

**Многомасштабное моделирование физических
явлений и систем
(часть II)**

Лекция 2

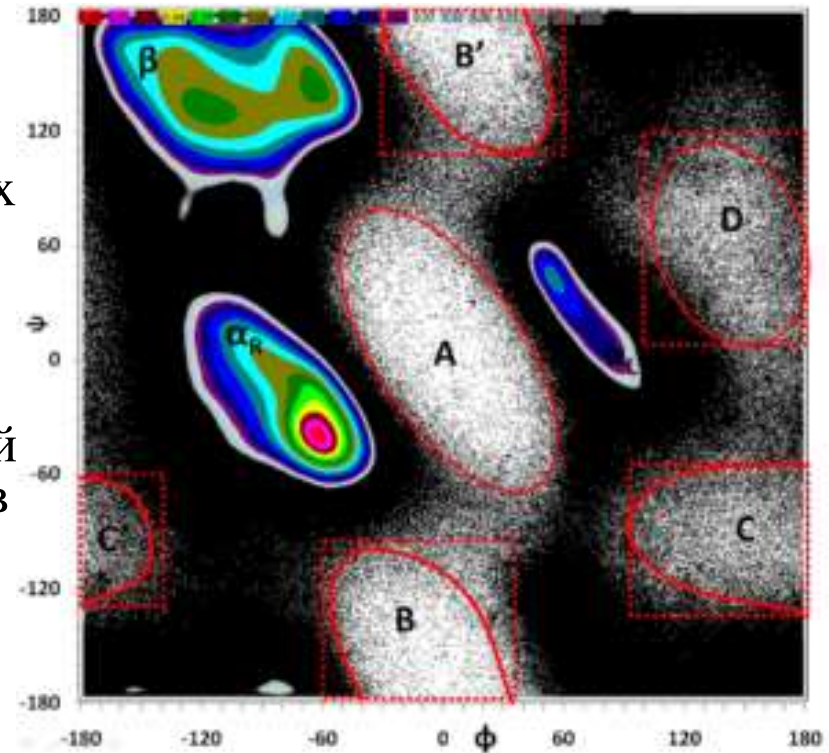
План лекции:

1. Применение методов машинного обучения в:

- физике белка,**
- биоинформатике,**
- физике твердого тела.**
- прикладной лингвистике,
- робототехнических науках
- биологических науках,
- биомедицинских науках,
- поведенческих науках

Применение методов машинного обучения в физике белка

- Взаимосвязь между различными уровнями структурами белка (символьной последовательностью аминокислот, вторичной структурой, пространственной (третичной) структурой) — одна из важных областей применения методов анализа «сверхбольших данных». Более 140,000 пространственных структур в [PDB](#) (protein data bank), которым соответствует более 1 млн. аминокислотных последовательностей суммарно включающие десятки миллионов аминокислотных остатков, каждый из которых уникальным образом влияет на стабильность всей молекулы белка: анализ комплексных взаимосвязей между этими уровнями лежит вне пределов «общепризнанных» подходов к [биоинформатике](#).



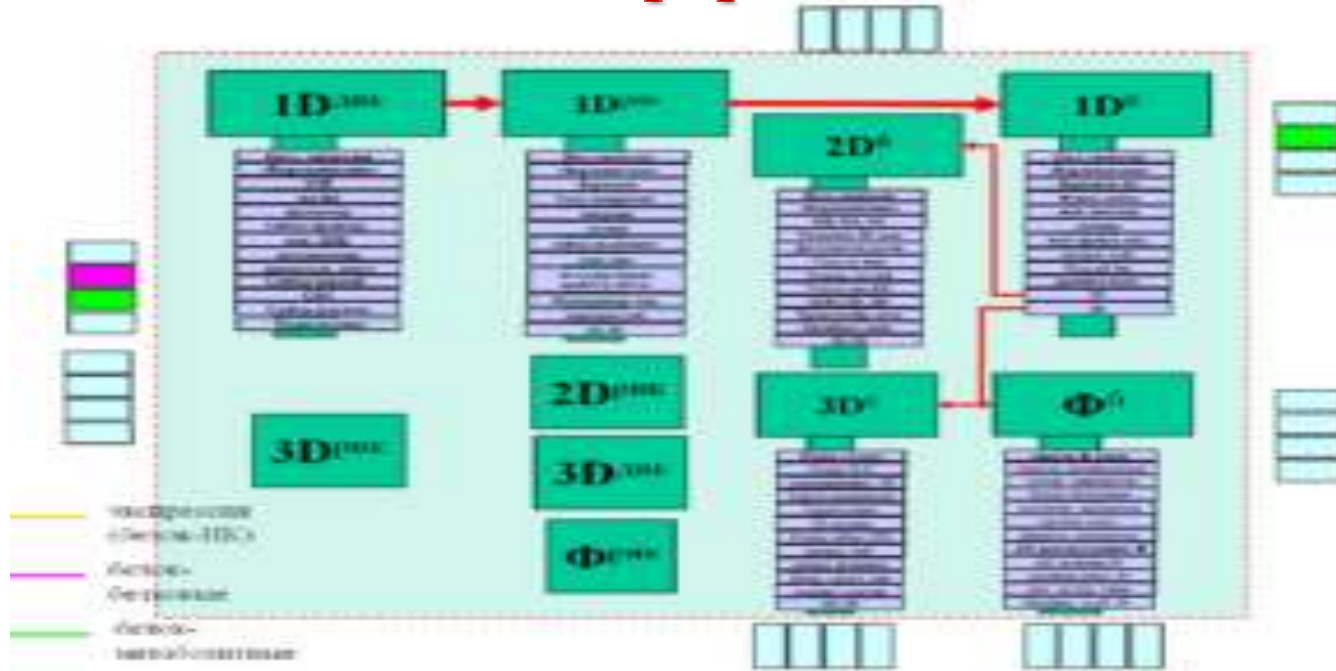
- <https://Bigdata-mining.ru>

Применение методов машинного обучения в физике белка

- Применение современных методов анализа сверхбольших данных является приоритетным и, пожалуй, наиболее [перспективным направлением исследований в области биоинформатики](#), математической и вычислительной биологии. Разрабатываемые в рамках алгебраического подхода методы анализа сверхбольших метрических конфигураций основаны на фундаментальных свойствах компактности и плотности метрических пространств, возникающих при формализации задач распознавания и классификации. Понятийный аппарат, вводимый в рамках разрабатываемого математического формализма, позволяет разрабатывать шкалируемые параллельные вычислительные алгоритмы для анализа «сгущений» точек в метрических конфигурациях большой размерности (миллионы, десятки миллионов точек).

Noncanonical and Strongly Disallowed Conformations of the Backbone in Polypeptide Chains of Globular Proteins

Применение методов машинного обучения в биоинформатике

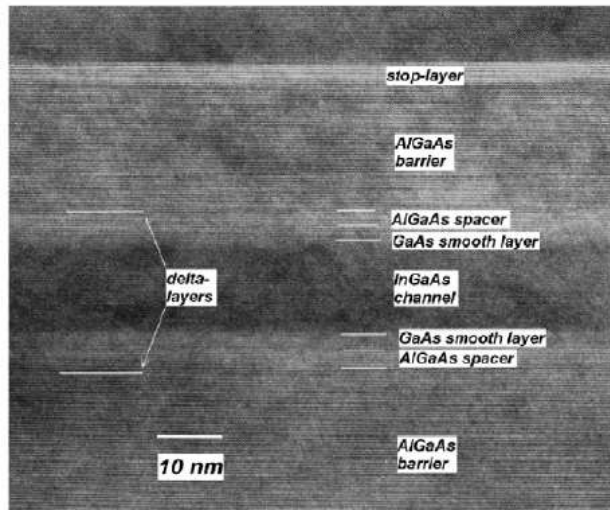


В биоинформатике уже более 40 лет существует широкий класс задач, связанных с обработкой данных сотен тысяч пространственных структур белков, миллионов аминокислотных последовательностей, миллиардов нуклеотидных последовательностей. Известные в западной биоинформатике методы основаны на многочисленных постулатах, подавляющее большинство из которых произвольны и не имеют четкого обоснования даже в самой проблемной области. В результате, повсеместно используемые в биоинформатике методы не обладают высокой специфичностью и селективностью распознавания и, с точки зрения практического биолога или врача, не представляют существенной практической ценности.

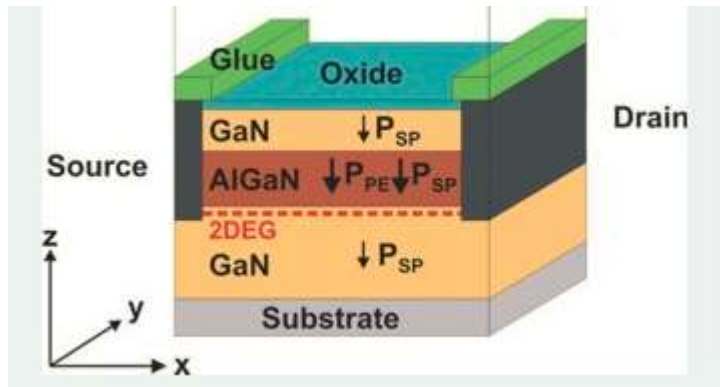
Применение методов машинного обучения в физике твердого тела

- Теория химической связи, [теория хемографов](#) (лекция 1) и методы метрического анализа данных позволяют разрабатывать системы прогнозирования свойств кристаллов. В частности, на основе метрических методов анализа информативности разнородных признаков описаний «больших данных» разработана система прогнозирования свойств сверхпроводимости кристаллических структур, показавшая достаточно высокое соответствие прогнозов эксперименту ($r=0.77$ в кросс-валидации) [1].
- Сверхпроводимость – свойство некоторых материалов обладать абсолютно нулевым электрическим сопротивлением при достижении ими температуры ниже определённого значения (т.н. критической температуры).
- **Сверхпроводимостью обладают** металлы и их сплавы, полупроводники, а также керамические материалы и тд. Существуют сверхпроводящие сплавы и материалы, у которых один из [элементов](#) или все [элементы](#), входящих в его состав, могут и не быть сверхпроводниками. Например, [сероводород](#), сплавы ртути с золотом и оловом.
- **Сверхпроводимость как явление** сопровождается несколькими эффектами. Определяющее значение имеют два из них: исчезновение электрического сопротивления и выталкивание магнитного потока (поля) из его объема. Поэтому важнейшее значение приобретает не только критический ток, но и критическое магнитное поле – определенное значение напряженности магнитного поля, по достижении которого сверхпроводник теряет свойство сверхпроводимости.
- **Явление сверхпроводимости** На практике. Если взять проводник, закольцевать его, сделав замкнутый [электрический](#) контур, охладить его до температуры ниже критической и подвести к нему электрический ток, а после чего убрать источник электрического тока, то электрический ток в таком проводнике будет существовать неограниченно долгое время.

Применение методов машинного обучения в физике твердого тела



Современное материаловедение оперирует сотнями тысяч кристаллических структур неорганических веществ, для которых известны сотни физических свойств. Для дизайна перспективных материалов с заданными свойствами крайне важны теоретические методы, которые бы позволяли прогнозировать комплекс физических свойств монокристаллических и поликристаллических фаз на основе информации о пространственной структуре кристаллов, концентрации дефектов и другой структурной информации.



Применение методов машинного обучения в физике твердого тела

Моделирование тепловых свойств наноразмерных структур в настоящий момент является востребованным направлением научных исследований, что связано с постоянно растущей скоростью работы микроэлектронных элементов, которые выделяют все большее количество энергии в виде тепла, которую требуется отводить, чтобы избежать перегрева и потери функциональных свойств устройств[1,2].

Для построения моделей теплопереноса в наноразмерных гетероструктурах свою эффективность показали методы на основе решения кинетического уравнения Больцмана для фононов[2]. Более детальное моделирование можно получить с использованием методов молекулярной динамики, но это сопряжено с более высокой вычислительной сложностью и нетривиальной задачей подбора оптимального потенциала[3]. Стоит отметить, что применение классических подходов для решения задач теплопроводности на основе закона Фурье для рассматриваемых наноструктур дает неудовлетворительные результаты, т.к. при подобном подходе игнорируются квантовомеханические эффекты материалов, что дает сильное рассогласование с экспериментальными данными[2].

Абгарян К.К., Колбин И.С.

ПОСТРОЕНИЕ МОДЕЛИ ЭФФЕКТИВНОГО КОЭФФИЦИЕНТА ТЕПЛОПРОВОДНОСТИ ДЛЯ НАНОРАЗМЕРНЫХ ГЕТЕРОСТРУКТУР
С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Применение методов машинного обучения в физике твердого тела

Технологии машинного обучения получили сильное развитие в последнее десятилетие, в том числе ведутся активные исследования в применении подобных подходов в задачах материаловедения[4,5]. При моделировании материалов хочется получить точные характеристики при этом используя при этом минимальное количество эмпирических данных. Подобными свойствами обладают первопринципные методы расчета, к сожалению они обладают существенной вычислительной сложностью, в связи с чем рассматриваемые системы как правило ограничены размером в сотни, иногда тысячи атомов. Было продемонстрировано, что подходы машинного обучения позволило проводить адаптированные первопринципные расчеты для систем из миллионов атомов на существующем в настоящий момент оборудовании за приемлемое время[5].

Применение методов машинного обучения в физике твердого тела

Целью настоящей работы является исследование применимости подходов машинного обучения для моделирования эффективного коэффициента теплопроводности наноразмерных гетероструктур, в частности сверхрешеток. Для этого предполагается генерация выборки на основе модели модального подавления[6] и обучения нейронных сетей на ней. На выходе алгоритма будут получены компактные нейросетевые модели, которые будут сравниваться по точности работы на тестовом наборе данных, для определения лучшей сети.

В качестве примера были построены нейросетевые модели для расчета эффективного коэффициента теплопроводности для бинарных гетероструктур: сверхрешеток GaAs/AlAs с различными периодами слоев. Данные для обучения были сгенерированы в программном пакете AlmaBTE 1.3.2[1], параметры материалов получены из открытой базы данных проекта. Выборка формировалась для различных комбинаций содержания GaAs и AlAs, толщин пленок, периодов сверхрешетки.

Абгарян К.К., Колбин И.С.

ПОСТРОЕНИЕ МОДЕЛИ ЭФФЕКТИВНОГО КОЭФФИЦИЕНТА ТЕПЛОПРОВОДНОСТИ ДЛЯ НАНОРАЗМЕРНЫХ ГЕТЕРОСТРУКТУР
С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Применение методов машинного обучения в физике твердого тела

Полученный массив данных был разделен на 3 части: 60% для обучения нейросетей, 20% для валидации (для избежания переобучения) и 20% как тестовая выборка, для оценки результирующей точности обученных моделей. Оптимизация нейросетей велась с использованием алгоритма RMSprop с шагом 0.0001 в среде Tensorflow 2.3[7].

В качестве модели для расчета использовались многослойные нейронные сети прямого распространения. В работе были рассмотрены сети с различным числом скрытых слоев, также проведены сравнения для варьирующего числа нейронов и распространенных активационных функций. Полученные в результате обучения сети сравнивались по среднеквадратичной ошибке и средней абсолютной ошибке на тестовой выборке. В итоге наиболее высокая точность на тестовой выборке была получена на трехслойной сети с ReLU активационной функцией, с 16 нейронами на каждом скрытом слое: средняя абсолютная ошибка 0.502 Вт/(м К), среднеквадратичная ошибка 0.544 Вт/(м К). Относительная среднеквадратичная ошибка при этом составляет порядка 3-5%, что позволяет говорить о достаточно хорошей точности выбранного подхода.

Абгарян К.К., Колбин И.С.

ПОСТРОЕНИЕ МОДЕЛИ ЭФФЕКТИВНОГО КОЭФФИЦИЕНТА ТЕПЛОПРОВОДНОСТИ ДЛЯ НАНОРАЗМЕРНЫХ ГЕТЕРОСТРУКТУР
С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Применение методов машинного обучения в физике твердого тела

Литература

1. *J. Carrete, B. Vermeersch, A. Katre, A. van Roekeghem, T. Wang, G. Madsen, N. Mingo.* AlmaBTE: a solver of the space-time dependent Boltzmann transport equation for phonons in structured materials. *Comp. Phys. Commun.* 220C, 351 (2017)..
2. *Хвесьюк В.И., Скрябин А.С.* Теплопроводность наноструктур. *Теплофизика высоких температур*, том 55, н. 3, с. 446–471 (2017).
3. *Абгарян К.К., Евтушенко Ю.Г., Мутигуллин И.В., Уваров С.И.* Молекулярно–динамическое моделирование начальных этапов процесса нитридации поверхности Si(111) в атмосфере NH₃. // *Известия вузов. Материалы электронной техники.* 2015. Т. 18. № 4. С.267-272
4. *Mortazavi, B., Podryabinkin, E. V., Roche, S., Rabczuk, T., Zhuang, X., & Shapeev, A. V.* Machine Learning Interatomic Potentials Enable First-Principles Multiscale Modeling of Lattice Thermal Conductivity in Graphene/Borophene Heterostructures. *Materials Horizons.* — 2020, *Materials Horizons.*
5. *Weile Jia et al.* Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning. — 2020, preprint <https://arxiv.org/pdf/2005.00223v1.pdf>.
6. *Vermeersch, B., Carrete, J., & Mingo, N.* (2016). Cross-plane heat conduction in thin films with ab-initio phonon dispersions and scattering rates. *Applied Physics Letters*, 108(19), 193104.
7. *Martín Abadi et al.* TensorFlow: A System for Large-Scale Machine Learning. — *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16).*

1. Произвести обзор современных тенденций по моделям теплопереноса в многослойных структурах
2. Разработать физико-математическую оптимизационную модель теплопереноса в многослойных полупроводниковых наноструктурах;
3. Применить разработанную оптимизационную модель к изучению процессов теплопереноса в арсенид галлиевых структурах типа AlAs/GaAs
4. Сравнить результаты моделирования с данными, полученными из экспериментов с помощью пакетных приложений.

