



# Thermal conductivity of nanoporous silicon: molecular dynamics simulations and machine learning prediction

V. V. Kuryliuk  and O. Ya. Olikh 

*Taras Shevchenko National University of Kyiv, Kyiv 01033, Ukraine*

E-mail: kuryliuk@knu.ua; olegolikh@knu.ua

Received August 8, 2025, published online November 17, 2025

The thermal conductivity of nanoporous silicon (*p*-Si) was investigated using equilibrium molecular dynamics simulations and machine learning techniques. Several interatomic potentials were evaluated, with the Tersoff potential selected for its stability and accuracy. Thermal conductivity values were computed over a wide range of temperatures and porosities. A **Symbolic Regression** algorithm was used to derive an analytical expression describing its dependence on temperature and porosity. Additionally, Random Forest and Gradient Boosting models were trained to reconstruct heat current autocorrelation functions and predict thermal conductivity, outperforming Support Vector Regression. The combined approach enables accurate, data-efficient prediction of thermal transport in porous nanostructures, supporting the design of materials for thermal management and energy conversion applications.

Keywords: nanoporous silicon, thermal conductivity, molecular dynamics, symbolic regression, machine learning.

## 1. Introduction

In recent years, nanostructured materials have attracted significant interest for their potential in thermal insulation and heat dissipation applications [1, 2]. For instance, as 3D integration advances in micro- and nanoelectronics, there is a growing need for low thermal conductivity materials that remain compatible with CMOS processes. Furthermore, for energetically autonomous systems, the development of effective nanomaterials exhibiting high thermoelectric conversion efficiency is also crucial [3]. The thermal behavior of nanomaterials is significantly influenced by their structural dimensions and the manner in which phonons propagate through their frameworks [4]. In such materials, thermal phonons are frequently scattered at surfaces and interfaces, thereby significantly lowering overall thermal conductivity [5]. It is noteworthy that nanoporous structures have attracted significant interest due to their large specific surface area and shortened phonon mean free path, which further suppresses heat transport [6, 7].

Nanoporous silicon (*p*-Si), produced by electrochemical anodization of crystalline silicon, has attracted significant attention because its optical, electrical, and thermal properties can be readily tuned by adjusting the porosity. This controllability enables a wide range of potential applications for *p*-Si in areas such as highly sensitive sensors [8], lithium-ion batteries [9], optoelectronics, micro- and nanoelectronics [10], as well as MEMS technologies [11]. It has been demonstrated that nanostructured silicon with

tailored porosity exhibits thermal conductivity that is two to three orders of magnitude lower than that of bulk crystalline silicon ( $\sim 140 \text{ W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$ ) [12]. This renders it an appealing candidate for utilization in thermal sensors [13], thermal insulation, and thermoelectric generators in silicon-based microsystems [14]. In order to gain deeper insight into the mechanisms of phonon transport in *p*-Si, it is crucial to undertake comprehensive thermal characterization. Furthermore, it is imperative to possess precise knowledge of the thermal behavior of this material to ensure its effective integration into practical devices.

Molecular dynamics (MD) simulations have become an indispensable tool for investigating thermal transport phenomena in nanostructured materials, providing valuable atomistic insights that are often inaccessible through experiments alone. In particular, MD methods have been extensively utilized to predict the thermal conductivity of porous silicon and analogous systems, elucidating the profound impact of nanoscale characteristics, such as pore size, shape, and distribution, on phonon scattering and heat conduction [15–18]. The integration of machine learning (ML) approaches with atomistic modelling has emerged as a promising strategy to efficiently predict and optimize the properties of complex materials by uncovering hidden structure-property relationships. Two primary approaches integrate ML with MD. The first one employs machine learning interatomic potentials in simulations, enabling the study of large-scale emergent phenomena in specific materials with

accuracy close to that of *ab initio* MD [19, 20]. The second approach uses MD simulation results from a limited number of cases as input data to build and train ML models; these models are then applied to a much wider range of predictions [21, 22]. Both methods have been successfully used to assess the thermal properties of various materials [23–27], including porous structures [28]. Previous studies have shown that such integrated frameworks can significantly accelerate the design of advanced nanostructured materials for thermal management and energy conversion applications. Unfortunately, to the best of our knowledge, no similar studies have been conducted on porous silicon.

In this study, molecular dynamics and machine learning methods were employed to calculate the thermal conductivity of nanoporous silicon. A series of interatomic interaction potentials was tested to predict heat transport in silicon. The equilibrium molecular dynamics approach was used to determine the dependence of the thermal conductivity on temperature and porosity. The resulting dataset was then processed and generalized using machine learning models, allowing the thermal conductivity of *p*-Si to be estimated over a wide range of temperatures and porosities.

## 2. Methods

The thermal conductivity of the *p*-Si was calculated via equilibrium molecular dynamics (EMD). All MD simulations were performed using the LAMMPS package [29]. The EMD simulation setup considered in this paper is schematically illustrated in Fig. 1. The simulation box is a cube consisting of  $L^3 = 10a \times 10a \times 10a$  unit cells, where  $a$  is the unit-cell lattice vector in the Si. The porosity of the structure was varied by changing the pore radius  $R$  and was calculated as  $\bar{p} = \frac{4\pi R^3}{3L^3}$ .

The pore was introduced as a spherical cavity by removing atoms from the silicon lattice and was placed at the center of the simulation box to ensure symmetry and minimize boundary effects. While real materials often exhibit randomly distributed pores of irregular shape, modern fabrication techniques also enable periodic porous structures [30, 31].

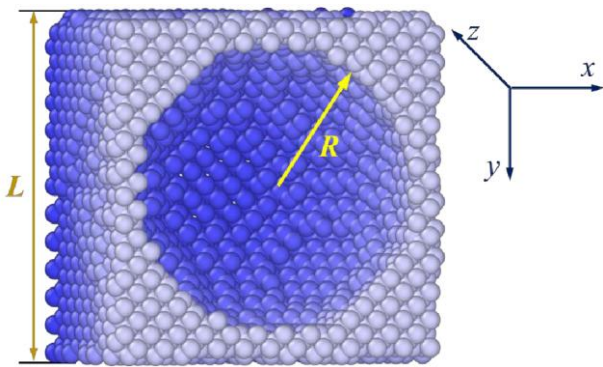


Fig. 1. (Color online) Schematic illustration of the EMD simulation setup:  $L$  is size of the Si supercell (cube),  $R$  is pore radius.

Furthermore, recent studies have demonstrated that the effective thermal conductivity of porous silicon is not only determined by the overall porosity but also strongly depends on the pore morphology and its spatial correlations. For a given porosity, different pore shapes (e.g., cylindrical, slit-like, or irregular) can lead to markedly different phonon scattering rates and thus to distinct thermal conductivities, often accompanied by anisotropy in heat transport [32, 33]. Moreover, interactions between neighboring pores, such as pore–pore coupling and percolation effects, have been shown to further reduce the thermal conductivity compared to isolated pores [34, 35]. In the present work, we restrict our attention to the case of spherical pores created by atom removal in molecular dynamics simulations, which provides a simplified but instructive model system. This allows us to isolate the effect of porosity itself, while future studies may extend to more complex pore geometries and correlated pore networks.

For each temperature and porosity, an initial simulation was performed in the isobaric–isothermal (NPT) ensemble to generate thermally equilibrated configurations. The system was then further equilibrated in an isochoric–isothermal (NVT) ensemble. In both the NPT and NVT simulations, the temperature was controlled using a Nosé–Hoover thermostat. The simulation time for each ensemble was 0.5 ns, and the integration of Newton’s equations of motion was carried out using the Verlet algorithm with a time step of 1 fs. After reaching equilibrium, the system was switched to a constant-volume and constant-energy (NVE) ensemble to determine the thermal conductivity  $k$  using the Green–Kubo method. In this method,  $k$  was estimated by integrating the heat current autocorrelation function (HCACF) as follows:

$$k = \frac{1}{3Vk_B T^2} \int_0^{t_c} \langle \mathbf{J}(t) \mathbf{J}(0) \rangle dt, \quad (1)$$

where  $V$  is the volume of the system,  $t$  is the time, the brackets  $\langle \rangle$  is the ensemble-averaged HCACF and  $\mathbf{J}$  is the heat current vector that is defined as

$$\mathbf{J}(t) = \frac{1}{V} \left[ \sum_{j=1}^N \mathbf{v}_j E_j - \sum_{\alpha=1}^2 h_{\alpha} \sum_{j=1}^{N_{\alpha}} \mathbf{v}_{\alpha j} \right] + \frac{1}{V} \left[ \frac{1}{2} \sum_{j=1}^N \sum_{\substack{i=1 \\ i \neq j}}^N \mathbf{r}_{ij} (\mathbf{v}_{ij} \cdot \mathbf{F}_{ij}) \right], \quad (2)$$

Here,  $\mathbf{v}_j$  and  $E_j$  are the velocity and energy of the  $j$ th particle;  $\mathbf{r}_{ij}$  and  $\mathbf{F}_{ij}$  denote the distance and the interaction force between the  $i$ th and  $j$ th particles;  $N$  is the total number of particles in the system; and  $h_{\alpha}$  is the average partial enthalpy of component  $\alpha$ . Obviously, that  $t_c$  value (the upper limit of the integral or the MD simulation time) must be sufficiently large for the system to reach equilibrium. In essence, the thermal conductivity  $k$  was determined as the

saturation value of the  $k(t_c)$  curve. To reduce statistical errors, our MD simulations were run for more than 10 ns on average, and five independent simulations were performed for each specific condition. For a structure with a given porosity, calculating the thermal conductivity on a cluster of 128 AMD EPYC cores at the same temperature typically took 10–12 hours.

The primary goal of applying Machine Learning (ML) methods was to enable the rapid estimation of the thermal conductivity of porous silicon across arbitrary values of porosity  $p$  ( $0 \leq p < 1$ ) and temperature  $T$ . A key consideration when selecting ML approaches is the limited availability of thermal conductivity data – only a few dozen values can realistically be obtained from MD simulations within a feasible timeframe for model development. As a result, conventional ML methods that require a training dataset containing tens or hundreds of thousands of samples are not suitable in this context.

As a result, two distinct approaches were used. The first approach employed Symbolic Regression (SR), an algorithm that derives a mathematical expression to describe a given dataset. Unlike conventional regression methods, which require predefined input variables and adjust numerical coefficients to achieve the best fit, SR simultaneously identifies both the structure of the equation and its parameters [36]. It uses evolutionary algorithms to uncover functional relationships and numerical coefficients between input and output variables. A key advantage of SR is its ability to produce interpretable results in the form of analytical expressions. Moreover, this method does not require large amounts of training data to achieve high predictive accuracy [36, 37].

SR was implemented in Python using the open-source PySR package [38]. Data preprocessing was limited to normalizing temperature values to 300 K. During training, the loss function was defined as the mean weighted squared error (MWSE) and was minimized:

$$\text{MWSE} = \frac{1}{N} \sum_{i=1}^N \omega_i (\hat{y}_i - y_i)^2. \quad (3)$$

Here,  $N$  represents the number of values in the training set,  $y_i$  denotes the true thermal conductivity (result of MD simulation), while  $\hat{y}_i$  indicates the predicted value;  $\omega_i$  is the weighting coefficient, in this study, statistical (reciprocal) weighting was applied, meaning  $\omega_i = 1/y_i$ . The use of weighting coefficients is motivated by the fact that the range of target values for thermal conductivity (TC) spans two orders of magnitude (from 1 to 226 W/mK). At the same time, the absolute errors of the values calculated using MD increase with thermal conductivity (see Fig. 3). The weighting was introduced to prevent the model from adapting primarily to large TC values (corresponding to low porosity and temperature).

In the second approach, classical algorithms — specifically Random Forest (RF), Gradient Boosting (GB), and Support Vector Regression (SVR) — were used to build models predicting the  $k(t_c)$  curve. This strategy significantly increased the size of the training dataset. The input descriptor set included porosity, temperature, and simulation time. To standardize the features and target values, they were normalized to have zero mean and unit standard deviation. The models were implemented using Python's Scikit-learn toolkit. Model parameters were optimized using the Optuna framework, which employed the TPE sampler and Hyperband pruner for efficient hyperparameter selection. A 5-fold cross-validation scheme was applied during model tuning. To determine thermal conductivity, the predicted  $k(t_c)$  curves were processed in the same way as those obtained from MD simulations.

The evaluation metrics used to assess forecast quality were the mean squared error (MSE), mean absolute percentage error (MAPE), and coefficient of determination ( $R^2$ ), as defined in Eqs. (4)–(6).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (4)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{y_i} \times 100\%, \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{(\hat{y}_i - \bar{y})^2}, \quad (6)$$

where  $\bar{y}$  is the mean of the true values. MSE was used for model tuning, while all three metrics were applied to evaluate the prediction quality of the trained models.

### 3. Results and discussion

#### 3.1. Evaluation and selection of interatomic potentials

The accuracy of MD simulations strongly depends on the choice of interatomic potential. Therefore, the initial stage of this study focused on evaluating several well-known potentials based on three key criteria: (i) computational performance, which is particularly relevant for high-throughput simulations and machine learning dataset generation; (ii) agreement between simulated and experimental temperature-dependent thermal conductivity of single-crystalline silicon; and (iii) potential stability for highly porous structures over a wide temperature range. Clearly, the latter criterion plays the most significant role in selecting an interatomic potential for calculating the thermal properties of a porous material. The tested potentials included MEAM [39], Tersoff [40], EDIP [41], Stillinger–Weber (SW) [42], modified Stillinger–Weber (SW-mod) [43], and a machine learning-based potential SNAP [44]. The results are summarized in Figs. 2 and 3.

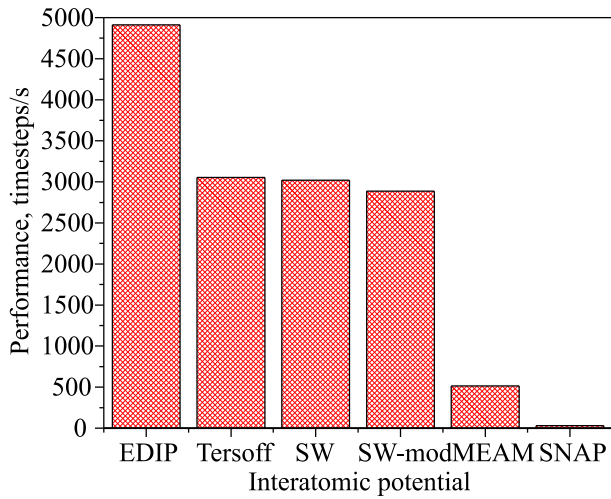


Fig. 2. (Color online) Comparison of the computational performance of various interatomic potentials for calculating the thermal conductivity of single-crystalline silicon. Structure size:  $10a \times 10a \times 10a$ ; number of cores per node: 128.

As shown in Fig. 2, the EDIP potential demonstrated the highest computational performance, reaching nearly 5000 MD steps per second on a 128-core node. Tersoff, SW, and SW-mod potentials achieved approximately 3000 steps per second, while MEAM exhibited ten times lower performance than EDIP. The SNAP potential was the slowest, with only ~30 steps per second under identical simulation and hardware conditions.

Thermal conductivity values calculated using MEAM, Tersoff, SW, and SW-mod consistently exceeded the experimental results reported in Ref. 45 across the entire temperature range, with deviations ranging from 48% to 106%. In contrast, the SNAP potential significantly underestimated the thermal conductivity, indicating inadequate parameterization for heat transport simulations. The smallest deviation between the calculated and experimental  $k(T)$  dependencies was obtained using the EDIP potential. However, this should not be regarded as an inherent advantage of

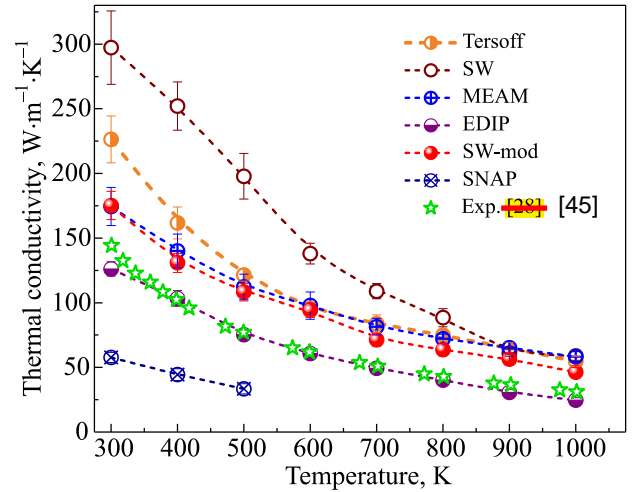


Fig. 3. (Color online) Calculated temperature dependence of thermal conductivity of single-crystalline silicon using different interatomic potentials, compared with experimental data adapted from [45].

EDIP, since thermal conductivity calculated via MD simulations is typically higher than experimental values due to the use of idealized model structures – defect-free and mono-isotopic.

Moreover, further testing of the potentials revealed that, for highly porous structures at elevated temperatures, the porous architecture became unstable with certain potentials, leading to pore collapse and the formation of nearly amorphous configurations (see Fig. 4). As shown in Table 1, similar instabilities were observed for the SW, SW-mod, and EDIP potentials.

As a result, the Tersoff potential was chosen for the remainder of the study. An additional argument in favor of the Tersoff potential is the calculated temperature dependence of  $k$ , which is often approximated as follows [46]:

$$k \propto T^{-b}. \quad (7)$$

Fitting the calculated  $k(T)$  for single-crystalline Si using Eq. (7) yielded  $b = 1.15 \pm 0.03$ , which is close to both the

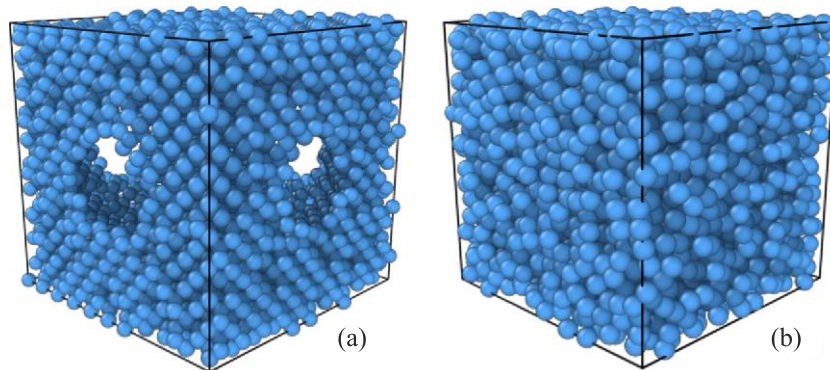


Fig. 4. (Color online) Atomic configurations of porous silicon with 50% porosity, obtained after structure equilibration at 800 K using Tersoff potentials (a) and EDIP (b).



Table 1. Stability of interatomic potentials in MD simulations of thermal conductivity for *p*-Si

Potential	Stability
MEAM	+
Tersoff	+
EDIP	-
Stillinger-Weber	-
modified Stillinger-Weber	-
SNAP	+

theoretical expectation ( $b = 1.0$  for semiconductors and dielectrics) and the experimental value ( $1.25 \pm 0.01$ ) [45]. For comparison, the values for other potentials are EDIP  $1.36 \pm 0.09$ , SW  $1.5 \pm 0.1$ , SW-mod  $1.08 \pm 0.05$ , and MEAM  $0.92 \pm 0.02$ .

### 3.2. Generation of molecular dynamics data for machine learning

The next stage of the study involved generating a dataset of MD simulation results for training machine learning models aimed at predicting the thermal properties of porous silicon structures. Two competing requirements had to be balanced. On the one hand, the simulated data needed to span a wide range of porosity and temperature values and capture all key dependencies of thermal conductivity on these parameters. On the other hand, limitations in computational resources had to be considered. Therefore, MD simulations were performed to obtain temperature-dependent thermal conductivity values at several porosity levels using 100 K intervals, as well as the dependence of thermal conductivity on pore volume fraction at room temperature. Figure 5(a) presents selected results for the temperature and porosity dependence of thermal conductivity, calculated using the Tersoff potential.

It should be noted that the observed trends are consistent with literature data: thermal conductivity decreases with increasing temperature, and the form of the  $k(T)$  dependence evolves as the system becomes increasingly disordered. A monotonic decrease in thermal conductivity with increasing porosity is also observed [Fig. 5(b)], attributed to increased phonon-boundary scattering at pore surfaces. Thus, this dependency aligns with general expectations, further supporting the validity of the performed calculations.

The presented data served as a foundation for the development and training of ML models for predicting the heat transport behavior of porous silicon. In addition, a set of *p*-Si thermal conductivity values was calculated to test the ML models at specific temperatures and porosities within the ranges of 300–1000 K and 0–80%, respectively. In summary, 29 thermal conductivity values (or  $k(t_c)$

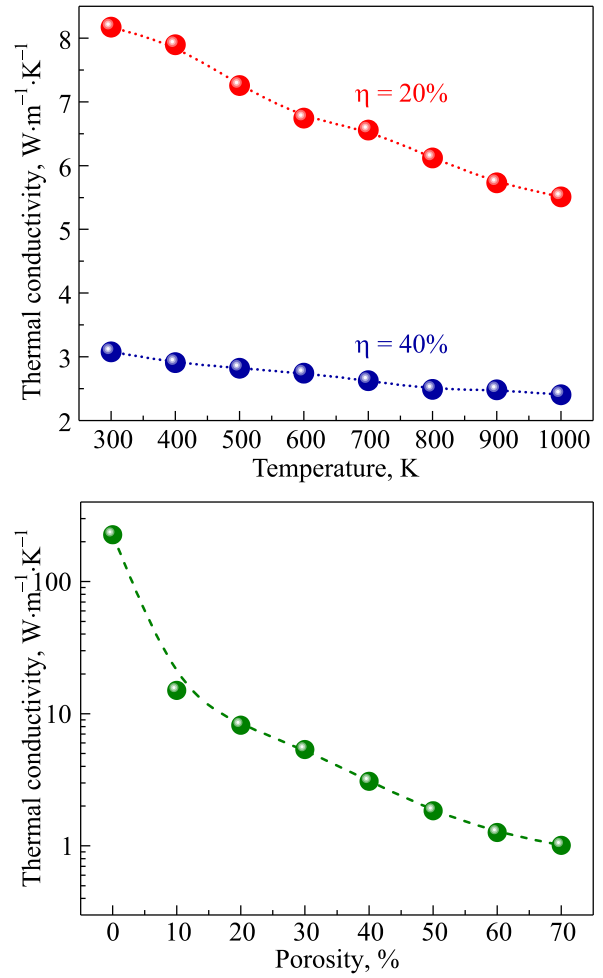


Fig. 5. (Color online) Dependence of thermal conductivity on temperature for *p*-Si with 20% and 40% porosity (a) and on porosity for *p*-Si at  $T = 300$  K (b). Calculations were performed using the Tersoff interatomic potential.

dependencies for classical ML algorithms) obtained from MD were used for training, and 9 for testing, corresponding to a 76:24 split.

### 3.3. Machine learning-based prediction of thermal transport properties in porous silicon

Using the results of MD simulations with the Tersoff potential (Figs. 3 and 5) and the Symbolic Regression algorithm, we derived an analytical expression describing the dependence of porous silicon's thermal conductivity on temperature and porosity. By adjusting parameters such as the population size, the maximum expression complexity, and the multiplicative coefficient penalizing complexity, we obtained a formula with a MAPE of no more than 2.89% and a median APE of 1.96% on the training set [see Fig. 6(a)]. The resulting analytical expression is:

$$k = \left( -7.024 + \frac{7.481}{p^{0.55} + (T_n - 0.105) / (\sin[T_n / (4.486 \cdot p - 0.251)] + 27.09)} \right) \exp(-p), \quad (8)$$

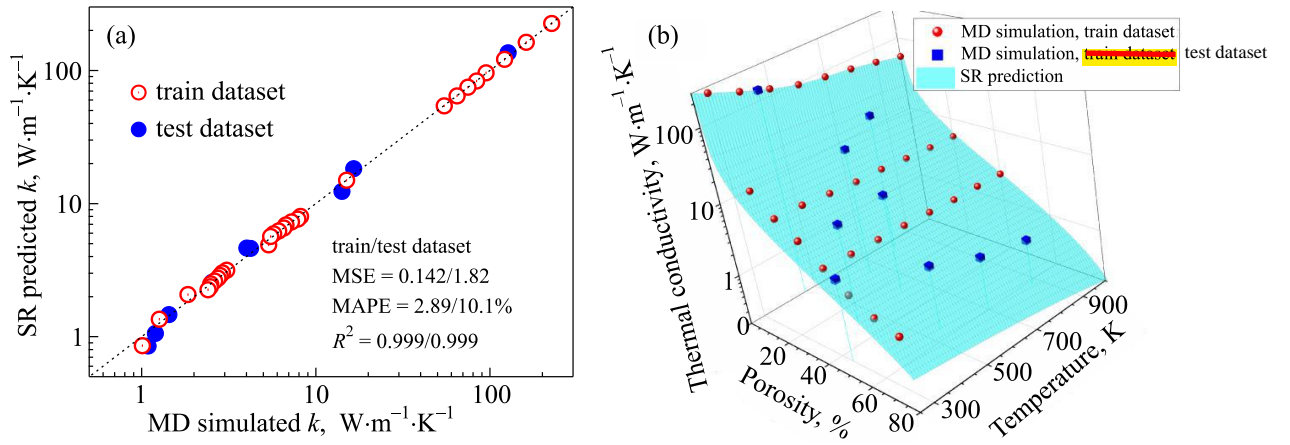


Fig. 6. (Color online) Comparison of SR predicted  $k$  and MD calculated  $k$  (true value) for  $p$ -Si. Open and closed markers indicate training and test datasets, respectively. The black lines are the identified lines serving as the references (a). Dependence of the thermal conductivity of porous silicon on temperature and porosity. The points represent the results of MD simulations, while the surface corresponds to Eq. (8) (b).

where  $T_n = T/300$ . For the test set, the average error is approximately 10%. Figure 6(b) presents the obtained  $k(T, p)$  dependence within the temperature range of 300–1000 K and the porosity range of 0–80%. The corresponding MD simulation results are also shown in the figure.

As shown in Fig. 6(b), the MD calculations do not fully cover the parameter space. However, in SR, the key objective is to identify an analytical expression that captures the underlying patterns in the data rather than merely approximating them numerically [36, 37]. Thus, SR does not require exact reproduction of all training values; its primary goal is to recover the correct functional form. Consequently, SR can potentially reconstruct the true equation even from incomplete data, which sets it apart from the “black-box” nature of standard ML models that generalize poorly without additional input. At the same time, validation of the derived expression against known physical laws or asymptotic behavior remains crucial, since the ability to reproduce trends beyond the training domain provides strong evidence of its correctness.

Due to the limited experimental data on porous silicon, the analysis is focused on the predicted  $k(T)$  and  $k(p)$  dependencies. At the same time, it is more appropriate to focus on the character of these dependencies rather than on the absolute thermal conductivity values, since passivation of the pore surface by air molecules can significantly alter  $k$ , and this factor is rarely controlled in experiments.

As noted above, an increase in temperature should correspond to a decrease in thermal conductivity [see Eq. (7)], and partial amorphization should modify the  $k(T)$  dependence with increasing disorder. In particular, experiments [46, 47] have shown that the exponent  $b$  decreases with increasing porosity. Figure 7(a) summarizes the values of  $b$  obtained from various sources. Overall, the results from both MD calculations and SR predictions generally agree with physical expectations and experimental data. A slight

increase in  $b$  at high porosity may indicate limitations in applying Eq. (8) for  $p > 0.5$ . However, the non-monotonic behavior of  $b(p)$  is also observed experimentally.

The Looyenga effective medium model is most often used [48, 49] to describe the dependence of thermal conductivity on porosity. This model accounts for percolation strength and assumes the following power dependence:

$$k \propto (1-p)^\alpha. \quad (7)$$

Moreover, the value of  $\alpha$  may depend on the type of conductivity of the semiconductor matrix [49] and is practically independent of temperature. Figure 7(b) presents the

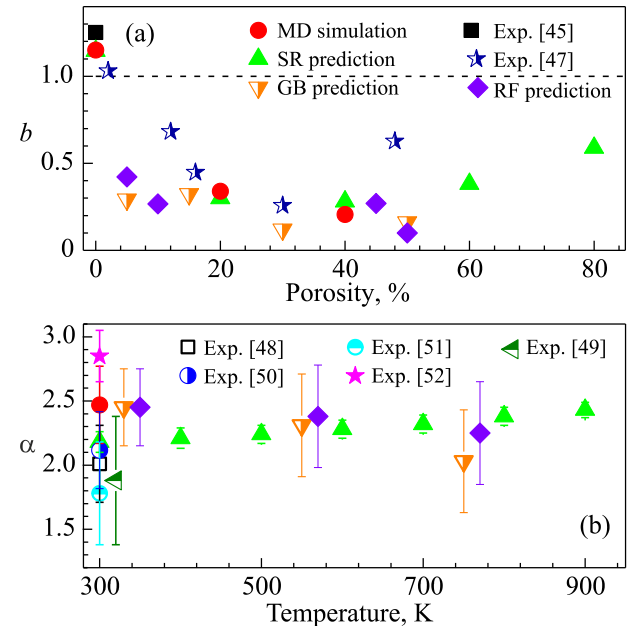


Fig. 7. (Color online) Degree indicators from Eqs. (7) (a) and Eqs. (9) (b), obtained from MD calculations, SR prediction, and experimental data. The dotted horizontal line in panel (a) represents the theoretical expectation for semiconductors and dielectrics.

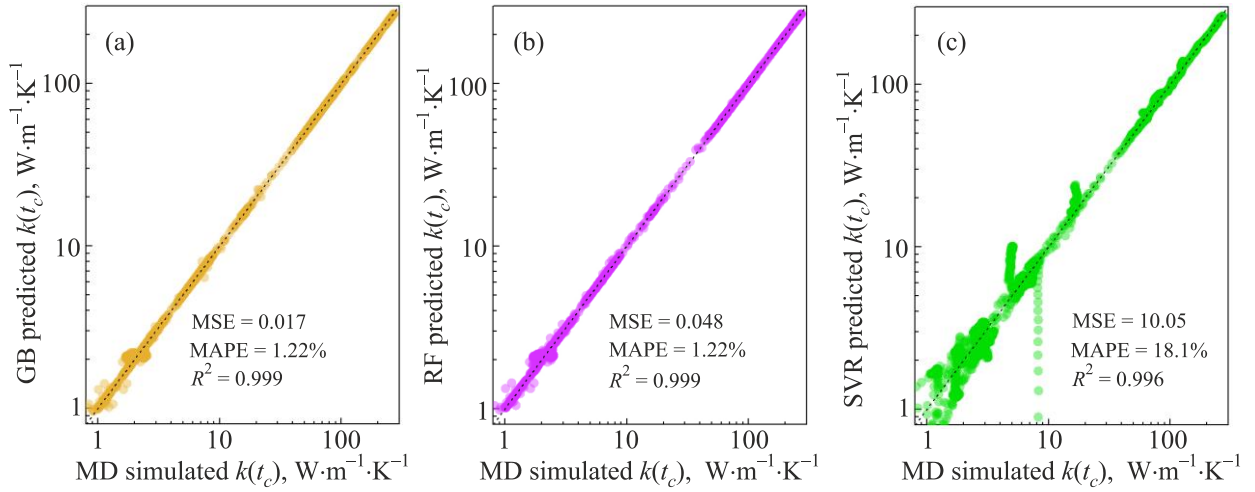


Fig. 8. (Color online) Comparison of  $k$  values at different  $t_c$ , calculated by MD simulations and predicted by Gradient Boosting (a), Random Forest (b), and Support Vector Regression (c) for the training dataset. The black lines indicate reference lines.

values of  $\alpha$  obtained by approximating  $k(p)$  dependencies from both our calculations and various experimental studies of  $p$ -type silicon. To the best of our knowledge, data for non-room temperatures are not available unfortunately. Nevertheless, the figure shows that the dependencies derived from Eq. (8) yield  $\alpha$  values with only weak temperature dependence, and their values near room temperature agree well with experimental data, consistent with physical expectations.

The initial data for each point in Figs. 3 and 5 consisted of MD simulation results for  $k$  at 10.000 values of  $t_c$ . This dataset was used for training classical ML models. The training results are shown in Fig. 8. It is evident that the performance metrics of SVR are lower than those of RF and GB, which demonstrate similar levels of accuracy.

Figure 9 shows the typical  $k(t_c)$  dependencies obtained from MD simulations and ML model predictions for the

training (a) and test (b) datasets. It should be noted that RF and GB reproduce the training curves  $k(t_c)$  with high accuracy; the deviations seen in Figures 8(a) and (b) are confined to the initial parts of the curves and do not affect the thermal conductivity values, which are determined in the saturation region. For the test set, however, both RF and GB tend to overestimate the thermal conductivity [Fig. 9(b)]. In contrast, SVR performs considerably worse, with large discrepancies from the reference values already visible for the training set and becoming even more pronounced for the test set.

In all cases, MD calculations were performed for the same  $t_c$  values, so the training set contained only these points. To test the robustness of the models, predictions of  $k(t_c)$  were also generated for perturbed data, where  $t_c$  values were multiplied by  $(1 + \epsilon)$ , where  $\epsilon$  is drawn from a normal distribution with mean 0 and standard deviation 0.05.

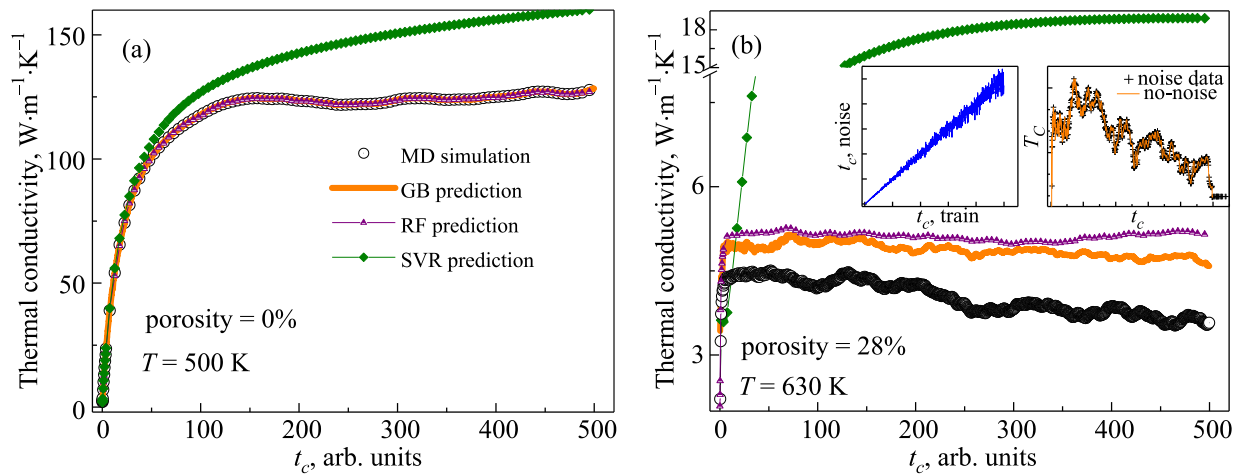


Fig. 9. (Color online) Dependencies of  $k$  on  $t_c$  from MD (black circles) and ML predictions for different porosity and temperature values in the training (a) and test (b) sets. The inset shows the dependence of noisy on noise-free  $t_c$  values (left) and GB predictions with noisy inputs (symbols) versus noise-free inputs (solid line) (right).

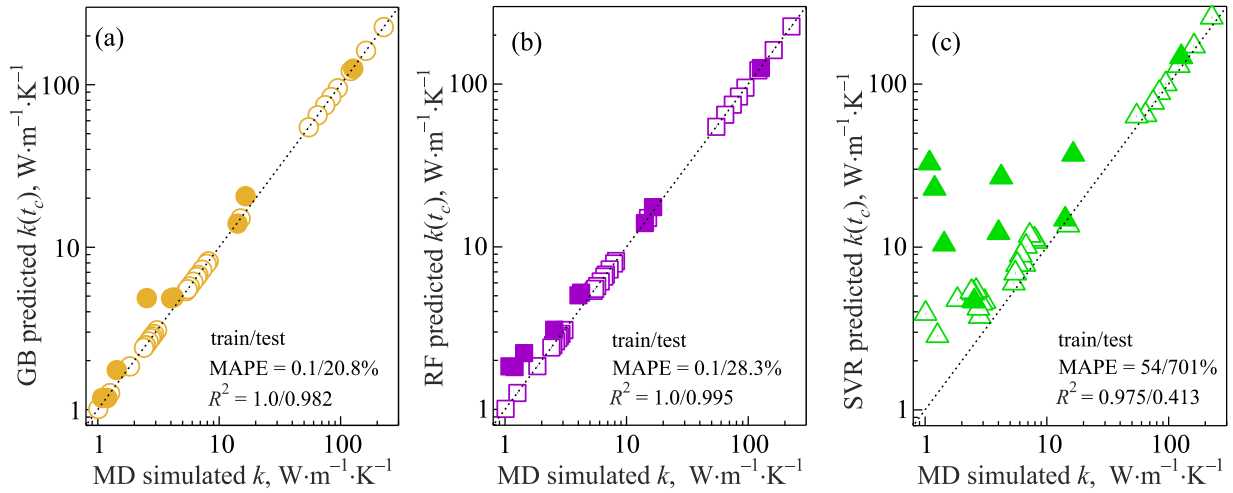


Fig. 10. (Color online) Comparison between MD-calculated thermal conductivity and predictions obtained from Gradient Boosting (a), Random Forest (b), and Support Vector Regression (c) for porous silicon. Open and closed markers indicate training and test datasets, respectively. Solid black lines represent reference trends.

The inset in Fig. 9(a) illustrates that adding noise does not alter the shape of the predicted curves. Although only GB results are shown, the same behavior was observed for the other algorithms.

Figure 10 summarizes the thermal conductivity values extracted from the  $k(t_c)$  curves for the training and test sets. Gradient Boosting provides the best accuracy on the test set (MAPE = 20.8%, median APE = 16.1%), although the errors are notably higher than for training data. Random Forest yields slightly larger deviations (MAPE = 28.3%, median APE = 23.6%), with the error increasing as  $k$  decreases (i.e., with rising temperature and/or porosity). In contrast, the poor performance of SVR, evident even on the training set, makes it unsuitable for the tasks addressed in this study.

RF and GB were also employed to predict  $k(T)$  and  $k(p)$  dependencies at  $T$  and  $p$  values beyond the MD simulation range (since, given the accuracy of the training curves, there was no reason to consider regions with MD results). The exponents of the power laws extracted from these predictions are summarized in Fig. 7. Similar to SR, both the trends and absolute values agree with general physical expectations, further confirming the feasibility of the proposed ML oriented approach for estimating the thermal conductivity of porous silicon.

#### 4. Conclusion

This study evaluated the suitability of several interatomic potentials for calculating the thermal conductivity of porous silicon via molecular dynamics simulations and confirmed the adequacy of the Tersoff potential. Thermal conductivity was then computed over a broad range of temperatures and porosity values. Based on these data, a Symbolic Regression algorithm was employed to derive an analytical expression describing the dependence of thermal conductivity on temperature (300–1000 K) and porosity (0–80%). Furthermore, it was demonstrated that Random

Forest and Gradient Boosting algorithms can reliably estimate thermal conductivity by reconstructing the time dependence of the heat current autocorrelation function. The realism of the ML model predictions is supported by the agreement between the predicted dependencies of thermal conductivity on temperature and porosity and the trends expected from physical models and experimental data. In contrast, Support Vector Regression yielded significantly less accurate results and was deemed unsuitable for this task.

#### Acknowledgments

The work was supported by the National Research Foundation of Ukraine (project No. 2023.05/0024).

1. S. Chatterjee, Hu. H. Paras, and M. A. Chakraborty, *Review of Nano and Microscale Heat Transfer: An Experimental and Molecular Dynamics Perspective*, *Processes* **11**, 2769 (2023).
2. Anto Zacharias, Rajesh Baby, Hanna J. Maria, and Sabu Thomas, *Preparation, characterization, and selection of nano-assisted phase change materials for thermal management and storage applications*, *Renew. Sustain. Energy Rev.* **210**, 115195 (2025).
3. V. Lacatena, M. Haras, J.-F. Robillard, S. Monfray, T. Skotnicki, and E. Dubois, *Phononic engineering of silicon using “dots on the fly” e-beam lithography and plasma etching*, *Microelectron. Eng.* **121**, 131 (2014).
4. Siqi Xie, Hongxin Zhu, Xing Zhang, and Haidong Wang, *A brief review on the recent development of phonon engineering and manipulation at nanoscales*, *Int. J. Extrem. Manuf.* **6**, 012007 (2024).
5. V. Kuryliuk, O. Tyvonovych, and S. Semchuk, *Impact of Ge clustering on the thermal conductivity of SiGe nanowires: atomistic simulation study*, *Phys. Chem. Chem. Phys.* **25**, 6263 (2023).



6. Mykola Isaiev, Yuliia Mankovska, Vasyl Kuryliuk, and David Lacroix, *Thermal transport properties of nanoporous silicon with significant specific surface area*, *Appl. Phys. Lett.* **122**, 172201 (2023).
7. D. Lacroix, M. I. Nkenfack, G. Pernot, and M. Isaiev, *Thermal properties of nanoporous materials, large scale modelling with the use of Monte Carlo phonon transport autocorrelation*, *J. Appl. Phys.* **134**, 025101 (2023).
8. Stephen E. Lewis, John R. DeBoer, James L. Gole, and Peter J. Hesketh, *Sensitive, selective, and analytical improvements to a porous silicon gas sensor*, *Sens. Actuators B: Chem.* **110**, 54 (2005).
9. Donghyeok Ma, Da Hoon Kim, Mi Yu, Ye Eun Cho, and Hansu Kim, *Porous silicon layer overlaid graphite anode materials for fast-charging of lithium-ion batteries*, *J. Power Sources* **645**, 237199 (2025).
10. T. Dzhabarov and A. Bayramov, *Porous Silicon and Solar Cells*, in: *Handbook of Porous Silicon*, L. Canham (eds.), Springer, Cham. (2017).
11. Gerhard Müller, Alois Friedberger, and Kathrin Knese, *Handbook of Silicon Based MEMS Materials and Technologies*, William Andrew Publishing (2010).
12. U. Bernini, S. Lettieri, P. Maddalena, R. Vitiello, and G. Di Francia, *Evaluation of the thermal conductivity of porous silicon layers by an optical pump-probe method*, *J. Phys.: Condens. Matter.* **13**, 1141 (2001).
13. Pritam Sharma, John Dell, Giacinta Parish, and Adrian Keating, *Engineering 1/f noise in porous silicon thin films for thermal sensing applications*, *Microporous Mesoporous Mater.* **324**, 111302 (2021).
14. Nguyen Van Toan, Yijie Li, Truong Thi Kim Tuoi, Nuur Syahidah Sabran, Jun Hieng Kiat, Ioana Voiculescu, and Takahito Ono, *Thermoelectric generator using nanoporous silicon formed by metal-assisted chemical etching method*, *Energy Convers. Manag.* **323**, 119268 (2025).
15. Jin Fang and Laurent Pilon, *Scaling laws for thermal conductivity of crystalline nanoporous silicon based on molecular dynamics simulations*, *J. Appl. Phys.* **110**, 064305 (2011).
16. Y. He and G. Galli, *Microscopic origin of the reduced thermal conductivity of nanoporous Si*, *Phys. Rev. Lett.* **108**, 215901 (2012).
17. Jin Fang and Laurent Pilon, *Tuning thermal conductivity of nanoporous crystalline silicon by surface passivation: A molecular dynamics study*, *Appl. Phys. Lett.* **101**, 011909 (2012).
18. Mykola Isaiev, Nataliia Kyrychenko, Vasyl Kuryliuk, and David Lacroix, *Features of phonon scattering by a spherical pore: Molecular dynamics insight*, *Appl. Phys. Lett.* **124**, 142202 (2024).
19. Ryan Jacobs, Dane Morgan, Siamak Attarian *et al.* *A practical guide to machine learning interatomic potentials – Status and future*, *Curr. Opin. Solid State Mater. Sci.* **35**, 101214 (2025).
20. Guanjie Wang, Changrui Wang, Xuanguang Zhang, Zefeng Li, Jian Zhou, and Zhimei Sun, *Machine learning interatomic potential: Bridge the gap between small-scale models and realistic device-scale simulations*, *Science* **27**, 109673 (2024).
21. Hoang-Giang Nguyen, Thanh-Dung Le, Hong-Giang Nguyen, and Te-Hua Fang, *Mechanical properties of AlCoCrCuFeNi high-entropy alloys using molecular dynamics and machine learning*, *Mater. Sci. Eng. R* **160**, 100833 (2024).
22. Omarelfarouq Elgack, Belal Almomani, Junaidi Syarif, Mohamed Elazab, Mohammad Irshaid, and Mohammad Al-Shabi, *Molecular dynamics simulation and machine learning-based analysis for predicting tensile properties of high-entropy FeNiCrCoCu alloys*, *J. Mater. Res. Technol.* **25**, 5575 (2023).
23. Y. Luo, M. Li, H. Yuan *et al.*, *Predicting lattice thermal conductivity via machine learning: A mini review*, *npj Comput. Mater.* **9**, 4 (2023).
24. Liben Guo, Yuanbin Liu, Lei Yang, and Bingyang Cao, *Lattice dynamics modeling of thermal transport in solids using machine-learned atomic cluster expansion potentials: A tutorial*, *J. Appl. Phys.* **137**, 081101 (2025).
25. Xianqing Liu, Fei Liang, Shule Liu, Gechuanqi Pan, Jing Ding, and Jianfeng Lu, *NaCl-KCl-CaCl<sub>2</sub> molten salts for high temperature heat storage: Experimental and deep learning molecular dynamics simulation study*, *Sol. Energy Mater. Sol. Cells* **280**, 113275 (2025).
26. Yao Du, Penghua Ying, and Jin Zhang, *Prediction and optimization of the thermal transport in hybrid carbon-boron nitride honeycombs using machine learning*, *Carbon* **184**, 492 (2021).
27. Prabudhya Roy Chowdhury and Xiulin Ruan, *Unexpected thermal conductivity enhancement in aperiodic superlattices discovered using active machine learning*, *npj Comput. Mater.* **8**, 12 (2022).
28. Jing Wan, Jin-Wu Jiang, and Harold S. Park, *Machine learning-based design of porous graphene with low thermal conductivity*, *Carbon* **157**, 262 (2020).
29. A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, *LAMMPS — a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales*, *Comp. Phys. Comm.* **271**, 10817 (2022).
30. S. Moser, C. Kenel, L. A. Wehner, R. Spolenak, and D. C. Dunand, *3D ink-printed, sintered porous silicon scaffolds for battery applications*, *J. Power Sour.* **507**, 230298 (2021).
31. Zaichun Liu, Xinhai Yuan, Shuaishuai Zhang, Jing Wang, Qinghong Huang, Nengfei Yu, Yusong Zhu, Lijun Fu, Faxing Wang, Yuhui Chen, and Yuping Wu, *Three-dimensional ordered porous electrode materials for electrochemical energy storage*, *NPG Asia Mater.* **11**, 12 (2019).
32. J. Ordóñez-Miranda and J. J. Alvarado-Gil, *Effect of the pore shape on the thermal conductivity of porous media*, *J. Mater. Sci.* **47**, 6733 (2012).

33. H. Babaei, A. J. H. McGaughey, and C. E. Wilmer, *Effect of pore size and shape on the thermal conductivity of metal-organic frameworks*, *Chem. Sci.* **8**, 583 (2017).
34. P. Ferrando-Villalba, L. D'Ortenzi, G. G. Dalkiranis *et al.* *Impact of pore anisotropy on the thermal conductivity of porous Si nanowires*, *Sci. Rep.* **8**, 12796 (2018).
35. K. Valalaki and A. G. Nassiopoulou, *Thermal conductivity of highly porous Si in the temperature range 4.2 to 20 K*, *Nanoscale Res. Lett.* **9**, 318 (2014).
36. Dimitrios Angelis, Filippas Sofos, and Theodoros E. Karakasidis, *Artificial intelligence in physical sciences: symbolic regression trends and perspectives*, *Arch. Comput. Methods Eng.* **30**, 3845 (2023).
37. F. Sofos, C. Dritselis, S. Misdanitis, T. Karakasidis, and D. Valougeorgis, *Computation of flow rates in rarefied gas flow through circular tubes via machine learning techniques*, *Microfluid. Nanofluidics* **27**, 85 (2023).
38. Miles Cranmer, *Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl*. [arXiv:2305.01582](https://arxiv.org/abs/2305.01582).
39. B.-J. Lee, *A modified embedded atom method interatomic potential for silicon*, *Calphad.* **31**, 95 (2007).
40. J. Tersoff, *Modeling solid-state chemistry: Interatomic potentials for multicomponent systems*, *Phys. Rev. B* **39**, 5566 (1989).
41. J. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov, and S. Yip, *Interatomic potential for silicon defects and disordered phases*, *Phys. Rev. B* **58**, 2539 (1998).
42. F. H. Stillinger and T. A. Weber, *Computer simulation of local order in condensed phases of silicon*, *Phys. Rev. B* **31**, 5262 (1985).
43. Y. Lee and G. S. Hwang, *Force-matching-based parameterization of the Stillinger–Weber potential for thermal conduction in silicon*, *Phys. Rev. B* **85**, 125204 (2012).
44. A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, *Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials*, *J. Comput. Phys.* **285**, 316 (2015).
45. A. V. Inyushkin, *Thermal conductivity of group IV elemental semiconductors*, *J. Appl. Phys.* **134**, 221102 (2023).
46. B. Graczykowski, A. El Sachat, J. S. Reparaz *et al.*, *Thermal conductivity and air-mediated losses in periodic porous silicon membranes at high temperatures*, *Nat. Commun.* **8**, 415 (2017).
47. D. Kojda, T. Hofmann, N. Gostkowska-Lekner, and K. Habicht, *Characterization and modeling of the temperature-dependent thermal conductivity in sintered porous silicon-aluminum nanomaterials*, *Nano Res.* **15**, 5663 (2022).
48. S. Lettieri, U. Bernini, E. Massera, and P. Maddalena, *Optical investigations on thermal conductivity in n- and p-type porous silicon*, *Phys. Status Solidi C* **2**, 3414 (2005).
49. G. Geseley, J. Linsmeieryx, V. Drachy, J. Frickey, and R. Arens-Fischer, *Temperature-dependent thermal conductivity of porous silicon*, *Phys. D: Appl. Phys.* **30**, 2911 (1997).
50. S. Erfantalab, G. Parish, and A. Keating, *Determination of thermal conductivity, thermal diffusivity and specific heat capacity of porous silicon thin films using the 3 $\omega$  method*, *Int. J. Heat Mass Transf.* **184**, 122346 (2022).
51. Z. Fang, M. Hu, W. Zhang, X. Zhang, and H. Yang, *Thermal conductivity and nanoindentation hardness of as-prepared and oxidized porous silicon layers*, *J. Mater. Sci.: Mater. Electron.* **19**, 1128 (2008).
52. R. Srinivasan, M. Jayachandran, and K. Ramachandran, *Photoacoustic studies on optical and thermal properties of p-type and n-type nanostructured porous silicon for (100) and (111) orientations*, *Cryst. Res. Technol.* **42**, 266 (2007).

**Теплопровідність нанопористого кремнію:  
моделювання методами молекулярної динаміки  
та прогнозування методами машинного навчання**

V. V. Kuryliuk, O. Ya. Olikh

Досліджено теплопровідність нанопористого кремнію (p-Si) із застосуванням методів рівноважної молекулярної динаміки та машинного навчання. Проаналізовано ефективність використання різних міжатомних потенціалів у молекулярно-динамічних розрахунках і обґрунтовано вибір потенціалу Tersoff завдяки його стабільності та точності. Значення теплопровідності p-Si обчислено в широкому діапазоні температур і пористостей. За допомогою алгоритму символічної регресії отримано аналітичний вираз, який описує залежність теплопровідності від температури та пористості. Показано, що моделі випадкового лісу та градієнтного бустингу, натреновані для реконструкції автокореляційних функцій теплового потоку та прогнозування теплопровідності, забезпечують вищу точність порівняно з моделями, побудованими на основі методу опорних векторів. Запропонований комбінований підхід забезпечує точне й ефективне прогнозування теплового транспорту в пористих наноструктурах і сприяє розробленню матеріалів для систем теплового менеджменту та енергетичних застосувань.

**Ключові слова:** нанопористий кремній, теплопровідність, молекулярна динаміка, символічна регресія, машинне навчання.