**S1 For the supplementary description of DPMD (Deep Potential Molecular Dynamics) calculations and methods.**

**1. Density functional theory (DFT) calculations**

All first-principles calculations were performed using the Vienna ab initio simulation package (VASP)[1-3]. The Projector Augmented Wave (PAW) method was used to characterise the interaction between electrons and ions. To ensure the accuracy of the calculations, the energy and force convergence criteria were set to 10−7 eV and 10−4 eV Å−1, respectively. The kinetic energy cutoff was fixed at 550 eV. The Brillouin zone was sampled using a 5 × 5 × 5 Monkhorst–Pack K-point grid. The phonon dispersion was obtained by using the finite displacement method as implemented in the Phonopy package. Lattice transport calculations: The lattice thermal conductivity was calculated by solving the phonon Boltzmann transport equation using the ShengBTE program[4, 5]. The second- and third-order interatomic force constant (IFC) matrices required for the calculations were obtained by applying the finite displacement method to a 2 × 2 × 2 supercell using Phonopy and by solving the Boltzmann equation iteratively using thirdorder.py[5], respectively. The formula for the lattice thermal conductivity is:

**2. DP-GEN iterations**

**Figure S1** shows the distribution of at different temperatures in the first four iterations[6]. And **Table S1** displays the percentage of the accurate, candidate, and failed groups of configurations in each iteration. In the first iteration, it is not surprising that the trajectories given by the preliminary models include lots of unreasonable configurations and high-temperature simulations blow up very quickly. A large fraction of the snapshots sampled in this iteration have a larger than 0.4 eV/Å **[Fig. S1(a)]**. A large portion of the candidates with fallen in the selection range and selected for labeling are from low-temperature simulations. This situation is drastically improved after just adding 300 labeled configurations to the training dataset. In the second iteration, most low-temperature snapshots are labeled as “accurate” and the majority of newly selected snapshots come from higher-temperature simulations. Going from the second iteration to the third and the fourth, although the time duration of the simulation is extended (i.e., 1000 fs, 5000 fs, and 10 000 fs, respectively), most snapshots have their value at a satisfactory level, demonstrating a quick convergence of the DP-GEN process. After four iterations, the models have converged in the original cell (50 atoms), i.e., the percentage of candidates being ~1%. The fifth to eighth iterations are performed with 2 × 2 × 2 supercells (200 atoms) with the percentage of candidates gradually decreasing to 0.6%. Then, the exploration moves to disorder structures from the ninth to the sixteenth iterations. Due to the similarity of the structures, only a small percentage of newly sampled ones are labeled as “candidate” in these iterations. In the last three iterations, the exploration is run with NVT using the experimental lattice parameters, the same setting we use in the calculation of diffusion coefficients. Still, a few candidates arise in these iterations, yet the DP models converge within three iterations. Around 4000 configurations in total are collected via DP-GEN automatically to train the DP mod.

**3. Deep potential generation**

**The content of this subsection is derived from the research by Zeng[7] et al., and further details can be found in the relevant literature.**

A Deep Potential (DP) model, denoted by , can be generally represented as

where is the fitting properties, is the fitting network (This explanation can be seen from the work of Zeng[7] et al), and is the descriptor (This explanation can be seen from the work of Zeng[7] et al). , with being the Cartesian coordinates and being the chemical species, denotes the degrees of freedom of the atom . The indices of the neighboring atoms (i.e., atoms within a certain cutoff radius) of atom are given by the notation . Note that the Cartesian coordinates can be either under the periodic boundary condition (PBC) or in vacuum (under the open boundary condition). The network parameters are denoted by , where and yield the network parameters of the descriptor (if any)

and those of the fitting network, respectively. From Eq. (1), one may compute the global property of the system by

where is the number of atoms in a frame. For example, if represents the potential energy contribution of atom , then gives the total potential energy of the frame. In the following text, is the expected maximum number of neighboring atoms, which is the same constant for all atoms over all frames. A matrix with a dimension of will be padded if the number of neighboring atoms is less than .

3.1 Neural networks

A neural network ( NN ) function is the composition of multiple layers ,

In the DeePMD-kit package, a layer may be one of the following forms depending on whether a ResNet is used and the number of nodes:

where is the input vector and is the output vector. and are weights and biases, respectively, both of which are trainable. can be either a trainable vector, which represents the "timestep" in the skip connection, or a vector of all ones , which disables the time step. is the activation function. In theory, the activation function can be any form, and the following functions are provided in the DeePMD-kit package: hyperbolic tangent (tanh), rectified linear unit (ReLU), ReLU6, softplus, sigmoid, Gaussian error linear unit (GELU), and identity. Among these activation functions, ReLU and ReLU6 are not continuous in the first-order derivative, and others are continuous up to the second-order derivative.

3.2 Descriptors

DeePMD-kit supports multiple atomic descriptors, including the local frame descriptor, the two-body and three-body embedding DeepPot-SE descriptor, the attention-based descriptor, and the hybrid descriptor that is defined as a combination of multiple descriptors. In the following text, we use to represent the atomic descriptor of the atom .  
a. Local frame. The local frame descriptor (sometimes simply called the DPMD model), which is the first version of the DP descriptor, is constructed by using either full information or radial-only information,

where are three Cartesian coordinates of the relative position between atoms and , i.e., in the local frame, and is its norm. In Eq. (6), the order of the neighbors is sorted in ascending order according to their distance to the atom . is transformed from the global relative coordinate through

where

is the rotation matrix constructed by

where denotes the operation of normalizing a vector. and are the two axis atoms used to define the axes of the local frame of atom , which, in general, are the two closest atoms, independently of their species, together with the center atom .

The limitation of the local frame descriptor is that it is not smooth at the cutoff radius and the exchanging of the order of two nearest neighbors [i.e., the swapping of and ], so its usage is limited. We note that the local frame descriptor is the only nonsmooth descriptor among all DP descriptors, and we recommend using other descriptors for the usual system.

b. Two-body embedding DeepPot-SE. The two-body embedding smooth edition of the DP descriptor is usually named DeepPot-SE descriptor. It is noted that the descriptor is a multi-body representation of the local environment of the atom . We call it "two-body embedding" because the embedding network takes only the distance between atoms and (see below), but it is not implied that the descriptor takes only the pairwise information between and its neighbors. The descriptor, using either full information or radial-only information, is given by

where is the coordinate matrix, and each row of can be constructed as

where is the relative coordinate and is its norm. The switching function is defined as

where switches from 0 at to 1 at the cutoff radius and switches from 1 at to 0 at . The switching function is smooth in the sense that the second-order derivative is continuous. The derivation process of the fifth-order polynomial can be found in the work of Zeng et al.

Each row of the embedding matrix consists of nodes from the output layer of an NN function of ,

where the NN function was given in Eq. (5) and the subscript " " is used to distinguish the NN from other NNs used in the DP model. In Eq. (15), the network parameters are not explicitly written. only takes first columns of to reduce the size of , and are hyperparameters provided by the user. Compared to the local frame descriptor, the DeepPot-SE is continuous up to the second-order derivative in its domain.

c. Three-body embedding DeepPot-SE. The three-body embedding DeepPot-SE descriptor incorporates bond-angle information, making the model more accurate. The descriptor can be represented as

where is defined by Eq. (13). Currently, only the full information case of is supported by the three-body embedding. Similar to Eq. (15), each element of comes from nodes from the output layer of an NN function,

where considers the angle form of two neighbors ( and ). The notation " " in Eq. (16) indicates the contraction between matrix and the first two dimensions of tensor . The network parameters are also not explicitly written in Eq. (17).  
d. Handling the systems composed of multiple chemical species. For a system with multiple chemical species , parameters of the embedding network are as follows chemical-specieswise in Eqs. (15) and (17):

Thus, there will be or embedding networks, where is the number of chemical species. To improve the performance of matrix operations, is divided into blocks of different chemical species. Each matrix with a dimension of is divided into corresponding blocks, and each block is padded to separately. The limitation of this approach is that when there are large numbers of chemical species, such as 57 elements in the OC2M dataset, the number of embedding networks will become 3249 or 57 , requiring large memory and decreasing computing efficiency.

For other specific information, please refer to the work of Zeng[7] et al

During the training process, the loss function should be minimized to optimize the parameters of fitting neural networks, given by

and 1 denote the batch size and the index of the training datasets. and are the energy, the atomic force and the virial stress predicted by the DP MODEL, respectively. and are the corresponding DFT results. The force and the virial tensor are calculated using energy by:

and represent the initial prefactor and the final prefactor, respectively. The learning rate, denoted as , is determined by the initial learning rate and the current step . is defined as

Here and are terms used to represent the number of steps at which decay occurs and the rate at which decay occurs, respectively.

**4. Training of the model**

The Deep Potential GENerator (DP-GEN)[8] package is utilized to produce a collection of training datasets that could comprehensively cover the necessary range of relevant configurational space in an efficient manner. DP-GEN performs four primary stages during each generating iteration. These stages include (1) creating the initial datasets, (2) training the trial DP models, (3) configuration exploration, and (4) labeling. In this study, the primary stages of DP-GEN are done using DeepMD-kit, LAMMPS, and Vienna Ab initio Simulation Package (VASP), respectively.

**5. Creating the initial datasets**

To construct the foundational datasets essential for our study, we embarked on a meticulous selection process of various structural configurations as precursors for AIMD simulations using the VASP. Our selection encompassed structures: Platinum HCP, FCC, and BCC samples, as well as pressure-perturbed Pt configurations. The compilation of initial datasets is systematically cataloged in Table S2, while Table S3 details the specific parameters that guided our AIMD simulations. Within these simulations, the energy, force, and virial tensor calculations were conducted using the Perdew, Burke, and Ernzerhof (PBE) exchange-correlation functional, grounded in the principles of DFT. Subsequently, the results from the AIMD simulations were seamlessly transformed into the DeePMD-kit format, courtesy of a conversion tool named dpdata. This meticulous process culminated in the formation of comprehensive initial datasets, laying a solid foundation for our subsequent analyses.

**6. Training the trial DPs models**

Different random seeds were used to generate four different trial DP models based on the initial datasets. The relevant hyperparameters used for training these DP models are shown in **Table S4**.

**7. Exploring the configurations**

To overcome potential limitations in structural diversity imposed by the initial temperature setting of , we execute the configuration exploration phase using LAMMPS[9]. In this phase, a trial DP model, trained from the initial training datasets, is employed for MD simulations. This strategy enables extensive exploration of the sample space, ensuring comprehensive coverage of structural configurations. Subsequently, the initial configurations were evolved within the NPT ensemble, and the exploration strategy was presented in **Table S5**. During each exploration process, the maximum deviation of the predicted forces in the four trial DP models ( is used as the criterion for selecting reasonable configurations:

is the average of trial DP models predicted forces, and is the predicted force on the ith atom. Only the reasonable configurations could be selected into VASP for single point DFT calculations. In this study, reasonable structures were chosen, whose is in a range of 0.05 to .

**8. Labelling the configurations and training the model**

The labelling stage is conducted utilizing VASP, primarily via singlepoint DFT calculations to generate labels. **Table S3** presents a comprehensive compilation of hyperparameters employed in these calculations. Then, the datasets generated by DP-GEN were utilized to train the DP MODEL of Pt. The DP model is trained by using DeePMD-kit, and the hyperparameters of training are illustrated in **Table S4**.

**9. Simulation and calculation settings**

We conducted MD simulations using LAMMPS to calculate both the thermal conductivity of . In the calculation of thermal conductivity, we utilize the ShengBTE code, which entails the combination of second-order and third-order interatomic force constants, followed by the solution of the phonon Boltzmann transport equation. The phonopy[10] code calculate the second-order interatomic force constants, and the lattice thermal conductivity is calculated using the NEMD method.

**10. Built-in Mechanism of DP-GEN for Overfitting Prevention**

DP-GEN dynamically generates training data through an iterative active learning process, ensuring that the training dataset covers a wide state space of the target physical system. This mechanism significantly reduces the likelihood of overfitting, as the model is consistently trained with high-quality, newly generated data rather than relying on a fixed, potentially biased initial dataset. In each iteration, DP-GEN employs multiple deep potential models to predict the target system and identifies inadequacies in the data by analyzing inconsistencies in model predictions (such as deviations in force or energy distributions). These inadequacies are then supplemented into the training dataset. This dynamic data augmentation approach increases the diversity of the training data, ensuring the model does not overfit to any specific state.

**11. Convergence Rate Assurance in DP-GEN**

The DP-GEN framework enhances model convergence efficiency through the following methods:

1. Active Learning Strategy: By evaluating prediction errors, it prioritizes generating the most valuable data for model learning, thereby accelerating model convergence.

2.Parallel Training and Validation: DP-GEN incorporates the capability for parallel training across multiple models, validating various potential data states in each iteration. This approach ensures that the training process converges rapidly in a stable direction.

3.Unified Convergence Criteria: The iterative process of DP-GEN is automatically terminated based on convergence conditions (thresholds for prediction errors and maximum force deviations). Training stops when the model reaches the specified accuracy. This automated strategy, based on global performance, further reduces errors arising from manual parameter settings.

**12. Overfitting Manifestations and Testing**

During training, DP-GEN monitors the model's performance on validation and test sets to ensure there is no overfitting. In this study, we paid particular attention to the maximum force deviations of the model under different temperature conditions (as shown in Fig. S1). Experimental results indicate that even under complex physical states (e.g., high temperature, extreme pressure), the deep potential models generated by DP-GEN maintain low deviations. This fully validates the model's broad applicability and robustness.

**13. Use RMSE values to ensure model reliability for different Pt stages (FCC, HCP, BCC)**

13.1 Phase-Specific Evaluation:

RMSE values were calculated separately for each Pt phase (FCC, HCP, BCC) to evaluate how well the DP model replicates the DFT-calculated energies and forces in distinct crystal structures. This phase-specific breakdown allowed us to identify any potential biases or inaccuracies in the model for particular configurations.

13.2 Training and Validation Consistency:

During the iterative training process, RMSE values were monitored for both the training and validation datasets. This ensured that the model generalized well and avoided overfitting to the training data, thereby maintaining accuracy across different phases.

13.3 Threshold Analysis:

RMSE values were compared to a predefined threshold (based on acceptable deviation levels from DFT benchmarks) to ensure the DP model's predictions were within a range considered reliable for molecular dynamics simulations. For all phases, the RMSE values for both energy and force were well below this threshold, confirming the model's robustness.

13.4 Comparison with Other Potentials:

To further validate reliability, RMSE values from the DP model were benchmarked against those from other potential models, such as EAM and MEAM. This comparison highlighted the superior performance of DP in capturing phase-specific properties with high fidelity.

By using these RMSE metrics across different Pt phases, we ensured the DP model's predictive reliability, demonstrating its capability to handle diverse crystal structures with accuracy comparable to DFT calculations.

**14. The influence of heating rate on phase transition behavior**

14.1 Shift in Melting Point:

Rapid heating rates tend to elevate the observed melting point compared to experimental conditions because the system has less time to reach thermodynamic equilibrium. As a result, the phase transition might occur at slightly higher temperatures than under slower, experimentally relevant heating rates.

14.2 Metastable States:

High heating rates can trap the system in metastable configurations, as the atoms may not have sufficient time to rearrange into lower-energy configurations. This could influence the pathway and dynamics of the phase transition observed in simulations.

14.3 Kinetic Effects:

The faster heating rate amplifies kinetic effects, potentially masking slow diffusion-driven phenomena or other equilibrium processes that occur under experimental conditions. This means the results focus more on the dynamic response rather than equilibrium melting behavior.

Despite these limitations, the high heating rate is a necessary trade-off in molecular dynamics simulations to make the calculations computationally feasible. The observed melting point trends remain consistent with experimental data, and the relative comparisons between DP, DFT, and other models retain their validity, providing meaningful insights into phase transition behavior.

**References**

[1] G Kresse, Furthmuller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B Condens Matter, 54(16): 11169-86. 1996, doi: 10.1103/physrevb.54.11169

[2] G Kresse, Furthmüller J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Computational Materials Science, 6(1): 15-50. 1996, doi: 10.1016/0927-0256(96)00008-0

[3] G Kresse, Hafner J. Ab initio molecular dynamics for liquid metals. Phys Rev B Condens Matter, 47(1): 558-61. 1993, doi: 10.1103/physrevb.47.558

[4] W Li, Carrete J, Katcho NA, Mingo N. ShengBTE: A solver of the Boltzmann transport equation for phonons. Computer Physics Communications, 185(6): 1747-58. 2014, doi: 10.1016/j.cpc.2014.02.015

[5] Z Han, Yang X, Li W, Feng T, Ruan X. FourPhonon: An extension module to ShengBTE for computing four-phonon scattering rates and thermal conductivity. Computer Physics Communications, 270. 2022, doi: 10.1016/j.cpc.2021.108179

[6] J Huang, et al. Deep potential generation scheme and simulation protocol for the Li(10)GeP(2)S(12)-type superionic conductors. J Chem Phys, 154(9): 094703. 2021, doi: 10.1063/5.0041849

[7] JZ Zeng, et al. DeePMD-kit v2: A software package for deep potential models. J Chem Phys, 159(5): 55-60. 2023, doi: 10.1063/5.0155600

[8] Y Zhang, et al. DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models. Computer Physics Communications, 253: 107206. 2020, doi: <https://doi.org/10.1016/j.cpc.2020.107206>

[9] AP Thompson, et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. Computer Physics Communications, 271: 108171. 2022, doi: 10.1016/j.cpc.2021.108171

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