МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ

ТАРАСА ШЕВЧЕНКА

ЗВІТ

по договору підряду № 25 ГФ051-02

за договором від «03» березня 2025 року № 93/0252

на виконання грантової підтримки НФДУ

за період з «03» березня 2025 року по «30» травня 2025 року

**«Аналітичні вирази для оцінки теплотранспортних властивостей мультишарових структур, отримані з використанням символьної регресії. Налаштовані моделі ансамблевого навчання на базі дерев рішень для оцінки теплотранспортних властивостей мультишарових кремнієвих структур (отримання аналітичних виразів для показника заломлення поруватого кремнію та взамозв’язку параметрів фотоакустичного сигналу з теплопровідністю мультишарових кремнієвих структур).»**

Науковий керівник \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Павло ЛІЩУК

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

*Зміст виконаної роботи*

Робота, виконана під час даного етапу може біти розділена на дві частини. Перша пов’язана з отриманням аналітичних виразів для опису показника заломлення поруватого кремнію (*p*-Si) для різних діапазонів довжин хвиль, друга – знаходження функціонального взаємозв’язку між параметрами фотоакустичного сигналу та теплопровідністю двошарової структури.

Основною задачею проєкту загалом є розробка принципів створення та характеризації поруватих кремнієвих наноструктур з оптимальними теплотранспортними властивостями. Одним зі шляхів керування такими властивостями є створення мультишарових структур типу Брегівских відбивачів, які складаються з прошарків *p*-Si з різним ступенем поруватості. Проте для оптимізації подібних структур необхідна інформація щодо показника заломлення поруватого кремнію *npSi*. Загалом, оцінка цієї величина може бути проведена відповідно до ізотропної моделі Бруггемана [D. A. G. Bruggeman; Berechnung verschiedener physikalisher Konstanten von heterogen Substanzen // *Ann. Phys.(Leipzig*). 1935, 24. рр. 634-664.]. Застосовність цієї моделі обумовлюється морфологією структур (співвідношенням розмірів пор та довжин хвиль видимого та інфрачервоного діапазонів) і була детально проаналізована у попередніх звітах. У випадку, коли дійсна частина коефіцієнта заломлення суттєво більша за уявну, згідно з цією моделлю:

, (1)

де *р* – ступінь поруватості, *nSi* – показник заломлення монокристалічного кремнію. В (1), вважається, що пори заповнені повітрям з показником заломлення рівним одиниці.

Загалом, рівняння (1) може бути розв’язано аналітично:

, (2)

де

. (3)

Проте значення *nSi* не є відомою функцією довжини хвилі, визначаються з експерименту і табульовані лише при певних значеннях довжин хвиль; більше того, для широкого спектрального діапазону необхідно використовувати декілька джерел. Як наслідок, розрахунок *npSi* з використанням (2)-(3) не є зручним і такий підхід не може бути використаний при аналітичній оптимізації теплотранспортних властивостей мультишарових структур.

Водночас, алгоритм символьної регресіїї (Symbolic Regression, SR) на відміну від багатьох інших алгоритмів машинного навчання, дозволяє отримати аналітичні вирази, спираючись на набір даних. Тому в нашій роботі цей алгоритм був застосований для отримання залежностей *npSi* = *f* (λ, *p*), де λ – довжина хвилі. Отримані вирази відповідали температурі 293 К та мали дозволяти максимально точно оцінювати показник заломлення паруватого кремнію для діапазону *р* ∈ [0..0,8]. Щодо діапазону довжин хвиль, то так як *nSi* змінюється в достатньо широких межах, то окремо були отримані більш загальний вираз для діапазону [530 нм … 12000 нм], та більш точний, але обмежений у використанні, для діапазону [600 нм … 6000 нм]. При створенні тренувального та тестового наборів даних величина показника заломлення розраховувалася з використанням виразів (2) та (3), причому значення *nSi* бралися з роботи [M.A. Green; Improved silicon optical parameters at 25C, 295 K and 300 K including temperature coefficients // *Prog. Photovolt. Res. Appl*., 2022, 30, pp. 164–179] при λ ≤ 1450 нм та з роботи [H. H. Li; Refractive index of silicon and germanium and its wavelength and temperature derivatives //J. Phys. Chem. Ref. Data , 1980; 9 (3), pp. 561–658] при λ > 1450 нм. Для кожного визначених діапазонів довжин хвиль окремо створювалися набори даних. Тренувальний містив 800 величин *npSi* , розрахованих для пар (λ, *p*), значення яких випадковим чином вибиралися зі вказаних діапазонів. Перевірка точності отриманих виразів здійснювалася та тестових наборах, які відповідали всім комбінаціям значень коефіцієнта поруватості з діапазону [0..0,8] з кроком 0,02 та значень довжини хвилі з кроком 30 нм (для діапазону [530 нм … 12000 нм]) чи 20 нм (для діапазону [600 нм … 6000 нм]).

Метриками, які використовувалися для оцінки якості прогнозів були

- середня відносна похибка МАРЕ:

, (4)

де N – кількість значень у тренувальному наборі,  – істинне значення показника заломлення, розраховане з використанням виразів (2) та (3),  – прогнозоване значення відповідно до запронованого виразу;

- середня квадратична похибка МSE:

, (5)

- середня абсолютна похибка МАE:

. (6)

Окрім метрик, які відображали середні значення похибки, для оцінки найбільших відхилень розглядалася також максимальна відносна похибка

, (7)

та медіанна відносна похибка для *АРЕ*MED, яка показувала менше якої величини становить похибка для 50% прогнозів.

Символьна регресія була реалізована на мові Pyton з використанням пакету PySR. Під час розрахунків параметр populations (кількість популяцій) дорівнював 36 (рекомендоване потроєне значення кількості ядер), population\_size (кількість особин у кожній популяції) – 1000, ncycles\_per\_iteration (кількість ітерацій на цикл) – 500, максимальна складність виразу (maxsize) – 25, мультиплікативний коефіцієнт для визначення покарання за складність (parsimony) - 10-5. Попередня обробка даних не застосовувалася.

Під час тренування мінімізувалася функція втрат, яка була середнім значенням зваженої квадратичної похибки:

 (8)

де  – ваговий коефіцієнт. Вигляд вагового коефіцієнта був одним з гіперпараметрів, які підбиралися для оптимізації роботи алгоритму; зокрема розглядалися варіанти  Як показали проведені дослідження, для діапазону [600 нм … 6000 нм] раціональним вибором є  (стандартне статистичне зважування), а для [530 нм … 12000 нм] – .

Нижче наведено два найкращі варіанти залежностей *npSi* = *f* (λ, *p*), отримані для діапазону довжин хвиль [600 нм … 6000 нм]:

 (9)

 (10)

де значення довжини хвилі очікується в нанометрах.

На рис.1 приведено результати тренування та тестування з використанням виразів (9) та (10). Як видно з наведених даних, для більшості значень поруватості та довжини хвилі визначені аналітичні вирази дають похибку, що є меншою 0.5%. У табл. 1 зведено величини метрик, отримані в даному випадку.

|  |  |
| --- | --- |
|  |  |
|  |  |
| Рис.1. Передбачення величин показника заломлення, отримані з використанням формул (9) (а) та (10) (в) на тренувальному наборі та розподіли величини абсолютної відносної похибки, отримані на тестовому наборі (б – формула (9), г – формула (10)). Діапазон довжин хвиль [600 нм … 6000 нм]. | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Табл.1. Метрики точності передбачень показника заломлення поруватого кремнію, відповідно до аналітичних виразів, отриманих з використанням символьної регресії. | | | | | | |
|  | тренувальний набір | тестовий набір | | | | |
| Формула | MWSE | *MAPE*, % | *MSE* | *MAE* | *АРЕ*max, % | *АРЕ*МED, % |
| [600 нм … 6000 нм] | | | | | | |
| (9) |  | 0,10 |  |  | 1,10 | 0,09 |
| (10) |  | 0,16 |  |  | 1,56 | 0,14 |
| [530 нм … 12000 нм] | | | | | | |
| (11) |  | 0.12 |  |  | 1,13 | 0.12 |
| (12) |  | 0.07 |  |  | 1,38 | 0.05 |

 (11)

 (12)

|  |  |
| --- | --- |
|  |  |
|  |  |
| Рис.1. Передбачення величин показника заломлення, отримані з використанням формул (11) (а) та (12) (в) на тренувальному наборі та розподіли величини абсолютної відносної похибки, отримані на тестовому наборі (б – формула (11), г – формула (12)). Діапазон довжин хвиль [530 нм … 12000 нм]. | |

Нижче наведені типові отримані виразі, які характеризуються найкращими величинами метрик:

На даному етапі виконання проєкту проводився пошук аналітичних виразів для опису температурних залежностей коефіцієнту теплопровідності p-Si з різним ступенем поруватості. При цьому використовувався алгоритм символьної регресіїї (Symbolic Regression, SR), який під час пошуку функціональної залежності між вхідними та вихідними параметрами та числових коефіцієнтів використовує еволюційні алгоритми. На відміну від багатьох інших алгоритмів машинного навчання, SR дозволяє отримати результати (аналітичні формули), які можуть бути легко інтерпретовані. Крім того, цей підхід не вимагає значних об’ємів тренувальних даних для досягнення високоточних прогнозів.

Основою для побудови розрахункових аналітичних виразів були значення ТП, отримані в результаті застосування MD. Відповідні розрахунки проводилися за допомогою пакету LAMMPS, коефіцієнт теплопровідності (ТС) визначали шляхом обчислення середнього ансамблевого значення автокореляційної функції теплового струму в рамках формалізму Гріна-Кубо. Міжатомна взаємодія описувалася за допомогою потенціалу Tersoff. Всього були використані значення ТС отримані для кремнію з поруватістю *p* 0, 0,2 і 0,4% та семи значень температури, рівномірно розподілених в діапазоні температур 400-1000 К, а також ТС для 300 К, розраховані для р від 0 до 0,7 з кроком 0,1. Всього було використано 29 значень. Завдання полягала в отриманні виразу, застосовного в діапазоні температур *Т* = 250-100 К для поруватостями *р* = 0-0,8.

Символьна регресія була реалізована за допомогою Pyton з використанням пакету PySR. Під час розрахунків параметр populations (кількість популяцій) дорівнював 36 (рекомендоване потроєне значення кількості ядер), population\_size (кількість особин у кожній популяції) – 1000, ncycles\_per\_iteration (кількість ітерацій на цикл) – 500. Максимальна складність виразу (maxsize) та мультиплікативний коефіцієнт для визначення покарання за складність (parsimony) варіювалися у різних запусках в діапазонах 15-25 та 10-6-10-3, відповідно. Попередня обробка даних полягала лише у нормуванні значення температури на 300 К.

Під час тренування мінімізувалася функція втрат, яка була середнім значенням зваженої квадратичної похибки:

 (1)

де N – кількість значень у тренувальному наборі,  – істинне значення коефіцієнта тепропровідності,  – прогнозоване значення,  – ваговий коефіцієнт. В нашому випадку використовувалося статистичне зважування, тобто .

Метриками, які використовувалися для оцінки якості прогнозів були

- середня відносна похибка МАРЕ:

, (2)

- середня квадратична похибка МSE:

, (3)

- середня абсолютна похибка МАE:

. (4)

Нижче наведені типові отримані виразі, які характеризуються найкращими величинами метрик:

 (5)

 (6)

 (7)

 (8)

де *Tn* = *Τ* / 300 K. Налаштування, при яких отримані вирази (5)-(8) вказані в Табл.1.

|  |  |  |  |
| --- | --- | --- | --- |
| Табл.1. Гіперпараметри, використані при отриманні виразів (5)-(8) | | | |
| Формула | maxsize | parsimony | час тренувань, год |
| (5) | 20 | 10-4 | 30 |
| (6) | 25 | 10-4 | 20 |
| (7) | 25 | 10-6 | 50 |
| (8) | 25 | 10-6 | 30 |

На рис.1 представлено порівняння прогнозованих та істинних (розрахованих шляхом застосування методу молекулярної динаміки) значень, а в Табл.2 – значення метрик.

|  |  |
| --- | --- |
|  |  |
|  |  |
| Рис.1. Діаграми розсіювання, що порівнюють значення коефіцієнта теплопровідності з тренувального набору та відповідні величини, отримані з використанням формул (5) (а), (6) (б), (7) (в) та (8) (г). Пунктирні прямі – лінії ідентичності, наведені для зручності. | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Табл.2. Метрики, отримані на тренувальному наборі, при використанні виразів (5)-(8) | | | | |
| Формула | MWSE | MAPE | MSE | MAE |
| (5) | 0,1000 | 3,862 | 1,471 | 0,623 |
| (6) | 0,0356 | 3,552 | 0,0890 | 0,221 |
| (7) | 0,0711 | 2,023 | 1,430 | 0,564 |
| (8) | 0,0261 | 2,889 | 0,142 | 0,241 |

На Рис.2 наведено значення коефіцієнту теплопровідності у всьому діапазоні змінних (температури та поруватості), обчисленні відповідно до виразів, отриманих з використанням символьної регресії. Враховуючи значення метрик, а також поведінку отриманих виразів при високих значеннях поруватості та температури, вважаємо, що вираз (8) найбільш придатних для розрахунків температурних залежностей коефіцієнта теплопровідності поруватого кремнію

|  |  |
| --- | --- |
|  |  |
|  |  |
| Рис.2. Залежності коефіцієнта ТП пористого кремнію від температури та пористості. Кульки представляють результати MD розрахунків, поверхні відповідають рівнянням (5) (а), (6) (б), (7) (в) та (8) (г). | |