Unraveling Thermal Transport Correlated with Atomistic Structures in Amorphous Gallium Oxide via Machine Learning Combined with Experiments

Yuanbin Liu[1], Huili Liang[2,3], Lei Yang[1], Guang Yang[1], Hongao Yang[1], Shuang Song[3], Zengxia Mei[2,3], Gábor Csányi[4,*], and Bingyang Cao[1,*]

[1]*Key Laboratory for Thermal Science and Power Engineering of Ministry of Education, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, China*

[2]*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*

[3]*Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China*

[4]*Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom*

[*]gc121@cam.ac.uk; [*]caoby@tsinghua.edu.cn;

Abstract: Thermal transport properties of amorphous materials are crucial for their emerging applications in energy and electronic devices. However, understanding and controlling thermal transport in disordered materials remains an outstanding challenge, owing to the intrinsic limitations of computational techniques and the lack of physically-intuitive descriptors for complex atomistic structures. Here, we show how combining machine learning-based models and experimental observations can help to accurately describe realistic structures, thermal transport properties, and structure–property maps for disordered materials, which is illustrated by a practical application on gallium oxide. First, we report the experimental evidence to demonstrate that machine learning interatomic potentials, generated in a self-guided fashion with minimum quantum-mechanical computations, enable the accurate modeling of amorphous gallium oxide and its thermal transport properties. Our atomistic simulations then reveal the microscopic changes in the short-range and medium-range order with density and elucidate how these changes can reduce localization modes and enhance coherences' contribution to heat transport. Finally, a physics-inspired structural descriptor for disordered phases is proposed, with which the underlying relationship between structures and thermal conductivities is predicted in a linear form. This work might shed light on the future accelerated exploration of novel thermal transport properties and

mechanisms in disordered functional materials.

1. Introduction

Thermal transport properties of amorphous materials are crucial for their emerging applications in thermoelectric devices[1], phase-change memory devices[2], flexible electronics[3], radiation detectors[4], artificial intelligence chips[5], thermal barrier coatings[6], and batteries[7]. For instance, low vibrational thermal conductivity (κ) in amorphous solids is advantageous to improve heat-to-electricity conversion efficiency in thermoelectric applications or sensitivity in gamma-ray detectors, but on the contrary, it may lead to serious heat dissipation problems for next-generation electronics or batteries where a large amount of Joule heat is generated. Clearly, these various technological aspects have triggered immense interest in accurately describing, understanding, and ultimately controlling thermal transport in amorphous materials.

Heat in nonmetallic crystals is mainly carried by phonons and thermal resistance arises from phonon scattering, while the absence of periodicity makes thermal behaviors in amorphous solids quite different by the strong localization of vibrational modes and the suppression of vibrational scattering length scales. To date, tremendous progress has been made in the theoretical formalism for describing heat conduction in disordered phases by explicitly considering off-diagonal terms of the heat current operator[8]. The landmark theoretical frameworks include the Allen-Feldman (AF) theory[9] and the more recently developed unified theory[8] as well as the quasi-harmonic Green-Kubo method[10]. Their accuracy, thanks to the development of powerful computer simulations, has been highlighted in some amorphous solids[11] or glass-like crystals with strong anharmonicity[8, 12].

Despite significant advances in theoretical formalisms, the complete understanding of heat conduction mechanisms in amorphous materials is still elusive because there remain notable challenges to the accurate modeling of disordered systems. For almost last three decades, molecular dynamics (MD) simulations, based on either density functional theory (DFT)[13] or classical force fields, have long been the central approach for modeling and understanding materials. Generally, the realistic modeling of amorphous materials requires long simulation times to reach experimentally relevant cooling rates (e.g., 10^{11} K/s for a-Si[14]) and large system sizes to ensure statistical reproduction of medium-range order[14-15]. On the one hand, although DFT-MD can correctly capture the bonding and structural subtleties of amorphous matter with quantum-mechanical accuracy, the limited system sizes (a few hundreds of atoms) and time scales (a few tens of ps) in DFT-MD simulations are inaccessible to the above modeling requirements due to formidable computational cost. On the other hand, empirically parameterized force fields are significantly cheaper than DFT, granting access to larger system sizes (up to millions of atoms) and longer timescales (up to 1000 ns). Nevertheless, their relatively poor predictive power and transferability have been identified in many studies[15b, 16]. Thus, the lack of reliable interatomic potentials has been the biggest obstacle in the realistic modeling of thermal transport in disordered systems.

Recently, machine learning (ML) techniques are emerging as a powerful tool to create a new generation of interatomic potentials by directly mapping the relationship between atomistic configurations and energies from a suitably chosen ensemble of quantum-mechanically computed reference data[17]. With comparable accuracy to DFT, but with linear scaling behavior in computational efficiency, ML interatomic potentials promise to address the aforementioned challenges in materials modeling. The emergent applications of ML potentials have enabled the unprecedented understanding of atomistic structures of a number of disordered materials, such as amorphous carbon (a-C)[16a, 18], silicon (a-Si)[14, 15b, 16b], phosphorus (a-P)[7a], hafnium dioxide (a-HfO_2)[19], and chalcogenides (e.g., a-GeTe, a-$Ge_2Sb_2Te_5$)[15a, 20]. In spite of the above progress, our knowledge regarding thermal transport in amorphous materials is still limited compared with crystalline materials[21]. Specifically, the following critical problems still remain unexplored:

(i) How to effectively generate a high-fidelity ML potential at a low cost to accurately describe thermal transport properties of amorphous materials?

(ii) How does heat conduction change across different topological network regimes? What are the underlying mechanisms?

(iii) How to quantitively describe the relationship between amorphous structures and thermal conductivities in terms of materials informatics?

Problem (i) arises from the large chemical space of disordered phases, which requires ML potentials to be flexible and general enough for a wide range of configurations. This poses a major challenge to efficiently sample training data, set appropriate regularization (to avoid overfitting), and find the best hyperparameters. Problem (ii) is the key to the complete understanding of nanoscale heat conduction mechanisms in disordered phases. Problem (iii) has always been a challenging bottleneck in designing functional materials with superb thermal properties (e.g., ultralow κ materials for thermoelectric devices or relatively high κ materials for electronic cooling). The challenge arises from the construction of a suitable descriptor (or fingerprint) for complex amorphous networks, which needs to sufficiently capture the feature of complex chemical environments and satisfy the fundamental physical requirements: smoothness and invariance to translation, rotation, and permutation.

To shed light on these problems, we report a combined ML-driven atomistic simulations and experimental study on structural features, thermal transport mechanisms, and structure–thermal property correlations in a representative amorphous metal oxide system, namely gallium oxide. Here, gallium oxide is chosen as the model material because of its emerging applications in next-generation high-power and flexible electronics[4, 22], where the thermal properties of gallium oxide are critical to the reliable operation of devices but remain unclear hitherto. Firstly, we introduce a powerful ML-based Gaussian approximation potential (GAP)[23] for gallium oxide systems, which can describe a broad range of bulk allotropes, including liquid, amorphous, and crystalline phases. We show how a first-principles reference database, and the data-effective GAP model can be constructed through iterative random structure searching and fitting in a largely automated fashion to minimize human assistance. The systematical experiments on the structures and thermal transport properties of amorphous gallium oxide are carried out to validate the overall capability of the GAP model. We then perform a bunch of GAP-driven simulations to explore the changes in the short- and medium-range structural order from low- to high-density regions of amorphous and stoichiometric gallium oxide (a-Ga_2O_3). A comprehensive investigation is followed to elucidate how atomistic structures and thermal transport properties are connected. Finally, we develop a physics-inspired descriptor for arbitrary amorphous material, with which the quantitative relationship between structure and thermal conductivity in a-GaO_x under both stoichiometric (x =

1.5) and non-stoichiometric ($x \neq 1.5$) conditions is successfully revealed. Our work exemplifies how ML-driven modeling can help the challenging study of thermal transport in disordered functional materials. The findings in this work, e.g., descriptor of amorphous structures or data-efficient workflow to generate GAP for disordered materials, can be readily extended to other systems of interest.

2. Results and Discussion

Thermal conductivities. To provide fundamental data for the structures and thermal properties of amorphous gallium oxide, we first deposit the thin films of *a*-Ga_2O_3 on silicon by radio frequency magnetron sputtering technique at room temperature (see Methods section for more details). We used various spectroscopy techniques including scanning transmission electron microscopy (STEM), energy-dispersive X-ray spectroscopy (EDS), X-ray reflection spectra (XRR), and X-ray diffraction (XRD) to characterize the film thickness, chemical compositions, density, and structures respectively. Figure 1a shows a high-angle annular dark field (HADDF) STEM image for the cross-sectional profile of the sample. The *a*-Ga_2O_3 film thickness is measured as 498.9 nm. Figure 1b further shows the atomistic structures near the heterogeneous interface between the *a*-Ga_2O_3 film and the silicon substrate at 2 nm resolution. It is solidly evidenced by electron diffraction patterns (inset in Figure 1b) that the amorphous phase of Ga_2O_3 is formed. The EDS mapping (Figures 1c and 1d) reveals a ~2 nm SiO_x interface layer formed between *a*-Ga_2O_3 and silicon. The formation of this SiO_x layer can be attributed to the chemical bonding of interfacial oxygen with silicon. The above STEM-EDS characterizations reflect the high quality of our sample. Furthermore, the density of *a*-Ga_2O_3 film is measured as 4.78 g cm^{-3} by the extracted values from XRR (Figure S1, Supporting Information), which is consistent with the previously experimental data[24]. The X-ray structure factor of *a*-Ga_2O_3 is extracted from the XRD data (Figure S2, Supporting Information), which will be used later to validate the quality of our GAP-generated structures.

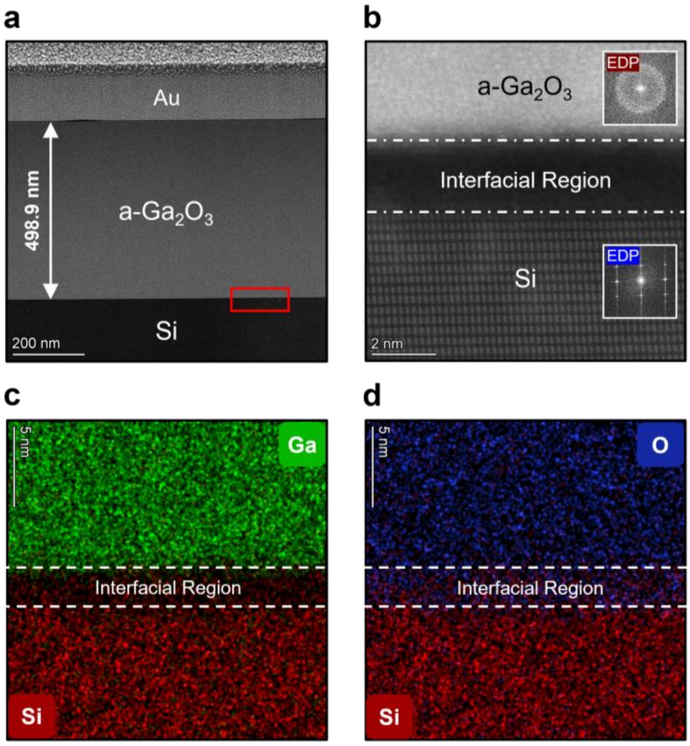

Figure 1. STEM-EDS characterizations of the *a*-Ga_2O_3 thin film which is fabricated by the reactive radio frequency magnetron sputtering process. a HAADF-STEM image showing the structure of an experimental sample and the film thickness of the *a*-Ga_2O_3 layer. **b** HAADF-STEM image showing the atomistic structures near the heterogeneous interface from the region highlighted by the red box in **a**. The two insets in **b** show the electron diffraction pattern (EDP) of *a*-Ga_2O_3 and Si from individual layers shown in the cross section. **c,d** Ga-Si and O-Si EDS profiles. The elemental map indicates the existence of the SiO_x interlayer between the *a*-Ga_2O_3 film and the Si substrate.

The temperature-dependent thermal conductivity of *a*-Ga_2O_3 is then measured by the three-sensor 3ω-2ω method (schematic in Figure 2a)[25], an electrical technique to heat the sample surface and detect the corresponding temperature response by designed metal strips[26], enabling the simultaneous extraction of thermal conductivities of both film and substrate as well as their thermal boundary resistance. Our measurement is demonstrated to have high sensitivity to thermal conductivities of both *a*-Ga_2O_3 and silicon (Figure S4, Supporting Information). It is noted that although *a*-Ga_2O_3 thin film is used here, its 498.9 nm thickness is large enough to ignore the size effect on thermal conductivity, which will be discussed later. Finally, the experimental thermal conductivity of *a*-Ga_2O_3 from 300 K to 400 K are depicted in Figure 2b with circles. The values of κ are from 1.25 ± 0.24 W m^{-1} K^{-1} at 300 K to 1.31 ± 0.26 W m^{-1} K^{-1} at 400 K, here under Debye temperature.

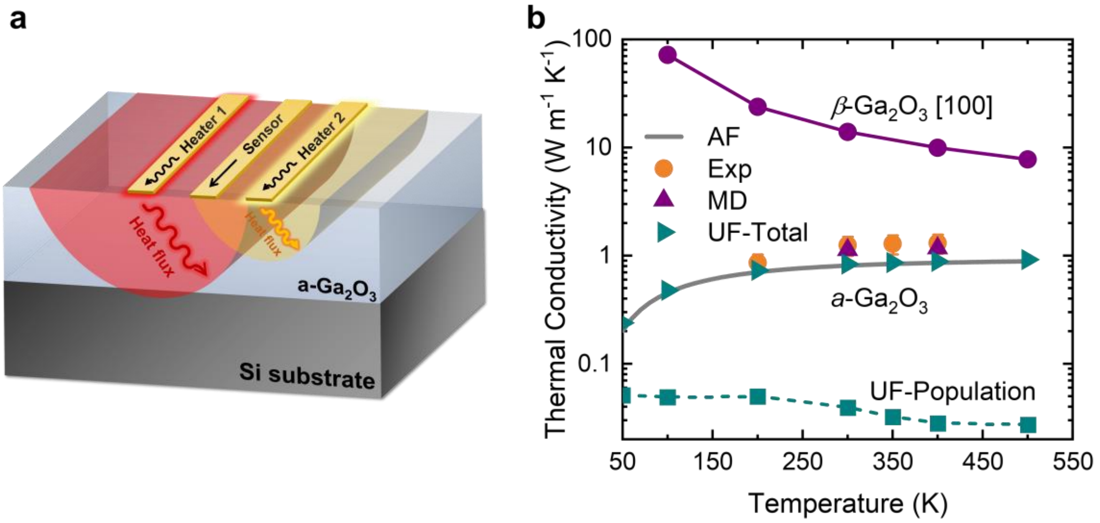

Figure 2. Thermal conductivity of *a*-Ga_2O_3. a Illustration of the three-sensor 3ω-2ω measurement method. The detailed measurement circuit and measurement sensitivity are available (Figure S3,4, Supporting Information). **b** Three-sensor 3ω-2ω measured thermal conductivity of *a*-Ga_2O_3 in comparison with values from ML-driven nonequilibrium molecular dynamics simulations, AF theory, and unified (UF) theory calculations. The darkcyan squares denote temperature-dependent thermal conductivities of populations, which are obtained from the UF theory calculations (see Equation 9). The error bars for the three-sensor 3ω-2ω data represent one standard deviation and are obtained via the analysis of uncertainty propagation combining numerical and systematic errors. The lattice thermal conductivity of β-Ga_2O_3 is obtained from the literature[22a].

To theoretically capture the atomistic structures (inaccessible to experiments) and κ of amorphous gallium oxide, we start by constructing an accurate and transferrable ML-based GAP model. Given the phase space of disordered and amorphous solids is much larger than that of crystals, ML potentials are required to be flexible and general enough for the exploration tasks in structure prediction of disordered phases. Aside from the regression task itself and structural representations, the generality of ML potentials largely depends on the diversity of atomic configurations in the reference database. Usually, relevant atomistic reference data are manfully chosen and tuned when fitting ML potentials, which inevitably requires both considerable domain experience and human efforts. We here use an inherently different way to generate reference data and construct the GAP model through iterative random structure searching (RSS) and fitting[27]. GAP-RSS is a self-guided learning method that enables the largely automated sampling of the few most relevant and diverse structures at each iterative step by a Boltzmann-biased flat histogram and leverage-score CUR[28]. Moreover, our GAP-RSS generated structures contain different compositions of GaO_x binary alloys, which substantially expands the applicability of our model to non-stoichiometric oxides. As a result, the GAP-RSS structures can widely spread over the disordered phase space, including from initially randomized unstable regions to relaxed low-energy regions. In addition, the search process only involves a relatively small number of single-point DFT calculations, leading to low computational costs. Our final ML model from the GAP-RSS iterations shows the required generality and accuracy in the reproduction of DFT potential surfaces for the desired disordered regions of configuration space, with a low root mean square error (RMSE) of 0.017 eV/atom for predicted energies in the testing datasets. The RMSEs of predicted forces in the testing datasets are 0.32 eV/Å for Ga atom and 0.26 eV/Å for O atom.

We now turn to evaluate the performance of the GAP model in modeling κ of amorphous gallium oxide. Firstly, we adopt the experimental density of 4.78 g cm^{-3} to generate the a-Ga_2O_3 model by canonical ensemble (NVT) melt-quench simulations with the LAMMPS package[29]. We then comprehensively study thermal transport in a-Ga_2O_3 by the AF theory, the unified theory, and non-equilibrium molecular dynamics (NEMD), respectively. It is worth noting that both the AF and unified theories consider the quantum effect of vibrations at low temperatures, namely energy hopping between localized vibrational eigenstates. The difference is that the AF theory intensively treats the off-diagonal contributions to κ from the heat-flux operator[30] under harmonic approximation, while the recently developed unified theory comprehensively accounts for both diagonal and off-diagonal contributions, as well as high-order anharmonicity. The diagonal terms correspond to the propagation of vibrational modes, referred to as populations[8], while the off-diagonal terms originate from the wave-like tunnelling and loss of coherence between different vibrational eigenstates, referred to as coherences[8]. In general, coherences' contribution to κ is negligible in simple crystals, but it would become significant in disordered materials. Distinct from the AF and unified theories, MD simulations assume that all the vibrations are fully occupied at any temperatures but entirely ignore the quantum effect of vibrations, leading to the unphysical prediction at low temperatures. Hence, NEMD will be only performed above the room temperature in this work. In the calculations of both AF and unified theories, the 500-atom as-quenched cell is used, which mirrors the maximum model size for DFT-quality MD. However, the 10,000-atom as-quenched cell is adopted in NEMD, which is far beyond the limit of what is presently possible for DFT-based MD.

The reasonable agreement between experimental and calculated κ (Figure 2b) provides the solid evidence for the accuracy and strength of the ML-based GAP potential in treat challenging exploration tasks. It is shown that populations' contribution to κ is negligible in disordered Ga_2O_3 and thermal transport is almost entirely driven by coherences, which also proves the weak size effect of heat transport in a-Ga_2O_3. Interestingly, the harmonic AF calculations are comparable to these of unified theory over the wide temperature range. It suggests that thermal transport in a-Ga_2O_3 is strongly influenced by disorder rather than anharmonicity. Furthermore, the large difference between thermal conductivities in both a-Ga_2O_3 and crystalline β-Ga_2O_3 materials clearly demonstrates the distinct thermal transport mechanisms between them. Compared with a-Ga_2O_3, populations become dominant in heat transport in β-Ga_2O_3. The strong localization of heat carriers makes the thermal conductivity of a-Ga_2O_3 one order of magnitude lower than that of β-Ga_2O_3 at room temperature. It is noted that the thermal conductivities predicted from the UF/AF calculations are lower than those from the NEMD simulations. This is because of the quantum effects below Debye temperature.

Structural properties. To shed light on the effect of atomistic networks on heat transport, structural properties need to be fully understood. To this purpose, we carry out a bunch of GAP-driven quench simulations to generate an ensemble of a-Ga_2O_3 models at various densities (Figure 3a). Due to the breaking of long-range translational symmetry, the short- and medium-range order has become the most prominent features of amorphous structures, which could crucially determine the macroscopic properties of materials. Here, we identify the short- and medium-range order in the generated structures by using the four types of quality indicators: total pair distribution functions (PDFs), coordination numbers, total angle-distribution functions (ADFs), and statistics of shortest-path rings. Initially, PDFs are used to measure the short-range order in a-Ga_2O_3 networks (Figure 3b). The first PDF peak is located at ≈ 1.9 Å and indicates average bond length of all kinds of atom pairs. As density increases, a significant phenomenon is that the intensity of the first PDF peak weakens, and the second PDF peak is blue-shifted for the high-density system (7.50 g cm^{-3}).

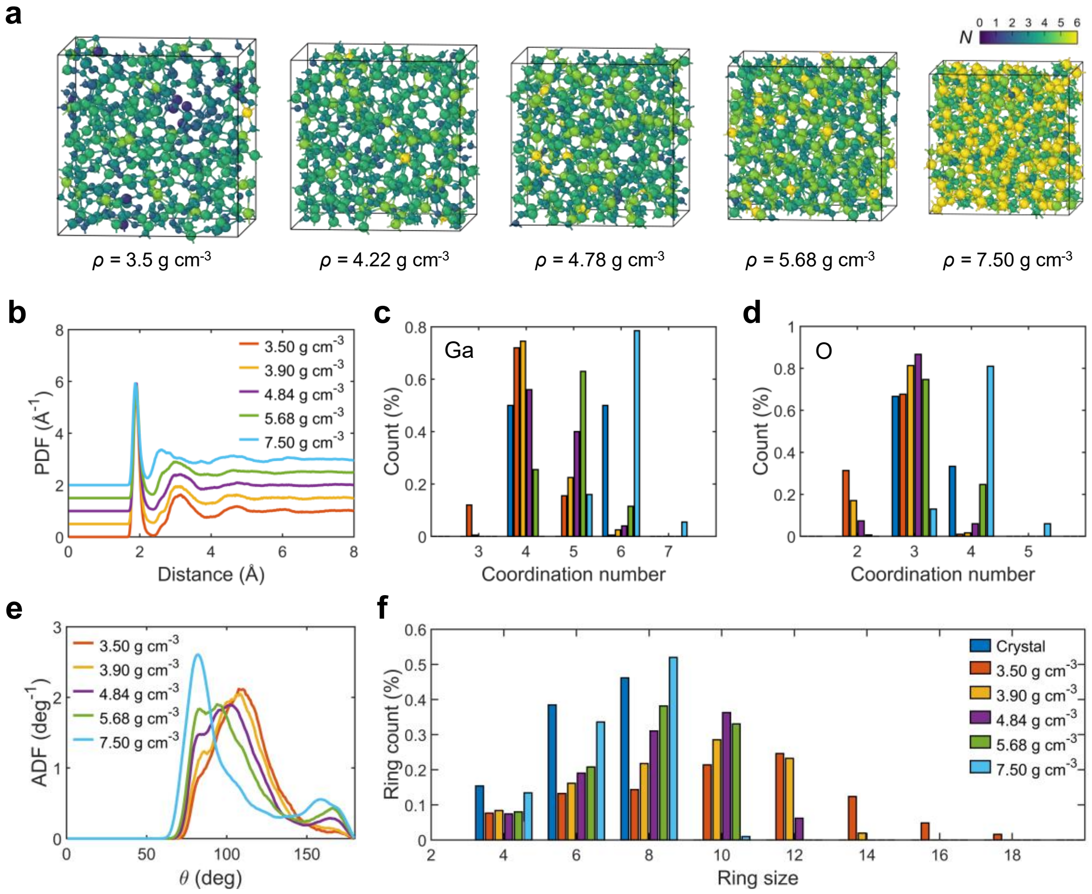

Figure 3. Structures of *a*-Ga_2O_3 at various densities, obtained from melt-quench simulations using the GAP-RSS model. **a** Snapshots of 500-atom models. They are visualized by OVITO[31]. Colour mapping indicates coordination numbers, *N*, with a maximum bond-length cutoff of 2.35 Å. Larger spheres represent Ga atoms, while smaller spheres represent O atoms. **b** Total atomic pair distribution functions (PDFs), measuring the short-range order in *a*-Ga_2O_3. **c,d** Average coordination-number histograms for Ga atom and O atom, respectively. **e** Angle-distribution functions (ADFs). **f** Distribution of shortest-path rings in amorphous and crystalline phases, which is used to characterize medium-range order in *a*-Ga_2O_3. Note that **c,d,f** share the same legends.

To reveal the local structural characteristics in finer detail, we evaluate the coordination number distribution using a maximum bond-length cutoff of 2.35 Å (Figure 3c,d). Generally, the average coordination numbers of Ga and O atoms gradually increase as density increases. The most probable coordination numbers of Ga and O atoms are 4- and 3-fold in *a*-Ga_2O_3 with lower density, respectively, while the fraction of 5- and 6-fold coordinated Ga atoms or 4-fold coordinated O atoms is more prominent in *a*-Ga_2O_3 with higher density. A reduced proportion of 2-fold coordinated atoms is clearly observed as density increases. As to crystalline β-Ga_2O_3, however, 50% of Ga atoms are 4- or 6-fold coordinated, while 1/3 and 2/3 O atoms are 4- and 3-fold coordinated, respectively. It indicates that *a*-Ga_2O_3 networks contain a certain number of coordination defects, e.g., "dangling bonds" and "floating bonds". The O–Ga–O ADFs of all simulated models are shown in Figure 3e. Compared with the low-density system, the high-density one has a sharper peak in the ADF. As evidenced by the shift of the maximum peak in the ADF, the proportion of tetrahedral-like environments is reduced with density, while the octahedral-like environments become dominant at high-density systems.

Beyond the nearest-neighbor environments, the medium-range order of *a*-Ga_2O_3 is characterized by means of ring statistics. To this end, Franzblau's shortest-path algorithm and ISAACS package[32] are used. We here only discuss even-membered rings and define an 2*n*-fold ring as the shortest path of alternating Ga−O heteropolar bonds. That is, a 2*n*-fold rings consists of 2*n* alternating Ga−O bonds. The quick decay of distributions for different structural models suggests that the 500-atom cell is already good to estimate the medium-range structural features (see Figure 3f). The 8-fold ring is the most probable motif at high densities which is similar to β-Ga_2O_3. By contrast, low-density amorphous structures have higher-order rings consisting of up to 18 atoms. The medium-size rings, 8−12 fold, have the largest relative probabilities to be building blocks for the amorphous materials. Remarkably, as the density of *a*-Ga_2O_3 increases, its medium-range ordered structure distribution is closer to that of β-Ga_2O_3.

Thermal transport mechanisms. It has been demonstrated that the atomistic structures of *a*-Ga_2O_3 strongly depend on density. Here, we proceed to discuss how density-induced variation in structures affects the thermal properties of *a*-Ga_2O_3. Three representative systems with different densities, viz. 3.5 g cm^{-3}, 4.78 g cm^{-3}, and 5.68 g cm^{-3}, are chosen for specific analysis. For a clear visualization, the thermal properties of the system with the experimental density of 4.78 g cm^{-3} will act as a benchmark to compare those of the sparse (3.5 g cm^{-3}) and dense (5.68 g cm^{-3})

systems.

We adopt the unified theory to calculate the thermal-conductivity density of states κ_ω and cumulative thermal conductivity κ_{CUM} at different densities (Figure 4a,b). It is shown that the total κ in the high-density system is larger than that in the low-density system. The variation in κ can be linked to the change of local chemical environments. As discussed above, increasing the density also increases atomic coordination numbers. This increase in the bonding number per atom will provide additional pathways through which vibrational modes could interact, leading to a gradual enhancement in thermal transport. It is also shown that the low vibrational modes below 10 THz contribute mostly to the total κ, and the structural variation exerts evident influence on the mode-wise thermal conductivity in the vibrational regimes approximately over 3 THz.

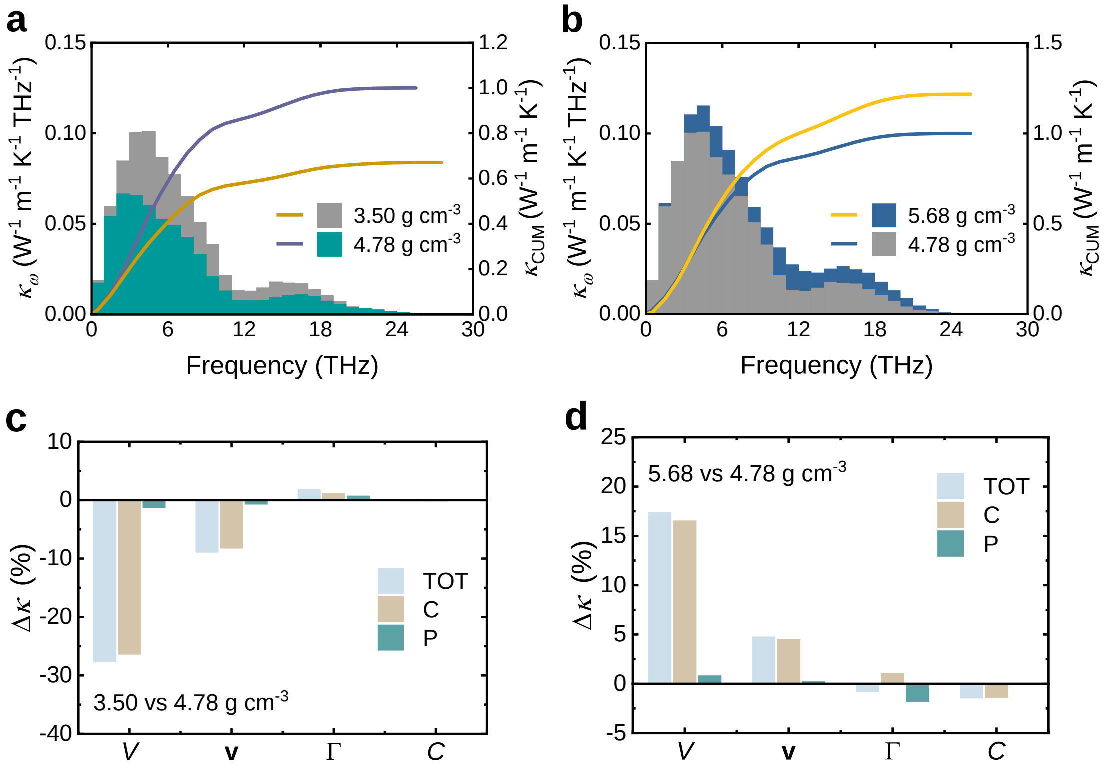

Figure 4. Vibrational properties and influencing factors. a,b Thermal conductivity density of states κ_ω and cumulative thermal conductivity κ_{CUM} of *a*-Ga_2O_3. **c,d** Role of cell volume (*V*), velocity operator (**v**), mode linewidth (Γ), and specific heat (*C*) in thermal transport as density changes. The system at 4.78 g cm^{-3} acts as a benchmark for the comparison.

We further estimate quantitively the role of four important macroscopic and microscopic physical factors in thermal transport as density changes, viz. cell volume (V), specific heat (C), velocity operator ($\mathbf{v}$), and mode linewidth (Γ). For example, we calculate the contribution of cell volume to the variation in κ by:

$$\Delta\kappa_V = \frac{\kappa(V, \mathbf{v}^b, \Gamma^b, C^b)}{\kappa(V^b, \mathbf{v}^b, \Gamma^b, C^b)} \tag{1}$$

where the superscript b indicates that the value of the corresponding factor is taken from the benchmark system. As previously assumed, the benchmark system is chosen as the one at 4.78 g cm^{-3}. The contributions from the other three factors can be evaluated by analogy. In addition to total κ, the variation in coherences' thermal conductivity κ_{C} and populations' thermal conductivity κ_{P} under the four factors is also established (Figure 4c,d). It is shown that cell volume plays the most significant role in changing κ among the four factors. Indeed, density directly changes cell volume due to their intrinsic inversely proportional relationship. Aside from cell volume, velocity operator is also identified as a relatively important factor affecting $\Delta\kappa$. The off-diagonal terms of velocity operator can reflect the coherences' contribution to κ. It indicates the coupling strength of vibrational modes at high-density is larger than that at low density. Compared with the former two factors, the influence of linewidth and specific heat on $\Delta\kappa$ is much less important, which again indicates the weak anharmonicity in a-Ga_2O_3. The overall results here provide concrete evidence for the dominant contribution from coherences to κ at different densities.

To gain deep insight into the mechanism of heat transport, using the unified theory, we carry out a vibrational mode analysis to explore the localization and diffusivity of modes, respectively. The inverse participation ratio (IPR) is an effective parameter to quantitively describe the localization extent for each mode, given by

$$\mathrm{IPR} = \frac{\sum_{i=1}^{N}\left(\sum_{\alpha=1}^{3} u_{i\alpha}^2\right)^2}{\left(\sum_{i=1}^{N}\left(\sum_{\alpha=1}^{3} u_{i\alpha}^2\right)\right)^2} \tag{2}$$

where N is the total number of atoms, $u_{i\alpha}$ is the eigenvector component for atom i in α direction.

For example, IPR equals 1 when a mode is fully localized on a single atomic site, while IPR equals $1/N$ when a fully delocalized mode that spans all atoms. In principle, it is impossible to define a specific IPR to distinguish localized modes, but a number of previous studies have reasonably defined localized modes as those with participation ratio less than 0.2 (corresponding to IPR $\geq$ 0.01 for a 500-atom supercell)[11b, 33]. We here follow this convention to define localized vibrational modes. With regard to mode diffusivities D, it describes the rate at which the heat carried by a vibration diffuses, which is extracted by recasting coherences' contribution of pairs of vibrational modes as $\kappa_C = \frac{1}{VN_\mathbf{q}} \sum C_{\mathbf{q},s} D_{\mathbf{q},s}$ (see Methods section for more details).

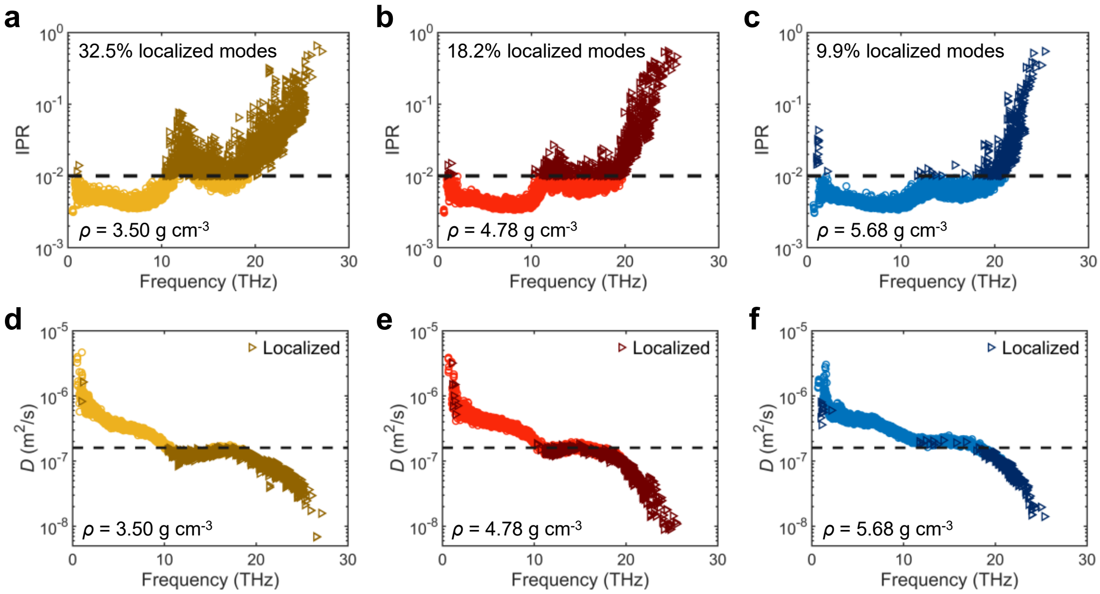

Figure 5. Vibrational mode analysis. a,b,c Inverse participation ratio (IPR) and **d,e,f** mode diffusivity (D) as a function of frequency for a-Ga_2O_3 at 3.50 g cm^{-3}, 4.78 g cm^{-3}, 5.68 g cm^{-3}, respectively. The dash lines in **d,e,f** denote the maximum diffusivity (1.6×10^{-7} m^2/s) of high-frequency localized modes.

The IPR spectra in a 500-atom supercell of a-Ga_2O_3 with different densities are shown in Figure 5a–c, while the corresponding mode diffusivities are plotted in Figure 5d–f. According to our IPR calculations, localized modes mainly stem from high-frequency regions. It is shown that ~32.5% of the vibrational modes are localized for the system at 3.50 g cm^{-3}, which indicates that a large part of modes is inactive in transferring heat, resulting in a low thermal conductivity. As density increases, the percentage of localized modes gradually decreases to ~10%. This suggests that more vibrational modes become delocalized with density. Moreover, the increase in mode diffusivities is clearly observed with density. The over-constrained bonding network in the high-density system could be responsible for the large percentage of delocalized modes and relatively high mode diffusivities, leading to the enhancement in heat transport.

Structure–thermal property correlations. It has become clear that the thermal properties of disordered materials are tied to the corresponding atomistic networks. Along with the above-discussed density, the composition of the binary alloys also plays an important role in thermal energy transport in disordered materials by altering the network connectivity, chemical bonding, and vibrational mode localization[11b, 34]. Arguably, tuning network topology through varying density or compositions could be a promising method towards the rational design of amorphous materials with superb thermal properties. For the accelerated screening of promising amorphous materials from the high-dimensional chemical space, as usually done in materials informatics, it is essential to obtain the quantitative structure–property relationship. Here, we aim to resolve the construction of structure–thermal property relationships for amorphous materials by finding suitable structure descriptors to map complex topological networks into a suitable representation which are strongly correlated with the target physical quantities, e.g., κ.

We firstly evaluate the correlation between the thermal conductivity of amorphous solids and the two heuristically determined structural parameters, viz. density and composition ratio. To this end, we calculate the thermal conductivity of 68 types of additional a-GaO$_x$ cells according to the AF theory which is used because of its relatively low computational cost and weak anharmonicity in a-GaO$_x$. The calculated systems have the wide density and composition ratio distributions, ranging from 3.5 g cm^{-3} to 6.5 g cm^{-3} and from 1.1 to 1.7, respectively. Figure 6a reveals the strong correlation between thermal conductivity and density. It is again suggested that low thermal conductivities occur in the most open structures. However, a simple descriptor only using density cannot explain structure–property relationships in bulks with various compositions (Figure 6a). This is because density is insufficient for uniquely characterizing a given amorphous structure. Consequently, different amorphous structures could have the same density but distinct bonding and thermal behavior. Figure 6b indicates the relationship between composition ratio and thermal conductivity. Increasing the heavy element Ga content in the alloy could further suppress thermal conductivity by localizing more vibrational modes. However, such relationships are highly non-linear and taking composition ratio as the only structural feature cannot account for the effect of density on thermal conductivity. Accordingly, neither density nor composition ratio is adequate for regressing a structure–property relation in amorphous solids.

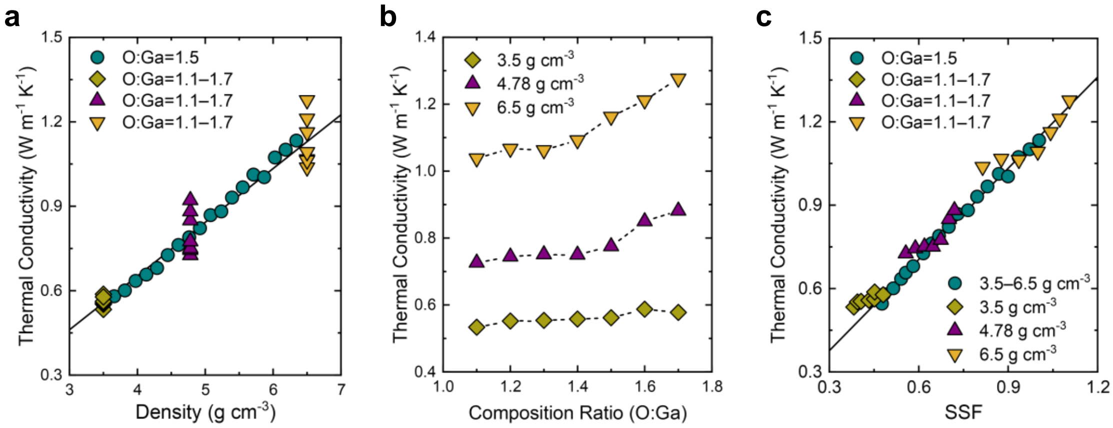

Figure 6. **Structure–thermal conductivity relationship in *a*-GaO$_x$. a–c** Thermal conductivity as a function of density, composition ratio, and SSF, respectively. The black solid curve is a linear fit to the data.

We now turn to a state-of-the-art many-body descriptor known as “smooth overlap of atomic positions” (SOAP)[35] which has been successfully applied to fit ML potentials[17c, 23, 36] and analyze atomic-scale structures[37]. For a given structure, the average global SOAP fingerprint is a high-dimensional vector $\mathbf{p} = \{p_{nn'l}\}$ by considering all components $l \leq l_{\max}$ and n, $n' \leq n_{\max}$, and each element reads

$$p_{nn'l} = \frac{\pi}{N^2}\sqrt{\frac{8}{2l+1}}\sum_m \sum_{i,j} \left(c^i_{nlm}\right)^* c^j_{n'lm} \tag{3}$$

where i and j denote atomic species, and c_{nlm} corresponds to the density expansion coefficients. Compared with the used structural indicators before, viz. PDFs, ADFs, and ring statistics, which just estimate one aspect of statistical characteristics of networks, the SOAP descriptor with a suitable cutoff radius could take all aspect of short- and medium-range structural environments into account. Most importantly, SOAP is an inherently density-sensitive descriptor because the SOAP features are fundamentally constructed by the atomic neighborhood density and power spectrum. Given the strong correlation between thermal conductivity and density, SOAP promises to be an excellent representation for amorphous materials.

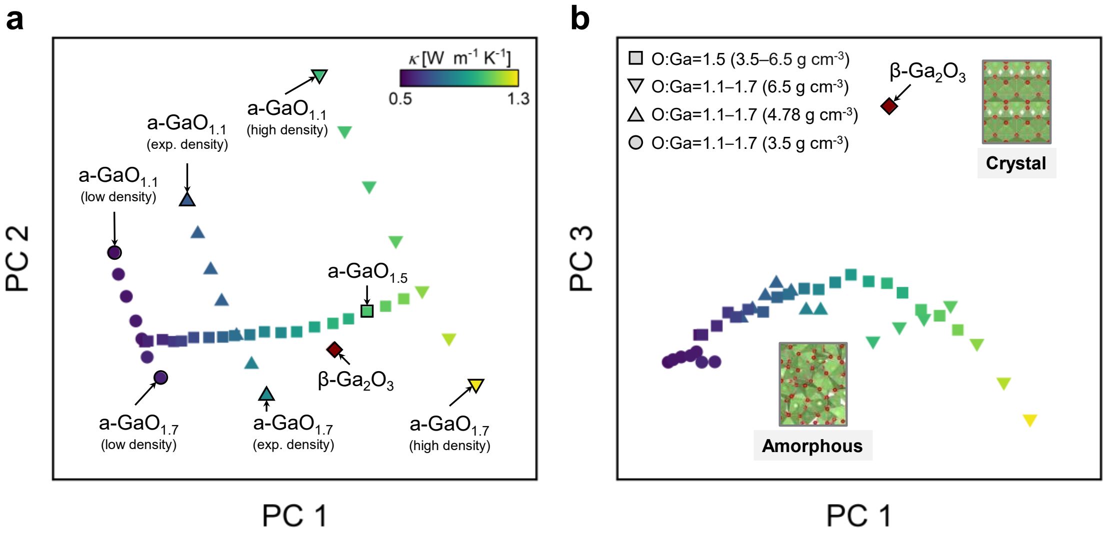

Figure 7. **PCA maps of amorphous and crystalline gallium oxide configurations using the SOAP fingerprints, colored according to thermal conductivity. a** First and second principal components. **b** First and third principal components.

To take insight into the high-dimensional SOAP representations for different *a*-GaO$_x$ systems and reveal their underlying correlations with κ, a ML-based dimensionality reduction technique, viz. principal component analysis (PCA)[38], is used to map each structure to a point in three-dimensional (3D) space for visual inspections. The resulting plot is shown in Figure 7, in which we take the crystalline β-Ga$_2$O$_3$ as a reference and encode the species by symbols and the thermal conductivity by color. The projection axes (also named "principal components", PCs) of panels a and b are formed by the first three eigenvectors of the design matrix in PCA. It can be clearly observed that the amorphous structures with different motifs are clearly separated on the PCA map (Figure 7a). A distinction between amorphous and crystalline phases is also clearly visible, a sign of their fundamental structural differences (Figure 7b). It is suggested that the SOAP descriptor is leading to groupings of gallium oxides sharing similarities in atomic structures and compositions.

The strong horizontal color gradient in Figure 7a indicates that PC1 is essentially equivalent to density, while the vertical color gradient suggests that PC2 is well correlated with amorphous compositions. Consequently, the structure–thermal conductivity trends can be subtly identified by the combination of projections of the SOAP elements on the PC1 and PC2 axes. These results qualitatively reveal the high-level interpretability of SOAP in the representation and κ prediction of amorphous materials. Nevertheless, given that the thermal datasets are usually very limited due to high cost, the quantitative regression of κ using the full SOAP fingerprint or its low-dimensional PCA projection is prone to overfitting or numerical instability.

To circumvent these issues, we attempt to search for a more suitable SOAP-based microscopic metric to identify the relationship between amorphous structures and thermal conductivity. In the perspective of atomic length scales, amorphous and crystalline materials appear similar in the short-range structural order, and both are assembled by the polyhedral building blocks (Figure 7b). The major differences in amorphous and crystalline structures lie in their intermediate and long-range structural order, such as the relative amount, connectivity, and orientation of different polyhedral units. Such differences in structural networks fundamentally change the way of heat transfer from vibrational populations in crystalline Ga_2O_3 to vibrational coherences in amorphous phases. Inspired by this, we develop a robust SOAP-based metric for describing a-GaO_x by quantitatively measuring the structural similarity between amorphous phases and the reference crystalline counterpart, viz. β-Ga_2O_3. More specially, we measure the similarity of chemical environment of an atom in amorphous phase to that of an identical atom in the crystal bulk by using the SOAP descriptor. Accordingly, the scaled but non-normalized similarity metric of each atomic environment in a-GaO_x is defined as:

$$q_i = \frac{1}{\gamma\chi} \max_{k\in\{1,\ldots,\chi\}} \mathbf{p}_i \cdot \mathbf{p}_{j,k} \tag{4}$$

where $\mathbf{p}_i$ and $\mathbf{p}_j$ are the SOAP vectors of atom i in amorphous phase and identical atom j in crystalline counterparts, χ denotes the number of Ga or O atom types in crystal bulk, and γ is a scale factor given by $\max_{k\in\{1,\ldots,\chi\}} \left|\mathbf{p}_{j,k}\right|$ here. For β-Ga_2O_3, there are two types of Ga atoms ($\chi = 2$) and three types of O atoms ($\chi = 3$). For an overall comparison of structures, the structural similarity factor (SSF) is then given by averaging the feature metric of each atom in the amorphous supercell:

$$\mathrm{SSF} = \frac{1}{N}\sum_{i=1}^{N} q_i \tag{5}$$

Here, a SOAP cutoff of 5.50 Å is used so that the intermediate length scale beyond the individual building blocks can be measured.

Having SSF as the unified metric to quantitatively describe atomistic structures of a-GaO_x, we proceed to show how the structure factor is linked to thermal conductivity. Figure 6c illustrates that the thermal conductivity is strongly correlated with SSF. The larger the value of SSF is, the higher the thermal conductivity is. Using linear fitting, a regression model can be quantitatively derived as $\kappa = 1.09 \times \mathrm{SSF} + 0.05$. Combing with the mechanism of atomistic structures affecting thermal transport, such a relationship can be reasonably explained by the physical intuition of SSF. That is, SSF inherits the sensitivity of the SOAP descriptor to density and composition. A larger SSF suggests a denser network that consists of higher average bonding numbers and basically dictates the enhancement in the coherences’ contribution to thermal energy transport as discussed above. In addition, SSF subtly quantifies the similarity of the crystal and amorphous structures in the medium-range order. The increase in the SSF values indicates that the medium-range order in amorphous solids is closer to that in crystalline counterparts, which can be clearly verified by the distributions of shortest-path rings as shown in Figure 3f. More ‘ordered’ medium-range networks could also contribute to heat transport.

It is worth noting the main difference between SOAP and SSF. The quantitative fitting of thermal conductivity of amorphous materials directly using the high-dimensional SOAP fingerprints usually involves a nonlinear regression process with multiple parameters and thus requires a large amount of data to avoid overfitting, while the strongly linear relationship between the single-parameter SSF metric and thermal conductivity can readily circumvent the overfitting issue, especially in the case of small datasets. Moreover, compared with SSF, the structure–thermal property mappings obtained with SOAP is difficult to interpret intuitively. Consequently, SSF enables us to establish a physically reasonable relationship between structures and thermal conductivities with a small amount of data to significantly save the computational cost. Such a relationship would facilitate the fast and accurate prediction of thermal conductivity directly from structural information for amorphous systems and eventually enable the accelerated discovery of functional materials with superb thermal properties.

3. Conclusions

In summary, we demonstrate that combing machine learning-based models and experimental observations can lead to an unprecedented level of understanding of atomistic structures, thermal transport properties, and structure–thermal property correlations for disordered materials, by a practical application on gallium oxide systems. With the GAP-RSS method, we have developed a powerful ML potential with near quantum-mechanics accuracy for the realistic modeling of *a*-GaO_x. It is shown that the GAP-RSS method enables the largely automatic generation of a suitable first-principles database and a flexible ML potential for disordered materials, remarkably reducing computational cost and human efforts. Comprehensive experimental observations on the structures and thermal conductivity of *a*-Ga_2O_3 are shown to be consistent with the computational results from the ML-driven large-scale MD simulations and different theoretical formalisms, validating the overall capability of the ML-driven GAP model. This is a significant milestone toward the realistic modeling of thermal properties of structurally complex functional materials. It is further demonstrated that the harmonic coupling of vibrational modes dominates heat conduction in *a*-Ga_2O_3 instead of phonon-like propagation and anharmonicity. Moreover, the extensive ML-driven atomistic simulations have been carried out to describe the variation in the short- and medium-range structural order from low- to high-density amorphous regions, as well as to elucidate that the over-constrained bonding network, the formation of numerous octahedral-like environments, and the decrease in the heavy element Ga content dictate the enhancement in the coherences' contribution to thermal energy transport in *a*-GaO_x. Finally, we have proposed an effective metric SSF derived from SOAP to quantify disordered structures by determining the similarity between chemical environments of both amorphous and crystalline phases. It is found that SSF is strongly correlated with thermal conductivity in *a*-GaO_x and enables the accurate and fast prediction of thermal conductivity in a physically reasonable way.

The findings of this work would be fundamental to the development of thermal management techniques for flexible electronic devices based on gallium oxide. Looking beyond this application, it generally shows the capability of ML models in tackling realistic problems out of reach for the first-principles methods. Given the complexity and significance of heat transport in disordered phases, this work might provide a starting point for the future accelerated exploration of novel thermal transport properties and mechanisms in other important disordered materials.

4. Methods

Experimental approach. The experimental *a*-Ga_2O_3 film is deposited on Si substrates by radio frequency magnetron sputtering at room temperature. Ga_2O_3 ceramic target is 5N pure. The Si substrate is ultrasonically cleaned in acetone and isopropanol successively and then dried with pure N_2. After cleaning, the Si substrates were immediately loaded into the growth chamber. Before deposition, a base pressure of ~3.0×10^{-4} Pa is achieved by turbopump. The sputtering lasts for 90 min in total under a sputtering power of 60 W and a total pressure of 0.4 Pa in pure Ar atmosphere. In this work, *a*-Ga_2O_3 samples with a fixed density are prepared. However, it is worth noting that *a*-Ga_2O_3 thin films with different densities can be grown by using pulsed laser deposition in an O_2 gas flow and controlling the deposition rate[39].

After obtaining the sample, its thermal conductivity is measured by the 3-sensor 3ω-2ω method[25]. In the scheme, three parallel metal stripes with a unified thickness of 100 nm (10 nm Cr/90 nm Au), including two heaters and a detector, are fabricated through electron beam lithography and sputtering on the surface of the sample. Two heaters with different widths are located on both sides of the sensor. Based on the sensitivity analysis, the width of the two heaters and detector are designed as 30 μm, 1.5 μm and 800 nm, while the distances between each heater and the sensor are 20 μm and 800 nm, respectively (Table S1, Supporting Information). The experiment AC/DC currents are generated by Keithley 6221 current sources, while the 1ω/2ω/3ω voltage signals are captured by SRS SR830 lock-in amplifiers.

Allen-Feldman theory. The AF theory accounts for the diffuson-mediated thermal conductivity:

$$\kappa_{\mathrm{AF}} = \frac{1}{V}\sum_i C_i(\omega_i)\frac{\pi V^2}{3\hbar^2\omega_i^2}\sum_{j\neq i}\left|S_{ij}\right|^2\delta\left(\omega_i-\omega_j\right) \tag{6}$$

where ω_i is the frequency of the *i*th mode, $C(\omega)$ is the frequency-dependent specific heat, δ is the Dirac delta function, $\hbar$ is the reduced Planck constant, and S_{ij} is the matrix element of the heat current operator which measures the thermal coupling between vibrational mode *i* and *j* based on their frequencies and spatial overlap of eigenvectors. The *δ* function is generally obtained by

$$\delta(\omega_i-\omega_j)=\frac{\eta/\pi}{(\omega_i-\omega_j)^2+\eta^2} \tag{7}$$

where η is a free broadening factor in the Lorentzian function. Previous research used several times the mean level spacing of vibrational modes frequency ($\Delta\omega_{\mathrm{ave}}$) as η. Here, we uniformly use 10 times of $\Delta\omega_{\mathrm{ave}}$ as the broadening factor in the Lorentzian function to calculate the AF thermal conductivity. The good agreement between the AF and UF thermal conductivities indicates the reasonableness of our chosen parameter.

Unified theory. In the unified theory, the total thermal conductivity includes both the populations' and coherences' contributions:

$$\kappa_{\mathrm{TOT}}=\kappa_{\mathrm{P}}+\kappa_{\mathrm{C}}. \tag{8}$$

Under single-mode approximation, the resulting formula for κ_{P} reads:

$$\kappa_{\mathrm{P}}=\frac{1}{3VN_q}\sum_{\mathbf{q},s}C_{\mathbf{q},s}\mathbf{v}_{\mathbf{q},ss}^2\frac{1}{\Gamma_{\mathbf{q},s}}. \tag{9}$$

where N_q is the number of sampled wavevectors, and $\mathbf{v}_{\mathbf{q},ss}$ and $\Gamma_{\mathbf{q},s}$ are the generalized group velocity and linewidth indexed by wavevector **q** and branch *s*, respectively. The coherences' thermal conductivity is expressed as

$$\kappa_{\mathrm{C}}=\frac{1}{VN_{\mathbf{q}}}\sum_{\mathbf{q},s\neq s'}\frac{\omega_{\mathbf{q},s}+\omega_{\mathbf{q},s'}}{4}\left(\frac{C_{\mathbf{q},s}}{\omega_{\mathbf{q},s}}+\frac{C_{\mathbf{q},s'}}{\omega_{\mathbf{q},s'}}\right)\frac{\left\|\mathbf{v}_{\mathbf{q},ss'}\right\|^2}{3}\frac{\frac{1}{2}\left(\Gamma_{\mathbf{q},s}+\Gamma_{\mathbf{q},s'}\right)}{\left(\omega_{\mathbf{q},s}-\omega_{\mathbf{q},s'}\right)^2+\frac{1}{4}\left(\Gamma_{\mathbf{q},s}+\Gamma_{\mathbf{q},s'}\right)^2} \tag{10}$$

where the mode specific heat is given by

$$C_{\mathbf{q},s}=\frac{\hbar^2\omega_{\mathbf{q},s}^2}{k_{\mathrm{B}}T^2}\bar{N}_{\mathbf{q},s}\left(\bar{N}_{\mathbf{q},s}+1\right) \tag{11}$$

where k_{B} is the Boltzmann constant and $\bar{N}_{\mathbf{q},s}=\left[\exp\left(\hbar\omega_{\mathbf{q},s}/k_{\mathrm{B}}T\right)-1\right]^{-1}$ is the Bose-Einstein distribution. The above mode properties are calculated by lattice dynamics in a large primitive cell with volume *V*. To describe the thermal diffusivity of modes, coherences' contribution needs to be recast as $\kappa_{\mathrm{C}}=\frac{1}{VN_{\mathbf{q}}}\sum C_{\mathbf{q},s}D_{\mathbf{q},s}$, where the diffusivity of a vibration reads

$$D_{\mathbf{q},s} = \sum_{s'} \frac{\omega_{\mathbf{q},s} + \omega_{\mathbf{q},s'}}{2\left(C_{\mathbf{q},s} + C_{\mathbf{q},s'}\right)} \left(\frac{C_{\mathbf{q},s}}{\omega_{\mathbf{q},s}} + \frac{C_{\mathbf{q},s'}}{\omega_{\mathbf{q},s'}} \right) \frac{\left\| \mathbf{v}_{\mathbf{q},ss'} \right\|^2}{3} \frac{\frac{1}{2}\left(\Gamma_{\mathbf{q},s} + \Gamma_{\mathbf{q},s'}\right)}{\left(\omega_{\mathbf{q},s} - \omega_{\mathbf{q},s'}\right)^2 + \frac{1}{4}\left(\Gamma_{\mathbf{q},s} + \Gamma_{\mathbf{q},s'}\right)^2}. \tag{12}$$

The calculation of the UF theory is carried out with the uniform 2×2×2 **q**-mesh, which is well converged with the tolerance in a thermal conductivity of ~0.02 W m^{-1} K^{-1} compared with a 3 × 3 × 3 **q**-mesh. It should be noted that the MD-based normal mode decomposition method is used to calculate the mode linewidth here. The full-order scattering is implicitly included in MD simulations. Hence, the thermal conductivity based on the unified theory using the extracted linewidths accounts for the full-order anharmonicity. More details on the calculation of the mode linewidth can be found in the literature[40].

GAP-Driven MD. In our NEMD simulations, the size of the simulated system is chosen as 30.0 Å × 30.0 Å × 144.7 Å. The periodic conditions are used in the *x* and *y* directions, and the fixed boundary condition is used in the *z* direction. First of all, the entire structure is fully relaxed by the NVT ensemble for 50 ps with a time step of 1 fs. Next, Langevin thermostats with a temperature difference of 20 K are applied for 400 ps to reach steady-state temperature gradient and heat flux in *a*-$\mathrm{Ga_2O_3}$. Then run NEMD for another 600 ps to collect time-averaged data. To prove that our simulated system is sufficiently large to eliminate the size effects in the vertical (*z*) direction, we calculate the MFP distribution of vibrational modes at different temperatures (Figure S5, Supporting Information). It can be clearly observed that the maximum MFP of dominant vibrational modes (< 10 THz) is less than the length of the simulated system (144.7 Å). In this case, the size effects in the vertical direction can be ignored[41]. Moreover, we perform the melt-quench MD in an NVT ensemble with a Nosé–Hoover thermostat to generate the different models with setup density. The melt-and-quench protocol used for GAP-MD is illustrated in Figure S6.

Reference data. All reference data for fitting the ML potential are generated by the GAP-RSS method. At each iteration, the initial 10,000 randomized structures will be independently generated by the ‘buildcell’ algorithm as in the established *ab initio* random structure searching framework[42]. The cells contain 4–20 gallium atoms and 6–30 oxygen atoms. The most diverse 100 initial structures will be chosen by leverage-score CUR and optimized by GAP-driven atomistic simulations with a preconditioned LBFGS algorithm[43]. The selection of the training configurations from full trajectories involves a Boltzmann-biased flat histogram and leverage-score CUR. Then the selected candidates for fitting will be evaluated by single-point DFT calculations. In this work, all DFT calculations are carried out with the Vienna Ab-Initio Simulation Package (VASP),[44] using the Perdew-Becke-Erzenhof functional[45] and a projector augmented wave (PAW) method.[46] A plane-wave basis cutoff, Gaussian smearing width, and the halting criterion for the self-consistent field iterations are set to 520 eV, 0.05 eV, and 10^{-7} eV, respectively. The automatic k-mesh generation with $L = 30$ is implemented, where L is a parameter to determine the number of divisions along each reciprocal lattice vector.

Listing 1: definition of the descriptor string used in the GAP fit.

```
gap={distance_Nb order=2 compact_clusters=T cutoff=8 covariance_type=ARD_SE
theta_uniform=1.0 sparse_method=uniform n_sparse=15 delta=1 n_species=2 species_Z={31 8}
add_species=T: soap cutoff=5 cutoff_transition_width=1 covariance_type=dot_product zeta=4
delta=0.2 atom_sigma=0.5 l_max=6 n_max=10 n_sparse=3000 sparse_method=cur_points
radial_scaling=0.5 f0=0.0 add_species=T: soap cutoff=3 cutoff_transition_width=1
covariance_type=dot_product zeta=4 delta=0.2 atom_sigma=0.3 l_max=6 n_max=10
n_sparse=300 sparse_method=cur_points radial_scaling=0.5 f0=0.0 add_species=T}
```

Data availability

The potential model described herein as well as the DFT reference data used for fitting the model are openly available at https://libatoms.github.io/GAP/data.html. The other data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The GAP code, which was used to carry out the fitting of the potential and the validation shown throughout this work, is freely available at https://github.com/libAtoms/GAP for noncommercial research. The GAP-RSS code is freely available at https://github.com/HongAo-Yang/ml_rss.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 51825601, U20A20301).

Conflicts of interest

The authors declare no competing interests.

References

[1] a)G. J. Snyder, M. Christensen, E. Nishibori, T. Caillat, B. B. Iversen, *Nat. Mater.* **2004**, 3, 458; b)M. He, Y. Zhao, B. Wang, Q. Xi, J. Zhou, Z. Liang, *Small* **2015**, 11, 5889.

[2] a)M. Wuttig, N. Yamada, *Nat. Mater.* **2007**, 6, 824; b)V. L. Deringer, R. Dronskowski, M. Wuttig, *Adv. Funct. Mater.* **2015**, 25, 6343.

[3] a)R. A. Street, *Adv. Mater.* **2009**, 21, 2007; b)G. S. Armatas, M. G. Kanatzidis, *Nature* **2006**, 441, 1122.

[4] a)H. Liang, S. Cui, R. Su, P. Guan, Y. He, L. Yang, L. Chen, Y. Zhang, Z. Mei, X. Du, *ACS Photonics* **2019**, 6, 351; b)S. Cui, Z. Mei, Y. Zhang, H. Liang, X. Du, *Adv. Opt. Mater.* **2017**, 5, 1700454.

[5] Y. Lu, A. Alvarez, C.-H. Kao, J.-S. Bow, S.-Y. Chen, I. W. Chen, *Nat. Electron.* **2019**, 2, 66.

[6] D. R. Clarke, S. R. Phillpot, *Mater. Today* **2005**, 8, 22.

[7] a)Y. Zhou, W. Kirkpatrick, V. L. Deringer, *Adv. Mater.* **2022**, 34, 2107515; b)I. Capone, J. Aspinall, E. Darnbrough, Y. Zhao, T.-U. Wi, H.-W. Lee, M. Pasta, *Matter* **2020**, 3, 2012; c)Y. Kim, Y. Park, A. Choi, N.-S. Choi, J. Kim, J. Lee, J. H. Ryu, S. M. Oh, K. T. Lee, *Adv. Mater.* **2013**, 25, 3045.

[8] M. Simoncelli, N. Marzari, F. Mauri, *Nat. Phys.* **2019**, 15, 809.

[9] P. B. Allen, J. L. Feldman, *Phys. Rev. B* **1993**, 48, 12581.

[10] L. Isaeva, G. Barbalinardo, D. Donadio, S. Baroni, *Nat. Commun.* **2019**, 10, 3853.

[11] a)P. B. Allen, J. L. Feldman, J. Fabian, F. Wooten, *Philos. Mag. B-Phys. Condens. Matter Stat. Mech. Electron. Opt. Magn. Prop.* **1999**, 79, 1715; b)K. Aryana, D. A. Stewart, J. T. Gaskins, J. Nag, J. C. Read, D. H. Olson, M. K. Grobis, P. E. Hopkins, *Nat. Commun.* **2021**, 12, 2817.

[12] a)Y. Xia, V. Ozoliņš, C. Wolverton, *Phys. Rev. Lett.* **2020**, 125, 085901; b)Y. Xia, V. I. Hegde, K. Pal, X. Hua, D. Gaines, S. Patel, J. He, M. Aykol, C. Wolverton, *Phys. Rev. X* **2020**, 10, 041029.

[13] a)P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, 136, B864; b)W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, 140, A1133.

[14] V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott, G. Csányi, *J. Phys. Chem. Lett.* **2018**, 9, 2879.

[15] a)F. C. Mocanu, K. Konstantinou, T. H. Lee, N. Bernstein, V. L. Deringer, G. Csányi, S. R. Elliott, *J. Phys. Chem. B* **2018**, 122, 8998; b)A. P. Bartók, J. Kermode, N. Bernstein, G. Csányi, *Phys. Rev. X* **2018**, 8, 041048.

[16] a)V. L. Deringer, G. Csányi, *Phys. Rev. B* **2017**, 95, 094203; b)V. L. Deringer, N. Bernstein, G. Csányi, C.

Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, S. R. Elliott, *Nature* **2021**, 589, 59.

[17] a)V. L. Deringer, M. A. Caro, G. Csányi, *Adv. Mater.* **2019**, 31, 1902765; b)Y. Liu, W. Hong, B. Cao, *Adv. Theory Simul.* **2022**, 5, 2200037; c)V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, G. Csányi, *Chem. Rev.* **2021**, 121, 10073; d)P. Friederich, F. Häse, J. Proppe, A. Aspuru-Guzik, *Nat. Mater.* **2021**, 20, 750.

[18] V. L. Deringer, C. Merlet, Y. Hu, T. H. Lee, J. A. Kattirtzi, O. Pecher, G. Csányi, S. R. Elliott, C. P. Grey, *Chem. Commun.* **2018**, 54, 5988.

[19] a)G. Sivaraman, A. N. Krishnamoorthy, M. Baur, C. Holm, M. Stan, G. Csányi, C. Benmore, Á. Vázquez-Mayagoitia, *npj Comput. Mater.* **2020**, 6, 104; b)G. Sivaraman, L. Gallington, A. N. Krishnamoorthy, M. Stan, G. Csányi, Á. Vázquez-Mayagoitia, C. J. Benmore, *Phys. Rev. Lett.* **2021**, 126, 156002.

[20] G. C. Sosso, G. Miceli, S. Caravati, J. Behler, M. Bernasconi, *Phys. Rev. B* **2012**, 85, 174103.

[21] a)X. Qian, S. Peng, X. Li, Y. Wei, R. Yang, *Mater. Today Phys.* **2019**, 10, 100140; b)P. Korotaev, I. Novoselov, A. Yanilkin, A. Shapeev, *Phys. Rev. B* **2019**, 100, 144308; c)B. Mortazavi, E. V. Podryabinkin, S. Roche, T. Rabczuk, X. Zhuang, A. V. Shapeev, *Mater. Horiz.* **2020**, 7, 2359.

[22] a)Y.-B. Liu, J.-Y. Yang, G.-M. Xin, L.-H. Liu, G. Csányi, B.-Y. Cao, *J. Chem. Phys.* **2020**, 153, 144501; b)M. J. Tadjer, *Science* **2022**, 378, 724.

[23] A. P. Bartok, M. C. Payne, R. Kondor, G. Csanyi, *Phys. Rev. Lett.* **2010**, 104, 4, 136403.

[24] Z. Han, H. Liang, W. Huo, X. Zhu, X. Du, Z. Mei, *Adv. Opt. Mater.* **2020**, 8, 1901833.

[25] G. Yang, B.-Y. Cao, *arXiv preprint arXiv:2208.08467* **2022**.

[26] a)S. M. Lee, D. G. Cahill, *J. Appl. Phys.* **1997**, 81, 2590; b)Y.-C. Hua, B.-Y. Cao, *J. Appl. Phys.* **2021**, 129, 125107.

[27] N. Bernstein, G. Csányi, V. L. Deringer, *npj Comput. Mater.* **2019**, 5, 99.

[28] M. W. Mahoney, P. Drineas, *Proc. Natl. Acad. Sci.* **2009**, 106, 697.

[29] S. Plimpton, *J. Comput. Phys.* **1995**, 117, 1.

[30] R. J. Hardy, *Phys. Rev.* **1963**, 132, 168.

[31] A. Stukowski, *Model. Simul. Mater. Sci. Eng.* **2009**, 18, 015012.

[32] a)D. S. Franzblau, *Phys. Rev. B* **1991**, 44, 4925; b)S. Le Roux, V. Petkov, *J. Appl. Crystallogr.* **2010**, 43, 181.

[33] a)J. Chen, G. Zhang, B. Li, *Nano Lett.* **2010**, 10, 3978; b)L. Yang, N. Yang, B. Li, *Nano Lett.* **2014**, 14, 1734;

c)J. Moon, A. J. Minnich, *RSC Adv.* **2016**, 6, 105154.

[34] A. Giri, C. J. Dionne, P. E. Hopkins, *npj Comput. Mater.* **2022**, 8, 55.

[35] A. P. Bartok, R. Kondor, G. Csanyi, *Phys. Rev. B* **2013**, 87, 16, 184115.

[36] a)A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti, *Sci. Adv.* **2017**, 3, e1701816; b)A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, 87, 184115.

[37] a)N. Bernstein, B. Bhattarai, G. Csányi, D. A. Drabold, S. R. Elliott, V. L. Deringer, *Angew. Chem.-Int. Edit.* **2019**, 58, 7057; b)S. De, A. P. Bartók, G. Csányi, M. Ceriotti, *Phys. Chem. Chem. Phys.* **2016**, 18, 13754; c)J. Mavračić, F. C. Mocanu, V. L. Deringer, G. Csányi, S. R. Elliott, *J. Phys. Chem. Lett.* **2018**, 9, 2985; d)B. Cheng, R.-R. Griffiths, S. Wengert, C. Kunkel, T. Stenczel, B. Zhu, V. L. Deringer, N. Bernstein, J. T. Margraf, K. Reuter, G. Csanyi, *Accounts Chem. Res.* **2020**, 53, 1981; e)A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, R. Q. Snurr, *Matter* **2021**, 4, 1578.

[38] S. Wold, K. Esbensen, P. Geladi, *Chemometrics Intell. Lab. Syst.* **1987**, 2, 37.

[39] J. Kim, T. Sekiya, N. Miyokawa, N. Watanabe, K. Kimoto, K. Ide, Y. Toda, S. Ueda, N. Ohashi, H. Hiramatsu, H. Hosono, T. Kamiya, *NPG Asia Mater.* **2017**, 9, e359.

[40] a)J. M. Larkin, A. J. H. McGaughey, *Phys. Rev. B* **2014**, 89, 144303; b)L. Yang, B.-Y. Cao, *physica status solidi (RRL) – Rapid Research Letters* **2022**, 16, 2200217.

[41] Y. Zhou, X. Zhang, M. Hu, *Phys. Rev. B* **2015**, 92, 195204.

[42] C. J. Pickard, R. J. Needs, *J. Phys.-Condes. Matter* **2011**, 23, 053201.

[43] D. Packwood, J. Kermode, L. Mones, N. Bernstein, J. Woolley, N. Gould, C. Ortner, G. Csányi, *J. Chem. Phys.* **2016**, 144, 164109.

[44] a)G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **1996**, 6, 15; b)G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, 59, 1758.

[45] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, 77, 3865.

[46] P. E. Blöchl, *Phys. Rev. B* **1994**, 50, 17953.