МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ

ТАРАСА ШЕВЧЕНКА

ЗВІТ

по договору підряду № 25ГФ013-02/10

за договором від «03» березня 2025 року № 150/0024

на виконання грантової підтримки НФДУ

за період з «31» травня 2025 року по «30» серпня 2025 року

**«Фізичний аналіз теплового транспорту в напівпровідникових структурах різної розмірності: нанонитки, нанокомпозити, мультишарові структури, розрахунок температурних залежностей коефіцієнта теплопровідності матеріалів, їх аналіз та оптимізація морфології для енергоефективних застосувань (розрахунок температурних залежностей коефіцієнта теплопровідності нанокомпозитів)»**

Науковий керівник \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ігор КОМАРОВ

Виконавець \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Олег ОЛІХ

*Зміст виконаної роботи*

Наноструктуровані матеріали активно досліджуються завдяки їхнім унікальним тепловим властивостям. У мікро- та наноелектроніці зростає потреба у матеріалах з низькою теплопровідністю, придатних для використання у CMOS-технологіях, а для автономних систем важливими є наноматеріали з високою ефективністю термоелектричного перетворення. Їхня теплова поведінка визначається характерними розмірами структури та процесами розсіянням фононів, яке у нанопористих матеріалах істотно зменшує теплопровідність. Особливе місце серед таких матеріалів займає нанопористий кремній. Регулювання його пористості дозволяє змінювати оптичні, електричні та теплові характеристики, що відкриває можливості для застосування у сенсорах, батареях, опто- та наноелектроніці, а також у MEMS-пристроях. Завдяки значно нижчій теплопровідності порівняно з масивним кремнієм, цей матеріал розглядають як перспективний для теплоізоляції, сенсорів і термоелектричних генераторів. Водночас детальне вивчення його теплової поведінки є необхідною умовою ефективного використання на практиці. Молекулярна динаміка (MD) є ключовим інструментом для дослідження теплопереносу в наноструктурованих матеріалах. Її широко застосовують для прогнозування теплопровідності пористого кремнію та подібних систем, зокрема для аналізу впливу розміру, форми й розподілу пор на розсіяння фононів. Останнім часом перспективним підходом стало поєднання MD з методами машинного навчання (ML). При цьому можна виділити два основних підходи: використання в розрахунках міжатомних потенціалів, які обчислюються з використанням ML, або ж навчання ML-моделей на обмеженій кількості MD-даних з подальшим поширенням прогнозів на ширший клас умов. Обидва підходи довели ефективність у вивченні теплових властивостей, зокрема пористих структур, і суттєво прискорюють розробку матеріалів для теплового менеджменту та енергетичних застосувань. Водночас досліджень із такою інтеграцією для пористого кремнію поки що не проводили.

Наші зусилля зосереджені на другому напрямі, зокрема на попередньому етапі був застосований алгоритм символьної регресії для отриманні аналітичного виразу, що описує теплопровідність поруватого кремнію *k* і застосовного в діапазоні температур *Т* = 250-1000 К для поруватостей *р* = 0-0,8:

 (1)

де *Tn* = *Τ* / 300 K. Зазначимо, що результати отримані з використанням дуже обмеженого набору значень теплопровідності, розрахованих з використанням MD.

На цьому етапі для вирішення подібної задачі було вирішено застосувати інший підхід. При MD обчисленнях *k* оцінювалося шляхом інтегрування функції автокореляції теплового потоку.

, (2)

де J – вектор теплового потоку. Значення *t*c (верхня межа інтеграла або час моделювання MD) повинно бути достатньо великим, щоб система досягла рівноваги. Фактично, *k* визначається як значення насичення кривої *k* ( *t*c ). Нами були використані класичні ML алгоритми, а саме Random Forest (RF), Gradient Boosting (GB) та Support Vector Regression (SVR) для побудови моделей, що прогнозують криву *k* ( *t*c ). Можливість застосування цих алгоритмів з’явилася завдяки суттєвому збільшення розміру навчального набору даних внаслідок використання різних значень *t*c.

Таким чином, набір вхідних дескрипторів включав пористість, температуру та час моделювання. Ознаки та цільова змінна були стандартизовані таким чинок, щоб дані кожного набору мали нульове середнє значення та одиничне стандартне відхилення. Моделі були реалізовані за допомогою інструментарію Scikit-learn на мові Python. Моделі були оптимізовані за допомогою фреймворку Optuna, в якому для ефективного вибору гіперпараметрів використовувалися семплер TPE та гіперсмуговий обрізувач. Під час налаштування моделей було застосовано схему 5-кратної перехресної валідації. Для визначення теплопровідності прогнозовані криві *k* ( *t*c ) оброблялися так само, як і криві, отримані в результаті MD-моделювання.

Метриками, що використовувалися для оцінки якості прогнозу, були середньоквадратична похибка (MSE), середня абсолютна процентна похибка (MAPE) та коефіцієнт детермінації (*R*2), як визначено рівняннями (3)-(5):

, (3)

, (4)

, (5)

де *N* – кількість значень у тренувальному наборі,  – істинне значення коефіцієнта тепропровідності,  – прогнозоване значення,  – середнє значенням істинних значень. Для налаштування моделі використовувався MSE, а для оцінки якості прогнозування навчених моделей застосовувалися всі три метрики.

Вихідні дані для кожного з 30 остаточних значень теплопровідності складалися з результатів MD-моделювання *k* при 10 000 значеннях tc. Цей набір даних був розділений на підмножини для навчання (80%) і тестування (20%) і використовувався для розробки ML-моделей. Результати навчання показано на рис. 1. Очевидно, що показники ефективності SVR нижчі за показники RF і GB, які демонструють подібний рівень точності.

|  |
| --- |
|  |
| Рис.1. Порівняння значень k при різних tc, розрахованих за допомогою MD-моделювання та передбачених методами Gradient Boosting (а), Random Forest (б) та Support Vector Regression (в) для навчального набору даних. Чорні суцільні лінії – лінії істинності. |

На рис.2а наведено приклад залежності *k* ( *t*c ), отриманої як в результаті MD моделювання, так і з передбачень навчених ML моделей для тих самих значень *T* і *p*, що використовувалися під час навчання. Очевидно, що RF і GB точно відтворюють поведінку *k* ( *t*c ), дозволяючи отримати значення теплопровідності з похибкою менше 1%. Ефективність SVR значно гірша. Загалом, для прогнозів теплопровідності при значеннях температури та пористості, для яких доступні розрахунки MD, MAPE становлять 1,05%, 0,98% та 26,8%, а значення R2 становлять 0,999, 0,999 та 0,977 для GB, RF та SVR відповідно.

|  |
| --- |
|  |
| Рис. 2. Залежності k від tc, розраховані за допомогою MD-моделювання (чорні кола) та отримані в результаті передбачень різних ML алгоритмів для значень пористості та температури, які або відповідають випадку MD-моделювання (a), або близькі до них (b). |

Однак основною метою розробки моделей є оцінка теплопровідності для значень *T* і *p*, що виходять за межі моделювання методом молекулярної динаміки. На рис.2b представлені типові криві *k* ( *t*c ), отримані за допомогою моделей, а також результати моделювання методом молекулярної динаміки для близьких значень температури і пористості. Криві, отримані за допомогою обох алгоритмів на основі дерев: 1) тісно збігаються між собою, 2) відповідають тенденціям, що узгоджуються з результатами MD, 3) дають значення теплопровідності, що відповідають загальним очікуванням, тобто підвищення температури та пористості призводить до зниження значень k. На відміну від цього, крива, отримана за допомогою SVR, не відтворює ці бажані характеристики.

ML Моделі були використані для оцінки значень теплопровідності для різних пар (*T*, *p*), не охоплених MD-моделюванням (див. рис. 3). Для порівняння, значення, отримані за допомогою рівняння (1), також показані на рисунку у поверхонь. Оскільки для цих комбінацій температури та пористості немає результатів MD, показники оцінки були обчислені відносно прогнозів моделі символьної регресії. Невеликі похибки навчального набору вказують на те, що SR забезпечує достатньо надійне наближення фізичної залежності. Тому порівняння прогнозів інших ML-моделей з рівнянням (1) дає змогу отримати значущу оцінку їхньої здатності узагальнювати навчальні дані. Це не повністю замінює тестування на MD або інших даних, але чітко демонструє поведінку та фізичну правдоподібність прогнозів моделі в широкому діапазоні параметрів, де прямі розрахунки недоступні.

|  |
| --- |
|  |
| Рис.3. Значення теплопровідності, визначені за кривими *k* ( *t*c ), передбаченими методами Gradient Boosting (a), Random Forest (b) та Support Vector Regression (c), у порівнянні із значеннями, обчисленими за формулою (1). Метрики обчислені у припущенні, що формула (1) надає справжні значення. |

Як можна побачити, результати моделей Random Forest і Gradient Boosting є співмірними, даючи значення MAPE 24% і 22% відповідно, а також медіанну абсолютну процентну похибку (APE) 19% для обох випадків. Найбільші відхилення спостерігаються при підвищених температурах і рівнях пористості. На відміну від цього, модель Support Vector Regression демонструє значно вищі похибки, з медіанною відносною похибкою 50%, що перевищує навіть середнє значення 48%. Ці результати вказують на те, що SVR є найменш підходящим підходом для оцінки теплопровідності шляхом реконструкції залежностей *k* ( *t*c ).