

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

Лекция 1. Интеграция многомасштабного моделирования и методов анализа данных: перспективы, проблемы и возможности применения в химии, физике твердого тела, биомедицинских и других науках

## **Цели курса:**

Получить практические навыки по формированию многомасштабных моделей и информационных систем в различных областях естественнонаучных исследований, с применением методов многомасштабного моделирования, навыков построения физико-математических моделей, методов анализа и управления данными, включая большие данные, программных методов и средств построения многоуровневых информационных систем, с базами знаний, по направлениям исследований из области естествознания.

# **Тема лекции 1:** Интеграция многомасштабного моделирования и методов анализа данных: перспективы, проблемы и возможности применения в естествознании

## **План лекции:**

1. Основные положения
2. Применение методов машинного обучения в химии. Теория хемографов и анализ больших данных в химии.
3. Многомасштабное моделирование иерархических хемосенсоров
4. Применение методов машинного обучения в физике твердого тела, физике белка, биоинформатике, прикладной лингвистике, робототехнике биологических, биомедицинских и поведенческих науках

## **Интеграция многомасштабного моделирования, методов анализа больших данных. Проблемы, перспективы, возможности в химии, физике, биологии, медицине и поведенческих науках**

- В настоящее время благодаря революционным технологиям физические, химические, биологические, биомедицинские и поведенческие науки собирают больше данных, чем когда-либо прежде. Существует острая необходимость в экономичных по времени и затратах стратегиях для анализа и интерпретации этих данных для улучшения здоровья людей, улучшения качества их жизни и др.
- Постоянное развитие машинного обучения в качестве мощного метода для интеграции мультимодальности, данных с множественной точностью и выявления корреляций между взаимосвязанными явлениями представляет особую возможность в этом отношении.
- Однако классические методы машинного обучения часто игнорируют фундаментальные законы физики и приводят к некорректным задачам или нефизическим решениям.
- Многомасштабное моделирование это успешная стратегия интеграции мультимасштабных, многофизических данных и раскрытия механизмов, объясняющих появление функциональных зависимостей.

## **Интеграция многомасштабного моделирования, методов анализа больших данных. Проблемы, перспективы, возможности в химии, физике, биологии, медицине и поведенческих науках**

•Однако, одно многомасштабное моделирование часто не позволяет эффективно комбинировать большие наборы данных из разных источников и с разным уровнем разрешения. Авторы отдельных научных публикаций, включая (\*) показывают, как машинное обучение и многомасштабное моделирование могут дополнять друг друга, создавая надежные прогностические модели, которые объединяют основную физику для управления некорректными задачами и исследования больших пространств проектирования. Они критически рассматривают текущую литературу, выделяют приложения и возможности, решают открытые вопросы и обсуждают потенциальные проблемы и ограничения в четырех всеобъемлющих актуальных областях:

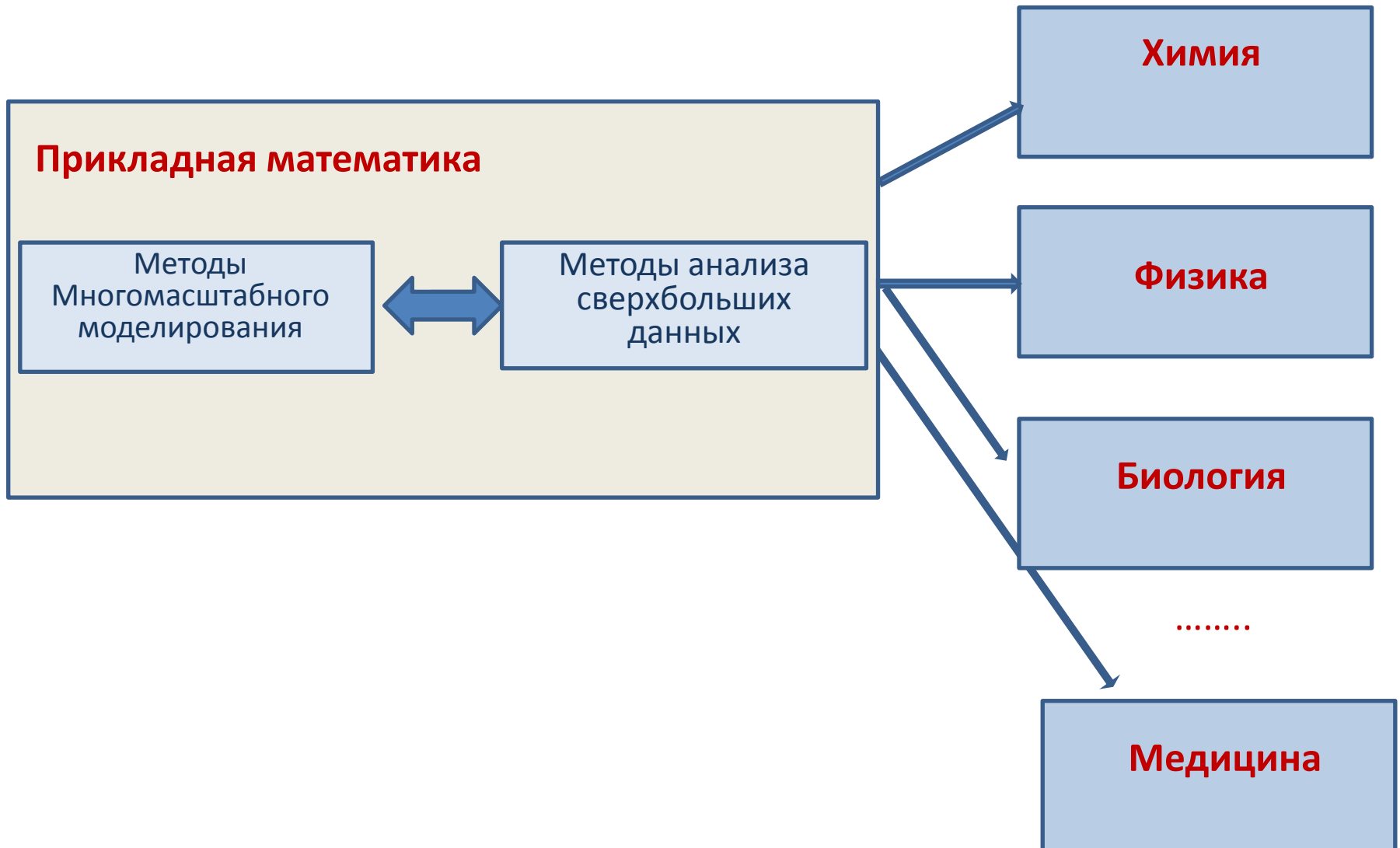
- обыкновенные дифференциальные уравнения, уравнения в частных производных, подходы, основанные на данных, и подходы, основанные на теоретическом физ-мат моделировании.

• Для достижения поставленных целей используется опыт в прикладной математике, информатике, вычислительной биологии, биофизике, биомеханике, инженерной механике, экспериментах и медицине.

• Междисциплинарный подход предполагает, что интеграция машинного обучения и многомасштабного моделирования может дать новое представление о механизмах создания новых материалов, процессов разработки новых лекарств, о причинах заболеваний, помочь в определении новых целей и стратегий лечения, а также в принятии решений на благо здоровья человека.

\* [Марк Альбер](#), [Уильям Р. Кэннон](#) Интеграция машинного обучения и многомасштабного моделирования: перспективы, проблемы и возможности в биологических, биомедицинских и поведенческих науках

# Интеграция многомасштабного моделирования, методов анализа больших данных. Проблемы, перспективы, возможности



# Основные положения

## Методы многомасштабного моделирования

Математической основой для разработки новейших методов многомасштабного моделирования являются:

- теоретико-множественный аппарат, применяемый для описания формализованной технологии многомасштабного моделирования,
- декомпозиция поставленной задачи, путем выделения масштабных уровней,
- использование на каждом уровне соответствующих пространственно-временных масштабов, моделей и приближений,
- применение теоретико-множественного подхода к синтезу многомасштабных композиций, лежащих в основе многомасштабных моделей
- математический аппарат теории принятия решений для выбора наилучших многомасштабных композиций для решаемой задачи,
- математические методы решения обратных задач,
- Модельно-ориентированный подход для создания программных систем

# Основные положения

## Методы анализа «сверхбольших данных»

Математической основой для разработки новейших подходов к анализу сверхбольших данных являются:

- топологический подход к анализу плохо-формализованных задач распознавания и классификации,
- метрический подход к анализу разнородных признаков описаний,
- теория анализа размеченных графов,
- алгебраический подход к синтезу корректных алгоритмов,
- комбинаторная теория разрешимости,
- теория классификации значений признаков,
- метрическая теория корректности и полноты алгоритмов.



## **Основные положения**

### **Методы анализа «сверхбольших данных»**

Математической основой для разработки новейших подходов к анализу сверхбольших данных являются:

- топологический подход к анализу плохо-формализованных задач распознавания и классификации,
- метрический подход к анализу разнородных признаков описаний,
- теория анализа размеченных графов,
- алгебраический подход к синтезу корректных алгоритмов,
- комбинаторная теория разрешимости,
- теория классификации значений признаков,
- метрическая теория корректности и полноты алгоритмов.

## Теория химической связи, теория хемографов и анализ сверхбольших данных из области хемоинформатики

- *Химическая связь* — это взаимодействие атомов, обуславливающее устойчивость молекулы или кристалла как целого.

### Химическая связь

определяется взаимодействием между заряженными частицами (ядрами и электронами). Современное описание химической связи проводится на основе квантовой механики. Основные характеристики химической связи-прочность, длина, полярность

- Для описания электронных облаков атомов и молекул в квантовой физике используется т.н. *волновая функция* — комплекснозначная функция, описывающая вероятность обнаружения объекта (в данном случае, электрона) в заданной точке в заданный момент времени. Взаимодействия волновых функций частиц в современной физике и химии описываются основным уравнением квантовой механики — *уравнением Шрёдингера*.

## Теория химической связи, теория хемографов и анализ сверхбольших данных из области хемоинформатики

Уравнение Шрёдингера предназначено для описания гамильтоновых систем (т.е. систем без диссипации энергии и без зависимости сил от скорости) и является аналогом 2-го закона Ньютона в классической механике. В трёхмерном случае (одна частица), уравнение Шрёдингера записывается следующим образом:

$$(I) \quad -\frac{\hbar^2}{2m} \left( \frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} \right) + U(x, y, z, t) \Psi = i\hbar \frac{\partial \Psi}{\partial t},$$

где  $\hbar = \frac{h}{2\pi}$  ( $h$  - постоянная Планка),  $m$  - масса частицы,  $U(x, y, z, t)$  - потенциальная энергия в точке  $(x, y, z)$  в момент времени  $t$ . В стационарном виде, когда  $U$  не является функцией времени, и при векторе координат  $\vec{r}$  произвольной размерности уравнение (I) принимает следующий вид:

$$(II) \quad -\frac{\hbar^2}{2m} \Delta \psi(\vec{r}) + U(\vec{r}) \psi(\vec{r}) = E \psi(\vec{r}),$$

где  $E$  - полная энергия частицы при движении, описываемом функцией  $\psi(\vec{r})$ .

## Теория химической связи, теория хемографов и анализ сверхбольших данных из области хемоинформатики

Вычисления геометрических конфигураций (конформаций, кристаллических структур) и других свойств молекул по уравнению Шрёдингера чрезвычайно затруднительны, когда необходимо анализировать сверхбольшие данные современной хемоинформатики (сотни миллионов молекул, миллионы активностей молекул).

Однако, анализ решений уравнения Шрёдингера для различных элементов периодической таблицы Менделеева (ТМ) атомарных систем позволил установить, что волновые функции (орбитали) атомов бывают четырёх основных типов: s-орбитали (сферические), p-орбитали (8-образные), d-орбитали и f-орбитали.

*Гибридными орбиталями* отдельных атомов называются линейные комбинации орбиталей различных типов, подставляемые в уравнение Шрёдингера. Понятие *гибридизации* данного атома (т.е. конкретного вида терма (кривые зависимости энергии от межъядерного расстояния (**термы**)) для гибридной орбитали атома) принципиально важно для описания конформаций молекул/кристаллов и позволило разработать теорию анализа *хемографов* (т.е. конечных размеченных графов без петель с кликовым числом не более 3).

## **Теория химической связи, теория хемографов и анализ сверхбольших данных из области хемоинформатики**

Методы теории анализа хемографов (понятие  $\chi$ -графа (хемогр́афа) – специальной разновидности размеченного графа для описания химической структуры молекул) успешно применяются в хемоинформатике при работе со сверхбольшими данными для прогнозирования фармакологических свойств молекул-лигандов на основе их молекулярных структур, а также для прогнозирования эффектов молекул-лигандов на транскриптом и протеом человека и др. Перспективным направлением приложения теории хемографов также является прогнозирование свойств неорганических кристаллов (в частности, свойств сверхпроводимости).



# Многомасштабное моделирование хемосенсоров

Описан многомасштабный подход для предсказательного моделирования иерархических наноструктурированных материалов для оптических хемосенсоров. Стратегия этого подхода базируется на иерархической структуре материала, в которой ключевым элементом является супрамолекулярный рецепторный центр (СРЦ), представляющий собой молекулу органического красителя вместе с ее ближайшим окружением. Структура, стабильность и спектральный отклик конкретного СРЦ определяются с использованием методов молекулярного моделирования. Структура компонентов СРЦ рассчитывается методами квантовой химии. Возможные конфигурации СРЦ рассчитываются с использованием классического силового поля для описания межмолекулярных взаимодействий компонент СРЦ. Для расчета уточненной геометрии, относительной стабильности, требуемых спектральных свойств СРЦ и его спектрального отклика (изменений в спектрах поглощения или флуоресценции) на взаимодействие с молекулами аналитов. Свободная энергия образования супрамолекулярных комплексов (СРЦ + аналит) может быть определена методом молекулярной динамики.

<sup>1</sup> Русская версия обзора, который выходит в свет на английском языке в издании Chemical Sensors: Simulation and Modeling Volume 4: Optical Sensors, Edited by Ghenadii Korotcenkov, Momentum Press, 2013, pp. 1–38. Публикуется с разрешения издательства Momentum Press.

# Многомасштабное моделирование хемосенсоров

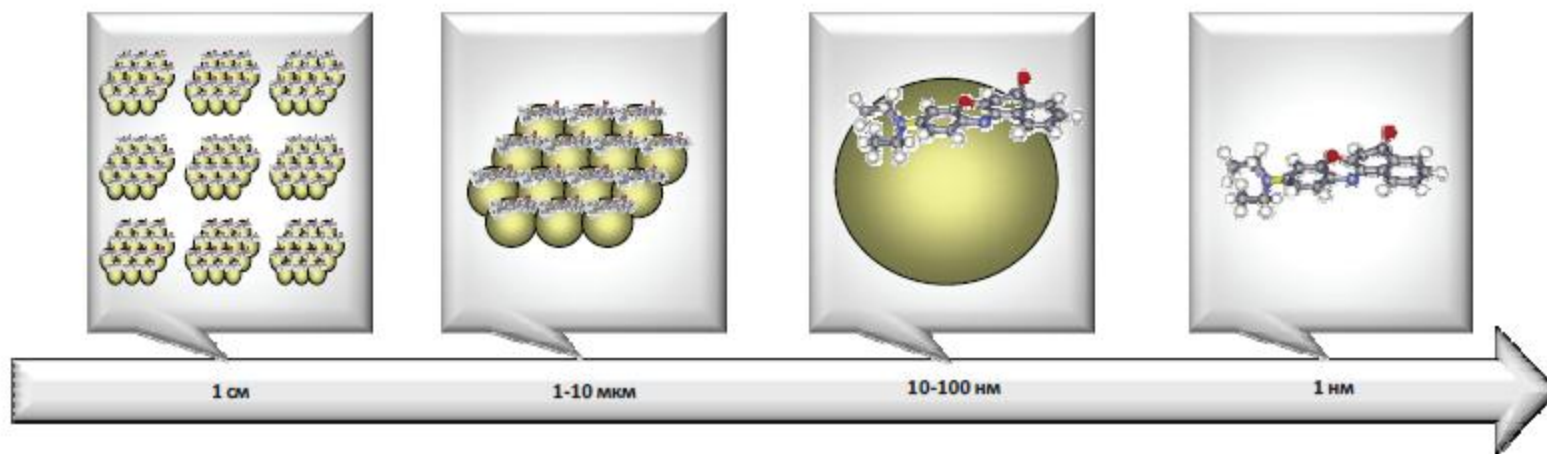


Рис. 1. Иллюстрация иерархически построенного наноструктурированного материала для оптического хемосенсора: (а) массив сенсорных элементов; (б) организованный ансамбль наночастиц, несущих рецепторные центры; (в) наночастица с молекулой органического красителя, адсорбированной на ее поверхности и образующей рецепторный центр; и (г) молекула индикаторного красителя (Нильского Красного), которая придает функциональные свойства сенсорному материалу.

# Многомасштабное моделирование хемосенсоров

На Рис. 1 схематически представлена структура иерархического наноструктурированного материала для оптического хемосенсора. Органическая "индикаторная молекула" (молекула красителя Нильского Красного на рисунке) лежит в основе конструкции сенсорного материала. Именно она отвечает за генерацию оптического сигнала при взаимодействии материала с аналитом. Таким образом, индикаторная молекула придает сенсорному материалу его главные функциональные свойства. индикаторная молекула вместе с ее локальным окружением образует "рецепторный центр". Локальное окружение может быть выбрано таким образом, чтобы усилить сенсорный сигнал и улучшить его избирательность. Рецепторные центры располагаются на поверхности или в объеме частиц (наночастиц), а сами наночастицы собираются в организованную структуру (которая может, в частности, обладать свойствами фотонного кристалла) [6,7] Флуоресцентные оптические хемосенсоры, в которых выходным сигналом является изменение флуоресцентных свойств сенсорного материала, принадлежат к числу наиболее перспективных, вследствие высокой чувствительности флуоресцентного сигнала к различным агентам и, как следствие, возможности детектирования крайне низких концентраций аналитов [35]. При разработке сенсоров необходимо знать заранее наиболее важные свойства материала, такие как энергии взаимодействия аналитов с рецепторным центром и изменения спектральных свойств вследствие этого взаимодействия. Такие свойства материалов могут быть предсказаны с использованием компьютерного моделирования.



## **Иерархия пространственно-временных масштабов и структура наноматериалов для нанофотоники**

Оптические хемосенсоры, использующие молекулярное распознавание молекул аналитов, основаны на эффекте изменения оптических свойств индикаторной молекулы при ее взаимодействии с аналитом. Индикаторная молекула обычно представляет собой молекулу красителя (или другого вещества, обладающего подходящими спектральными свойствами). Следовательно, материал оптического хемосенсора содержит индикаторные молекулы, а сигнал сенсора представляет собой изменение его оптических свойств (положения и интенсивностей спектральных полос) при его взаимодействии с аналитом. Такое изменение оптических свойств молекулы или более сложной системы (супрамолекулярной, наноразмерной и т.д. системы) будет также называться ее оптическим откликом. Наконец, оптический отклик сенсора при его взаимодействии с газом, содержащим некоторую концентрацию данного аналита, будет служить в качестве его аналитического сигнала.

Использование иерархически построенного, наноструктурированного материала усиливает чувствительность и повышает селективность оптического хемосенсора. Такой материал обладает очень большой эффективной поверхностью, которая обеспечивает наиболее эффективное взаимодействие аналита с рецепторными центрами, легкость проникновения аналита в материал и возможность конструирования архитектуры материала на всех уровнях масштаба.

## **Структура наноматериалов для оптических хемосенсоров, От молекулы к супрамолекулярному центру, к наночастице и наноматериалу**

На низшем, молекулярном уровне индикаторная молекула является основным сенсорным элементом такого материала. Выбор индикаторной молекулы определяется ее способностью (селективно) взаимодействовать с молекулой аналита и ее способностью (селективно) изменять свои оптические свойства (в первую очередь, интенсивность флуоресценции) при этом взаимодействии. Предполагается, что такой выбор может быть сделан на основе имеющихся данных о химической природе и физических оптических свойствах известных красителей и других светопоглощающих или светоиспускающих соединениях.

На следующем, супрамолекулярном уровне предполагается создание ("конструирование") супрамолекулярного рецепторного центра. Супрамолекулярный рецепторный центр (СРЦ) включает индикаторную молекулу и ее ближайшее окружение. Это окружение (архитектура СРЦ) выбирается (конструируется) таким образом, чтобы оптимизировать взаимодействие между молекулой аналита и СРЦ, максимизировать селективность этого взаимодействия и как можно более усилить оптический отклик сконструированного таким образом СРЦ.

## **Структура наноматериалов для оптических хемосенсоров, От молекулы к супрамолекулярному центру, к наночастице и наноматериалу**

На более высоком, наноразмерном уровне создается наночастица, содержащая СРЦ на своей поверхности или в ее объеме. Предполагается, что известны структура как самой наночастицы, так и ее поверхности. Построение сенсорного материала из наночастиц обеспечивает увеличение поверхности материала, доступной для аналитов, легкое проникновение молекул аналита к рецепторным центрам, а также возможность создания специальной структуры (архитектуры) материала с наилучшими условиями для рассеяния (или поглощения) света, например, путем создания фотонной решетки.

На следующем иерархическом, микро уровне образуется двух- или трехмерный ансамбль однородных наночастиц, упорядоченных некоторым образом. Этот ансамбль содержит индикаторные молекулы только одного определенного типа. Индикаторные молекулы этого типа характеризуются определенным значением отклика на определенную молекулу аналита. Именно такой ансамбль и составляет основу конструируемого материала для оптических хемосенсоров. Из набора таких ансамблей можно построить массив (матрицу) ансамблей. Такая матрица представляет собой основной сенсорный элемент оптического хемосенсора, рассматриваемого как устройство (макро уровень). Каждый элемент ( $A_i$ ) такой матрицы может характеризоваться своей собственной величиной отклика ( $S_i$ ) на взаимодействие с определенной молекулой аналита. Анализ всей

## **Структура наноматериалов для оптических хемосенсоров, От молекулы к супрамолекулярному центру, к наночастице и наноматериалу**

совокупности откликов  $\{S_i\}$  от каждого элемента матрицы при ее взаимодействии с данной многокомпонентной газовой смесью позволяет получить информацию о составе смеси и о концентрациях заданных аналитов в смеси.

Основной задачей многомасштабного моделирования является предсказание и конструирование структуры материала и его отклика на каждом уровне масштаба. Результаты моделирования на каждом уровне переносятся на следующий уровень и, таким образом, реализуется схема многомасштабного (многоуровневого) моделирования. Эта схема соответствует конструированию материала снизу-вверх. В действительности, эта схема включает и некоторые обратные связи между различными уровнями, когда результаты высшего уровня используются для уточнения результатов нижнего уровня моделирования (подход "сверху-вниз"). Такой комбинированный подход (снизу-вверх/сверху-вниз) широко используется для моделирования и теоретического предсказания структуры и свойств наноматериалов для широкого спектра различных применений (оптические хемосенсоры, светоизлучающие и фотовольтаические материалы; материалы для фотонных кристаллов; среды оптической памяти; и т.д.)



## **Структура наноматериалов для оптических хемосенсоров, От молекулы к супрамолекулярному центру, к наночастице и наноматериалу**

Основная особенность такого многомасштабного подхода состоит в том, что моделирование стартует, начиная с самого низшего, молекулярного уровня, с использованием атомистических неэмпирических (первопринципных) методов, основанных на фундаментальных законах атомных и молекулярных взаимодействий. На каждом следующем уровне результаты предыдущего уровня используются как входные параметры, причем моделирование основывается на физических моделях и методах, специально разработанных для соответствующих временных и пространственных масштабов.

В соответствии со сказанным выше, многомасштабный подход к моделированию оптических материалов должен включать четыре основных уровня: молекулярный уровень (1–2 нм), супрамолекулярный уровень (2–10 нм), уровень наночастиц (10–200 нм) и уровень ансамблей наночастиц (10–50 мкм), см. Рис. 2.

Сигнал, генерируемый оптическим хемосенсором для детектирования аналита (или набора аналитов) представляет собой изменение оптических характеристик (как правило, спектров люминесценции) сенсорного материала при его взаимодействии с аналитом. Это изменение должно определяться специальным детектором. Характеристики сгенерированного сигнала и, следовательно, его интенсивность и селективность определяются физическими параметрами и процессами на каждом иерархическом уровне. Поэтому соответствующие физические характеристики должны моделироваться на каждом иерархическом уровне материала.

# Иерархия пространственно-временных масштабов и структура наноматериалов для нанофотоники

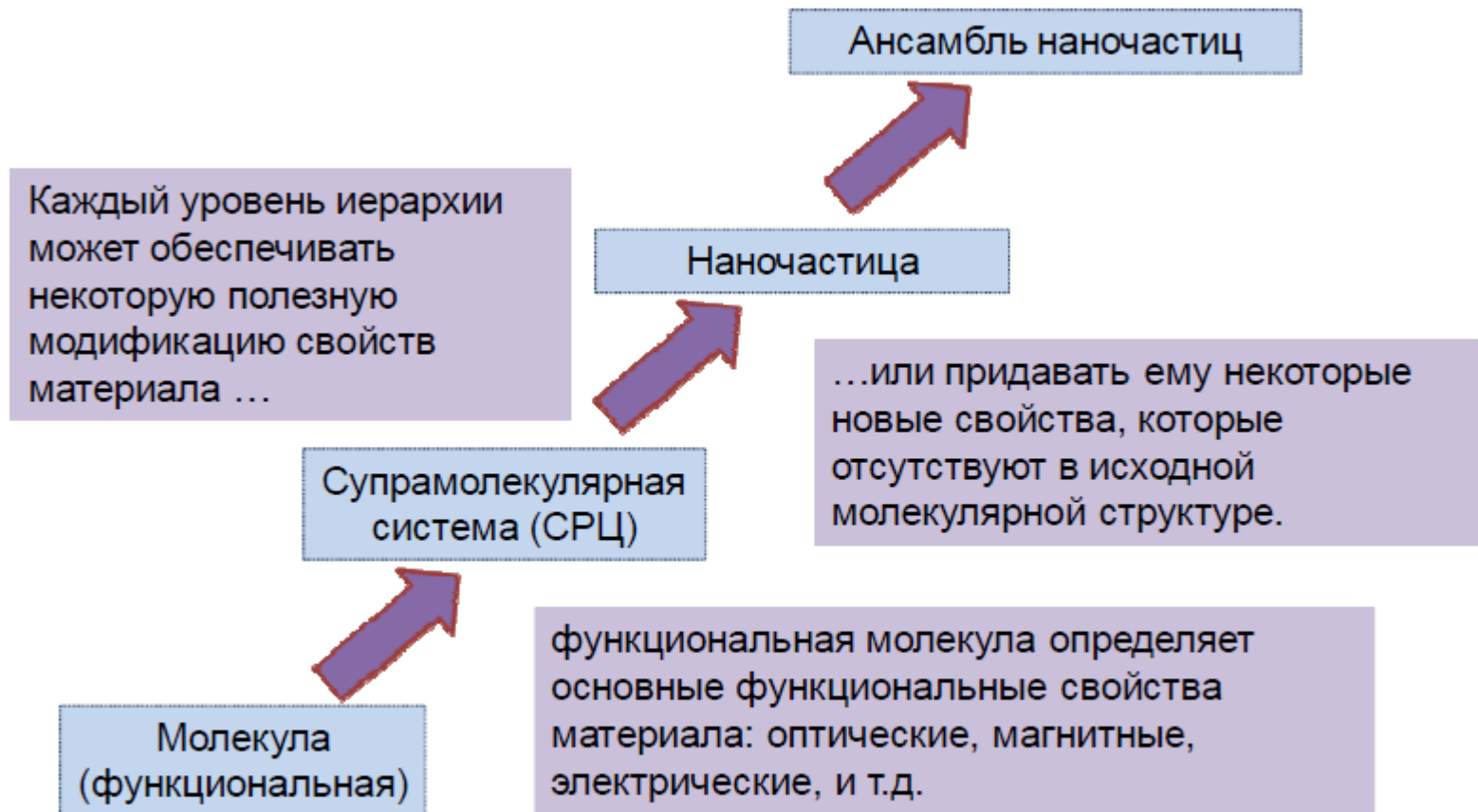


Рис. 2. Иерархические уровни функционального материала для применений в нанофотонике.