\documentclass[a4paper,12pt]{article}

\usepackage[utf8]{inputenc}

\usepackage[ukrainian]{babel}

\usepackage{amsmath}

\usepackage{geometry}

\usepackage{graphicx}

\geometry{top=2cm, bottom=2cm, left=3cm, right=1.5cm}

\usepackage{hyperref} % Пакет для роботи з посиланнями

\begin{document}

\begin{center}

\textbf{Київський національний університет імені Тараса Шевченка} \\

\textbf{Фізичний факультет} \\

\textbf{Кафедра загальної фізики}

\end{center}

\vspace{2cm}

\begin{center}

\textbf{\Large Звіт} \\

\textbf{\large про науково-виробничу практику із фізики наноматеріалів}

\end{center}

\vspace{1cm}

\noindent

\textbf{Тема роботи}: «Параметризацiя рухливостi носiїв заряду у кремнiї з використанням машиного навчання»

\vspace{2cm}

\noindent

\begin{flushright}

\textbf{Студент 2 курсу магiстратури} \\

освiтньої програми «Фiзика наносистем» \\

кафедри загальної фiзики \\

Кущ Іван

\vspace{1cm}

\textbf{Керiвник практики:} \\

завiдувач кафедри загальної фiзики, д. ф.-м. н. \\

Олiх Олег

\vspace{1cm}

\textbf{Науковий керiвник:} \\

завiдувач кафедри загальної фiзики, д. ф.-м. н. \\

Олiх Олег

\end{flushright}

\vspace{2cm}

\begin{center}

\textbf{Київ – 2024}

\end{center}

\section{План}

науково-виробничої практики з фізики наноматеріалів Куща.І.О.

(ОКР “магістр” , ІІ курс)

\begin{table}[h!]

\centering

\begin{tabular}{|c|p{6cm}|c|c|}

\hline

\textbf{\#} & \textbf{Вид роботи} & \textbf{Кабінет} & \textbf{Час виконання} \\ \hline

1 & Знайомство з методами розрахунку рухливості& 432 Кабінет & 25.01.2024–17.02.2024 \\ \hline

2 & Розрахунок рухливості носіїїв заряду за допомогою методу Klassen & 432 Кабінет & 17.02.2024–14.03.2024 \\ \hline

3 & Параметризація рузливості носіїїв заряду за допомогою Випадкового лісу & 432 Кабінет & 14.03.2024–01.04.2024 \\ \hline

4 & Параметризація рузливості носіїїв заряду за допомогою Символічної регресії & 432 Кабінет & 01.04.2024–20.04.2024 \\ \hline

5 & Написання звіту & 432 Кабінет & 20.04.2024–20.05.2024 \\ \hline

\end{tabular}

\caption{Розклад виконання робіт}

\label{tab:rozklad}

\end{table}

\section{Методики розрахунку рухливості}

\subsection{Наближення Klassen}

Метод наближення Klassen є одним із підходів для оцінки рухливості носіїв заряду в напівпровідниках, враховуючи різні механізми розсіювання. Він передбачає паралельне складання впливу розсіювання на ґратці та розсіювання на домішках. Рухливість \(\mu\) визначається через зворотну величину, що є сумою зворотних рухливостей, обумовлених кожним із цих механізмів. Це дозволяє точніше моделювати поведінку електронів та дірок при різних умовах температури і концентрації домішок. Метод використовує численні коефіцієнти, які визначають температурну залежність та взаємодію з концентрацією носіїв заряду.

У напівпровіднику визначається через паралельне складання впливу різних механізмів розсіювання. Це виражається через зворотну рухливість:

\[

\mu^{-1} = \mu\_L^{-1} + \mu\_{DA}^{-1}

\]

де:

\(\mu\_L\) — рухливість, обмежена розсіюванням на ґратці (фононах),

\(\mu\_{DA}\) — рухливість, обмежена розсіюванням на домішках (іонізовані домішки).

Рухливість, обмежена розсіюванням на ґратці:

\[

\mu\_L = \mu\_{\text{max}} \cdot \left(\frac{300}{T}\right)^{2.25}

\]

Рухливість, обмежена розсіюванням на домішках:

\[

\mu\_{DA} = \frac{\mu\_{\text{max}}^2}{\mu\_{\text{max}} - \mu\_{\text{min}}} \cdot \frac{N\_{\text{sc}}}{N\_{\text{eff}}} \cdot \left(\frac{N\_{\text{eff}}}{s\_c}\right)^{\alpha} \cdot \left(\frac{T}{300}\right)^{3\alpha - 1.5}

+ \frac{\mu\_{\text{max}} \cdot \mu\_{\text{min}}}{\mu\_{\text{max}} - \mu\_{\text{min}}} \cdot \frac{n + p}{N\_{\text{eff}}} \cdot \left(\frac{300}{T}\right)^{0.5}

\]

Для електронів (\( n \)):

\[

\mu\_{\text{max}} = 0.1414, \quad \mu\_{\text{min}} = 0.00685, \quad N\_{\text{eff}} = 9.2 \times 10^{22}

\]

Для дірок (\( p \)):

\[

\mu\_{\text{max}} = 0.04705, \quad \mu\_{\text{min}} = 0.000449, \quad N\_{\text{eff}} = 2.28 \times 10^{23}

\]

Для електронів (\( n \)):

\[

N\_{\text{sc}} = N\_d + N\_a + p

\]

Для дірок (\( p \)):

\[

N\_{\text{sc}} = N\_d + N\_a + n

\]

Для електронів (\( n \)):

\[

N\_{\text{eff}} = N\_d + N\_a \cdot G\_n(P\_n, T) + \frac{P}{F\_n(P\_n, T)}

\]

Для дірок (\( p \)):

\[

N\_{\text{eff}} = N\_a + N\_d \cdot G\_p(P\_p, T) + \frac{n}{F\_p(P\_p, T)}

\]

Концентрація донорів (\( N\_d \)) та акцепторів (\( N\_a \)) — це параметри, що визначають кількість іонізованих атомів у напівпровіднику, які можуть передавати або отримувати електрони:

\begin{itemize}

\item \( N\_d \) — концентрація донорів, тобто атомів або молекул, які здатні віддавати електрони в зону провідності, створюючи електронні носії заряду.

\item \( N\_a \) — концентрація акцепторів, тобто атомів або молекул, які здатні приймати електрони з валентної зони, створюючи дірки (недостаючі електрони) в кристалі.

\end{itemize}

Ці концентрації суттєво впливають на властивості напівпровідника, оскільки вони визначають кількість носіїв заряду в системі, а також її електричні та оптичні характеристики.

\[

G\_n = 1 - \left( \frac{S\_1}{S\_2 + \left( \frac{m\_0}{m\_e} \cdot \frac{T}{300} \right)^{S\_4} \cdot P\_n} \right)^{S\_3} + \frac{S\_5}{\left( \frac{m\_e}{m\_0} \cdot \frac{300}{T} \right)^{S\_7}} \cdot P\_p^{S\_6}

\]

\begin{table}[h!]

\centering

\begin{tabular}{|c|c|}

\hline

\textbf{Коефіцієнт} & \textbf{Значення} \\

\hline

\( S\_1 \) & 0.89233 \\

\( S\_2 \) & 0.41372 \\

\( S\_3 \) & 0.19778 \\

\( S\_4 \) & 0.28227 \\

\( S\_5 \) & 0.005978 \\

\( S\_6 \) & 1.80618 \\

\( S\_7 \) & 0.72169 \\

\hline

\end{tabular}

\caption{Значення коефіцієнтів для коефіцієнта іонізації}

\end{table}

\[

F\_n = \frac{r\_1 \cdot P\_n^{r\_6} + r\_2 + r\_3 \cdot \frac{m\_e}{m\_0}}{P\_n^{r\_6} + r\_4 + r\_5 \cdot \frac{m\_e}{m\_0}}

\]

\begin{table}[h!]

\centering

\begin{tabular}{|c|c|}

\hline

\textbf{Коефіцієнт} & \textbf{Значення} \\

\hline

\( r\_1 \) & 0.7643 \\

\( r\_2 \) & 2.2999 \\

\( r\_3 \) & 6.5502 \\

\( r\_4 \) & 2.3670 \\

\( r\_5 \) & -0.01552 \\

\( r\_6 \) & 0.6478 \\

\hline

\end{tabular}

\caption{Значення коефіцієнтів для функції \( F\_n \)}

\end{table}

\[

P = \frac{1}{P\_b + P\_c}

\]

Для електронів (\( n \)):

\[

P\_b = \frac{3.83}{\left( \frac{1.36 \cdot 10^{26}}{(N\_{\text{dop}} + P) \cdot \left( \frac{T}{300} \right)^2} \right) \cdot \frac{m\_e}{m\_0}}

\]

\[

P\_c = \frac{2.46}{\left( 3.97 \cdot 10^{19} \cdot \left( \frac{T}{300Z} \right)^3 \cdot \frac{1}{N\_{\text{dop}}} \right)^{2/3}}

\]

Для дірок (\( p \)):

\[

P\_b = \frac{3.83}{\left( \frac{1.36 \cdot 10^{26}}{(N\_{\text{dop}} + n) \cdot \left( \frac{T}{300} \right)^2} \right) \cdot \frac{m\_e}{m\_0}}

\]

Для електронів (\( Z\_n \)):

\[

Z\_n = 1 + \frac{1}{0.21 + \left( \frac{4 \cdot 10^{26}}{N\_{\text{dop}}} \right)^2}

\]

Для дірок (\( Z\_p \)):

\[

Z\_p = 1 + \frac{1}{0.5 + \left( \frac{7.2 \cdot 10^{26}}{N\_{\text{dop}}} \right)^2}

\]

\cite{bib:vdm\_scan}

\cite{bib:PDG\_18}

\subsection{Наближення Arorra}

Метод Arorra є альтернативним підходом для оцінки рухливості носіїв заряду в напівпровідниках, який також враховує вплив різних механізмів розсіювання, таких як розсіювання на ґратці та на домішках. Однак, на відміну від методу Klassen, метод Arorra більше фокусується на температурних залежностях, пропонуючи інші математичні вирази для рухливості. В Arorra, основна увага приділяється розрахунку ефективної рухливості, враховуючи фактори, що змінюються в залежності від температури та концентрації носіїв заряду, зокрема акцепторів і донорів.

Однією з головних відмінностей між методами є те, що метод Klassen передбачає більш комплексну модель, яка поєднує різні механізми розсіювання в одну формулу з використанням зворотних рухливостей. У той час як метод Arorra може використовувати більш окремі підходи для кожного механізму розсіювання, зокрема, для температурної залежності рухливості електронів та дірок. Крім того, метод Arorra часто включає інші коефіцієнти, які специфічні для конкретних умов і матеріалів, що може давати точніші результати в певних випадках.

\[

\mu = \mu\_{\text{min}} + \frac{\mu\_{\text{max}}}{1 + \left( \frac{N}{N\_0} \right)^{\alpha}}

\]

Ця формула виражає рухливість носіїв заряду в залежності від концентрації \( N \). Коефіцієнт \( \mu\_{\text{min}} \) є мінімальною рухливістю, тоді як \( \mu\_{\text{max}} \) — максимальною рухливістю. Параметр \( N\_0 \) визначає критичну концентрацію, після досягнення якої відбувається насичення рухливості.

\[

N = N\_{\text{a}^{-}} + N\_{\text{d}^{+}}

\]

\( N \) виражає загальну концентрацію носіїв заряду , як суму концентрацій акцепторів \( N\_{\text{a}^{-}} \) та донорів \( N\_{\text{d}^{+}} \).

\[

\mu\_{\text{min}} = \mu\_{\text{min}}^{\text{eff}} \cdot \left( \frac{T}{T\_{\text{eff}}} \right)^{\beta\_1}

\]

\[

\mu\_0 = \mu\_0^{\text{eff}} \cdot \left( \frac{T}{T\_{\text{eff}}} \right)^{\beta\_2}

\]

\[

N\_0 = N\_0^{\text{eff}} \cdot \left( \frac{T}{T\_{\text{eff}}} \right)^{\beta\_3}

\]

\[

\alpha = \alpha\_0^{\text{eff}} \cdot \left( \frac{T}{T\_{\text{eff}}} \right)^{\beta\_4}

\]

\textbf{Значення параметрів:}

\[

T\_{\text{eff}} = 300 \, \text{K}, \quad \alpha\_0 = 0.88, \quad \beta\_1 = -0.57

\]

Для \( n \) (електронів):

\[

\beta\_2 = -2.33

\]

Для \( p \) (дірки):

\[

\beta\_2 = -2.23

\]

\[

\beta\_3 = 2.4, \quad \beta\_4 = -0.146

\]

\begin{table}[h!]

\centering

\begin{tabular}{|c|c|c|}

\hline

\textbf{Параметр} & \textbf{n} & \textbf{p} \\ \hline

$\mu\_{\text{min (eff)}}$ & 88 & 54,3 \\ \hline

$\mu\_{0 \, \text{(eff)}}$ & 1252 & 407 \\ \hline

$N\_{0 \, \text{(eff)}}$ & $1,26 \times 10^{17}$ & $2,35 \times 10^{17}$ \\ \hline

\end{tabular}

\caption{Параметри для основних носіїв}

\label{tab:parametry}

\end{table}

Для розрахунку рухливості не основних зарядів змінюються параметри

\begin{table}[h!]

\centering

\begin{tabular}{|c|c|c|}

\hline

\textbf{Параметр} & \textbf{$\mu\_e \, \text{(Teff = 292K)}$} & \textbf{$\mu\_n \, \text{(Teff = 295K)}$} \\ \hline

$\mu\_{\text{min}}$ & 130 & 232 \\ \hline

$\mu\_0$ & 370 & 1180 \\ \hline

$N\_0$ & $8 \times 10^{17}$ & $8 \times 10^{16}$ \\ \hline

$\alpha$ & 1,25 & 0,9 \\ \hline

\end{tabular}

\caption{Значення параметрів неосновних носіїв заряду для $\mu\_e$ та $\mu\_n$}

\label{tab:parametry\_non\_main\_carriers}

\end{table}

\subsection{Наближення Fletcher}

Ще один із відомих методів пошуку рухливості основних носіїв заряду є методі Fleatcher . Основна перевага в цьому методі полягає в доволі простому обчисленні .

\[

\frac{1}{\mu\_{n,\text{fl}}} = \frac{1}{\mu\_{\text{in},n}} + \frac{1}{\mu\_{\text{cc}}}

\]

\[

\frac{1}{\mu\_{p,\text{fl}}} = \frac{1}{\mu\_{\text{in},p}} + \frac{1}{\mu\_{\text{cc}}}

\]

\[

\mu\_{\text{cc}} = \frac{\left(\frac{T}{T\_{\text{eff}}}\right)^{3/2} F\_1}{\sqrt{n p} \cdot \ln\left(1 + \left(\frac{T}{T\_{\text{eff}}}\right)^2 \cdot (n p)^{-1/3} F\_2\right)}

\]

\text{Де:}

\begin{itemize}

\item $T$ — температура решітки,

\item $\mu\_{\text{in},n}$ і $\mu\_{\text{in},p}$ — вхідна рухливість електронів і дірок,

\item $n$ — концентрація електронів,

\item $p$ — концентрація дірок,

\item $F\_1$ (одиниця SI: с$^2$A/(м$^3$кг)) і $F\_2$ (одиниця SI: 1/м$^2$) — властивості матеріалу,

\item $T\_{\text{eff}}$ — характеристична температура матеріалу.

\end{itemize}

\section{Параметризація рухливості за допомогою Random Forest}

Метод Random Forest — це ансамблевий алгоритм машинного навчання, який будує кілька рішень дерев на випадкових підмножинах даних і комбінує їх для покращення точності та стабільності моделі. Кожне дерево в лісі приймає рішення незалежно, і кінцевий результат визначається голосуванням або середнім значенням передбачень усіх дерев. Цей метод допомагає знижувати ризик перенавчання, підвищує точність прогнозів і може працювати з великою кількістю вхідних змінних.

В коді використовувався метод пошуку найкращих гіперпараметрів Optuna

Optuna — це потужна бібліотека для автоматичного налаштування гіперпараметрів, яка використовує інтелектуальні методи пошуку для оптимізації моделей машинного навчання.

ptuna використовує методи послідовного вибору гіперпараметрів, базуючись на попередніх результатах. Це дозволяє скоротити кількість ітерацій у порівнянні з випадковим пошуком.

Код та результати обчислень знаходяться за

\href{https://github.com/kushchIvan/RFR-PYSR}{посиланням}.

\begin{figure}

\centering

\includegraphics[width=0.5\linewidth]{fig1.png}

\caption{лінія ідеальної відповідності Випадкового лісу }

\label{fig:enter-label}

\end{figure}

Модель передбачення демонструє загалом хороший рівень точності. На основі графіка, що порівнює передбачені значення з реальними, видно, що більшість точок розташовані близько до лінії ідеального співпадіння, що свідчить про здатність моделі ефективно описувати основні закономірності даних.

Однак, на графіку також помітні окремі значні відхилення (виброси).

\section{Параметризація рухливості за допомогою Symbolic Regression }

PySR (Python Symbolic Regression) є потужним інструментом для виконання символічної регресії, яка полягає у пошуку математичних виразів, що найкраще описують вхідні дані. PySR використовує еволюційні алгоритми для генерування та оптимізації формул, що дозволяє знаходити компактні та інтерпретовані математичні моделі[3].

Основний компонент пакету - PySRRegressor, який надає інтерфейс для тренування та оцінки моделей символічної регресії. PySRRegressor автоматично шукає найкращу формулу шляхом ітеративного еволюційного процесу, що включає генерацію можливих виразів, їх оцінку та вибір найкращих серед них на основі заданих критеріїв, таких як точність та складність моделі.

Протягом семестру я працював над кількома наборами даних, застосовуючи PySRRegressor для розробки точних та інтерпретованих моделей. Цей процес включав попередню обробку даних, налаштування параметрів PySRRegressor, а також аналіз отриманих результатів для забезпечення найкращої можливої апроксимації даних. Завдяки цьому я здобув глибокі знання в області символічної регресії та її практичного застосування для аналізу даних.

Навчання проходило в пару етапів . Перший етап поялгає в предобробці данних . Розбиття на тренувальний набір, для основного тренування моделі , та тестування , на якому відбувається оцінка вже навченної моделі.

Потім відбувається створеннч самоїї моделі та саме навчання за ним .

Весь код та результати з результатами знаходяться за

\href{https://github.com/kushchIvan/RFR-PYSR}{посиланням}.

Окремої уваги хочеться звенути на створенні оператори та залучення їх ,

а саме "Zved" та "PowInv" . Пакет Pysr дозволяє створбювати свої оператори для начання моделі , для більш швидкого та точного підбору апроксимуючих формул. Ці формули ,"Zved" та "PowInv", зумовленні тереотичними данними , що були представленні в звіті.

І представляють собою :

\begin{equation}

\text{Zved}(x, y) = \frac{x \cdot y}{x + y}

\end{equation}

\begin{equation}

\text{PowInv}(x) = \frac{1.0}{1.0 + x}

\end{equation}

Було виконано декілька спроб начання з різними параметрами та часом навчання . В ході роботи було виявленно , що для нашого набору данних для основний приріст точності відбувається 30 годинним завчанням , після проходження порогу в 30 годин , приріст точності дуже поступається прирісту складності та інтерпретації запропонованих формул.Основним показником було вирішино брати MAE(mean absolut error), через великий діапазон самих значень , саме цей критерій допомагав найбільш точно та наглядно інтерпретувати результати моделі. Найкращий результат якого змогла досгти модель полягає в MAE = 56,2749 . Що є доволі хорошим та значущим результатом .

Сама запропопована формує має такий вигляд :

\[

e^{\left(\frac{44.2}{\log{\left(T - 1.82 \right)}}\right)^{\sin{\left(\log{\left(\left(\frac{T - 63.1}{0.328 Nd + T^{6.15}}\right)^{-0.0606} \right)} \right)}}}

\]

З результату видно , що модель користується запропонованими нашими функціями , що дає потенційно гарні результати .

\begin{figure}[h!]

\centering

\includegraphics[width=\textwidth]{fig2.png}

\caption{графік розсіювання передбачуваних значень , отриманих на моделі символічної регресії }

\label{fig:enter-label}

\end{figure}

\section{Висновок}

Протягом цієї практики я вивчав теоретичні основи методів розрахунку рухливості основних та неосновних носіїв заряду в кремнії, зокрема методи Классена, Аврори та Флетчера. Це забезпечило розуміння фундаментальних фізичних принципів та математичних підходів, які лежать в основі цих методів.

Окрім теоретичного опрацювання, значна увага була приділена практичній частині, а саме застосуванню методів машинного навчання для параметризації даних. Використання алгоритмів Random Forest та символічної регресії дозволило порівняти ефективність різних підходів для моделювання та аналізу залежностей, пов’язаних із рухливістю носіїв заряду. Такі методи дали змогу отримати точні результати та виявити ключові закономірності в даних.

\section\*{Список літератури}

\begin{thebibliography}{99}

\bibitem{klaassen1}

D.B.M. Klaassen. A Unified Mobility Model for Device Simulation. Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands.

\bibitem{klaassen2}

D.B.M. Klaassen. A Unified Mobility Model for Device Simulation--II. Temperature Dependence of Carrier Mobility and Lifetime. Philips Research Laboratories, P.O. Box 80000, 5600 JA Eindhoven, The Netherlands.

\bibitem{sofos2023}

Sofos F., Dritselis C., Misdanitis S., Karakasidis T., Valougeorgis D. Computation of flow rates in rarefied gas flow through circular tubes via machine learning techniques. 2023.

\bibitem{kronberger2024}

Kronberger G., Burlacu B., Kommenda M., Winkler S. M., Affenzeller M. Symbolic Regression. Chapman \& Hall, 2024. 312 с.

\bibitem{pysr}

PySR: High-Performance Symbolic Regression in Python and Julia. Available at: \texttt{https://astroautomata.com/PySR/} (accessed June 3, 2024).

\end{thebibliography}

\end{document}