [40]) in SiO₂ system.

empirical-potential-like speed while retaining near-DFT-level fidelity. In MD of amorphous SiO₂ structures (Fig. 2e), the model sustains 1.25×10^7 atom-steps s⁻¹ on a single RTX 4090 card (Fig. 2e). For comparison, a Tersoff-SiO₂ [40] potential achieves 2.14×10^7 atom-steps s⁻¹ on the same GPU. NEP-FET thus attains empirical-potential-like speed despite its more complex many-body descriptors and larger effective cutoff. The resulting speed-accuracy balance enables systematic exploration of heat transport in realistic FET geometries.

Temperature maps: Heat transport was evaluated using nonequilibrium MD (NEMD). Heat sources were positioned near the drain-side hotspot following finite-element pre-analysis predictions [8, 27, 28], with cold reservoirs defined at the top surface and at the substrate. Fixed boundary regions were introduced adjacent to the cold zones to ensure numerical stability and proper energy dissipation within the NEMD (see Methods for details). The resulting cross-sectional temperature distributions for the FinFET and GAAFET devices are shown in Fig. 3a,b, together with temperature line profiles extracted along the indicated cutlines. In the GAAFET, heat remains tightly confined to the channel region and decays sharply away from the hotspot. Pronounced temperature discontinuities are observed across dielectric interfaces, signaling significant interfacial thermal resistance. In contrast, the FinFET shows a more laterally extended and uniform temperature field. The smoother gradients highlight more efficient heat spread-

ing into the fin and stronger thermal coupling to the underlying substrate compared with the fully surrounded GAA geometry.

Effective thermal conductivity: An effective thermal conductivity, k_{eff} , was extracted once a steady-state temperature field was reached (procedure described in Methods). From Fig. 3e, we obtain $k_{\text{eff}} = 3.37 \text{ W m}^{-1} \text{ K}^{-1}$ for the FinFET and $2.13 \text{ W m}^{-1} \text{ K}^{-1}$ for the GAA device, meaning that the FinFET exhibits a 58.2% higher effective thermal conductivity. This enhancement arises from its larger proportion of high-conductivity crystalline Si along the primary heat-conduction paths and the improved structural continuity available for heat flow. Together, these factors enable more efficient thermal spreading and reduce the overall thermal resistance relative to the fully encapsulated GAA architecture.

Atomic heat flux analysis: From per-atom heat fluxes (find the details in the Methods), we construct microscopic heat-flow maps across the device. In Fig. 4 (labeled region A–F refer to the annotated regions), the GAA device shows peak heat flux concentrated near the hotspot, with dominant lateral channels towards regions A and B, especially towards the gate-proximal side (region A), where the TiN interconnects assist heat removal (region C). Region D contributes most of the downward flux into the substrate. Region (E), composed of amorphous SiO₂, exhibits strong phonon scattering; heat leakage through this path is minimal compared with crystalline Si (region F) and TiN (regions D and G). The

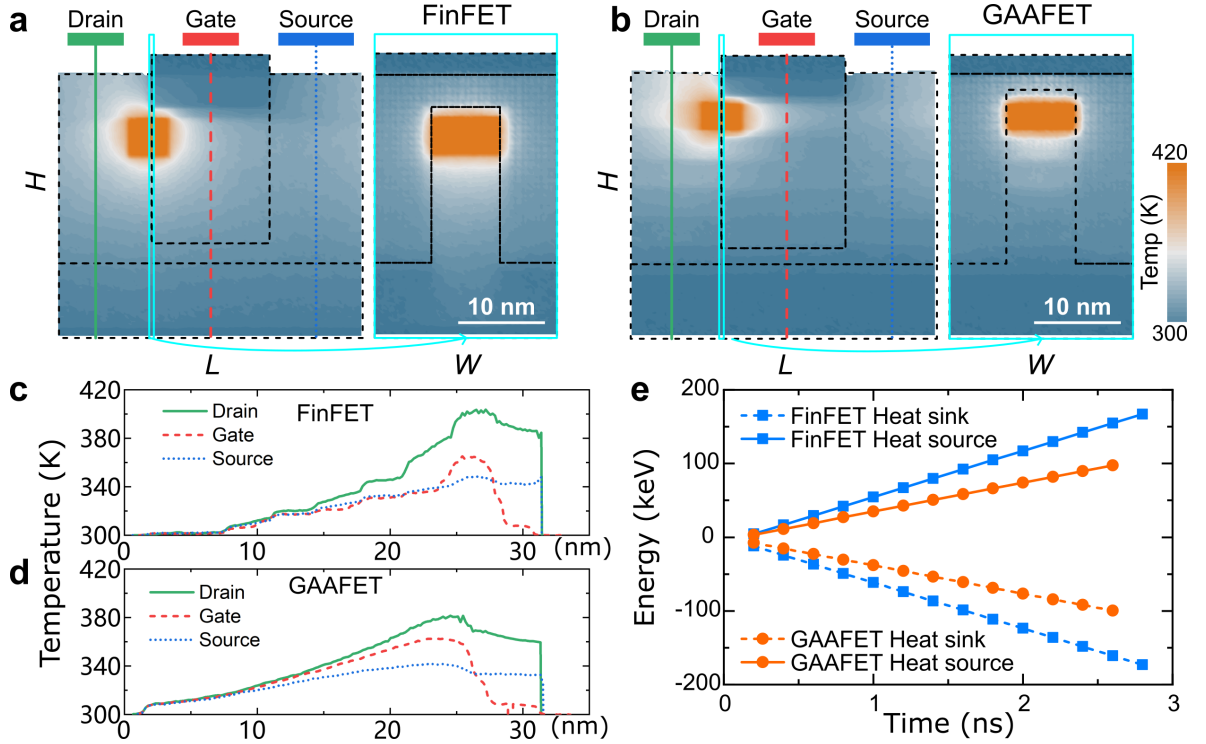


FIG. 3. **Thermal-property model of NEP-FET devices.** (a, b) Steady-state temperature maps for FinFET and GAAFET, respectively. Heat-source locations are defined in Fig. 1b,d. The sections are taken normal to the two device interfaces; light-blue boxes delineate the regions from which the special cross-sections are extracted. Color encodes local temperature (see color bar). Green solid, red dashed, and blue dotted traces label the source, gate and drain regions, respectively. Position-dependent temperature profiles are plotted for (c) FinFET and (d) GAAFET. (e) Heat-flow distributions in the heat-source and heat-sink regions for both devices.

left inset of Fig. 4a quantifies the z -directed flux beneath the hotspot. The substrate-directed component is substantially smaller in the GAA device ($\sim 8.52 \text{ keV ns}^{-1}$, $\sim 22.3\%$ of total) than in the FinFET (versus $\sim 39.1 \text{ keV ns}^{-1}$, $\sim 62.5\%$), consistent with nongray BTE trends [8]. Consequently, in our GAA model, the majority of heat dissipates laterally through the gate region, which motivates the use of high-thermal-conductivity gate stacks. In contrast, the FinFET exhibits a more substrate-dominated dissipation pathway, with pronounced heat flow from the fin base into the underlying Si and STI regions, leading to a deeper and more vertically extended hotspot (see Supplementary Materials Fig. S7 for details). The upper inset of Fig. 4a shows the flux across the shaded gate section: a large fraction is carried by TiN, while SiO₂ contributes negligibly. The lower inset of Fig. 4a reports the flux across the device core, revealing a comparatively uniform in-plane distribution away from interfaces. Analogous spatial patterns are observed in the FinFET, but with a larger substrate-coupled fraction and a broader hotspot footprint along the fin sidewalls.

Thermal stress: 4b presents the atomic von Mises stress [42] in the GAA device (see Methods for details), highlighting thermo-mechanical responses near the hotspot. Amorphous SiO₂ sustains the highest stresses, indicating significant built-in strain; HfO₂ near the gate

also shows pronounced stress accumulation, whereas crystalline Si remains comparatively low, consistent with its role as a structural backbone. The blue boxes in 4b mark cross-sectional stress maps through the highlighted regions, and the accompanying line plots compare stress across the drain, gate, and source. Stresses are consistently more concentrated in the gate region and peak near the hotspot, underscoring coupled thermal and mechanical reliability constraints.

III. CONCLUSIONS

We have introduced NEP-FET, a computational framework that bridges the critical gap between quantum-mechanical accuracy and device-scale thermal simulations for advanced transistors. By leveraging a targeted MLP trained on a compact yet representative dataset, this approach enables fully atomistic simulations of realistic FET geometries, encompassing millions of atoms, with DFT-level fidelity on a single GPU.

Our framework provides direct access to previously inaccessible atomistic fields, including spatially resolved temperature maps, per-atom heat flux, and thermal stress distributions. Applying NEP-FET to comparative studies of FinFET and GAA architectures reveals funda-

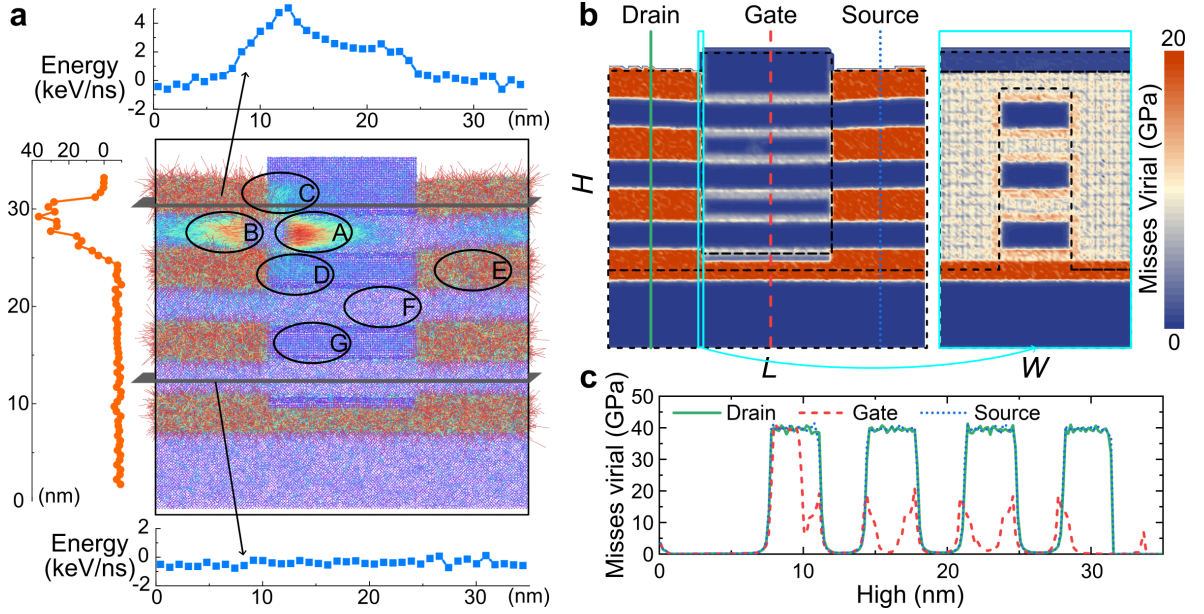


FIG. 4. **Atomic-scale analysis of GAAFET device properties.** (a) Per-atom heat-flux map of the device. Three dotted-line profiles show how the heat flux varies with position across the three regions: the left profile reports the global heat-flux-position relation for the entire device, whereas the upper and lower profiles give the corresponding relations for the grey-shaded subregions in the map. (b) Per-atom von Mises stress [42] distribution, with color indicating magnitude (see color bar). (c) Von Mises stress analysis of structures in different cross sections. Positional annotations within the corresponding regions use colors and line styles consistent with the profiles.

mental insights into nanoscale heat dissipation. We find that the GAA transistor exhibits more localized thermal hotspots and a significantly lower effective thermal conductivity compared to the FinFET, a consequence of its more confined architecture and greater interfacial density, which impedes efficient heat spreading towards the substrate. Atomic-scale analysis further elucidates the dominant heat-flow pathways and identifies regions of pronounced thermomechanical stress, particularly within amorphous gate dielectrics.

NEP-FET thus establishes a systematic and predictive pathway for investigating phonon-mediated heat transport and its coupling with mechanical response in advanced electronic devices. This capability is pivotal for addressing the escalating challenge of self-heating and for guiding the thermally aware design of future high-performance transistors.

IV. METHODS

This workflow combines a curated, interface-rich training database, DFT labeling, a fine-tuned NEP model, and large-scale NEMD on device geometries generated with the FETMOD module. The framework yields spatially resolved temperature, heat flux and stress fields, from which effective thermal conductivity and thermo-mechanical reliability metrics are extracted.

Training datasets. The devices studied here involve

five atomic species (O, Si, Hf, Ti and N). We first extracted 28 280 structures containing only these elements from the OMat24 dataset [29]. A preliminary NEP model was fitted to this raw subset and used in a two-step “train-predict-clean” protocol: for each structure we compared NEP-predicted and reference DFT energies and removed outliers with large discrepancies. After two iterations, this yielded 22 721 bulk configurations with consistent DFT and NEP energies. To obtain a compact and diverse bulk subset, we then applied farthest-point sampling (FPS) in descriptor space, selecting 4 000 representative bulk structures for model selection and hyperparameter tuning (the results for PCA are shown in Supplementary Materials S5). Because OMat24 [29] mainly contains periodic bulk phases and thus lacks explicit device-like non-bulk environments, we constructed a supplementary interface/surface dataset comprising six heterointerfaces (a-HfO₂/TiN, a-SiO₂/a-HfO₂, a-SiO₂/TiN, a-SiO₂/a-Si, a-Si/SiO₂, and a-Si/a-HfO₂) and five surface families (a-HfO₂, a-Si, a-SiO₂, TiN surfaces, and TiN nanowires), for a total of 11 classes of device-relevant local environments. Starting from the general-purpose NEP89 model [34], we fine-tuned a five-element NEP and used it to run NPT MD trajectories (0.5 fs time step, 1.5 ns length) for 11 representative interface/surface structures. For each trajectory, 50 frames were sampled at equal time intervals to reduce temporal correlations, and single-point DFT calculations (energies, forces and virials) were performed for all 550 configurations. After discarding a small number of pathological cases and adding

a few additional structures from targeted melt-quench tests, this yielded 616 interface and surface configurations with DFT labels. Finally, the bulk set of 4 000 configurations and the 616 supplementary interface/surface configurations were merged, followed by a second FPS step to define a compact training core. This yielded a final set of 4 616 fully validated configurations used for NEP-FET training, consistent with the dataset statistics reported in the main text.

DFT calculations. All reference single-point DFT calculations were carried out with the VASP package (version 6.3.0) using the projector-augmented-wave (PAW) method [43]. Exchange-correlation effects were treated within the generalized-gradient approximation using the Perdew–Burke–Ernzerhof (PBE) functional [44, 45]. A plane-wave kinetic-energy cutoff of 500 eV (`ENCUT` = 500) and the “Accurate” precision setting (`PREC` = `Accurate`) were employed to suppress aliasing errors. Brillouin-zone integrations used the automatic k -point generation scheme (`KSPACING` = 0.2, `KGAMMA` = `.TRUE.`), corresponding to a reciprocal-space sampling density of $\sim 0.2 \text{ \AA}^{-1}$. Electronic self-consistency was converged to 10^{-6} eV in total energy (`EDIFF` = 1×10^{-6}) and a maximum of 150 electronic iterations (`NELM` = 150) was allowed, which proved sufficient for robust convergence across the chemically and structurally diverse configurations in the training database. All calculations were performed in a non-spin-polarized setup (`ISPIN` = 1), with fixed ionic positions (`NSW` = 1, `IBRION` = -1). Partial occupancies were treated using the tetrahedron method with Blöchl corrections (`ISMEAR` = -5). Additional technical details of the reference data calculations are provided in Supplemental Notes S2 and S3.

NEP training. The NEP model for the five-component FET material system was trained using the fourth-generation NEP (NEP4) scheme [35, 36, 46]. Following the standard NEP formalism, the total energy is expressed as a sum of site energies U_i depending on atom-centered descriptors within a finite cutoff radius. In practice, we used a radial cutoff of 6.0 Å and an angular cutoff of 5.0 Å (`cutoff` 6 5), Chebyshev expansion orders $n_{\max} = 4$ for both radial and angular channels with 8 basis functions per channel (`n_max` 4 4, `basis_size` 8 8), and angular-momentum limits $l_{\max} = (4, 2, 1)$ for the multi-body channels (`l_max` 4 2 1). Short-range nuclear overlap was regularized by adding a Ziegler-Biersack-Littmark repulsive term smoothly truncated at 2 Å (`zbl` 2). (see Supplementary Materials S3 for details)

The atomic energy mapping was implemented as a fully connected feedforward neural network with a single hidden layer of 80 neurons (`neuron` 80) and hyperbolic-tangent activation. Energies, forces and virial stresses were fitted simultaneously by minimizing a composite loss function defined as a weighted sum of the root-mean-square errors (RMSEs) of energies (per atom), forces and virials, plus $\mathcal{L}_2$ regularization on the network parameters. We set equal weights for energies, forces and

virials ($\lambda_e = \lambda_f = \lambda_v = 1$; `lambda_e` 1, `lambda_f` 1, `lambda_v` 1), retaining the default $\mathcal{L}_2$ coefficient. Optimization used the separable natural evolution strategy built into NEP4 with a population size of 50 candidate parameter vectors per generation (`population` 80), and a maximum of 4×10^5 generations (`generation` 200000). The final NEP-FET model was chosen as the parameter set with the lowest validation loss and attains the energy, force and stress errors reported in Fig. 2.

Device construction and NEMD simulations.

Atomistic FinFET and GAAFET geometries were generated using the FETMOD module. The gate metal was TiN, while the drain, source and substrate were crystalline Si. STI and spacer layers were SiO_2 , and the fin sidewall dielectric was HfO_2 . For the FinFET, the source width (D_W) was 8 nm, the fin height 18 nm, the HfO_2 shell thickness 1 nm, the gate width 14 nm and the gate-substrate spacing 2 nm. For the GAA device, each channel (Fin) had a height of 4 nm and a pitch of 5 nm. The primary simulation box spanned $35 \times 22 \times 35 \text{ nm}^3$ and contained approximately 2.0×10^6 atoms; additional sizes were generated by scaling while preserving aspect ratios (see Supplementary S4 for details). Si substrates were oriented along (001) with channels aligned to $\langle 110 \rangle$.

Amorphous oxide blocks were generated via melt-quench MD and then relaxed using the NEP-FET potential. These pre-relaxed slabs were joined to the crystalline Si and TiN regions with 0.3–0.5 nm interfacial relaxation windows; residual atomic overlaps were removed by soft-core pushes followed by energy minimization. For selected individual interfaces (e.g., Si/ SiO_2 , $\text{SiO}_2/\text{HfO}_2$, HfO_2/TiN), we also built simplified hetero-structure cells to benchmark ITC against reference calculations (Sec. II).

All MD simulations were performed with GPUMD [35] using the NEP-FET model. Periodic boundary conditions were applied in the lateral directions. Along the transport direction, we included fixed boundary layers and Langevin thermostat regions to set up non-equilibrium heat flow. Each assembled device structure was first equilibrated in the isothermal-isobaric (NPT) ensemble at the target temperature and zero external pressure for 0.5 ns, followed by 0.25 ns of canonical (NVT) equilibration. The barostat was applied only along periodic directions so as not to distort free surfaces. The time step was 0.5 fs, which we verified to be sufficient for stable energy conservation and accurate heat-flux evaluation.

In the production NEMD runs (4.0 ns), two spatially separated local Langevin thermostats [20] defined a hot and a cold region along the cross-plane direction, imposing a temperature difference of 60 K. This setup generated a steady non-equilibrium state with approximately constant heat flux across the device. For simplified interface cells, the corresponding temperature drop ΔT across the interface was used to evaluate the ITC G via

$$G = \frac{dE/dt}{A\Delta T}, \quad (1)$$

where S is the interfacial cross-sectional area and dE/dt is the time-averaged energy exchange rate between thermostats and heat baths.

Temperature fields and effective thermal conductivity. Local temperature fields were obtained from atomic kinetic energies after removing bin-wise drift. The simulation cell was voxelized into cubic bins of edge length 0.5 nm. For each bin centered at position $\mathbf{r}$, the instantaneous kinetic temperature was computed from the atomic kinetic energies in that bin after subtracting the local center-of-mass velocity. Bins that overlapped with the Langevin thermostat regions were excluded from subsequent analyses.

Temperature maps were time-averaged over the steady-state window (typically 20 snapshots uniformly sampled over the last 4 ns). For visualization, the fields were lightly smoothed using a Gaussian kernel with width $\sigma = 0.5$ nm; all quantitative values reported in the main text were extracted from the raw data. One-dimensional temperature profiles were obtained along annotated cut-lines through the device. Interfacial temperature jumps were estimated by linear fits to bulk-like regions on either side of the interface, extrapolated to the nominal interface plane.

The effective thermal conductivity along the transport direction was computed as

$$\kappa_{\text{eff}} = \frac{QL}{\Delta T}, \quad (2)$$

where Q is the steady heat flux (net thermostat power divided by the cross-sectional area normal to the transport direction), L is the separation between the centroids of the hot and cold slabs, and ΔT is the fitted temperature difference between these slabs. The thermal boundary conductance of a given interface was similarly obtained as $G = J/\Delta T_{\text{int}}$, where ΔT_{int} is the fitted temperature jump across that interface.

Microscopic heat flux. Microscopic heat current for atom i is evaluated using the following definition [35],

$$\mathbf{J}_i = \sum_{j \neq i} \mathbf{r}_{ij} \left(\frac{\partial U_j}{\partial \mathbf{r}_{ji}} \cdot \mathbf{v}_i \right), \quad (3)$$

For a group of atoms within a volume V , the heat flux is

$$\mathbf{Q} = \frac{1}{V} \sum_i \mathbf{J}_i, \quad (4)$$

where U_j is the site energy of atom j , $\mathbf{v}_i$ is the velocity of atom i , and $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. The GPUMD implementation partitions many-body contributions in a manner consistent with the NEP energy decomposition. Local heat-flux vectors were coarse-grained on the same voxel grid as the temperature fields by summing $\mathbf{Q}_i$ over atoms in each bin and dividing by the bin volume, and then time-averaged over the steady-state window. The one-dimensional heat-flux profiles shown in the main text

correspond to area averages over transverse directions, retaining only the component along the transport direction.

Stress fields and von Mises stress. Local stresses were obtained from the microscopic virial tensor. Assuming Newton's third law holds, the per-atom virial is [35]

$$\mathbf{W}_i = \sum_{j \neq i} \mathbf{r}_{ij} \otimes \frac{\partial U_j}{\partial \mathbf{r}_{ji}}. \quad (5)$$

The Cauchy stress tensor within a voxel of volume V_{bin} was then defined as

$$\boldsymbol{\sigma} = -\frac{1}{V_{\text{bin}}} \sum_{i \in \text{bin}} \mathbf{W}_i. \quad (6)$$

Voxel-level stress tensors were time-averaged over the steady-state portion of the trajectory to obtain smooth continuum-like stress fields.

From the stress components, we computed the von Mises stress as [42]

$$\sigma_{\text{vM}} = \frac{1}{\sqrt{2}} \left[\frac{(\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2}{2} + 6(\sigma_{xy}^2 + \sigma_{yz}^2 + \sigma_{zx}^2) \right]^{1/2}. \quad (7)$$

To isolate thermo-elastic contributions, we report stress differences relative to an isothermal 300 K reference simulation performed under identical boundary conditions, excluding bins that overlap with the thermostats.

Data availability:

The training and test datasets and the trained machine-learned potential models will be made available upon publication.

Code availability:

The source code and documentation for GPUMD are available at <https://github.com/brucefan1983/GPUMD> and <https://gpumd.org>, respectively. The source code for FETMOD is available at <https://github.com/Kick-H/FETMOD>.

Declaration of competing interest:

The authors declare that they have no competing interests.

CONTRIBUTIONS

KX proposed the initial idea, constructed the dataset, trained the NEP model, and performed all the MD calculations. KX and TL performed the benchmark tests.

YX and GW carried out the density functional theory calculations. DD, HG, and XG provided partial computational support. XK, LT, XW, GZ, and JX drafted the manuscript. All authors discussed the results and contributed to the writing of the manuscript.

ACKNOWLEDGMENTS

This work was supported by the Department of Science and Technology of Liaoning Province (No. JYTMS20231613).

-
- [1] K. Bhol, B. Jena, and U. Nanda, Journey of mosfet from planar to gate all around: A review, [Recent Patents on Nanotechnology](#) **16**, 326 (2022).
 - [2] L. V. Kumar, G. P. Kumar, and B. Karthikeyan, Scaling of field-effect transistors, in [Field Effect Transistors](#) (John Wiley & Sons, Ltd, 2025) Chap. 4, pp. 63–72.
 - [3] G. Chen, Non-fourier phonon heat conduction at the microscale and nanoscale, [Nature Reviews Physics](#) **3**, 555 (2021).
 - [4] X. Wang, C. Liu, Y. Wei, S. Feng, D. Sun, and H. Cheng, Three-dimensional transistors and integration based on low-dimensional materials for the post-moore’s law era, [Materials Today](#) **63**, 170 (2023).
 - [5] K. Zheng, S. Ghosh, and S. Granick, Exceptions to fourier’s law at the macroscale, [Proceedings of the National Academy of Sciences](#) **121**, e2320337121 (2024).
 - [6] A. Beardo, W. Chen, B. McBennett, T. Karimzadeh Sabet, E. E. Nelson, T. H. Culman, H. C. Kapteyn, J. L. Knobloch, M. M. Murnane, and I. Dabo, Nanoscale confinement of phonon flow and heat transport, [npj Computational Materials](#) **11**, 172 (2025).
 - [7] A. J. Gabourie, C. A. Polanco, C. J. McClellan, H. Su, M. Malakoutian, Ç. Köroğlu, S. Chowdhury, D. Donadio, and E. Pop, in [2024 IEEE International Electron Devices Meeting \(IEDM\)](#) (2024) pp. 1–4.
 - [8] Y. Sheng, S. Wang, Y. Hu, J. Xu, Z. Ji, and H. Bao, Integrating first-principles-based non-fourier thermal analysis into nanoscale device simulation, [IEEE Transactions on Electron Devices](#) **71**, 1769 (2024).
 - [9] Y. Hu, Y. Shen, and H. Bao, Ultra-efficient and parameter-free computation of submicron thermal transport with phonon boltzmann transport equation, [Fundamental Research](#) **4**, 907 (2024).
 - [10] Y. Sheng, Y. Xia, J. Xu, S. Wang, P. Ren, Z. Ji, and H. Bao, Multiscale thermal simulation for gaafet with first-principles-based boltzmann transport equation, [IEEE Transactions on Electron Devices](#) **72**, 4700 (2025).
 - [11] N. Schlünzen, J.-P. Joost, and M. Bonitz, Achieving the scaling limit for nonequilibrium green functions simulations, [Phys. Rev. Lett.](#) **124**, 076601 (2020).
 - [12] Y. Guo and M. Wang, Lattice boltzmann modeling of phonon transport, [Journal of Computational Physics](#) **315**, 1 (2016).
 - [13] C. Zhang and Z. Guo, A transient heat conduction phenomenon to distinguish the hydrodynamic and (quasi) ballistic phonon transport, [International Journal of Heat and Mass Transfer](#) **181**, 121847 (2021).
 - [14] S. A. Ali, G. Kollu, S. Mazumder, P. Sadayappan, and A. Mittal, Large-scale parallel computation of the phonon boltzmann transport equation, [International Journal of Thermal Sciences](#) **86**, 341 (2014).
 - [15] Z. Shomali, B. Pedar, J. Ghazanfarian, and A. Abbassi, Monte-carlo parallel simulation of phonon transport for 3d silicon nano-devices, [International Journal of Thermal Sciences](#) **114**, 139 (2017).
 - [16] H. Honarvar, J. L. Knobloch, T. D. Frazer, B. Abad, B. McBennett, M. I. Hussein, H. C. Kapteyn, M. M. Murnane, and J. N. Hernandez-Charpak, Directional thermal channeling: A phenomenon triggered by tight packing of heat sources, [Proceedings of the National Academy of Sciences](#) **118**, e2109056118 (2021).
 - [17] L. Yuan, Z. Weidong, P. J. Huan, W. Qi, and J. Jian-Hua, Strong lattice anharmonicity and glass-like lattice thermal conductivity in nitrohalide double antiperovskites: A case study based on machine-learning potentials, [Thermo-X](#) **1** (2025).
 - [18] T. Liang, P. Ying, K. Xu, Z. Ye, C. Ling, Z. Fan, and J. Xu, Mechanisms of temperature-dependent thermal transport in amorphous silica from machine-learning molecular dynamics, [Phys. Rev. B](#) **108**, 184203 (2023).
 - [19] K. Xu, A. J. Gabourie, A. Hashemi, Z. Fan, N. Wei, A. B. Farimani, H.-P. Komsa, A. V. Krashenninnikov, E. Pop, and T. Ala-Nissila, Thermal transport in mos2 from molecular dynamics using different empirical potentials, [Phys. Rev. B](#) **99**, 054303 (2019).
 - [20] Z. Li, S. Xiong, C. Sievers, Y. Hu, Z. Fan, N. Wei, H. Bao, S. Chen, D. Donadio, and T. Ala-Nissila, Influence of thermostatting on nonequilibrium molecular dynamics simulations of heat conduction in solids, [The Journal of Chemical Physics](#) **151**, 234105 (2019).
 - [21] C. Shao and H. Bao, A molecular dynamics investigation of heat transfer across a disordered thin film, [International Journal of Heat and Mass Transfer](#) **85**, 33 (2015).
 - [22] H. Bao, J. Chen, X. Gu, and B. Cao, A review of simulation methods in micro/nanoscale heat conduction, [ES Energy and Environment](#) **1**, 16 (2018).
 - [23] X. Gu, Z. Fan, and H. Bao, Thermal conductivity prediction by atomistic simulation methods: Recent advances and detailed comparison, [Journal of Applied Physics](#) **130**, 210902 (2021).
 - [24] J. Chen, X. Xu, J. Zhou, and B. Li, Interfacial thermal resistance: Past, present, and future, [Rev. Mod. Phys.](#) **94**, 025002 (2022).
 - [25] T. Liang, K. Xu, P. Ying, W. Jiang, M. Han, X. Wu, W. Ouyang, Y. Yao, X. Zeng, Z. Ye, Z. Fan, and J. Xu, Probing the ideal limit of interfacial thermal conductance in two-dimensional van der waals heterostructures, (2025), [arXiv:2502.13601 \[physics.comp-ph\]](#).
 - [26] E. Prasnikař, M. Ljubic, A. Perdić, and J. Borisek, Machine learning heralding a new development phase in molecular dynamics simulations, [Artificial Intelligence Review](#) **57**, 102 (2024).
 - [27] Y. Hu, R. Jia, J. Xu, Y. Sheng, M. Wen, J. Lin, Y. Shen, and H. Bao, Giftbte: an efficient deterministic solver for non-gray phonon boltzmann transport equation, [Journal of Physics: Condensed Matter](#) **36**, 025901 (2023).

- [28] Y. Hu, Y. Shen, and H. Bao, Ultra-efficient and parameter-free computation of submicron thermal transport with phonon boltzmann transport equation, *Fundamental Research* **4**, 907 (2024).
- [29] L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, C. L. Zitnick, and Z. W. Ulissi, Open materials 2024 (omat24) inorganic materials dataset and models, *arXiv* (2024), 2410.12771 [cond-mat.mtrl-sci].
- [30] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, F. Bigi, S. M. Blau, V. Cărare, M. Ceriotti, S. Chong, J. P. Darby, S. De, F. Della Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, E. Fako, F. Falcioni, A. C. Ferrari, J. L. A. Gardner, M. J. Gawkowski, A. Genreith-Schriever, J. George, R. E. A. Goodall, J. Grandel, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. H. Ho, S. Hofmann, C. Holm, J. Jaafar, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, P. Kourtis, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, C. Lin, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. A. M. Rosset, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, C. Sutton, V. Svahn, T. D. Swinburne, J. Tilly, C. van der Oord, S. Vargas, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, T. Wolf, F. Zills, and G. Csányi, *A foundation model for atomistic materials chemistry*, (2025).
- [31] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modeling, *Nature Machine Intelligence* **5**, 1031 (2023).
- [32] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, *Nature Computational Science* **2**, 718 (2022).
- [33] D. Zhang, A. Peng, C. Cai, W. Li, Y. Zhou, J. Zeng, M. Guo, C. Zhang, B. Li, H. Jiang, T. Zhu, W. Jia, L. Zhang, and H. Wang, *A graph neural network for the era of large atomistic models*, (2025), arXiv:2506.01686 [physics.comp-ph].
- [34] T. Liang, K. Xu, E. Lindgren, Z. Chen, R. Zhao, J. Liu, E. Berger, B. Tang, B. Zhang, Y. Wang, K. Song, P. Ying, N. Xu, H. Dong, S. Chen, P. Erhart, Z. Fan, T. Ala-Nissila, and J. Xu, Nep89: Universal neuroevolution potential for inorganic and organic materials across 89 elements, (2025), arXiv:2504.21286 [cond-mat.mtrl-sci].
- [35] K. Xu, H. Bu, S. Pan, E. Lindgren, Y. Wu, Y. Wang, J. Liu, K. Song, B. Xu, Y. Li, T. Hainner, L. Svensson, J. Wiktor, R. Zhao, H. Huang, C. Qian, S. Zhang, Z. Zeng, B. Zhang, B. Tang, Y. Xiao, Z. Yan, J. Shi, Z. Liang, J. Wang, T. Liang, S. Cao, Y. Wang, P. Ying, N. Xu, C. Chen, Y. Zhang, Z. Chen, X. Wu, W. Jiang, E. Berger, Y. Li, S. Chen, A. J. Gabourie, H. Dong, S. Xiong, N. Wei, Y. Chen, J. Xu, F. Ding, Z. Sun, T. Ala-Nissila, A. Harju, J. Zheng, P. Guan, P. Erhart, J. Sun, W. Ouyang, Y. Su, and Z. Fan, GPUMD 4.0: A high-performance molecular dynamics package for versatile materials simulations with machine-learned potentials, *Materials Genome Engineering Advances* **3**, e70028 (2025).
- [36] Z. Fan, Z. Zeng, C. Zhang, Y. Wang, K. Song, H. Dong, Y. Chen, and T. Ala-Nissila, Neuroevolution machine learning potentials: Combining high accuracy and low cost in atomistic simulations and application to heat transport, *Physical Review B* **104**, 104309 (2021).
- [37] K. Xu, T. Liang, N. Xu, P. Ying, S. Chen, N. Wei, J. Xu, and Z. Fan, Nep-mb-pol: a unified machine-learned framework for fast and accurate prediction of water's thermodynamic and transport properties, *npj Computational Materials* **11**, 279 (2025).
- [38] L. Wang, A. R. Brown, M. Nedjalkov, C. Alexander, B. Cheng, C. Millar, and A. Asenov, in *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)* (2015) pp. 112–115.
- [39] L. Chen, N. Kumari, and Y. Hou, Thermal resistances of crystalline and amorphous few-layer oxide thin films, *AIP Advances* **7**, 115205 (2017).
- [40] S. Munetoh, T. Motooka, K. Moriguchi, and A. Shintani, Interatomic potential for si-o systems using tersoff parameterization, *Computational Materials Science* **39**, 334 (2007).
- [41] T. Schaul, T. Glasmachers, and J. Schmidhuber, in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11 (Association for Computing Machinery, New York, NY, USA, 2011) pp. 845–852.
- [42] D. Wang, J. Lee, K. Holland, T. Bibby, S. Beaudoin, and T. Cale, Von mises stress in chemical-mechanical polishing processes, *Journal of The Electrochemical Society* **144**, 1121 (1997).
- [43] P. E. Blöchl, Projector augmented-wave method, *Phys. Rev. B* **50**, 17953 (1994).
- [44] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [45] M. Mohr, J. Maultzsch, E. Dobardžić, S. Reich, I. Milošević, M. Damnjanović, A. Bosak, M. Krisch, and C. Thomsen, Phonon dispersion of graphite by inelastic x-ray scattering, *Phys. Rev. B* **76**, 035439 (2007).
- [46] K. Song, R. Zhao, J. Liu, Y. Wang, E. Lindgren, Y. Wang, S. Chen, K. Xu, T. Liang, P. Ying, N. Xu, Z. Zhao, J. Shi, J. Wang, S. Lyu, Z. Zeng, S. Liang, H. Dong, L. Sun, Y. Chen, Z. Zhang, W. Guo, P. Qian, J. Sun, P. Erhart, T. Ala-Nissila, Y. Su, and Z. Fan, General-purpose machine-learned potential for 16 elemental metals and their alloys, *Nature Communications* **15**, 10208 (2024).