

ООО «ТРАНЗИСТ ВИДЕО», резидент Сколково

На правах рукописи

Купервассер Олег Юрьевич

Применение принципов новой кибернетики в молекулярной физике:
учет влияния окружающей среды/наблюдателя
с помощью методов математического и компьютерного моделирования

05.13.18 — Математическое моделирование, численные методы и комплексы программ

Диссертация на соискание ученой степени
доктора физико-математических наук

Москва 2016

Оглавление

Введение.....	3
Актуальность темы.....	3
Цели диссертационной работы.....	10
Положения, выносимые на защиту.....	10
Научная новизна.....	11
Практическая ценность.....	12
Апробация.....	12
Публикации.....	12
Личный вклад автора.....	13
Объем и структура работы.....	13
Глава 1. Системы, рассчитываемые на основе известной физики подсистем.....	14
Глава 1.1 Первая группа проблем, где учет влияния внешней среды/наблюдателя, рассчитывается на основе точных законов физики – идеальной динамики.....	14
Глава 1.1.1 Континуальная модель среды I: Алгоритм для построения гладкой молекулярной поверхности.....	14
Глава 1.1.2 Континуальная модель среды II: Универсальный адаптивный алгоритм для триангуляции гладкой молекулярной поверхности.....	38
Глава 1.1.3 Континуальная модель среды III: Вычисление аналитических градиентов параметров поверхностных элементов на молекулярных поверхностях по координатам атомов.....	55
Глава 1.1.4 Континуальная модель среды IV: Вычисление аналитических градиентов для энергии сольватации по координатам атомов.....	75
Глава 1.1.5 Укрупненные поверхностные элементы и условия нормировки для столбцов и строк матриц метода COSMO.....	89
Глава 1.1.6 Итерационный метод расчета полярной составляющей энергии Гиббса растворения молекул при условии плавного изменения диэлектрической проницаемости раствора.....	104
Глава 1.2 Вторая группа проблем учета влияния внешнего шума/наблюдателя на основе наблюдаемой динамики макропараметров, рассчитываемой из идеальной динамики и независимой от амплитуды шума.....	115
Глава 1.2.1 Случайный шум и динамика полюсов при распространении неустойчивого фронта пламени.....	116
Глава 1.2.2 Проблема Сафмана-Тейлора без поверхностного натяжения в фильтрационном горении: Образование единственного "пальца" с шириной, равной половине ширины канала.....	137
Глава 2 Эмерджентные системы, не рассчитываемые аналитически на основе известной физики подсистем.....	153
Глава 2.1 Третья группа проблем, где учет внешнего шума/наблюдателя за счет модели «черного ящика» и построение его наблюдаемой динамики происходит из реального/численного эксперимента.....	153
Глава 2.1.1 Введение.....	153
Глава 2.1.2 Постановка задачи.....	155
Глава 2.1.3 Предварительные обозначения.....	155
Глава 2.1.4 Общая форма $P(A/x_1, x_2)$	156
Глава 2.1.5 Ограничения на функции $J(a, b)$ и $\bar{J}(a, b)$	158
Глава 2.1.6 Определение расстояния.....	159
Глава 2.1.7 Ограничение на основные функции.....	159
Глава 2.1.8 Оптимизация.....	161

Глава 2.1.9 Среднее расстояние между предложенной аппроксимацией функции $P(A/x_1, x_2) - \Gamma(\alpha, \beta, \theta)$ и реальной функцией $P(A/x_1, x_2)$	164
Глава 2.1.10 Случай более чем двух переменных A и X	167
Глава 2.1.11 Выводы	168
Глава 2.2 Четвертая группа проблем непредсказуемой динамики, где учет внешнего шума/наблюдателя происходит за счет интуиции наблюдателя.....	168
Глава 2.2.1 Введение	169
Глава 2.2.2 Непредсказуемая динамика.	170
Глава 2.2.3 Синергетические модели локальных корреляций.....	172
Глава 2.2.4 Синергетические модели глобальных корреляций, охватывающих всю вселенную.	177
Глава 2.2.5 Практическое применение непредсказуемых систем для искусственного интеллекта	181
Глава 3 Разрешение основных парадоксов физики на основе учета влияния окружающей среды и наблюдателя	188
Глава 3.1 Введение.....	188
Глава 3.2 Макроскопическая энтропия, закон увеличения энтропии и аргумент о стирании памяти наблюдателя	189
Глава 3.2.1 Релевантность аргумента о стирании памяти наблюдателя для классической механики.....	189
Глава 3.2.2 Правильное определение термодинамической энтропии в парадоксе Лошмидта	190
Глава 3.3 Парадокс Шредингеровского кота и спонтанная редукция.	191
Глава 3.4 Синхронизация стрел времени/декогеренция	195
Глава 3.5 Закон увеличения энтропии и синхронизации стрел времени/ декогеренция в гравитационной теории.	197
Глава 3.5.1 Черные Дыры	197
Глава 3.5.2 Кротовые норы	199
Заключение.....	202
Библиография.....	204

Введение

Актуальность темы.

Основу современных точных наук составляют математика и теоретическая физика. Они позволяют построить математические модели многих систем, а затем реализовать их в виде компьютерных программ. Методы этого построения универсальны и зачастую независимы от тематики поставленной задачи и применимы во многих областях науки. Узкие специалисты в своих областях могут решить многие специфические проблемы, но, тем не менее, существуют и другие проблемы, которые требуют глубокой математической подготовки и использования универсальных методов математики, физики и компьютерного моделирования. Одним из таких универсальных методов является кибернетика и ее современная версия – *новая кибернетика*, кибернетика второго порядка.

Ключевой идеей в кибернетике была обратная связь [1]. Это замкнутый цикл – объект через посредство других объектов влияет на самого себя. В новой кибернетике [2-5] эта идея получила дальнейшее развитие – кибернетики (исследователи, наблюдатели) и кибернетика (исследуемая, наблюдаемая система) влияют друг на друга, образуя обратную связь. Новая кибернетика иначе известна как кибернетика кибернетики или кибернетика второго порядка. Хейнц фон Фёрстер видит первопричину появления кибернетики второго порядка в попытках классических кибернетиков сконструировать модель мозга [2]: «Мозг требует найти теорию мозга, которая для достижения полноты должна включать и создающего эту теорию. И даже более удивительно, создатель этой теории должен взять в расчет и самого себя при этом. Переходя в рамки кибернетики, это можно сформулировать так: кибернетик, рассматривая свою область должен принимать в расчет и свою собственную активность. Кибернетика в таком случае становится кибернетикой кибернетики или кибернетикой второго порядка».

Можно сформулировать набор основных черт новой кибернетики, связанных с этой идеей:

- Новый тип обратной связи [2]: Связь наблюдаемой системы с собой через наблюдателя. Это также и связь наблюдателя с собой через наблюдаемую систему. Часто это связь мала и ее можно пренебречь. Но во многих важных случаях это сделать нельзя. Это связано с неустойчивостью и хаотичностью многих реальных систем, обсуждаемой ниже
- Взаимовлияние и корреляции наблюдателя и наблюдаемой системы [2, 6]: изучаемую (наблюдаемую) систему нельзя рассматривать в отрыве от исследователя (наблюдателя). Существуют взаимодействия и корреляции между ними, которые часто необходимо учитывать, несмотря на их кажущуюся малость. Это связано с неустойчивостью и хаотичностью многих реальных систем, обсуждаемой ниже
- Относительность наблюдателей [6,7]: разные наблюдатели могут видеть наблюдаемую систему по-разному, однако это не ведет к противоречию.
- Невозможность полного самоописания [6]: система не может полностью описать и предсказать (понять) саму себя. На самом деле, пытаясь делать полное самописание системы, мы приходим к противоречию. Например, для описания системы мы используем чернила. Но при самоописании они тоже входят в систему. Т.е. их тоже надо описать. Для этого нужны другие чернила. Их тоже надо описать. И т.д. до бесконечности. Тем не менее, не полное, а частичное самописание возможно.
- Сложность [8]: реальные системы состоят из большого количества частей, которые взаимодействуют друг другом и окружением

- Неустойчивость и хаотичность [9]: многие реальные системы крайне чувствительны к слабым внешним воздействиям. Это приводит к важности учета даже малого взаимодействия наблюдателя с наблюдаемой системой.
- Эмерджентность [10]: у целого часто появляются новые свойства, которые нельзя предвидеть на основе даже полного знания частей и взаимодействия между ними (принципиальная эмерджентность). Действительно, мы знаем, что наблюдаемая система взаимодействует с наблюдателем. Это взаимодействие приводит к тому, что наблюдаемая система становится незамкнутой и подверженной непредсказуемому внешнему шуму. Следовательно, полная система должна включать наблюдателя. Если же мы включаем наблюдателя в систему, она становится непредсказуемой вследствие невозможности полного самоописания.
- Целостность и взаимосвязь (корреляции и взаимодействия) [6, 10, 11, 12]: наш мир это не набор случайных тел и событий, а довольно целостный набор взаимодействующих (пусть иногда и очень слабо) и коррелированных друг с другом объектов. Т.е. наш мир скорее напоминает взаимосвязанный организм, чем комплекс случайных событий (голографическая модель Вселенной или голографическая модель мозга, например)
- Научная непредсказуемость [6,10]: многие системы нельзя во всех деталях описать чисто логическими, научными методами, например математическими моделированием. Многие системы, хоть и непредсказуемы, но описываются вероятностными законами. Здесь идет речь о более сильной непредсказуемости, которая не описывается даже теорией вероятностей. Часто эта невозможность принципиальна, и не связана со сложностью системы. На самом деле, взаимодействие наблюдателя и наблюдаемой системы приводит к тому, что наблюдаемая система становится незамкнутой и подверженной непредсказуемому внешнему шуму. Если же мы включаем наблюдателя в систему, она становится непредсказуемой вследствие невозможности полного самоописания
- Главные параметры системы [6,13]: часто сложная система включает много параметров, описывающих ее. Однако во многих случаях динамику системы и ее взаимосвязи можно достаточно точно описать, используя лишь небольшое количество параметров. Они называются главными параметрами системы. Главные параметры – один из методов борьбы со сложностью и непредсказуемостью системы. Примерами главных параметров служат термодинамические переменные в физике, или «фичеры» - характерные черты многопиксельных объектов, используемые в теории распознавания образов.
- Интуиция (озарение) [6,14]: В психологии интуиция проявляется, когда решение проблемы приходит само быстро и внезапно. То, что невозможно достичь научными методами, можно достичь интуицией. Интуиция по определению не вытекает из науки. Интуиция – метод преодоления научной непредсказуемости системы *на практике*. На самом деле, наблюдатель взаимодействует с окружающим миром, коррелирован с ним и является его частью. Это позволяет ему, просто действуя по интуиции (из неких необъяснимых внутренних побуждений), достигать тех целей, которых невозможно в принципе достичь чисто научными методами.
- Междисциплинарность [15]: очень похожие свойства и явления происходят в системах, относящихся к совершенно разным областям знаний.

Проиллюстрируем основные черты новой кибернетики с помощью примеров из молекулярной физики. Почему именно молекулярной физики? В последние годы во всех промышленно развитых странах происходит развитие макро и нанотехнологий, и выход практической науки о материалах на атомный и молекулярный уровень. Идет активное исследование мезоскопических систем – промежуточных между микроскопическими и макроскопическими. Появляются такие устройства как квантовые и молекулярные

компьютеры. Также происходит развитие биотехнологий, моделирование химических и физических процессов в живых организмах, где процессы на микроуровне имеют сильную обратную связь с процессами на макроуровне. Высокомолекулярные соединения – наиболее часто встречающиеся мезосистемы.

В работе разрабатываются практические примеры новой кибернетики: учета влияния внешнего окружения/наблюдателя на мезосистемы. При анализе термодинамических парадоксов мезосистем широко подчеркивается важность учета даже малого шума при анализе их динамики.

Мы можем разбить проблему учета влияния внешнего окружения/наблюдателя на мезосистемы на следующие пять главных групп.

Первые две группы относятся к случаю, когда система еще достаточно простая и может быть рассчитана на основе известных законов физики элементов (подсистем), из которых эта система построена

Первая группа проблем - это проблемы, где влияние внешней среды/наблюдателя значительно, но его можно точно рассчитать на основе знания законов существующей физики. В качестве примера используется задача расчета влияния окружающей среды на энергию растворенных молекул и ее производных (градиентов) по координатам атомов (континуальная модель растворимости).

Мезосистемы и высокомолекулярные соединения не находятся, как правило в пустом пространстве, а окружены некоторой средой. Это молекулы в растворе, коллоидные частицы в жидкости или геле, наночастицы в нанокompозитных материалах [16, 17]. Поэтому необходимо уметь рассчитывать влияние окружающей среды. Для этого в данной работе анализируется широкий спектр континуальных моделей среды.

Континуальное означает, что окружающую среду мы рассматриваем как непрерывную диэлектрическую среду, а не состоящую из отдельных молекул. Рассчитанная энергия и ее градиенты могут быть использованы для расчета взаимодействия молекул. Этот расчет взаимодействия может использоваться при компьютерной разработке лекарств или расчете нанокompозитных материалов, квантовых спектров молекул и т.д. [16, 17]. Полученное влияние на основе точных законов физики влияние окружающей среды/наблюдателя описывается *идеальной динамикой*.

Для учета влияния внешней среды/наблюдателя, рассчитываемого на основе точных законов физики (идеальной динамики), разработана континуальная модель среды для полимерных молекул и их комплексов (нанокompозитные материалы, коллоидные частицы, растворы полимеров). Описание внешней среды оценивается с помощью оценки влияния этой среды на энергию растворенной молекулы. При этом та часть этой энергии, которая определяется короткодействующим взаимодействием и сосредоточена на поверхности молекул, оценивается очень просто введением поверхностного натяжения. Очень трудоемок (с точки зрения компьютерного времени, памяти и точности оценки) расчет дальнедействующих сил – электростатического взаимодействия зарядов молекул с окружающим раствором, который рассматривается как диэлектрик. При этом для упрощения расчета мы будем рассматривать этот диэлектрик не как набор молекул, а как непрерывную среду. Для учета взаимодействия с диэлектриком мы должны разделить пространство на область, занимаемой молекулами, и область, занимаемая средой. Эти области могут быть разделены четкой резкой границей или плавным переходом. Были разработаны оба метода:

- 1) Был развит алгоритм построения гладкой поверхности раздела с помощью усовершенствованной программы TAGSS. Этот метод позволяет получить поверхность раздела путем первичной обкатки атомов растворенных молекул сферой с радиусом, равным среднему радиусу молекулы растворителя; а затем устранения сингулярностей полученной поверхности путем вторичной обкатки внутренней части этой поверхности сферой малого переменного радиуса. Был разработан адаптивный

механизм для триангуляции как уменьшением поверхностных элементов (в случае, например, поверхности с мелкими деталями), так и метод расчета укрупненных поверхностных элементов методом COSMO (аналогичный разработанному ранее методу для PCM). Был проведен на этой основе расчет энергии сольватации с помощью методов COSMO, SGB, PCM. Для расчета конфигурации растворенных молекул или взаимодействия растворенных молекул нужно найти минимум энергии растворенных молекул (включающей энергию растворения), сдвигая атомы, входящие в молекулы. Для поиска этого минимума энергии необходимо знать производные (градиенты) от энергии растворения по положению атомов. Численное нахождение этих градиентов очень трудоемко (с точки зрения компьютерного времени, памяти и точности оценки). Поэтому огромное значение имеет возможность найти такие градиенты аналитически. Получение разделяющей поверхности методом обкатки сферами позволяет аналитически рассчитать градиенты параметров поверхностных элементов при сдвиге атомов, а из этого и получить аналитически градиенты энергии. Была написана и запатентована программа DISOLV, проводящая расчеты энергии растворения и ее аналитические градиенты.

- 2) Был разработан метод расчета электростатической энергии взаимодействия для метода плавного перехода между молекулами растворителя и раствором. Получена оценка энергии растворения методом итераций.

Вторая группа проблем относится к случаю, когда динамика системы хоть и меняется при наличии шума, но не зависит значительно от амплитуды шума для широкого интервала значений этих амплитуд, когда шум становится выше некоторой хоть малой, но конечной пороговой величины. Динамика системы скорее определяется характерными макропараметрами системы. Мы иллюстрируем это важное свойство, часто встречающееся в мезосистемах, на конкретных примерах уравнений, описывающих динамику фронта горения. А именно, это задача движения фронта пламени в предварительно подготовленной смеси газов и задача фильтрационного горения. При этом динамика описывается не полным набором переменных системы (микропараметрами), а их макропараметрами. Это небольшое число переменных, являющимися функциями микропараметров. Поведение этих макропараметров устойчиво к малому шуму и описывается законами *наблюдаемой динамики, полученной на основе идеальной динамики*.

Для иллюстрации второй группы проблем используется задача распространении фронта пламени в теории горения [18]. При этом рассматриваются два случая.

В первом случае предварительно смешиваются реагирующие газы, а затем поджигается экзотермическая химическая реакция между ними. При этом выводятся упрощенные уравнения движения фронта пламени из уравнений Навье-Стокса, уравнений термодинамики и химических уравнений. Мы рассматриваем две геометрии распространения пламени – геометрия в канале, описываемая уравнением Михельсона-Сивашинского; радиальная геометрия, описываемая уравнением Сивашинского-Филянда-Франкеля.

Во втором случае мы рассматриваем фильтрационное горение – горение пористой среды при движении реагирующего с ней газа в этой пористой среде. При этом, при поджигании, газ взаимодействует с пористой средой, создавая пламя. Движение фронта пламени описывается уравнением лапласовского роста. В работе изучается лапласовский рост в канале.

Таким же уравнением описывается движение поверхности раздела двух жидкостей в канале, когда менее вязкая жидкость толкает более вязкую. Отличие между этими двумя примерами лапласовского роста заключается в типе регуляризации движения, предотвращающего образование сингулярностей (бесконечной кривизны): для движения поверхности раздела жидкостей - это поверхностное натяжение; для фильтрационного горения – это диффузия газа.

Имеется замечательный результат – все описанные выше уравнения для движения фронта имеют точные решения в виде суммы функций, которые имеют постоянное и конечное число полюсов в комплексной плоскости. Вся динамика этих решений описывается движением полюсов в комплексной плоскости. Таким образом, вместо сложной системы нелинейных интегро-дифференциальных уравнений для описания движения фронта пламени, мы получаем конечное число обыкновенных дифференциальных уравнений для полюсов в комплексной плоскости. Любое начальное состояние можно представить с любой степенью точности в виде суммы конечного числа полюсных функций. При этом набор всех полюсных функций является переполненным.

- 1) В теории горения предварительно подготовленной смеси газов детально проанализированы полюсные решения в канале и при радиальном распространении фронта пламени. Полюса имеют ясный физический смысл – это изломы на фронте пламени. Глубина излома определяется близостью полюса к действительной оси, а положение излома вдоль канала (вдоль его конечной ширины) или вдоль угловой координаты для радиального случая совпадает с положением полюса на действительной оси. В случае периодических граничных условий для достаточно большого числа полюсов в канале существует устойчивое решение в виде гигантского излома, состоящего из полюсов, число которых пропорционально ширине канала. Этот гигантский излом может быть расположен в любой точке вдоль горизонтальной оси канала и может дрейфовать вдоль оси канала из-за шума. В случае более физически верных условий непротекания мы получаем гигантские половины изломов, примыкающие к стенкам канала, высота которых будет флюктуировать из-за шума, что мы и видим в эксперименте. В радиальном случае, из-за постоянного расширения фронта, такого устойчивого решения нет. Найденные аналитические полюсные решения обладают рядом свойств. Это постоянная скорость волнового фронта, независимая от ширины канала или среднего радиуса фронта в радиальном случае; постоянное число изломов для радиального случая или один гигантский излом в канале. Однако эти свойства не наблюдаются в физическом эксперименте и численном моделировании уравнений. В физическом эксперименте и численном моделировании мы получаем рост скорости фронта горения (ускорение) при увеличении ширины канала или радиальном расширении пламени, хотя есть некоторые свидетельства возможного насыщения этой скорости. При распространении фронта пламени в канале у основания гигантского излома (в точке наиболее удаленной от этого излома) появляются постоянно маленькие изломы, которые затем бегут паровозиком к гигантскому излому и исчезают на нем. Для радиального пламени постоянно рождаются новые изломы на фронте пламени, поддерживая ячеистую структуру пламени. Однако введение в уравнение малого шума приводит к появлению источника полюсов. На основании такого источника полюсов объяснено ускорение фронта пламени в случае газовой смеси при радиальной и сферической геометрии объяснено увеличение скорости фронта при увеличении ширины прямоугольного канала; объяснена ячеистая структура фронта пламени. При этом мы можем определить ряд естественных макропараметров, описывающих поведение системы (например, степень, задающая зависимость скорости от времени; среднее число полюсов, появляющихся в единицу времени и т.д.). Причем динамика этих макропараметров хоть и определяется шумом, но почти не зависит от амплитуды и вида этого шума, когда этот шум выше некоторого малого порога и в широком интервале больших шумов. В работе описывается и исследуется эта динамика макропараметров.
- 2) Фильтрационное горение в прямоугольном канале описывается уравнением лапласовского роста. Это решение имеет полюсные решения. Более того, обыкновенные дифференциальные уравнения для полюсов можно аналитически

проинтегрировать – найти интегралы движения. Полюса в случае фильтрационного горения имеют ясный физический смысл – это каналы, образуемые участками фронта пламени. При этом константы, входящие в интегралы движения для полюсов описывают ширину, направление и координаты основания для этих каналов. В результате слияния двух полюсов образуется канал с шириной, равной сумме слившихся каналов. Для общего случая начальных условий каналов произвольной ширины, наклона и для периодических граничных условий, слияние всех каналов на фронте пламени дает один единый канал. (Исключение возможны для вырожденных случаев существования групп полюсов с нулевым наклоном каналов. Образующиеся из групп полюсов каналы будут параллельны и не смогут пересечься и дать единый канал.) Для более физически верных условий непротекания мы получаем две половинки канала на стенках полного канала. Ширина этого канала (полуканалов для условий непротекания) может быть любой – от нулевой до ширины прямоугольного канала и будет меняться при наличии шума. К сожалению, ситуация может быть и иной – в результате движения полюсов за конечное время могут появляться сингулярности с бесконечной кривизной, после чего описание движения полюсами становится невозможным. Для предотвращения появления таких сингулярностей (для регуляризации) обычно вводят поверхностное натяжение. Введение такой регуляризации приводит к серьезной проблеме, поскольку делает полюсное решение невозможным. Можно было бы попытаться ввести простую регуляризацию, просто отбрасывая начальные условия с полюсами, приводящие к сингулярностям. Простая регуляризация имеет свои проблемы. Действительно, введение регуляризации поверхностным натяжением приводит к результатам, наблюдаемым в эксперименте и численном моделировании (при наличии поверхностного натяжения). Это следующие результаты: ширина образующегося в асимптотике единого канала равна половине ширине прямоугольного канала; появляются новые полюса даже в случае решений, не приводящих к сингулярностям. Однако эти результаты не наблюдаются для чистых полюсных решений с простой регуляризацией. Такой шум стирает разницу между двумя типами регуляризации и приводит к результатам, совпадающим с результатами, наблюдаемым в эксперименте (при наличии поверхностного натяжения) и численном моделировании. При этом шум позволяет сохранить аналитический полюсный вид решений. Он приводит к появлению новых полюсов (каналов) даже для случая решений без сингулярностей. Появление новых полюсов (шума для поверхностного натяжения) приводит к флуктуациям ширины единого канала во времени для обеих регуляризаций. Кроме того для случая простой регуляризации минимальный шанс на появления сингулярностей имеется в случае, когда асимптотическая ширина канала близка половине ширины полного канала (когда ширина единого асимптотического канала близка к нулю или ширине полного канала сингулярности появляются для почти 100% случаев и будут отбрасываться простой регуляризацией). Поэтому ширина канала для простой регуляризации (как и для поверхностного натяжения) будет флуктуировать около половины величины прямоугольного канала. Возможность получения физически обоснованного полюсного решения при введении внешнего шума и простой регуляризации обсуждается в этой работе.

Другие две группы проблем относятся к случаю, когда система слишком сложна для этого и подобное аналитическое описание невозможно из-за ее сложности или *в принципе*. Это и есть, так называемое, явление *эмерджентности*.

Третья группа проблем относится к случаю, когда учет влияния окружения/наблюдателя на взаимодействие молекул крайне сложен для расчета на базе основных законов физики ее подсистем. Поэтому разрабатываются феноменологические и статистические методы расчета взаимодействия на основе модели «черного ящика». Мы не можем смоделировать поведение ящика на основе изучения его внутренней структуры,

а получаем его модель на основе изучения связи его входа и выхода. В молекулярной физике это QSAR и байесовские модели. Мы приводим анализ причин оптимальности наивной байесовской модели [17], используемой при поиске количественных соотношений структура-свойство в молекулярной физике (Quantitative Structure-Activity Relationship (QSAR)). Подобная система также определяется малым числом макропараметров (входы-выходы системы) и описывается законами *наблюдаемой динамики, однако полученной на основе экспериментального наблюдения, а не идеальной динамики*. При этом эксперименты могут быть как реальными в пробирке, так и численными на компьютере.

Для анализа описанных выше трех групп проблем широко использовались сложные методы математического и компьютерного моделирования.

Четвертая группа проблем относится к случаю, когда описанные эти методы математического и компьютерного моделирования учета внешнего шума не работают из-за слишком высокой сложности системы или *принципиальной невозможности* их использования [19]. Такая принципиальная невозможность возникает, когда мы делаем попытку учесть влияние окружения/наблюдателя путем включения в изучаемую и описываемую систему этого окружения/наблюдателя. Однако такую систему, включающую как наблюдаемую систему, так и наблюдателя с окружением принципиально невозможно наблюдать и рассчитывать во всей ее полноте, если рассматривать все ее микропараметры. На самом деле для такого расчета и описания мы вынуждены использовать ее же параметры (точнее параметры наблюдателя и его «записи»), т.е. *часть* параметров системы должны описать *все* ее поведение. Ранее в *наблюдаемой динамике* такое описание было возможно за счет использования небольшого числа макропараметров. Но могут существовать сложные системы, где выделение такого малого набора макропараметров будет невозможно. Динамику таких систем мы будем называть *непредсказуемой динамикой*. Однако это не означает, что наблюдатель бессилён в этой ситуации. Наблюдатель является органичной частью таких систем, также как шестеренка является неотъемлемой частью часового механизма. Поэтому опираясь на свою *интуицию (а не научный анализ)*, наблюдатель может принимать правильные для достижения его целей решения. Так же шестеренка часов, двигаясь по своим внутренним законам, обеспечивает слаженную работу часов, хоть и не понимает и не может проанализировать свою роль в обеспечении этой слаженной работы часового механизма.

Мы даем примеры непредсказуемых систем, такие как квантовые, фрактальные и молекулярные компьютеры. Рассмотрено практическое применение непредсказуемых систем в области искусственного интеллекта.

Пятая группа проблем относится к базовым парадоксам физики мезосистем [19]. На самом деле, указанные выше принципы новой кибернетики важны не только для решения конкретных физических задач, но и для принципиальных базовых принципов физики. Для мезоскопических систем уже нельзя игнорировать противоречия и парадоксы между обратимыми микроскопическими уравнениями физики и термодинамической физикой макроскопических систем. К этим парадоксам относятся парадокс между необратимостью роста энтропии и обратимостью уравнений динамики, противоречие между наблюдаемой редукцией волнового пакета и уравнениями квантовой механики, а также их аналоги в квантовой и релятивистской теории гравитации. Ранее, когда мы имели дело с чисто микроскопическими или макроскопическими системами эти парадоксы можно было просто игнорировать, относя их к «философии». Сейчас, когда активно разрабатывается физика мезоскопических систем, которая имеет не только теоретическое, но и большее практическое значение, игнорирование этих парадоксов приводит к непониманию фундаментальных процессов протекающих в мезосистемах и грубейшим ошибкам, наносящим реальный практический вред. Поэтому последняя глава данной диссертации посвящена анализу и разрешению этих парадоксов.

Особенно замечательно, что все эти базовые физические парадоксы разрешаются на основе тех же единых принципов новой кибернетики, которые были использованы и для разрешения проблем конкретных динамических систем, описанных выше. Первый метод опирается на влияние внешнего наблюдателя или окружающей среды, которое приводит к разрушению корреляций в системе и приводит к синхронизации стрел времени. Второе приближение основывается на ограничениях в возможности самопознания и самоописания системы, когда мы включаем как наблюдателя, так и окружение в рассматриваемую систему.

Цели диссертационной работы

Целью диссертации является разработка теоретических и компьютерных моделей мезосистем методами новой кибернетики для учета внешнего влияния/шума со стороны внешней среды/наблюдателя, в том числе теория растворимости молекулярных соединений, теория горения, понимание и разрешение базовых парадоксов физики. С помощью принципов новой кибернетики анализируются проблемы молекулярной физики, требующие применения сложных математических методов и расчетов, связанных с учетом влияния окружения на молекулярные системы.

Положения, выносимые на защиту

На защиту выносятся результаты, полученные для молекулярных/мезосистем с помощью методов новой кибернетики, основанных на учете влияния внешнего шума/воздействия со стороны внешнего окружения/наблюдателя:

- 1) Для систем, рассчитываемые на основе известной физики подсистем разработаны следующие практические применения указанной теории: решены математические и компьютерные модели с высокой универсальностью, позволившие решить следующие задачи:
 - (a) Для первой группы проблем, (где учет влияния внешней среды/наблюдателя, рассчитывается на основе точных законов физики - идеальной динамики) разработана континуальная модель среды для полимерных молекул и их комплексов (нанокомпозитные материалы, коллоидные частицы, растворы полимеров): Был развит существующий алгоритм построения гладкой поверхности и усовершенствована программа TAGSS; разработан адаптивный механизм для триангуляции; доведены до конца методы расчета аналитических градиентов параметров поверхностных элементов при сдвиге атомов; проведен на этой основе расчет энергии сольватации и аналитических градиентов энергии сольватации при сдвиге атомов в методах COSMO, SGB, PCM; написана и запатентована программа DISOLV, проводящая эти расчеты; проведен расчет диагональных элементов в методе COSMO; доведена до конца точная аналитическая интегральная модель растворителя
 - (b) Для второй группы проблем (где учет влияния внешнего шума/наблюдателя осуществляется на основе наблюдаемой динамики макропараметров, рассчитываемой из идеальной динамики и независимой от амплитуды шума) детально проанализированы полюсные решения в теории горения смеси газов в прямоугольном канале; проанализировано влияние малых шумов на распространение фронта горения и на его основании объяснено ускорение фронта пламени в случае газовой смеси для радиальной и сферической геометрии; объяснено увеличение скорости фронта при увеличении ширины

прямоугольного канала; объяснена ячеистая структура фронта пламени; показаны естественные методы регуляризации полюсного решения в теории фильтрационного горения, позволяющие сохранить полученное точное аналитическое полюсное решение, объяснена асимптотика решения в виде одного пальца, равного половине ширины канала.

- 2) Для эмерджентных систем, т.е. систем, не рассчитываемых аналитически на основе известной физики подсистем, получены следующие результаты:
 - (a) Для третьей группы проблем (где учет внешнего шума/наблюдателя за счет модели «черного ящика» и построение его наблюдаемой динамики происходит из реального/численного эксперимента) доказана оптимальность наивной байесовской модели, используемой при поиске количественных соотношений структура-свойство в молекулярной физике (Quantitative Structure-Activity Relationship (QSAR))
 - (b) Для четвертой группы проблем непредсказуемой динамики, где учет внешнего шума/наблюдателя происходит за счет интуиции наблюдателя, определены границы применимости научных методов предсказания систем и роль интуиции наблюдателя для непредсказуемых систем; приведены примеры непредсказуемых систем
- 3) Развита основа термодинамики, позволяющие решить базовые парадоксы современной физики на базе новой кибернетики, учитывающей влияния внешнего шума со стороны внешней среды/наблюдателя:
 - (a) Противоречие между ростом энтропии и обратимым характером физических законов (парадоксы Пуанкаре и Лошмидта)
 - (b) Парадокс Шредингеровского Кота (или парадокс редукции волнового пакета)
 - (c) Информационный парадокс черных дыр
 - (d) Парадокс дедушки для временных кротовых нор

Научная новизна

В работе даны оригинальные и сложные методы решения задач молекулярной физики, связанные с учетом влияния окружения на молекулярные системы.

Развита оригинальные полуаналитические модели и компьютерные программы для решения этих задач. Указанные модели позволяют глубоко понять свойства задач и получить новые интересные результаты:

- 1) Разработанные новые оригинальные модели расчета энергии растворимости молекул и ее аналитических градиентов
- 2) Проанализировано влияние малых шумов на распространение фронта горения – ускорение фронта пламени в случае газовой смеси объяснено увеличением скорости фронта при увеличении ширины прямоугольного канала; объяснена ячеистая структура фронта пламени; для фильтрационного горения объяснена асимптотика решения в виде одного пальца, равного половине ширины прямоугольного канала
- 3) Доказана оптимальность наивной байесовской модели, используемой при поиске количественных соотношений структура-свойство в молекулярной физике (Quantitative Structure-Activity Relationship (QSAR)).
- 4) Определены границы применимости научных методов предсказания систем и роль интуиции для непредсказуемых систем. Приведены примеры непредсказуемых систем
- 5) В работе даны оригинальные методы решения парадоксов, возникающих из-за разницы между макроскопическим и микроскопическим описанием мира. Эти

методы важны для глубокого понимания физики мезосистем, в том числе и высокомолекулярных соединений.

Практическая ценность

Указанные методы, модели и алгоритмы могут быть использованы при создании нанокompозитных материалов [16], при компьютерной разработке лекарств [17], при расчете спектров хемосенсоров, создании квантовых и молекулярных компьютеров, квантовых и молекулярных машин [19], разработке химических реакторов, основанных на горении [18].

Апробация

Основные положения диссертации докладывались на следующих семинарах и конференциях:

- 1) Семинар Теоретического отдела Объединенного института высоких температур РАН 19 ноября 2015 года
<http://www.youtube.com/watch?v=yEFsE4DBTzw>
- 2) Семинар имени К.И. Бабенко в Институте прикладной математики им. Келдыша 14 мая 2015 года
- 3) Четыре семинара в МГУ в 2014-2015 годах, Механико-математический факультет (ММФ МГУ), 1-н на кафедре волновой и газовой динамики и 3-ри на кафедре дифференциальных уравнений;
- 4) The 17th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2013, Florida, USA;
- 5) Доклад на «Российском междисциплинарном семинаре по темпорологии», Институт исследования природы времени 19 марта 2013, МГУ
<http://www.chronos.msu.ru/old/seminar/rautumn12.pdf>
- 6) 2 семинара «Disolv - моделирование растворимости молекул», «Основные парадоксы статистической классической физики и квантовой механики» группы «Молекулярное моделирование и разработка биологически активных соединений» в НИВЦ МГУ, 2010;
- 7) 2 семинара: «Континуальные модели растворителя. Подсчет свободной энергии растворения»; «Конструктор супрамолекул: инструмент для многомасштабного моделирования наноструктур с заданными свойствами» на II Всероссийской конференции «Многомасштабное моделирование процессов и структур в нанотехнологиях». 27–29 мая 2009. М.: МИФИ, 2009;
<http://www.twirpx.com/file/537141/>
- 8) "What is life?", January 4th 2007 The Israel Society for Astrobiology and the Study of the Origin of Life, ILASOL Conference 2006 in Weizmann Institute of Science
- 9) 82nd Statistical Mechanics Meeting, Department of Mathematics, Rutgers University, December 10-12, 1999;

Публикации

По теме диссертационной работы опубликованы 18 печатных работ в изданиях из перечня ВАК, получен один патент (программа ЭВМ DISOLV), выпущены три монографии.

Личный вклад автора

Алгоритмы и программа DISOLV для расчета энергии взаимодействия со средой и ее аналитических производных была на 60% сделана автором. При этом 90% этого вклада связано с расчетом именно энергий и их производных и только 10% с расчетом самой гладкой поверхности вокруг молекулы.

Работа по усовершенствованию метода COSMO на 90% сделана автором.

Итерационный метод для интегральной модели среды на 50% сделан автором.

Вклад автора в модели теории горения составляет 60%.

Вклад автора в теорию оптимальности наивной байесовской модели составляет 90%.

Описанная теория разрешения парадоксов физики мезосистем развита на 90% автором.

Объем и структура работы

Работа состоит из трех глав, посвященных получению новых результатов в молекулярной физике с помощью методов новой кибернетики, связанных с учетом влияния окружения на молекулярные системы и разрешению физических парадоксов.

В *Главе 1* рассмотрены системы, рассчитываемые на основе известной физики подсистем (идеальной динамики).

В *Главе 1.1* мы иллюстрируем первую группу проблем, где учет влияния внешней среды/наблюдателя, рассчитывается на основе точных законов физики (идеальной динамики) на примере континуальной модели среды для теории полимерных молекул и их кластеров.

В *Главе 1.2* подчеркивается влияние даже малого шума на динамику физической системы. Мы анализируем влияния малых шумов на динамику в теории горения для смеси газов и фильтрационного горения.

В *Главе 2* мы рассматриваем эмерджентные системы, т.е. системы, не рассчитываемые аналитически на основе известной физики подсистем.

В *Главе 2.1* мы рассматриваем третью группу проблем, где учет внешнего шума/наблюдателя происходит за счет модели «черного ящика». В этой главе доказана оптимальность наивной байесовской модели, используемой при поиске количественных соотношений структура-свойство в молекулярной физике (Quantitative Structure-Activity Relationship (QSAR)).

В *Главе 2.2* для четвертой группы проблем непредсказуемой динамики, где учет внешнего шума/наблюдателя происходит за счет интуиции наблюдателя, определены границы применимости научных методов предсказания систем и роль интуиции наблюдателя для непредсказуемых систем. Приведены примеры непредсказуемых систем. Рассмотрено практическое применение непредсказуемых систем в области искусственного интеллекта.

В *Главе 3* предложены методы решения основных парадоксов физики на основе методов новой кибернетики (учета влияния внешнего шума от окружения/наблюдателя) – парадокс роста энтропии, парадокс редукции волнового пакета в квантовой механике, информационный парадокс черных дыр и парадокс дедушки для временных кротовых нор.

Глава 1. Системы, рассчитываемые на основе известной физики подсистем

Для систем, рассчитываемые на основе известной физики подсистем разработаны следующие практические применения указанной теории: решены математические и компьютерные модели с высокой универсальностью, позволившие решить задачи учета влияния внешней среды/наблюдателя.

Глава 1.1 Первая группа проблем, где учет влияния внешней среды/наблюдателя, рассчитывается на основе точных законов физики – идеальной динамики

В этой главе мы иллюстрируем первую группу проблем, где учет влияния внешней среды/наблюдателя, рассчитывается на основе точных законов физики (идеальной динамики) на примере континуальной модели среды для теории полимерных молекул и их кластеров. Была разработана континуальная модель среды для полимерных молекул и их комплексов (нанокompозитные материалы, коллоидные частицы, растворы полимеров) [20]: был развит существующий алгоритм построения гладкой поверхности и усовершенствована программа TAGSS; разработан адаптивный механизм для триангуляции; доведены до конца методы расчета аналитических градиентов параметров поверхностных элементов при сдвиге атомов; проведен на этой основе расчет энергии сольватации и аналитических градиентов энергии сольватации при сдвиге атомов в методах COSMO, SGB, PCM; написана и запатентована программа DISOLV, проводящая эти расчеты; проведен расчет диагональных элементов в методе COSMO; доведена до конца точная аналитическая интегральная модель растворителя

Глава 1.1.1 Континуальная модель среды I: Алгоритм для построения гладкой молекулярной поверхности

В данной работе представлен полный и исчерпывающий алгоритм построения гладкой молекулярной поверхности исключенного объема растворителя - SES, а также поверхности, доступной растворителю – SAS [20, 21]. Эти поверхности играют роль границы между областями молекулы и растворителя. Основой алгоритма является первичная и вторичная обкатки молекул. Оригинальность данной работы состоит в создании *полного и уточненного* алгоритма вторичной обкатки, который позволяет создать оптимально гладкую поверхность SES любой молекулы или системы из молекул, обкатывая любые нерегулярности и близкие к ним ситуации, возникающие при первичной обкатке. Для этого используется адаптивное критическое расстояние, характеризующее максимально допустимую нерегулярность поверхности. Главная задача, которую будет решать полученная поверхность и которая будет рассмотрена в дальнейших частях - это расчет энергии сольватации и ее градиентов для континуальных моделей растворителя. Также она может быть использована для демонстрационных целей в молекулярных редакторах.

1.1.1.1 Введение

Цель данной части – дать полное и исчерпывающее описание алгоритма, позволяющего построить оптимально гладкую молекулярную поверхность методами первичной и вторичной обкатки. Мы сможем использовать этот алгоритм далее (а) в молекулярных редакторах для демонстрационных целей или (б) для получения энергии сольватации молекулы (разница между значениями свободной энергии молекулы в растворе и вакууме) и аналитических градиентов этой энергии. Мы реализуем эту гладкость с помощью первичной обкатки молекулы сферами (внешнюю поверхность

молекулы сферами с радиусом равным размеру молекулы растворителя) и вторичной обкатки (внутреннюю поверхность сферой с переменным радиусом, устраняя все оставшиеся нерегулярности).

Нужно подчеркнуть особую важность получения гладкой поверхности для этих целей. Действительно, в случае молекулярного редактора гладкость поверхности необходима для её триангуляции и дальнейшего наглядного изображения, при котором нас не будут отвлекать не имеющие физического смысла нерегулярности.

В случае определения электростатической составляющей энергии сольватации и ее аналитических производных гладкость поверхности является необходимым условием устойчивости алгоритма – фиктивные нерегулярности поверхности приводят к накоплению на них фиктивного заряда, что приводит к неустойчивости алгоритма и большим численным ошибкам [22-24].

Поэтому получение сглаженной поверхности – не только интересная математическая задача, но и практически важная проблема. На основе изложенного в данной работе алгоритма первичной и/или последующей вторичной обкатки были созданы программы PQMS [25], MSMS [26], программа Тотрова и Абагына [27], SIMS [28], TAGSS [29-32] и её усовершенствованная версия, вошедшая как часть в программу DISOLV [20,31-32]. На основе этих программ и алгоритмов были созданы успешно работающий молекулярный редактор [32] и выдан патент на программу для расчета энергии сольватации и её производных [33].

В чем заключается новизна данной работы по отношению к другим работам [25,26,27,29,30,34-38], анализирующим первичную и/или вторичную обкатку молекулярной поверхности? Это, во-первых, полнота – рассмотрены все возможные случаи нерегулярности и показана возможность их сглаживания. Во-вторых, мы стремимся провести сглаживание оптимально. Это означает, что мы стремимся получить не только гладкость, но и в максимальной возможной мере избежать появления участков хоть и уже гладких, но, тем не менее, слишком близких к нерегулярностям (т.е. избегать появления участков с очень малым радиусом кривизны и очень узких «каналов»). Насколько мы знаем, полностью и исчерпывающее эта задача решена лишь в [21].

Однако следует отметить, что без дополнительных ограничений задача построения гладкой поверхности может быть решена тривиально - просто описанием сферы вокруг молекулы. Каковы же эти ограничения?

Во-первых, радиусы атомов это хорошо определенные величины, которые не могут произвольно меняться. В случае использования данного алгоритма для целей изображения молекулы в молекулярном редакторе, радиусы их атомов определяются размерами электронных оболочек. (Это, так называемые, радиусы Ван-дер-Ваальса [39,40]. Для радиусов Ван-дер-Ваальса в литературе имеется различные наборы величин [39-41]). В случае использования их для вычисления энергии сольватации, начальные («огрубленные») значения радиусов задаются как и в первом случае, но затем уточняются таким образом, чтобы вычисленные энергии сольватации молекул соответствовали их экспериментальным значениям.

Во-вторых, если молекула находится в растворе, то радиус молекул растворителя тоже хорошо определенная величина, которая не может произвольно меняться. Этот радиус безусловно влияет на поверхность раздела между раствором и самой молекулой. Подбираются этого радиус из тех же описанных выше соображений, что и радиус атомов молекул.

Заметим, что указанные выше радиусы достаточно сильно ограничивают возможность варьирования поверхности молекулы. Но, тем не менее, они оставляют достаточно возможностей для того, чтобы сделать её гладкой.

Причем мы хотим сделать её не просто гладкой, но и оптимально гладкой. Под оптимальностью в данном случае понимается, что мы (а) ищем не только гладкую

поверхность, но и стремимся уменьшить ее кривизну, не теряя детали поверхности, а также (b) стремимся увеличить декартово расстояние между *несоседними участками поверхности*. Это участки, близкие пространственно, но удаленные при измерении расстояния между ними вдоль поверхности. Для оценки степени гладкости задается *максимальное критическое расстояние*, на которое алгоритмом разрешается приблизиться несоседним участкам поверхности. Если это расстояние сближения меньше критического, возникает ситуация близкая к нерегулярности, которая требует вторичной обкатки.

Существует два типа поверхности, окружающей молекулу [42]:

SAS (Solvent Accessible Surface) - поверхность доступная растворителю образуется центрами молекул растворителя, касающихся молекулы субстрата. Число молекул растворителя, касающихся поверхности молекулы пропорционально площади SAS.

SES (Solvent Excluded Surface) - поверхность исключённого из растворителя объёма. Объем, занимаемый растворителем, лежит *вне* объема, ограниченного этой поверхностью. Сам субстрат полностью лежит *внутри* этого объема.

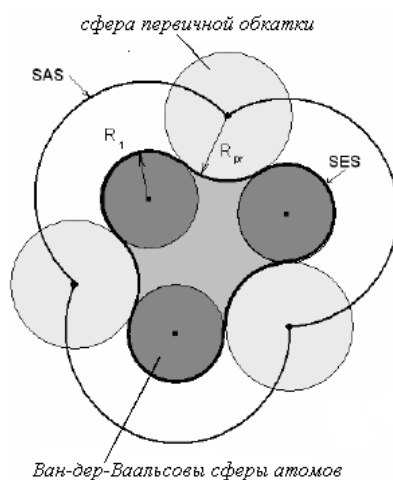


Рис. 1. Первичная обкатка поверхности. См. [20].

Поверхность SAS можно получить, обкатывая молекулу субстрата молекулой растворителя и отмечая положения ее центра. Обкатка – это движение молекулы растворителя вдоль поверхности субстрата, при котором она последовательно касается всех доступных ей точек субстрата. (Рис. 1) Для упрощения молекулу растворителя можно заменить сферой обкатки (сфера, описанная вокруг молекулы растворителя). [43,44].

Поверхность SES молекулы можно описывать следующим образом [45]:

- (1) сглажено, заменяя ее простыми формами типа [46-52]
 - (a) сфера
 - (b) эллипсоид
 - (c) цилиндр
- (2) детально, давая все изгибы поверхности молекулы
 - (a) покрытием из Ван-дер-ваальсовых сфер вокруг атомов
 - (b) покрытием из сфер вокруг химических групп атомов
 - (c) как в предыдущих двух методах, но заполняя остающееся пустое пространство внутри SES фиктивными сферами (это было реализовано в программе GEOPOL) [34-37]
 - (d) Поверхностью уровня для электронной плотности молекулы [53], определяемой через квантовую механику, или иные типы функций, используемые для построения поверхности уровня [54-58]. Данный метод сталкивается с серьезными трудностями:

поверхность уровня может быть также не гладкой; она задаётся не явно, что затрудняет ее триангуляцию; труден подбор функций, дающий поверхность близкую к реальной и определяемой радиусами Ван-де-Ваальса атомов.

(е) соединяя сферы, описанные в (а) или (б) участкам вогнутых и выпуклых поверхностей [27,38].

Наиболее гладкая и реалистичная поверхность получается методом (2)(е), который и рассматривается в этой части. Получить SES этим методом можно так же как и SAS, обкатывая внешнюю сторону молекулы сферой и беря (а) положения точек касания с атомами (б) отрезки геодезических линий сферы обкатки, проходящими через две точки касания сферы обкатки и атомов (в) сегменты нижней части обкатывающей сферы, ограниченные отрезками геодезических линий сферы обкатки, проходящими через точки касания сферы обкатки и атомов (*первичная обкатка*, Рис. 1). Такое определение поверхности SES было впервые предложено в работе [59]. Предложенный метод получил развитие в работах [27], [60-64]. Однако полученная таким образом поверхность SES может оказаться негладкой [26]. Для дальнейшего сглаживания можно обкатать внутреннюю поверхность молекулы вторично (*вторичная обкатка*, Рис. 2 и 3), что предложено впервые в [38].

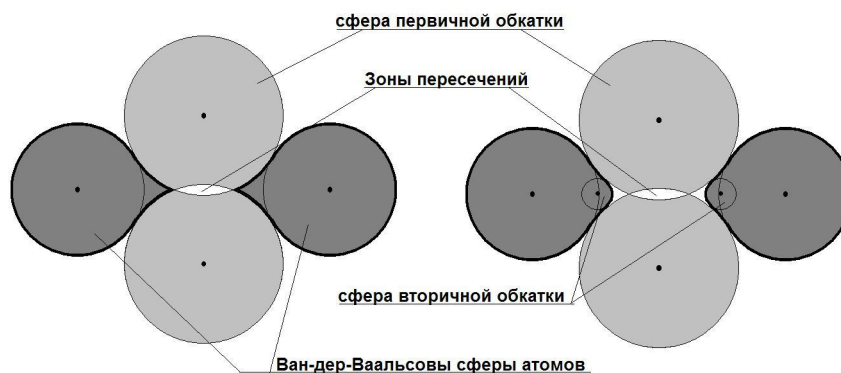


Рис. 2 Вторичная обкатка поверхности. См.[20].

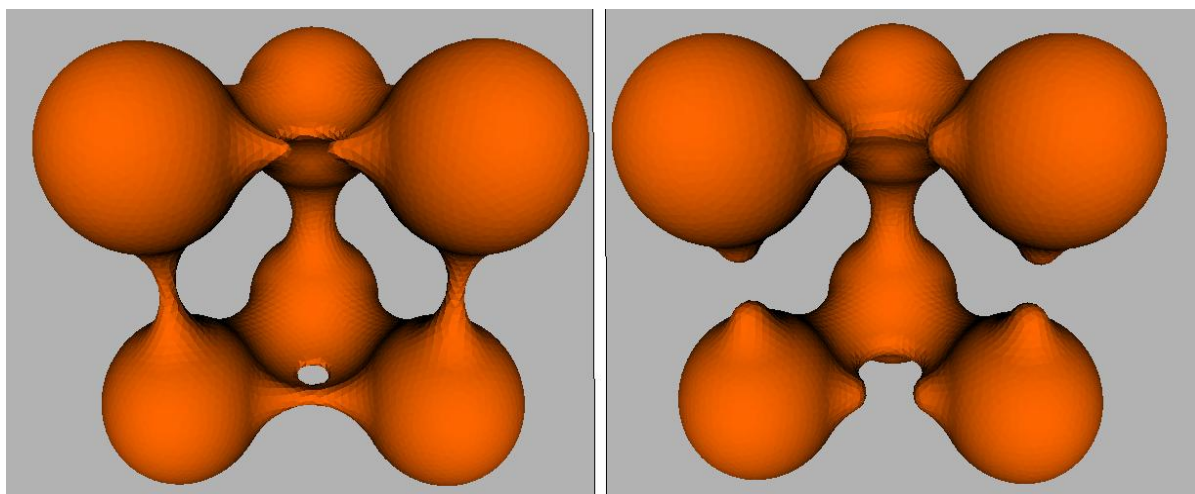


Рис. 3 Применение метода вторичной обкатки для геометрической конфигурации многих атомов.

Может возникнуть вопрос – почему мы уверены, что вторичная обкатка успешно сгладит все нерегулярности и не потребуются третичная и т.д. обкатки? Аргумент здесь

следующий: мы не имели возможность менять радиусы атомов и сферы первичной обкатки. Это заранее заданные фиксированные величины. Радиус же вторичной обкатки – произвольно выбираемая величина, причем даже для обкатки разных участков одной и той же поверхности он может быть выбран разным. Интуитивно ясно, что выбрав радиус вторичной обкатки сколь угодно малым, мы способны «сгладить» все, что угодно. Действительно, в пределе бесконечно малого радиуса, нет проблем для гладкой обкатки любой нерегулярности. Однако мы стремимся не к просто получению гладкой, но оптимально гладкой поверхности. Т.е. она не должна включать элементов хоть и гладких, но близких к нерегулярным. Поэтому следует добиваться вторичной обкатки с радиусом максимально возможным, исходя из геометрических ограничений поверхности. Это делает алгоритм вторичной обкатки достаточно нетривиальным.

Поверхность обкатки состоит из двух типов поверхностных сегментов – сферических и тороидальных. Они подразделяются на фрагменты пяти подтипов. Это сферические элементы ван-дер-ваальсовых сфер атомов, вогнутые сферические элементы первичной обкатки, тороидальные фрагменты первичной и вторичной обкаток обкатки и выпуклые сферические элементы вторичной обкатки. (Рис. 4)

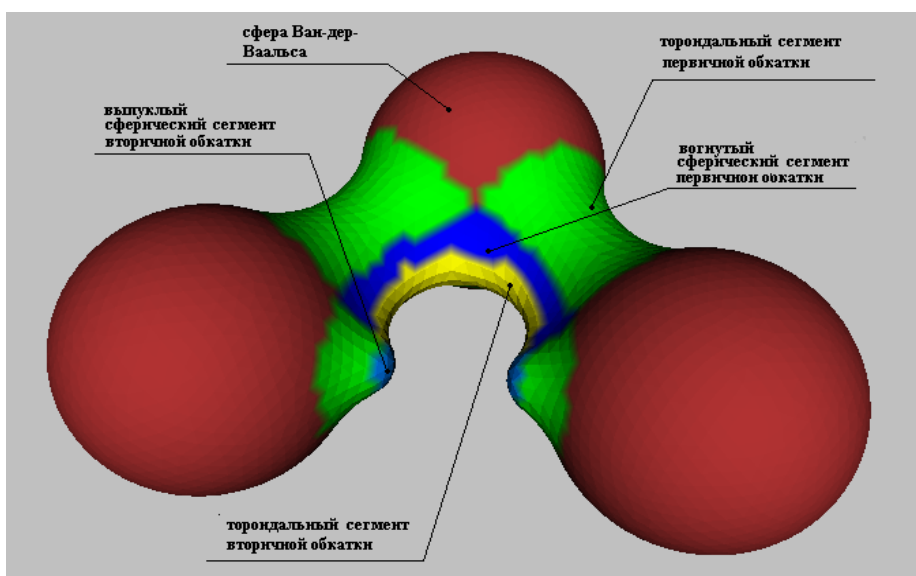


Рис. 4 Поверхность молекулы состоит из фрагментов пяти типов. Это сферические элементы ван-дер-ваальсовых сфер атомов, вогнутые сферические элементы первичной обкатки, тороидальные фрагменты первичной и вторичной обкаток обкатки и выпуклые сферические элементы вторичной обкатки.

Радиусы первичной и вторичной обкатки и критическое расстояние имеют ясный физический смысл. Радиус первичной обкатки равен радиусу сферы, описанной вокруг молекулы растворителя. Радиус вторичной обкатки и определенное выше критическое расстояние определяют минимальную кривизну границы молекулы. Нижние пределы для величины радиуса вторичной обкатки и максимального критического расстояния (определенного выше) связаны с «расплыванием» электронных облаков, которые не могут давать слишком острые углы, узкие перешейки и каналы, могущие возникать при первичной обкатке, вследствие соотношения неопределенности Гейзенберга между координатой и импульсом электрона.

В работе [38] (где впервые предложена вторичная обкатка) рассматривается обкатка *не всех типов* нерегулярностей, которые могут возникать при первичной обкатке (b) не рассматривается ситуация «сцепленных» между собой нерегулярностей первичной обкатки (c) отсутствует рассмотрение ситуаций близких к нерегулярным (d) отсутствует решение проблемы нерегулярностей уже вторичной обкатки.

Эти недостатки были преодолены в алгоритме программы TAGSS (Triangulate Area Grid of Smooth Surface) [29-32]. А именно: (a) новый тип нерегулярности, рассмотренный там – пересечение трех первичных сфер (тройная вторичная точка). (b) рассматривались *большие (конфликтные) группы* нерегулярностей, которые обкатывались совместно (c) вторичную обкатку стоит делать и для регулярных участков поверхности, но близкими по свойствам к нерегулярным. Это участки поверхности с малыми декартовыми пространственными расстояниями (меньше некоего *заранее заданного параметра* - критического расстояния) между удаленными, несоседними участками поверхности (т.е. такие участки поверхности, что декартово пространственное расстояние между ними намного больше расстояния вдоль поверхности), узкие каналы или перешейками (d) нерегулярности могут возникнуть уже и для вторичной обкатки. Но убрать их можно просто непрерывно уменьшая радиус сферы вторичной обкатки до тех пор, пока эти нерегулярности не исчезнут.

Однако описанные методы также имели проблемы: (i) если диаметр сферы вторичной обкатки становился меньше критического расстояния, то приходилось принимать определенное выше критическое расстояние за ноль и обкатывать только истинные нерегулярности, меняя радиус вторичной обкатки до необходимой величины. Последнее же приводило к появлению участков со слишком высокой кривизной. (ii) формируемые слишком *большие (конфликтные) группы* нерегулярностей приводили к слишком долгой работе алгоритма и ухудшали гладкость поверхности, повышенная вероятность ситуации, требующей уменьшения радиуса вторичной обкатки (iii) не были полностью рассмотрены все случаи нескольких полостей (а именно, поверхности, соответствующие атомным группам, лежащим внутри крупных полостей, или случай наличия нескольких молекул).

Дальнейшие шаги по преодолению этих проблем были сделаны в [20,33,65]. Было, соответственно: (i) введено *адаптивное переменное* критическое расстояние, которое могло постепенно уменьшаться в зависимости от типа поверхности (ii) сформулирован корректный и оптимальный механизм формирования *маленьких (конфликтных) групп нерегулярностей*, что позволило повысить гладкость поверхности и быстроту вторичной обкатки (iii) Рассмотрен случай нескольких молекул или случай включения атомных групп в молекулярные полости.

Та часть алгоритма этой части работы, которая относится к вторичной обкатке, использует достижения, о которых говорилось выше [20,33,65] и более подробно описывает, а также развивает их. Такая вторичная обкатка оптимальным образом сглаживает все возможные нерегулярности или близкие к ним ситуации, которые возникают при первичной обкатке. А также разбивает поверхностные атомы и сегменты поверхности на группы, соответствующие разным молекулярным поверхностям, относящимся к разным молекулам или включениям в молекулярные полости.

На основе методов первичной и/или вторичной обкатки были построены программы , PQMS [37], MSMS [26], программа Тотрова и Абагына [27], SIMS [28], TAGSS [29-32] и её усовершенствованная версия, вошедшая как часть в программу DISOLV [20,33,65]. Существуют и более сложные методы получения гладких молекулярных поверхностей, чем механизм вторичной обкатки [57-62, 66,67].

1.1.1.2 Этапы построения поверхностей молекул

1.1.1.2.1 Приём входных данных, заполнение внутренних структур. Входные и выходные данные.

Пусть у нас имеется система, состоящая из нескольких молекул в растворителе. Нам известны координаты и типы атомов, входящих в систему и их радиусы Ван-дер-ваальса. Первая стадия алгоритма – определение параметров обкатки (первичной и вторичной).

Определение параметров обкатки позволяет получить формальное описание поверхности молекулы в виде наборов координат положения и ориентации сферических и тороидальных фрагментов, а также геометрической связанности этих фрагментов друг с другом. Формируются односвязные поверхности молекул, входящих в систему. Далее находятся параметры SES и SAS. Входными данными для алгоритма являются массивы координат и Ван-дер-ваальсовых радиусов атомов. Важными параметрами являются *радиус первичной сферы обкатки* и *максимальный радиус вторичной сферы обкатки*. Задается и *максимальное критическое расстояние*. Это декартово расстояние, на которое алгоритмом разрешается приблизиться *несоседним участкам поверхности*. Несоседние участки поверхности – это участки удаленные друг от друга при измерении расстояния между ними вдоль поверхности. Если их декартово расстояние сближения меньше критического, возникает ситуация близкая к нерегулярности, которая требует вторичной обкатки.

Радиус первичной обкатки определяется радиусом сферы, описанной вокруг молекулы растворителя. Максимальный радиус вторичной обкатки задается пользователем. Он должен быть меньше половины радиуса первичной обкатки и меньше половины каждого из Ван-дер-ваальсовых радиусов атомов. Максимальное критическое расстояние должно быть меньше максимального диаметра вторичной обкатки. Эти максимальные величины адаптивные и могут уменьшаться самим алгоритмом.

Выходными данными являются массивы параметров, описывающих поверхностные тороидальные и сферические сегменты, поверхностные атомы. Определение параметров обкатки позволяет получить формальное описание гладкой поверхности молекулы в виде наборов координат положения и ориентации сферических и тороидальных фрагментов, а также геометрической связанности этих фрагментов друг с другом.

1.1.1.2.2 Определение собственного базиса молекулы и переход в него

Для простоты рассмотрения мы переходим в собственный базис молекулы. Центр этого базиса находится в центре инерции молекулы. Оси совпадают с главными осями инерции молекулы. Всем атомом при этом вычислении придается одинаковый «вес». Радиус-вектор центра инерции:

$$\mathbf{r}_c = \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_{old})_i, \quad (1)$$

где N – число атомов в молекуле, $(\mathbf{r}_{old})_i$ - вектор i -того атома.

Тензор, определяющий моменты инерции вычисляется по следующей формуле:

$$F = \frac{1}{N} \sum_{i=1}^N \begin{vmatrix} (\mathbf{r}_{old})_{i,x}^2 & (\mathbf{r}_{old})_{i,x}(\mathbf{r}_{old})_{i,y} & (\mathbf{r}_{old})_{i,x}(\mathbf{r}_{old})_{i,z} \\ (\mathbf{r}_{old})_{i,y}(\mathbf{r}_{old})_{i,x} & (\mathbf{r}_{old})_{i,y}^2 & (\mathbf{r}_{old})_{i,y}(\mathbf{r}_{old})_{i,z} \\ (\mathbf{r}_{old})_{i,z}(\mathbf{r}_{old})_{i,x} & (\mathbf{r}_{old})_{i,z}(\mathbf{r}_{old})_{i,y} & (\mathbf{r}_{old})_{i,z}^2 \end{vmatrix}. \quad (2)$$

Диагонализация матрицы F дает нам диагональную матрицу S :

$$S = G^T F G. \quad (3)$$

Переход в собственный базис дается формулой:

$$\mathbf{r}_i = G((\mathbf{r}_{old})_i - \mathbf{p}_c) \quad (4)$$

Обратное преобразование дается формулой:

$$\mathbf{r}_{old} = G^T \mathbf{r} + \mathbf{p}_c. \quad (5)$$

1.1.1.2.3 Разбиение пространства вокруг и внутри молекулы на перекрывающиеся кубические области

Разбиение пространства вокруг и внутри молекулы на кубические области представляет собой процесс запоминания индексов атомов молекулы и поверхностных элементов в наборе массивов, каждый из которых соответствует кубической области заданного размера и определённого положения.

Разбиение на кубические области производится с целью упрощения в дальнейшем процедур поиска индекса атома или элемента поверхности ближайшего к точке с заданными координатами. Кубы строятся и для того, чтобы увеличить скорость перебора поверхностных элементов и атомов при построении поверхности. Слишком удаленные элементы или атомы рассматривать обычно не имеет смысла. Перебор осуществляется только в пределах построенных кубов, что сужает пространство поиска.

Размер кубической области есть постоянный внутренний параметр программы, который выбирается исходя из известного диапазона возможных значений Ван-дер-ваальсовых радиусов атомов. Расположения и ориентация кубических областей привязаны к осям собственной системы координат. Следует отметить, что кубические области перекрываются между собой. Это позволяет для любой точки или сегмента найти кубическую область таким образом, чтобы эта точка или сегмент лежали вблизи центра этой области.

1.1.1.2.4 Первичная и вторичная обкатка

Определение параметров обкатки позволяет получить формальное описание гладкой поверхности молекулы в виде наборов координат положения и ориентации сферических и тороидальных фрагментов, а также геометрической связанности этих фрагментов друг с другом.

1.1.1.2.4.1 Первичная обкатка

Первичная обкатка получается катанием шара - зонда первичной обкатки, имитирующего молекулы растворителя, по группе атомов, формирующих молекулу [59]. Она служит для образования молекулярных поверхностей двух типов: SAS и SES. Поверхность SAS описывается положениями центра шар – зонда, имитирующего молекулы растворителя (алгоритм её получения из SES мы рассмотрим в пункте 1.1.1.2.7), а SES – состоит из нижних фрагментов шара - зонда первичной обкатки [59] в результате обкатки вокруг молекулы. Это (а) положения точек касания с атомами (б) отрезки геодезических линий сферы обкатки, проходящими через две точки касания сферы обкатки и атомов (в) сегменты нижней части обкатывающей сферы, ограниченные отрезками геодезических линий сферы обкатки, проходящими через точки касания сферы обкатки и атомов (Рис. 1).

Поверхность SES первичной обкатки состоит из сферических и тороидальных фрагментов трех типов (Рис.4):

- 1) В том случае, когда шар первичной обкатки соприкасается только с одним атомом – это выпуклые сферические *фрагменты Ван-дер-ваальсовых сфер атомов*

- 2) В том случае, когда шар первичной обкатки соприкасается только с двумя атомами – вогнутые тороидальные фрагменты, образующиеся вращением шара первичной обкатки вокруг этих атомов. Будем называть их *полные первичные торы* - если сфера обкатки может совершить полный круг, не натываясь на другие атомы; в ином случае - *первичными осями*.
- 3) В том случае, когда шар первичной обкатки соприкасается с тремя атомами (или даже большим числом атомов в вырожденных случаях, например, для ароматических колец) - это вогнутые сферические фрагменты, образуемые поверхностью шара первичной обкатки, неподвижно зафиксированного этими атомами (точнее - сегменты нижней части обкатывающей сферы, ограниченные отрезками геодезических линий сферы обкатки, проходящими через точки касания сферы обкатки и атомов). Для наиболее типичного невырожденного случая трех опорных атомов – это вогнутый сферический треугольник. Будем называть их для краткости *первичными тройными точками*.

Алгоритм, описывающий первичную обкатку следующий:

1 Шаг: Находим *большие первичные конфликтные группы*, которые формируются следующим образом:

Берем любой атом и находим все соседние атомы, лежащие от него на расстоянии менее диаметра первичной обкатки, затем всех их подобных соседей и т.д. Набор всех таких атомов образует первичную большую конфликтную группу. Затем берем любой атом, не вошедший в эту группу, и строим следующую группу. Так до тех пор, пока не исчерпаем все атомы. Первичные большие конфликтные группы образованы несвязанными между собой наборами молекул. Все большие первичные конфликтные группы обкатываются независимо друг от друга. Делаем это следующим образом:

2 Шаг: Находим *большие первичные конфликтные группы*, состоящие из одного изолированного атома - они допускающие полную обкатку вокруг них без всяких препятствий.

3 Шаг: Находим все *полные первичные торы* обкатки для текущей большой первичной конфликтной группы, состоящей более чем из одного атома. Они формируются следующим образом:

- Находим все пары атомов из большой группы, которые могут быть на полный круг обкатаны шаром первичной обкатки без того, чтобы он наткнулся при этом на другие атомы.
- Если ширина тора в самом узком месте меньше максимального критического расстояния, определенного выше, или тор самопересекается, то делим тор на две независимые половинки. Это связано с тем, что при самопересечении или дальнейшей вторичной обкатке он распадется на два несвязанных куска.
- Для каждого элемента (атом, полный тор или его половинка) запоминаем его соседние элементы (атом, полный тор или его половинка).

4 Шаг: Находим все *первичные тройные точки* большой группы следующим образом:

- Перебираем все тройки атомов, расстояние между которыми меньше диаметра шара первичной обкатки.

- Для каждой тройки атомов существует либо две тройные точки, либо их нет совсем.
- Если тройные точки существуют, проверяем, что образующая их сфера первичной обкатки не пересекает другие атомы. Подобные тройные точки отбраковываем. В случае просто касания других атомов тройная точка является вырожденной.
- Таким образом, перебираем все возможные тройки атомов.
- Для каждого элемента (атом, первичная тройная точка, полный первичный тор или его половинка) запоминаем его соседние элементы (атом, первичная тройная точка, полный первичный тор или его половинка).

5 Шаг: Отбрасываем из рассмотрения в дальнейшем все атомы, не вошедшие в предыдущие группы, поскольку они находятся внутри объема молекулы и не влияют на обкатку. Оставшиеся атомы будем называть поверхностными.

6 Шаг: Находим *малые первичные конфликтные группы*, входящие в большие группы, которые формируются следующим образом:

- Берем любую найденную выше первичную тройную точку.
- Если тройная точка не вырождена, то все ее три опорных атома соединены торами, образованные полным поворотом сферы обкатки вокруг них. Однако другие атомы могут препятствовать такой полной обкатке. При этом на каждом торе может быть несколько несвязанных сегментов, при обкатке которых первичный шар-зонд упирается в два разных атома. Такие независимые сегменты мы назвали ранее *первичными осями*. Они ограничены двумя крайними положениями сферы первичной обкатки, при которых она упирается в другие атомы.
- Выбираем любой атом, на который опирается исходная тройная точка.
- Находим все тройные точки и все первичные оси, опирающиеся на этот же атом, которые образуются при полной круговой обкатке вокруг этого атома первичного шара, находящимся первоначально в исходной тройной точке, до тех пор, пока при обкатке он не встречает препятствий. Включаем их в малую группу.
- Следует отметить, что тор, пересекаясь с атомом, вырезает из него круглые сегменты. После того, как все торы сделают это «обрезание», оставшаяся необрезанная площадь представляет собой один или несколько несвязанных многоугольных сферических сегментов. Каждый такой атомный сегмент будет тоже рассматривать как независимый элемент. Мы будем вносить такой атомный сегмент в ту малую конфликтную группу, которая определяется опирающимися на этот атомный сегмент тройными точками и осями.
- Если на этот атомный сегмент опирается половинка полного тора – включаем его в малую группу.
- Продолжаем брать опорные атомы тройных точек малой группы до тех пор, пока не исчерпаем их все. Формирование первой малой конфликтной группы закончено.
- Далее берем тройную точку, не вошедшую в уже сформированную малую группу, и формируем новую малую группу. Так до исчерпания всех свободных тройных точек.
- Оставшиеся свободные атомы (вместе с опирающимися на них половинками тора, если таковые есть) также образуют независимые малые группы. Формирование малых конфликтных групп закончено.
- Далее перебираем все пары атомов, соединенные полным тором. Если оба сегмента атомов, на который опирается полный тор, принадлежат одной и той же

малой группе – включаем тор в эту малую группу. Если разным малым группам – объединяем эти две малые группы в одну и включаем в нее этот полный тор.

- Для каждого элемента (атом, тройная точка, полный тор или его половинка, сферические и тороидальные сегменты) запоминаем его соседние элементы.

7 Шаг: Фильтрация «паразитных» поверхностей.

Полученные малые первичные конфликтные группы включают одну или несколько независимых замкнутых поверхностей, которые образуются при первичной обкатке. Это:

- 1) Внешняя поверхность молекулы.
- 2) Внутренняя полость, попав в которую, шар первичной обкатки не может выбраться наружу. Внутренние полости обычно имеют нормаль, направленную внутрь полости. Следовательно, написав формулу для нахождения объема полости

$$V = \frac{1}{3} \oint_S (\mathbf{r} \cdot \mathbf{n}) dS$$

можно видеть, что объем в этом случае получается отрицательным.

Это позволяет отфильтровать такие полости

- 3) Внешние поверхности группы атомов, «запертых» во внутреннюю полость. Эту поверхность можно отфильтровать по следующим признакам – ее внутренние атомы являются также и внутренними атомами для внешней поверхности молекулы. При этом средний радиус внешней поверхности больше.
- 4) Случай, когда группы атомов могут распадаться на несколько несвязанных молекул (его обработка рассмотрена в пункте 2.5). Это происходит, когда первичные оси разрываются при самопересечении или вторичной обкаткой. Для полной обработки таких ситуаций необходимо вначале провести вторичную обкатку.

Процедура первичной обкатки иногда может приводить к нежелательным самопересечениям поверхности и изломам SES. Эти нерегулярности можно разделить на два типа, причем нерегулярность может быть одновременно образована несколькими элементами разного или одинакового типа:

- 1) Самопересечение тора первичной обкатки.
- 2) Пересечение вогнутых сферических фрагментов первичной обкатки – излом.

При этом могут возникать ситуации хоть и гладкие, но близкие к этим нерегулярным элементам поверхности. Это происходит, когда

- 1) ширина тора в самом узком месте меньше некоего заранее эмпирически выбранного критического расстояния
- 2) расстояние между вогнутыми сферическими сегментами меньше этого же критического расстояния.

Для преодоления этих проблем применяется метод вторичной обкатки [20,29-33,38, 65].

1.1.1.2.4.2 Вторичная обкатка

Мы выбираем шар-зонд вторичной обкатки с диаметром, меньшим половины радиуса первичной обкатки и меньшим половины радиусов всех атомов. Этим шаром мы обкатываем *внутреннюю* поверхность молекулы, полученную в результате первичной

обкатки, только в тех местах, где возникают нерегулярности или близкие к нерегулярности элементы поверхности.

Но обкатывая эти места, мы неизбежно должны обкатать и все связанные с ними другие фрагменты первичной обкатки. Так, например, обкатывая шаром вторичной обкатки пересечение вогнутых сферических сегментов первичной обкатки, мы неизбежно должны обкатать и торы первичной обкатки, которые с ними соприкасаются. Другие возможные ситуации будут описаны ниже. Так образуются наборы фрагментов первичной обкатки, которые нуждаются во вторичной обкатке и объединяются в *конфликтные группы вторичной обкатки*.

При обкатке таких конфликтных групп могут возникать новые нерегулярности (или близкие к ним ситуации), связанные с пересечением или сближением на расстояние меньшего критического фрагментов шаров вторичной обкатки. Однако, третичная обкатка уже не нужна! Мы можем ликвидировать эти новые нерегулярности (или близкие к ним ситуации) путем плавного уменьшения радиуса вторичной обкатки до той величины, когда эти проблемы исчезают. Также может возникнуть необходимость уменьшения критического расстояния, определяющего ситуации близкие к нерегулярным. Эта необходимость возникает при уменьшении диаметра шара вторичной обкатки (проводимого для ликвидации этих новых нерегулярностей или близких к ним ситуаций) до величины меньшей этого критического расстояния. Диаметр шара вторичной обкатки должен оставаться всегда больше этого критического расстояния.

Поверхность SES вторичной обкатки, сглаживающая нерегулярность, состоит из сферических и тороидальный фрагментов трех типов Рис. 4:

- 1) В том случае, когда шар вторичной обкатки соприкасается только с двумя вогнутыми сферическими сегментами первичной обкатки – выпуклые тороидальные фрагменты вторичной обкатки, образующиеся вращением шара вторичной обкатки вокруг этих сферических сегментов – вторичные торы. Сегменты вторичных торов называются вторичными осями.
- 2) В том случае, когда шар первичной обкатки соприкасается с тремя (и более в вырожденных случаях) сферическими сегментами первичной обкатки - выпуклые сферические фрагменты, образуемые поверхностью шара вторичной обкатки, неподвижно зафиксированного этими первичными сферами – вторичные тройные точки.
- 3) В случае, когда образуется самопересечение тора первичной обкатки или близкая к этому ситуация узкого перешейка тора – возникает пара устойчивых положений шара вторичной обкатки вблизи пары точек самопересечения или узкого перешейка тора. Выпуклые сферические сегменты шара вторичной обкатки, сглаживающие нерегулярность образует третий тип – пары точек устойчивого положения. См. Рис. 2.

Мы видим, что первые два типа сглаживания аналогичны типам, возникающим при первичной обкатке, и появляются при вторичной обкатке первичных сфер. Третий тип новый и возникает вследствие обкатки первичных торов, которых не было при первичной обкатке.

На Рис. 2, 3, 4 показана процедура сглаживания изломов и удаления области самопересечения при помощи метода вторичной обкатки. Она происходит следующим образом. Для каждой найденной ранее малой *первичной* конфликтной группы вторичная обкатка производится независимо:

- 1) Обкатываем полные торы, с шириной перешейка меньше критического или самопересекающиеся. Образуется пара точек устойчивого положения. Вторичная конфликтная группа не возникает.
- 2) Обкатываем вторичным шаром-зондом пересекающиеся или лежащие на расстоянии меньше критического первичные тройные точки. Если ничто не препятствует обкатке, образуется полный вторичный тор и вторичная конфликтная группа не возникает.
- 3) В случаях, отличных от двух предыдущих, возникают сцепленные нерегулярности и образуются их вторичные конфликтные группы. Формируем вторичные конфликтные группы первичных тройных точек следующим образом: Берем любую первичную тройную точку и находим все соседние первичные тройные точки, лежащих от нее на расстоянии менее диаметра вторичной обкатки, затем все их подобные соседи и т.д. Комплекс всех таких первичных тройных точек образует *большую вторичную конфликтную группу*. Затем берем любую первичную тройную точку, не вошедшую в эту группу, и строим следующую группу. Так до тех пор, пока не исчерпаем все первичные тройные точки. Внутри таких *больших вторичных конфликтных групп* будем искать *малые вторичные конфликтные группы* связанных нерегулярностей или близких к нерегулярностям ситуаций.
- 4) Обкатываем вторичным шаром-зондом пересекающиеся или лежащие на расстоянии Δ меньше критического две первичные тройные точки. Начинаем с пересекающихся тройных точек (или с минимального Δ при их отсутствии). (Далее для формирования следующих малых вторичных конфликтных групп берем тройные точки в порядке увеличения Δ .) Запоминаем это текущее расстояние Δ (при отсутствии пересечения) и радиус вторичной обкатки в массив «резервных» критических величин. При обкатке образуется вторичные оси. Движению вторичного шара-зонда вдоль этих осей может что-то препятствовать.
 - Движению вдоль этой вторичной оси может помешать другая первичная тройная точка из большой вторичной конфликтной группы (В принципе, может быть и вырожденный случай более одной такой первичной тройной точки, *одновременно* останавливающих обкатку). При этом образуется уже вторичная тройная точка, опирающаяся на три первичные тройные точки. Также образуются новые, выходящие из нее вторичные оси. Включаем пары первичных тройных точек (служащих опорой для этой вторичной тройной точки и для исходящих из нее вторичных осей) в малую вторичную конфликтную группу для их дальнейшей обкатки.
 - Торы первичной обкатки из большой конфликтной группы, примыкающие к этой паре первичных тройных точек, могут служить препятствием для обкатки вторичного тора. Включаем эти торы и связанные с ними первичные оси в малую вторичную конфликтную группу для их дальнейшей обкатки.
 - При обкатке такого тора образуется пара точек устойчивого положения. Мы должны обкатать и все пары первичных тройных точек, примыкающие к такому тору. Включаем эти пары первичных тройных точек в малую вторичную конфликтную группу для их дальнейшей обкатки.
 - Обкатываем все (как уже запомненные ранее, так и образующиеся по тому же принципу снова) пары первичных тройных точек и первичных осей (входящие в формируемую малую конфликтную группу) до их полного исчерпания.
- 5) Может возникнуть ситуация, когда при вторичной обкатке, описанной в пункте 4:
 - а. Могут попасть под вторичную обкатку нерегулярность или близкая к ней ситуация, *уже* входящая в другую, ранее образованную малую конфликтную группу.

- b. Образуется вторичная ось с перешейком, меньше критического (ширина Δ) или самопересекающаяся.
- c. Пара сфер устойчивого положения пересечется или сблизится на расстояние (Δ), меньшее критического.
- d. Пара вторичных тройных точек пересечется или сблизится на расстояние (Δ), меньшее критического.

Тогда уменьшаем непрерывно диаметр вторичной обкатки до того максимально большого диаметра D_{new} , когда описанная выше ситуация с обкаткой ранее обкатанной нерегулярности, либо нерегулярность или близкая к ней ситуация исчезает. Далее возвращаемся в начало пункта 4 и формируем заново конфликтную группу. При этом критическое расстояние и диаметр вторичной обкатки выбираем следующим образом:

- a. Пусть этот диаметр D_{new} *больше* текущего критического расстояния. Тогда выбираем новый диаметр обкатки равный ему, а критическое расстояние оставляем прежним. Если минимальное расстояние Δ , приведшее к проблеме, больше нуля, то запоминаем величину этого расстояния и найденный радиус вторичной обкатки в массив «резервных» критических расстояний.
- b. Пусть этот диаметр D_{new} *меньше* текущего критического расстояния и максимальное «резервное» критическое расстояние *больше* этого нового диаметра вторичной обкатки. Тогда выбираемое заново критическое расстояние равно максимальному «резервному», а радиус вторичной обкатки – также «резервный» соответствующий выбранному критическому расстоянию. Из «резерва» выбрасываются эти величины.
- c. Пусть этот диаметр D_{new} *меньше* текущего критического расстояния и максимальное «резервное» критическое расстояние *меньше* этого нового диаметра вторичной обкатки. Тогда выбираемые заново критическое расстояние и диаметр вторичной обкатки берутся равными этому диаметру D_{new} вторичной обкатки.
- d. Если одно из найденных критических расстояний слишком мало (например, меньше допустимого машинной точностью), то увеличиваем радиусы атомов на 0.1 Å и повторяем первичную и вторичные обкатку.

Продолжаем это процесс до тех пор, пока не будут сформированы все малые вторичные конфликтные группы и определены критические расстояния и диаметры вторичной обкатки для них. Для каждого элемента (атом, тройная точка, полный тор или его половинка, первичные и вторичные сферические и тороидальные сегменты) запоминаем его соседние элементы.

1.1.1.2.5 Формирование независимых групп поверхностных атомов и сегментов, каждая из которых соответствует одной и только одной поверхности

Здесь мы даем алгоритм для случая, когда группы атомов могут распадаться на несколько несвязанных молекул (Это, например, случай докинга (присоединения) лиганда к субстрату). При этом происходит формирование независимых групп поверхностных атомов и сегментов, каждая из которых соответствует одной и только одной поверхности.

Малые *первичные* конфликтные группы могут распасться на подгруппы, соответствующие разным поверхностям. Это происходит, когда первичные оси разрываются при самопересечении или сферами устойчивого положения вторичной обкатки.

Однако такой разрыв однозначно происходит лишь только для такого типа первичной оси, где две ее смежные первичные тройные точки (ограничивающие ее угловые размеры) лежат на трех или более одинаковых опорных атомах. Другие типы первичных осей могут не разорваться, поскольку оказываются связанными возникшей после вторичной обкатки *сетью* из одной или более вторичных осей. Мы устанавливаем новые связи атомов через эту вновь возникшую сеть.

Делая анализ связанность (т.е. разбивая атомы малых первичных конфликтных групп на другие подгруппы из-за разрыва первичных осей (торов) и иногда восстанавливая связанность этих подгрупп через сеть вторичных торов), находим те первичные группы поверхностных атомов и сегментов, которые будут соответствовать разным поверхностям.

1.1.1.2.6 Определение параметров сферических и тороидальных сегментов

1.1.1.2.6.1 Определение параметров выпуклых сферических сегментов поверхностных атомов

Выпуклые сферические фрагменты являются частями Ван-дер-ваальсовых сфер атомов, которые обращены к растворителю. Для описания этих сферических фрагментов для каждого атома используется набор векторов описывающих направления на соседние контактные атомы и набор значений «запрещённых» углов связанных с каждым из направлений. Контактными считаются атомы, имеющие расстояние между своими сферами меньше диаметра сферы первичной обкатки $2R_{pr}$. А «запрещённый» угол – это угол, под которым из центра атома виден радиус дуговой траектории центра сферы обкатки при касании пары атомов – текущего и второго контактного. Та часть Ван-дер-ваальсовской сферы данного атома, которая лежит вне всех «конфликтных конусов» является искомым выпуклым сферическим фрагментом поверхности молекулы. Следует отметить, что запрещенные углы могут разбивать атомную сферу не только на один, но и на два или даже более несвязанных между собой сферических фрагментов S_R .

Пусть r_0 - вектор координат текущего атома, а r_i - вектор координат контактного атома; R_0 - радиус текущего атома, а R_i - радиус контактного атома;

Для каждого атома определены массивы единичных векторов \vec{d}_i и массивы запрещённых углов: γ_i . Вектор направления определяется из условия:

$$d_i = \frac{r_i - r_0}{|r_i - r_0|}, \quad (6)$$

а запрещённый угол:

$$\cos(\gamma_i) = \frac{b^2 - a^2 - c^2}{2ac}, \quad a = R_0 + R_{pr}, \quad b = R_i + R_{pr}, \quad c = |r_i - r_0|, \quad (7)$$

где индексом '0' обозначен текущий атом, а индексом 'i' - контактный атом. Алгоритм определения параметров выпуклых сферических элементов состоит в переборе всех пар контактных атомов и вычислении для каждого из атомов набора направляющих векторов и соответствующих запрещённых углов.

1.1.1.2.6.2 Определение массива координат центров первичных вогнутых сферических фрагментов (первичная сферическая перетяжка) между тройками атомов

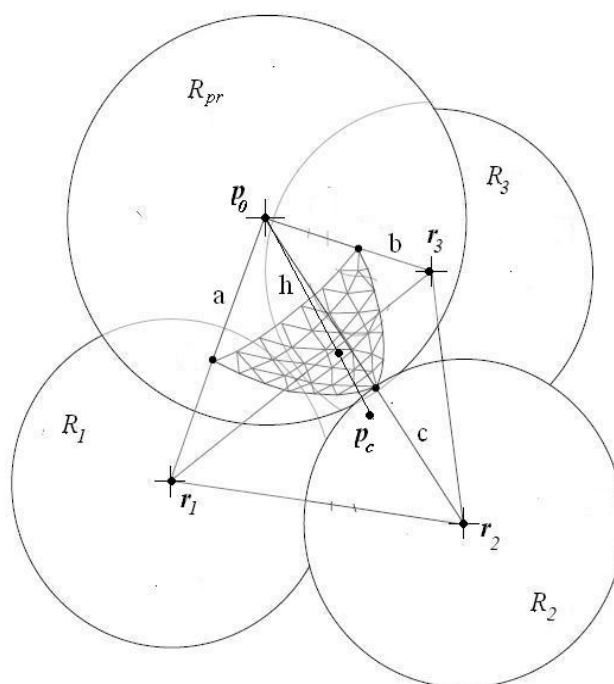


Рис. 5 Геометрия задачи об определении центра сферы касающейся трёх атомов.

Вогнутые сферические элементы поверхности SES (Рис. 5) имеют форму сферического треугольника и образуются при контакте сферы обкатки одновременно с тремя атомами. (Кроме редких вырожденных случаев. При этом точки контактов сферы обкатки со сферами атомов определяют вершины сферического треугольника являющегося частью сферы обкатки.)

Расчёт координат вектора \mathbf{p}_0 центра сферы R_{pr} опирающейся на три атомами с координатами $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ производится по нижеследующим формулам.

Боковые стороны пирамиды:

$$a = R_1 + R_{pr}; b = R_2 + R_{pr}; c = R_3 + R_{pr}, \quad (8)$$

где h - высота пирамиды, \mathbf{p}_c - радиус-вектор основания высоты пирамиды.

$$\mathbf{z} = \frac{[(\mathbf{r}_3 - \mathbf{r}_1) \times (\mathbf{r}_2 - \mathbf{r}_1)]}{\|[(\mathbf{r}_2 - \mathbf{r}_1) \times (\mathbf{r}_3 - \mathbf{r}_1)]\|},$$

(9)

$$\mathbf{r} = \mathbf{p}_c \pm \mathbf{z} h, \quad (10)$$

$$\begin{cases} (\mathbf{r} - \mathbf{r}_1)^2 = a^2 \\ (\mathbf{r} - \mathbf{r}_2)^2 = b^2 \\ (\mathbf{r} - \mathbf{r}_3)^2 = c^2 \end{cases} \Leftrightarrow \quad (11)$$

$$\begin{cases} (\mathbf{p}_c - \mathbf{r}_1)^2 = a^2 - h^2 \\ (\mathbf{p}_c - \mathbf{r}_2)^2 = b^2 - h^2 \\ (\mathbf{p}_c - \mathbf{r}_3)^2 = c^2 - h^2 \end{cases} \Rightarrow \quad (12)$$

$$\begin{cases} (\mathbf{r}_2 - \mathbf{r}_1) \cdot \mathbf{p}_c = (a^2 - b^2 - \bar{r}_1^2 + \bar{r}_2^2) / 2, \\ (\mathbf{r}_3 - \mathbf{r}_1) \cdot \mathbf{p}_c = (a^2 - c^2 - \bar{r}_1^2 + \bar{r}_3^2) / 2 \end{cases} \quad (13)$$

$$\mathbf{p}_c = \mathbf{r}_1 + \alpha[-\mathbf{z} \times (\mathbf{r}_2 - \mathbf{r}_1)] + \beta[\mathbf{z} \times (\mathbf{r}_3 - \mathbf{r}_1)] \quad (14)$$

$$\mathbf{p}_c = \mathbf{r}_1 + [\mathbf{z} \times (\gamma \mathbf{r}_1 - \alpha \mathbf{r}_2 + \beta \mathbf{r}_3)] \quad (15)$$

$$\begin{cases} \alpha = \frac{(a^2 - c^2 + (\mathbf{r}_3 - \mathbf{r}_1)^2)}{2 \|[(\mathbf{r}_2 - \mathbf{r}_1) \times (\mathbf{r}_3 - \mathbf{r}_1)]\|} \\ \beta = \frac{(a^2 - b^2 + (\mathbf{r}_2 - \mathbf{r}_1)^2)}{2 \|[(\mathbf{r}_2 - \mathbf{r}_1) \times (\mathbf{r}_3 - \mathbf{r}_1)]\|} \\ \gamma = \alpha - \beta = \frac{(b^2 - c^2 - (\mathbf{r}_2 - \mathbf{r}_1)^2 + (\mathbf{r}_3 - \mathbf{r}_1)^2)}{2 \|[(\mathbf{r}_2 - \mathbf{r}_1) \times (\mathbf{r}_3 - \mathbf{r}_1)]\|} \end{cases} \quad (16)$$

$$[(\mathbf{r}_3 - \mathbf{r}_1) \times (\mathbf{r}_2 - \mathbf{r}_1)] = [(\mathbf{r}_2 - \mathbf{r}_3) \times (\mathbf{r}_1 - \mathbf{r}_3)] = [(\mathbf{r}_1 - \mathbf{r}_2) \times (\mathbf{r}_3 - \mathbf{r}_2)] \quad (17)$$

$$\mathbf{p}_c = \frac{(\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3)}{3} + \frac{1}{6} \left[\begin{aligned} & \left[\frac{3(b^2 - c^2) + (\mathbf{r}_3 - \mathbf{r}_1)^2 - (\mathbf{r}_2 - \mathbf{r}_1)^2}{\|[(\mathbf{r}_3 - \mathbf{r}_1) \times (\mathbf{r}_2 - \mathbf{r}_1)]\|^2} [[(\mathbf{r}_3 - \mathbf{r}_1) \times (\mathbf{r}_2 - \mathbf{r}_1)] \times \mathbf{r}_1] + \right. \\ & \left. + \frac{3(c^2 - a^2) + (\mathbf{r}_1 - \mathbf{r}_2)^2 - (\mathbf{r}_3 - \mathbf{r}_2)^2}{\|[(\mathbf{r}_1 - \mathbf{r}_2) \times (\mathbf{r}_3 - \mathbf{r}_2)]\|^2} [[(\mathbf{r}_1 - \mathbf{r}_2) \times (\mathbf{r}_3 - \mathbf{r}_2)] \times \mathbf{r}_2] + \right. \\ & \left. + \frac{3(a^2 - b^2) + (\mathbf{r}_2 - \mathbf{r}_3)^2 - (\mathbf{r}_1 - \mathbf{r}_3)^2}{\|[(\mathbf{r}_2 - \mathbf{r}_3) \times (\mathbf{r}_1 - \mathbf{r}_3)]\|^2} [[(\mathbf{r}_2 - \mathbf{r}_3) \times (\mathbf{r}_1 - \mathbf{r}_3)] \times \bar{\mathbf{r}}_3] \right] \end{aligned} \right] \quad (18)$$

$$h^2 = \frac{a^2 + b^2 + c^2 - (\mathbf{r}_1^2 + \mathbf{r}_2^2 + \mathbf{r}_3^2) + 2\mathbf{p}_c \cdot (\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3)}{3} - \mathbf{p}_c^2 \quad (19)$$

Для каждой сферы первичной обкатки определены массивы единичных векторов \mathbf{d}_i и массивы запрещённых углов: γ_i . Они связаны с вторичными торами обкатки вогнутых сферических элементов. Вектор направления определяется из условия:

$$\mathbf{d}_i = \frac{\mathbf{r}_i - \mathbf{p}_0}{|\mathbf{r}_i - \mathbf{p}_0|} \quad (20)$$

а запрещённый угол:

$$\cos(\gamma_i) = \frac{b'^2 - a'^2 - c'^2}{2a'c'}, \quad a' = R_{pr} + R_{sec}, \quad b' = R_{pr} + R_{sec}, \quad c' = |\mathbf{r}_i - \mathbf{p}_0|, \quad (21)$$

где индексом '0' обозначена текущая первичная сфера, а индексом 'i' - контактная первичная сфера.

Алгоритм определения параметров вогнутых сферических элементов состоит в переборе всех пар вогнутых сферических элементов, для которых проводится вторичная обкатка, и вычислении для каждого из вогнутых сферических элементов набора направляющих векторов и соответствующих запрещённых углов.

1.1.1.2.6.3 Определение массивов параметров тороидальных фрагментов между парами атомов

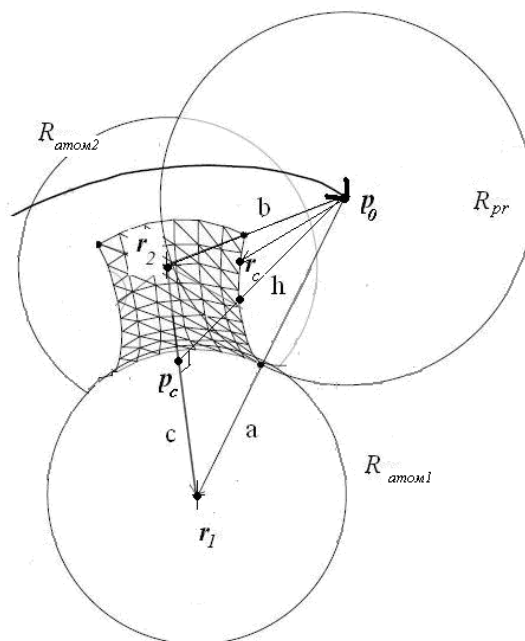


Рис. 6 Обкатка двух атомов первичным шаром-зондом. Формирование первичного тора.

Первичные тороидальные фрагменты поверхности (Рис. 6) образуются в процессе обкатки первичной сферой обкатки двух атомов. При этом центр сферы обкатки описывает круговую дугу или полную окружность в зависимости от расположения других атомов. Тороидальный фрагмент заключён между точками касания сферы двух атомов и положениями сферы обкатки, в которых она упирается в какой-либо другой атом, то есть положениями касания сферы обкатки одновременно трёх атомов.

Для описания тороидальных фрагментов используются следующие параметры.

Центр \mathbf{p}_c и радиус h окружности, вокруг которой описывает дугу центр сферы обкатки.

Локальный базис x, y, z , ось z которого направлена параллельно прямой соединяющей центры опорных атомов, а оси x и y определяют плоскость дуги центра сферы обкатки, при этом ось x направлена на один из атомов, который ограничивает обкатку вокруг данной пары атомов.

Углы α и β относительно оси x начала и конца дуговой траектории центра сферы обкатки.

Угол γ , описывает максимальное угловое отклонение от прямой, определяемой осью z , радиус-вектора из текущей точки тора в текущий центр сферы обкатки $\mathbf{p}_0 - \mathbf{r}_s$. При этом отклонении она ещё не вырезается из тора одной из сфер устойчивого положения, если была необходимость во вторичной обкатке тора

Для элементов находящихся на тороидальной перетяжке между атомами \mathbf{r}_1 и \mathbf{r}_2 :
 R_{pr} -радиус первичной сферы обкатки, R_{sec} -радиус вторичной сферы обкатки, R_1 – радиус первого атома, R_2 – радиус второго атома
 Промежуточные расчёты проводятся по нижеследующим формулам:

$$a = R_1 + R_{pr}; b = R_2 + R_{pr}; c = |\mathbf{r}_2 - \mathbf{r}_1|, \quad (22)$$

c - расстояние между атомами.
 Радиус круговой траектории центра сферы обкатки:

$$h = \frac{1}{2c} \sqrt{4a^2c^2 - (a^2 + c^2 - b^2)^2}. \quad (23)$$

Центр круговой траектории:

$$\mathbf{p}_c = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2) + \frac{(\mathbf{r}_2 - \mathbf{r}_1)(a^2 - b^2)}{2c^2}. \quad (24)$$

Локальный ортонормированный базис

$$\mathbf{z} = \frac{(\mathbf{r}_2 - \mathbf{r}_1)}{c}, \quad (\mathbf{x} \cdot \mathbf{z}) = 0, \quad (\mathbf{y} \cdot \mathbf{z}) = 0, \quad (\mathbf{x} \cdot \mathbf{y}) = 0. \quad (25)$$

Если атом с координатами \vec{r}_3 ограничивает обкатку вокруг атомов \vec{r}_1 и \vec{r}_2 , то определение базисных векторов задано выражениями:

$$\mathbf{l} = \frac{\mathbf{r}_3 - \mathbf{p}_c}{|\mathbf{r}_3 - \mathbf{p}_c|}, \quad \mathbf{y} = [\mathbf{z} \times \mathbf{l}], \quad \mathbf{x} = [\mathbf{z} \times \mathbf{y}], \quad (26)$$

Расчёт углов α и β производится по формулам:

$$\alpha = \arccos\left(\mathbf{x} \cdot \frac{\mathbf{p}_{0,\alpha} - \mathbf{p}_c}{|\mathbf{p}_{0,\alpha} - \mathbf{p}_c|}\right), \quad \beta = \arccos\left(\mathbf{x} \cdot \frac{\mathbf{p}_{0,\beta} - \mathbf{p}_c}{|\mathbf{p}_{0,\beta} - \mathbf{p}_c|}\right), \quad (27)$$

где $\mathbf{p}_{0,\alpha}$ и $\mathbf{p}_{0,\beta}$ - координаты центра сферы обкати при касании трёх атомов на начале и конце дуги обкатки вокруг пары атомов. Расчёт угла γ производится по формуле

$$\sin(\gamma) = \frac{h}{R_{pr} + R_{sec}}. \quad (28)$$

1.1.1.2.6.4 Определение массива координат центров и радиусов вторичных сферических фрагментов (вторичная сферическая перетяжка) между тройками первичных сферических фрагментов

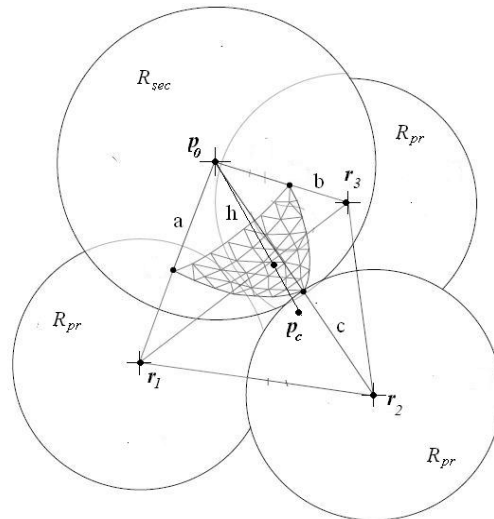


Рис. 7 Геометрия задачи об определении центра вторичного шара-зонда, касающегося трёх первичных сфер обкатки.

Расчёт координат вектора p_0 центра вторичной сферы опирающейся на три первичные сферы обкатки (Рис. 7) с координатами r_1, r_2, r_3 производится по нижеследующим формулам. Радиус-вектора положений центров опорных первичных сфер r_1, r_2, r_3 определяют основание пирамиды, а центр вторичного шара-зонда r ее вершину. Этот случай похож на случай первичной сферической перетяжки. Он даже проще, поскольку все ребра пирамиды, выходящие из ее вершины в текущем случае равны.

Радиус-вектор центра шара-зонда можно разложить по двум векторам - это радиус-вектор основания высоты пирамиды p_c и единичный вектор, перпендикулярный основанию пирамиды z :

$$r = p_c \pm z h . \quad (29)$$

Эти два радиус-вектор определяются двумя формулами

$$z = \frac{[(r_3 - r_1) \times (r_2 - r_1)]}{\|[(r_2 - r_1) \times (r_3 - r_1)]\|} , \quad (30)$$

$$p_c = \frac{(r_1 + r_2 + r_3)}{3} + \frac{1}{6} \left[\left(\frac{(r_3 - r_1)^2 - (r_2 - r_1)^2}{\|[(r_3 - r_1) \times (r_2 - r_1)]\|^2} [[(r_3 - r_1) \times (r_2 - r_1)] \times r_1] + \right. \right. \\ \left. \left. + \frac{(r_1 - r_2)^2 - (r_3 - r_2)^2}{\|[(r_1 - r_2) \times (r_3 - r_2)]\|^2} [[(r_1 - r_2) \times (r_3 - r_2)] \times r_2] + \right. \right. \\ \left. \left. + \frac{(r_2 - r_3)^2 - (r_1 - r_3)^2}{\|[(r_2 - r_3) \times (r_1 - r_3)]\|^2} [[(r_2 - r_3) \times (r_1 - r_3)] \times r_3] \right) \right], \quad (31)$$

$$h^2 = (R_{pr} + R_{sec})^2 - (r_1^2 + r_2^2 + r_3^2) / 3 - p_c^2 + 2p_c \cdot \frac{(r_1 + r_2 + r_3)}{3}, \quad (32)$$

где R_{sec} - радиус сферы вторичной обкатки.

1.1.1.2.6.5 Определение массива координат центров и радиусов пар вторичных сферических фрагментов (сферы устойчивых положений), сглаживающих торы первичной обкатки при их самопересечении или наличии узких перешейков

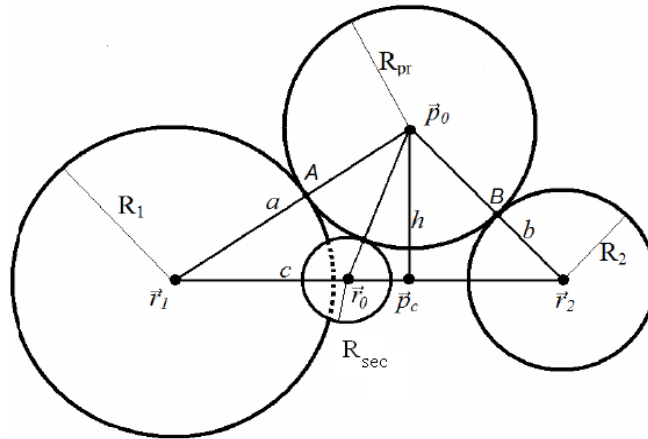


Рис. 8 Вторичные сферы устойчивых положений. См. [20].

Для точек, находящихся на поверхности сфер устойчивых положений (Рис. 8), расстояние центра сферы до центра тора d :

$$d = \sqrt{(R_{pr} + R_{sec})^2 - h^2}.$$

(33)

Радиус-вектора r_0 центров сфер устойчивых положений:

$$r_0 = p_c \pm dz.$$

(34)

1.1.1.2.6.6 Определение массива координат центров и радиусов вторичных сферических фрагментов (вторичная сферическая перетяжка) между тройками первичных сферических фрагментов

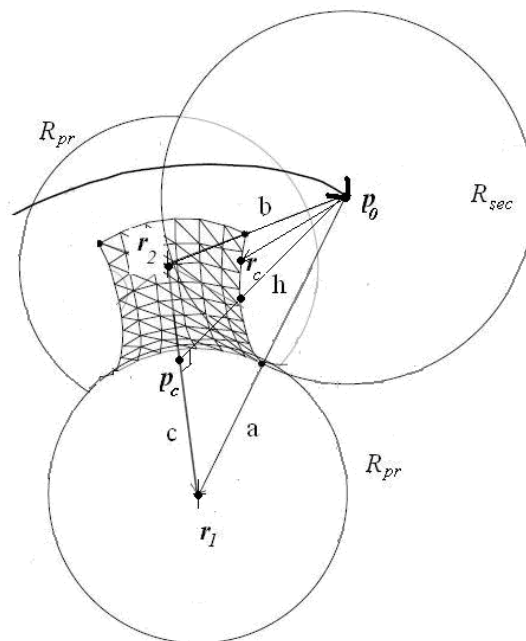


Рис. 9 Вторичная обкатка двух сфер первичной обкатки. Торы вторичной обкатки.

Тороидальные вторичные фрагменты (Рис. 9) поверхности образуются в процессе обкатки вторичной сферой обкатки двух первичных сферических сегментов. При этом центр вторичной сферы обкатки описывает круговую дугу или полную окружность в зависимости от расположения других первичных сфер или первичных торов. Тороидальный вторичный фрагмент заключён между точками касания вторичной сферы двух первичных сфер и положениями вторичной сферы обкатки, в которых она упирается в какой-либо другую (или даже несколько *одновременно*) первичную сферу или первичный тор.

Для описания вторичных тороидальных фрагментов используются следующие параметры.

Центр p_c и радиус h окружности, вокруг которой описывает дугу центр сферы вторичной обкатки.

Локальный базис x, y, z , ось z которого направлена параллельно прямой соединяющей центры опорных первичных сфер, а оси x и y определяют плоскость дуги центра вторичной сферы обкатки. При этом ось x направлена либо на одну из сфер первичной обкатки, которая ограничивает обкатку вокруг данной пары атомов сфер первичной обкатки, либо (для ограничивающего движение первичного тора) на центр вторичной сферы, ограничивающий вращение.

Углы α и β относительно оси x начала и конца дуговой траектории центра сферы обкатки. Плоскость Рис. 9 задают центры сфер первичной обкатки r_1 и r_2 и точка поверхности r_s : R_{sec} -радиус сферы вторичной обкатки, R_{pr} - радиус сферы первичной обкатки.

Стороны треугольника, образуемого центрами сфер и их производные:

$$a = b = R_{pr} + R_{sec} ; c = |r_2 - r_1|. \quad (35)$$

Высота треугольника или, что тоже, радиус круговой траектории центра сферы вторичной обкатки:

$$h = \sqrt{a^2 - \frac{c^2}{4}}. \quad (36)$$

Центр круговой траектории центра сферы вторичной обкатки или, что тоже, радиус-вектор точки основания высоты треугольника p_c :

$$p_c = \frac{r_1 + r_2}{2}. \quad (37)$$

x, y, z - локальный базис парной перетяжки (центр координат в основании высоты p_c , ось z в направлении от первой опорной первичной сферы ко второй):

$$z = \frac{r_2 - r_1}{c}, \quad (x \cdot z) = 0, \quad (y \cdot z) = 0, \quad (x \cdot y) = 0. \quad (38)$$

Если первичная сфера с координатами \vec{r}_3 ограничивает обкатку вокруг первичных сфер r_1 и r_2 , то определение базисных векторов задано выражениями:

$$l = \frac{r_3 - p_c}{|r_3 - p_c|}, \quad y = [z \times l], \quad x = [z \times y]. \quad (39)$$

Если вторичной сфера ограничивает обкатку вокруг первичных сфер r_1 и r_2 выбрана для задания направления оси x , то определение базисных векторов задано выражениями:

$$x = \frac{p_{0,\alpha} - p_c}{|p_{0,\alpha} - p_c|}, \quad y = [z \times x], \quad (40)$$

где $p_{0,\alpha}$ - радиус-вектор центра ограничивающей вторичной сферы.

Расчёт углов α и β производится по формулам:

$$\alpha = \arccos\left(x \cdot \frac{p_{0,\alpha} - p_c}{|p_{0,\alpha} - p_c|}\right), \quad \beta = \arccos\left(x \cdot \frac{p_{0,\beta} - p_c}{|p_{0,\beta} - p_c|}\right), \quad (41)$$

где $p_{0,\alpha}$ и $p_{0,\beta}$ - координаты центра вторичной сферы обкати при касании трёх первичных сфер на начале и конце дуги обкатки вокруг пары первичных сфер или координаты центров вторичных сфер устойчивого положения.

Проекция нормали на локальный базис для торов вторичной обкатки (нормаль направлена внутрь тора):

$$\begin{aligned} \mathbf{n}_s \cdot \mathbf{z} &= -\gamma_s, \\ (\mathbf{n}_s \cdot \mathbf{x}) &= ((-\sqrt{\mathbf{n}_{sx}^2 + \mathbf{n}_{sy}^2} \mathbf{e}_s) \cdot \mathbf{x}) = -\alpha_s \sqrt{1 - \gamma_s^2}, \\ (\mathbf{n}_s \cdot \mathbf{y}) &= ((-\sqrt{\mathbf{n}_{sx}^2 + \mathbf{n}_{sy}^2} \mathbf{e}_s) \cdot \mathbf{y}) = -\beta_s \sqrt{1 - \gamma_s^2}. \end{aligned} \quad (42)$$

$$\mathbf{n}_s = -\alpha_s \sqrt{1 - \gamma_s^2} \mathbf{x} - \beta_s \sqrt{1 - \gamma_s^2} \mathbf{y} - \gamma_s \mathbf{z}. \quad (43)$$

1.1.1.2.7 Поверхность SAS

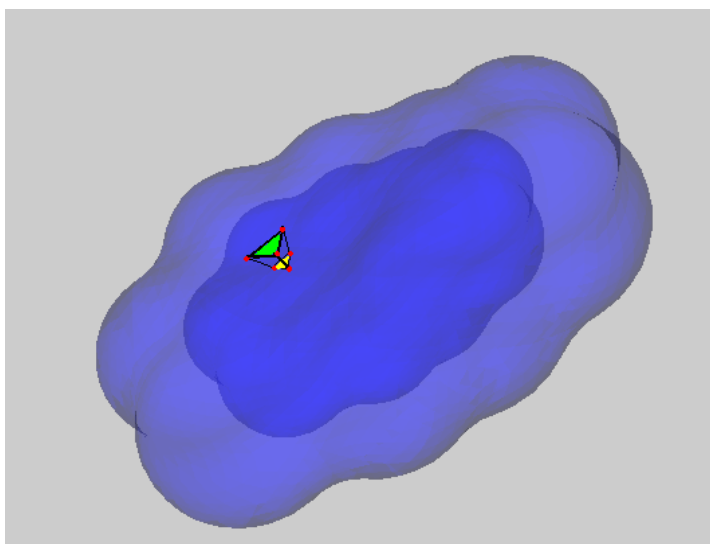


Рис. 10 Преобразование поверхности типа SES в поверхность типа SAS

Пусть имеется поверхность после только *первичной* обкатки. Тогда поверхность SAS (Рис. 10) получается из нее переносом всех точек поверхности вдоль внешней нормали на расстояние, равное радиусу первичной обкатки (R_{pr}). При этом сегменты сфер атомов с радиусом R_{atom} преобразуются в сегменты сфер SAS с радиусом $R_{atom} + R_{pr}$, торы первичной обкатки выражаются в линии пересечения сфер, сферические сегменты тройных точек вырождаются в точки пресечения трех (или более при вырождении) сфер.

1.1.1.3 Выводы

Представленный алгоритм и основанные на нем программы позволяют быстро и надежно строить максимально гладкую поверхность вокруг любых молекул, включая белки, а затем делать ее триангуляцию. Поверхность получается раскрашенной в зависимости от типа ближайших атомов, что удобно для визуализации, и триангулированной, что позволяет не только вычислять ее поверхность и ограниченный ею объем, но и решать с хорошей точностью интегральные уравнения, используемые в континуальной модели растворителя. Полученная с помощью данной программы поверхность может применяться как для целей визуализации молекулы, это особенно актуально для больших белковых молекул, так и для целей вычисления сольватационных вкладов в энергию взаимодействия молекул друг с другом при наличии внешней среды.

Глава 1.1.2 Континуальная модель среды II: Универсальный адаптивный алгоритм для триангуляции гладкой молекулярной поверхности

В данной работе представлен алгоритм, описывающий оригинальные и универсальные принципы триангуляции гладкой молекулярной поверхности, исключая растворитель (SES), полученной первичной и вторичной обкаткой, и производной от нее поверхности, доступной растворителю (SAS) [68]. Эти поверхности играют роль границы между областями молекулы и растворителя.

Оригинальность данной работы состоит в создании универсального и адаптивного алгоритма триангуляции. Универсальность алгоритма триангуляции заключается в том, что он пригоден не только для поверхности обкатки, состоящей из фрагментов торов и сфер, но любой гладкой поверхности, в том числе и для любой поверхности уровня. Адаптивность этого алгоритма заключается в том, что размер элемента триангуляции может меняться в зависимости от его местоположения, отражая даже небольшие, но гладкие особенности поверхности; предотвращая «перескок» на несоседние, но близкие участки поверхности, исключая «обрезание» узких перешейков и каналов. Это достигается уменьшением шага сетки триангуляции до величины меньшей двух главных радиусов кривизны поверхности, а также вблизи «активного центра» - места сближения на малое расстояние несоседних участков поверхности.

Полученная триангулированная поверхность может быть использована для демонстрационных целей в молекулярных редакторах (сам алгоритм подходит для триангуляции любой гладкой поверхности, например поверхности уровня) а также и для расчета энергии сольватации для континуальных моделей растворителя.

1.1.2.1 Введение

В данной работе описывается универсальный адаптивный алгоритм триангуляции поверхности и формирования поверхностных элементов. Этот алгоритм будет использоваться в дальнейшем для расчета градиента энергии сольватации по сдвигам атомов на основе нахождения градиентов параметров поверхностных элементов. Важной особенностью и отличием данного алгоритма от многих других алгоритмов триангуляции молекулярной поверхности является то, что границы треугольных поверхностных элементов, лежащих вблизи границ тороидальных и сферических сегментов молекулярной поверхности, в общем случае не лежат на границе между этими сегментами. Это значительно упрощает расчет градиентов параметров таких граничных поверхностных элементов.

Приведем краткий обзор работ, посвященных триангуляции поверхности молекулы. Для построения поверхности молекула субстрата все его атомы окружаются сферами радиуса Ван-дер-Ваальса [21,43-45]. Для получения гладкой поверхности, необходимой для сходимости метода, обычно производится первичная и вторичная «обкатка» поверхности сферами. Затем производится триангуляция поверхности – разбиение на маленькие треугольники. Из этих треугольников в дальнейшем и формируются поверхностные элементы.

Существует два типа гладких поверхностей, окружающих молекулу, построение которых было подробно обсуждено в предыдущих работах [21,29-33,65]:

SAS (Solvent Accessible Surface) - поверхность доступная растворителю образуется центрами молекул растворителя, касающихся молекулы субстрата.

SES (Solvent Excluded Surface) - поверхность исключённого из растворителя объёма. Объем, занимаемый растворителем, лежит *вне* объема, ограниченного этой поверхностью. Сам субстрат полностью лежит *внутри* этого объема

Дальнейшими шагами является триангуляция поверхности и построение на основе этих треугольников поверхностных элементов. Существует широкий класс алгоритмов, предназначенных для триангуляции молекулярных поверхностей. Большая часть из них не является универсальной и привязана к конкретному алгоритму построения поверхности и типу образующихся сегментов поверхности [34-37,69-75]. Универсальные алгоритмы «Марширующие Кубы» (Marching Cubes (MC)) и его усовершенствованный аналог «Дуальное оконтуривание» (Dual contouring (DC)) не учитывают индивидуальных свойств гладкости поверхности, что осложняет построение равномерной сетки [76-78].

В данной работе приводится универсальный алгоритм триангуляции, пригодный не только для триангуляции поверхности обкатки, но и любой поверхности уровня. Он относится к классу алгоритмов – метод распространяющегося фронта (Advancing front technique (AFT)) с дискретизацией поверхности напрямую (Direct surface meshing) [79-84]. Иные используемые названия - метод распространения через предиктор-корректор (continuation method through the predictor-corrector) [85-88], или последовательный процесс формирования дискретных элементов (Mesh Generation using Processing Sequences) [89] или метод «мощения» (Paving: method) [90].

Генерирование триангуляционной сетки осуществляется методом последовательного добавления треугольников на поверхность. При этом используется алгоритм проецирования произвольной точки пространства на ближайшую точку поверхности. В дальнейшем на основе полученных треугольников строятся поверхностные элементы на SES и SAS и приводятся формулы, определяющие параметры этих поверхностных элементов – их координаты, нормали и площади.

Впервые описываемый алгоритм был применен в рамках программы TAGSS (Triangulate Area Grid of Smooth Surface) [29-32]. Он отличается от уже описанных выше алгоритмов некоторыми специфическими сценариями преодоления конфликтов, возникающих при добавлении новых треугольников. Однако ошибки в сценариях добавления нового треугольника приводили к частым сбоям программы. Кроме того, отсутствовал механизм адаптации сетки к изгибам поверхности и наличию близких, но несоседних участков поверхности. Это также приводило к сбоям алгоритма и излишней грубости триангуляции. Эти проблемы были преодолены в рамках усовершенствованной версии программы TAGSS, вошедшей как часть в программу DISOLV [20,33,65]. В этих работах был также добавлен метод построения расчета площади неплоских треугольников на торе, рассмотрен случай нескольких замкнутых поверхностей (кроме внутренних полостей, уже рассмотренных в TAGSS, это случаи нескольких молекул или включений в полости), проведен расчет параметров поверхностных элементов на SAS.

В данной работе мы описываем подробности алгоритма, не изложенные в предыдущих работах [20,33,65] и его дальнейшие усовершенствования. На основе данного алгоритма были создана усовершенствованная версия программы TAGSS, вошедшая как часть в программу DISOLV [20,33,65]

1.1.2.2 Этапы триангуляции поверхностей молекул

1.1.2.2.1 Входные и выходные данные алгоритма

Входными данными для программы являются массивы координат и Ван-дер-ваальсовых радиусов атомов, а также массивы параметров, описывающих поверхностные тороидальные и сферические сегменты, возникшие в результате первичной и вторичной обкатки. Задается также максимальный шаг триангуляции.

Выходными данными являются массивы структур данных описывающих сетку триангуляции поверхности молекулы, массивы параметров описывающих поверхностные элементы, значение площади поверхности.

1.1.2.2.2 Определение собственного базиса молекулы и переход в него. Разбиение пространства вокруг и внутри молекулы на перекрывающиеся кубические области

Под собственным базисом молекулы понимается базис, соответствующий системе координат с центром в геометрическом центре молекулы и осями, направленными вдоль главных осей геометрического тензора инерции молекулы.

Разбиение пространства вокруг и внутри молекулы на кубические области представляет собой процесс запоминания индексов атомов, поверхностных сегментов молекулы и поверхностных элементов в наборе массивов, каждый из которых соответствует кубической области заданного размера и определённого положения. Следует отметить, что кубические области перекрываются между собой

Разбиение на кубические области производится с целью упрощения в дальнейшем процедур поиска индекса атома или элемента поверхности ближайшего к точке с заданными координатами.

Подробности этого алгоритма изложены в работе [21]

1.1.2.2.3 Проверка, что все атомы лежат внутри одной и только одной из внешних молекулярных поверхностей

Все атомы должны лежать внутри одной и только одной из внешних молекулярных поверхностей, которую находит алгоритм построения поверхности [21]. Поэтому после нахождения молекулярных поверхностей разумно совершить проверку этого свойства. Мы можем для этого использовать следующий интеграл:

$$\frac{1}{4\pi} \oint \frac{\mathbf{r} - \mathbf{R}}{|\mathbf{r} - \mathbf{R}|^3} d\mathbf{S} = \begin{cases} 0 & \text{если атом вне интегральной поверхности} \\ 1 & \text{если атом внутри интегральной поверхности} \end{cases} \quad (1)$$

где \mathbf{R} – радиус-вектор атома.

1.1.2.2.4 Алгоритм триангуляции

1.1.2.2.4.1 Алгоритм проецирования: для заданной точки в пространстве определение ближайшей проекции на поверхность молекулы

Найденные параметры основных фрагментов поверхности используются в алгоритме определения ближайшей проекции заданной точки в пространстве на поверхность молекулы, которая, в свою очередь, используется для построения сетки триангуляции поверхности.

Для заданной точки в пространстве определяется номер кубической области, в которой она лежит, далее производится поиск ближайших поверхностных атомов из этой кубической области. Если поверхностных атомов в кубической области найдено не было – поиск среди всех поверхностных атомов. Для найденного ближайшего поверхностного атома производится проверка попадания точки в один из его запрещённых конусов. Если точка не попала ни в один из запрещённых конусов ближайшего атома, то производится проекция этой точки на сферу атома. Если же такое попадание имеет место, то производится перебор сначала первичных тороидальных фрагментов, окружающих данный атом.

Для каждого их перебираемых тороидальных первичных фрагментов производится попытка проецировать точку на его поверхность, с фиксированием расстояния между исходной точкой и точкой проекции. Если проекция попадает в область, запрещенной

вторичной обкаткой, то рассматриваются два выпуклых вторичных сферических сегмента двух сфер устойчивого положения, ограничивающих первичный тор.

Для каждого их перебираемых первичных фрагментов (а в случае необходимости и вторичных фрагментов) производится попытка проецировать точку на его поверхность, с фиксированием расстояния между исходной точкой и точкой проекции. В конце перебора всех первичных торов или соприкасающихся вторичных фрагментов, выбирается наиболее близкая проекция из всех найденных.

Если таковую найти не удалось, начинается перебор вогнутых первичных сферических элементов, окружающих данный атом. Производится попытка проецировать точку на их поверхность, с фиксированием расстояния между исходной точкой и точкой проекции. Если проекция попадает в область, запрещенной вторичной обкаткой, то производится перебор сначала вторичных тороидальных, а затем и выпуклых вторичных сферических элементов, окружающих данный первичный сегмент.

Для каждого их перебираемых первичных фрагментов (а в случае необходимости и вторичных фрагментов) производится попытка проецировать точку на его поверхность, с фиксированием расстояния между исходной точкой и точкой проекции. В конце перебора всех первичных сферических элементов или соприкасающихся вторичных фрагментов, выбирается наиболее близкая проекция из всех найденных.

Используется два типа проецирования: проецирование на сферу с заданным центром и проецирование на тороидальный фрагмент. Если задан центр сферы p_0 радиуса R и некоторая точка r , то проекция на сферу r_s вычисляется по формуле:

$$r_s = p_0 + \frac{r - p_0}{|r - p_0|} R. \quad (2)$$

При этом нормаль поверхности n_s в точке r_s задаётся выражением:

$$n_s = \pm \frac{p_0 - r_s}{|p_0 - r_s|}, \quad (3)$$

где знак «+» выбирается для вогнутых фрагментов, а знак «-» для выпуклых.

Для тороидальных фрагментов определение проекции точки r вычисляется по формулам:

$$\begin{aligned} r_s &= p_0 + \frac{r - p_0}{|r - p_0|} R_{pr}, \\ p_0 &= p_c + h(\cos(\alpha')x + \cos(\beta')y), \\ \cos(\alpha') &= \frac{(r - p_c) \cdot x}{|r - p_c| \sin(\gamma')}, \\ \cos(\beta') &= \frac{(r - p_c) \cdot y}{|r - p_c| \sin(\gamma')}, \\ \cos(\gamma') &= \frac{(r - p_c) \cdot z}{|r - p_c|}. \end{aligned} \quad (4)$$

При этом производится проверка значения угла α' - лежит - ли оно внутри интервала $(\alpha\beta)$ соответствующего свободной обкатке вокруг пары атомов. Для первичного тора

обкатки проверяется (при наличии вторичной обкатки), что угол γ' меньше критического угла γ , определяемого вторичными сферами устойчивого положения. Для нормали используется выражение:

$$n_s = \pm \frac{p_0 - r_s}{|p_0 - r_s|}, \quad (5)$$

где + для первичных торов, а – для вторичных торов.

1.1.2.2.4.2 Механизмы адаптивного регулирования шага сетки

Пусть L – адаптивный шаг сетки. Шаг сетки - это радиус окружности L , описанной вокруг равностороннего треугольника с высотой, равной $Rch = 1.5L$. Этот треугольник определяет максимальный размер триангуляции. Этот максимальный размер L^{max} задается пользователем. Однако он должен меняться вблизи тех мест, где два несоседних участка поверхности подходят близко друг к другу. Такие места задаются с помощью «центров адаптации». «Центр адаптации» – это точка, вблизи которой происходит изменение (адаптация) стандартного максимального шага сетки. Критическое расстояние от «центров адаптации» – расстояние от поверхностного элемента до центра адаптации, при котором происходит изменение шага сетки. Для каждого центра адаптации определяется его максимальный шаг сетки. Центры адаптации для молекулярной поверхности обкатки [21] подразделяются на следующие виды

- 1) центры самых узких мест (шириной δ) первичных торов (центр их «перешейков»).
- 2) центры самых узких мест (шириной δ) вторичных торов (центр их «перешейков»).
- 3) место максимального сближения (на расстояние δ) первичных сферических сегментов.
- 4) место максимального сближения (на расстояние δ) вторичных сферических сегментов.

Критическое расстояние от «центров адаптации» - т.е. «радиус адаптации» определяется следующей формулой:

$R^a \geq 2(Rch \cdot R_{mr})^{1/2}$, где R_{mr} – максимальный радиус кривизны в месте максимального сближения. Заметим, что $R_{mr} \leq R_{pr}$, где R_{pr} первичный радиус обкатки. Поэтому можно положить $R^a = 2(Rch \cdot R_{pr})^{1/2}$.

Изменение шага сетки на расстоянии меньшего R^a дается формулой: $L = \delta/4$

Каждому кубу соотносятся все «центры адаптации» лежащие от него на расстоянии меньшем или равным R^a .

Следует отметить, что поверхность имеет тонкости структуры, которые мы не хотим потерять из-за большого шага сетки. В связи с этим, мы уменьшаем шаг сетки до размеров меньших половины главных радиусов кривизны в вершинах соответствующего треугольника сетки. Кроме того, все вершины одного треугольника должны лежать либо на одном и том же, либо на соседних по отношению друг к другу сегментах (тороидальных или сферических). Подробнее часть алгоритма, обеспечивающая эти условия, будет описана в сценариях триангуляции.

1.1.2.2.4.3 Определение положения затравочного треугольника

Положение затравочного треугольника определяет начало построения сетки триангуляции. С точки зрения алгоритма построения это положение может быть выбрано произвольно. Задаётся некоторая удалённая от молекулы точка в пространстве, для

которой определяется соответствующая проекция на поверхность. Затем вокруг найденной точки проекции выстраивается равносторонний треугольник с размером радиуса описанной окружности примерно равной величине заданной пользователем. Данные о первом треугольнике помещаются в массивы структур данных описывающих сетку триангуляции. В результате работы процедуры построения затравочного треугольника образуется сетка триангуляции, состоящая из одного треугольника, трёх вершин и трёх граничных рёбер.

1.1.2.2.4.4 Алгоритм триангуляции после создания затравочного треугольника. Генерирование сетки триангуляции методом последовательного добавления треугольников

1.1.2.2.4.4.1 Алгоритм генерирования сетки

Алгоритм генерирования сетки имеет следующий вид:

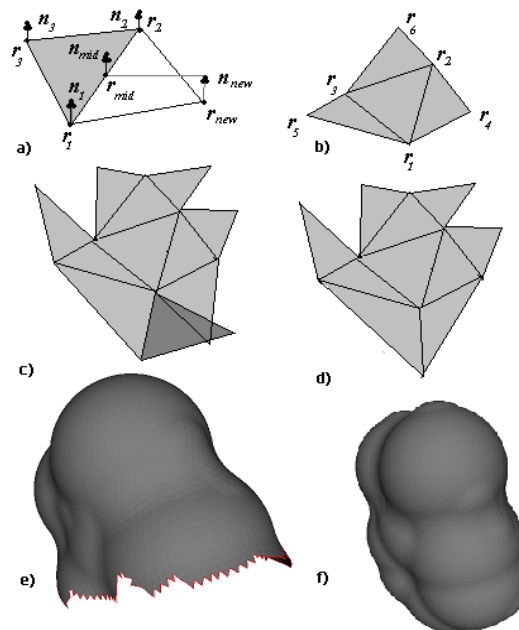


Рис. 1 Описание метода триангуляция поверхности последовательным добавлением новых треугольников.

- (a) Добавление нового треугольника: Определение новой точки поверхности (r_4), которая вместе двумя концевыми точками (r_1, r_2) ребра (r_1) образует новый треугольник.
- (b) После обхода всех рёбер первого треугольника получаем фрагмент поверхности из 4 треугольников и новое множество граничных рёбер.
- (c) Последовательно производим обход граничных рёбер и добавляем к каждому по новому треугольнику. При этом может возникнуть наложение треугольников.
- (d) Для случаев наложения треугольников применяется ряд сценариев для устранения конфликтов, например, сшивание двух треугольников общим ребром.
- (e) Повторяя процедуру обхода граничных ребер, выполняется послойное построение поверхности. После каждой процедуры обхода образуется новый набор граничных рёбер. Граничные рёбра могут образовывать несколько замкнутых ломаных.
- (f) Построенная поверхность. Поверхность считается построенной, когда не остаётся ни одного граничного ребра.

Алгоритм генерирования сетки триангуляции поверхности (Рис. 1) представляет собой циклический процесс обхода граничных рёбер с добавлением к каждому граничному ребру нового треугольника. Работа алгоритма начинается с обхода граничных рёбер затравочного треугольника, которые образуют текущий набор граничных рёбер. При этом формируется новый набор граничных рёбер, который формирует последующий набор текущих рёбер. По завершении обработки текущего набора, номера рёбер из последующего набора перемещаются в массив текущих граничных рёбер, после чего массив индексов последующих граничных рёбер обнуляется. Цикл обработки повторяется заново. В процессе добавления треугольников неизбежно возникают случаи их наложения и пересечения. Для разрешения конфликтных ситуаций в процессе добавления очередного треугольника производится анализ геометрической конфигурации соседних треугольников и выбор одного из возможных способов добавления треугольника или видоизменения геометрической конфигурации текущих граничных рёбер. При этом текущее множество граничных рёбер может разбиваться на несколько множеств геометрически связанных граничных рёбер, каждое из которых образует замкнутую ломаную. Таким образом, граница уже триангулированных областей – множество, состоящее из одной или нескольких замкнутых ломаных. Делаем перебор (цикл), по элементам этого множества ломаных, затем перебор (цикл) по всем ребрам ломаных до тех пор, пока это множество ломаных не исчезнет. На каждом шаге этих циклов мы применяем один из сценариев добавления треугольника или видоизменения геометрической конфигурации текущих граничных рёбер. Процесс генерирования сетки оканчивается тогда, когда не остаётся ни одного граничного ребра.

В результате работы процедуры для генерирования сетки триангуляции получаем заполненные массивы трёх типов структур данных (Point (вершины), Edge (углы), Triangle (треугольники)), описывающих сетку триангуляции и связи между этими элементами.

1.1.2.2.4.4.2 Процедура обхода рёбер текущей замкнутой ломаной. Цикл перебора текущих рёбер и выбор подходящего сценария

Список всех возможных сценариев приведен ниже описания алгоритма:

1. Выбираем текущее ребро следующим образом:
Предположим что ранее, после добавления нового треугольника текущая замкнутая ломаная разделилась на две новые замкнутые ломаные с одной общей вершиной. Тогда в качестве текущего ребра выбираем одно из рёбер, имеющих одним из концов общую точку этих двух образовавшихся ломаных. Целью такого выбора является пространственное разделение двух ломаных. Действительно, добавление новых треугольников приводит к исчезновению их общей точки. После образования двух не связанных ломаных прекращаем выполняемый цикл обработки текущей ломаной и идем на начальный пункт перебора всех ломаных. Если разделения ломанной на две не произошло, то берем следующее ребро после текущего в текущем множестве рёбер.
2. Пусть текущее множество, включающее текущее ребро, состоит из трех рёбер. Тогда используем сценарий 10 (и идем на пункт 1).
3. Пусть текущее множество, включающее текущее ребро, состоит из четырех рёбер.
Пусть у образованного ими четырехугольника имеется «особая» вершина, обладающая двумя свойствами. Во-первых, из нее выходит только два ребра. Во-вторых, на нее опирается только один треугольник, образованный этими двумя ребрами. Если двухгранный угол между этой треугольной гранью с «особой» вершиной и гранью с противоположной ей вершиной четырехугольника меньше $\pi/6$, то используем сценарий 9 (и идем на пункт 1), а если больше – то сценарий 8 (и идем на пункт 1). Если «особой» точки нет, то используем сценарий 7 (и идем на пункт 1).

4. Один из углов текущего ребра с одним из двух смежных ребер из текущего множества не маленький ($\alpha_1 > \pi/9$), а с другим маленький ($\alpha_2 < \pi/9$):
Если разные концы текущего ребра и смежного ребра, образующие малый угол не связаны еще одним путем из двух ребер, то используем сценарий 6 (и идем на пункт 1). Иначе используем сценарий 5 (и идем на пункт 1).
5. Оба угла α_1 и α_2 – маленькие ($\alpha_1 < \pi/9$ и $\alpha_2 < \pi/9$), то используем сценарий 1 (и идем на пункт 1).
6. Строим «новый» треугольник и «новую» точку по сценарию 2, но пока не подтверждаем их построение в качестве узла и треугольника сетки.
7. Пусть углы β_1 и β_2 – углы между ребрами, смежными к текущему ребру, и соответствующими смежными к ним сторонами «нового» треугольника
8. Пусть хотя бы один из углов β_1 или α_1 мал, а углы β_2 и α_2 – велики. ($[(\beta_1 < \pi/6$ или $\alpha_1 < 2\pi/9)$ и $(\beta_2 > \pi/6$ и $\alpha_2 > 2\pi/9)]$). Или, наоборот, хотя бы один из углов β_2 или α_2 мал, а углы β_1 и α_1 – велики ($[(\beta_2 < \pi/6$ или $\alpha_2 < 2\pi/9)$ и $(\beta_1 > \pi/6$ и $\alpha_1 > 2\pi/9)]$). Тогда применяем сценарий 4 для текущего ребра и смежного ребра, образующего малый угол (и идем на пункт 1).
9. Пусть углы β_1 и β_2 малы ($\beta_1 < \pi/6$ и $\beta_2 < \pi/6$). Тогда применяем сценарий 4 для текущего ребра и смежного ребра, образующего меньший угол (и идем на пункт 1).
10. Пусть все определенные выше углы $\beta_1, \alpha_1, \beta_2$ и α_2 велики ($\beta_1 > \pi/6, \alpha_1 > 2\pi/9, \beta_2 > \pi/6$ и $\alpha_2 > 2\pi/9$). Пусть \mathbf{n}_{mid} – нормали к поверхности в точке проекции середины текущего ребра, \mathbf{n}_{new} – нормаль к поверхности в «новой» точке «нового» треугольника. Пусть γ – угол между \mathbf{n}_{new} и \mathbf{n}_{mid} . Если угол γ велик ($\gamma > \pi/2$), то применяем сценарий 1 (и идем на пункт 1).
11. Пусть хотя бы один из определенных выше углов $\beta_1, \alpha_1, \beta_2$ и α_2 мал ($\beta_1 < \pi/6$, или $\alpha_1 < 2\pi/9$, или $\beta_2 < \pi/6$, или $\alpha_2 < 2\pi/9$). Тогда выполняем сценарий 1 (и идем на пункт 1).
12. Находим «первую особую» точку, близкую к двум вершинам текущего ребра по формуле:

$$\mathbf{r}^{(1)} = \frac{\mathbf{r}_1 + \mathbf{r}_2 + 2\mathbf{r}_{new}}{4}, \quad (6)$$

\mathbf{r}_{new} – радиус-вектор «новой» точки.

$\mathbf{r}_1, \mathbf{r}_2$ – радиус-вектора двух вершин текущего ребра. Если для всех точек (в кубической области, содержащей «первую особую» точку) Формируем множество всех узлов сетки, которые обладают следующими свойствами:

а. Лежат в кубической области, содержащей «первую особую» точку

в. Не совпадают с вершинами текущего ребра и вершинами двух смежных с ним граничных ребер

Находим «вторую особую» точку. Эта точка из найденного множества, имеющая минимальное расстояние $\|r_{min}\|$ до «первой особой» точки.

13. Если расстояние $\|r_{min}\| > Rch$, то подтверждаем построение «нового» треугольника по сценарию 2 (и идем на пункт 1). Если $\alpha_1 < \pi/2$ или $\alpha_2 < \pi/2$, то выполняем сценарий 1 (и идем на пункт 1).
14. Если «вторая особая» точка не лежит на текущей границе, то выполняем сценарий 1 (и идем на пункт 1).
15. Если $\|r_{min}\|$ меньше половины текущего ребра, то выполняем сценарий 1 (и идем на пункт 1).
16. Пусть \mathbf{n}_m – нормаль во «второй особой» точке. Если угол между \mathbf{n}_{new} и \mathbf{n}_m больше $\pi/2$, то выполняем сценарий 1 (и идем на пункт 1).
17. Выполняем сценарий 3. Образуются две замкнутые ломаные граничных ребер с общей точкой. Идем на пункт 1.

1.1.2.2.4.4.3 Основные сценарии добавления нового треугольника или обработки текущих граничных рёбер

Основные сценарии (Рис. 3) добавления нового треугольника или обработки текущих граничных рёбер:

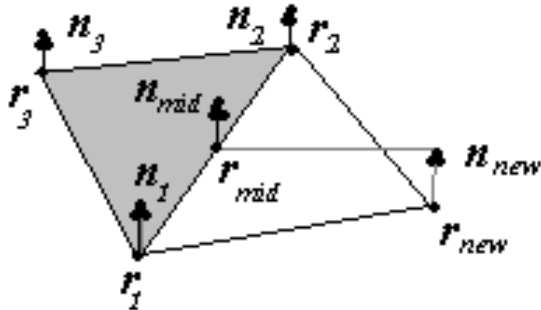


Рис. 2 Определение новой точки поверхности (r_{new}), которая вместе двумя концевыми точками (r_1, r_2) ребра образует новый треугольник.

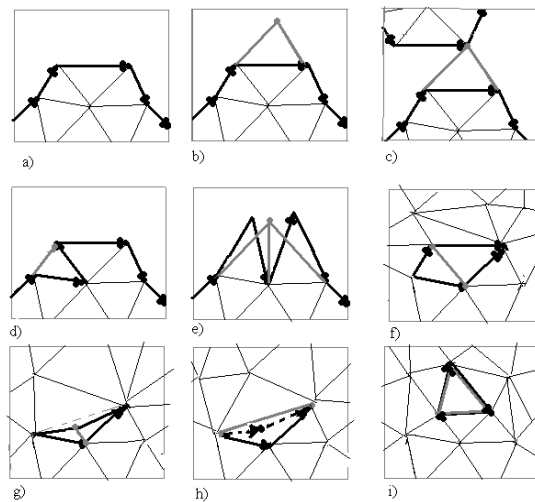


Рис. 3 Схематическая иллюстрация геометрических конфигураций сетки триангуляции для применяемых сценариев обработки граничных рёбер. Жирными отрезками со стрелками указаны граничные рёбра текущего набора и направление их обход. Пунктиром указаны новые граничные или внутренние рёбра.

(a) Ничего не делать. Когда все другие сценарии не сработали. *Сценарий 1*

(b) Треугольник образуется текущим граничным ребром и двумя новыми граничными рёбрами. Пунктиром показаны новые граничные рёбра. *Сценарий 2*

(c) Построение треугольника, геометрически объединяющего или разделяющего текущие граничные массивы. *Сценарий 3*

(d) Треугольник образуется текущим и последующим (предыдущим) граничным рёбрами текущего граничного набора и одним новым граничным ребром. Пунктиром показано новое граничное ребро образующее новый треугольник. *Сценарии 4-5.*

(e) Слияние двух граничных рёбер образующих острый угол между собой в одно внутреннее ребро. *Сценарий 6.*

- (f) Цикл из четырех ребер разбивается на два треугольника новым построенным пятым ребром, соединяющий противоположные тупые углы четырехугольника. *Сценарий 7*
- (g) Цикл из четырех ребер разбивается на два треугольника дополнительным пятым ребром, соединяющий «особую» вершину с противоположной ей вершиной. *Сценарий 8*
- (h) «Особая» вершина цикла из четырех ребер выбрасывается. Из оставшихся трех вершин формируем треугольник. *Сценарий 9*
- (i) Построение треугольника закрывающего граничное множество, состоящее из трёх рёбер. *Сценарий 10*

- 1) Ничего не делать. Когда все другие сценарии не сработали.
- 2) Строится «новый» треугольник, опирающийся на текущее ребро. (Рис. 2) Делается это следующим образом. Проецируем центр текущего ребра на поверхность. Из этой точки строим вектор, перпендикулярный нормали в этой точке и вектору текущего ребра и длиной равный $Rch = 1.5L$. Проецируем полученную точку на поверхность. Это «новая» точка. Строим «новый» треугольник из текущего ребра и полученной «новой» точки. L – это адаптивный шаг сетки. Адаптивный шаг сетки - это радиус окружности, описанной вокруг равностороннего треугольника с высотой $Rch = 1.5L$. Этот треугольник определяет максимальный размер триангуляции в данном месте. Адаптация шага сетки идет по следующему алгоритму. Если к кубу, определяемого серединой текущего ребра, не относится ни один «центр адаптации», то этот шаг сетки определяется заранее заданным максимальным размером $L = L^{max}$. Если такие центры есть, то считается расстояние от каждого из этих «центров адаптации», относящихся к кубу, до середины текущего ребра $\|r_j\|$, ($j=1, \dots, N_a$ – номер центра адаптации). Находим среди этих расстояний те, которые меньше критического для соответствующего «центра адаптации» $\|r_j\| < R_j^a$. Каждому такому «центру адаптации» j соответствует свой шаг сетки $L = L_j^a$. Выбираем среди них минимальный шаг $L = L^{min}$. Каждой двух вершин текущего ребра (впрочем, как и любой точке поверхности) соответствует два главных радиуса кривизны R^{g1} , R^{g2} . Для тора это радиусы двух образующих тора в этой точке (один из них всюду одинаков, другой увеличивается от центра к краям тора). Для сферического сегмента оба главных радиуса равны его радиусу. Корректируем шаг сетки, чтобы он был не больше половины этих радиусов для обеих вершин:

$$L = \min\left(L^{min}, R_1^{g1}/2, R_1^{g2}/2, R_2^{g1}/2, R_2^{g2}/2\right). \quad (7)$$

Далее строим «новую» точку с таким шагом сетки. Пусть «новая» точка лежит на сегменте (тороидальном или сферическом) S_N . Вершины текущего ребра лежат на сегментах S_1, S_2 . Проверяем, что S_N либо совпадает с S_1 или S_2 , либо является соседним для обоих из них. Для «новой» точки находим ее два главных радиуса кривизны R_N^{g1}, R_N^{g2} . Проверяем, что $L < R_N^{g1}/2$, $L < R_N^{g2}/2$. Если хотя бы одно из этих условий не выполняется, то шаг сетки уменьшается следующим образом:

$$L = \min\left(L/2, R_N^{g1}/2, R_N^{g2}/2\right). \quad (8)$$

Далее строим «новую» точку с таким шагом сетки и снова проверяем описанные выше условия. Этот процесс продолжается до тех пор, пока эти условия не выполняются.

- 3) «Новая» точка - а, построенная по методу 2, «сливается» в одну точку со «второй особой» точкой (узлом сетки) - б. «Слияние» идет по формуле:

$$\mathbf{r} = \frac{(N_a \mathbf{r}_a + N_b \mathbf{r}_b)}{(N_a + N_b)}, \quad (9)$$

$N_a=2$ - число ребер, выходящих из точки а; N_b - число ребер, выходящих из точки б;

\mathbf{r}_a – радиус-вектор точки а;

\mathbf{r}_b - радиус-вектор точки б;

\mathbf{r} - радиус-вектор образующейся в результате слияния точки;

- 4) Строится новый треугольник. Он образуется, во-первых, текущим ребром. Во-вторых, граничным ребром, смежным к текущему ребру и имеющим с ним малый угол. И, в-третьих, одним новым граничным ребром, построенным напротив этого маленького угла. Пусть а - «новая точка», построенная по сценарию 2. Вторая точка б смежного ребра (не принадлежащая текущему ребру) сдвигается по формуле:

$$\mathbf{r} = \frac{(N_a \mathbf{r}_a + N_b \mathbf{r}_b)}{(N_a + N_b)}, \quad (10)$$

$N_a=2$ - число ребер, выходящих из точки а; N_b - число ребер, выходящих из точки б;

\mathbf{r}_a – радиус-вектор точки а; \mathbf{r}_b - радиус-вектор начального положения точки б; \mathbf{r} - радиус-вектор нового положения точки б;

- 5) Строится новый треугольник, образуемый текущим ребром, одним из граничных ребер, смежных к нему и имеющий с ним малый угол, и одним новым граничным ребром, построенным напротив этого маленького угла.
- 6) Слияние двух смежных граничных ребер, имеющих общую точку и образующих малый острый угол между собой в одно внутреннее (не граничное) ребро. Две различающиеся точки смежных ребер (точки а и б) «сливаются» в одну точку. Радиус-вектор получающейся точки считается по следующей формуле:

$$\mathbf{r} = \frac{(N_a \mathbf{r}_a + N_b \mathbf{r}_b)}{(N_a + N_b)}, \quad (11)$$

N_a - число ребер, выходящих из точки а; N_b - число ребер, выходящих из точки б;

\mathbf{r}_a – радиус-вектор точки а;

\mathbf{r}_b - радиус-вектор точки б;

\mathbf{r} - радиус-вектор образующейся в результате слияния точки;

- 7) Цикл из четырех ребер разбивается на два треугольника новым построенным пятым ребром, соединяющий противоположные тупые углы четырехугольника.
- 8) Цикл из четырех ребер разбивается на два треугольника дополнительным пятым ребром, соединяющий «особую» вершину с противоположной ей вершиной.
- 9) «Особая» вершина цикла из четырех ребер выбрасывается. Из оставшихся трех вершин формируем треугольник.
- 10) Построение треугольника закрывающего граничное множество, состоящее из трёх ребер. Проводится специальная проверка, чтобы не применять этот метод для первого шага – когда обрабатывается граничное множество затравочного треугольника.

Наиболее важен первый сценарий. Здесь производится вычисление координат новой вершины. Метод этого вычисления пояснён на Рис. 2.

На Рис. 3 схематично представлены геометрические конфигурации применения некоторых важных сценариев. Граничные рёбра считаются направленными отрезками – в направлении, совпадающем с направлением их обхода.

1.1.2.2.4.4.4 Окончательная «утряска» сетки

Для создания более однородной сетки используется механизм «утряски». Перебираются все узлы сетки. Для всех ближайших соседей текущего узла находится «центр тяжести». Затем он проецируется на поверхность и заменяет собой этот текущий узел. Процесс утряски может проводиться несколько раз.

1.1.2.2.4.5 Формирование массивов параметров поверхностных элементов на основании данных о триангуляции поверхности

На основании данных о сетке триангуляции поверхности производится формирование массивов данных о поверхностных элементах. Каждый поверхностный элемент характеризуется координатами в пространстве, направлением нормали поверхности и его площадью. Координаты и нормаль соответствуют узлам, полученной сетки триангуляции. Каждый такой узел рассматривается как центр будущего многоугольного поверхностного элемента. А для вычисления площади треугольного поверхностного элемента применяются следующие методы:

1.1.2.2.4.5.1 Формулы для площадей треугольных поверхностных элементов

Приведем формулы для площадей треугольных поверхностных элементов. Их можно разделить на три типа – сферические, тороидальные и граничные, чьи точки принадлежат разным поверхностным фрагментам.

1) *Площадь сферического треугольника. (Рис. 4)*

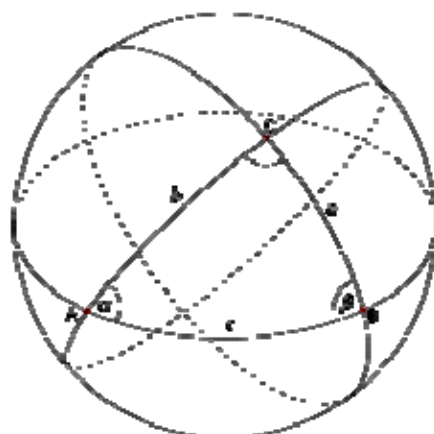


Рис. 4 Сферический треугольник

Сумма s углов сферического треугольника α, β, γ :

$$s = \alpha + \beta + \gamma \quad (12)$$

всегда меньше 3π и больше π . Величина

$$\varepsilon = s - \pi \quad (13)$$

называется сферическим избытком. Площадь сферического треугольника S определяется по формуле Жирара:

$$S = R^2 \varepsilon, \quad (14)$$

где R - радиус сферы.

2) Площадь тороидального треугольника. (Рис. 5)

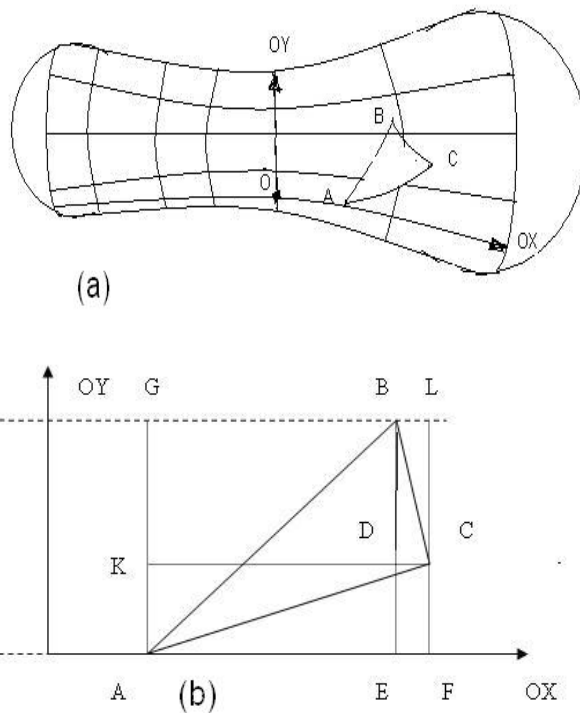


Рис. 5 (a)Тороидальный треугольник ABC (b) Его отображение на плоскость, пропорции искажены

Рассмотрим тороидальный треугольник ABC на торе.

Определим криволинейную систему координат, образованной образующими тора двух типов:

OX, OY –оси, идущие вдоль образующих (OY – круги, перпендекулярные оси тора, OX – точки касания между тором и сферой обкатки)

При этом возьмем некие три точки A, B, C. Базисные оси, идущие вдоль образующих OX, OY могут всегда быть выбраны следующим образом:

- 1) Точка C лежит между точками A и B по оси OY.
- 2) Положительное направление оси OX от отрезка [AB] к точке C

- 3) Для точек А (x_A, y_A) и В (x_B, y_B) $x_B > x_A, y_B > y_A$
- 4) Линии, соединяющие пары точек АС, АВ, СВ мы можем выбрать не всегда совпадающими с геодезическими тора. Выберем их как отрезки линий между точками, которые являются образом геодезических линий цилиндра, получающимся отображением тора на этот цилиндр. Это отображение делается так : образующие тора ОХ отображаются на образующие цилиндра, параллельные оси цилиндра, с сохранением длины, образующие тора ОУ отображаются на образующие цилиндра, перпендикулярные оси цилиндра, с сохранением угловых размеров.
- 5) α_A - угловое положение вдоль образующей ОХ $x_A = R_{OX} \alpha_A$; R_{OX} - радиус сферы обкатки тора; h - расстояние от центра сферы обкатки до прямой, соединяющей опорные сферы тора.
- 6) φ_{AB} - угловое расстояние между точками А и В вдоль образующей ОУ

Тогда площадь тороидального треугольников и четырехугольников:

Площадь квадрата EAGB:

$$S_{EAGB} = \varphi_{AB} R_{OX} (h\alpha - R_{OX} \sin(\alpha)) \Big|_{\alpha_A}^{\alpha_B} . \quad (15)$$

Площадь треугольников ABE, BCD и ACF:

$$S_{ABE} = \varphi_{AB} \frac{R_{OX}}{\alpha_B - \alpha_A} \left(h \frac{(\alpha - \alpha_A)^2}{2} - R_{OX} ((\alpha - \alpha_A) \sin(\alpha) + \cos(\alpha)) \right) \Big|_{\alpha_A}^{\alpha_B} , \quad (16)$$

$$S_{BCD} = \varphi_{BC} \frac{R_{OX}}{\alpha_B - \alpha_C} \left(h \frac{(\alpha - \alpha_C)^2}{2} - R_{OX} ((\alpha - \alpha_C) \sin(\alpha) + \cos(\alpha)) \right) \Big|_{\alpha_B}^{\alpha_C} , \quad (17)$$

$$S_{ACF} = \varphi_{AC} \frac{R_{OX}}{\alpha_A - \alpha_C} \left(h \frac{(\alpha - \alpha_C)^2}{2} - R_{OX} ((\alpha - \alpha_C) \sin(\alpha) + \cos(\alpha)) \right) \Big|_{\alpha_A}^{\alpha_C} . \quad (18)$$

Площадь треугольника ABC:

$$S_{ABC} = S_{ABE} - S_{ACF} + \text{sign}(x_C - x_B)(S_{EDCF} + S_{BCD}) . \quad (19)$$

3) Площади треугольников, чьи вершины не принадлежат одному и тому же поверхностному фрагменту.

Треугольник образован вершинами r_1, r_2, r_3 . Используем для расчета их площади S формулу площади плоского треугольника.

Стороны треугольника:

$$a = r_2 - r_1 \quad b = r_3 - r_1 . \quad (20)$$

Площадь треугольника S:

$$S = \frac{1}{2} |[\mathbf{a} \times \mathbf{b}]|. \quad (21)$$

1.1.2.2.4.5.2 Формирование многоугольных поверхностных элементов и определение их параметров

На основе треугольных поверхностных элементов формируются *многоугольные* поверхностные элементы. Центры многоугольных элементов совпадают с вершинами треугольных элементов. Вершины многоугольных элементов – центры сторон треугольных элементов, выходящие из центров многоугольных элементов и центры тяжести (точка пересечения медиан) этих треугольных элементов.

Площадь многоугольного элемента S^M - сумма третей от площадей треугольных элементов S^{tr}_i с вершиной в центре многоугольного элемента. Для каждого узла сетки триангуляции производится суммирование площадей тех треугольников, для которых данный узел является вершиной. Полученная суммарная площадь делится на три, полученный результат запоминается как площадь поверхностного элемента с координатами и нормалью соответствующих данному узлу сетки триангуляции

$$S^M = \frac{\sum_i S^{tr}_i}{3}. \quad (22)$$

1.1.2.2.4.5.3 Поверхностные элементы на SAS

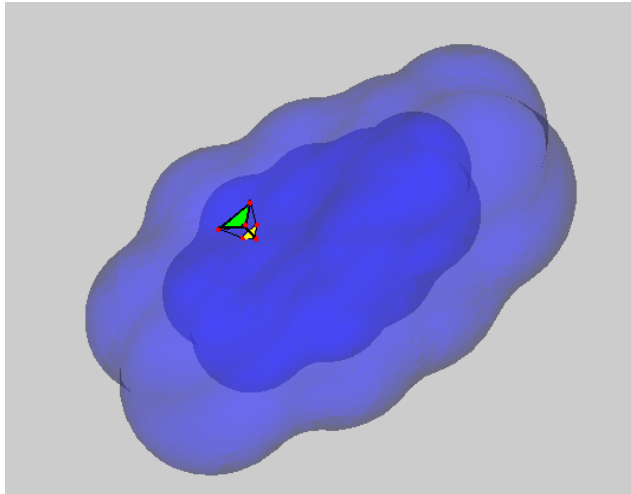


Рис. 6 Преобразование поверхности типа SES в поверхность типа SAS

Пусть имеется поверхность после первичной обкатки.

Пусть имеется треугольник на поверхности SES ($\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$)

Образ этого треугольника на SAS ($\mathbf{r}_{1n}, \mathbf{r}_{2n}, \mathbf{r}_{3n}$) (Рис.6)

$$\mathbf{r}_{1n} = \mathbf{r}_1 + \mathbf{n}_1 R_{pr}, \quad (23)$$

$$\mathbf{r}_{2n} = \mathbf{r}_2 + \mathbf{n}_2 R_{pr}, \quad (24)$$

$$\mathbf{r}_{3n} = \mathbf{r}_3 + \mathbf{n}_3 R_{pr}, \quad (25)$$

где R_{pr} – радиус первичной обкатки, $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ -нормали в соответствующий точках.

Тогда

- 1) поверхностный элемент на первичной сфере, опирающейся на три атома, отображается в точку. Его площадь на SAS нулевая
- 2) поверхностный элемент на торе, опирающимся на два атома, отображается в линию. Его площадь на SAS нулевая
- 3) поверхностный элемент на атоме отображается в сферический треугольник. Его площадь на SAS пропорциональна его площади на SES: $S_{ses} \left((R_{atom} + R_{pr}) / R_{atom} \right)^2$
- 4) граничный поверхностный элемент между тором и первичной сферой, опирающейся на три атома – нулевая площадь
- 5) граничный треугольный поверхностный элемент, опирающийся на первичный тор и атом, первичную сферу и атом или первичные тор, сферу и атом – площадь S считается по формуле для трех точек ($\mathbf{r}_{1n}, \mathbf{r}_{2n}, \mathbf{r}_{3n}$) :

$$\mathbf{a} = \mathbf{r}_{2n} - \mathbf{r}_{1n}, \quad (26)$$

$$\mathbf{b} = \mathbf{r}_{3n} - \mathbf{r}_{1n}, \quad (27)$$

$$S = |\mathbf{a} \times \mathbf{b}| / 2. \quad (28)$$

1.1.2.2.4.5.4 Вычисление площади и объема поверхностей SES и SAS

Вычисление площади поверхности производится суммированием площадей всех поверхностных элементов

$$S = \oint_S dS. \quad (29)$$

Вычисление объема поверхности производится по следующей формуле:

$$V = \frac{1}{3} \oint_S (\mathbf{r} \cdot \mathbf{n}) dS. \quad (30)$$

\mathbf{r} – радиус-вектор текущей точки поверхности, \mathbf{n} - ее нормаль

1.1.2.3 Выводы

Представленный алгоритм и основанная на нем программа позволяют быстро и надежно строить триангуляцию гладкой поверхности. Поверхность получается раскрашенной в зависимости от типа ближайших атомов, что удобно для визуализации, и триангулированной, что позволяет не только вычислять ее поверхность и ограниченный ею объем, но и решать с хорошей точностью интегральные уравнения, используемые в континуальной модели растворителя. Полученная с помощью данной программы поверхность может применяться как для целей визуализации молекулы, это особенно

актуально для больших белковых молекул, так и для целей вычисления сольватационных вкладов в энергию взаимодействия молекул друг с другом при наличии внешней среды.

Алгоритм триангуляции может быть легко обобщен на случай любой гладкой поверхности, в том числе гладкой поверхности уровня.

Глава 1.1.3 Континуальная модель среды III: Вычисление аналитических градиентов параметров поверхностных элементов на молекулярных поверхностях по координатам атомов.

Задача нахождения аналитических градиентов (производных по координатам атомов) энергии сольватации распадается на две подзадачи: на первом этапе мы ищем параметры поверхностных элементов (три координаты, три компонента единичного вектора нормали и площадь) и их производные, на втором этапе мы дифференцируем энергию и выражаем ее через производные от матриц, описывающих задачу. При этом производные матричных элементов выражаются через производные параметров поверхностных элементов SES (поверхность исключённого из растворителя объёма) или SAS (поверхность доступная растворителю). Целью данной работы является нахождение именно этих аналитических градиентов от параметров поверхностных элементов [91].

1.1.3.1 Введение

Целью данной работы является нахождение аналитических градиентов параметров поверхностных элементов на поверхности исключенного объема и поверхности доступной растворителю. Под аналитическими градиентами подразумеваются частные производные от параметров по координатам атомов, составляющих молекулу. Нахождение этих градиентов необходимо для нахождения аналитических градиентов свободной энергии сольватации. Эти градиенты могут использоваться в дальнейшем для поиска глобального минимума энергии, что необходимо, например, в компьютерном моделировании лекарств.

Прежде чем искать градиенты поверхностных элементов, необходимо определить для какой поверхности это делается. Существуют два вида поверхности вокруг молекулы. Во-первых, это SES (Solvent Excluded Surface) - поверхность исключённого из растворителя объёма. Объем, занимаемый растворителем, лежит *вне* объема, ограниченного этой поверхностью. Сам субстрат полностью лежит *внутри* этого объема. Во-вторых, это SAS (Solvent Accessible Surface) - поверхность доступная растворителю образуется центрами молекул растворителя, касающихся молекулы субстрата. Первый тип поверхности используется для расчета поляризационной составляющей энергии сольватации, а второй – для расчета кавитационной и ванн-дер-ваальсовской составляющих [45].

Проведенные ранее работы по аналитическим градиентам параметров SES относились к алгоритму, реализованному в программе GEPOLE [37], где пустоты в молекуле заполняются фиктивными сферами. Для гладкой поверхности, полученной первичной и вторичной обкаткой алгоритмом, описанным в [20,21,29-33,65,68], такая работа продельвается в подробностях впервые в настоящей части. В работах [20,21,29-33,65,68] уже описаны алгоритмы построения гладкой поверхности, метод ее триангуляции и разбиения на поверхностные элементы (Рис.1) и методы нахождения их параметров. Здесь же остановимся подробно на нахождении аналитических градиентов. Расчеты аналитических градиентов поверхностных элементов, полученной методами первичной и вторичной обкатки [20,21,29-33,65,68] реализованы в программе DISOLV [20,21,33,65,68]

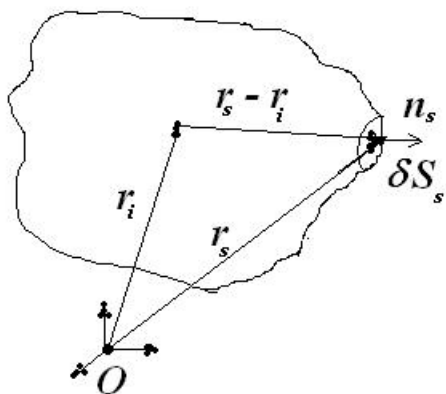


Рис.1 Элемент поверхности относительно начала координат.

1.1.3.2 Аналитические градиенты параметров поверхностных элементов

Найдем формулы, описывающие производные от параметров поверхностных элементов (координат, нормалей, площадей) по координатам атомов для SES и SAS. Рассмотрим вначале поверхность SES. У нас имеются два типа сегментов этой поверхности – фрагменты сфер и фрагменты торов. Рассмотрим, как меняются при сдвиге атомов параметры поверхностного элемента, полностью лежащего на одном из таких фрагментов.

1.1.3.2.1 Поверхностный элемент SES лежит на сфере. (Рис.2)

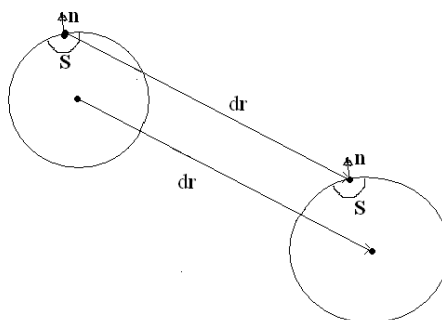


Рис.2 Изменение параметров поверхностного элемента сферы при сдвиге сферы. Видно, что меняется только его координаты и их изменения равны сдвигу центра сферы. Площадь и нормаль постоянны.

Изменения координат поверхностных элементов равны сдвигу центра сферы. Площадь и нормаль постоянны.

$$\begin{aligned} \frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} &= \frac{\partial \mathbf{r}_m}{\partial x_i^k}, \\ \frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} &= 0, \\ \frac{\partial S_s}{\partial R_{i,x^k}} &= 0, \end{aligned} \tag{1}$$

где \mathbf{n}_s - нормаль поверхностного элемента; S_s - площадь поверхностного элемента; \mathbf{r}_s - радиус-вектор координат поверхностного элемента; \mathbf{r}_m - центр сферического сегмента; R_{i,x^k} - k -ая компонента радиус-вектора R_i ($k=1,2,3$); i - номер сдвигаемого атома ($0 < i < N$); N число атомов.

1.1.3.2.2 Поверхностный элемент SES лежит на торе

Градиенты параметров поверхностных элементов являются линейной суперпозицией градиентов для следующих двух случаев:

1.1.3.2.2.1 Сдвиг одной из опорных сфер (r_2) просходит перпендекулярно оси тора (Рис.3).

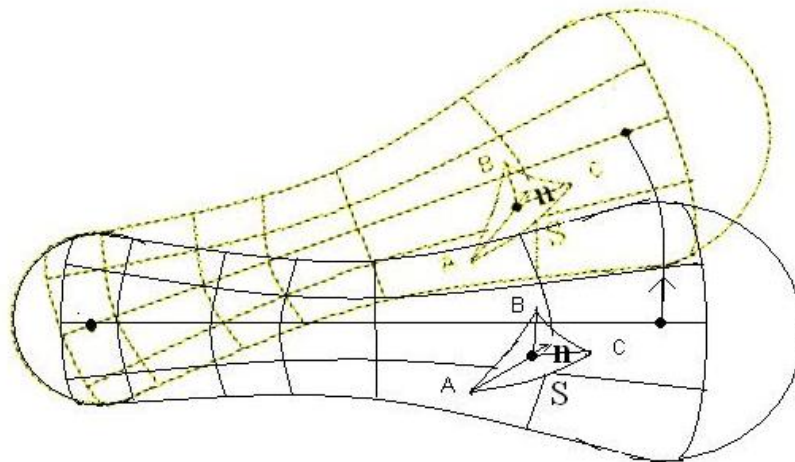


Рис.3 Поворот тора вместе с поверхностным элементом, соответствующий вертикальному смещению центра одной из опорных сфер. Площадь поверхностного элемента не меняется, его центр и нормаль поворачиваются на тот же угол.

Такой малый сдвиг эквивалентен просто повороту всей системы из 2-ух сфер и тора вокруг центра неподвижной опорной сферы r_1 . Центр поверхностного элемента поворачивается на тот же угол вокруг той же оси, его нормаль также вращается на тот же угол. Площадь же остается неизменной.

$$\frac{\partial \varphi_s}{\partial R_{i,x^k}} = \frac{[(r_2 - r_1) \times \frac{\partial (r_2 - r_1)}{\partial x_i^k}]}{(r_2 - r_1)^2},$$

$$\frac{\partial (r_s - r_1)}{\partial R_{i,x^k}} = \left[\frac{\partial_s \varphi}{\partial x_i^k} \times (r_s - r_1) \right],$$

$$\frac{\partial n_s}{\partial R_{i,x^k}} = \left[\frac{\partial \varphi_s}{\partial x_i^k} \times n_s \right],$$

$$\frac{\partial S_s}{\partial R_{i,x^k}} = 0,$$

(2)

r_1 - координата неподвижного атома, r_2 - координата вращающегося атома, φ_s - вектор угла поворота.

1.1.3.2.2.2 Сдвиг одной из опорных сфер (r_2) просходит вдоль оси тора (Рис.4).

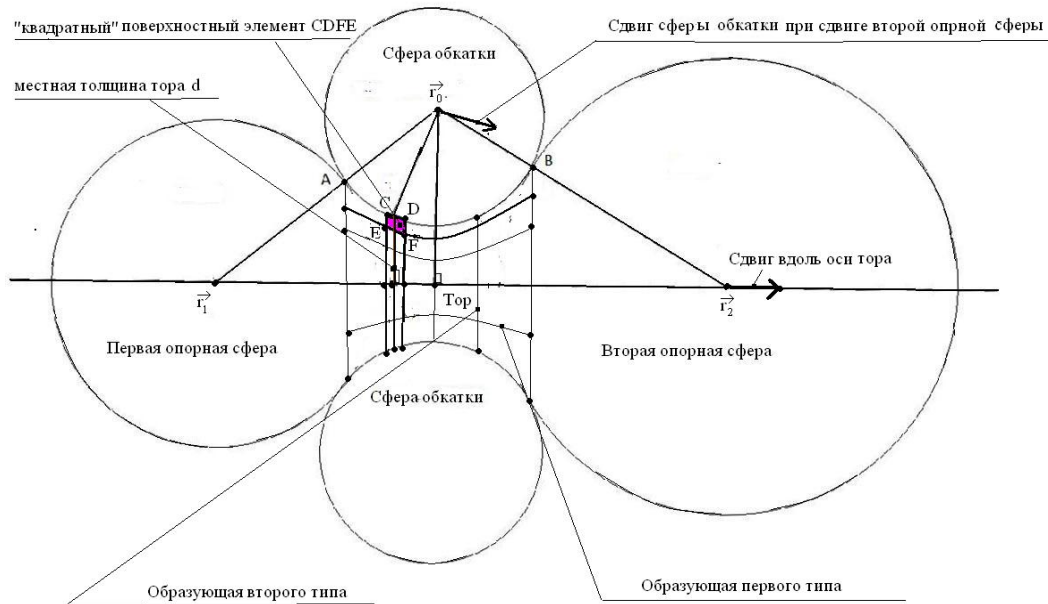


Рис.4. Тор обкатки и две опорные сферы. Не происходит смещения элементов относительно сферы обкатки при ее поступательном сдвиге, происходящем при сдвиге второй опорной сферы вдоль оси тора. Сам тор «утоњшается» (длина образующих второго типа уменьшается), а длина образующих первого типа возрастает за счет растяжения граничных поверхностных элементов. При этом длина отрезков образующей первого типа для поверхностных элементов, лежащих полностью на торе, остается неизменной. Число поверхностных элементов не меняется.

При малом сдвиге второй опорной сферы (см. Рис.4) в сторону от первой опорной сферы вдоль оси тора этот тор обкатки удлинняется. При этом происходит поступательное (без вращения) смещение сферы обкатки тора. При этом сдвиге нормали не меняют направление, а смещение всех точек образующей первого типа (это линия касания тора сферой обкатки) будет такое же, как смещение центра сферы обкатки.

Отсюда градиенты нормалей и центров поверхностных элементов вычисляются по формулам:

$$\begin{aligned} \frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} &= \frac{\partial \mathbf{r}_0}{\partial x_i^k}, \\ \frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} &= 0, \end{aligned} \quad (3)$$

где \mathbf{r}_0 – смещение центра сферы обкатки.

Рассмотрим маленький «квадратный» поверхностный элемент, образованный двумя отрезками образующих тора первого типа (уже определены выше) и двумя отрезками образующих тора второго типа. Образующие тора второго типа – это окружности, образованные пересечением плоскостей, перпендикулярных оси тора с самим тором. При поступательном смещении сферы обкатки тора, длина отрезков, лежащих на образующей первого типа, не меняется. Длина отрезков образующих второго типа уменьшается из-за “утонышения” тора и в той же пропорции, что и местная толщина тора. В итоге и площадь «квадратного» поверхностного элемента (равного произведению длины отрезка образующей первого типа на длину отрезка образующей второго типа) меняется в той же пропорции:

$$\frac{\partial S_s}{\partial R_{i,x^k}} = \frac{\partial d}{d} \cdot S_s, \quad (4)$$

d – местная толщина тора. Любой малый поверхностный элемент можно разбить на набор малых «квадратных» элементов. Следовательно формула (45) верна и для него.

1.1.3.2.3 Поверхностный элемент SES лежит на границе между сегментами разного типа

При поступательном смещении сферы обкатки тора одновременно происходит удлинение образующей первого типа. Однако при этом длины любых отрезков образующей первого типа, лежащие до смещения полностью на торе, не меняются. За счет чего же происходит ее удлинение? Здесь важно вспомнить, что число поверхностных элементов фиксировано. При поступательном смещении сферы обкатки тора происходит растягивание *граничных* поверхностных элементов. Граничные элементы - это элементы, вершины которых лежат на сегментах разного типа, т.е. как на торе, так и на сфере. Изменение их площади и дает искомое удлинение образующих первого типа. Используем формулу для площади треугольника по трем точкам ($S_s = \frac{1}{2} |(\mathbf{r}_1 - \mathbf{r}_2) \times (\mathbf{r}_3 - \mathbf{r}_2)|$), а затем берем градиент.

Таким образом, мы находим выражение для градиента площади граничного элемента поверхности через градиенты координат для точек SES. Эти градиенты координат подставляются через уже найденные выше формулы (42-44). Градиенты его нормали и координат определяются теми же формулами в зависимости от типа сегмента, на котором лежит центр поверхностного элемента.

1.1.3.2.4 Поверхностный элемент SAS

Поверхностные элементов на SAS получены сдвигом вдоль нормали треугольных элементов SES. Снова используя формулу для площади треугольника по трем точкам, а затем, беря градиент, мы находим выражение для градиента площади поверхностного элемента на SAS через градиенты координат и нормалей для точек SES, которые подставляются через уже найденные выше формулы (42-44).

1.1.3.2.5 Итог

Мы видим, что зная смещение центров сфер, мы можем найти изменения параметров поверхностных элементов, исходя из изложенных выше принципов, а, следовательно, и найти их градиенты.

1.1.3.3 Детальные формулы для аналитических градиентов на различных типах сегментов

Приведем ниже подробные и детальные формулы для аналитических градиентов параметров поверхностных элементов, лежащих на различных типах первичных и вторичных сфер и торов, а также атомах. Эти типы определены в работах [21,68].

1.1.3.3.1 Поверхностные элементы на атомах

Производные от параметров поверхностных элементов (координат, нормалей, площадей) для точек находящихся на поверхности атома с координатами центра \mathbf{R}_m полностью определяются через градиент центра атома, как говорилось выше:

$$\begin{aligned}\frac{\partial \mathbf{R}_m}{\partial R_{i,x^k}} &= \delta_m^k \mathbf{e}^k \\ \frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} &= \delta_m^k \mathbf{e}^k, \\ \frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} &= 0, \\ \frac{\partial S_s}{\partial R_{i,x^k}} &= 0,\end{aligned}\tag{5}$$

- то есть точки перемещаются вместе с центром атома.

\mathbf{e}^k - базисный орт в направлении координаты x^k , член множества $\{\mathbf{e}^0, \mathbf{e}^1, \mathbf{e}^2\}$,

δ_{ij} - дельта-функция

1.1.3.3.2 Поверхностные элементы, находящиеся на сфере первичного шара-зонда, при его опоре на три атома. (Рис.5)

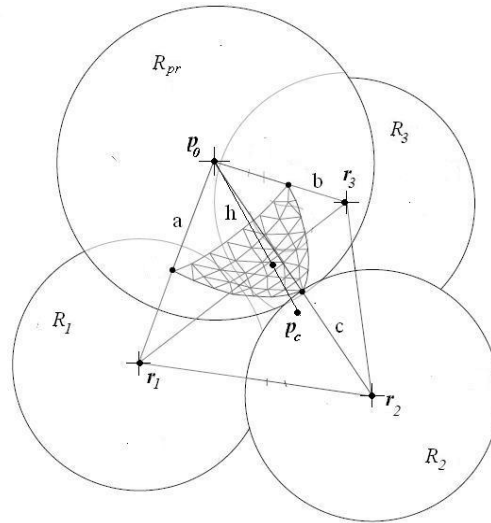


Рис.5 Обкатка при опоре на три атома.

Производные от параметров поверхностных элементов (координат, нормалей, площадей), находящихся на поверхности сферы первичного шара-зонда и опирающегося на три атома, полностью определяются градиентом центра этого шара-зонда. Поэтому вначале вычислим положение и градиенты центра этого шара-зонда. Определим положение центра шара-зонда.

Обозначения на Рис. 5:

$$a = R_1 + R_{pr}; b = R_2 + R_{pr}; c = R_3 + R_{pr}, \quad (6)$$

\mathbf{r} - координаты центра сферы первичной обкатки R_{pr}
 a, b, c - стороны пирамиды, R_1, R_2, R_3 - радиусы опорных сфер, R_{pr} - радиус сферы первичной обкатки; $\mathbf{r}_1 = \mathbf{R}_m, \mathbf{r}_2 = \mathbf{R}_n, \mathbf{r}_3 = \mathbf{R}_l$ - радиус-векторы опорных сфер.

Радиус-вектор, определяющий положение шара-зонда, является решением системы трех векторных уравнений:

$$\begin{cases} (\mathbf{r} - \mathbf{R}_m)^2 = a^2 \\ (\mathbf{r} - \mathbf{R}_n)^2 = b^2 \\ (\mathbf{r} - \mathbf{R}_l)^2 = c^2 \end{cases}, \quad (7)$$

Радиус-вектор центра шара-зонда можно разложить по двум векторам - это радиус-вектор основания высоты пирамиды \mathbf{p}_c и вектор, перпендикулярный основанию пирамиды \mathbf{z} :

$$\mathbf{r} = \mathbf{p}_c \pm zh. \quad (8)$$

Эти два радиус-вектор определяются двумя формулами;

$$z = \frac{[(\mathbf{R}_l - \mathbf{R}_m) \times (\mathbf{R}_n - \mathbf{R}_m)]}{\|[(\mathbf{R}_n - \mathbf{R}_m) \times (\mathbf{R}_l - \mathbf{R}_m)]\|}, \quad (9)$$

$$p_c = \frac{(\mathbf{R}_m + \mathbf{R}_n + \mathbf{R}_l)}{3} + \frac{1}{6} \left[\begin{array}{l} \left(\frac{3(b^2 - c^2) + (\mathbf{R}_l - \mathbf{R}_m)^2 - (\mathbf{R}_n - \mathbf{R}_m)^2}{\|[(\mathbf{R}_l - \mathbf{R}_m) \times (\mathbf{R}_n - \mathbf{R}_m)]\|^2} [(\mathbf{R}_l - \mathbf{R}_m) \times (\mathbf{R}_n - \mathbf{R}_m)] \times \mathbf{R}_m \right) + \\ + \frac{3(c^2 - a^2) + (\mathbf{R}_m - \mathbf{R}_n)^2 - (\mathbf{R}_l - \mathbf{R}_n)^2}{\|[(\mathbf{R}_m - \mathbf{R}_n) \times (\mathbf{R}_l - \mathbf{R}_n)]\|^2} [(\mathbf{R}_m - \mathbf{R}_n) \times (\mathbf{R}_l - \mathbf{R}_n)] \times \mathbf{R}_n \right) + \\ + \frac{3(a^2 - b^2) + (\mathbf{R}_n - \mathbf{R}_l)^2 - (\mathbf{R}_m - \mathbf{R}_l)^2}{\|[(\mathbf{R}_n - \mathbf{R}_l) \times (\mathbf{R}_m - \mathbf{R}_l)]\|^2} [(\mathbf{R}_n - \mathbf{R}_l) \times (\mathbf{R}_m - \mathbf{R}_l)] \times \mathbf{R}_l \end{array} \right], \quad (10)$$

Отсюда высота пирамиды находится по формуле:

$$h^2 = (a^2 + b^2 + c^2)/3 - (\mathbf{R}_m^2 + \mathbf{R}_n^2 + \mathbf{R}_l^2)/3 - p_c^2 + 2p_c \cdot \frac{(\mathbf{R}_m + \mathbf{R}_n + \mathbf{R}_l)}{3}. \quad (11)$$

Производная центра шара-зонда.

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial R_{i,x^k}} &= \frac{(r_{x^k} - R_{m,x^k}) \cdot \delta_m}{(\mathbf{r} - \mathbf{R}_m) \cdot [(\mathbf{r} - \mathbf{R}_n) \times (\mathbf{r} - \mathbf{R}_l)]} [(\mathbf{r} - \mathbf{R}_n) \times (\mathbf{r} - \mathbf{R}_l)] + \\ &+ \frac{(r_{x^k} - R_{n,x^k}) \cdot \delta_n}{(\mathbf{r} - \mathbf{R}_n) \cdot [(\mathbf{r} - \mathbf{R}_m) \times (\mathbf{r} - \mathbf{R}_l)]} [(\mathbf{r} - \mathbf{R}_m) \times (\mathbf{r} - \mathbf{R}_l)] + \\ &+ \frac{(r_{x^k} - R_{l,x^k}) \cdot \delta_l}{(\mathbf{r} - \mathbf{R}_l) \cdot [(\mathbf{r} - \mathbf{R}_n) \times (\mathbf{r} - \mathbf{R}_m)]} [(\mathbf{r} - \mathbf{R}_n) \times (\mathbf{r} - \mathbf{R}_m)] \end{aligned} \quad (12)$$

Отсюда для параметров поверхностного элемента:

$$\begin{aligned} \frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} &= \frac{\partial \mathbf{r}}{\partial R_{i,x^k}}, \\ \frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} &= 0, \\ \frac{\partial S_s}{\partial R_{i,x^k}} &= 0. \end{aligned} \quad (13)$$

1.1.3.3.3 Поверхностные элементы, находящиеся на сфере вторичного шара-зонда, при его опоре на три сферы первичной обкатки. (Рис.6)

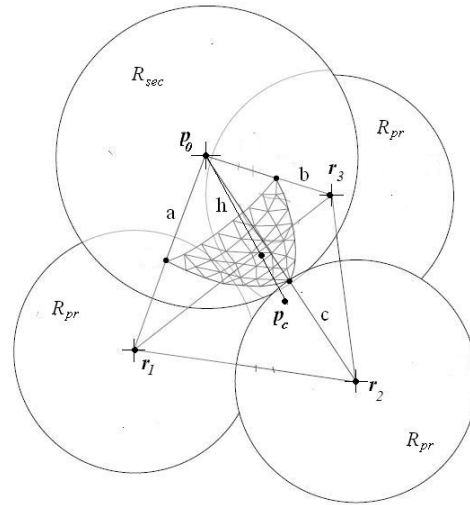


Рис.6 Обкатка при опоре на три сферы первичной обкатки.

Этот случай похож на предыдущий. Он даже проще, поскольку все ребра пирамиды, выходящие из ее вершины, равны. Радиус-вектора, определяющие положения опорных сфер r_1 , r_2 , r_3 и центр вторичного шара-зонда r образуют пирамиду. Однако смещаться при смещении одного атома могут центры уже всех трех первичных опорных сфер, поэтому для нахождения градиентов параметров используем принцип суперпозиции по этим смещениям.

Радиус-вектор центра шара-зонда можно разложить по двум векторам - это радиус-вектор основания высоты пирамиды p_c и вектор, перпендикулярный основанию пирамиды z :

$$r = p_c \pm zh. \quad (14)$$

Эти два радиус-вектор определяются двумя формулами

$$z = \frac{[(r_3 - r_1) \times (r_2 - r_1)]}{\|[(r_2 - r_1) \times (r_3 - r_1)]\|}, \quad (15)$$

$$p_c = \frac{(r_1 + r_2 + r_3)}{3} + \frac{1}{6} \left[\begin{aligned} & \left(\frac{(r_3 - r_1)^2 - (r_2 - r_1)^2}{\|[(r_3 - r_1) \times (r_2 - r_1)]\|^2} [[(r_3 - r_1) \times (r_2 - r_1)] \times r_1] + \right. \\ & \left. + \frac{(r_1 - r_2)^2 - (r_3 - r_2)^2}{\|[(r_1 - r_2) \times (r_3 - r_2)]\|^2} [[(r_1 - r_2) \times (r_3 - r_2)] \times r_2] + \right. \\ & \left. + \frac{(r_2 - r_3)^2 - (r_1 - r_3)^2}{\|[(r_2 - r_3) \times (r_1 - r_3)]\|^2} [[(r_2 - r_3) \times (r_1 - r_3)] \times r_3] \right). \end{aligned} \right] \quad (16)$$

Высота пирамиды определяется формулой:

$$h^2 = (R_{pr} + R_{sec})^2 - (r_1^2 + r_2^2 + r_3^2) / 3 - p_c^2 + 2p_c \cdot \frac{(r_1 + r_2 + r_3)}{3}, \quad (17)$$

где R_{sec} - радиус сферы вторичной обкатки.

Смещение одного атома может приводить к смещению всех трех первичных опорных сфер. Производная центра шара-зонда определяется суперпозицией смещений, каждая из компонент которой соответствует сдвигу только одной из опорных сфер.

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial R_{i,x^k}} = & \frac{(\mathbf{r}-\mathbf{r}_1) \cdot \frac{\partial \mathbf{r}_1}{\partial R_{i,x^k}}}{(\mathbf{r}-\mathbf{r}_1) \cdot [(\mathbf{r}-\mathbf{r}_2) \times (\mathbf{r}-\mathbf{r}_3)]} [(\mathbf{r}-\mathbf{r}_2) \times (\mathbf{r}-\mathbf{r}_3)] + \\ & + \frac{(\mathbf{r}-\mathbf{r}_2) \cdot \frac{\partial \mathbf{r}_2}{\partial R_{i,x^k}}}{(\mathbf{r}-\mathbf{r}_2) \cdot [(\mathbf{r}-\mathbf{r}_1) \times (\mathbf{r}-\mathbf{r}_3)]} [(\mathbf{r}-\mathbf{r}_1) \times (\mathbf{r}-\mathbf{r}_3)] + \\ & + \frac{(\mathbf{r}-\mathbf{r}_3) \cdot \frac{\partial \mathbf{r}_3}{\partial R_{i,x^k}}}{(\mathbf{r}-\mathbf{r}_3) \cdot [(\mathbf{r}-\mathbf{r}_2) \times (\mathbf{r}-\mathbf{r}_1)]} [(\mathbf{r}-\mathbf{r}_2) \times (\mathbf{r}-\mathbf{r}_1)]. \end{aligned} \quad (18)$$

Градиенты центров опорных первичных сфер, входящих в эту формулу, определяются ранее найденной формулой (13).

Отсюда для параметров поверхностного элемента:

$$\begin{aligned} \frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} &= \frac{\partial \mathbf{r}}{\partial R_{i,x^k}}, \\ \frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} &= 0, \\ \frac{\partial S_s}{\partial R_{i,x^k}} &= 0. \end{aligned} \quad (19)$$

1.1.3.3.4 Точки на поверхности вторичных сфер устойчивых положений (Рис.7)

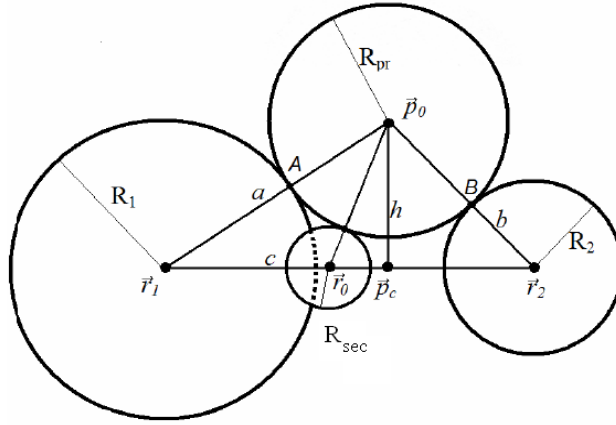


Рис.7 Вторичные сферы устойчивых положений. См. [20].

Производные от параметров поверхностных элементов (координат, нормалей, площадей) для точек находящихся на поверхности сфер устойчивых положений полностью определяется градиентом центров этих сфер. Поэтому вначале вычислим градиенты центров этих сфер. Рассмотрим треугольник, образованный центрами двух опорных атомов и любой из сфер первичной обкатки:

$$a = R_1 + R_{pr}; \quad b = R_2 + R_{pr}; \quad c = |\mathbf{R}_m - \mathbf{R}_n|, \quad (20)$$

$$a, b - const \quad \frac{\partial a}{\partial R_{i,x^k}} = \frac{\partial b}{\partial R_{i,x^k}} = 0, \quad (21)$$

$$\frac{\partial c}{\partial R_{i,x^k}} = \frac{(\delta_{in} - \delta_{im}) \cdot (R_{n,x^k} - R_{m,x^k})}{c}, \quad (22)$$

a, b, c – стороны треугольника, R_{pr} - радиус сферы первичной обкатки; $\mathbf{r}_1 = \mathbf{R}_m, \mathbf{r}_2 = \mathbf{R}_n$ - радиус-векторы опорных атомов.

Высота треугольника и ее градиент:

$$h = \frac{1}{2c} \sqrt{4a^2c^2 - (a^2 + c^2 - b^2)^2}, \quad (23)$$

$$\frac{\partial h}{\partial R_{i,x^k}} = \left(\frac{b^2 + a^2 - c^2}{2h} - h \right) \cdot \frac{\partial c}{\partial R_{i,x^k}}. \quad (24)$$

Расстояние от центра сферы устойчивого положения до основания высоты треугольника и ее градиент:

$$d = \sqrt{(R_{pr} + R_{sec})^2 - h^2}, \quad (25)$$

$$\frac{\partial d}{\partial R_{i,x^k}} = - \frac{h \cdot \frac{\partial h}{\partial R_{i,x^k}}}{d}, \quad (26)$$

где R_{sec} - радиус сферы вторичной обкатки.

Единичный орт вдоль оси тора обкатки от первого атома ко второму и его градиент:

$$\mathbf{z} = \frac{(\mathbf{R}_n - \mathbf{R}_m)}{c}, \quad (27)$$

$$\frac{\partial \mathbf{z}}{\partial R_{i,x^k}} = \frac{(\delta_n - \delta_m) \mathbf{e}^k}{c} - \frac{(\mathbf{R}_n - \mathbf{R}_m)}{c^2} \frac{\partial c}{\partial R_{i,x^k}}, \quad \left(\left(\frac{\partial \mathbf{z}}{\partial R_{i,x^k}} \cdot \mathbf{z} \right) = 0 \right). \quad (28)$$

Радиус вектор точки основания высоты треугольника \mathbf{p}_c и его градиент:

$$\mathbf{p}_c = \frac{1}{2}(\mathbf{R}_m + \mathbf{R}_n) + \frac{(\mathbf{R}_n - \mathbf{R}_m)(a^2 - b^2)}{2c^2}, \quad (29)$$

$$\frac{\partial \mathbf{p}_c}{\partial R_{i,x^k}} = \frac{1}{2}(\delta_n + \delta_m) \mathbf{e}^k + \frac{(\delta_n - \delta_m)(a^2 - b^2)}{2c^2} \mathbf{e}^k - (\mathbf{R}_n - \mathbf{R}_m)(a^2 - b^2) \frac{((\delta_n - \delta_m) \cdot (\mathbf{R}_{n,x^k} - \mathbf{R}_{m,x^k}))}{c^4}. \quad (30)$$

\mathbf{r}_0 - радиус-вектора центров двух сфер устойчивого положения и их градиенты:

$$\mathbf{r}_0 = \mathbf{p}_c \pm d\mathbf{z},$$

$$\frac{\partial \mathbf{r}_0}{\partial R_{i,x^k}} = \frac{\partial \mathbf{p}_c}{\partial R_{i,x^k}} \pm \left(\frac{\partial d}{\partial R_{i,x^k}} \cdot \mathbf{z} + d \cdot \frac{\partial \mathbf{z}}{\partial R_{i,x^k}} \right). \quad (31)$$

Отсюда для параметров поверхностного элемента:

$$\begin{aligned} \frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} &= \frac{\partial \mathbf{r}_0}{\partial R_{i,x^k}}, \\ \frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} &= 0, \\ \frac{\partial S_s}{\partial R_{i,x^k}} &= 0. \end{aligned} \quad (32)$$

1.1.3.3.5 Первичная обкатка двух атомов. Торы первичной обкатки (Рис.8).

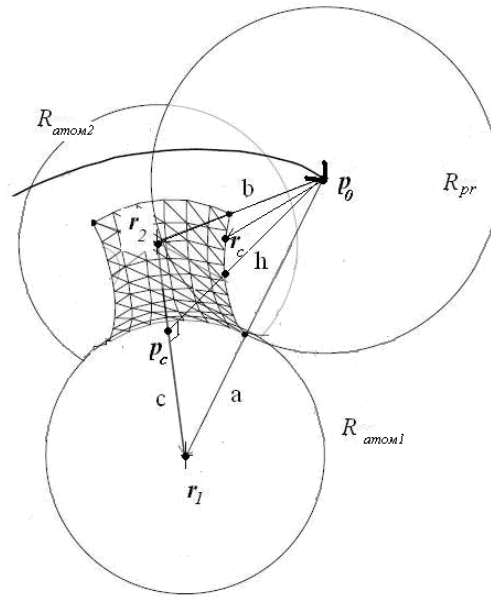


Рис.8 Торы первичной обкатки при опоре на два атома.

Частично необходимые формулы уже выведены в предыдущем пункте.

Пусть x, y, z - локальный базис тора (центр координат лежит в основании высоты). Ось z определена как в предыдущем пункте

$$(x \cdot z) = 0, (y \cdot z) = 0, (x \cdot y) = 0, (x \cdot x) = 1, (y \cdot y) = 1, (z \cdot z) = 1. \quad (33)$$

Нормаль поверхностного элемента разлагается по локальному базису:

$$n_s = \alpha_s \sqrt{1 - \gamma_s^2} x + \beta_s \sqrt{1 - \gamma_s^2} y + \gamma_s z, \quad (34)$$

$$\alpha_s^2 + \beta_s^2 = 1, \quad (35)$$

$$\frac{\partial \gamma_s}{\partial R_{i,x^k}} = 0, \quad \frac{\partial \alpha_s}{\partial R_{i,x^k}} = 0, \quad \frac{\partial \beta_s}{\partial R_{i,x^k}} = 0. \quad (36)$$

Задания $\alpha_s, \beta_s, \gamma_s$ определяют положение центра поверхностного элемента на торе.

e_s - единичный вектор из основания высоты к центру сферы обкатки, определяемой текущими α_s, β_s

$$e_s = \alpha_s x + \beta_s y. \quad (37)$$

Его градиент:

$$\frac{\partial e_s}{\partial R_{i,x^k}} = - \left(e_s \cdot \frac{\partial z}{\partial R_{i,x^k}} \right) z. \quad (38)$$

Отсюда градиент нормали поверхностного элемента для торов первичной обкатки:

$$\frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} = \sqrt{1 - \gamma_s^2} \frac{\partial \mathbf{e}_s}{\partial R_{i,x^k}} + \gamma_s \cdot \frac{\partial \mathbf{z}}{\partial R_{i,x^k}}. \quad (39)$$

Для градиента радиус-вектора поверхностного элемента:

$$\frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} = \frac{\partial \mathbf{p}_c}{\partial R_{i,x^k}} + \frac{\partial h}{\partial R_{i,x^k}} \cdot \mathbf{e}_s + (h - R_{rp} \sqrt{1 - \gamma_s^2}) \cdot \frac{\partial \mathbf{e}_s}{\partial R_{i,x^k}} - R_{rp} \gamma_s \cdot \frac{\partial \mathbf{z}}{\partial R_{i,x^k}}. \quad (40)$$

Для градиента площадей поверхностных элементов торов первичной обкатки:

$$\frac{\partial S_s}{\partial R_{i,x^k}} = \frac{\frac{\partial h}{\partial R_{i,x^k}}}{(h - R_{rp} \sqrt{1 - \gamma_s^2})} \cdot S_s. \quad (41)$$

1.1.3.3.6 Вторичная обкатка двух сфер первичной обкатки. Торы вторичной обкатки (Рис.9).

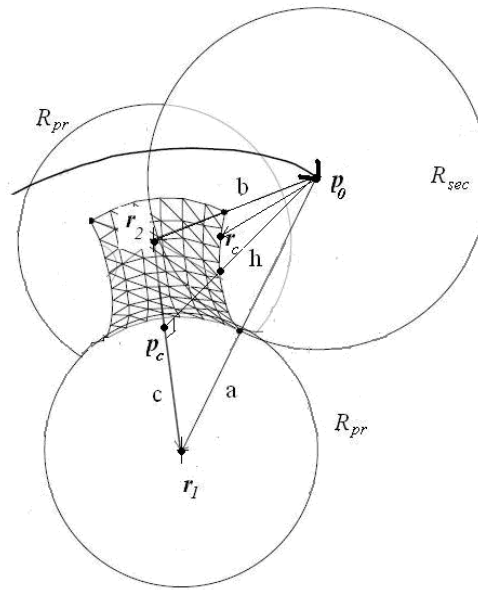


Рис.9 Торы вторичной обкатки при опоре на две первичные сферы.

Этот случай похож на предыдущий. Он даже проще – опорные сферы имеют одинаковый радиуса и треугольник, образованный центрами опорных сфер и сферой вторичной обкатки является равносторонним. Однако есть и отличие – все нормали направлены в направлении, *противоположным* центру текущей сферы обкатки.

Стороны треугольника, образованного центрами двух сфер первичной обкатки и сферой вторичной обкатки и их градиенты:

$$a = b = R_{pr} + R_{sec}; \quad c = |\mathbf{r}_2 - \mathbf{r}_1|, \quad (42)$$

$$\frac{\partial c}{\partial R_{i,x^k}} = \frac{\left(\frac{\partial r_2}{\partial R_{i,x^k}} - \frac{\partial r_1}{\partial R_{i,x^k}}\right) \cdot (r_2 - r_1)}{c}. \quad (43)$$

Высота и ее градиент:

$$h = \frac{1}{2} \sqrt{4a^2 - c^2}, \quad (44)$$

$$\frac{\partial h}{\partial R_{i,x^k}} = -\frac{c}{4h} \cdot \frac{\partial c}{\partial R_{i,x^k}}. \quad (45)$$

Радиус вектор точки основания высоты треугольника \mathbf{p}_c и ее производная:

$$\mathbf{p}_c = \frac{1}{2} (\mathbf{r}_1 + \mathbf{r}_2), \quad (46)$$

$$\frac{\partial \mathbf{p}_c}{\partial R_{i,x^k}} = \frac{1}{2} \left(\frac{\partial \mathbf{r}_1}{\partial R_{i,x^k}} + \frac{\partial \mathbf{r}_2}{\partial R_{i,x^k}} \right).$$

$\mathbf{x}, \mathbf{y}, \mathbf{z}$ - локальный базис

$$\mathbf{z} = \frac{(\mathbf{r}_2 - \mathbf{r}_1)}{c}, \quad (\mathbf{x} \cdot \mathbf{z}) = 0, \quad (\mathbf{y} \cdot \mathbf{z}) = 0, \quad (\mathbf{x} \cdot \mathbf{y}) = 0. \quad (47)$$

Производные локального базиса

$$\frac{\partial \mathbf{z}}{\partial R_{i,x^k}} = \frac{\left(\frac{\partial \mathbf{r}_2}{\partial R_{i,x^k}} - \frac{\partial \mathbf{r}_1}{\partial R_{i,x^k}} \right)}{c} - \frac{(\mathbf{r}_2 - \mathbf{r}_1)}{c^2} \frac{\partial c}{\partial R_{i,x^k}}. \quad (48)$$

Проекция нормали на локальный базис для торов вторичной обкатки (нормаль направлена внутрь тора):

$$\mathbf{n}_s = -\alpha_s \cdot \sqrt{1 - \gamma_s^2} \cdot \mathbf{x} - \beta_s \cdot \sqrt{1 - \gamma_s^2} \cdot \mathbf{y} - \gamma_s \cdot \mathbf{z}. \quad (49)$$

Производные радиус-вектора центра текущей сферы обкатки, единичного вектора из основания высоты к центру текущей сферы обкатки, векторов локального базиса $\mathbf{x}, \mathbf{y}, \mathbf{z}$:

$$\mathbf{e}_s = \alpha_s \mathbf{x} + \beta_s \mathbf{y}, \quad (50)$$

$$\frac{\partial \mathbf{e}_s}{\partial R_{i,x^k}} = - \left(\mathbf{e}_s \cdot \frac{\partial \mathbf{z}}{\partial R_{i,x^k}} \right) \mathbf{z}. \quad (51)$$

Отсюда градиент нормали поверхностного элемента для торов вторичной обкатки:

$$\frac{\partial \mathbf{n}_s}{\partial R_{i,x^k}} = - \sqrt{1 - \gamma_s^2} \frac{\partial \mathbf{e}_s}{\partial R_{i,x^k}} - \gamma_s \cdot \frac{\partial \mathbf{z}}{\partial R_{i,x^k}}. \quad (52)$$

Для градиента радиус-вектора поверхностного элемента:

$$\frac{\partial \mathbf{r}_s}{\partial R_{i,x^k}} = \frac{\partial \mathbf{p}_c}{\partial R_{i,x^k}} + \frac{\partial h}{\partial R_{i,x^k}} \cdot \mathbf{e}_s + (h - R_{\text{sec}} \sqrt{1 - \gamma_s^2}) \cdot \frac{\partial \mathbf{e}_s}{\partial R_{i,x^k}} - R_{\text{sec}} \gamma_s \cdot \frac{\partial \mathbf{z}}{\partial R_{i,x^k}}. \quad (53)$$

Для градиента площадей поверхностных элементов торов вторичной обкатки:

$$\frac{\partial S_s}{\partial R_{i,x^k}} = \frac{\frac{\partial h}{\partial R_{i,x^k}}}{(h - R_{\text{sec}} \sqrt{1 - \gamma_s^2})} \cdot S_s. \quad (54)$$

1.1.3.3.7 Рассмотрим граничный многоугольный поверхностный элемент, который состоит из треугольников, чьи вершины не принадлежат одному и тому же поверхностному фрагменту

Нормали и координаты радиус-вектора многоугольного поверхностного элемента определяются типом фрагмента поверхности, на которой лежит его центр. По формулам, относящимся к этому фрагменту, вычисляются и градиенты радиус-вектора и нормалей этого поверхностного элемента.

С его площадью сложнее. Ее величина равна трети суммы площадей составляющих его треугольников. Соответственно, ее градиент

$$\frac{\partial S^M}{\partial R_{i,x^k}} = \frac{\sum_j \frac{\partial S_j^{\text{tr}}}{\partial R_{i,x^k}}}{3}. \quad (55)$$

Пусть треугольник образован вершинами $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$

$$\begin{aligned} \mathbf{a}_j &= \mathbf{r}_2 - \mathbf{r}_1 \\ \mathbf{b}_j &= \mathbf{r}_3 - \mathbf{r}_1 \end{aligned} \quad (56)$$

Его площадь

$$\mathbf{s} = [\mathbf{a}_j \times \mathbf{b}_j],$$

(57)

$$S_j^{tr} = \frac{1}{2} |\mathbf{s}|.$$

(58)

Градиент его площади:

$$\frac{\partial S_j^{tr}}{\partial R_{i,x^k}} = \frac{1}{4} \frac{(\frac{\partial \mathbf{a}_j}{\partial R_{i,x^k}} \cdot \mathbf{a}_j) \mathbf{b}_j^2 + (\frac{\partial \mathbf{b}_j}{\partial R_{i,x^k}} \cdot \mathbf{b}_j) \mathbf{a}_j^2 - (\mathbf{a}_j \cdot \mathbf{b}_j) ((\frac{\partial \mathbf{a}_j}{\partial R_{i,x^k}} \cdot \mathbf{b}_j) + (\mathbf{a}_j \cdot \frac{\partial \mathbf{b}_j}{\partial R_{i,x^k}}))}{S_j^{tr}},$$

(59)

$$\frac{\partial \mathbf{a}_j}{\partial R_{i,x^k}} = \frac{\partial \mathbf{r}_2}{\partial R_{i,x^k}} - \frac{\partial \mathbf{r}_1}{\partial R_{i,x^k}},$$

$$\frac{\partial \mathbf{b}_j}{\partial R_{i,x^k}} = \frac{\partial \mathbf{r}_3}{\partial R_{i,x^k}} - \frac{\partial \mathbf{r}_1}{\partial R_{i,x^k}}.$$

(60)

Мы предположили выше, что градиенты нормалей граничных элементов могут рассчитываться на основе градиента нормали к поверхности, проходящей через центр этого элемента. Однако можно посчитать это величину и иначе. Многоугольный поверхностный элемент состоит из треугольных граней. Можно посчитать нормаль как среднюю от нормалей всех ее граней (взвешенную по площадям). На основе этого затем посчитать и производную нормали, что мы и сделаем ниже. Практические расчеты показали, что эти производные приводят к меньшим численным ошибкам в градиенте.

Однако для самой энергии меньшие ошибки получаются для нормалей, взятых в центре поверхностного элемента.

Это все связано с тем, что многоугольные элементы состоят из треугольников, относящихся к разным поверхностным фрагментам, Их площади сильно меняются и вершины сильно сдвигаются *после* искажения поверхностной сетки сдвигом атома, по которому берется градиент. Поэтому и средняя нормаль сильно отличается от нормали в центре *после* искажения поверхностной сетки сдвигом атома.

Опишем пересчет градиентов нормалей граничных поверхностных элементов, описанный выше.

Нормаль элемента ищем как векторную среднюю от нормалей к треугольным элементам с вершиной в центральной точке многоугольного элемента. Веса пропорциональны площадям треугольников.

$$\mathbf{a}_{gi} = \mathbf{r}_{gi} - \mathbf{r}_g, \tag{61}$$

где

\mathbf{r}_g – радиус-вектор центра текущего поверхностного элемента

\mathbf{r}_{gi} – радиус-вектора центров соседних поверхностных элементов.

$i=1, \dots, N_{eg}$, N_{eg} – число соседних поверхностных элементов.

\mathbf{n}_g - прежняя нормаль в центре поверхностного элемента

$$\mathbf{a}_{gi_next} = \begin{cases} \mathbf{a}_{g(i+1)} & \text{for } 1 \leq i < N_{eg} \\ \mathbf{a}_{g1} & \text{for } i = N_{eg} \end{cases} . \quad (62)$$

Новая нормаль:

$$\mathbf{n}_{g(new)} = \frac{\sum_{i=1}^{N_{eg}} (\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}) \text{sign}(\mathbf{n}_g \cdot [\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}])}{\left| \sum_{i=1}^{N_{eg}} (\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}) \text{sign}(\mathbf{n}_g \cdot [\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}]) \right|} . \quad (63)$$

Градиент нормали считаем через градиенты вершин треугольников по формуле, аналогичной формуле для градиента площади граничного поверхностного элемента.

$$L_{gi} = \text{sign}(\mathbf{n}_g \cdot [\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}]), \quad (64)$$

$$\mathbf{s}_g = \sum_i L_{gi} [\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}], \quad (65)$$

$$\frac{\partial \mathbf{a}_{gi}}{\partial R_{m,x^k}} = \frac{\partial \mathbf{r}_{gi}}{\partial R_{m,x^k}} - \frac{\partial \mathbf{r}_g}{\partial R_{m,x^k}}, \quad (66)$$

$$\frac{\partial \mathbf{a}_{gi_next}}{\partial R_{m,x^k}} = \frac{\partial \mathbf{r}_{gi_next}}{\partial R_{m,x^k}} - \frac{\partial \mathbf{r}_g}{\partial R_{m,x^k}}, \quad (67)$$

$$\mathbf{S}_{gi} = \left(\left[\frac{\partial \mathbf{a}_{gi}}{\partial R_{m,x^k}} \times \mathbf{a}_{gi_next} \right] + \left[\mathbf{a}_{gi} \times \frac{\partial \mathbf{a}_{gi_next}}{\partial R_{m,x^k}} \right] \right) L_{gi}, \quad (68)$$

$$\frac{\partial \mathbf{n}_{g(new)}}{\partial R_{m,x^k}} = \frac{\sum_{i=1}^{N_{eg}} (\mathbf{S}_{gi} - \mathbf{n}_{g(new)} (\mathbf{n}_{g(new)} \cdot \mathbf{S}_{gi}))}{|\bar{\mathbf{s}}_g|} . \quad (69)$$

1.1.3.4 Поверхностные элементы на SAS и их аналитические градиенты.

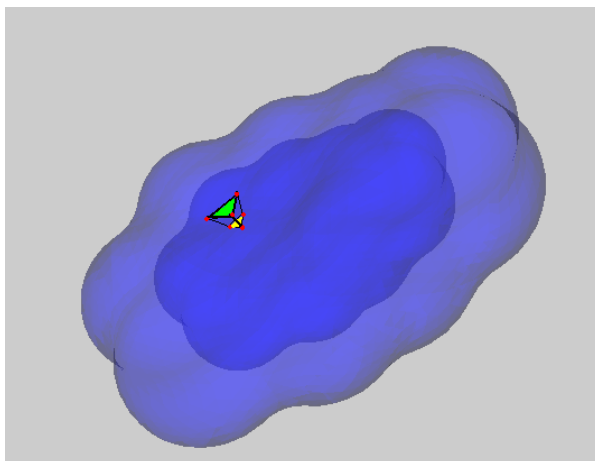


Рис.10 Преобразование поверхности типа SES в поверхность типа SAS.

Пусть имеется поверхность SES после первичной обкатки.

Пусть имеется треугольник на первичной поверхности SES (r_1, r_2, r_3)

Образ этого треугольника на SAS (r_{1n}, r_{2n}, r_{3n}) (Рис.10)

$$r_{1n} = r_1 + n_1 R_{pr}, \quad (70)$$

$$r_{2n} = r_2 + n_2 R_{pr}, \quad (71)$$

$$r_{3n} = r_3 + n_3 R_{pr}, \quad (72)$$

где R_{pr} – радиус первичной обкатки, n_1, n_2, n_3 – нормали в соответствующий точках.

- 1) поверхностный элемент на первичной сфере, опирающейся на три атома, отображается в точку. Его градиент площади нулевой
- 2) поверхностный элемент на торе, опирающимся на два атома, отображается в линию. Его градиент площади нулевой
- 3) поверхностный элемент на атоме отображается в сферический треугольник. Градиент площади по смещению атомов равен нулю
- 4) граничный поверхностный элемент между тором и первичной сферой, опирающейся на три атома. Имеет нулевую площадь и, следовательно, его градиент площади нулевой
- 5) граничный треугольный поверхностный элемент, опирающийся на первичный тор и атом, первичную сферу и атом или первичные тор, сферу и атом – площадь считается по формуле (58) для трех точек (r_{1n}, r_{2n}, r_{3n}). Градиент его площади считаем по формуле:
- 6)

$$\frac{\partial s_j}{\partial R_{m,x^k}} = \frac{\left((b \cdot b) \left(\frac{\partial a}{\partial R_{m,x^k}} \cdot a \right) + (a \cdot a) \left(\frac{\partial b}{\partial R_{m,x^k}} \cdot b \right) - (a \cdot b) \left(\left(\frac{\partial a}{\partial R_{m,x^k}} \cdot b \right) + \left(\frac{\partial b}{\partial R_{m,x^k}} \cdot a \right) \right) \right)}{S_4}, \quad (73)$$

где стороны треугольного поверхностного элемента и их градиенты:

$$\mathbf{a} = \mathbf{r}_{2n} - \mathbf{r}_{1n} , \quad (74)$$

$$\mathbf{b} = \mathbf{r}_{3n} - \mathbf{r}_{1n} , \quad (75)$$

$$\frac{\partial \mathbf{a}}{\partial R_{m,x^k}} = \frac{\partial \mathbf{r}_{2n}}{\partial R_{m,x^k}} - \frac{\partial \mathbf{r}_{1n}}{\partial R_{m,x^k}} , \quad (76)$$

$$\frac{\partial \mathbf{b}}{\partial R_{m,x^k}} = \frac{\partial \mathbf{r}_{3n}}{\partial R_{m,x^k}} - \frac{\partial \mathbf{r}_{1n}}{\partial R_{m,x^k}} , \quad (77)$$

$$S_4 = 2|\mathbf{a} \times \mathbf{b}| , \quad (78)$$

градиенты $\mathbf{r}_{1n}, \mathbf{r}_{2n}, \mathbf{r}_{3n}$:

$$\frac{\partial \mathbf{r}_{1n}}{\partial R_{m,x^k}} = \frac{\partial \mathbf{r}_{1n}}{\partial R_{m,x^k}} + \frac{\partial \mathbf{n}_1}{\partial R_{m,x^k}} R_{pr} , \quad (79)$$

$$\frac{\partial \mathbf{r}_{2n}}{\partial R_{m,x^k}} = \frac{\partial \mathbf{r}_{2n}}{\partial R_{m,x^k}} + \frac{\partial \mathbf{n}_2}{\partial R_{m,x^k}} R_{pr} , \quad (80)$$

$$\frac{\partial \mathbf{r}_{3n}}{\partial R_{m,x^k}} = \frac{\partial \mathbf{r}_{3n}}{\partial R_{m,x^k}} + \frac{\partial \mathbf{n}_3}{\partial R_{m,x^k}} R_{pr} . \quad (81)$$

Индекс $j=1, \dots, N$ пробегает по всем треугольным поверхностным элементам с ненулевым градиентом на SAS.

1.1.3.5 Компьютерная проверка расчетов

Была проведена компьютерная проверка найденных выше формул с помощью программы DISOLV [20,33,65,68]:

- 1) Строились две поверхности обкатки – исходная поверхность молекулы и поверхность, полученная при малом сдвиге одного из атомов. Сдвиг узлов сетки рассчитывался с помощью найденных выше градиентов координат поверхностных элементов путем домножения их на указанный сдвиг атома. Затем эти узлы проецировались на поверхность, полученную после сдвига атома. Таким образом, получались две новые сетки – сдвинутая и ее проекция на новую поверхность. При правильных градиентах разница в положении узлов этих двух сеток должна была быть мала, много меньше разницы между ними и исходной сеткой на исходной молекулярной поверхности. Кроме того, уменьшение этой разницы между двумя новыми сетками должно происходить нелинейно (и более быстро, чем линейно) при уменьшении сдвига атома, поскольку определяется вторыми производными. Все это действительно наблюдалось при компьютерных расчетах. Это подтверждает верность формул для производных координат поверхностных элементов.
- 2) Искались численные градиенты площадей поверхностных элементов через разницу площадей поверхностных элементов на старой и полученной новой сетке. Эта разница

сравнивалась с градиентом площади поверхностного элемента, домноженного на сдвиг. Хорошее совпадение результатов подтверждает верность формул для производных площадей поверхностных элементов.

- 3) Делалось все то же самое, что и в пункте 1). Но кроме этого новая спроецированная сетка сдвигалась на величину нормали к новой поверхности. Новая сетка, полученная до проецирования, тоже сдвигалась на длину пересчитанной нормали (старая нормаль плюс ее градиент, умноженный на сдвиг атома). Эти две новые сдвинутые сетки сравнивались между собой по тем же принципам, что изложены в пункте 1).

1.1.3.6 Выводы

В работе проведены расчеты аналитических градиентов параметров поверхностных элементов молекулярной поверхности, полученной методами первичной и вторичной обкатки в соответствии с алгоритмами, изложенными в [20,21,29-33,65,68].

Приведем проблемные случаи, которые могут возникать при расчете градиентов и решении задачи оптимизации геометрии молекул с их использованием:

- Вырождение - тройная точка опирается не на три, а на большее число атомов (например, ароматические кольца)
- Происходит резкая перестройка поверхности из-за изменения радиуса вторичной обкатки или критического расстояния самой программой при автоматической его настройке
- Происходит резкая перестройка поверхности из-за появления или исчезновения тора обкатки (первичного или вторичного) при сдвиге тройных точек или изменения узкого перешейка тора на расстояниях близких к критическому
- Происходит резкая перестройка поверхности из-за появления или исчезновения тройных точек первичных или вторичных.

Как результат могут быть скачки градиентов и «колебания» оптимизатора вблизи вышеописанных точек

Компьютерный расчет дает хорошее совпадение с ожидаемыми результатами и подтверждает верность формул для градиентов нормалей поверхностных элементов.

Глава 1.1.4 Континуальная модель среды IV: Вычисление аналитических градиентов для энергии сольватации по координатам атомов

В данной работе мы описываем методы нахождения аналитических градиентов (производных) по координатам атомов от энергии сольватации [92]. Это делается как для неполярной составляющей энергии, так и для полярной составляющей энергии, полученной методами PCM, COSMO и SGB. Эти градиенты ищутся с использованием аналитических градиентов параметров (координат, нормалей и площадей) поверхностных элементов. Поверхностные элементы расположены как на оптимально гладкой молекулярной поверхности исключенного объема (SES), так и производной от нее поверхности доступной растворителю (SAS). Эти поверхности находятся методами первичной и вторичной обкатки, согласно алгоритму, описанному в [20,21,29-33,65,68,91].

1.1.4.1 Введение

Свободная энергия сольватации ΔG_s в континуальных моделях растворителя представляется в виде суммы трех составляющих [93]

$$\Delta G_s = \Delta G_{cav} + \Delta G_{np} + \Delta G_{pol}, \quad (1)$$

$$\Delta G_{nonpol} = \Delta G_{cav} + \Delta G_{np},$$

где

ΔG_{pol} – полярная составляющая, связанная с поляризацией диэлектрика-раствора

ΔG_{cav} – кавитационная составляющая, связанная с образованием полости в растворителе

ΔG_{np} – составляющая, определяемая ванн-дер-ваальсовым взаимодействием

ΔG_{pol} – неполярная составляющая, являющаяся суммой предыдущих двух.

Приведем краткий обзор работ, посвященных нахождению аналитических градиентов (производных) энергии сольватации, которые берутся по координатам атомов. Подобные градиенты необходимы, например, при финальной оптимизации положения и геометрии молекулы - лиганда в модели доркинга.

Задача нахождения аналитических градиентов распадается на две подзадачи:

- Нахождение аналитических градиентов параметров (координат, нормали, площади) поверхностных элементов молекулярной поверхности.

Эта задача для алгоритма GEPOI построения поверхности (путем заполнения пустого пространства фиктивными шарами) рассмотрена в [37]. Алгоритм, реализующий построение гладкой поверхности методами первичной и вторичной обкатки, описан в [20,21,29-33,65,68,91].

- На основе полученных выше градиентов параметров поверхностных элементов рассчитываются аналитические градиенты матриц, входящих в уравнения модели (для COSMO [94,95], для PCM [96,97]) . На основе уже градиентов этих матриц затем рассчитываются и аналитические градиенты энергии сольватации.

Следует отметить, что для метода PCM описанная в [96,97] методика нахождения градиентов требует обращения одной из матриц уравнения PCM. Если же число поверхностных элементов велико, то и размер этой матрицы велик. Хранение такой матрицы требует большой компьютерной памяти и значительно повышает вычислительное время задачи. Кроме того сами алгоритмы обращения матриц [45,98] также требуют много памяти и затрат времени. Однако для метода с укрупненными поверхностными элементами [99] обратная матрица мала, ее вычисление и использование не проблематично. Используя методику в [100] можно итерационно найти обратную матрицу. Используя методы, изложенные в [96,97], можно уже найти и сами аналитические градиенты.

В данной работе мы описываем решение второй подзадачи – нахождение аналитических градиентов для полярной составляющей энергии сольватации для методов PCM [39], COSMO [94] и SGB [101] , а также для неполярной составляющей энергии сольватации [39]. Для этого используются аналитические градиенты параметров поверхностных элементов молекулярных поверхностей, полученных методами первичной и вторичной обкатки [20,21,29-33,65,68,91]

Существуют два вида поверхности вокруг молекулы [20,21,29-33,65,68,91]. Во-первых, это SES (Solvent Excluded Surface) - поверхность исключённого из растворителя объёма. Объем, занимаемый растворителем, лежит *вне* объёма, ограниченного этой поверхностью. Сам субстрат полностью лежит *внутри* этого объёма. Алгоритм [20,21,29-33,65,68,91] методами первичной и вторичной обкатки стоит гладкую поверхность SES. Во-вторых, это SAS (Solvent Accessible Surface) - поверхность доступная растворителю, которая образуется центрами молекул растворителя, касающихся молекулы субстрата при первичной обкатке. Первый тип поверхности используется для расчета полярной составляющей энергии сольватации (действительно, на SAS многие участки молекулы, существенные для электростатического взаимодействия, имеют нулевую или очень малую площадь [102]), а второй – для расчета кавитационной и ванн-дер-

ваальсовской составляющих. Расчеты аналитических градиентов энергии, полученной методами PCM, COSMO и SGB реализованы в программе DISOLV [20,21,29-33,65,68,91]

1.1.4.2 Физические модели

1.1.4.2.1 Полярная часть энергии сольватации

Для расчета полярной части энергии берутся поверхностные элементы на SES.

1.1.4.2.1.1 Модель PCM

1.1.4.2.1.1.1 Основные уравнения для энергии PCM

Дадим теперь точное описание используемых нами моделей. Начнем с модели *PCM (Polarized Continuum Model)* [39,45,100,101]. Рассмотрим описанную выше задачу нахождения полярной части энергии сольватации. При этом раствор Ω_{ex} заменяется диэлектриком с известной диэлектрической проницаемостью ϵ_{ex} . Растворенная молекула рассматривается как полость Ω_{in} внутри этого диэлектрика, заполненная либо вакуумом $\epsilon_{in}=1$, либо диэлектриком с диэлектрической проницаемостью $\epsilon_{in}\neq 1$. Внутри полости находятся точечные заряды Q_i , находящиеся в центрах атомов \mathbf{R}_i на некотором удалении от границы Σ . Эти заряды, приближенно аппроксимирующие распределение заряда внутри молекулы, мы находили методами силового поля MMFF94 [103-105]. Возможна ситуация, когда имеется не одна, а несколько полостей, соответствующих нескольким молекулам. При этом влияние диэлектрика-раствора полностью определяется поверхностным зарядом $\sigma(\mathbf{r})$ на границе полости. Точное линейное уравнение, связывающее этот поверхностный заряд с положением и величиной зарядов внутри поверхности, называется PCM. Для нахождения полярной компоненты энергии сольватации мы вначале считаем энергию E_1 , необходимую для перенесения молекулы из диэлектрика ϵ_{in} в вакуум (поверхностный заряд, который при этом образуется, имеет плотность $\sigma_{vac}(\mathbf{r})$). Затем энергию E_2 , необходимую для перенесения молекулы из диэлектрика ϵ_{in} в диэлектрик ϵ_{ex} . Энергия сольватации перехода одного моля субстрата из вакуума в растворитель ϵ_{ex} равна разнице этих двух энергий: $\Delta G_{pol} = E_2 - E_1$ и описывается следующей формулой:

$$\Delta G_{pol} = \frac{1}{2} \sum_i Q_i \int_{\Sigma} \frac{\sigma(\mathbf{r}) - \sigma_{vac}(\mathbf{r})}{|\mathbf{R}_i - \mathbf{r}|} dS. \quad (2)$$

Для $\epsilon_{in}=1$ поверхностный заряд в вакууме $\sigma_{vac}=0$ и, следовательно

$$\Delta G_{pol} = \frac{1}{2} \sum_i Q_i \int_{\Sigma} \frac{\sigma(\mathbf{r})}{|\mathbf{R}_i - \mathbf{r}|} dS. \quad (3)$$

Поскольку ϵ_{in} это эмпирически подбираемый параметр, мы выбирали его в расчетах равным $\epsilon_{in}=1$, чтобы исключить из рассмотрения вакуумную составляющую энергии и упростить расчеты.

Интегральное уравнение PCM для поверхностного заряда имеет следующий вид:

$$\sigma(\mathbf{r}) = \frac{\epsilon_{in} - \epsilon_{ex}}{2\pi(\epsilon_{in} + \epsilon_{ex})} \left(\sum_i \frac{Q_i ((\mathbf{r} - \mathbf{R}_i) \cdot \mathbf{n})}{\epsilon_{in} |\mathbf{r} - \mathbf{R}_i|^3} + \int_{\Sigma} \frac{\sigma(\mathbf{r}') ((\mathbf{r} - \mathbf{r}') \cdot \mathbf{n})}{|\mathbf{r} - \mathbf{r}'|^3} dS' \right). \quad (4)$$

Последний интеграл является сингулярным при $r=r'$ и по определению равен:

$$\int_{\varepsilon} \frac{\sigma(r')((r-r') \cdot n)}{|r-r'|^3} dS' = \lim_{\delta \rightarrow 0} \int_{\varepsilon / (|r-r'| < \delta)} \frac{\sigma(r')((r-r') \cdot n)}{|r-r'|^3} dS' . \quad (5)$$

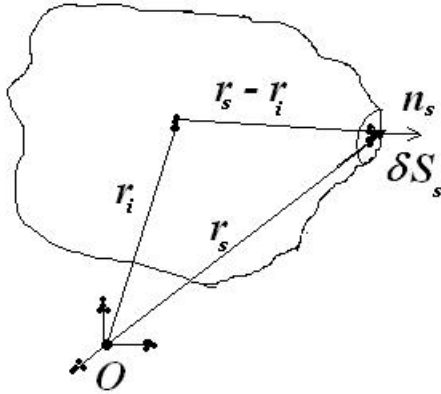


Рис.11 Элемент поверхности относительно начала координат.

После дискретизации поверхности (Рис.1) (т.е. после разбиения её на малые поверхностные элементы, несущие поверхностный заряд) это линейное уравнение может быть записано в матричной форме:

$$Aq = BQ , \quad (6)$$

A – матрица, зависящая от параметров поверхностных элементов

B – матрица, зависящая от геометрических параметров поверхностных элементов и положения зарядов внутри молекулы.

q – столбец зарядов поверхностных элементов

Q – столбец зарядов внутри молекулы.

Определим параметры малых поверхностных элементов и найдем в дальнейшем, как через них выражаются элементы описанных выше матриц:

M - число поверхностных элементов

N - число зарядов внутри полости

q_j - заряд j -ого поверхностного элемента.

S_j – площадь поверхностного элемента

r_j – радиус-вектор центра поверхностного элемента

n_j – нормаль в центре поверхностного элемента

Q_i – заряды внутри молекулы

R_i – радиус-вектор заряда i внутри молекулы

$\varepsilon = \varepsilon_{out} / \varepsilon_{in}$

Элементы b_{ij} матрицы B :

$$b_{ij} = \frac{n_i(r_i - R_j)}{\|r_i - R_j\|^3} S_i \frac{1 - \varepsilon}{4\pi(\varepsilon + 1)} \frac{1}{\varepsilon_{in}} . \quad (7)$$

Для столбцов матрицы B выполняется тождество

$$\sum_{i=1}^M b_{ij} = \frac{1-\varepsilon}{1+\varepsilon} \frac{1}{\varepsilon_{in}}. \quad (8)$$

Элементы матрицы A :

$$a_{ij} = \begin{cases} \frac{n_i(\mathbf{r}_i - \mathbf{r}_j)}{\|\mathbf{r}_i - \mathbf{r}_j\|^3} S_i \frac{\varepsilon - 1}{4\pi(\varepsilon + 1)} & i \neq j \\ a_{jj} = \frac{\varepsilon}{\varepsilon + 1} - \sum_{i \neq j} a_{ij} & i = j \end{cases}. \quad (9)$$

Для столбцов матрицы A выполняется тождество:

$$\sum_{i=1}^M a_{ij} = \frac{\varepsilon}{\varepsilon + 1}. \quad (10)$$

Из (8) и (10) автоматически следует тождество, связывающее суммарный поверхностный заряд с суммарным зарядом внутри полости:

$$\sum_{j=1}^M q_j = - \left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{ex}} \right) \sum_{i=1}^N Q_i. \quad (11)$$

Использование (8) для коррекции численных ошибок автоматически приводит к очень точному выполнению указанного выше тождества.

Каждая из двух компонент энергии в (2) или энергия (3) могут быть записаны в матричной форме:

$$\Delta G_{pol} = Q^T D q \quad (12)$$

Где матричные элементы D :

$$d_{ij} = \frac{1}{2} \left(\frac{1}{\|\mathbf{R}_i - \mathbf{r}_j\|} \right) \quad (13)$$

1.1.4.2.1.1.2 Аналитические производные для матриц

R_{i,x^k} - x^k -ая компонента радиус-вектора R_i

x^k член множества $\{x^0, x^1, x^2\}$

i – номер сдвигаемого атома ($0 < i < N$);

\mathbf{e}^k - базисный орт в направлении координаты x^k , член множества $\{\mathbf{e}^0, \mathbf{e}^1, \mathbf{e}^2\}$

δ_{ij} – дельта-функция

Производная элементов матрицы A по координате одного из атомов R_{m,x^k} :

$$\left\{ \frac{\partial a_{ij}}{\partial R_{m,x^k}} = \frac{1}{4\pi} \frac{\varepsilon - 1}{\varepsilon + 1} \left(\begin{array}{l} \frac{\mathbf{n}_i \cdot (\mathbf{r}_i - \mathbf{r}_j)}{\|\mathbf{r}_i - \mathbf{r}_j\|^3} \frac{\partial S_i}{\partial R_{m,x^k}} + \\ \frac{\frac{\partial \mathbf{n}_i}{\partial R_{m,x^k}} \cdot (\mathbf{r}_i - \mathbf{r}_j)}{\|\mathbf{r}_i - \mathbf{r}_j\|^3} S_i + \\ \mathbf{n}_i \cdot \left(\frac{\partial \mathbf{r}_i}{\partial R_{m,x^k}} - \frac{\partial \mathbf{r}_j}{\partial R_{m,x^k}} \right) \frac{S_i}{\|\mathbf{r}_i - \mathbf{r}_j\|^3} - \\ 3 \frac{\mathbf{n}_i \cdot (\mathbf{r}_i - \mathbf{r}_j)}{\|\mathbf{r}_i - \mathbf{r}_j\|^5} \left((\mathbf{r}_i - \mathbf{r}_j) \cdot \left(\frac{\partial \mathbf{r}_i}{\partial R_{m,x^k}} - \frac{\partial \mathbf{r}_j}{\partial R_{m,x^k}} \right) \right) S_i \end{array} \right) \right. \text{for } i \neq j, \quad (14)$$

$$\left. \frac{\partial a_{jj}}{\partial R_{m,x^k}} = - \sum_{i \neq j} \frac{\partial a_{ij}}{\partial R_{m,x^k}} \quad \text{for } i = j. \right.$$

Отсюда следует, что

$$\sum_{i=1}^M \frac{\partial a_{ij}}{\partial R_{m,x^k}} = 0. \quad (15)$$

Производная элементов матрицы В по координате одного из атомов R_{m,x^k} :

$$\frac{\partial b_{ij}}{\partial R_{m,x^k}} = \frac{1}{4\pi} \frac{1 - \varepsilon}{1 + \varepsilon} \frac{1}{\varepsilon_{in}} \left(\begin{array}{l} \frac{\mathbf{n}_i \cdot (\mathbf{r}_i - \mathbf{R}_j)}{\|\mathbf{r}_i - \mathbf{R}_j\|^3} \frac{\partial S_i}{\partial R_{m,x^k}} + \frac{\frac{\partial \mathbf{n}_i}{\partial R_{m,x^k}} \cdot (\mathbf{r}_i - \mathbf{R}_j)}{\|\mathbf{r}_i - \mathbf{R}_j\|^3} S_i + \\ \mathbf{n}_i \cdot \left(\frac{\partial \mathbf{r}_i}{\partial R_{m,x^k}} - \delta_{jm} \mathbf{e}^k \right) \frac{S_i}{\|\mathbf{r}_i - \mathbf{R}_j\|^3} - \\ 3 \frac{\mathbf{n}_i \cdot (\mathbf{r}_i - \mathbf{R}_j)}{\|\mathbf{r}_i - \mathbf{R}_j\|^5} \left((\mathbf{r}_i - \mathbf{R}_j) \cdot \left(\frac{\partial \mathbf{r}_i}{\partial R_{m,x^k}} - \delta_{jm} \mathbf{e}^k \right) \right) S_i \end{array} \right). \quad (16)$$

Из формул (8) следует, что

$$\sum_{i=1}^M \frac{\partial b_{ij}}{\partial R_{m,x^k}} = 0. \quad (17)$$

Производная элементов матрицы D .

$$\frac{\partial d_{ij}}{\partial R_{m,x^k}} = -\frac{1}{2} \frac{(\mathbf{R}_i - \mathbf{r}_j) \cdot \left(\delta_{im} \mathbf{e}^k - \frac{\partial \mathbf{r}_j}{\partial R_{m,x^k}} \right)}{\|\mathbf{R}_i - \mathbf{r}_j\|^3}. \quad (18)$$

1.1.4.2.1.1.3 Аналитические градиенты энергии для РСМ

Введем понятие зеркальных зарядов поверхностных элементов, которые являются решением следующего уравнения:

$$A^T q^* = D^T Q. \quad (19)$$

Физический смысл зеркальных зарядов q^* следующий:

Пусть имеется распределение зарядов внутри полости, создающие поверхностный заряд $q_{surf} = BQ$ и потенциалы поверхностных элементов q^* , которые находятся из (19). Тогда эта система зарядов имеет ту же энергию, что и настоящие поверхностные заряды q и настоящие потенциалы поверхностных элементов $D^T Q$. Использование зеркальных зарядов позволяет избежать операции обращения матрицы при нахождении аналитических градиентов энергии.

Энергия через параметры и матрицы выражается несколькими способами:

$$\Delta G_{pol} = Q^T Dq = Q^T DA^{-1}BQ = (q^*)^T BQ = Q^T B^T q^* = (q^*)^T Aq. \quad (20)$$

Производная энергии по смещению одного из атомов R_{m,x^k} :

$$\frac{\partial \Delta G_{pol}}{\partial R_{m,x^k}} = \frac{\partial (Q^T DA^{-1}BQ)}{\partial R_{m,x^k}} = Q^T \frac{\partial D}{\partial R_{m,x^k}} q - (q^*)^T \frac{\partial A}{\partial R_{m,x^k}} q + (q^*)^T \frac{\partial B}{\partial R_{m,x^k}} Q. \quad (21)$$

Из формул (15), (17) и (21) следует, что изменение всех компонент столбца q^* на постоянную величину не меняет значения производной энергии. Отсюда обнуление среднего значения зеркального заряда

$$q_i^* = q_i^* - \frac{\sum_{i=1}^M q_i^*}{M} \quad (22)$$

не влияет на энергию, но позволяет уменьшить численные ошибки и ошибки дискретизации.

1.1.4.2.1.2 Модель COSMO

1.1.4.2.1.2.1 Основные уравнения для энергии COSMO

Для случая больших значений $\epsilon \gg 1$ используется модель COSMO ("COnductor-like Screening MOdel") [94]. В этой модели поверхностный заряд находится для бесконечной проницаемости $\epsilon = \infty$. Этот случай эквивалентен замене окружающего диэлектрика на

проводник. Преимущество COSMO над PCM состоит в том, что в линейном матричном уравнении столбец поверхностных зарядов умножается на симметричную, положительно определенную матрицу, элементы которой не зависят от нормалей поверхностных элементов. Для такой матрицы нахождение энергии и аналитических градиентов можно сделать быстрее и точнее.

Уравнение COSMO в интегральной форме записывается в следующем виде:

$$\int_{\Sigma} \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS + \sum_{i=1}^N \frac{Q_i / \epsilon_{in}}{|\mathbf{r} - \mathbf{R}_i|} = 0. \quad (23)$$

Уравнение COSMO в матричном виде после дискретизации поверхности записывается в следующем виде

$$A^C q = -\frac{2}{\epsilon_{in}} D^T Q, \quad (24)$$

где матрица A^C симметрична

$$(A^C)^T = A^C, \quad (25)$$

и ее элементы:

$$\begin{cases} a^C_{ij} = \left(\frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right) & i \neq j, \\ a^C_{ii} = C_s \frac{1}{\sqrt{S_i}} & C_s = 2\sqrt{3.83}, i = j. \end{cases} \quad (26)$$

Матрица D была определена выше в методе PCM.

Для COSMO из теоремы Остроградского-Гаусса электростатики следует тождество, связывающее суммарный поверхностный заряд с суммарным зарядом внутри полости:

$$\sum_{j=1}^M q_j = -\frac{1}{\epsilon_{in}} \sum_{i=1}^N Q_i. \quad (27)$$

Полярная энергия, получаемая для металлической поверхности $\left(\frac{1}{2} \sum_i Q_i \int_{\Sigma} \frac{\sigma(\mathbf{r})}{|\mathbf{R}_i - \mathbf{r}|} dS \right)$, умножается на корректирующий фактор C_f :

$$C_f = \left(\frac{\epsilon - 1}{\epsilon + \frac{1}{2}} \right). \quad (28)$$

В итоге полярная энергия сольватации:

$$\Delta G_{pol} = \frac{1}{2} C_f \sum_i Q_i \int_{\Xi} \frac{\sigma(\mathbf{r})}{|\mathbf{R}_i - \mathbf{r}|} dS \quad (29)$$

Относительная погрешность найденной энергии имеет величину порядка $1/(2\varepsilon)$.

Энергия в матричной форме

:

$$\Delta G_{pol} = C_f Q^T D q = -C_f 2Q^T D (A^C)^{-1} D^T Q = -\frac{C_f}{2} q^T (A^C) q \quad (30)$$

1.1.4.2.1.2.2 Аналитические производные для матриц COSMO

Производная матрицы A^C :

$$\left\{ \begin{array}{l} \frac{\partial a_{ij}^C}{\partial R_{m,x^k}} = -\frac{(\mathbf{r}_i - \mathbf{r}_j) \left(\frac{\partial \mathbf{r}_i}{\partial R_{m,x^k}} - \frac{\partial \mathbf{r}_j}{\partial R_{m,x^k}} \right)}{\|\mathbf{r}_i - \mathbf{r}_j\|^3} \text{ for } i \neq j, \\ \frac{\partial a_{ii}^C}{\partial R_{m,x^k}} = -2\sqrt{3.83} \frac{\partial S_i}{2\sqrt{S_i^3}} = -a_{ii}^C \frac{\partial R_{m,x^k}}{2S_i} \text{ for } i = j. \end{array} \right. \quad (31)$$

Матрица D и ее производная найдена ранее в разделе о PCM.

1.1.4.2.1.2.3 Аналитические производные для энергии COSMO

Производная энергии (30) по координате одного из атомов R_{m,x^k} :

$$\frac{1}{C_f} \frac{\partial \Delta G_{pol}}{\partial R_{m,x^k}} = -2 \frac{\partial (Q^T D (A^C)^{-1} D^T Q)}{\partial R_{m,x^k}} = 2q^T \frac{\partial D^T}{\partial R_{m,x^k}} Q + \frac{1}{2} q^T \frac{\partial A^C}{\partial R_{m,x^k}} q \quad (32)$$

1.1.4.2.1.3 Модель SGB

1.1.4.2.1.3.1 Основные уравнения для энергии SGB

Выражение для электростатического взаимодействия в рамках эмпирической модели SGB [101] находится напрямую через параметры зарядов внутри полости и так называемые Борновские радиусы атомов a_i :

$$\Delta G_{pol} = -\frac{1}{2} \cdot \frac{1}{\varepsilon_m} \frac{1}{1 + \frac{1}{2\varepsilon}} \left(1 - \frac{1}{\varepsilon} \right) \cdot \sum_{i,j} \frac{Q_i \cdot Q_j}{\sqrt{|\mathbf{R}_{i,j}|^2 + a_i \cdot a_j \cdot \exp\left(-\frac{|\mathbf{R}_{i,j}|^2}{c \cdot a_i \cdot a_j}\right)}}, \quad (33)$$

Где $\mathbf{R}_{i,j} = \mathbf{R}_i - \mathbf{R}_j$; c - эмпирическая константа равная 8.

Борновские радиусы атома i находятся через интегралы по поверхности полости:

$$a_i = \frac{1}{2 \left(\sum_n A_n \cdot I_n^i - A_0 \right)}, \quad (34)$$

A_n – находимые эмпирически константы, I_n – интегралы по поверхности полости вида:

$$I_n^i = \left[\oint \frac{(\mathbf{n}_s \cdot (\mathbf{r}_s - \mathbf{R}_i)) dS}{|\mathbf{r}_s - \mathbf{R}_i|^n} \right]^{1/n-3} \quad n \geq 4, \quad (35)$$

или, выражая интегралы через малые поверхностные элементы j ,

$$I_n^i = \left[\sum_j \frac{(\mathbf{n}_j \cdot (\mathbf{r}_j - \mathbf{R}_i)) S_j}{|\mathbf{r}_j - \mathbf{R}_i|^n} \right]^{1/n-3} \quad n \geq 4. \quad (36)$$

Мы ограничивались четырьмя интегралами $n=4,5,6,7$. Эмпирически подобраны значения параметров (для размерности расстояния в А):

$$A_0 = -8.191 \text{ 1/А}, A_1 = 152.661, A_2 = -238.555, A_3 = 151.700, A_4 = -1.074$$

1.1.4.2.1.3.2 Аналитические градиенты энергии SGB

Полный дифференциал энергии при неизменных зарядах и диэлектрических постоянных:

$$\frac{\partial \Delta G_{pol}}{\partial R_{m,x^k}} = -\frac{1}{2} \cdot \frac{1}{\varepsilon_{in}} \frac{1}{1 + \frac{1}{2\varepsilon}} \left(1 - \frac{1}{\varepsilon} \right) \cdot \sum_{i,j} Q_i \cdot Q_j \cdot \frac{\partial g_{i,j}}{\partial R_{m,x^k}}, \quad (37)$$

где функция $g_{i,j}$ и ее производные определяются формулами:

$$g_{i,j} = \frac{1}{\sqrt{|\vec{R}_{i,j}|^2 + a_i \cdot a_j \cdot \exp\left(-\frac{|\mathbf{R}_{i,j}|^2}{c \cdot a_i \cdot a_j}\right)}}, \quad (38)$$

$$\frac{\partial g_{i,j}}{\partial R_{m,x^k}} = \frac{-g_{i,j}^3}{2} \left[\frac{\partial a_i}{\partial R_{m,x^k}} a_j + a_i \frac{\partial a_j}{\partial R_{m,x^k}} \right] \left(1 + \frac{|\mathbf{R}_{i,j}|^2}{c \cdot a_i \cdot a_j} \right) \exp\left(-\frac{|\mathbf{R}_{i,j}|^2}{c \cdot a_i \cdot a_j}\right) + \left[2((R_{i,x^k} - R_{j,x^k}) \cdot (\delta_{im} - \delta_{jm})) \left(1 - \frac{1}{c} \exp\left(-\frac{|\mathbf{R}_{i,j}|^2}{c \cdot a_i \cdot a_j}\right) \right) \right]. \quad (39)$$

Дифференциалы для борновских радиусов атомов:

$$\frac{\partial a_i}{\partial R_{m,x^k}} = \frac{-1}{2 \left(\sum_n A_n \cdot I_n^i - A_0 \right)^2} \cdot \sum_n A_n \cdot \frac{\partial I_n^i}{\partial R_{m,x^k}}. \quad (40)$$

Производные I_n^i -интегралов (для дискретного приближения значений интегралов, s – индекс поверхностного элемента):

$$\frac{\partial I_n^i}{\partial R_{m,x^k}} = \frac{\sum_s \frac{\partial J_{s,n}^i}{\partial R_{m,x^k}}}{n-3} \frac{1}{[I_n^i]^{n-4}} \quad n \geq 4, \quad (41)$$

где $J_{s,n}^i$ и их производные ищутся по формулам:

$$J_{s,n}^i = \frac{(\mathbf{n}_s \cdot (\mathbf{r}_s - \mathbf{R}_i)) S_s}{|\mathbf{r}_s - \mathbf{R}_i|^n}, \quad (42)$$

$$\frac{\partial J_{s,n}^i}{\partial R_{m,x^k}} = \frac{\left[\left[\left(\mathbf{n}_s \cdot \left(\frac{\partial \mathbf{r}_s}{\partial R_{m,x^k}} - \delta_{im} \mathbf{e}^k \right) \right) + \left(\frac{\partial \mathbf{n}_s}{\partial R_{m,x^k}} \cdot (\mathbf{r}_s - \mathbf{R}_i) \right) \right] S_s + (\mathbf{n}_s \cdot (\mathbf{r}_s - \mathbf{R}_i)) \frac{\partial S_s}{\partial R_{m,x^k}}}{|\mathbf{r}_s - \mathbf{R}_i|^n} - \frac{n(\mathbf{n}_s \cdot (\mathbf{r}_s - \mathbf{R}_i)) \left(\left(\frac{\partial \mathbf{r}_s}{\partial R_{m,x^k}} - \delta_{im} \mathbf{e}^k \right) \cdot (\mathbf{r}_s - \mathbf{R}_i) \right) S_s}{|\mathbf{r}_s - \mathbf{R}_i|^{n+2}} \right]. \quad (43)$$

1.1.4.2.2 Неполярная компонента энергии сольватации

1.1.4.2.2.1 Основные уравнения для неполярная компоненты энергии

Для расчета неполярной части энергии берутся поверхностные элементы на SAS.

Для подсчета неполярной части энергии используем формулу Bordner, Cavasotto, Abagyan [39]:

$$E_{SAS} = \sum_j \sigma_j S_j + b, \quad (44)$$

где $j=1, \dots, N_s$ по всем поверхностным элементам на SAS с ненулевой площадью.

Для воды:

$$\sigma_j = \sigma = 0.00378 \text{ - в ккал/(моль } \text{Å}^2)$$

$$b=0.698 \text{ - в ккал/моль} \quad (45)$$

1.1.4.2.2 Градиенты неполярной составляющей энергии сольватации

Градиент энергии неполярной составляющей:

$$\frac{\partial \Delta G_{nonpol}}{\partial R_{m,x^k}} = \sum_{j=1}^N \sigma_j \frac{\partial s_j}{\partial R_{m,x^k}} \quad (46)$$

Индекс $j=1, \dots, N_g$ пробегает по всем поверхностным элементам на SAS с ненулевым градиентом.

1.1.4.3 Компьютерная проверка расчетов

Была проведена компьютерная проверка найденных выше формул с помощью программы DISOLV [21,33,65,68,91]:

1.1.4.3.1 Сравнение численных и аналитических градиентов для методов PCM, COSMO и SGB (Рис.2)

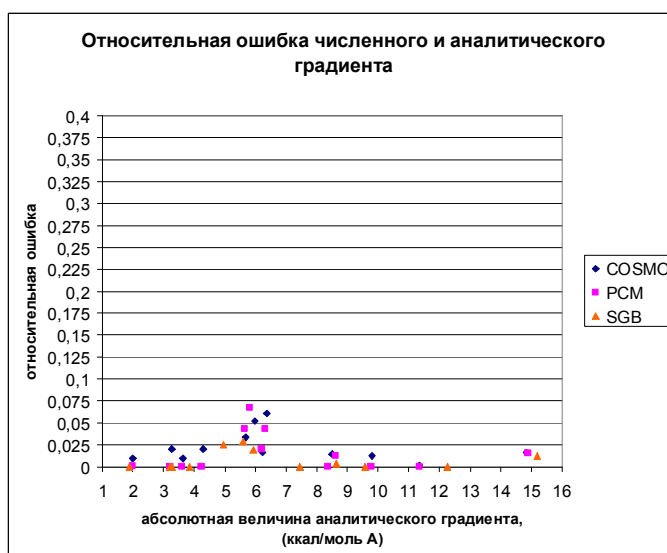


Рис.12 Относительная ошибка численного и аналитического градиента для методов PCM, COSMO и SGB (для шага сетки 0.1) По оси x – величина численного градиента энергии (COSMO или SGB), по оси y – относительная погрешность энергии по отношению к аналитическому градиенту (COSMO или SGB, соответственно):

$$\delta^{COSMO} = \frac{|\nabla_{R_j} E_{num} - \nabla_{R_j} E_{anal}|}{|\nabla_{R_j} E_{num}|}$$

Для метода PCM, COSMO и SGB получены следующие результаты по сравнению численных градиентов с аналитическими:

1) Для PCM:

относительная погрешность численной производной по отношению к аналитической для градиентов больших 2 ккал/(моль А) меньше 6,6 % и в среднем равна 2% Для объемных

атомов погрешность численной производной по отношению к аналитической для градиентов больших 2 ккал/(моль А) меньше 0.03%.

2) Для SGB:

относительная погрешность численной производной по отношению к аналитической для градиентов больших 2 ккал/(моль А) меньше 3% и в среднем равна 1,5%. Для объемных атомов погрешность численной производной по отношению к аналитической для градиентов больших 2 ккал/(моль А) меньше 0.03%.

3) Для COSMO:

относительная погрешность численной производной по отношению к аналитической для градиентов больших 2 ккал/(моль А) меньше 6% и в среднем равна 3%

Для объемных атомов погрешность численной производной по отношению к аналитической для градиентов больших 2 ккал/(моль А) меньше 0.2%.

Ошибки численных градиентов объемных атомов определяются погрешностью вычисления энергии, а поверхностных атомов – прежде всего сдвигом сетки, происходящем при сдвиге атома.

1.1.4.3.2 Сравнение аналитических градиентов для COSMO и SGB с PCM

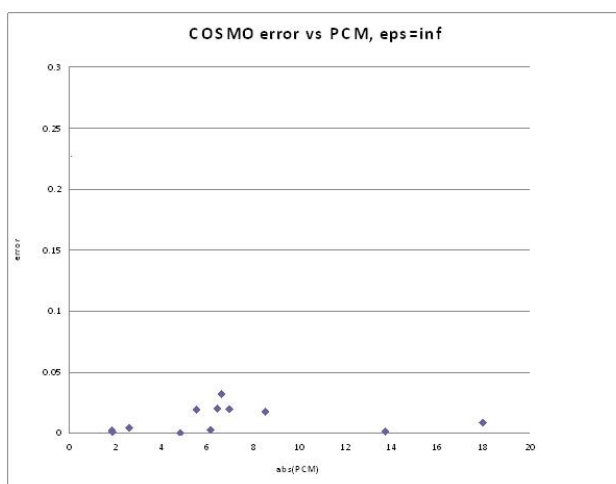


Рис.13 Относительная ошибка аналитических градиентов COSMO в сравнении с PCM для $\epsilon \rightarrow \infty$. По оси x – величина градиента PCM, по оси y – относительная

погрешность по отношению к методу COSMO:
$$\delta = \frac{|\nabla_{R_j} E_{COSMO} - \nabla_{R_j} E_{PCM}|}{|\nabla_{R_j} E_{PCM}|}$$
.

1) аналитические градиенты для COSMO по сравнению с PCM (Рис.3):

а) Для диэлектрической проницаемости стремящейся к бесконечности $\epsilon \rightarrow \infty$ и размера шага сетки стремящегося к нулю, аналитические градиенты для COSMO и PCM должны точно совпадать (т.к. COSMO является пределом PCM при бесконечной диэлектрической проницаемости и бесконечно малом шаге сетки). Действительно, для COSMO для $\epsilon \rightarrow \infty$ относительная погрешность аналитической производной по отношению к аналитической производной для PCM и для градиентов больших 2 ккал/(моль А) меньше 2.5% и в среднем равна 1.2%.

Максимальные значения относительной погрешности производной соответствует поверхностным атомам.

- б) Для COSMO для $\epsilon=78.5$ (вода) относительная погрешность аналитической производной по отношению к аналитической производной для PCM немного выше, а именно, для градиентов больших 2 ккал/(моль А) меньше 4% и в среднем равна 2%.

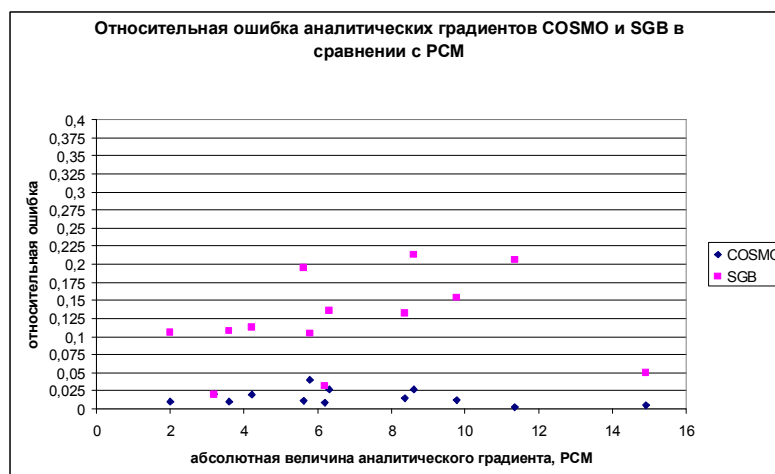


Рис. 14 Относительная ошибка аналитических градиентов COSMO и SGB в сравнении с PCM для $\epsilon=78.5$. По оси x – величина аналитического градиента PCM, по оси y – относительная погрешность по отношению к методу (COSMO или SGB):

$$\delta = \frac{|\nabla_{R_j} E_{(COSMO, SGB)} - \nabla_{R_j} E_{PCM}|}{|\nabla_{R_j} E_{PCM}|}$$

- 2) Сравним аналитические градиенты для SGB с PCM (Рис.4):

Для SGB для $\epsilon=78.5$ (вода) Относительная погрешность аналитической производной по отношению к аналитической производной для PCM много выше, а именно, для градиентов больших 2 ккал/(моль А) она меньше 23% и в среднем равна 13%. Это неудивительно, поскольку SGB это приближенная эмпирическая модель.

1.1.4.4 Выводы

В работе проведены расчеты аналитических градиентов энергии на основе методов PCM, COSMO и SGB для молекулярной поверхности, полученной методами первичной и вторичной обкатки в соответствии с алгоритмами, изложенными в [20,21,29-33,65,68,91]. Приведем проблемные случаи, которые могут возникать при расчете градиентов и решении задачи оптимизации геометрии молекул с их использованием:

- Вырождение - тройная точка опирается не на три, а на большее число атомов (например, ароматические кольца)
- Происходит резкая перестройка поверхности из-за изменения радиуса вторичной обкатки или критического расстояния самой программой при автоматической его настройке

- Происходит резкая перестройка поверхности из-за появления или исчезновения тора обкатки (первичного или вторичного) при сдвиге тройных точек или изменения узкого перешейка тора на расстояниях, близких к критическому
- Происходит резкая перестройка поверхности из-за появления или исчезновения тройных точек первичных или вторичных.

Как результат могут быть скачки градиентов и «колебания» оптимизатора вблизи вышеописанных точек.

Были проведены компьютерные расчеты для подтверждения правильности формул для градиентов энергии, полученные разными методами. Хорошее совпадение аналитических градиентов между собой для разных методов, а также хорошее совпадение численных градиентов с аналитическими для метода COSMO подтверждает правильность найденных формул.

Глава 1.1.5 Укрупненные поверхностные элементы и условия нормировки для столбцов и строк матриц метода COSMO

В более ранней работе Тотрова и Абагына были найдены условия нормировки столбцов матрицы PCM и разработан метод укрупненных поверхностных элементов. Мы разработали аналогичные методы для метода COSMO [106]. Эти методы позволяют ввести более крупные поверхностные элементы без потери точности, а также позволяют находить расчетно-быстрые приближенные значения энергии сольватации и радиусов Борна для метода SGB (*Surface Generalized Born*). Кроме того, предложенные в данной работе корректировки численных результатов позволили значительно увеличить точность вычислений.

1.1.5.1 Введение

Хорошо известно, что адекватный учет взаимодействия заряженной молекулы и окружающего ее растворителя (чаще всего воды) очень важен при моделировании различных химических систем, в частности, биомолекул. Существует несколько различных подходов к моделированию растворителя. В данной работе мы остановим внимание на моделях, в которых растворитель рассматривается как сплошная среда с однородной диэлектрической проницаемостью, а молекула – как область внутри этой среды также с однородной диэлектрической проницаемостью, причем на границе области диэлектрическая проницаемость терпит разрыв. При таком подходе решение электростатической задачи сводится к нахождению поверхностных зарядов, индуцированных внесением заряженной молекулы в диэлектрик. Ниже приводится краткое описание двух таких моделей – PCM [99] и COSMO [94].

В работе рассматриваются условия нормализации для столбцов (строк) матриц COSMO (*COnductor-like Screening MOdel*) [94]. Эти методы позволяют найти точные значения диагональных элементов симметричной матрицы COSMO и обеспечить выполнение условия, связывающего заряд внутри полости с поверхностным зарядом. Также рассмотренные условия позволяют разработать приближенный метод укрупненных поверхностных элементов для COSMO, аналогичный методу укрупненных поверхностных элементов для PCM (*Polarized Continuum Model*) [45,99]. Этот метод должен давать лучший результат, чем в случае PCM, поскольку его матричные элементы зависят лишь от координат поверхностных элементов и менее чувствительны к ошибкам укрупнения размеров поверхностных элементов.

Эта часть построена следующим образом. В первой части нашей работы мы даем краткое описание условий нормировки матриц PCM и метод укрупненных поверхностных элементов, разработанный в [99]. Во второй части мы определяем условия нормировки

для матриц COSMO и находим формулы для диагональных элементов симметричной матрицы. В третьей части мы определяем метод укрупненных элементов для COSMO.

1.1.5.2 PCM-метод с укрупненными поверхностными элементами для $\varepsilon=\infty$

Метод PCM [45,99] (модель поляризуемой диэлектрической среды) представляет собой метод нахождения точного решения электростатической задачи. В этой задаче точечные заряды, соответствующие заряженным атомам молекулы растворяемого вещества, окружены непрерывным диэлектриком с заданной диэлектрической проницаемостью ε . В этом случае электрическое поле в диэлектрической среде будет равно некому электрическому полю. Это то поле, которое создавалось бы в вакууме зарядами внутри и поверхностными зарядами, индуцированными молекулой растворяемого вещества на поверхности, разделяющей диэлектрик-растворитель и молекулу. Таким образом, вместо решения уравнения Пуассона в трехмерном пространстве необходимо найти поверхностные заряды.

Рассмотрим точную постановку задачи о взаимодействии точечных зарядов внутри диэлектрической полости с диэлектриком. Пусть имеются две односвязные области пространства, разделённые замкнутой поверхностью Ξ . Внутренняя область Ω_{in} имеет диэлектрическую проницаемость $\varepsilon_{in}=1$. Внешняя область Ω_{ex} диэлектрическую проницаемость $\varepsilon=\infty$. Стоит отметить, что метод PCM вполне успешно может быть применен и в случае конечной диэлектрической проницаемости внешней среды. Однако для единообразия с дальнейшим описанием изложим его в предельном случае $\varepsilon=\infty$. Во внутренней части расположена система зарядов Q_i , находящихся в точках \mathbf{R}_i . В результате взаимодействия зарядов с диэлектриком на границе раздела областей индуцируется поверхностный заряд плотности σ . Электростатический потенциал может быть выражен как сумма потенциалов точечных зарядов внутри и поверхностных зарядов:

$$\varphi(\mathbf{r}) = \sum_i \frac{Q_i}{\varepsilon_{in} |\mathbf{r} - \mathbf{R}_i|} + \int_{\Xi} \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' . \quad (1)$$

Производная вдоль нормали на поверхности с внутренней стороны

$$\left(\frac{\partial \varphi_{in}(\mathbf{r})}{\partial n} \right) = \left(- \int_{\Xi} \frac{\sigma(\mathbf{r}')((\mathbf{r} - \mathbf{r}') \cdot \mathbf{n}(\mathbf{r}'))}{|\mathbf{r} - \mathbf{r}'|^3} dS' + 2\pi\sigma \right) - \sum_i \frac{Q_i((\mathbf{r} - \mathbf{R}_i) \cdot \mathbf{n}(\mathbf{r}))}{\varepsilon_{in} |\mathbf{r} - \mathbf{R}_i|^3} . \quad (2)$$

Поверхностный интеграл является сингулярным при $\mathbf{r}=\mathbf{r}'$, в этом случае определим его следующим образом:

$$\int_{\Xi} \frac{\sigma(\mathbf{r}')((\mathbf{r} - \mathbf{r}') \cdot \mathbf{n}(\mathbf{r}'))}{|\mathbf{r} - \mathbf{r}'|^3} dS' = \lim_{\delta \rightarrow 0} \int_{\Xi / (|\mathbf{r} - \mathbf{r}'| < \delta)} \frac{\sigma(\mathbf{r}')((\mathbf{r} - \mathbf{r}') \cdot \mathbf{n}(\mathbf{r}'))}{|\mathbf{r} - \mathbf{r}'|^3} dS' . \quad (3)$$

Из теоремы Остроградского – Гаусса

$$4\pi\sigma = \frac{\partial \varphi_{in}(\mathbf{r})}{\partial n} . \quad (4)$$

Из уравнений (2) и (4) получаем интегральное уравнение для σ :

$$4\pi\sigma(\mathbf{r}) = \left(- \int_{\varepsilon} \frac{\sigma(\mathbf{r}')((\mathbf{r}-\mathbf{r}') \cdot \mathbf{n}(\mathbf{r}'))}{|\mathbf{r}-\mathbf{r}'|^3} dS' + 2\pi\sigma(\mathbf{r}) \right) - \sum_j \frac{Q_j((\mathbf{r}-\mathbf{R}_j) \cdot \mathbf{n}(\mathbf{r}))}{\varepsilon_{in}|\mathbf{r}-\mathbf{R}_j|^3}. \quad (5)$$

Для численного решения этого интегрального уравнения введем дискретизацию – разобьем поверхность на N малых поверхностных элементов, которые можно считать точечными зарядами. Пусть q_i – заряд i -ого поверхностного элемента, $q = \{q_{ij}\}_{i=1,N}$ – вектор-столбец зарядов поверхностных элементов, S_i – площадь i -ого поверхностного элемента, \mathbf{r}_i – вектор, определяющий положение центра i -ого поверхностного элемента, \mathbf{n}_i – вектор нормали к i -ому поверхностному элементу, проведенный через его центр и направленный в сторону диэлектрика-растворителя. Число точечных зарядов внутри молекулы обозначим за M , $Q = \{Q_{jj}\}_{j=1,M}$ – вектор-столбец зарядов атомов молекулы. Тогда уравнение (5) может быть записано в матричной форме:

$$4\pi q = ((-Aq + 2\pi q) + 2\pi q) + BQ,$$

Здесь первый интеграл для удобства записывается в виде $(-Aq + 2\pi q)$. Тогда после сокращения одинаковых слагаемых получаем уравнение

$$Aq = BQ, \quad (6)$$

где $A = \{a_{ij}\}$ – квадратная матрица размера $N \times N$, $B = \{b_{ij}\}$ – матрица размера $N \times M$, зависящая от геометрических параметров поверхностных элементов и положения зарядов Q_i атомов молекулы. Матричные элементы матрицы B имеют вид:

$$b_{ij} = - \frac{\mathbf{n}_i \cdot (\mathbf{r}_i - \mathbf{R}_j)}{|\mathbf{r}_i - \mathbf{R}_j|^3} S_i. \quad (7)$$

Для столбцов матрицы B должно выполняться следующее тождество:

$$\sum_{i=1}^N b_{ij} = -4\pi. \quad (8)$$

Это тождество может быть использовано для коррекции численных ошибок дискретизации, возникающие при разбиении на конечные поверхностные элементы:

$$b_{ij} \rightarrow -4\pi \frac{b_{ij}}{\sum_{i=1}^N b_{ij}}.$$

Использование такой коррекции в программе DISOLV [20,33,65] дало значительное увеличение точности. Матричные элементы матрицы A имеют следующий вид [45,99]:

$$a_{ij} = \begin{cases} \frac{n_i(\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} S_i, & i \neq j, \\ a_{jj} = 4\pi - \sum_{i \neq j} a_{ij}, & i = j. \end{cases} \quad (9)$$

Геометрический смысл $2\pi - \sum_{i \neq j} a_{ij}$, – поправка на кривизну поверхности ($a_{jj}=2\pi$ для плоского малого элемента). Из выражения (9) следует условие нормировки для столбцов матрицы A:

$$\sum_i a_{ij} = 4\pi \cdot \quad (10)$$

Из уравнения (6) путем суммирования всех строк, а затем использованием тождеств (8) и (10), получаем тождество, связывающее суммарный поверхностный заряд с суммарным зарядом внутри полости:

$$\sum_{j=1}^N q_j = -\sum_{i=1}^M Q_i \cdot \quad (11)$$

Для убыстрения РСМ может быть использован метод Тотрова и Абагына с увеличенными поверхностными элементами [99]. Он описывается теми исходными интегральными уравнениями (6-10), что и обычная РСМ модель, но при дискретизации поверхностные элементы берутся существенно более крупными. Введем эти укрупненные поверхностные элементы следующим образом. Рассмотрим все поверхностные атомы. Объединим все малые поверхностные элементы в группы, состоящие из поверхностных элементов, наиболее близких к данному поверхностному атому.

Будем считать, что поверхностная плотность заряда одинакова во всех точках для любого укрупненного поверхностного элемента. Тогда уравнение РСМ с укрупненными поверхностными элементами в матричной форме можно записать в следующем виде:

$$Rq^{big} = EQ, \quad (12)$$

где q^{big} – столбец зарядов укрупненных поверхностных элементов.

$E=\{e_{ij}\}$ – матрица размера $N_{\{surf\}} \times M$, $R=\{R_{ij}\}$ – квадратная матрица размера $N_{\{surf\}} \times N_{\{surf\}}$. $N_{\{surf\}}$ – число укрупненных поверхностных элементов.

Матричные элементы для укрупненных элементов будут вычисляться по формулам:

$$\left\{ \begin{array}{l} R_{jk} = \frac{\left(\sum_{l_j} \sum_{m_k} a_{l_j m_k} S_{m_k} \right)}{\sum_{m_k} S_{m_k}}, \quad j \neq k, \\ R_{kk} = 4\pi - \sum_{j \neq k} R_{jk}, \quad j = k, \\ e_{ji} = \sum_{l_j} b_{l_j i}. \end{array} \right. \quad (13)$$

Суммирование здесь проводится по индексам l_j, m_k , которые соответствуют всем малым поверхностным элементам, попавшим в один крупный поверхностный элемент с номером $j(k)$. Здесь S_{l_j} - площадь малых поверхностных элементов, $a_{l_j m_k}$ - матричные элементы для малых поверхностных элементов (9), а матричные элементы $b_{l_j i}$ для малых поверхностных элементов вычисляются по формуле (7).

Для элементов e_{ji} выполняется условие нормализации, аналогичное (8)

$$\sum_j e_{ji} = -4\pi. \quad (14)$$

Использование его для коррекции ошибок

$$e_{ij} \rightarrow -4\pi \frac{e_{ij}}{\sum_{i=1}^N e_{ij}}$$

в программе DISOLV [20,33,65] дало значительное увеличение точности.

1.1.5.3 Вывод уравнения COSMO

Метод COSMO (модель экранирования, подобного экранированию в проводниках) – метод решения электростатической задачи, описанной в начале предыдущего раздела. В этой задаче предполагается, что диэлектрическая проницаемость внутри молекулы равна 1, а диэлектрическая проницаемость внешней среды бесконечна: $\epsilon_{out} = \infty$. В этом предельном случае внешнюю среду можно считать проводником. Таким образом, задача сводится к рассмотрению полости внутри бесконечного проводника, что, разумеется, упрощает ее.

Как и в предыдущем разделе, разобьем поверхность на малые элементы, расположение в точках \mathbf{r}_i и имеющие заряд q_i . Из общих соображений электростатики потенциал поверхности и внешней области одинаков и принимается за ноль. Отсюда получаем уравнение COSMO [94]:

$$\sum_{k=1}^N \frac{q_k}{|\mathbf{r}_m - \mathbf{r}_k|} + \left(\sum_i \frac{Q_i}{|\mathbf{r}_m - \mathbf{R}_i|} \right) + \phi_{mm} = 0, \quad (15)$$

где ϕ_{mm} - собственный потенциал поверхностного элемента. В матричной форме его можно переписать так:

$$A_{(COSMO)} q = -B_{(COSMO)} Q, \quad (16)$$

где $A_{(COSMO)}$ - симметричная матрица с элементами

$$a_{mk}^{(COSMO)} \approx |\mathbf{r}_m - \mathbf{r}_k|^{-1}, m \neq k, \quad (17)$$

$$a_{mm}^{(COSMO)} = \frac{\phi_{mm}}{q_m},$$

а элементы матрицы $B_{(COSMO)}$ задаются следующим выражением:

$$b_{mj}^{(COSMO)} \approx |\mathbf{r}_m - \mathbf{R}_j|^{-1}. \quad (18)$$

Для малого плоского поверхностного элемента собственный потенциал выражается следующей формулой [20,33,65,94]:

$$\phi_{mm} \approx q_m \frac{2\sqrt{3.83}}{\sqrt{S_i}}. \quad (19)$$

Для малых почти плоских поверхностных элементов это выражение дает хорошее приближение, как свидетельствуют расчеты с помощью программы DISOLV [20,33,65].

Мы, однако, заинтересованы в более точном выражении, применимом для собственного потенциала более крупного поверхностного элемента. Это выражение должно учитывать его кривизну и выражаться через параметры самих поверхностных элементов аналогично формуле (10).

Найдем точное условие нормировки для столбцов (строк) симметричной матрицы $A_{(COSMO)}$ и $B_{(COSMO)}$, которое позволит нам получить точное значение диагональных поверхностных элементов. Эта задача сводится к зеркальной электростатической задаче - нахождению распределения ненулевого заряда q_m^{Norm} по исходной полости, заполненной металлом, с вакуумом вовне, как мы увидим ниже. Такое распределение заряда находится из следующих условий:

- I) Напряженность электрического поля на поверхности полости (с внутренней стороны) равна нулю.
- II) Напряженность электрического поля внутри полости (в том числе в точках, где были размещены точечные заряды) равна нулю.
- III) Потенциал внутри полости (в том числе в точках, где были размещены точечные заряды) равен константе. Потенциал на бесконечности принимаем за ноль.
- IV) Потенциал на поверхности полости равен той же константе.

Алгоритм решения этой задачи подробно изложен в Приложении.

1.1.5.3.1 Коррекция ошибок дискретизации для матрицы $B_{(COSMO)}^T$ из условия III)

Из-за ошибок дискретизации имеем лишь приближенное равенство:

$$B_{(COSMO)}^T q^{Norm} \approx 1_M C_{val}. \quad (20)$$

, где 1_M – столбец из M единиц

В развернутом виде:

$$\sum_m b_{(COSMO)mi} q_m^{Norm} \approx C_{val} \quad \forall i. \quad (21)$$

Отсюда определяем точное значение константы C_{val} :

$$C_{val} = \frac{1}{M} \sum_{i=1}^M \sum_m b_{(COSMO)mi} q_m^{Norm}. \quad (22)$$

Коррекция ошибок дискретизации для матрицы $B_{(COSMO)}$ производится следующим образом:

$$b_{(COSMO)mi} = b_{(COSMO)mi} \frac{C_{val}}{\sum_m b_{(COSMO)mi} q_m^{Norm}}. \quad (23)$$

Теперь равенства (20-21) выполняются точно, а не приближенно.

1.1.5.3.2 Подсчет диагональных элементов матрицы $A_{(COSMO)}$

Условие IV) можно записать в матричной форме:

$$A_{(COSMO)} q^{Norm} = I_N C_{val}, \quad (24)$$

где I_N - столбец из N единиц. Тогда из этого условия можно найти диагональные элементы матрицы A и, соответственно, собственные потенциалы поверхностных элементов:

$$a_{(COSMO)ii} \approx \frac{2\sqrt{3.83}}{\sqrt{S_i}}, \text{ если } q_i^{Norm} \ll \frac{\sum_j q_j^{Norm}}{N}, \quad (25)$$

Иначе

$$\sum_{j \neq i} a_{(COSMO)ij} q_j^{Norm} + a_{(COSMO)ii} q_i^{Norm} = C_{val} \quad (26)$$

и

$$a_{(COSMO)ii} = \frac{C_{val} - \sum_{j \neq i} a_{(COSMO)ij} q_j^{Norm}}{q_i^{Norm}}. \quad (27)$$

1.1.5.3.3 Условие нормализации поверхностных зарядов COSMO

Домножая обе части уравнения (16) на $(q^{Norm})^T$, получаем:

$$(q^{Norm})^T A_{(COSMO)} q = -(q^{Norm})^T B_{(COSMO)} Q. \quad (28)$$

Следовательно, из (20) и (24)

$$I_N q = -I_M Q. \quad (29)$$

В развернутом виде:

$$\sum_{j=1}^N q_j = -\sum_{i=1}^M Q_i. \quad (30)$$

Использование (23) и (27) приводит к уменьшению численных ошибок. Кроме того, оно ведет к автоматическому выполнению теоретического условия на суммарный заряд поверхности (30).

1.1.5.3.4 Алгоритм решения

Уравнение COSMO для зарядов поверхностных элементов решалось с помощью итерационной схемы метода сопряженных градиентов, используемого для линейных уравнений, определяемых симметричной положительной определенной матрицей [107]. Метод сопряженных элементов минимизирует энергию, описываемую следующей квадратичной формой

$$\Delta G_{pol}(q_v) = \frac{1}{2} q_v^T A_{(COSMO)} q_v - f^T q_v, \quad (31)$$

где $f = -B_{(COSMO)} Q$ - столбец правой части уравнения COSMO, q_v - столбец зарядов поверхностных элементов. Минимум квадратичной формы q соответствует решению уравнения COSMO.

Начальные шаги итерации выбираются так:

$$q^{(0)} = 0, p^{(0)} = f, r^{(0)} = f. \quad (32)$$

Шаги итерации:

$$\left\{ \begin{array}{l} q^{(i+1)} = q^{(i)} + \alpha_i p^{(i)}, \\ q^{(i+1)} = q^{(i)} - \alpha_i A p^{(i)}, \left(\alpha_i = \frac{(r^{(i)})^T r^{(i)}}{(p^{(i)})^T A p^{(i)}} \right), \\ p^{(i+1)} = r^{(i+1)} + \beta_i p^{(i)}, \left(\beta_i = \frac{(r^{(i+1)})^T r^{(i+1)}}{(r^{(i)})^T r^{(i)}} \right), \end{array} \right. \quad (33)$$

где $q^{(i)}$ - столбец зарядов поверхностных элементов, $p^{(i)}$ и $r^{(i)}$ - столбцы вспомогательных переменных.

Используя выражение для энергии (31) на каждом шаге итерации, получаем ошибку энергии по отношению к ее значению в минимуме, даваемым столбцом q :

$$\Delta G_{pol}^{(i)} = -\frac{1}{2} (q^{(i)})^T r^{(i)} + \frac{1}{2} q^T r^{(i)}. \quad (34)$$

Верхняя граница ее модуля определяется по формуле:

$$|\Delta G_{pol}^{(i)}| < \frac{1}{2} \|r^{(i)}\| \left(\sum_j |Q_j| \right) + \frac{1}{2} |(q^{(i)})^T r^{(i)}| \quad (35)$$

Мы использовали здесь следующее неравенство: $\sum_i |q_i| \leq |Q_n| + |Q_p|$, где Q_n и Q_p – соответственно, сумма отрицательных и положительных зарядов в полости. Действительно, любая силовая линия, выходящая из заряда на поверхности, может закончиться только на заряде внутри полости, поскольку поверхность полости и внешняя область эквипотенциальны; это условие нарушится, если силовая линия закончиться на полости или уйдет во внешнюю область. Поэтому величина положительной компоненты заряда полости не превышает абсолютную величину суммы всех отрицательных зарядов в полости. Аналогичное утверждение верно и для отрицательной компоненты.

1.1.5.3.5 Производные диагональных элементов матрицы $A_{(COSMO)}$

Часто возникает потребность посчитать производную диагональных элементов матрицы $A_{(COSMO)}$ по какому-нибудь параметру α , например, среднему радиусу молекулы, координатам атомов и т.д. Прямое дифференцирование выражения (27) весьма сложно, поскольку само распределение поверхностных зарядов q^{Norm} сложным образом зависит от параметра α . Однако производную диагональных элементов матрицы $A_{(COSMO)}$ можно найти по-другому:

$$a_{(COSMO)ii} \approx \frac{C_s}{\sqrt{S_i}} \quad (36)$$

$$\frac{\partial a_{(COSMO)ii}}{\partial \alpha} = -C_s \frac{\frac{\partial S_i}{\partial \alpha}}{2\sqrt{S_i^3}} \approx -a_{(COSMO)ii} \sqrt{S_i} \frac{\frac{\partial S_i}{\partial \alpha}}{2\sqrt{S_i^3}} = -\frac{a_{(COSMO)ii}}{2S_i} \frac{\partial S_i}{\partial \alpha}$$

Отметим, что в полученное выражение не входит константа C_s , для которой имеются только приближенные оценки.

1.1.5.4 Укрупненные поверхностные элементы для COSMO

1.1.5.4.1 Расчет величины поверхностных элементов

Укрупненные элементы формируются так же, как в методе Тоторова и Абагяна для укрупненных поверхностных элементов РСМ[99]. Затем необходимо найти уравнение COSMO укрупненных поверхностных элементов. Делаем это совершенно аналогично методу Тоторова и Абагяна [99] для укрупненных поверхностных элементов РСМ:

- 1) Объединяем малые поверхностные элемента в более крупные. Плотность заряда на всех малых элементах, относящихся к одному крупному, считаем одинаковой
- 2) Находим новые матричные элементы для L (число поверхностных атомов) больших поверхностных элементов:

$$Yq_{(COSMO)}^{big} = -TQ, \quad (37)$$

$$\left\{ \begin{array}{l} y_{jk} = \frac{\left(\sum_{l_j} \sum_{m_k} a_{(COSMO)l_j m_k} S_{m_k} \right)}{\sum_{m_k} S_{m_k}} \quad j \neq k \cdot \\ t_{ji} = \sum_{l_j} b_{(COSMO)l_j i} \end{array} \right. \quad (38)$$

С целью нахождения диагональных элементов матрицы Y и нормализации матрицы T снова нужно решить задачу о распределении ненулевого заряда $(q^{Norm})^{big}$. Для решения этой задачи вполне достаточно следствий из условия II) предыдущего пункта - силы в точках, где были размещены точечные заряды, равны нулю. При его реализации мы снова разбиваем большие элементы на малые, считая поверхностную плотность заряда в границах больших поверхностных элементов постоянной:

$$\sum_j (q_j^{Norm})^{big} \sum_{l_j} \frac{1}{|R_i - r_{l_j}|^3} (R_i - r_{l_j}) = 0 \quad \forall i. \quad (39)$$

Для решения уравнений (39) уже применим метод ортогонализации Грама - Шмидта для нахождения $(q^{Norm})^{big}$, поскольку число поверхностных элементов невелико. Можно, однако, использовать и итерационный метод из Приложения. Из условия III) получаем:

$$T^T (q^{Norm})^{big} = I_M C_{val}, \quad (40)$$

где I_M – вектор единиц длиной M . В развернутом виде уравнение (40) выглядит следующим образом:

$$\sum_m t_{mi} (q_m^{Norm})^{big} \approx C_{val} \quad \forall i. \quad (41)$$

Это уравнение выполняется приближенно из-за ошибок дискретизации. Отсюда

$$C_{val} = \frac{1}{M} \sum_{i=1}^M \sum_m t_{mi} (q_m^{Norm})^{big}. \quad (42)$$

Коррекция ошибок дискретизации для матрицы T производится по формуле

$$t_{mi} = t_{mi} \frac{C_{val}}{\sum_m t_{mi} (q_m^{Norm})^{big}}. \quad (43)$$

Из условия IV) имеем:

$$Y(q^{Norm})^{big} = I_L C_{val}, \quad (44)$$

где I_L - столбец из единиц длиной L . Тогда из этого условия можно найти диагональные элементы матрицы. Y и, соответственно, собственные потенциалы поверхностных элементов:

$$\sum_{j \neq i} y_{ij} (q_j^{Norm})^{big} + y_{ii} (q_i^{Norm})^{big} = C_{val}, \quad (45)$$

откуда для диагональных элементов

$$y_{ii} = \frac{C_{val} - \sum_{j \neq i} y_{ij} (q_j^{Norm})^{big}}{(q_i^{Norm})^{big}}. \quad (46)$$

Домножая обе части уравнения (37) на $((q^{Norm})^{big})^T$, получаем:

$$((q^{Norm})^{big})^T Y q_{(COSMO)}^{big} = -((q^{Norm})^{big})^T T Q, \quad (47)$$

Получим условие нормализации для зарядов

$$(I_L)^T q_{(COSMO)}^{big} = -(I_M)^T Q. \quad (48)$$

Это же можно записать в развернутом виде:

$$\sum_{j=1}^L q_{(COSMO)j}^{big} = -\sum_{i=1}^M Q_i. \quad (49)$$

1.1.5.4.2 Производные диагональных элементов матрицы Y

Введем матрицу T_{surf} для тех номеров i , которые относятся к поверхностным атомам:

$$t_{(surf)ij} = t_{ij}. \quad (50)$$

Производные нормирующих поверхностных зарядов по какому-либо параметру

$\frac{\partial (q^{Norm})^{big}}{\partial \alpha}$ находятся из уравнения (40):

$$T_{surf}^T \frac{\partial (q^{Norm})^{big}}{\partial \alpha} = -\frac{\partial T_{surf}^T}{\partial \alpha} (q^{Norm})^{big}. \quad (51)$$

Производная диагональных элементов матрицы Y (из (44)):

$$\sum_{j \neq i} \frac{\partial y_{ij}}{\partial \alpha} (q_j^{Norm})^{big} + \frac{\partial y_{ii}}{\partial \alpha} (q_i^{Norm})^{big} = - \sum_{j \neq i} y_{ij} \frac{\partial (q_j^{Norm})^{big}}{\partial \alpha} - y_{ii} \frac{\partial (q_i^{Norm})^{big}}{\partial \alpha}. \quad (52)$$

Отсюда получаем:

$$\frac{\partial y_{ii}}{\partial \alpha} = - \frac{\sum_{j \neq i} y_{ij} \frac{\partial (q_j^{Norm})^{big}}{\partial \alpha} + y_{ii} \frac{\partial (q_i^{Norm})^{big}}{\partial \alpha} + \sum_{j \neq i} \frac{\partial y_{ij}}{\partial \alpha} (q_j^{Norm})^{big}}{(q_i^{Norm})^{big}}. \quad (53)$$

где производные зарядов находятся из (51).

1.1.5.5 Выводы

В данной работе получены условия нормализации для столбцов (строк) матриц модели COSMO. Из них находятся точные выражения для диагональных элементов основной матрицы COSMO. Для этих целей необходимо решить электростатическую задачу распределения заряда по металлической поверхности. Разработан метод укрупненных элементов для COSMO, аналогичный подобному методу в PCM [99]. Указанный метод позволит приближенно решать уравнение COSMO и находит радиусы Борна для модели SGB[101].

Приложение 1 Алгоритм подсчета нормалей по координатам поверхностных элементов

В моделях COSMO часто не задаются нормали, а лишь координаты поверхностных элементов. В этом случае нормали могут быть найдены на основе следующего метода. Нормаль элемента ищем как векторную среднюю нормалей к треугольным элементам с вершиной в центральной точке многоугольного элемента с весами, пропорциональными площадям треугольников

$$\mathbf{a}_{gi} = \mathbf{r}_{gi} - \mathbf{r}_g, \quad (54)$$

где \mathbf{r}_g – радиус-вектор центра текущего поверхностного элемента, \mathbf{r}_{gi} – радиус-вектора центров соседних поверхностных элементов, $i=1, \dots, N_{eg}$ – номера соседних поверхностных элементов.

Введем следующее обозначение:

$$\mathbf{a}_{gi_next} = \begin{cases} \mathbf{a}_{g(i+1)} & \text{for } 1 \leq i < N_{eg} \\ \mathbf{a}_{g1} & \text{for } i = N_{eg} \end{cases}. \quad (55)$$

На первом шаге алгоритма.

$$\mathbf{n}_g = \frac{\sum_{i=1}^{N_{eg}} (\mathbf{a}_{gi} \times \mathbf{a}_{gi_next})}{\left| \sum_{i=1}^{N_{eg}} (\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}) \right|}$$

Нормаль текущего поверхностного элемента:

$$\mathbf{n}_g = \frac{\sum_{i=1}^{N_{eg}} (\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}) \text{sign}(\mathbf{n}_{near} [\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}])}{\left| \sum_{i=1}^{N_{eg}} (\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}) \text{sign}(\mathbf{n}_{near} [\mathbf{a}_{gi} \times \mathbf{a}_{gi_next}]) \right|} \quad (56)$$

где \mathbf{n}_{near} - нормаль соседнего поверхностного элемента, используется для согласования направлений нормалей.

При таком вычислении все нормали будут направлены в одну сторону – либо вовнутрь, либо наружу. Однако нам необходимо, чтобы они все были направлены наружу. Для этого объема поверхности вычисляется по следующей формуле:

$$V = \frac{1}{3} \oint_S (\mathbf{r} \cdot \mathbf{n}) dS \quad (57)$$

где \mathbf{r} – радиус-вектор текущей точки поверхности, \mathbf{n} - ее нормаль. Если $V < 0$, то для всех нормалей меняем знак:

$$\mathbf{n}_g = -\mathbf{n}_g. \quad (58)$$

Приложение 2 Алгоритм нахождения распределения ненулевого заряда q^{Norm} по исходной полости, заполненной металлом, и вакуумом вне нее

Опишем алгоритм решения «зеркальной» задачи – нахождения распределения заряда по исходной полости, заполненной металлом, и вакуумом вне нее. Напомним четыре условия, которым должно удовлетворять такое распределение заряда:

- I) Напряженность электрического поля на поверхности полости (с внутренней стороны) равна нулю.
- II) Напряженность электрического поля внутри полости (в том числе в точках, где были размещены точечные заряды) равна нулю.
- III) Потенциал внутри полости (в том числе в точках, где были размещены точечные заряды) равен константе C_{val} . Потенциал на бесконечности принимаем за ноль
- IV) Потенциал на поверхности полости равен той же константе C_{val} .

Тогда уравнения, определяемые условиями I) и II), запишутся как условия ортогональности вектора зарядов известной системе векторов.

Из условия I) для направлений, *перпендикулярных* нормали поверхностного элемента:

$$\left[\sum_{i \neq j} q_j^{Norm} \frac{(\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} \times \mathbf{n}_i \right] = 0. \quad (59)$$

Условия I) для направлений *вдоль* нормалей поверхностных элементов уравнения для поверхности:

$$\left(\frac{\partial \varphi_{in}(\mathbf{r})}{\partial n} \right) = 0. \quad (60)$$

При разбиении на поверхностные элементы из (2) при (60) и при $Q=0$ в матричной форме уравнение (2) выглядит следующим образом (аналогично получению уравнения (6) из (5)):

$$- Aq^{Norm} + 4\pi q^{Norm} = 0. \quad (61)$$

То же самое можно записать через матричные элементы a_{ij} :

$$- \sum_{j \neq i} a_{ij} q_j^{Norm} - a_{ii} q_i^{Norm} + 4\pi q_i^{Norm} = 0 \quad (62)$$

или в развернутом векторном виде:

$$- \mathbf{n}_i S_i \sum_{j \neq i} q_j^{Norm} \frac{(\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} + q_i^{Norm} \left(\sum_{j \neq i} \frac{(\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_j - \mathbf{r}_i|^3} \mathbf{n}_j S_j \right) = 0. \quad (63)$$

Для плоских малых элементов это вырождается в следующее уравнение:

$$-n_i S_i \sum_{j \neq i} q_j^{Norm} \frac{(\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} + q_i^{Norm} 2\pi = 0. \quad (64)$$

Из условия II) для точек расположения центров атомов получаем:

$$\sum_m \frac{q_m^{Norm}}{|\mathbf{R}_i - \mathbf{r}_m|^3} (\mathbf{R}_i - \mathbf{r}_m) = 0 \quad \forall i. \quad (65)$$

Используем итерационный метод для нахождения q^{Norm} из уравнений (58), (62) и (64). Это, по сути, нахождение вектора, ортогонального данному набору векторов.

- (1) В качестве начального приближения мы выбираем вектор $q^{(0)Norm} = S$, где S - столбец из площадей поверхностных элементов, и нормируем его на единицу: $\|q^{(0)Norm}\| = 1$
- (2) Из условия эквипотенциальности поверхности все заряды поверхностных элементов имеют один знак, который выберем положительным $q^{Norm} > 0$. Отсюда $(q^{(0)Norm})^T \cdot q^{Norm} > 0$, значит, у начального приближения есть ненулевая проекция на искомый вектор, что делает его приемлемым для итераций.
- (3) Процесс ортогонализации Грамма – Шмидта [108] будет слишком затратный. Поэтому сделаем процесс поиска итерационным:
 - а) Нормализуем на 1 все $3N+3M$ векторов, входящие в (58), (62) и (64). Назовем этот набор векторов d_k $k=1, \dots, 3N+3M$; $\|d_k\|=1$. Текущий вектор $q_j^{(k-1)Norm}$ будем менять последовательно с каждым вектором из набора: $q_j^{(k)Norm} = q_j^{(k-1)Norm} - ((q_j^{(k-1)Norm})^T d_k) d_k$, $k = 1, \dots, 3N + 3M$. Если какая-то из компонент этого вектора получается меньше нуля $(q_j^{(k)Norm})_m < 0$, то обнуляем ее $(q_j^{(k)Norm})_m = 0$.
 - б) Если $\|q_j^{(3N+3M)Norm}\| \approx 1$, то прекращаем итерации, если нет – возвращаемся на пункт (а) алгоритма с $q_{j+1}^{(0)Norm} = q_j^{(3N+3M)Norm} / \|q_j^{(3N+3M)Norm}\|$.

Глава 1.1.6 Итерационный метод расчета полярной составляющей энергии Гиббса растворения молекул при условии плавного изменения диэлектрической проницаемости раствора

На основе ранее развитой модели формулируется итерационный метод расчета полярной составляющей энергии Гиббса растворения, предполагающий плавное изменение диэлектрической проницаемости как между субстратом и растворителем, так и внутри растворителя [109]. Метод развит в приближении локальной связи между векторами смещения \mathbf{D} и напряженности электрического поля \mathbf{E} : $\mathbf{D} = \varepsilon(\mathbf{r})\mathbf{E}$.

1.1.6.1 Введение

Расчет полярной компоненты энергии Гиббса сольватации молекулы G_{solv} представляет собой важную задачу, для решения которой было создано большое число различных методов [45,94,110,111]. Многие из них используют для расчета G_{solv} построение молекулярной поверхности [27,38,45,60,94], внутри которой диэлектрическая проницаемость имеет одно значение (как правило, близкое к единице), а вовне — другое. Эти методы отличает быстрота и точность. Но они имеют и много недостатков: во-первых, универсальные и точные алгоритмы построения гладких молекулярных поверхностей очень сложны и занимают много компьютерной памяти и времени [27,38,60]. Во-вторых, при малом сдвиге атомов может происходить значительная перестройка поверхности, что приводит к нефизическим скачкам G_{solv} при оптимизации структуры молекулы в растворе. Кроме того, в реальных физических средах изменение диэлектрической проницаемости происходит непрерывно и гладко, а не скачком.

Другой возможный подход состоит в том, чтобы напрямую решать уравнение Пуассона-Больцмана. При этом мы разбиваем все пространство на ячейки. Однако это связано с большими затратами компьютерной памяти и времени [112]. В настоящей работе на основе развитой ранее модели [110,111] дается компактная формулировка метода для расчета G_{solv} в среде с диэлектрической проницаемостью $\varepsilon(\mathbf{r})$, которая является дифференцируемой функцией координат пространства. Предлагаемый метод позволяет рассчитывать энергию сольватации для плавно меняющейся диэлектрической проницаемости путем итерационного процесса, где на каждой итерации берется интеграл лишь по области, где существует ненулевой градиент диэлектрической проницаемости. Это позволяет значительно снизить затраты метода по сравнению с прямым решением уравнения Пуассона-Больцмана. Быстрота сходимости метода обеспечивает хорошие результаты уже для первой итерации.

Теоретические основы метода были заложены в работах [113,114]. Впервые метод с непрерывной диэлектрической проницаемостью был предложен в [111]. В этой работе предлагалось такое решение:

$$\mathbf{E} = -\frac{\nabla\psi_0}{\varepsilon}. \quad (1)$$

где ψ_0 — электростатическое поле, которое создавалось бы молекулой в отсутствие диэлектрика. Однако это решение не удовлетворяет уравнениям Максвелла для электростатики - ротор напряженности электрического поля не равен нулю [114]:

$$\nabla \times \mathbf{E} = \frac{\nabla \varepsilon \times \nabla \psi_0}{\varepsilon^2} \neq 0, \text{ если векторы } \nabla \varepsilon \text{ и } \nabla \psi_0 \text{ не сонаправлены.}$$

На самом деле выражение (1) является лишь *первым* приближением (достаточным, впрочем, на практике) для напряженности электрического поля, которая в действительности представляется полученным нами итерационным рядом. Дальнейшее развитие предложенный метод получил в работе [110], однако основной акцент в ней сделан на возможность нелокальной связи между векторами смещения и напряженности электрического поля.

Эта часть работы построена следующим образом. Вначале приводится описание новой итерационной схемы. Затем приводятся две модели сглаживания диэлектрической среды – одна уже ранее рассмотренная в работе [111], приводящая к разрывам в градиенте, а другая новая, позволяющая проводить бесконечное дифференцирование. В качестве доказательства работоспособности метода в третьей части приведены данные из работы [111] - результаты расчетов G_{solv} для некоторого набора молекул с учетом только первого члена полученного разложения (1) и их сравнение с результатами эксперимента и расчетов другим методом, которые иллюстрируют очень быструю сходимость предлагаемого ряда. Также дается расчет двух следующих членов разложения, подтверждающий этот вывод.

Полярная компонента энергии Гиббса, которая ищется в данной работе, включает не только чисто электростатическую энергию: в нее также входят как компонента, связанная с частичным упорядочиванием диполей вдоль поля (энтропийная составляющая), так и неэлектростатическая энергия «растяжения» диполей вдоль поля. Однако никакого их дополнительного расчета не требуется – данные компоненты уже включены в вычисленную в этой работе энергию. Поскольку даже в серьезной литературе [113] встречается недостаточно ясное понимание этого факта, мы сочли нужным добавить Приложение, проясняющее этот вопрос.

1.1.6.2 Итерационная схема метода

Пусть имеется молекула субстрата, помещенного в раствор. Плотность распределения заряда внутри этой молекулы ρ_0 , а диэлектрическая проницаемость равна 1. Окружающий раствор имеет диэлектрическую проницаемость ε_{out} , но переход на границе субстрата сглаживается. Основные уравнения электростатики:

$$\mathbf{D} = \varepsilon(\mathbf{r}) \mathbf{E}, \quad (2)$$

$$\nabla \cdot \mathbf{D} = 4\pi \rho_0, \quad (3)$$

$$\mathbf{E} = -\nabla \psi. \quad (4)$$

где \mathbf{E} - вектор напряженности электрического поля, \mathbf{D} - вектор смещения электрического поля, ψ - потенциал электрического поля. Из последнего уравнения непосредственно следует, что $\nabla \times \mathbf{E} = 0$.

Под влиянием электрического поля зарядов внутри молекулы происходит поляризация окружающего ее диэлектрика, вследствие чего возникает ненулевое распределение так называемых связанных зарядов в диэлектрике. Обозначим эту плотность распределения ρ . Тогда выполняется следующее соотношение:

$$\nabla \cdot \mathbf{E} = 4\pi(\rho_0 + \rho). \quad (5)$$

Как известно, зная распределение зарядов внутри молекулы ρ и диэлектрическую проницаемость в каждой точке пространства, можно однозначно найти потенциал электрического поля $\psi(\mathbf{r})$ при граничных условиях

$$\lim_{r \rightarrow \infty} r \psi(\mathbf{r}) = \text{const},$$

$$\lim_{r \rightarrow \infty} r^2 |\nabla \psi(\mathbf{r})| = \text{const}.$$

а, следовательно, и вектор напряженности \mathbf{E} . Из уравнения (5) тогда можно найти плотность связанных зарядов ρ , а через плотность выразить энергию взаимодействия молекулы и раствора.

Схему для расчета свободной энергии сольватации можно построить, сохранив локальную связь (1). Запишем полный потенциал в виде:

$$\psi = \frac{\psi_0}{\varepsilon} + \varphi, \quad (6)$$

где ψ_0 является решением следующего уравнения

$$\nabla^2 \psi_0 = -4\pi\rho_0, \quad (7)$$

то есть является потенциалом поля, которое создавалось бы в отсутствие диэлектрика. Тогда для напряженности поля \mathbf{E} и смещения \mathbf{D} из уравнений (4) и (2) получим:

$$\mathbf{E} = -\frac{\nabla \psi_0}{\varepsilon} + \frac{\psi_0 \nabla \varepsilon}{\varepsilon^2} - \nabla \varphi, \quad (8)$$

$$\mathbf{D} = -\nabla \psi_0 + \frac{\psi_0 \nabla \varepsilon}{\varepsilon} - \varepsilon \nabla \varphi.$$

С учетом (7) и (8) условие (3) можно переписать так:

$$\nabla \left(\frac{\psi_0 \nabla \varepsilon}{\varepsilon} - \varepsilon \nabla \varphi \right) = 0. \quad (9)$$

Поскольку дивергенция выражения в скобках равна 0, то это выражение является ротором некоторого векторного поля \mathbf{B} :

$$\nabla \times \mathbf{B} = \frac{\psi_0 \nabla \varepsilon}{\varepsilon} - \varepsilon \nabla \varphi. \quad (10)$$

Отсюда можно выразить $\nabla \varphi = \frac{\psi_0 \nabla \varepsilon}{\varepsilon^2} - \frac{\nabla \times \mathbf{B}}{\varepsilon}$ и подставить в уравнение (8):

$$\mathbf{E} = -\frac{\nabla \psi_0}{\varepsilon} + \frac{\nabla \times \mathbf{B}}{\varepsilon}. \quad (11)$$

Тогда ротор напряженности, который равен нулю, может быть следующим образом выражен через векторное поле \mathbf{B} :

$$\frac{\nabla \varepsilon \times \nabla \psi_0}{\varepsilon} + \varepsilon \nabla \times \left(\frac{\nabla \times \mathbf{B}}{\varepsilon} \right) = \frac{\nabla \varepsilon \times \nabla \psi_0}{\varepsilon} - \frac{\nabla \varepsilon \times (\nabla \times \mathbf{B})}{\varepsilon} + \nabla \times (\nabla \times \mathbf{B}) = 0. \quad (12)$$

Последнее равенство представляет собой уравнение на векторное поле \mathbf{B} . Перепишем его в виде линейного интегрального уравнения:

$$\begin{aligned} \nabla \times \mathbf{B} &= \frac{1}{4\pi} \int \left[\frac{-\frac{\nabla' \varepsilon \times \nabla' \psi_0}{\varepsilon} + \frac{\nabla' \varepsilon \times (\nabla' \times \mathbf{B})}{\varepsilon}, \mathbf{r} - \mathbf{r}' \right]}{|\mathbf{r} - \mathbf{r}'|^3} d^3 \mathbf{r}' = \\ &= \frac{1}{4\pi} \int \left[\nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|}, \left(-\frac{\nabla' \varepsilon \times \nabla' \psi_0}{\varepsilon} + \frac{\nabla' \varepsilon \times (\nabla' \times \mathbf{B})}{\varepsilon} \right) \right] d^3 \mathbf{r}'. \end{aligned} \quad (13)$$

Один из возможных путей решения (13) связан с представлением \mathbf{B} в виде суммы $\mathbf{B} = \mathbf{B}_1 + \mathbf{B}_2 + \dots$, в которой ротор каждого последующего слагаемого существенно меньше ротора предыдущего. При этом считается, что $\nabla \varepsilon$ и $\nabla \times \mathbf{B}_1$ имеют одинаковый (первый) порядок малости. Исходя из предположения, что $\nabla \times \mathbf{B}$ мало по сравнению с $\nabla \psi_0$, первый член ряда может быть выбран следующим образом:

$$\nabla \times \mathbf{B}_1 = \frac{1}{4\pi} \int \left[\frac{-\frac{\nabla' \varepsilon \times \nabla' \psi_0}{\varepsilon}, \mathbf{r} - \mathbf{r}' \right]}{|\mathbf{r} - \mathbf{r}'|^3} d^3 \mathbf{r}' = \frac{1}{4\pi} \int \left[\nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|}, -\frac{\nabla' \varepsilon \times \nabla' \psi_0}{\varepsilon} \right] d^3 \mathbf{r}'. \quad (14)$$

Подставляя (14) в (12) и сохраняя члены одного порядка малости, получим уравнение для следующего члена ряда:

$$-\frac{\nabla \varepsilon \times (\nabla \times \mathbf{B}_1)}{\varepsilon} + \nabla \times (\nabla \times \mathbf{B}_2) = 0. \quad (15)$$

Отсюда можно выразить второй член:

$$\nabla \times \mathbf{B}_2 = \frac{1}{4\pi} \int \left[\frac{\frac{\nabla' \varepsilon \times (\nabla' \times \mathbf{B}_1)}{\varepsilon}, \mathbf{r} - \mathbf{r}' \right]}{|\mathbf{r} - \mathbf{r}'|^3} d^3 \mathbf{r}' = \frac{1}{4\pi} \int \left[\nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|}, \frac{\nabla' \varepsilon \times (\nabla' \times \mathbf{B}_1)}{\varepsilon} \right] d^3 \mathbf{r}'. \quad (16)$$

Таким образом, получаем следующее рекуррентное соотношение для членов ряда:

$$\nabla \times \mathbf{B}_{n+1} = \frac{1}{4\pi} \int \left[\frac{\frac{\nabla' \varepsilon \times (\nabla' \times \mathbf{B}_n)}{\varepsilon}, \mathbf{r} - \mathbf{r}' \right]}{|\mathbf{r} - \mathbf{r}'|^3} d^3 \mathbf{r}'. \quad (17)$$

Используя это соотношение, можно найти вектор $\nabla \times \mathbf{B}$. С учетом (11) можно теперь выразить плотность связанных зарядов через этот вектор из уравнения (5):

Опишем, как должна меняться диэлектрическая проницаемость, чтобы выполнялось условие малости ее градиента. При этом известно, что внутри молекулы она равна 1, а в растворителе - $1 + 4\pi\{\chi_\infty + \chi_{in}\}$. Здесь χ_∞, χ_{in} – константы, большие нуля, отвечающие за быструю (электронную) и медленную (ядерную) диэлектрические проницаемости. За основу расчета диэлектрической проницаемости около субстрата используем модель непрерывной среды, разработанной в [111] (рис. 1):

$$\varepsilon(\mathbf{r}) = 1 + 4\pi z(\mathbf{r})\{\chi_\infty + \chi_{in}\varphi(z(\mathbf{r}))\}, \quad (21)$$

где $z(\mathbf{r})$ – некая функция, однозначно определяемая расстоянием от точки пространства \mathbf{r} до поверхности молекулы, которая подробно описывается ниже, а $\varphi(z) = \exp\left\{\frac{(z-1)}{z_0}\right\}$ (z_0 – константа, большая нуля) (см. рис. 2). В этом случае диэлектрическая проницаемость плавно меняется от 1 внутри молекулы до $1 + 4\pi(\chi_\infty + \chi_{in})$ в растворителе.

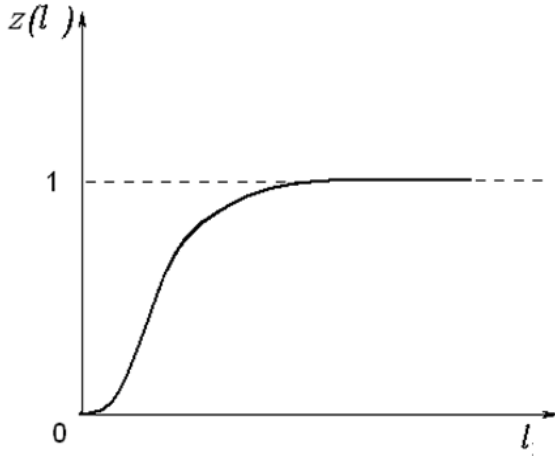


Рис. 2 Поведение функции $z(l)$. Из работы [111].

Зададим теперь поведение функции $z(\mathbf{r})$. Пусть молекулярная поверхность задается сферами Ван-дер-ваальса с радиусами R_j и центрами в атомах, радиус-векторы которых $\mathbf{R}_j; j=1, \dots, N$ – текущий номер атома. Пусть $l(\mathbf{r})$ – расстояние от точки \vec{r} до молекулярной поверхности. Если точка \mathbf{r} не лежит внутри одного из атомов, т.е. если $\|\mathbf{r} - \mathbf{R}_j\| > R_j$ для всех j , тогда $l(\mathbf{r}) = \min_j l_j(\mathbf{r})$, где $l_j(\mathbf{r}) = \|\mathbf{r} - \mathbf{R}_j\| - R_j$, иначе

$$l(\mathbf{r}) = 0. \quad (22)$$

В работе [111] предлагается следующая формула для $z(\mathbf{r})$:

$$z(\mathbf{r}) = 1 + \exp\left\{-\frac{2l(\mathbf{r})}{\delta}\right\} - 2 \exp\left\{-\frac{l(\mathbf{r})}{\delta}\right\}, \quad (23)$$

где δ – малая константа, большая нуля, имеющая смысл ширины перехода. Однако такой выбор $z(\mathbf{r})$ плох тем, что производная этой функции имеет разрыв в точках, где расстояния до двух или более атомов одинаковы.

Для более быстрой сходимости метода в настоящей работе рекомендуется использовать другую гладкую функцию $z(\mathbf{r})$, которая является бесконечно дифференцируемой в любой точке:

$$z(\mathbf{r}) = \frac{\sum_j \exp\left\{-\frac{\delta}{l_j(\mathbf{r})}\right\} \left(1 - \exp\left\{-\frac{\delta_1^2}{(l_j(\mathbf{r}) - l(\mathbf{r}))^2}\right\}\right)}{\sum_j \left(1 - \exp\left\{-\frac{\delta_1^2}{(l_j(\mathbf{r}) - l(\mathbf{r}))^2}\right\}\right)}, \quad (24)$$

причем константы δ , δ_1 удовлетворяют условию $\delta_1 \ll \delta < R_j$.

1.1.6.4 Сравнение результатов расчета полярной энергии, проведенного предлагаемым в работе методом, с экспериментом и с результатами расчета методом РСМ.

В работе [111] для приближения $\nabla \times \mathbf{B} = 0$ были проведены численные расчеты энергии растворения для ряда молекул, которые сравнивались с результатами расчета методом РСМ [45] и экспериментальными данными. Для расчета диэлектрической проницаемости использовалась формулы (21-23) со следующими параметрами:

$$\begin{aligned} z_0 &= 0.5, \\ \delta &= 0.32 \text{ \AA}, \\ \varepsilon_{out} &= 1 + 4\pi\{\chi_\infty + \chi_{in}\} = 78, \\ \varepsilon_\infty &= 1 + 4\pi\chi_\infty = 1.77, \end{aligned}$$

Заряды атомов были взяты из квантово-механических расчетов методом ХФ (6-31G(d,p) базис) согласно ESP схеме зарядовой локализации [115]. Экспериментальные данные приводятся из работ [116,117]. Радиусы были взяты из Таблицы 5 [117] и домножались на $k=0.92$ для предлагаемого метода и на $k=1.06$ для РСМ.

Для выделения полярной составляющей из экспериментальной величины энергии Гиббса растворения ионизированной молекулы вычиталась экспериментальная энергия Гиббса растворения аналогичной по структуре неионизированной молекулы, при этом разница неполярных составляющих близка к нулю. Полученный результат сравнивался с соответствующей разницей рассчитанных полярных компонент энергии Гиббса растворения этих двух молекул. Сравнение полярных энергий Гиббса (см. рис. 3) свидетельствуют о быстрой сходимости рассматриваемого метода – уже первый член полученного нами ряда обеспечивает достаточную точность модели.

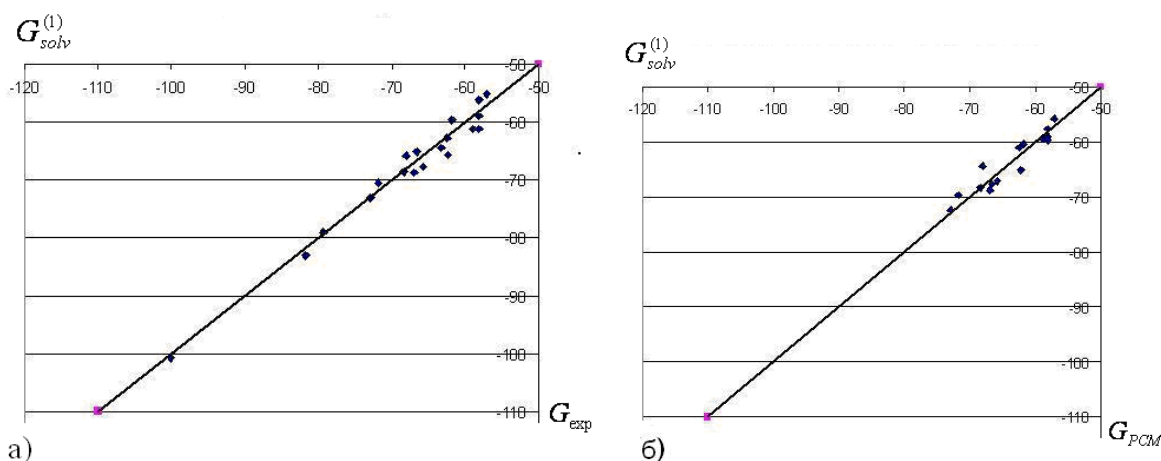


Рис. 3. Для ионов аммониевого типа корреляция величин а) $G_{solv}^{(1)}$ и G_{exp} , среднеквадратичное отклонение 1.75 ккал/моль; б) $G_{solv}^{(1)}$ и G_{PCM} , среднеквадратичное отклонение 1.68 ккал/моль; здесь $G_{solv}^{(1)}$ - энергии Гиббса растворения, полученная при расчете в рамках только одной (первой) итерации, G_{PCM} - то же в рамках РСМ, G_{exp} - пересчитана из эксперимента. Все энергетические величины даны в ккал/моль. Использованы данные из работы [111].

Поскольку может возникнуть вопрос, насколько сильный вклад могут вносить дополнительные члены разложения, был проведен численный расчет первого ($G_{solv}^{(1)}$ из формулы (20)) и двух последующих членов разложения ($G_{solv}^{(2)}$ из формул (14,18,19), $G_{solv}^{(3)}$ из формул (16, 18, 19)) для пяти ионов, уже рассмотренных ранее в [111]. Соответствующие интегралы нужно численно подсчитывать лишь в узкой области вблизи поверхности молекулы. Вне этой области подынтегральная функция близка к нулю, из-за малой величины градиента ϵ . В точках разрыва подынтегральной функции интегралы имеют хорошую математическую сходимость. Эти факты значительно облегчают задачу численного интегрирования. Для расчета заряда атомов в отличие от [111] использовалась модель силового поля MMFF94. Соответственно радиусы из Таблицы 5 [117] домножались на $k=0.8$. Остальные параметры брались теми же, что и в [111]. Результат расчета (в ккал/моль) приводится в Таблице 1.

Таблица 1. Расчетные и экспериментальные энергии сольватации

Ион	$\Delta G_{эксн}[111]$	$G_{solv}^{(1)}$	$G_{solv}^{(2)}$	$G_{solv}^{(3)}$	$\Delta G_{solv}=G_{solv}^{(2)}+G_{solv}^{(3)}$
$(CH_3)_2NH_2^+$	-65,9	-65,89	-3,22	1,67	-1,55
$(CH_3)_3NH^+$	-58,9	-60,05	-2,33	0,66	-1,67
$CH_3(CH_2)_2NH_3^+$	-68,8	-71,30	-8,61	6,01	-2,60
$CH_3CH_2NH_3^+$	-70,4	-72,57	-6,27	5,49	-0,78
$CH_3NH_3^+$	-73,1	-75,23	-3,03	2,02	-1,01

Из приведенных результатов расчета можно видеть, что итоговая поправка, вносимая следующими двумя членами разложения, действительно мала и имеет тот же порядок величины, что и среднеквадратичное отклонение расчетов $G_{solv}^{(1)}$ от эксперимента $\Delta G_{эксн}$ (1.75 ккал/моль, см. подпись к рис. 3). Результаты расчета первой поправки $G_{solv}^{(1)}$ на рис. 3 взяты из [111] и очень близки по величине к полученным в данной работе.

1.1.6.5 Выводы

В данной работе предложена компактная формулировка итерационного метода расчета полярной составляющей энергии Гиббса растворения G_{solv} в случае непрерывного распределения диэлектрической проницаемости раствора. Подобное распределение дает более физически верную модель растворителя на границе с субстратом, предотвращает скачки G_{solv} при сдвиге атомов, позволяет легко рассчитать ситуации с плавным изменением диэлектрической проницаемости растворителя. Предложена гладкая, бесконечно дифференцируемая модель для диэлектрической проницаемости на границе субстрата с растворителем. Сравнение численного расчета уже для первой итерации $G_{solv}^{(1)}$ дает хорошее совпадение с экспериментом и расчетом G_{solv} на основе метода РСМ.

Из результатов расчета двух следующих членов разложения можно видеть, что их сумма мала и имеет тот же порядок величины, что и среднеквадратичное отклонение первой поправки от эксперимента.

Приложение

Часто возникает вопрос о том, какая величина вычисляется с помощью диэлектрической проницаемости (полярная компонента энергии сольватации) – потенциальная электростатическая внутренняя энергия или энергия Гиббса [113,118]. В данном приложении содержится рассмотрение этого вопроса для модельной системы, подтверждающее последнее.

Рассмотрим вначале вопрос качественно. Энергию Гиббса можно записать так:

$$G_{\text{free}} = U_{\text{el}} + U_{\text{nonel}} - TS + pV, \quad (25)$$

где U_{el} – потенциальная электростатическая энергия:

$$U_{\text{el}} = \int \frac{E^2}{8\pi} dV, \quad (26)$$

U_{nonel} – неэлектростатическая энергия (т.е. потенциальная энергия натяжения «пружинок», которая присутствует, когда диполь в диэлектрике описывается как два заряда, соединенные пружинкой, и отсутствует для жестких вращающихся диполей),

TS – энтропийная составляющая. Энтропия уменьшается при растворении, вследствие упорядочивания диполей при поляризации среды.

pV – малосущественная компонента, поскольку объем почти не меняется при растворении.

Полярная компонента энергии сольватации определяется формулой:

$$G_{\text{pol}} = \int \frac{\varepsilon E^2}{8\pi} dV. \quad (27)$$

Из того, что она не совпадает с U_{el} (26), уже очевидно, что она не может быть потенциальной электростатической энергией. С другой стороны, можно проиллюстрировать качественно, что G_{pol} обладает свойствами свободной энергии, т.е. при растворении

$$\Delta G_{\text{free}} = \Delta G_{\text{pol}}. \quad (28)$$

Действительно, для $T=0$ и жестких диполей при растворении из (25):

$$\Delta G_{\text{free}} = \Delta U_{\text{el}} = \int \frac{E_2^2}{8\pi} dV - \int \frac{E_1^2}{8\pi} dV \quad (29)$$

и из (27)

$$\Delta G_{\text{pol}} = \Delta U_{\text{el}} = \int \frac{\varepsilon E_2^2}{8\pi} dV - \int \frac{E_1^2}{8\pi} dV. \quad (30)$$

Если $\Delta G_{\text{free}} = \Delta G_{\text{pol}}$ верно, то это возможно только для $\varepsilon=1$.

Действительно, жесткие диполи при абсолютном нуле все полностью ориентированы в одном направлении и линейный отклик поляризации среды на поле отсутствует – отсюда $\varepsilon=1$, как и ожидалось из (29) и (30)

Для $T=0$ и диполей на «пружинках» при растворении из (25):

$$\Delta G_{\text{free}} = \Delta U_{\text{el}} + \Delta U_{\text{nonel}} = \int \frac{E_2^2}{8\pi} dV - \int \frac{E_1^2}{8\pi} dV + \Delta U_{\text{nonel}} \quad (31)$$

и из (27)

$$\Delta G_{\text{pol}} = \Delta U_{\text{el}} = \int \frac{\epsilon E_2^2}{8\pi} dV - \int \frac{E_1^2}{8\pi} dV. \quad (32)$$

Если $\Delta G_{\text{free}} = \Delta G_{\text{pol}}$ верно, то ожидается ненулевой линейный отклик поляризации среды $\epsilon > 1$ за счет ΔU_{nonel} . И действительно, ничто не мешает таким диполям растягиваться и при абсолютном нуле, давая линейный отклик поляризации среды на поле, т.е. $\epsilon > 1$, как и ожидалось из (31) и (32)

Опишем ситуацию для жестких диполей количественно. Рассмотрим изменение свободной энергии системы диполей, помещенных в полное электрическое поле (собственное + внешнее) напряженностью E . Предположим, что поле слабо меняется на величине диполя. Тогда можно записать для изменения свободной энергии ΔG :

$$\Delta G = G(E) - G(0) = RT \ln \int \exp(-\beta(\sum -d_i E(r_i) \cos \vartheta_i)) d\Gamma \quad (33)$$

В (33) интегрирование ведется по всему фазовому пространству. Здесь d_i модуль i -ого дипольного момента, $E(r_i)$ - напряженность полного поля в центре i -ого дипольного момента, $\cos \vartheta_i$ - угол между внешним полем и дипольным моментом. Суммирование в экспоненте ведется по всем диполям. Отметим, что в (33) не используется предположение об отсутствии взаимодействия наведенных дипольных моментов.

Возьмем вариацию от (33) по функции напряженности внешнего поля. Получим

$$\frac{\delta(\Delta G)}{\delta E} = \frac{\int \sum -d_i \cos \vartheta_i \exp(-\beta(\sum -d_i E(r_i) \cos \vartheta_i)) d\Gamma}{\int \exp(-\beta(\sum -d_i E(r_i) \cos \vartheta_i)) d\Gamma} \quad (34)$$

Видно, что (34) – это просто среднее значение проекции индуцированного дипольного момента на направление напряженности внешнего поля. Тогда получаем

$$\frac{\delta(\Delta G)}{\delta E} = p = (\epsilon - 1)\epsilon_0 E \quad (35)$$

Здесь мы использовали то, что индуцированный дипольный момент единицы объема p связан с диэлектрической восприимчивостью χ как

$$p = \chi \epsilon_0 E. \quad (36)$$

а диэлектрическая проницаемость с восприимчивостью как

$$\varepsilon = 1 + \chi. \quad (37)$$

Интегрируя уравнение (35), получим:

$$\Delta G = \frac{(\varepsilon - 1)\varepsilon_0 E^2}{2} + C. \quad (38)$$

Константа интегрирования C определяется из условия $\Delta G(\varepsilon = 1) = 0$, получаем $C=0$.
Окончательно,

$$\Delta G = \frac{(\varepsilon - 1)\varepsilon_0 E^2}{2}, \quad (39)$$

что и требовалось доказать.

Глава 1.2 Вторая группа проблем учета влияния внешнего шума/наблюдателя на основе наблюдаемой динамики макропараметров, рассчитываемой из идеальной динамики и независимой от амплитуды шума.

В данной главе мы описываем вторую группу проблем учета влияния внешнего шума от окружения/наблюдателя на основе наблюдаемой динамики макропараметров, рассчитываемой из идеальной динамики и независимой от амплитуды шума на примере теории горения и распространения фронта пламени.

Вторая группа проблем была рассмотрена на примерах из теории горения: детально проанализированы полюсные решения в теории горения смеси газов и в фильтрационном горении; проанализировано влияние малых шумов на распространение фронта горения и на его основании объяснено увеличение скорости фронта при увеличении ширины прямоугольного канала (ускорение фронта пламени в случае газовой смеси при радиальной и сферической геометрии); объяснена ячеистая структура фронта пламени; показаны естественные методы регуляризации полюсного решения в теории фильтрационного горения (позволяющие сохранить полученное точное аналитическое полюсное решение); объяснена асимптотика решения в виде одного пальца, равного половине ширины прямоугольного канала.

Приведем несколько конкретных примеров ошибок, сделанных из-за непонимания основ новой кибернетики (учета влияния внешнего шума от окружения/наблюдателя), например, в теории горения смеси газов и теории фильтрационного горения.

Сивашинский в своей работе (Rahibe M., Aubry N., Sivashinsky G.I., Lima R. // Phys. Rev. E - 1995. - V. 52, N 4 - PP. 3675–3686) утверждал, что идеальная динамика полюсов, описывающих горение в смеси газов, приводит к ускорению фронта пламени. Он полагал, что это ускорение не вызвано шумом, поскольку оно не меняется при уменьшении шума и зависит лишь от свойств самой системы. Но ведь также наблюдаемая динамика, связанная с шумом, не зависит от него в широком интервале значений.

Танвир в своей работе (Siegel M., Tanveer S. // Phys. Rev. Lett. –1996. - V.76. - P. 419-422) нашел различное поведение асимптотического решения в виде «пальца» для уравнений, описывающих фильтрационное горение. Он не понял, что эта разница исчезает при наличии малого (в том числе численного) шума, приводящего к новой наблюдаемой динамике.

Глава 1.2.1 Случайный шум и динамика полюсов при распространении неустойчивого фронта пламени

Мы исследуем проблему распространения пламени. В этой проблеме изучаются неустойчивые фронты, которые «морщатся» на многих масштабах. Используется аналитический метод движения полюсов в плоскости комплексной переменной для описания процесса неустойчивого роста при произвольных начальных состояниях и возмущениях. Мы аргументируем, что воздействие случайного шума является очень значительным и что его ни в коем случае нельзя игнорировать в достаточно больших системах. Мы представляем модели, которые приводят к законам подобия в отношении скорости и ускорения фронта как функции размера системы и уровня шума, и аналитические аргументы, которые объясняют эти результаты через динамику шума полюсов. Дается детальное описание избыточного количества полюсов в системе; описание числа полюсов, которые появляются в системе за единицу времени; а также описание времени существования полюса. Это позволяет нам понять зависимость параметров системы от шума [119].

1.2.1.1 Введение

В этой части рассматривается очень интересная проблема описания нелинейной фазы развития гидродинамической неустойчивости пламени. Эта проблема может быть рассмотрена для одномерного случая 1D (распространение в канале), двухмерного случая 2D (для цилиндра), трехмерного случая 3D (для сферы). Можно составить прямые числовые модели на основе уравнений Навье-Стокса, включая химическую кинетику в форме закона Аррениуса [120]. Для двухмерного и трехмерного случаев мы можем рассматривать экспериментально наблюдаемые эффекты [121,122] самоускорения фронта расходящегося пламени, образование ячеистой структуры и другие эффекты.

Можно получить гораздо более простое уравнение распространения фронта горения [123,124]. Это приближенная модель Михельсона-Сивашинского. Модель Михельсона-Сивашинского предполагает очень серьезные ограничения, такие, как малая величина коэффициента расширения газа и, соответственно, потенциальный поток в продуктах горения и непрореагировавшей еще смеси, слабая нелинейность; имеют также место предположения о стабилизирующем эффекте кривизны фронта горения и линейной зависимости от кривизны фронта горения и другие. В наших предыдущих работах [125-130] расчеты шумовой составляющей выполнялись в рамках как одномерной, так и двухмерной формулировок проблемы, а недавно Карлин и Сивашинский выполнили в [131,132] это же для случаев 1D, 2D и 3D.

Интерес к этой модели объясняется, во-первых, тем, что несмотря на серьезные ограничения, с помощью этой модели можно качественно описать сценарий гидродинамической неустойчивости и, в частности, самоускорение фронта расходящегося пламени, образование ячеистой структуры и другие эффекты. Во-вторых, эта нелинейная модель имеет точные решения, которые могут истолковываться на основе разложения на полюса [133-136]. Метод разложения на полюса получил развитие в последующих работах. Были исследованы соотношения между полюсным решением и разложением в ряд Фурье [137], между полем потока газа и полюсным решением [138]. Более подробное изложение можно найти в наших предыдущих работах [125-130] и содержащихся в них соответствующих ссылках.

Нужно отметить, что самый простой (одномерный) случай распространения фронта горения является очень важным. Это было главной причиной того, что при написании

этой работы мы детально рассматривали именно одномерный случай. Исследование этого случая дает нам качественное и количественное понимание динамики полюсов. Это понимание является базисом для рассмотрения более сложных двух- и трехмерных случаев. В нашей работе [127] это показано достаточно четко. Ячеистая структура и экспонента ускорения для двухмерного случая были найдены на основе результатов одномерного случая.

Дальнейшее развитие наших результатов дается в работах [139-141]. В очень интересной работе [139] шумовая составляющая рассматривалась в форме набора полюсов. Найденные численные результаты для одномерного случая хорошо соотносятся с теоретическими результатами нашей работы.

Может существовать четыре таких режима горения пламени. Режим II не может наблюдаться из-за числового шума. Чтобы объяснить присутствие режима III, мы выполним численные вычисления и попытаемся дать аналитическое объяснение таких величин, как избыточное число полюсов в системе, число полюсов, которое появляется в системе за единицу времени, время существования полюса [130]. Разработанная теория может объяснить малую зависимость параметров проблемы от шума в режиме III.

Остальная часть работы построена следующим образом. Мы начинаем с представления уравнений движения и разложения на полюса в геометрии канала (Раздел 2). В следующем Разделе 3 описывается ускорение фронта горения, динамика полюсов и шум. В заключительном разделе (Раздел 4) подводятся итоги и делаются выводы.

1.2.1.2 Уравнения движения и разложения на полюса в геометрии канала

Известно, что плоское пламя, свободно распространяющееся через изначально неподвижные однородные горючие смеси, является по своей природе нестабильным. Известно, что при таких видах горения развиваются характерные структуры, которые включают в себя точки излома, и что в нормальных экспериментальных условиях фронт горения ускоряется во времени. В работе [123] была предложена модель в измерениях $1+1$, которая имеет прямое отношение к распространению фронтов горения в каналах ширины \tilde{L} . Решение записывается через положение $h(x, t)$ фронта горения над осью x . После соответствующих масштабирований уравнение принимает следующую форму:

$$\frac{\partial h(x, t)}{\partial t} = \frac{1}{2} \left[\frac{\partial h(x, t)}{\partial x} \right]^2 + \nu \frac{\partial^2 h(x, t)}{\partial x^2} + I\{h(x, t)\} + 1. \quad (1)$$

Областью определения будет интервал $0 < x < \tilde{L}$, ν - некоторый параметр, и мы используем периодические граничные условия. Функционал $I[h(x, t)]$ - это преобразование Гилберта производной, которая удобно определяется через члены пространственного преобразования Фурье

$$h(x, t) = \int_{-\infty}^{\infty} e^{ikx} \hat{h}(k, t) dk \quad (2)$$

$$I[h(k, t)] = |k| \hat{h}(k, t) \quad (3)$$

Для введения разложения на полюса будет удобно перемасштабировать область определения до $0 < \theta < 2\pi$. Выполнив это перемасштабирование и обозначив получившиеся величины теми же обозначениями, мы получим

$$\frac{\partial h(x, t)}{\partial t} = \frac{1}{2} \left[\frac{\partial h(x, t)}{\partial \theta} \right]^2 + \frac{\nu}{L^2} \frac{\partial^2 h(x, t)}{\partial \theta^2} + \frac{1}{L} I\{h(\theta, t)\} + 1. \quad (4)$$

В этом уравнении $L = \tilde{L}/2\pi$. Далее мы меняем переменные на $u(\theta, t) \equiv \partial h(\theta, t)/\partial \theta$. Получаем

$$\frac{\partial u(x, t)}{\partial t} = \frac{u(x, t)}{L^2} \frac{\partial u(x, t)}{\partial \theta} + \frac{\nu}{L^2} \frac{\partial^2 u(x, t)}{\partial \theta^2} + \frac{1}{L} I\{u(\theta, t)\}. \quad (5)$$

Хорошо известно, что решение этого уравнения для движения плоского фронта линейно нестабильно. Линейный спектр в представлении k :

$$\omega_k = |k|/L - ik^2/L^2. \quad (6)$$

Имеется характерный масштаб k_{max} , на котором последняя нестабильная мода

$$k_{max} = L\nu. \quad (7)$$

Нелинейные эффекты стабилизируются в новом стабильном состоянии, которое обсуждается ниже.

Отличительная характеристика решений этого уравнения - появление изломообразных структур в развивающихся фронтах. Соответственно, представление через моды Фурье будет очень неэффективным. Однако, оно будет весьма эффективным при представлении решений как суммы функций с полюсами в плоскости комплексных переменных. Ниже мы покажем, что положение изломов на фронте определяется действительной координатой полюса, в то время как высота излома определяется мнимой координатой. Более того, можно увидеть, что динамика развивающегося фронта можно эффективно описать через динамику полюсов. На основании [126, 133, 134, 136] мы разлагаем решения $u(\theta, t)$ на функции, которые зависят от N полюсов, положение которых $z_j(t) \equiv x_j(t) + iy_j(t)$ в плоскости комплексной переменной является зависящим от времени:

$$u(\theta, t) = \nu \sum_{j=1}^N \cot[\theta - z_j(t)] + c.c. = \nu \sum_{j=1}^N 2 \sin[\theta - x_j(t)] \cosh[y_j(t)] - \cos[\theta - x_j(t)], \quad (8)$$

$$h(\theta, t) = 2\nu \sum_{j=1}^N \ln[\cosh(y_j(t)) - \cos(\theta - x_j(t))] + C(t). \quad (9)$$

В (9) $C(t)$ является функцией времени. Функция (9) – это суперпозиция квази-изломов (т.е. изломов, закругленных на вершине). Действительная часть положения полюса (т.е. x_j) является координатой максимума квази-излома (на интервале $[0, 2\pi]$), а мнимая часть положения полюса (т.е. y_j) характеризует глубину квази-излома. По мере убывания y_j глубина излома возрастает. Так как $y_j \rightarrow 0$, глубина расходится к бесконечности. И наоборот, когда $y_j \rightarrow \infty$, глубина уменьшается до нуля.

Главное преимущество этого представления состоит в том, что распространение и сминание фронта может описываться через динамику полюсов. Подставляя (8) в (5), мы получаем следующие простые дифференциальные уравнения положения полюсов:

$$-L^2 dz_j/dt = [\nu \sum_{k=1, k \neq j}^{2N} \cot(z_j - z_k/2) + iL2\text{sign}[\text{Im}(z_j)]] \quad (10)$$

Отметим, что в (8), вследствие комплексного сопряжения, у нас имеется $2N$ полюсов, которые мы расположили парами так, что для $j < N$ $z_{j+N} = \bar{z}_j$. Во второй сумме в (8) каждая пара полюсов составляет один член. В ур.(10) мы снова используем $2N$ полюсов, так как все из них взаимодействуют друг с другом. Мы можем записать динамику полюсов через реальную и воображаемую части x_j и y_j . В силу упорядочивания парами, достаточно записать уравнение либо $y_j > 0$, либо для $y_j < 0$. Выберем первый случай. Уравнения для положения полюсов будут следующими:

$$-L^2 \frac{dx_j}{dt} = \nu \sum_{k=1, k \neq j}^N \sin(x_j - x_k) [[\cosh(y_j - y_k) - \cos(x_j - x_k)]^{-1} + [\cosh(y_j + y_k) - \cos(x_j - x_k)]^{-1}] \quad (11)$$

$$L^2 \frac{dy_j}{dt} = \nu \sum_{k=1, k \neq j}^N (\sinh(y_j - y_k) \cosh(y_j - y_k) - \cos(x_j - x_k) + \sinh(y_j + y_k) \cosh(y_j + y_k) - \cos(x_j - x_k)) + \nu \coth(y_j) - L \quad (12)$$

Отметим, что если начальное состояние дифференциального уравнения (5) разлагается на конечное число полюсов, уравнения движения сохраняют это число как функцию времени. С другой стороны, это может быть нестабильной ситуацией для дифференциального уравнения в частных производных, и шум может изменить число полюсов. Этот вопрос исследуется далее в разделе 3. Сейчас вернемся к обсуждению решения уравнений динамики полюсов для стабильного состояния.

1.2.1.2.1 Качественные свойства стационарного решения

Решение стабильного состояния для распространения фронта горения в каналах ширины 2π было представлено в работе [134]. Используя эти результаты, мы можем сразу же перевести обсуждение для канала ширины L . Основные результаты можно подытожить в следующем виде:

1. Имеется только одно стабильное стационарное решение, которое геометрически представлено гигантским изломом (либо эквивалентно одним «пальцем») и аналитически $N(L)$ полюсами, которые выровнены на одной линии, параллельной мнимой оси. Наличие этого решения станет яснее после следующих примечаний.
2. Имеется притяжение между полюсами вдоль действительной оси. Это очевидно из ур.(11), в котором знак dx_j/dt всегда определяется $\sin(x_j - x_k)$. Получающаяся в результате динамика приводит к слиянию всех полюсов вдоль оси x , положения же вдоль оси y которых остаются конечными и различными.
3. Положения y являются различными, и положения полюсов выравниваются вдоль одной линии один выше другого $y_{j-1} < y_j < y_{j+1}$ с максимумом, равным $y_{N(L)}$. Это можно понять из ур.(12), в котором взаимодействие, как видно, является отталкиванием на коротких диапазонах, но изменяет знак на более длинных диапазонах.

4. Если к такому решению добавляется дополнительный полюс, этот (или иной) полюс будет выталкиваться в сторону бесконечности вдоль мнимой оси. Если система имеет меньше чем $N(L)$ полюсов, она нестабильна для добавления полюсов, и любой шум будет подводить систему к этому единственному состоянию. Число $N(L)$ будет
- 5.

$$N(L) = \left[\frac{1}{2} \left(\frac{L}{\nu} + 1 \right) \right], \quad (13)$$

где [...] - целая часть. Чтобы увидеть это, рассмотрим систему с N полюсами и такую, чтобы все величины y_j удовлетворяли условию $0 < y_j < y_{\max}$. Теперь добавим один дополнительный полюс, координаты которого $z_a \equiv (x_a, y_a)$ с $y_a \gg y_{\max}$. Из уравнения движения для y_a (12) можно увидеть, что все слагаемые суммы имеют порядок единицы, так как описываются членом уравнения $\cot(y_a)$. Соответственно, уравнение движения y_a будет приблизительно следующим:

$$\frac{dy_a}{dt} \approx \nu \frac{2N+1}{L^2} - \frac{1}{L}. \quad (14)$$

Поведение этого полюса зависит от числа других полюсов. Если N слишком большое, полюс стремится к бесконечности, в то время если N небольшое, полюс будет стремиться к действительной оси. Условием движения к бесконечности будет $N > N(L)$, где $N(L)$ определено в (13). С другой стороны, координата y полюсов не может «попасть» в ноль. Ноль является линией отталкивания, а полюса отталкиваются от нуля с бесконечной скоростью. Чтобы увидеть это, рассмотрим полюс, y_j которого приближается к нулю. Для любого конечного L член $\coth(y_j)$ безгранично возрастает, в то время как все другие члены уравнения (12) остаются ограниченными.

6. Высота излома пропорциональна L . Распределение положений полюсов вдоль линии с постоянным x было подробно рассмотрено в [134].

Обратимся к решению со всеми этими свойствами в виде излома Чуала-Фриша-Хенона [Thual-Frisch-Henon (TFH)].

1.2.1.2.2 Нелинейная стабильность

Можно прибегнуть к интуиции для обсуждения вопроса стабильности к *большим* возмущениям. Другими словами, мы можем захотеть добавить в систему полюса с конечными величинами y и узнать, что с ними будет. Сначала в этом подразделе мы покажем, что полюса, начальная величина y которых меньше $y_{\max} \sim \log(L^2/\nu^2)$, будут притягиваться к действительной оси. Сценарий будет аналогичным таковому, описанному в последнем параграфе.

Предположим, что мы генерируем стабильную систему с гигантским изломом при $\theta_c = 0$ с полюсами, распределяемыми вдоль оси y до y_{\max} . Мы знаем, что сумма всех сил, которые действуют на верхний полюс, является нулевой. Далее рассмотрим дополнительный полюс, введенный в положение (π, y_{\max}) . Из уравнения (12) очевидно, что силы, действующие на этот полюс, будут тянуть его вниз. С другой стороны, если его начальное положение существенно выше y_{\max} , сила, действующая на него будет

отталкиванием в направлении к бесконечности. Мы видим, что этот простой аргумент определяет y_{\max} как типовой масштаб для нелинейной стабильности.

Далее мы оцениваем y_{\max} и даем толкование наших результатов через *амплитуду* возмущения фронта горения. Мы показываем, что положение верхнего полюса колеблется между минимальной величиной и бесконечностью по мере изменения L . Мы хотим оценить характерный масштаб минимальной величины $y_{\max}(L)$. С этой целью мы используем результат работы [134], касающийся стабильного распределения положений полюсов в стабильной большой системе. Параметрирование в [134] отличается от нашего; чтобы перейти от нашего параметрирования в ур.(5) к их параметрированию, нужно перемасштабировать u на L^{-1} и t на L . Параметр ν в их параметрировании будет соответствовать ν/L в нашем. В соответствии с [134], число полюсов между y и $y+dy$ определяется $\rho(y)dy$, где плотность $\rho(y)$ равна

$$\rho(y) = L\pi^2\nu \ln[\coth(|y|/4)]. \quad (15)$$

Для оценки минимальной величины y_{\max} нам нужно, чтобы «хвост» распределения $\rho(y)$, интегрированный между этой величиной и бесконечностью, допускал только один полюс. Другими словами,

$$\int_{y_{\max}}^{\infty} dy\rho(y) \approx 1. \quad (16)$$

Раскладывая (15) для больших y и напрямую интегрируя напрямую результат в (16), получим в конечном итоге

$$y_{\max} \approx 2 \ln[4L\pi^2\nu] \quad (17)$$

Для большой величины L этот результат составляет $y_{\max} \approx \ln(L^2\nu^2)$. Если мы сейчас добавим еще один полюс в положение (θ, y_{\max}) , это эквивалентно возмущению решения $u(\theta, t)$ с функцией $e^{-y_{\max}} \sin(\theta)$, как это можно увидеть непосредственно из (8). Соответственно, мы делаем вывод, что система нестабильна к возмущениям *большим*, чем

$$u(\theta) \sim \nu^3 \sin(\theta)/L^2. \quad (18)$$

Это указывает на очень сильную зависимость чувствительности к внешним возмущениям от размера системы для решения в виде гигантского излома. Это будет важным фактором и в нашем обсуждении систем с шумом.

1.2.1.3 Ускорение фронта горения, динамики полюсов и шума

Главной мотивацией этого раздела является наблюдение, что в радиальной геометрии те же уравнения движения приводят к ускорению фронта горения. Цель этого раздела – аргументировать, что данное явление вызвано генерированием шума, связанного с появлением новых полюсов. Более того, мы намерены показать, что в радиальной геометрии можно многое узнать об ускорении путем рассмотрения влияния шума на распространение пламени в канале. В работе [134] было показано, что любое

начальное состояние, представленное в полюсах, приходит к единственному стационарному состоянию. Этим стационарным состоянием является гигантский излом, распространяющийся с постоянной скоростью $v = 1/2$ с точностью до малых коррекций порядка $1/L$. В свете нашего обсуждения в последнем разделе мы ожидаем, что любое достаточно равномерное начальное состояние приведет к тому же стационарному состоянию. Соответственно, если в динамике канала конечной ширины нет шума, то ускорение фронта горения невозможно. Что случается, если добавить шум в систему?

Чтобы быть конкретными, введем дополнительную составляющую гладкого шума $\eta(\theta, t)$ в уравнение движение (5), где

$$\eta(\theta, t) = \sum_k \eta_k(t) \exp(ik\theta), \quad (19)$$

а амплитуды Фурье η_k коррелируют в соответствии с

$$\langle \eta_k(t) \eta_{k'}^*(t') \rangle = \frac{f}{L} \delta_{k,k'} \delta(t-t'). \quad (20)$$

Теперь рассмотрим результат численного моделирования динамики, возбуждаемой шумом, а затем вернемся к теоретическому анализу.

1.2.1.3.1 Моделирование с шумом

В предшествующих численных исследованиях [124, 142] не был введен шум в явной управляемой форме. Позже мы аргументируем, что некоторые из явлений, которые встречаются при этих моделированиях, могут быть приписаны (нерегулируемому) числовому шуму. Мы выполнили численное моделирование в ур.(5), используя псевдоспектральный метод. Шаг по времени был выбран по схеме Adams-Bashforth со вторым порядком точности по времени. Аддитивный белый шум был сгенерирован в пространстве Фурье путем выбора η_k для каждого k из равномерного распределения в интервале $[-\sqrt{2\frac{f}{L}}, \sqrt{2\frac{f}{L}}]$. Мы рассмотрели среднюю скорость фронта в стабильном состоянии как функции L для постоянной f и как функция f для постоянной L . Нами были обнаружены следующие интересные явления, которые подытоживаются здесь ниже:

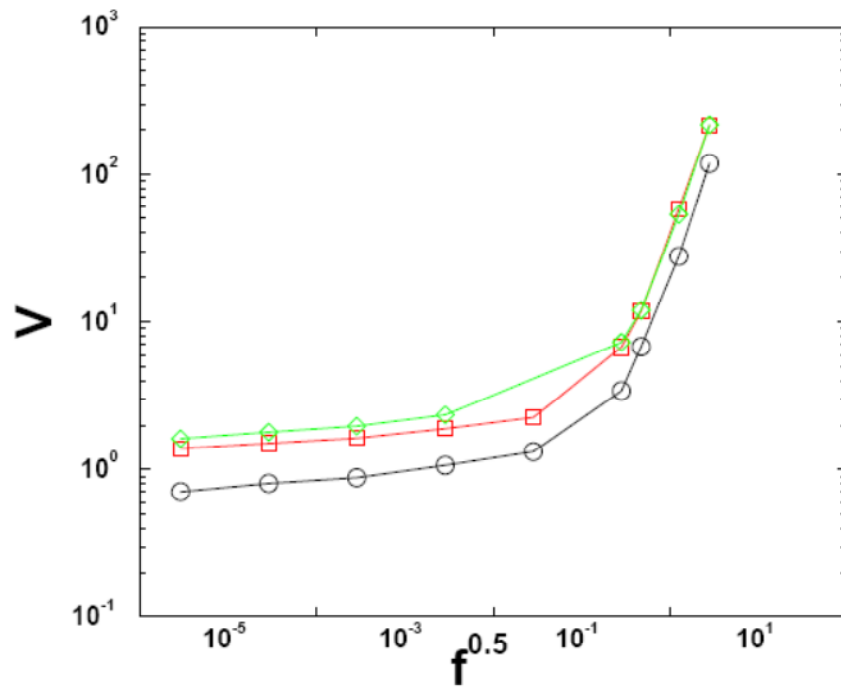


Рис.1: Зависимость средней скорости v от шума $f^{0.5}$ для $L=10, 40, 80$.

1. На рис. 1 мы видим два разных режима поведения средней скорости v как функции шума $f^{0.5}$ для фиксированных систем размера L . Для шума f меньшего, чем некоторая постоянная величина f_{cr}

$$v \sim f^{\xi} . \tag{21}$$

Для этих величин f данная зависимость очень слабая, и $\xi \approx 0.02$. Для больших величин f зависимость от шума будет гораздо сильнее.

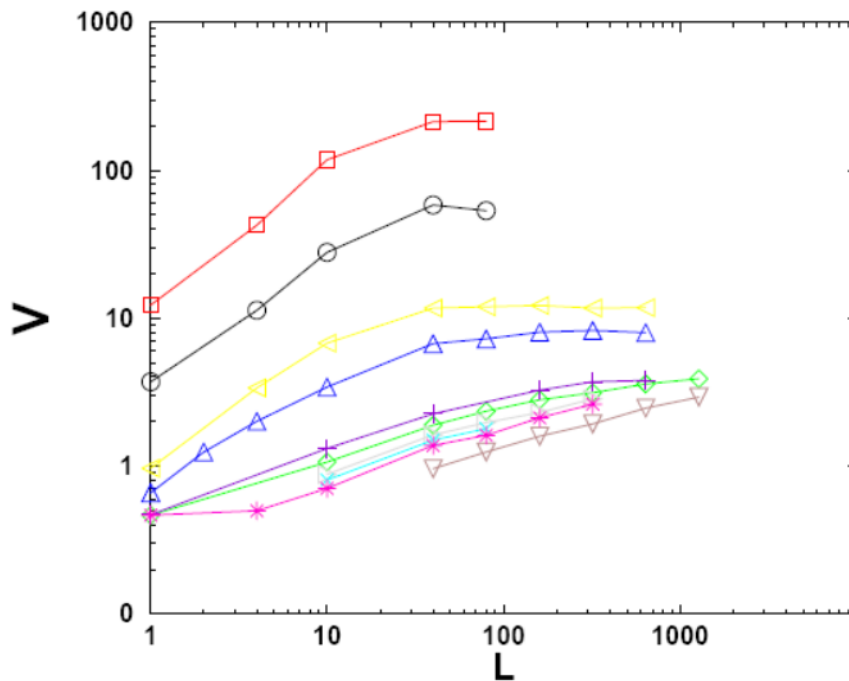


Рис.2: Зависимость средней скорости v от размера системы L для $f^{0.5} = 0, 2.7 \cdot 10^{-6}, 2.7 \cdot 10^{-5}, 2.7 \cdot 10^{-4}, 2.7 \cdot 10^{-3}, 2.7 \cdot 10^{-2}, 2.7 \cdot 10^{-1}, 0.5, 1.3, 2.7$.

На рис. 2 мы видим рост средней скорости v как функции размера системы L . После определенных величин L мы видим насыщение скорости. Для режима $f < f_{cr}$ рост скорости может быть изложен как

$$v \sim L^\mu, \quad \mu \approx 0.40 \pm 0.05. \quad (22)$$

1.2.1.3.2 Вычисление количества полюсов в системе

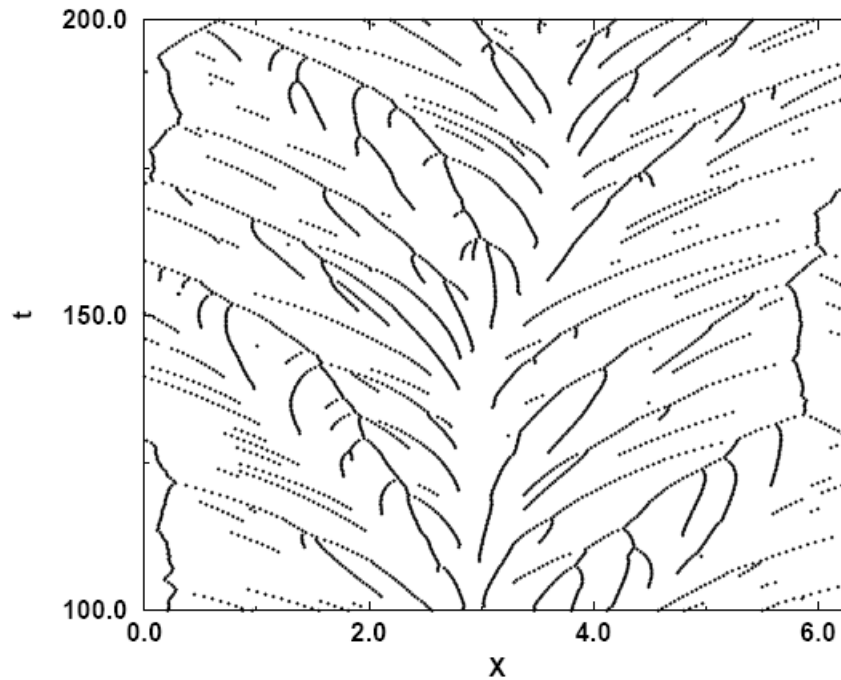


Рис.3 Зависимость положений изломов от времени. $L = 80$ $\nu = 0.1$ $f = 9 \cdot 10^{-6}$

Здесь мы собираемся заняться решением интересной проблемы – найти количество полюсов, которое существует в нашей системе за пределами гигантского излома. Мы можем сделать это следующим образом: вычислить количество изломов (точек минимума или точек изгиба) и их положения в интервале $\theta: [0, 2\pi]$ в каждый момент времени и изобразить положения изломов как функцию времени (см. рис.3).

Предположим, что наша система почти все время находится в «квазистабильном» состоянии, т.е. каждый новый излом, который появляется в системе, включает только один полюс. Мы находим интересные явления, которые мы опишем ниже:

1. Мы можем найти среднее число полюсов, которые существуют в нашей системе за пределами гигантского излома. (Для этого мы вычисляем количества изломов в определенный момент времени и затем исследуем историю каждого излома (кроме гигантского излома). Т.е. исследуем сколько начальных изломов принимают участие в образовании этого излома, а затем после усредняем по отношению к различным моментам времени.) Давайте обозначим это среднее число полюсов как ∂N . Мы можем выделить четыре режима, которые определяются в отношении зависимости этого числа от шума f :
 - (i) Режим I: Такой малый шум, что в нашей системе не существует полюсов за пределами гигантского излома.
 - (ii) Режим II: Сильная зависимость числа полюсов ∂N от шума f .
 - (iii) Режим III: Насыщение числа полюсов ∂N от шума f , т.е. мы наблюдаем очень слабую зависимость этого количества от шума.

$$\partial N \sim f^{0.03} \tag{23}$$

Насыщенная величина δN определяется следующей формулой

$$\delta N \approx N(L)/2 \approx \frac{1}{4} \frac{L}{\nu} \quad (24)$$

где $N(L) \approx \frac{1}{2} \frac{L}{\nu}$ - число полюсов в гигантском изломе.

- (iv) Режим IV: Мы снова наблюдаем сильную зависимость числа полюсов δN от шума f .

$$\delta N \sim f^{0.1} \quad (25)$$

Из-за числового шума мы наблюдаем в большинстве моделей только режимы III и IV. В дальнейшем, если мы не отметим что-либо отличное, мы будем обсуждать режим III.

2. Путем вычисления нового количества изломов мы можем найти число полюсов, которое появляется в системе за единицу времени $\frac{dN}{dt}$. В режиме III

$$\frac{dN}{dt} \sim f^{0.03} \quad (26)$$

Зависимость от L and ν определяется как

$$\frac{dN}{dt} \sim L^{0.8} \quad (27)$$

$$\frac{dN}{dt} \sim \frac{1}{\nu^2} \quad (28)$$

А в режиме IV -

$$\frac{dN}{dt} \sim f^{0.1} \quad (29)$$

1.2.1.3.3 Теоретическое обсуждение воздействия шума

1.2.1.3.3.1 Порог неустойчивости к дополнительному шуму. Переход из режима I в режим II

Сначала мы представим теоретические аргументы, которые объясняют чувствительность решения гигантского излома к воздействию дополнительного шума. Эта чувствительность очень сильно возрастает при увеличении размера системы L . Для рассмотрения этого мы снова прибегнем к соотношению между анализом линейной стабильности и динамикой полюсов.

Аддитивный шум вводит возмущения по всеми k -векторами. Ранее мы показали, что самый нестабильный режим – это компонента $A_1 \sin(\theta)$ с $k=1$. Соответственно, наиболее эффективное шумовое возмущение - это $\eta_1 \sin(\theta)$, которое потенциально может привести к росту наиболее нестабильного режима. Будет или нет этот режим расти, зависит от амплитуды шума. Чтобы рассмотреть это более четко, вернемся к описанию в виде полюсов. Для небольших величин амплитуды A_1 мы представим $A_1 \sin(\theta)$ как решение в виде единственного полюса с функциональной формой $\nu e^{-y} \sin \theta$. Положение y определяется из $y = -\log |A_1| / \nu$, а положение θ таково: $\theta = \pi$ для положительного A_1 и $\theta = 0$ для отрицательного A_1 . Для очень малой A_1 полюса будет двигаться к бесконечности, независимо от его положения θ ; динамика симметрична при замене $A_1 \rightarrow -A_1$, когда y достаточно велико. С другой стороны, когда величина A_1 возрастает, симметрия нарушается, и положение θ и знак A_1 становятся очень важными. Если $A_1 > 0$, то имеет место пороговая величина y , ниже которой полюс устремляется вниз. С другой стороны, если $A_1 < 0$, а $\theta = 0$, отталкивание от полюсов гигантского излома возрастает в уменьшением y . Тем самым мы подразумеваем, что с качественной точки зрения динамика A_1 характеризуется асимметричным «потенциалом» в соответствии с

$$\dot{A}_1 = -\frac{\partial V(A_1)}{\partial A_1}, \quad (30)$$

$$V(A_1) = \lambda A_1^2 - a A_1^3 + \dots \quad (31)$$

В результате линейного анализа стабильности нам известно, что $\lambda \approx \nu/L^2$, ср. с ур.(14). Далее, мы знаем, что порог для нелинейной стабильности имеет место при $A_1 \approx \nu^3/L^2$, ср. с ур. (18). Это определяет, что величина коэффициента $a \approx 2/3\nu^2$. Величина «потенциала» в максимуме будет

$$V(A_{\max}) \approx \nu^7/L^6. \quad (32)$$

Воздействие шума на развитие режима $A_1 \sin \theta$ можно понять из следующего стохастического уравнения

$$\dot{A}_1 = -\frac{\partial V(A_1)}{\partial A_1} + \eta_1(t). \quad (33)$$

Хорошо известно [143], что для такой динамики скорость R проникновения через «потенциальный» барьер для небольшого шума пропорциональна следующему:

$$R \sim \nu L^2 \exp^{-\nu^7/L^6}. \quad (34)$$

Делаем вывод, что любой, даже самый малый шум, становится действенным, когда увеличивается размер системы и когда уменьшается ν . Если мы возбуждаем систему с помощью шума амплитуды $\frac{f}{L}$, система всегда может быть чувствительной к шуму, когда

его размер превышает критическую величину L_c , которая определяется по формуле $\frac{f}{L_c} \sim \nu^7/L_c^6$. Эта формула определяет переход из режима I (новых полюсов нет) в режим

II. Для $L > L_c$ шум будет вводить в систему новые полюса. Для системы больших размеров даже численный шум, возникающий при моделировании, может иметь макроскопическое влияние.

Появление новых полюсов должно увеличивать скорость фронта. Скорость пропорциональна среднему $(u/L)^2$. Новые полюса приведут к деформированию гигантского излома дополнительными более мелкими изломами на крыльях гигантского излома, увеличивая u^2 . После увеличения амплитуды шума на фронте появляется все больше и больше мелких изломов, и скорость неизбежно возрастает. Это явление обсуждается и количественно в разделе 3.

1.2.1.3.3.2 Проверка асимметричной формы «потенциала»

Из уравнений движения для полюсов мы можем найти распределение полюсов в гигантском изломе [134]. Если нам известно распределение полюсов в гигантском изломе, мы можем затем найти форму «потенциала» и выполнять численную проверку выражений для величин λ , A_{\max} и $\frac{\partial V(A_1)}{\partial A_1}$, которые обсуждались ранее. Связь между амплитудой A_1

и положением полюса y определяется $A_1 = 4\nu e^{-y}$, а связь между функцией потенциала $\frac{\partial V(A_1)}{\partial A_1}$ и положением полюса y определяется формулой $\frac{\partial V(A_1)}{\partial A_1} = 4\nu \frac{dy}{dt} e^{-y}$, где $\frac{dy}{dt}$ может быть определено из уравнения движения полюсов. Мы можем определить A_{\max} как значение в нуле функции $\frac{\partial V(A_1)}{\partial A_1}$, а λ может быть определена как $\frac{1}{2} \frac{\partial^2 V(A_1)}{\partial A_1^2}$ для $A_1 = 0$.

Для набора величин $L = 2n\nu$ были выполнены численные измерения, где n является целым числом и $n > 2$. Для наших численных измерений мы используем постоянную $\nu = 0.005$ и переменную L , где L изменяется в интервале [1,150]; или переменную ν , которая изменяется в интервале [0.005,0.05], и константу $L = 1$. Получены следующие результаты:

1. Формула для $\frac{A_{\max} L^2}{\nu^3}$

$$\frac{A_{\max} L^2}{\nu^3} \approx 6.5. \quad (35)$$

2. Формула для $\frac{A_{\max}}{A_{N(L)}}$

$$\frac{A_{\max}}{A_{N(L)}} \approx 0.465. \quad (36)$$

где $A_{N(L)}$ определяется положением верхнего полюса.

3. Формула для $\frac{\lambda L^2}{\nu}$

$$\frac{\lambda L^2}{\nu} = 0.5. \quad (37)$$

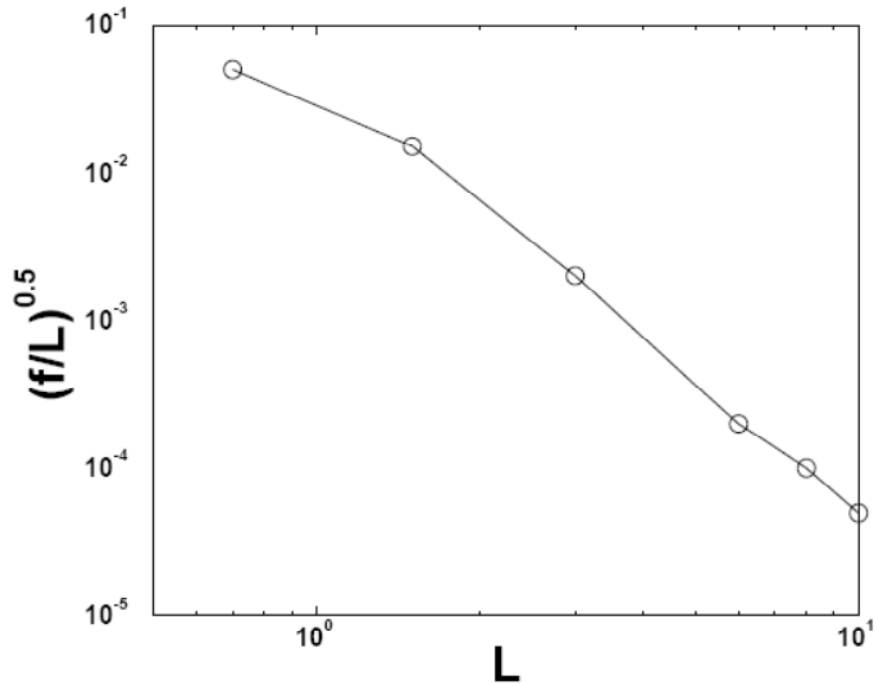


Рис.4: Первая нечетная собственная функция, полученная путем традиционного анализа стабильности.

4. Мы также выполняем проверку границы между режимом I (новых изломов нет) и режимом II (появляются новые изломы). На рис. 4 показана зависимость f/L_c от L_c . Можно видеть, что $f/L_c \sim 1/L_c^6$. Эти результаты хорошо соотносятся с теорией.

1.2.1.3.3.3 Стабильное состояние при наличии шума и его разрушение с увеличением шума и размера системы

В этом подразделе мы обсуждаем реакцию решения в виде гигантского излома на изменение шума, который способен ввести большое число избыточных полюсов в дополнение к тем, что уже имелись в гигантском изломе. Обозначим избыточное число полюсов как δN . Первый вопрос, который мы задаем себе – насколько трудно ввести сейчас дополнительный полюс, когда уже имеется заданное избыточное значение δN . С этой целью мы оцениваем эффективный потенциал $V_{\delta N}(A_1)$, который аналогичен (31), но при этом принимается во внимание наличие избыточного числа полюсов. Основное приближение, которое мы здесь используем – это то, что фундаментальная форма решения в виде гигантского излома не испытывает серьезного видоизменения из-за наличия избыточного числа полюсов. Конечно, это приближение количественно нарушается уже при одном избыточном полюсе. Однако, что касается качественной

стороны, то оно хорошо сохраняется до тех пор, пока избыточное число полюсов имеет порядок изначального числа $N(L)$ полюсов для решения в виде гигантского излома. Другое приближение состоит в том, что остальные линейные моды не играют никакой роли в этом случае. Соответственно, в этой точке мы ограничим обсуждение ситуацией $\delta N \ll N(L)$ (режим II).

Для оценки параметра λ в эффективном потенциале мы рассматриваем динамику полюса, положение y которого (y_a) гораздо больше y_{\max} . В соответствии с ур.(14), динамика будет следующей

$$\frac{dy_a}{dt} \approx \frac{2\nu(N(L) + \delta N)}{L^2} - \frac{1}{L} \quad (38)$$

Так как член $N(L)$ сокращается вместе с членом L^{-1} (ср. раздел 2.1), у нас остается член, описывающий отталкивание, который в эффективном потенциале переводится в форму

$$\lambda = \frac{\nu \delta N}{L^2}. \quad (39)$$

Затем мы оцениваем величину потенциала в точке перехода между притяжением и отталкиванием. В последней подстановке мы видели, что чужеродный полюс должен быть вставлен ниже y_{\max} , чтобы он притягивался к действительной оси. Сейчас нам нужно вставить новый полюс ниже положения имеющегося полюса, индекс которого $N(L) - \delta N$. Это положение оценивается как в разделе 2.2 путем применения функции распределения TFH (15). Получаем

$$y_{\delta N} \approx 2 \ln \left[\frac{4L}{\pi^2 \nu \delta N} \right]. \quad (40)$$

Как и раньше, здесь предполагается пороговая величина для амплитуды решения с единичным полюсом $A_{\max} \sin \theta$, которая получается из уравнивания $A_{\max} = \nu e^{-y_{\delta N}}$. Соответственно, в этом случае мы имеем $A_{\max} \sim \nu^3 (\delta N)^2 / L^2$. Снова используя кубическое представление для эффективного потенциала, мы получаем $a = 2/(3\nu^2 \delta N)$ и

$$V(A_{\max}) = \frac{1}{3} \frac{\nu^7 (\delta N)^5}{L^6}. \quad (41)$$

Повторив вычисления скорости перехода через барьер потенциала, мы получаем в этом случае

$$R \sim \nu \delta N L^2 \exp^{-\nu^7 (\delta N)^5 / L^6}. \quad (42)$$

Для заданной амплитуды шума $\frac{f}{L}$ всегда имеется величина L и ν , для которой скоростью проникновения R будет порядка $O(1)$ до тех пор, пока величина δN не слишком велика. Когда δN возрастает, скорость перехода падает, и в конце концов уже ни одного дополнительного полюса не может проникнуть в систему. Типовое число δN

для постоянных величин параметров оценивается с помощью приравнивания аргумента в экспоненте к единице:

$$\delta N \approx \left(\frac{f}{L} L^6 / \nu^7 \right)^{1/5}. \quad (43)$$

Как видим, δN зависит от шума f весьма существенно. В режиме III такого не наблюдается. Давайте отыщем условия перехода из режима II в режим III, где мы наблюдаем насыщение δN относительно шума f .

(i) Мы используем для амплитуды полюсного решения, которая на самом деле равна $\frac{2\nu \sin \theta}{\cosh(y_{\delta N}) - \cos \theta}$, выражение $A_{\max} = 4\nu e^{-y_{\delta N}}$, но это верно только для большого

числа $y_{\delta N}$. Для $y_{\delta N} < 1$ лучшим приближением будет $A_{\max} = \frac{4\nu}{y_{\delta N}^2}$. Из уравнения

(40) мы можем вывести, что граничная величина $y_{\delta N} = 1$ соответствует $\delta N \approx N(L)/2$

(ii) Мы используем выражение $y_{\delta N} \approx 2 \ln \left[\frac{4L}{\pi^2 \nu \delta N} \right]$, но для большой величины δN можно таким же образом найти лучшее приближение $y_{\delta N} \approx \frac{\pi^2 \nu}{2L} (N(L) - \delta N) \ln \left[\frac{8eL}{\pi^2 \nu (N(L) - \delta N)} \right]$. Эти выражения дают примерно тот же результат для $\delta N \approx N(L)/2$.

Из (i) и (ii) мы можем сделать следующие выводы.

(a) переход из режима II в режим III имеет место для состояния примерно $\delta N \approx N(L)/2$

(b) используя новое выражение в (i) и (ii) для амплитуды A_{\max} и $y_{\delta N}$, мы можем

вывести выражение для шума $\frac{f}{L}$ в режиме III

$$\frac{f}{L} \sim V(A_{\max}) \sim \lambda A_{\max}^2 \sim \frac{\nu \delta N}{L^2} \left(\frac{4\nu}{y_{\delta N}^2} \right)^2 \sim \frac{L^2}{\nu} \frac{\delta N}{(N(L) - \delta N)^4} \quad (44)$$

Это выражение определяет очень медленную зависимость δN от шума

$\frac{f}{L}$ для $\delta N > N(L)/2$, которая объясняет шумовое насыщение δN для

режима III.

(c) Форма решения в виде гигантского излома определяется полюсами, которые близки к нулю вдоль оси y . Для режима III $N(L)/2$ полюса, имеющие положение $y < y_{\delta N=N(L)/2} = 1$, остаются на своих местах. Такой результат объясняет, почему решение в виде гигантского излома не может быть серьезно изменено для режима III.

Из ур.(43) с помощью граничного условия

$$\delta N \approx N(L)/2 \quad (45)$$

можно получить пограничный шум f_b между режимами II и III

$$f_b \sim \nu^2 \quad (46)$$

Основное уравнение, описывающее динамику полюсов, будет следующим

$$\frac{dN}{dt} = \frac{\partial N}{T} \quad (47)$$

где $\frac{dN}{dt}$ - это число полюсов, которое появляется за единицу времен в нашей системе, ∂N - это избыточное число полюсов, T – среднее время существования полюса (между появлением и слиянием с гиганстким изломом). Используя результат численного моделирования для $\frac{dN}{dt}$ и (45), мы можем вывести для T

$$T = \frac{\partial N}{\frac{dN}{dt}} \sim \nu L^{0.2} \quad (48)$$

Таким образом, время существования пропорционально ν и очень слабо зависит от размера системы L .

Кроме того, время существования полюса определяется временем существования полюсов, находящихся в том же изломе. Из точки максимума линейной части ур. (1), мы можем найти среднюю характерную длину

$$\lambda_m \sim \nu \quad (49)$$

которая определяет размер наших изломов. Этот результат был подтвержден численными вычислениями, выполненными в [144]. В самом деле, это видно из рис. 9 в [144]. Среднее число полюсов в изломе

$$n_{\text{big}} \approx \frac{\lambda_m}{2\nu} \sim \text{const} \quad (50)$$

не зависит от L и ν . Среднее число изломов будет

$$N_{\text{big}} \sim \frac{\partial N}{n_{\text{big}}} \sim \frac{L}{\nu}. \quad (51)$$

Предположим, что в главном минимуме системы существует некий излом. Время существования полюса в таком изломе определяется тремя частями.

(I) Время формирования излома. Это время пропорционально размеру излома (с \ln -коррекциями) и числу полюсов в изломе (из уравнений движения полюсов)

$$T_1 \sim \lambda_m n_{\text{big}} \sim \nu \quad (52)$$

(II) Время, которое излом находится в окрестности минимума. Это время определяется как

$$T_2 \sim \frac{a}{v} \quad (53)$$

где a - окрестность минимума, при которой сила со стороны гигантского излома будет меньше, чем сила от флуктуаций избыточного числа полюсов δN , а v - скорость полюса в этой окрестности. Флуктуации избыточного числа полюсов δN выражаются как

$$N_{\Pi} = \sqrt{\delta N}. \quad (54)$$

Из этого результата и уравнений движения полюсов выводим, что L

$$v \sim \frac{v}{L} N_{\Pi} \sim \frac{v}{L} \sqrt{\frac{L}{v}} \sim \sqrt{\frac{v}{L}}. \quad (55)$$

Скорость, создаваемая от гигантского излома определяется с помощью

$$v \sim \frac{v}{L} N(L) \frac{a}{L} \sim \frac{a}{L}. \quad (56)$$

Соответственно, уравнивая эти два уравнения, мы получаем

$$a \sim \sqrt{vL}. \quad (57)$$

Соответственно, для T_2 получаем

$$T_2 \sim \frac{a}{v} \sim L. \quad (58)$$

(III) Время притяжения к гигантскому излому. Из уравнений движения для полюсов получаем

(IV)

$$T_3 \sim L \ln\left(\frac{L}{a}\right) \sim L \ln \sqrt{L} \sim L. \quad (59)$$

Исследованный домен размера системы был получен как

$$T_1 \gg T_2, T_3 \quad (60)$$

Поэтому полным временем существования будет

$$T = T_1 + T_2 + T_3 \sim v + sL, \quad (61)$$

где s - константа и

$$0 < s \ll 1. \quad (62)$$

Этот результат качественно и частично количественно объясняет зависимость (48). Из (48), (47), (45) мы видим, что в режиме III $\frac{dN}{dt}$ насыщается с размером системы L .

1.2.1.3.4 Ускорение фронта горения из-за шума

В этом разделе мы оценим экспоненты масштабирования, которые характеризуют скорость фронта горения как функцию размера системы. Наши аргументы в этом разделе будут менее солидные, чем предыдущие, но тем не менее мы считаем, что нам удалось достичь определенного качественного физического понимания, которое определяет взаимодействие между шумом и нестабильностью и которая дает в результате ускорение фронта горения.

Для оценки скорости фронта горения нам нужно записать уравнение среднего $\langle \frac{dh}{dt} \rangle$, для данного произвольного числа N полюсов в системе. Это уравнение следует напрямую из (4):

$$\langle \frac{dh}{dt} \rangle = \frac{1}{L^2} \frac{1}{2\pi} \int_0^{2\pi} u^2 d\theta. \quad (63)$$

После подстановки (8) в (63) мы получаем, используя (11) и (12)

$$\langle \frac{dh}{dt} \rangle = 2\nu \sum_{k=1}^N \frac{dy_k}{dt} + 2 \left(\frac{iN}{L} - \frac{\nu^2 N^2}{L^2} \right). \quad (64)$$

Оценки второго и третьего членов в этом уравнении будут самыми простыми. Записав $N = N(L) + \delta N(L)$ и помня, что $N(L) \sim L/\nu$ и $\delta N(L) \sim N(L)/2$, получаем, что эти члены вносят вклад в $O(1)$. Вклад от первого члена происходит только тогда, когда поток полюсов является асимметричным. Так как шум вводит полюса на конечной величине y_{\min} , в то время как отброшенные полюса стремятся к бесконечности и исчезают на границе нелинейности, определенной положением самого высокого полюса

$$y_{\max} \approx 2 \ln[4L\pi^2\nu]. \quad (65)$$

то мы имеем асимметрию, которая вносит вклад в скорость фронта. Для оценки первого члена давайте определим

$$d\left(\sum \frac{dy_k}{dt}\right) = \sum_1^{l+dl} \frac{dy_k}{dt} \quad (66)$$

где $\sum_1^{l+dl} \frac{dy_k}{dt}$ - сумма по полюсам на интервале $y : [l, l+dl]$. Мы можем записать:

$$d\left(\sum \frac{dy_k}{dt}\right) = d\left(\sum \frac{dy_k}{dt}\right)_{\text{up}} - d\left(\sum \frac{dy_k}{dt}\right)_{\text{down}} \quad (67)$$

Где $d(\sum \frac{dy_k}{dt})_{up}$ - поток полюсов направлен вверх, и $d(\sum \frac{dy_k}{dt})_{down}$ - поток полюсов направлен вниз.

Для этих потоков можем записать:

$$d(\sum \frac{dy_k}{dt})_{up}, d(\sum \frac{dy_k}{dt})_{down} \leq \frac{dN}{dt} dl \quad (68)$$

Соответственно для первого члена

$$\begin{aligned} 0 &\leq \sum_{k=1}^N \frac{dy_k}{dt} = \int_{y_{min}}^{y_{max}} \frac{d(\sum \frac{dy_k}{dt})}{dl} dl = \int_{y_{min}}^{y_{max}} \frac{d(\sum \frac{dy_k}{dt})_{up} - d(\sum \frac{dy_k}{dt})_{down}}{dl} dl \\ &\leq \frac{dN}{dt} (y_{max} - y_{min}) \leq \frac{dN}{dt} y_{max} \end{aligned} \quad (69)$$

Из-за медленной (\ln) зависимости y_{max} от L и ν , член $\frac{dN}{dt}$ определяет порядок нелинейности для первого члена. Этот член равен нулю для симметричного потока полюсов и достигает максимума для максимального асимметричного потока полюсов. Сравнение $\nu \sim L^{0.42} f^{0.02}$ и $\frac{dN}{dt} \sim L^{0.8} f^{0.03}$ подтверждает этот расчет.

1.2.1.4 Итоги и выводы

Основные две посылки этого раздела будут следующими:

- (i) Имеется значимое взаимодействие между нестабильностью развивающегося фронта и случайным шумом;
- (ii) Это взаимодействие и его предпосылки можно понять качественно и иногда количественно, используя описание через комплексные полюса.

Описание с помощью полюсов является естественным в этом контексте, во-первых, из-за того, что оно обеспечивает точное (и эффективное) представление стабильного состояния без шума. Как только удастся описать также *возмущения* в этом стабильном состоянии через полюса, можно добиться ясного языка для изучения взаимосвязи между шумом и нестабильностью.

Основными концептуальными этапами этой работы являются следующие: сначала нужно установить, что решение в виде стабильного состояния, которое характеризуется $N(L)$ полюсами, которые выровнены вдоль мнимой оси, будет слабо-устойчивым к шуму в периодическом массиве L величин. Для всех величин L стабильное состояние является нелинейно неустойчивым к шуму. Основной и наиболее важный эффект шума заданной амплитуды f - это введение избыточного числа полюсов $\delta N(L, f)$ в систему. Наличие этого дополнительного числа полюсов объясняет дополнительное «сморщивание» фронта горения на верху гигантского излома и наблюдаемое ускорение фронта горения. Рассмотрев появление шума новых полюсов, мы находим экспоненты масштабирования для наблюдаемых законов подобия как функции амплитуды шума и размера системы.

«Фазовая диаграмма» как функция L и f в этой системе включает в себя четыре режима. В первом режиме, который обсужден в Разделе 3.3.1, шум слишком мал для того, чтобы оказать какое-то воздействие на решение в виде гигантского излома. Второй режим

(с очень малым избыточным числом полюсов) не может наблюдаться из-за числового шума и обсуждается только теоретически. В третьем режиме шум вводит избыточные полюса, которые дополняют гигантский излом боковым изломами. В этом режиме мы обнаруживаем законы подобия для скорости в виде функции L и f , и мы достаточно хорошо продвигаемся в понимании экспонент масштабирования. В четвертом режиме шум достаточно велик, чтобы создавать небольшие подобные структуры, которые нельзя точно описать через индивидуальные полюса. Из наших численных данных, которые имеются в этом режиме, шерховатости фронта горения растут из структур малого масштаба таким способом, который напоминает *стабильные*, инициируемые шумом модели роста, например, модель Кадара-Париси-Занга (Kardar-Parisi-Zhang).

Одной из наших основных мотиваций для этого исследования была необходимость понимания явлений, наблюдаемых в радиальной геометрии при расширении фронта горения. Полный анализ этой проблемы здесь не представлен. Однако следует отметить, что понимание явлений в канальной геометрии, достигнутое выше, во многом позволяет прояснить и эту проблему. В самом деле, в радиальной геометрии фронт горения ускоряется, а изломы умножаются и образуют иерархическую структуру по мере течения времени. Так как радиус (и типовой масштаб) в этой системе возрастает все время, новые полюса будут добавляться в систему даже при исчезающе малом шуме. Маргинальная стабильность, обнаруживаемая выше, также имеет место в этом случае, и система позволит ввести дополнительные полюса в результате шума. Результаты, обсуждаемые в работе [126], в комбинации с пониманиями проблемы, позволили разработать теорию радиального роста. Эта теория была предложена в работе [127].

В радиальном случае [129] у нас была серьезная нерешенная проблема, но недавно эта проблема была успешно решена Карлиным и Сивашинским [131,132]. Для цилиндрического случая проблемы распространения фронта горения при отсутствии шума (а, точнее, только при наличии только численного шума, работы [145-149]), Сивашинский с помощью численных методов показал, что фронт горения постоянно ускоряется. В продолжение этого расчетного времени не наблюдалось никаких признаков насыщения. Увеличение расчетного времени – это трудная задача. Поэтому наличие или отсутствие насыщения скорости в цилиндрическом случае, как следствие уравнения движения фронта горения, было открытой проблемой до появления работ [131,132].

Для лучшего понимания зависимости скорости фронта горения от его радиуса для цилиндрического случая, с помощью численных методов была проанализирована также аналогичная зависимость скорости фронта горения от ширины канала (для плоского случая). Рост скорости наблюдается также и в отсутствие явного шума (а только при наличии численного шума), и не наблюдается никакого насыщения скорости. Введение явного Гауссового шума приводит к появлению точки насыщения, позволяет исследовать изменение уровня насыщения как функции амплитуды шума и экстраполировать эти результаты для малого численного шума (рис. 2).

Соответственно, введение Гауссового шума в численный расчет также и для цилиндрического случая может опять-таки привести к появлению точки насыщения и позволит исследовать ее поведение как функции амплитуды шума, а затем экстраполировать результаты на малый численный шум. Это исследование было действительно выполнено, и недавно насыщение наблюдали Карлин и Сивашинский [131,132] для одно-, двух- и трехмерного случаев.

И наконец, успех этого подхода в случае распространения пламени показывает, что Лапласовский рост можно анализировать, используя те же идеи. Основной интерес представляет Лапласовский рост в канале, для которого, как известно, имеется решение в виде «пальца» в стабильном состоянии. Известно, что стабильность такого решения в виде «пальца» под воздействием шума быстро уменьшается при увеличении ширины канала. Кроме того, известно, что такой шум вносит дополнительные геометрические

особенности на вершину «пальца». Здесь имеется достаточно сходных черт, которые указывают на то, что тщательный анализ аналитической теории может пролить свет на обе проблемы.

Глава 1.2.2 Проблема Сафмана-Тейлора без поверхностного натяжения в фильтрационном горении: Образование единственного "пальца" с шириной, равной половине ширины канала

1.2.2.1 Введение

Проблема образования структуры является сегодня одной из наиболее быстроразвивающихся ветвей нелинейной теории [119,125-128, 134,135,150-161].

Уравнение двухмерного Лапласовского роста описывает широкий круг физических задач, например - фильтрационное горение в пористой среде; вытеснение холодной жидкости (газа) в Хеле-Шоу канале такой же нагретой жидкостью (газом); происходящее в канале затвердевание жидкости в твердое тело и вытеснение им этой жидкости [152-154].

Это уравнение допускает красивые аналитические решения ([162] и ссылки в ней). Наиболее физически значимые решения – это решение в форме логарифмических полюсов («Тем не менее, значительный подкласс чисто логарифмических решений хорошо определен для всех положительных времен и описывает не сингулярную динамику поверхности при нулевом поверхностном натяжении» [162]). Однако подобное уравнение может приводить к возникновению сингулярностей за конечное время. Чтобы их предотвратить и регуляризовать проблему, в уравнение вводят член с поверхностным натяжением.

К сожалению, при этом аналитическое решение в форме полюсов становится невозможным. Кроме того, поверхностным натяжением объясняют появление асимптотического решения в виде пальца с шириной, равной половине ширине канала, наблюдающегося также в экспериментах.

В данной работе приводится математический механизм регуляризации, позволяющий избежать сингулярностей, получить желаемую асимптотику в виде пальца с шириной, равной половине ширине канала, сохранив при этом аналитическое решение в форме полюсов.

Сохранение аналитического характера решения крайне важно – оно позволит легко проанализировать решения уравнения двухмерного Лапласовского роста и объяснить качественно или количественно его поведение.

Автор искренне надеется, что эта работа сыграет ту же роль для уравнения двухмерного Лапласовского роста, что и работа [134] для теории горения смеси газов. Приведенное в [134] аналитическое решение и его асимптотика дали толчок к развитию теории горения смеси газов и позволило объяснить качественно или количественно поведение фронта горения газов [119,125-128,135].

Маковский, Алдушин [152-154] рассмотрели плоские, однородно распространяющиеся волны горения, приводимые в движение фильтрацией газа, содержащего окислитель, который реагирует со сгораемой пористой средой, через которую перемещается. Обычно эти волны оказывались нестабильными относительно гидродинамических возмущений, для направленного как вперед (попутный поток), так и назад (обратный поток) фильтрационного горения (ФГ), в котором направление потока газа будет соответственно тем же, что и направление распространения волны горения, либо противоположным ему.

Основной механизм, приводящий к нестабильности – это уменьшение сопротивления потоку в районе продуктов горения, что объясняется увеличением

пористости в данном районе. Другой дестабилизирующий эффект в направленном вперед ФГ – это образование газообразных продуктов в ходе реакции. В обратном ФГ этот эффект является стабилизирующим. В случае неустойчивости плоского фронта возникает альтернативный режим распространения в форме «пальца», который распространяется с постоянной скоростью. Область «пальца», заполненная продуктами горения, отделяется от участка без горения фронтом, в котором происходят химические реакции и передача тепла и массы.

В работах Маковского, Алдушина [152-154] было продемонстрировано, что решение проблемы фильтрационного горения с помощью «пальца» может быть охарактеризовано как решение проблемы Сафмана-Тейлора, которая изначально была сформулирована для описания вытеснения одной жидкости другой, имеющей меньшую вязкость, в пористой среде или в конфигурации Хеле-Шоу. Известно, что у проблемы ST имеется целое семейство решений в виде «пальца», при этом каждый «палец» характеризуется собственной скоростью и каждый занимает разную часть пористого канала, через который он распространяется. Скалярное поле, управляющее эволюцией поверхности раздела, является гармонической функцией. Поэтому, естественно назвать процесс *Лапласовским ростом*.

Математическая проблема Лапласовского роста без поверхностного натяжения имеет семейство точных аналитических решений, выражающиеся через логарифмические полюса в плоскости комплексной переменной.

Главная проблема такого решения – это имеющиеся сингулярности, возникающие за конечное время. Во избежание таких сингулярностей обычно применяется поверхностное натяжение, отличное от нуля ([155- 159]). Поверхностное натяжение также дает в результате четко определенное асимптотическое решение: единственный «палец» с шириной, равной половине ширины канала. Для регуляризации здесь могут использоваться также и другие методы (см. [160, 161] и ссылки в них).

Решения проблемы Сафмана-Тейлора без поверхностного натяжения аналогичны решению уравнения горения в случае сгорания смеси газов и образования ячеистой структуры пламени [119,125-128,134,135]. В самом деле, в обоих случаях решения могут быть преобразованы в набор обычных дифференциальных уравнений. Этот набор описывает движение полюсов в плоскости комплексной переменной.

Следует особо подчеркнуть, что фильтрационное горение и горение в смеси газов описываются разной физикой; уравнения двумерного Лапласовского роста и уравнения Михельсона-Сивашинского описывают совершенно разную математику. Более того, если для уравнения Михельсона-Сивашинского полюса входят в тригонометрические функции, для уравнения двумерного Лапласовского роста они входят в логарифмические функции. Аналогия тут не поверхностная, а глубокая – несмотря на различие в физике и математике, сложные интегро-дифференциальные уравнения имеют простое аналитическое решение в виде полюсов и, более того, даже поведение этих полюсов схоже. Однако, ненулевое поверхностное натяжение для проблемы Сафмана-Тейлора делает невозможным это замечательное аналитическое решение.

Другая проблема, упомянутая и обсужденная в работах проф. Маковского и д-ра А.П. Алдушина ([152-154]), заключается в том, что для фильтрационного горения в пористой среде поверхностное натяжение не может быть введено для математической проблемы, рассматриваемой Сафманом и Тейлором. Здесь зона химической реакции, диффузии тепла и массы сжимается до просто поверхности раздела, которая отделяет участок без горения от участка, где горения происходит. В данной ситуации нет скачков давления на поверхности раздела, поэтому здесь нельзя ввести поверхностное натяжение [152-154]. То есть, мы должны найти решение без ввода поверхностного натяжения. Далее проф. Маковский и д-р А.П. Алдушин ([152-154]) пишут, что однако разумно ожидать, что эффекты, даваемые поверхностным натяжением, могут быть получены другими

физическими возмущениями, релевантными конкретной рассматриваемой специфической проблеме. Например, в проблеме горения, как возмущение может быть рассмотрен эффект диффузии [152-154].

В работах проф. Маковского и д-ра А.П. Алдушина ([153-154]) были предложены критерии приводящие к выбору нужного семейства решений (это единственный «палец» с шириной, равной половине ширины канала). Эти критерии основываются на рассмотрении самой проблемы Сафмана-Тейлора, а не ее модификаций, полученных, например, путем добавления поверхностного натяжения к модели и далее принятия исчезновения поверхностного натяжения в пределе.

К сожалению, из работ [152-154] не понятно, *почему* Лапласовский рост без поверхностного натяжения (а именно, какой именно его *математический* механизм регуляризации) приводит к удовлетворению найденных критериев и, соответственно, дает известное асимптотическое решение (единственный «палец» с шириной, равной половине ширины канала). Найденные критерии не выведены теоретически из уравнений движения. Не показано, какой именно *математический* механизм регуляризации приводит к этому результату. Эти критерии являются лишь догадкой авторов, основанной на знании верной асимптотики из эксперимента.

Мы в данной части получаем верное асимптотическое решение и предлагаем *математический* механизм регуляризации, не основанный на введении поверхностного натяжения. Во-первых, мы можем ввести в систему небольшой шум. (Он может рассматриваться как поток полюсов из бесконечности). Во-вторых, для регуляризации проблемы мы будем отбрасывать все новые полюса, которые могут дать сингулярности, возникающие за конечное время. Можно строго доказать, что асимптотическим решением для такой проблемы будет единственный «палец». Более того, качественное рассмотрение показывает, что палец с величиной ширины канала $1/2$ является статистически стабильным. Соответственно, все свойства такого решения будут полностью такими же, что и для решения с ненулевым поверхностным натяжением при численном шуме.

Остальная часть работы построена следующим образом. В следующем Разделе 1.2.2.2 описывается формирование асимптотического единственного «пальца» Сафмана-Тейлора без поверхностного натяжения. Затем мы представляем аргументы по поводу формирования «пальца» Сафмана-Тейлора с шириной, равной половине ширины канала (Раздел 1.2.2.3). В заключительном разделе (Раздел 1.2.2.4) подводятся итоги и делаются выводы.

1.2.2.2 Формирование асимптотического единственного «пальца» Сафмана-Тейлора без поверхностного натяжения

В отсутствие поверхностного натяжения, которое воздействует в сторону стабилизации коротковолновых возмущений поверхности раздела, проблема двухмерного (2D) Лапласовского роста описывается следующим образом:

$$(\partial_x^2 + \partial_y^2)u = 0. \quad (1)$$

$$u|_{\Gamma(t)} = 0, \partial_n u|_{\Sigma} = 1. \quad (2)$$

$$v_n = \partial_n u|_{\Gamma(t)}. \quad (3)$$

Здесь $u(x, y; t)$ - скалярное поле, $\Gamma(t)$ - движущаяся поверхность раздела, Σ - неподвижная внешняя граница, ∂_n - компонент градиента, нормальный по отношению к

границе (т.е. производная по нормали), и v_n - проекция скорости фронта на нормаль к границе.

Введем физические граничные условия «без потока». Это означает отсутствие потока через боковые стенки канала. Для этого требуется, чтобы движущаяся поверхность раздела была ортогональна стенкам канала. Тем не менее, в отличие от случая с периодическими граничными условиями, две граничные точки движущейся поверхности на стенках канала не обязательно должны иметь одинаковую координату по вертикали. Однако, эту задачу можно свести к периодической, где период равен удвоенной ширине канала. Но только половина этой периодической полосы может рассматриваться как физический канал, в то время как вторая половина является нефизическим зеркальным образом.

Далее мы вводим конформное отображение с временной зависимостью f , отображающее нижнюю полуплоскость «математической» плоскости, $\xi \equiv \zeta + i\eta$ на область физической плоскости, $z \equiv x + iy$, где Лапласовское уравнение (1) определено как $\xi \xrightarrow{f} z$. Также мы выдвигаем требование, чтобы $f(\xi, t) \approx \xi$ для $\xi \rightarrow \zeta - i\infty$. Соответственно, функция $z = f(\zeta, t)$ описывает движущуюся поверхность раздела. Из уравнений (1), (2), (3) для функции $f(\xi, t)$ мы получаем Уравнение Лапласовского роста

$$\text{Im}\left(\frac{\partial f(\xi, t)}{\partial \xi} \overline{\frac{\partial f(\xi, t)}{\partial t}}\right) = 1 \Big|_{\xi=\zeta-i0}, f_{\xi} \Big|_{\zeta-i\infty} = 1. \quad (4)$$

Давайте искать решение Ур. (4) в следующей форме:

$$f(\xi, t) = \lambda \xi - i\tau(t) - i \sum_{l=1}^N \alpha_l \log(e^{i\xi} - e^{i\xi_l(t)}), \quad (5)$$

$$z(\zeta, t) = f(\zeta, t) = \lambda \zeta - i\tau(t) - i \sum_{l=1}^N \alpha_l \log(e^{i\xi} - e^{i\xi_l(t)}), \quad (6)$$

$$\alpha = \sum_{l=1}^N \alpha_l = 1 - \lambda, \quad (7)$$

где $\tau(t)$ - некоторая действительная функция времени, α_l - комплексная константа, $\xi_l = \zeta_l + i\eta_l$ обозначает положение полюса с номером l и N - число полюсов.

Для нашего граничного условия «без потока» мы должны добавить условие, что для каждого полюса $\xi_l = \zeta_l + i\eta_l$ с α_l имеется полюс $\xi_l = -\zeta_l + i\eta_l$ с $\overline{\alpha_l}$.

Соответственно, из этого условия и ур. (7) мы можем заключить, что λ является действительной константой.

Ниже мы докажем, что необходимым условием для отсутствия сингулярности, возникающей за конечное время, является следующее условие

$$-1 < \lambda < 1, \quad (8)$$

Также для функции $F(i\xi, t) = if(\xi, t)$ в случае граничного условия «без потока»:

$$\overline{F(i\xi, t)} = F(i\overline{\xi}, t) \quad (9)$$

Мы хотим доказать, что конечным состоянием процесса будет только один «палец», при условии, что в ходе эволюций полюсного решения не появляются сингулярности, возникающие за конечное время.

1.2.2.2.1 Асимптотическое поведение полюсов в математической плоскости

Приведенный здесь вывод аналогичен [163], но здесь мы также используем граничные условия «непротекания» (по аналогии с [164]). Главная цель этой главы – исследовать асимптотическое поведение полюсов в математической плоскости. Мы хотим продемонстрировать, что для времени $t \rightarrow \infty$, все полюса в случае периодических граничных условий сливаются в один полюс или смещаются к двум граничным точкам поверхности раздела на стенках канала в случае граничных условий «непротекания». Уравнением для поверхности раздела будет

$$f(\xi, t) = \lambda \xi - i\tau(t) - i \sum_{l=1}^N \alpha_l \log(e^{i\xi} - e^{i\xi_l(t)}),$$

$$\sum_{l=1}^N \alpha_l = 1 - \lambda, -1 < \lambda < 1. \quad (10)$$

Путем подстановки ур. (10) в Уравнение Лапласовского роста

$$\text{Im}\left(\frac{\partial f(\xi, t)}{\partial \xi} \frac{\partial \overline{f(\xi, t)}}{\partial t}\right) = 1 \Big|_{\xi=\zeta-i0} \quad (11)$$

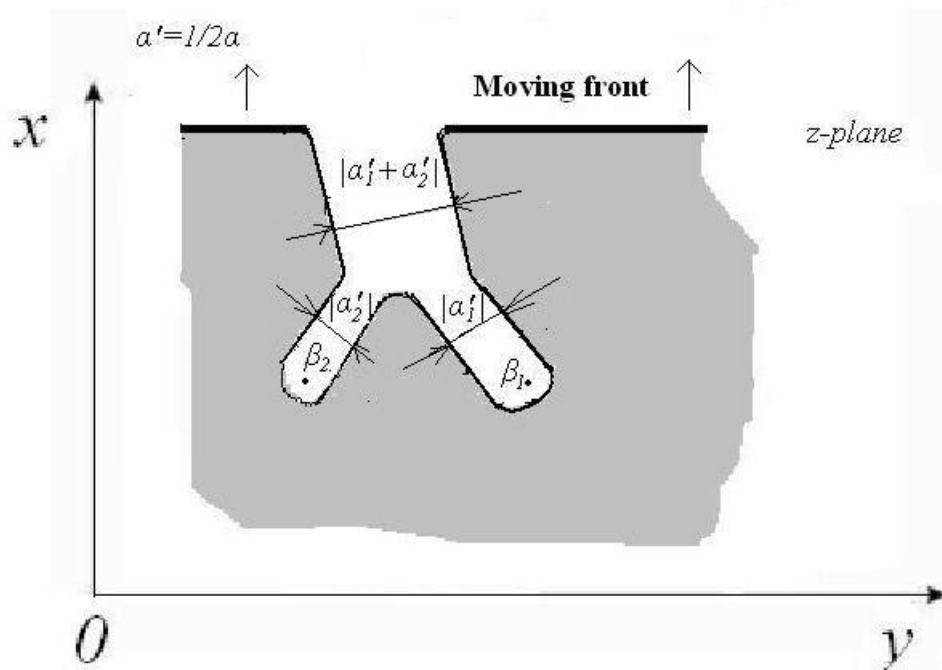


Рис.1: Геометрическая интерпретация комплексных констант движения $\alpha'_k = \frac{1}{2} \alpha_k$ и β_k ;

$$k = 1, \dots, N.$$

$$\alpha' = 1/2\alpha$$

мы можем найти уравнения движения для полюсов (рис. 1):

$$\beta_1 = \tau(t) + (1 - \sum_{k=1}^N \overline{\alpha_k}) \log \frac{1}{a_1} + \sum_{k=1}^N \overline{\alpha_k} \log \left(\frac{1}{a_1} - \overline{a_k} \right) = \text{const} \quad (12)$$

и

$$\tau = t - \frac{1}{2} \sum_{k=1}^N \sum_{l=1}^N \overline{\alpha_k} \alpha_l \log(1 - \overline{a_k} a_l) + C_0, \quad (13)$$

где $a_1 = e^{i\xi_1}$, C_0 является константой. Из уравнения (12) мы можем вывести

$$(1 - \lambda)\tau - \sum_{l=1}^N \alpha_l \log a_1 + \sum_{k=1}^N \sum_{l=1}^N \overline{\alpha_k} \alpha_l \log(1 - \overline{a_k} a_l) = \text{const}. \quad (14)$$

Из уравнений (13) и (14) мы получаем

$$\text{Im} \left(\sum_{l=1}^N \alpha_l \log a_1 \right) = \text{const} \quad (15)$$

и

$$t = \left(\frac{1 + \lambda}{2} \right) \tau + \frac{1}{2} \text{Re} \left(\sum_{l=1}^N \alpha_l \log a_1 \right) + C_1/2, \quad (16)$$

где C_1 и α_l - константы, $\xi_1(t)$ - положение полюсов, $a_1 = e^{i\xi_1(t)}$.

В Приложении А мы докажем из уравнения (13), что $\tau \mapsto \infty$, если $t \mapsto \infty$, при условии, что за конечное время не возникает сингулярности.

Уравнения движения полюса из уравнения (12) будут следующими:

$$\tau + i\overline{\xi_k} + \sum_1 \alpha_1 \log(1 - e^{i(\xi_1 - \overline{\xi_k})}) = \text{const}, \quad (17)$$

или в иной форме:

$$\zeta_k + \sum_1 (\alpha_1'' \log |1 - e^{i(\xi_1 - \overline{\xi_k})}| + \alpha_1' \arg(1 - e^{i(\xi_1 - \overline{\xi_k})})) = \text{const}, \quad (18)$$

$$\tau + \eta_k + \sum_1 (\alpha_1' \log |1 - e^{i(\xi_1 - \overline{\xi_k})}| - \alpha_1'' \arg(1 - e^{i(\xi_1 - \overline{\xi_k})})) = \text{const}, \quad (19)$$

где

$$\xi_1 = \zeta_1 + i\eta_1, \eta_1 > 0. \quad (20)$$

$$\alpha_1 = \alpha_1' + i\alpha_1'' . \quad (21)$$

Выполним следующее преобразование

$$\arg(1 - e^{i(\zeta_1 - \bar{\zeta}_k)}) = \arg([1 - e^{i(\zeta_1 - \zeta_k)} e^{-(\eta_1 + \eta_k)}]) = \arg[1 - a_{lk} e^{i\varphi_{lk}}] \quad (22)$$

где

$$\varphi_{lk} = \zeta_1 - \zeta_k, a_{lk} = e^{-(\eta_1 + \eta_k)} \quad (23)$$

Здесь $\arg[1 - a_{lk} e^{i\varphi_{lk}}]$ - однозначная функция φ_{lk} , т.е.

$$-\frac{\pi}{2} \leq \arg[1 - a_{lk} e^{i\varphi_{lk}}] \leq \frac{\pi}{2}. \quad (24)$$

Умножим ур. (19) на α_k'' и ур. (18) на α_k' и перенесем их разность в следующее уравнение:

$$\alpha_k' \zeta_k - \alpha_k'' \tau + \sum_{l \neq k} ((\alpha_1'' \alpha_k' - \alpha_k'' \alpha_1') \log |1 - e^{i(\zeta_1 - \bar{\zeta}_k)}| + (\alpha_1' \alpha_k' + \alpha_1'' \alpha_k'') \arg(1 - e^{i(\zeta_1 - \bar{\zeta}_k)})) = \text{const}. \quad (25)$$

Мы хотим исследовать асимптотическое поведение полюсов $\tau \mapsto \infty$. В этом уравнении у нас есть расходящиеся члены $\alpha_k'' \tau$. Из ур. (25) это расхождение может компенсировать только члены $\log |1 - e^{i(\zeta_1 - \bar{\zeta}_k)}|$. Необходимое условие для этого - $\eta_k \mapsto 0$ для $\tau \mapsto \infty, 1 \leq k \leq N$.

Можно предположить, что для $t \mapsto \infty$, имеются N' не слившихся групп полюсов ($N' \leq N$) ($\varphi_{lk} \mapsto 0$ для всех членов группы). N' в данный момент произвольно и даже может быть равным N . N_l - это номер полюсов в каждой группе, $1 \leq l \leq N'$.

Для каждой группы путем суммирования уравнений (25) по всем группам полюсов мы получаем

$$\alpha_k^{\text{gr}'} \zeta_k^{\text{gr}'} - \alpha_k^{\text{gr}''} \tau + \sum_{l \neq k} ((\alpha_1^{\text{gr}''} \alpha_k^{\text{gr}'} - \alpha_k^{\text{gr}''} \alpha_1^{\text{gr}'}) \log |1 - e^{i(\zeta_1^{\text{gr}'} - \bar{\zeta}_k^{\text{gr}'})}| + (\alpha_1^{\text{gr}'} \alpha_k^{\text{gr}'} + \alpha_1^{\text{gr}''} \alpha_k^{\text{gr}''}) \arg(1 - e^{i(\zeta_1^{\text{gr}'} - \bar{\zeta}_k^{\text{gr}'})})) = \text{const}, \quad (26)$$

где

$$\alpha_1^{\text{gr}''} = \sum_k^{N_l} \alpha_k'', \quad (27)$$

$$\alpha_1^{\text{gr}'} = \sum_k^{N_l} \alpha_k'. \quad (28)$$

Мы полагаем, что нет слияния групп для $\tau \mapsto \infty$, поэтому мы можем исследовать движение полюсов со следующим допущением:

$$|\zeta_1^{\text{gr}'} - \zeta_k^{\text{gr}'}| \gg \eta_1^{\text{gr}'} + \eta_k^{\text{gr}'}, 1 \leq l, k \leq N. \quad (29)$$

Для $l \neq k$, $\eta_k^{gr} \mapsto 0$, $\varphi_{lk}^{gr} = \zeta_l^{gr} - \zeta_k^{gr}$ мы получаем

$$\log |1 - e^{i(\zeta_l^{gr} - \zeta_k^{gr})}| \approx \log |1 - e^{i(\zeta_l^{gr} - \zeta_k^{gr})}| = \log 2 + \frac{1}{2} \log \sin^2 \frac{\varphi_{lk}^{gr}}{2} \quad (30)$$

и

$$\arg(1 - e^{i(\zeta_l^{gr} - \zeta_k^{gr})}) \approx \arg(1 - e^{i(\zeta_l^{gr} - \zeta_k^{gr})}) = \frac{\varphi_{lk}^{gr}}{2} + \pi n - \frac{\pi}{2}. \quad (31)$$

Мы выбираем n в ур.(31) таким образом, чтобы ур. (24) удовлетворялось. Подставляя эти результаты в уравнение (26), получаем

$$C_k = \alpha_k^{gr'} \zeta_k^{gr} - \alpha_k^{gr''} \tau + \sum_{l \neq k} [(\alpha_l^{gr''} \alpha_k^{gr'} - \alpha_k^{gr''} \alpha_l^{gr'}) \log |\sin \frac{\varphi_{lk}^{gr}}{2}| + (\alpha_l^{gr'} \alpha_k^{gr''} + \alpha_l^{gr''} \alpha_k^{gr'}) \frac{\varphi_{lk}^{gr}}{2}]. \quad (32)$$

1.2.2.2.2 Теорема о слияния полюсов

Из ур. (32) мы можем заключить

- (i) Путем суммирования уравнений (32) по всем группам полюсов (или сразу из ур. (15)) мы получаем

$$\sum_k \alpha_k^{gr'} \zeta_k^{gr} = \text{const}. \quad (33)$$

- (ii) Для $|\varphi_{lk}^{gr}| \mapsto 0, 2\pi$, мы получаем $\log \left| \sin \frac{\varphi_{lk}^{gr}}{2} \right| \mapsto \infty$, что означает, что полюса не могут проходить друг через друга;
- (iii) Из (ii) мы заключаем, что $0 < |\varphi_{lk}^{gr}| < 2\pi$
- (iv) Из (i) и (iii) мы видим, что $\zeta_k^{gr} \mapsto \infty$ невозможно;
- (v) В ур. (32) мы должны компенсировать второй расходящийся член. Из (iv) и (iii) мы можем получить это, только если $\alpha_l^{gr''} = \sum_k^{N_1} \alpha_k'' = 0$ для всех l .

Соответственно из ур. (32) получаем

$$\sum_k^{N_1} \alpha_k'' = 0, \quad (34)$$

$$\dot{\varphi}_{lk}^{gr} = 0, \quad (35)$$

$$\varphi_{lk}^{gr} \neq 0, \quad (36)$$

$$\dot{\zeta}_k^{gr} = 0. \quad (37)$$

Для асимптотического движения полюсов в группе N_m мы получаем из уравнений (34), (35), (36), (37) подставив главные члены в уравнения (16), (17)

$$\tau = \frac{2}{\lambda + 1} t, \quad (38)$$

$$\dot{\tau} + \sum_1^{N_m} \alpha_1 \frac{\dot{\eta}_k + \dot{\eta}_1 + i(\dot{\zeta}_k - \dot{\zeta}_1)}{\eta_k + \eta_1 + i(\zeta_k - \zeta_1)} = 0. \quad (39)$$

Решение для этих уравнений:

$$\eta_k = \eta_k^0 e^{-\frac{1}{\alpha_m^{\text{gr}}} \frac{2}{1+\lambda} t}, \quad (40)$$

$$\varphi_{lk} = \varphi_{lk}^0 e^{-\frac{1}{\alpha_m^{\text{gr}}} \frac{2}{1+\lambda} t}, \quad (41)$$

$$\dot{\zeta}_k = 0. \quad (42)$$

Соответственно, мы можем сделать вывод, что для компенсации расходящегося члена нам потребуется

$$\alpha_1^{\text{gr}''} = \sum_k^{N_l} \alpha_k'' = 0, \quad (43)$$

$$\alpha_1^{\text{gr}'} (1 + \lambda) > 0 \quad (44)$$

для всех l .

1.2.2.2.3 Конечный результат

Для периодических граничных условий уравнение. (43) верно для всех полюсов, таким образом мы получаем $N' = 1$, $m=1$ и $N_m=N$.

Следовательно, единственное решение будет:

$$\eta_k = \eta_k^0 e^{\frac{2}{1-\lambda^2} t}, \quad (45)$$

$$\varphi_{lk} = \varphi_{lk}^0 e^{-\frac{2}{1-\lambda^2} t}, \quad (46)$$

$$\dot{\zeta}_k = 0, \quad (47)$$

$$1 - \lambda^2 > 0. \quad (48)$$

Для граничных условий непротекания у нас имеются пары полюсов, для которых условие (43) выполняется. Поэтому все эти пары должны сходиться. В силу симметрии проблемы, эти полюса могут сходиться только на границах канала $\zeta = 0, \pm\pi$. Соответственно, мы получаем две группы полюсов на границах $N' = 2$, $m = 1, 2$, $N_1 + N_2 = N$, $\alpha_1^{gr'} + \alpha_2^{gr'} = 1 - \lambda$. (В принципе, возможны вырожденные случаи, когда условие (43) выполняется и для иных групп полюсов при некоторых значениях величин $\alpha_1^{gr''}$. Но такие ситуации крайне маловероятны и редки.)

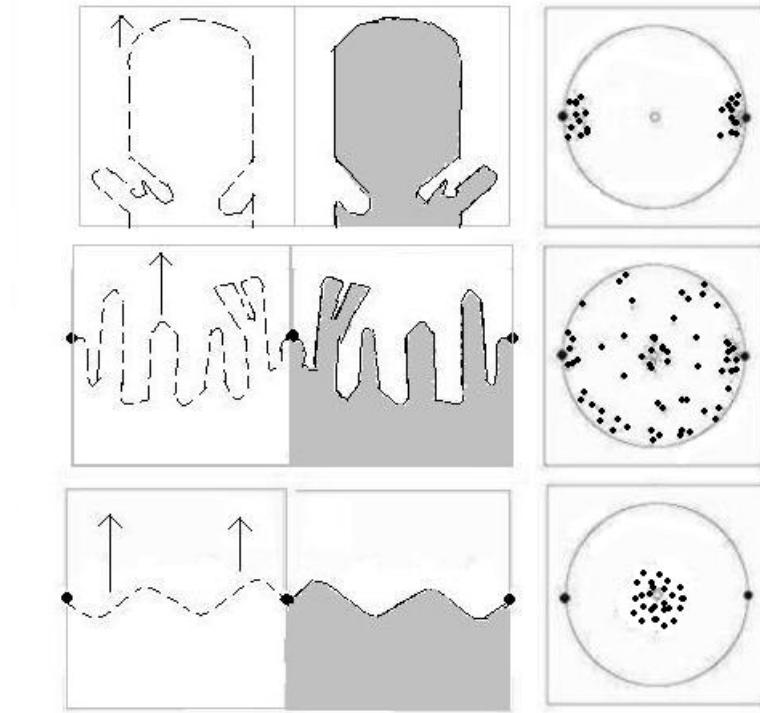


Рис.2: Три последовательных фазы вязкого пальцеобразования в ячейке Хеле-Шоу: начальная (слева), промежуточная (в центре), и асимптотическая (справа). Физическая плоскость z показана на верхних картинках, а на нижних картинках изображается распределение движущихся полюсов $a_k(t)$ в единичном круге $|\omega| < 1$ на математической плоскости ω . Маленький круг в центре обозначает область отталкивания для полюсов, $\omega = 0$, а Большой круг обозначает аттрактор $\omega = 1$ для полюсов, чья динамика отражена в (12-13).

Далее, мы получаем решение (на двух границах канала, рис. 2):

$$\eta_k^{(1)} = \eta_k^{(1),0} e^{-\frac{1}{\alpha_1^{gr'} + \lambda} t} \quad (49)$$

$$\varphi_{lk}^{(1)} = \varphi_{lk}^{(1),0} e^{-\frac{1}{\alpha_1^{gr'} + \lambda} t}, \quad (50)$$

$$\zeta_k^{(1)} = 0; \quad (51)$$

$$\eta_k^{(2)} = \eta_k^{(2),0} e^{-\frac{1}{\alpha_2^{gr'}(1+\lambda)} t}, \quad (52)$$

$$\varphi_{lk}^{(2)} = \varphi_{lk}^{(2),0} e^{-\frac{1}{\alpha_2^{gr'}(1+\lambda)} t}, \quad (53)$$

$$\zeta_k^{(2)} = \pm \pi; \quad (54)$$

$$\alpha_1^{gr'}(1+\lambda) > 0, \quad (55)$$

$$\alpha_2^{gr'}(1+\lambda) > 0. \quad (56)$$

Путем суммирования уравнений (55) и (56) и применив ур. (7), получаем:

$$(1-\lambda)(1+\lambda) = 1-\lambda^2 > 0. \quad (57)$$

Это позволяет непосредственно получить ранее сформулированное условие (8) для λ . $(\lambda+1)/2 = 1-\alpha/2$ имеет явно выраженный физический смысл. Это часть канала, занятая движущейся жидкостью. Мы видим, что в отсутствие сингулярности, возникающей за конечное время, и при $t \mapsto \infty$, мы получаем единственный «палец» с шириной $(\lambda+1)/2$.

1.2.2.3 Формирование пальца Сафмана-Тейлора с шириной, равной половине ширины канала

Случай Лапласовского роста в канале без поверхностного натяжения был детально рассмотрен Минеевым-Вайнштейном и Доусоном [165]. В этом случае имеется красивое аналитическое решение этой проблемы. Более того, они подразумевали, что все основные эффекты в случае исчезающе малого поверхностного натяжения можно также получить без поверхностного натяжения. Это позволит применить к исчезающе малому поверхностному натяжению мощные аналитические методы, разработанные для случая без поверхностного натяжения. Однако эта гипотеза не может быть принята без дополнительных допущений.

Первое возражение касается сингулярностей, возникающих за конечное время, для некоторых начальных состояний. Фактически, для того, чтобы преодолеть эту трудность, и был введен в уравнение новый член, описывающий поверхностное натяжение. Появление этого члена с поверхностным натяжением приводит к утрате аналитического решения. Однако регуляризацию можно выполнить гораздо более простым образом – просто отбросить начальные состояния, которые приводят к таким сингулярностям.

Второе возражение дается в работе Сиегеля и Танвира [166]. В ней показано, что при численном моделировании (подтвержденного некоторыми полуаналитическими расчетами с введением «дочерней сингулярности») в случае с любым (даже исчезающе малым) поверхностным натяжением «палец» с любой начальной толщиной в итоге занимает $1/2$ ширины канала за конечное время, не зависящее от величины малого поверхностного натяжения. Аналитическое же решение в случае без поверхностного натяжения дает в результате постоянную толщину «пальца», равную его начальной размеру, который может быть произвольным. Тем не менее, Сиегель и Танвир не принимали во внимание тот простой факт, что числовой шум вносит небольшое возмущение либо в начальное состояние, либо даже в ходе роста «пальца», которое

эквивалентно появлению удаленных полюсов, и с учетом этого возмущения аналитическое решение с постоянной толщиной «пальца» будет нестабильным.

Минеев-Вайнштейн [167] показал, что подобные полюсные возмущения могут привести, при определенных начальных состояниях, к расширению ширины пальца до ширины, полученной Сиегелем и Танвиром. Данный позитивный аспект работы [167] был отмечен Саркисяном и Левиным в их Комментариях [168] и в Отклике на их Комментарий Минеевым-Вайнштейном [169]. Подытоживая сказанное, мы можем отметить, что для совпадения результатов решения задач как с поверхностным натяжением, так и без него, необходимо ввести постоянный источник новых удаленных полюсов: это может быть либо внешний шум, либо бесконечное число полюсов в начальном состоянии. Какой из этих методов является предпочтительным – это пока еще открытый вопрос.

Конечно, этот дополнительный шум будет вносить и полюса, приводящие и к решениям, отличным от решения найденного Сиегелем и Танвиром [168]. Но эти решения при наличии шума будут возникать как при наличии поверхностного натяжения, так и без него. Таким образом, введение шума стирает разницу между уравнениями с/без поверхностного натяжения.

В случае распространения фронта пламени было показано [119,125-128], что внешний шум необходим для объяснения увеличения скорости пламени с увеличением размеров системы: бесконечное число полюсов в начальном состоянии не может дать такой результат. Интересно было бы понять, такова ли ситуация для Лапласовского роста в канале. Один из основных результатов Лапласовского роста в канале с небольшим поверхностным натяжением – это формирование «пальца» Сафмана-Тейлора с толщиной, равной $1/2$ значениям толщины канала. И чтобы использовать аналитический результат, полученный для нулевого поверхностного натяжения, необходимо доказать, что образование «пальца» с толщиной, равной $1/2$ значениям толщины канала, также имеет место и без поверхностного натяжения.

В нашей совместной с Минеевым-Вайнштейном работе [163] показано, что асимптотическое решение будет в виде «пальца» произвольной толщины. Это верно для любого конечного числа полюсов и при почти всех разрешенных начальных состояниях (разрешенных в том смысле, что они не приводят к сигнуляриям, возникающими за конечное время), за исключением небольшого числа некоторых вырожденных начальных состояний. Следует отметить, что решения и асимптотика, которые можно найти в [163] для любого конечного числа полюсов, являются идеализацией. Они, однако, обладают реальным смыслом для любых конечных интервалов времени между появлением новых полюсов. Эти полюса появляются в системе или за счет внешнего шума, или связаны с появлением в системе ранее близких к бесконечности полюсов из начального состояния. Начальное состояние может включать бесконечное число таких полюсов. Теорема, доказанная в [163], может быть применена еще раз для нового текущего конечного набора новых появившихся и старых полюсов, и снова получается асимптотика, которая является опять-таки «пальцем», но уже с возможной новой толщиной, отличной от прежней. Таким образом, введение источника новых полюсов приводит, в результате, к возможному дрейфу толщины «пальца», но не к изменению типа самого решения в виде «пальца».

Нужно отметить, что вместо периодических граничных условий можно ввести более реалистичные граничные условия «непротекания» [164]. Граничные условия «непротекания» - это условия, которые запрещают поток вещества через стенку канала. Граничные условия «непротекания» вносят полезные ограничения на параметры полюсов. Эти условия объясняют почему сумма всех комплексных параметров α_i для полюсов дает действительную величину α для полюсных решений (5) в [167].

Введение граничного условия «непротекания», тем не менее, оставляет в силе все полученные в [163] результаты для случая периодических граничных условий. Более того, граничное условие «непротекания» не вносит при этом ничего принципиально нового.

Этот факт обоснован в данной работе ниже. Так, например, в работе [164] была повторно выведена теорема об асимптотике в виде одного «пальца» для граничных условий «непротекания» (См. ссылку 14 in [167] и ссылку 20 (и соответствующий ей текст) в [163]). Но теорема об асимптотике в виде одного «пальца» уже была ранее доказана в работе [163] для периодических граничных условий.

Минеев-Вайнштейн в [167] пытается доказать, что асимптотика для Лапласовского роста в канале с нулевым поверхностным натяжением представляет собой единственный «палец» с толщиной, равной $1/2$ толщины канала. Но это доказательство является ошибочным. На самом деле, Минеев-Вайнштейн использует в [167] некий метод для доказательства и демонстрации неустойчивости по отношению к появлению новых удаленных полюсов «пальца» с толщиной, отличной от $1/2$. Однако, тот же самый метод может быть использован и для доказательства неустойчивости «пальца» с толщиной равной $1/2$, что подробно обосновывается в Комментариях Касадемунта, Магдалено и Алмгрена [170,171]. Касадемунта и Магдалено пишут, что возмущение решения в виде пальца (тонее, члена $\lambda\zeta$ в конформном отображении $z(\zeta, t)$), рассматриваемое Минеевым-Вайнштейном [167] представляет частный случай более общего возмущения:

$$\lambda\zeta \approx (\lambda - \lambda_0)\zeta - i \sum_{k=1}^N \delta_k \log(e^{i\zeta} - \varepsilon_k), \quad (58)$$

где ε_k мало, а $\sum_{k=1}^N \delta_k = \lambda_0$.

Возмущение рассматриваемое Минеевым-Вайнштейном в [167] соответствует случаю $\lambda = \lambda_0$:

$$\lambda\zeta \approx -i \sum_{k=1}^N \delta_k \log(e^{i\zeta} - \varepsilon_k), \quad (59)$$

Палец толщиной $1/2$ соответствует случаю $\mu = 0$. Такой палец будет устойчив к возмущению Минеева-Вайнштейна в [167], но неустойчив к более общему решению Касадемунта и Магдалено в [170]. Это опровергает тезис [167], что решение в виде пальца $1/2$ устойчиво и, соответственно, является асимптотикой всех других неустойчивых решений.

Ответ на это возражение приводится Минеевым-Вайнштейном в Отклике [172]. Он показывает, что возмущение Касадемунта и Магдалено легко преобразуется к виду возмущения Минеева-Вайнштейна добавлением полюса, равного нулю. Отсюда Минеев-Вайнштейн делает два вывода:

- 1) Возмущение Касадемунта и Магдалено неустойчиво. Действительно, для доказательства неустойчивости достаточно показать неустойчивость по отношению к хотя бы одному возмущению. Таким возмущением является добавочный полюс. Поскольку возмущение Касадемунта и Магдалено неустойчиво и переходит в возмущение Минеева-Вайнштейна, делается вывод, что его не имеет смысл использовать.
- 2) Возмущение Касадемунта и Магдалено является лишь «узким» подмножеством возмущения Минеева-Вайнштейна для случая добавочного полюса в нуле. Отсюда делается вывод, что возмущение Касадемунта и Магдалено «много реже встречается», чем возмущение Минеева-Вайнштейна. Поэтому возмущение Касадемунта и Магдалено не имеет смысл использовать.

Первое из возражений парадоксальным образом работает скорее против аргументов самого Минеева-Вайнштейна. Действительно, если «для доказательства

неустойчивости достаточно показать неустойчивость по отношению к хотя бы одному возмущению», то неустойчивости пальца $1/2$ по отношению к возмущению Касадемунта и Магдалено вполне достаточно для доказательства неустойчивости пальца $1/2$. Соответственно, он не может быть асимптотикой.

Второе возражение содержит простую математическую ошибку. Действительно, для бесконечных множеств подмножество может быть равным самому множеству. Так квадраты натуральных чисел составляют лишь подмножество натуральных чисел. Но между обоими множествами существует очевидное взаимно-однозначное соответствие. В данном случае каждому возмущению Минеева-Вайнштейна, приводящему к пальцу $1/2$:

$$\lambda \zeta \approx -i \sum_{k=1}^N \delta_k \log(e^{i\zeta} - \varepsilon_k), \quad (60)$$

можно сопоставить возмущение Касадемунта и Магдалено:

$$0\zeta \approx (0 - (-\lambda))\zeta - i \sum_{k=1}^N (-\delta_k) \log(e^{i\zeta} - \varepsilon_k), \quad (61)$$

возвращающее толщину пальца от $1/2$ к исходному значению. Добавим в подтверждение выше сказанного, что в нашей совместной работе с Минеевым-Вайнштейном [163] было показано, что для конечного числа полюсов «палец» любой толщины возможен в качестве асимптотического решения. Однако существует фактор, который может разрушать описанное выше взаимно-однозначное соответствие между возмущениями Минеева-Вайнштейна, приводящему к пальцу $1/2$, и возмущениями Касадемунта и Магдалено, возвращающее толщину пальца от $1/2$ к исходному. Дело в том, что часть этих возмущений будут приводить к сингулярности, возникающей за конечное время, вместо асимптотики в виде пальца. Положим, что подобные сингулярности чаще встречаются для возмущений Касадемунта и Магдалено (возвращающих толщину пальца от $1/2$ к исходной) чем для соответствующих возмущений Минеева-Вайнштейна (приводящих к пальцу $1/2$). Тогда это приводит к выделенной роли пальца толщиной $1/2$. Именно такую ситуацию мы и предполагаем ниже (см. Рис. 3).

Тем не менее, это не означает, что привилегированную роль «пальца» толщиной $1/2$ нельзя доказать в случае отсутствия поверхностного натяжения; это означает только то, что такое доказательство не приведено в [167]. Попробуем привести эти правильные аргументы здесь. Общее полюсное решение (5) в работе [167] характеризуется действительным параметром α который является суммой комплексных параметров α_i для полюсов. Толщина асимптотического пальца – это простая функция α . А именно, его толщина составляет $1 - \alpha/2$ от ширины канала. Величина $\alpha = 1$ соответствует толщине $1/2$. Толщина «пальца» должна быть между 0 и 1. Соответственно, возможная величина α находится в интервале между 0 и 2: ($0 < \alpha < 2$). Величина $\alpha = 1$, соответствующая ширине пальца $1/2$, находится точно по середине этого интервала. Что происходит с вполне возможными начальными состояниями полюсов, когда α лежит вне пределов от 0 до 2? Они «не допустимы» из-за описанных выше сингулярностей, возникающих за конечное время [163]. Также, и часть решений для α , лежащего в рамках интервала $0 < \alpha < 2$, приводит к сингулярностям.

Нахождение точных необходимых и достаточных условия, определяющих, приводит то или иное начальное состояние полюсов к сингулярности, все еще является открытой проблемой. Назовем начальные состояния с конечным числом полюсов, приводящие к сингулярности, «недопустимыми». Каждому α соответствует некий процент «недопустимых» начальных состояний от всего числа возможных полюсных

начальных состояний с конечным числом полюсов. Каков этот процент для различных α в интервале $0 < \alpha < 2$?

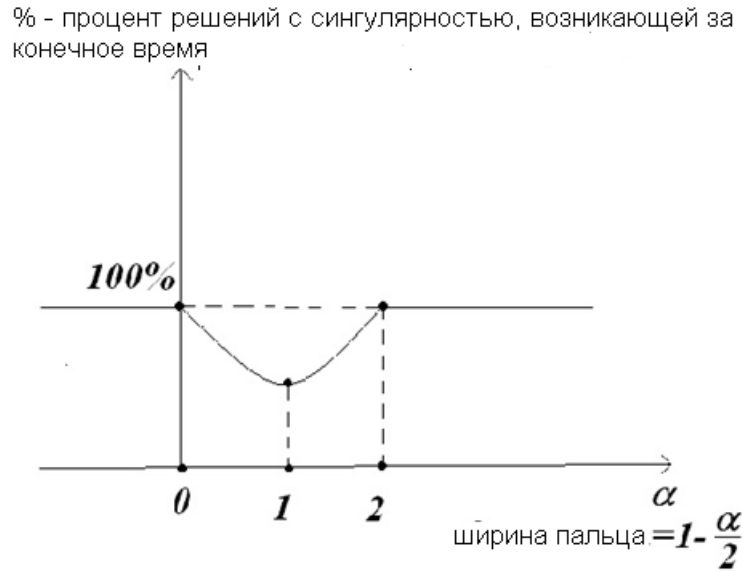


Рис.3: Рассматриваются решения с конечным числом полюсов. На графике показан процент сингулярных решений от общего числа решений как функция от α . Максимальная величина функции равна 100 процентам и соответствует $\alpha \leq 0$ или $\alpha \geq 2$. Минимальная величина функции находится в средней точке $\alpha = 1$, между $\alpha = 0$ (ширина пальца равна 1) и $\alpha = 2$ (ширина пальца равна 0). Таким образом, минимуму соответствует $\alpha = 1$ и ширина пальца равная $1/2$. В общем случае, ширина пальца составляет $1 - \alpha/2$. Здесь за 1 принимается ширина какнала.

Ответ изображен на рис. 3. Из соображений непрерывности и симметрии относительно $\alpha = 1$ (соответствует толщине пальца $1/2$) можно заключить, что это распределение имеет минимум в точке $\alpha = 1$. Величина $\alpha = 1$ является равноудаленной от обеих границ интервала $0 < \alpha < 2$. Процент «недопустимых» начальных состояний увеличивается по мере приближения к границам $\alpha = 2$ или 0, и достигает 100 процентов от всех полюсных решений решений за пределами этих границ. Таким образом, толщина $1/2$ является наиболее вероятной, так как для этой величины процент "недопустимых" (т.е. приводящих к сингулярности) начальных состояний достигает минимума.

Появление новых полюсов из шума приводит к дрейфу толщины пальца, однако эти колебания толщины происходят в окрестности наиболее вероятного среднего значения, равного $1/2$. Подобный результат получается и в случае пальца Сафмана-Тейлора с очень малым поверхностным натяжением при наличии некоторого внешнего шума.

Сформулируем кратко наши выводы. Для детального рассмотрения устойчивости решения необходимо явным образом рассматривать шум, который можно представить, как поток полюсов из нуля. Если такого шума явно нет в компьютерной задаче, его заменяет численный шум, связанный с конечной точностью вычислений. При наличии такого шума можно рассматривать не асимптотическую, а стохастическую устойчивость пальца $1/2$. Она возникает при регуляризации решения отбрасыванием полюсов из шума, могущих приводить к сингулярности за конечное время. При таком отбрасывании оказывается, что для пальца $1/2$ вероятность появления полюсов уменьшающих или увеличивающих его толщину одинакова. Для пальца же с толщиной отличной от $1/2$ появление плюсов, сдвигающих его толщину к $1/2$ более вероятно.

Интересно с этой точки зрения рассмотреть результаты Кесслера и Левине [173]. Там рассматривается случай асимметричного поверхностного натяжения, стремящегося к нулю. Показано, что для очень малого поверхностного натяжения асимптотой является не палец $1/2$, а случайный шум. Отсюда делается вывод о бессмысленности анализа решения в виде пальца $1/2$ для случая поверхностного натяжения стремящегося к нулю. Однако авторы делают ту же ошибку, что и Сиегель и Танвир в [166]. Необходимо сравнивать решения с/без поверхностного натяжения, вводя не только малое поверхностное натяжение, но и не нулевой шум (роль которого при вычислениях может играть и численный шум). Для малого поверхностного натяжения шум, при достаточной величине, может приводить сразу к появлению новой сингулярности, до того или сразу после того как прежняя уже сглажена малым поверхностным натяжением. Это и приводит к описанному в этой части случайному решению [174]. Т.е. для получения асимптотики в виде пальца $1/2$ нужно не только малое поверхностное натяжение, но и малый шум, что не учтено в [173].

Нужно отметить, что сформулированные аргументы являются лишь качественными, также необходимы и более строгие доказательства. Первый шаг в этом направлении сделан в статьях [174, 175]. К сожалению, анализ асимптотического решения в [174] проведен для уравнения Лапласовского роста в отсутствие шума, что физически бессмысленно. Для асимптотического решения с шумом и регуляризацией (путем отбрасыванием сингулярных и нефизических решений) результаты [174] скорее подтверждают выводы этой работы и могут служить преемным шагом в их математической формализации.

1.2.2.4 Выводы

Аналитическое полюсное решение для Лапласового роста иногда дает сингулярности, возникающие за конечное время. Однако, существует красивое решение этой проблемы. Во-первых, мы вводим в систему некоторый небольшой шум. Этот шум может считаться потоком полюсов из бесконечности. Во-вторых, для регуляризации проблемы мы отбросим все новые полюса, которые могут создать сингулярности, возникающие за конечное время. Можно строго доказать, что асимптотическим решением для такой системы будет единственный «палец». Более того, качественное рассмотрение показывает, что палец с шириной $1/2$ канала является статистически стабильным. Соответственно, все свойства такого решения являются совершенно такими же, что и свойства решения с малым ненулевым поверхностным натяжением при наличии численного шума.

Удивительно, что иная проблема, описывающая распространение фронта пламени горения в смеси газов (несмотря на абсолютно разные физические и математические уравнения для этих двух случаев) имеет также аналитические полюсные решения и демонстрирует качественно похожее поведение [119, 125-128, 134, 135].

Приложение А

Нам нужно доказать, что $\tau \mapsto \infty$, если $t \mapsto \infty$ при условии, что за конечное время не возникает сигнуляльность. Формула для τ будет следующей:

$$\tau = t + [-12 \sum_{k=1}^N \sum_{l=1}^N \overline{\alpha_k} \alpha_l \log(1 - \overline{a_k} a_l)] + C_0, \quad (62)$$

где $|a_l| < 1$ для всех l .

Докажем, что второй член в этой формуле больше нуля:

$$\begin{aligned} -12 \sum_{k=1}^N \sum_{l=1}^N \overline{\alpha_k} \alpha_l \log(1 - \overline{a_k} a_l) &= -12 \sum_{k=1}^N \sum_{l=1}^N \overline{\alpha_k} \alpha_l \sum_{n=1}^{\infty} (-\overline{a_k} a_l)^n n = 12 \sum_{n=1}^{\infty} \ln \left(\sum_{k=1}^N \overline{\alpha_k} (\overline{a_k})^n \right) \left(\sum_{l=1}^N \alpha_l (a_l)^n \right) = \\ 12 \sum_{n=1}^{\infty} \ln \left(\sum_{l=1}^N \alpha_l (a_l)^n \right) \left(\sum_{l=1}^N \alpha_l (a_l)^n \right) &> 0 \end{aligned} \quad (63)$$

Итак, второй член в ур. (62) всегда больше нуля и, соответственно, $\tau \mapsto \infty$, если $t \mapsto \infty$ при условии, что за конечное время не возникает сигнуляльность.

Глава 2 Эмерджентные системы, не рассчитываемые аналитически на основе известной физики подсистем.

Для эмерджентных систем, т.е. систем, не рассчитываемых аналитически на основе известной физики подсистем, получены результаты для учета внешнего шума/наблюдателя.

Глава 2.1 Третья группа проблем, где учет внешнего шума/наблюдателя за счет модели «черного ящика» и построение его наблюдаемой динамики происходит из реального/численного эксперимента.

В этой главе мы рассматриваем третью группу проблем. Для третьей группы проблем учет внешнего шума от окружения/наблюдателя происходит за счет модели «черного ящика». Построение его наблюдаемой динамики происходит из реального или численного эксперимента. В главе доказана оптимальность наивной байесовской модели, используемой при поиске количественных соотношений структура-свойство в молекулярной физике (Quantitative Structure-Activity Relationship (QSAR))

Глава 2.1.1 Введение

Байесовы Классификаторы широко в настоящее время используются для распознавания, идентификации и получения знания. Области применения - например, обработка изображения, медицина, химия (QSAR) [176] на Рис. 1.

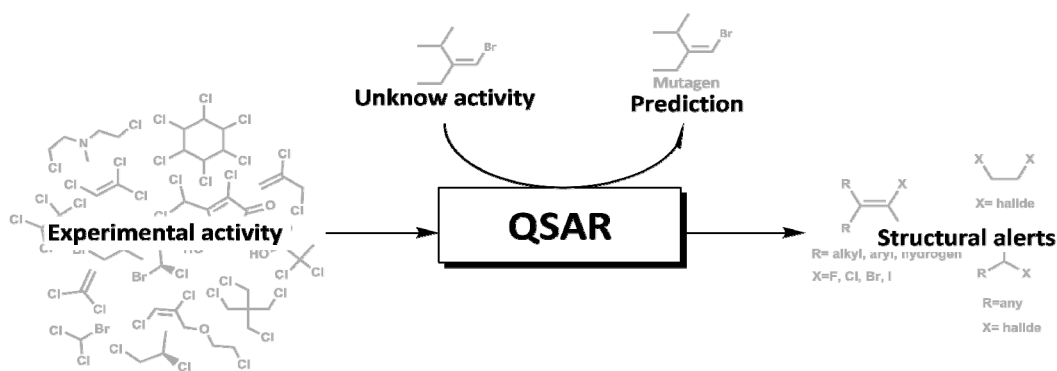


Рис. 1 QSAR

Позвольте нам дать пример использования методов QSAR в статьях [177,178]: «Молекулярное распознавание и связывание, выполняемого белками, являются фоном всех биохимических процессов в живой клетке. В частности, обычный механизм действия лечебного препарата - эффективное закрепление и ингибирование деятельности целевого белка. Прямое моделирование молекулярных взаимодействий в комплексах ингибитор-белок - основа современного метода компьютерной разработки лекарств, однако такое моделирование является чрезвычайно сложной проблемой ... В предлагаемом методе, подобие для связываемого участка осуществляется распознаванием существования химически и пространственно аналогичных областей по отношению к области связывания. Мы представляем новый метод определения местного подобия для участка связывания, основанный на анализе полной окружающей среды белка для фрагментов лиганда. Сравнение области связывания изучаемого белка (цель) с трехмерной структурой другого белка (аналога) в комплексе с лигандом позволяет фрагментам лиганда от аналогичного комплекса быть соотношенным с их возможным положением в целевом участке, так, чтобы полная окружающая среда белка для фрагмента и его аналога были подобны. Соответствующие окружения - подобные области и фрагменты, примыкающие к целевому участку, рассматриваются как типовые образцы связывания. Набор таких целевых образцов связывания, полученных из базы данных аналоговых комплексов, формирует подобную облаку структуру (облако фрагментов), которая является сильным инструментом для компьютерной разработки лекарств»

Особое значение Байесовские Классификаторы имеют в медицинской диагностике и биоинформатике. Очень хороший пример может быть найден в статье [179].

Однако, таинственным путем Наивный Байесовский Классификатор обычно дает очень хорошее представление распознавания. Более сложные модели Байесового Классификатора не могут улучшить его значительно [176]. Мы даем простое доказательство оптимальности Наивного Байесовского Классификатора, которое может объяснить этот интересный факт.

В статье [180] авторы объясняют это замечательное свойство. Однако, они используют некоторое предположение (Zero-One Loss), которое уменьшает универсальность и общность этого рассмотрения. Мы даем более общее доказательство оптимальности Наивного Байесовского Классификатора [176]. Последующее интересное развитие задачи было сделано в [181] (2004), [182] (2006). Однако, к сожалению эта статья не включает анализа предыдущей [176].

Позвольте нам сформулировать коротко основную задачу, которую мы пытаемся решить. Предположив, что у нас есть набор некоторых объектов и набор переменных, которые характеризуют эти объекты. Для каждого объекта мы знаем вероятностное распределение для каждой переменной. Однако, у нас нет никакой информации о корреляциях переменных. Теперь, предположим, что мы знаем значения переменных для

некоторого объекта из набора объектов. Какова вероятность, что эта выборка соответствует некоторому объекту? Это - типичная задача распознавания при условии неполноты информации.

Позвольте нам рассмотреть самый простой случай, когда никакие корреляции не существуют между переменными. В этом случае, Наивная Байесовская модель - точное решение задачи. Мы докажем, что для случая, когда мы ничего не знаем о корреляциях - Наивная Байесовская модель не точное, но оптимальное решение в некотором смысле. Более детально, мы докажем, что Наивная Байесовская модель дает минимальную среднюю ошибку по всем возможным моделям корреляций. Мы предполагаем в этом доказательстве, что у всех моделей корреляций есть равная вероятность. Мы думаем, что этот результат может объяснить описанную выше таинственную оптимальность Наивной Байесовской модели.

Материал организован следующим образом. В разделе 2.1.2 мы даем точное математическое определение задачи для двух переменных и двух объектов. В разделе 2.1.3 мы определяем обозначения. В разделе 2.1.4 мы даем общую форму условной вероятности для всех возможных корреляций наших переменных. В разделе 2.1.5 мы находим ограничения функций, описывающих корреляции. В разделе 2.1.6 мы даем определение расстоянию между двумя вероятностями (корреляционными) моделями. В разделе 2.1.7 мы находим ограничения для наших основных функций. В разделе 2.1.8 мы решаем свою главную задачу - мы доказываем оптимальность Наивной Байесовской модели для равномерного распределения всех возможных корреляций. В разделе 2.1.9 мы находим среднюю ошибку между Наивной Байесовской моделью и фактической моделью при равномерном распределении всех возможных корреляций. В разделе 2.1.10 мы рассматриваем случай более чем двух переменных и объектов. Последний раздел - заключение.

Глава 2.1.2 Постановка задачи

Пусть A - случайная величина, со значениями $0,1$. Предположим, что априорная вероятность $P(A) = P(A = 1)$ известна и обозначим ее θ . Пусть X_1, X_2 - две случайные величины, со значениями в некотором множестве, например $]-\infty; +\infty[$. Мы имеем следующую информацию: $X_1 = x_1$ и $X_2 = x_2$ (полученную из измерения). Кроме того, у нас есть две системы - "классификаторы", которые для данных x_1 и x_2 дают:

$$P(A = 1/X_1 = x_1) = P(A/x_1) \doteq \alpha,$$

$$P(A = 1/X_2 = x_2) = P(A/x_2) \doteq \beta.$$

Мы хотим оценить вероятности $P(A = 1/X_1 = x_1, X_2 = x_2) = P(A/x_1, x_2)$ в терминах α, β и θ . Более определенно, мы хотим найти функцию $\Gamma_{\text{opt}}(\alpha, \beta, \theta)$, которая в среднем является наилучшим приближением для $P(A/x_1, x_2)$ в смысле, определенным явно в продолжении (см. рис. 2).

Глава 2.1.3 Предварительные обозначения

$\rho_{X_1, X_2}(x_1, x_2)$ - совместная ФПВ (функция плотности вероятности) X_1 и X_2 .

$\rho_{X_1, X_2/A}(x_1, x_2) \doteq h(x_1, x_2)$ - совместная ФПВ X_1 и X_2 для $A = 1$. В терминах $h(x_1, x_2)$ и

θ мы можем написать $P(A/x_1, x_2)$ следующим образом:

$$P(A/x_1, x_2) = \frac{\theta h(x_1, x_2)}{\theta h(x_1, x_2) + (1 - \theta) \bar{h}(x_1, x_2)}, \quad (1)$$

Где

$\bar{h}(x_1, x_2) \doteq \rho_{X_1, X_2/A}^-(x_1, x_2)$ - совместная ФПВ X_1 и X_2 для $A = 0$.

Мы имеем:

$$\rho_{X_1}(x_1) = \int_{-\infty}^{+\infty} \rho_{X_1, X_2}(x_1, x_2) dx_2,$$

$$\rho_{X_2}(x_2) = \int_{-\infty}^{+\infty} \rho_{X_1, X_2}(x_1, x_2) dx_1,$$

$$h_1(x_1) \doteq \rho_{X_1/A}(x_1) = \int_{-\infty}^{+\infty} h(x_1, x_2) dx_2,$$

$$h_2(x_2) \doteq \rho_{X_2/A}(x_2) = \int_{-\infty}^{+\infty} h(x_1, x_2) dx_1,$$

$$\bar{h}_1(x_1) \doteq \rho_{X_1/A}^-(x_1) = \int_{-\infty}^{+\infty} \bar{h}(x_1, x_2) dx_2,$$

$$\bar{h}_2(x_2) \doteq \rho_{X_2/A}^-(x_2) = \int_{-\infty}^{+\infty} \bar{h}(x_1, x_2) dx_1.$$

Глава 2.1.4 Общая форма $P(A/x_1, x_2)$

Определим функции $g(x_1, x_2)$ и $\bar{g}(x_1, x_2)$

$$g(x_1, x_2) \doteq \frac{h(x_1, x_2)}{h_1(x_1)h_2(x_2)},$$

$$\bar{g}(x_1, x_2) \doteq \frac{\bar{h}(x_1, x_2)}{\bar{h}_1(x_1)\bar{h}_2(x_2)}.$$

Заметим, что если X_1 и X_2 являются условно независимыми, т.е.

$$h(x_1, x_2) = \rho_{X_1 X_2/A}(x_1, x_2) = \rho_{X_1/A}(x_1) \rho_{X_2/A}(x_2) = h_1(x_1) h_2(x_2),$$

тогда

$$g(x_1, x_2) = \bar{g}(x_1, x_2) = 1.$$

Определим следующую монотонно неубывающую функцию плотности вероятности:

$$H_1(x_1) \doteq \int_{-\infty}^{x_1} h_1(z) dz,$$

$$H_2(x_2) \doteq \int_{-\infty}^{x_2} h_2(z) dz,$$

$$\bar{H}_1(x_1) \doteq \int_{-\infty}^{x_1} \bar{h}_1(z) dz,$$

$$\bar{H}_2(x_2) \doteq \int_{-\infty}^{x_2} \bar{h}_2(z) dz.$$

Отметим, что, так как $H_1(x_1), H_2(x_2), \bar{H}_1(x_1)$ и $\bar{H}_2(x_2)$ являются монотонными (здесь, можно предположить, что $h_1(x_1), h_2(x_2), \bar{h}_1(x_1), \bar{h}_2(x_2) > 0$, так, что $H_1(x_1), H_2(x_2), \bar{H}_1(x_1)$ и $\bar{H}_2(x_2)$ монотонно увеличиваются. Это ограничение, как будет показано в продолжении, является лишним), то существуют обратные функции $H_1^{-1}(x_1), H_2^{-1}(x_2), \bar{H}_1^{-1}(x_1)$ и $\bar{H}_2^{-1}(x_2)$. Мы можем поэтому определить:

$$J(a, b) \doteq g(H_1^{-1}(a), H_2^{-1}(b)),$$

$$\bar{J}(a, b) \doteq g(\bar{H}_1^{-1}(a), \bar{H}_2^{-1}(b)),$$

Для полной ясности, мы с этого момента обозначим

$$J \doteq J(H_1(x_1), H_2(x_2)) = g(H_1^{-1}(H_1(x_1)), H_2^{-1}(H_2(x_2))) = g(x_1, x_2),$$

$$\bar{J} \doteq \bar{J}(\bar{H}_1(x_1), \bar{H}_2(x_2)) = g(\bar{H}_1^{-1}(\bar{H}_1(x_1)), \bar{H}_2^{-1}(\bar{H}_2(x_2))) = g(x_1, x_2).$$

По определению

$$h(x_1, x_2) = Jh_1(x_1)h_2(x_2), \quad (2)$$

$$\bar{h}(x_1, x_2) = \bar{J}\bar{h}_1(x_1)\bar{h}_2(x_2). \quad (3)$$

Мы теперь имеем:

$$h_1(x_1) \doteq \rho_{x_1/A}(x_1) = \rho_{x_1}(x_1)P(A/x_1)P(A) = \frac{\alpha\rho_{x_1}(x_1)}{\theta}, \quad (4)$$

$$h_2(x_2) \doteq \rho_{x_2/A}(x_2) = \rho_{x_2}(x_2)P(A/x_2)P(A) = \frac{\beta\rho_{x_2}(x_2)}{\theta}, \quad (5)$$

$$\bar{h}_1(x_1) \doteq \rho_{x_1/\bar{A}}(x_1) = \rho_{x_1}(x_1)P(\bar{A}/x_1)P(\bar{A}) = \frac{(1-\alpha)\rho_{x_1}(x_1)}{1-\theta}, \quad (6)$$

$$\bar{h}_2(x_2) \doteq \rho_{x_2/\bar{A}}(x_2) = \rho_{x_2}(x_2)P(\bar{A}/x_2)P(\bar{A}) = \frac{(1-\alpha)\rho_{x_2}(x_2)}{1-\theta}. \quad (7)$$

Следовательно, из (2),(3)

$$h(x_1, x_2) = J \frac{\alpha\beta\rho_{x_1}(x_1)\rho_{x_2}(x_2)}{\theta^2},$$

$$\bar{h}(x_1, x_2) = \bar{J} \frac{(1-\alpha)(1-\beta)\rho_{x_1}(x_1)\rho_{x_2}(x_2)}{(1-\theta)^2}.$$

Сейчас из (1)

$$P(A/x_1, x_2) = \frac{\frac{J}{\theta} \alpha \beta \rho_{x_1}(x_1) \rho_{x_2}(x_2)}{\frac{J}{\theta} \alpha \beta \rho_{x_1}(x_1) \rho_{x_2}(x_2) + \frac{\bar{J}}{(1-\theta)} (1-\alpha)(1-\beta) \rho_{x_1}(x_1) \rho_{x_2}(x_2)} \quad (8)$$

$$= \frac{\alpha \beta}{\alpha \beta + \frac{\bar{J}}{J} \frac{\theta}{1-\theta} (1-\alpha)(1-\beta)}.$$

Заметим, что в случае условной независимости $J = \bar{J} = 1$ и (8) становится точным решением $\Gamma(\alpha, \beta, \theta) = P(A/x_1, x_2)$.

Глава 2.1.5 Ограничения на функции $J(a, b)$ и $\bar{J}(a, b)$

Мы имеем

$$h_1(x_1) = \int_{-\infty}^{+\infty} J(H_1(x_1), H_2(x_2)) h_1(x_1) h_2(x_2) dx_2. \quad (9)$$

Следовательно

$$1 = \int_{-\infty}^{+\infty} J(H_1(x_1), H_2(x_2)) h_2(x_2) dx_2 = \int_0^1 J(H_1(x_1), H_2(x_2)) dH_2(x_2). \quad (10)$$

Таким образом, мы имеем следующее условие

$$\int_0^1 J(a, b) db = 1, \quad (11)$$

И аналогично

$$\int_0^1 J(a, b) da = 1. \quad (12)$$

Аналогично мы имеем

$$\int_0^1 \bar{J}(a, b) da = 1 \quad (13)$$

$$\int_0^1 \bar{J}(a, b) db = 1.$$

Очевидно

$$J(a, b), \bar{J}(a, b) \geq 0, \quad (14)$$

$$\int_0^1 \int_0^1 J(a, b) da db = \int_0^1 \int_0^1 \bar{J}(a, b) da db = 1. \quad (15)$$

Множество всех решений (11), (12), (13), (14), (15) вместе с (8) определяет множество всей возможной реализаций $P(A/x_1, x_2)$

Дадим пример решения (11), (12) и (14), (15):

Пусть $\rho(x)$ функция такая что $\rho(x) \geq 0$ и $\int_0^1 \rho(x) dx = 1$

Тогда

$$J(a, b) = \begin{cases} \rho(a - b) & , a \geq b \\ \rho(a - b + 1) & , a < b \end{cases}$$

удовлетворяет (11),(12) и (14),(15).

Глава 2.1.6 Определение расстояния

Мы определяем расстояние между предложенной аппроксимацией функции $P(A/x_1, x_2) - \Gamma(\alpha, \beta, \theta)$ и настоящей функцией $P(A/x_1, x_2)$ следующим образом:

$$\|\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2)\| \doteq \iint_{-\infty}^{+\infty} \rho_{X_1 X_2}(x_1, x_2) [\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2)]^2 dx_1 dx_2 .$$

Сейчас мы имеем из (2),(3) и (4),(5), (6),(7)

$$\begin{aligned} \rho_{X_1 X_2}(x_1, x_2) &= \theta h(x_1, x_2) + (1 - \theta) \bar{h}(x_1, x_2) = \theta h_1(x_1) h_2(x_2) + (1 - \theta) \bar{J} h_1(x_1) \bar{h}_2(x_2) = \\ &= [J\alpha\beta\theta + \bar{J}(1 - \alpha)(1 - \beta)(1 - \theta)] \rho_{X_1}(x_1) \rho_{X_2}(x_2) , \end{aligned}$$

$$\begin{aligned} \|\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2)\| &= \\ \iint_{-\infty}^{+\infty} \rho_{X_1}(x_1) \rho_{X_2}(x_2) [J\alpha\beta\theta + \bar{J}(1 - \alpha)(1 - \beta)(1 - \theta)] (\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2))^2 dx_1 dx_2 & \quad (16) \\ = \int_0^1 \int_0^1 \left[\frac{J\alpha\beta}{\theta} + \frac{\bar{J}(1 - \alpha)(1 - \beta)}{(1 - \theta)} \right] (\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2))^2 dF_1(x_1) dF_2(x_2) . \end{aligned}$$

где

$$F_1(x_1) = \int_{-\infty}^{x_1} \rho_{X_1}(z) dz ,$$

$$F_2(x_2) = \int_{-\infty}^{x_2} \rho_{X_2}(z) dz .$$

Глава 2.1.7 Ограничение на основные функции

Мы рассмотрим в дальнейшем все функции с аргументами $1 \geq F_1, F_2 \geq 0$, а не x_1, x_2 . У нас есть шесть функций от F_1, F_2 , которые определяют (16): $J, \bar{J}, H_1, H_2, \alpha, \beta$. Позвольте нам, написать другие функции помощью этих функций и найти ограничения для этих функций.

(i)

$$\alpha = P(A/x_1) = \theta h_1(x_1) / \rho_{X_1}(x_1) = \theta \frac{dH_1}{dx_1} \frac{dF_1}{dx_1} = \theta \frac{dH_1}{dF_1} .$$

Таким же образом

$$\beta = \theta \frac{dH_2}{dF_2}.$$

Мы знаем, что функции H_1, F_1, H_2, F_2 являются суммирующими функциями для функций распределения x_1, x_2 , соответственно. Эти функции - монотонно неубывающие функции и изменяются от 0 до 1 по определению функции распределения. Поэтому, мы можем получить следующие ограничения для функций H_1, H_2 , поскольку функции F_1, F_2

существуют: $H_1(1) = H_2(1) = 1$.
 $H_1(0) = H_2(0) = 0$,

$$0 \leq \alpha = \theta \frac{dH_1}{dF_1}, \beta = \theta \frac{dH_2}{dF_2} \leq 1,$$

$$0 \leq \theta \leq 1,$$

(ii)

$$\begin{aligned} \bar{H}_1(x_1) &= \int_{-\infty}^{x_1} \bar{h}_1(x_1) = \int_{-\infty}^{x_1} \frac{(1-\alpha)\rho_{x_1}(x_1)}{1-\theta} dx_1 = \frac{1}{1-\theta} \int_{-\infty}^{x_1} - \frac{\theta}{1-\theta} \int_{-\infty}^{x_1} \alpha \rho_{x_1}(x_1) \theta dx_1 \\ &= \frac{F_1}{1-\theta} - \frac{\theta}{1-\theta} H_1(x_1). \end{aligned}$$

Таким же образом

$$\bar{H}_2(x_2) = \frac{F_2}{1-\theta} - \frac{\theta}{1-\theta} H_2(x_2),$$

(iii)

$$J(H_1(F_1), H_2(F_2)):$$

$$J(H_1(F_1), H_2(F_2)) \geq 0$$

$$\int_0^1 J(a, b) db = 1$$

$$\int_0^1 J(a, b) da = 1,$$

$$\bar{J}(\bar{H}_1(F_1), \bar{H}_2(F_2)):$$

$$\bar{J}(\bar{H}_1(F_1), \bar{H}_2(F_2)) \geq 0$$

$$\int_0^1 \bar{J}(a, b) db = 1$$

$$\int_0^1 \bar{J}(a, b) da = 1,$$

(iv)

$$P(A/x_1, x_2) = \frac{\frac{J\alpha\beta}{\theta}}{\frac{J\alpha\beta}{\theta} + \frac{\bar{J}(1-\alpha)(1-\beta)}{1-\theta}}. \quad (17)$$

Глава 2.1.8 Оптимизация

Мы находим наилучшее приближение $\Gamma(\alpha, \beta, \theta)$ следующим образом

$$\min_{\Gamma(\alpha, \beta, \theta)} E[|\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2)|] \rightarrow \Gamma(\alpha, \beta, \theta),$$

где ожидаемое значение (или ожидание, или математическое ожидание, или среднее, или первый момент) $E[...]$ взято относительно совместной ФПВ и для возможной реализации:

$J, \bar{J}, \alpha, \beta, H_1, H_2$ для данных F_1 и F_2 .

Ради краткости мы обозначаем:

$$C \doteq \frac{J\alpha\beta}{\theta} + \frac{\bar{J}(1-\alpha)(1-\beta)}{(1-\theta)},$$

$$D \doteq \frac{J\alpha\beta}{\theta}.$$

Затем из (17) и (16)

$$\begin{aligned} & \|\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2)\|^2 = \int_0^1 \int_0^1 C(\Gamma(\alpha, \beta, \theta) - D/C)^2 dF_1 dF_2 \\ & = \int_0^1 \int_0^1 dF_1 dF_2 [D^2 C + \Gamma^2(\alpha, \beta, \theta) C - 2\Gamma(\alpha, \beta, \theta) D]. \end{aligned} \quad (18)$$

Таким образом

$$\begin{aligned} & \min_{\Gamma(\alpha, \beta, \theta)} E[|\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2)|] = \\ & \min_{\Gamma(\alpha, \beta, \theta)} E\left[\int_0^1 \int_0^1 dF_1 dF_2 [D^2 C + \Gamma^2(\alpha, \beta, \theta) C - 2\Gamma(\alpha, \beta, \theta) D]\right] = \\ & \min_{\Gamma(\alpha, \beta, \theta)} E\left[\int_0^1 \int_0^1 dF_1 dF_2 [D^2 C]\right] + \min_{\Gamma(\alpha, \beta, \theta)} E\left[\int_0^1 \int_0^1 dF_1 dF_2 [\Gamma^2(\alpha, \beta, \theta) C - 2\Gamma(\alpha, \beta, \theta) D]\right] = \\ & \text{Const} + \min_{\Gamma(\alpha, \beta, \theta)} E\left[\int_0^1 \int_0^1 dF_1 dF_2 [\Gamma^2(\alpha, \beta, \theta) C - 2\Gamma(\alpha, \beta, \theta) D]\right]. \end{aligned} \quad (19)$$

Остается рассчитать ожидаемую величину в (19)

Мы делаем следующее очевидное предположение

$$\begin{aligned} & \rho_{J, \bar{J}, \alpha, \beta, H_1, H_2 / F_1, F_2} (J, \bar{J}, \alpha, \beta, H_1, H_2 / F_1, F_2) = \\ & \rho_{J/H_1, H_1} (J/H_1, H_2) \rho_{\bar{J}/\bar{H}_1, \bar{H}_1} (\bar{J}/\bar{H}_1, \bar{H}_2) \rho_{\alpha F_1} (\alpha/F_1) \rho_{H_1/\alpha, F_1} (H_1/\alpha, F_1) \rho_{\beta F_2} (\beta/F_2) \rho_{H_2/\beta, F_2} (H_2/\beta, F_2). \end{aligned} \quad (20)$$

8.1 Лемма 1

$$E[J(a, b)] = \int_0^{+\infty} \rho_{J(a,b)/a,b}(J(a, b)/a, b) J(a, b) dJ = 1,$$

$$E[\bar{J}(a, b)] = \int_0^{+\infty} \rho_{\bar{J}(a,b)/a,b}(\bar{J}(a, b)/a, b) \bar{J}(a, b) d\bar{J} = 1.$$

Доказательство:

Позвольте нам рассмотреть функцию: $\rho_{J(a,b)/a,b}$. Функция $J(a, b)$ определена на квадрате $0 \leq a, b \leq 1$. Позвольте нам сделать выборку функции J на этом квадрате его делением его на небольших квадраты (i, j) и определять значение функции J_{ij} на каждом квадрате i, j . Тогда следующие ограничения на функцию $J(***)$ могут быть написаны:

$$J_{ij} \geq 0,$$

$$\frac{1}{N} \sum_{i=1}^N J_{ij} = 1,$$

$$\frac{1}{N} \sum_{j=1}^N J_{ij} = 1.$$

Здесь $i = 1, \dots, N, j = 1, \dots, N$

Все матрицы (J_{ij}) , которые удовлетворяют эти условия, имеют равную вероятность.

Определим функцию плотности вероятности

$$\rho(J_{11}, \dots, J_{ij}, \dots, J_{NN}).$$

Эта функция плотности вероятности должна быть симметричной относительно перестановки строк и столбцов в матрице (J_{ij}) , потому что у плотности распределения имеет равную вероятность для всех матриц (J_{ij}) , которые удовлетворяют вышеупомянутым условиям. Действительно, эти условия также симметричны относительно перестановки строк и столбцов в матрице (J_{ij}) . Из условий симметрии, которые определяют эту функцию (ρ) относительно перестановки строк и столбцов в матрице (J_{ij}) , мы можем заключить, что эта функция (ρ) также не преобразовывается относительно таких перестановок.

Рассмотрим функцию $\rho_{u/ij}(u/ij)$, которая является дискретной версией $\rho_{J(a,b)/a,b}(J(a, b)/a, b)$

:

$$\rho_{u/ij}(u/ij) = \int \dots \int_0^{+\infty} \rho(J_{11}, \dots, J_{nk}, \dots, J_{ij} = u, \dots, J_{NN}) \prod_{(lm) \neq (ij)} dJ_{lm}.$$

Позвольте нам переставлять строки и столбцы (J_{ij}) таким способом, что элемент J_{ij} будет заменен элементом J_{nk} , и функция, $\rho(J_{11}, \dots)$ не будет меняться после этого. Тогда из предыдущего уравнения мы получаем:

$$\rho_{u/ij}(u/ij) = \int \dots \int_0^{+\infty} \rho(J_{11}, \dots, J_{nk} = u, \dots, J_{ij}, \dots, J_{NN}) \prod_{(lm) \neq (nk)} dJ_{lm} = \rho_{u/nk}(u/nk).$$

$$\rho_{u/ij}(u/ij) = \int \dots \int_0^{+\infty} \rho(J_{11}, \dots, J_{nk} = u, \dots, J_{ij}, \dots, J_{NN}) \prod_{(lm) \neq (nk)} dJ_{lm} = \rho_{u/nk}(u/nk).$$

Из этого уравнения мы можем заключить, что $\rho_{u/ij}(u/ij)$ не зависит от ij поэтому $\rho_{J/ab}(J/ab)$ не зависит от ab и $\rho_{J/ab}(J/ab) = \rho_J(J)$,

и

$$E[J(a, b)] = \int_0^{+\infty} \rho_J(J) J dJ = \text{Const}, \quad (21)$$

из

$$\int_0^1 \int_0^1 J(a, b) da db = 1,$$

Мы можем заключить, что

$$\int_0^1 \int_0^1 E[J(a, b)] da db = 1.$$

Таким образом, мы получаем, что $\text{Const} = 1$ в уравнении (21).

8.2 Лемма 2

Функции плотности вероятности α и β не зависят от F_1 и F_2 .

$$\rho_{\alpha/F_1}(\alpha/F_1) = \rho_\alpha(\alpha),$$

$$\rho_{\beta/F_2}(\beta/F_2) = \rho_\beta(\beta).$$

Доказательство:

Позвольте нам сделать выборку функции $\alpha(F_1)$, деля область определения этой функции $F_1, [0, 1]$ на интервалах $1/N, N \gg 1$. Тогда имеем следующие ограничения для $\alpha_k, k = 1, \dots, N, 0 \leq \alpha_k \leq 1$,

$$\frac{1}{N} \sum_{k=1}^N \alpha_k = \int_0^1 \theta dN_1(F_1) dF_1 = \theta.$$

Все столбцы (α_k) , которые удовлетворяют этим условиям имеют равную вероятность. Позвольте рассмотреть соответствующую функцию $\rho(\alpha_1, \dots, \alpha_k, \dots, \alpha_1, \dots, \alpha_N)$. Из условий симметрии, которые определяют эту функцию относительно перестановок $\alpha_k \rightarrow \alpha_1$, функция $\rho(\alpha_1, \dots, \alpha_k, \dots, \alpha_1, \dots, \alpha_N)$ также не преобразовывается относительно таких перестановок. Таким образом, мы можем написать

$$\rho_k(u) = \int_0^1 \rho(\alpha_1, \dots, \alpha_k = u, \dots, \alpha_1, \dots, \alpha_N) \prod_{n \neq k} d\alpha_n = \int_0^1 \rho(\alpha_1, \dots, \alpha_k, \dots, \alpha_1 = u, \dots, \alpha_N) \prod_{n \neq 1} d\alpha_n = \rho_1(u).$$

Из этих уравнений, мы можем заключить, что функция $\rho_{\alpha/F_1}(\alpha/F_1)$ не зависит от F_1 .

$$\rho_{\alpha/F_1}(\alpha/F_1) = \rho_\alpha(\alpha).$$

Из (20) мы получаем

$$\begin{aligned}
& E[\Gamma^2(\alpha, \beta, \theta)C - 2\Gamma(\alpha, \beta, \theta)D] = \\
& \int_0^1 \int_0^1 \rho_\alpha(\alpha) \rho_\beta(\beta) d\alpha d\beta \int_0^1 \int_0^1 \rho_{H_1/\alpha, F_1}(H_1/\alpha, F_1) \rho_{H_2/\beta, F_2}(H_2/\beta, F_2) dH_1 dH_2 \\
& \int_0^\infty \int_0^\infty \rho_J(J) \rho_{\bar{J}}(\bar{J}) [\Gamma^2(\alpha, \beta, \theta) \left[\frac{J\alpha\beta}{\theta} + \frac{\bar{J}(1-\alpha)(1-\beta)}{1-\theta} \right] - 2\Gamma(\alpha, \beta, \theta) \frac{J\alpha\beta}{\theta}] dJ d\bar{J} \\
& = \int_0^1 \int_0^1 \rho_\alpha(\alpha) \rho_\beta(\beta) d\alpha d\beta [\Gamma^2(\alpha, \beta, \theta) \left[\frac{E[J]\alpha\beta}{\theta} + \frac{E[\bar{J}](1-\alpha)(1-\beta)}{1-\theta} \right] - 2\Gamma(\alpha, \beta, \theta) \frac{E[J]\alpha\beta}{\theta}].
\end{aligned}$$

Позвольте нам определить

$$\begin{aligned}
\bar{C} &= \frac{\alpha\beta}{\theta} + \frac{(1-\alpha)(1-\beta)}{1-\theta}, \\
\bar{D} &= \frac{\alpha\beta}{\theta}.
\end{aligned}$$

Согласно Лемме 1, $E[J] = E[\bar{J}] = 1$. Отсюда

$$\begin{aligned}
E[\Gamma^2(\alpha, \beta, \theta)C - 2\Gamma(\alpha, \beta, \theta)D] &= \int_0^1 \int_0^1 [\Gamma^2(\alpha, \beta, \theta)\bar{C} \\
&- 2\Gamma(\alpha, \beta, \theta)\bar{D}] \rho_\alpha(\alpha) \rho_\beta(\beta) d\alpha d\beta.
\end{aligned}$$

Остается найти

$$\min_{\Gamma(\alpha, \beta, \theta)} \int_0^1 \int_0^1 dF_1 dF_2 \int_0^1 \int_0^1 d\alpha d\beta \rho_\alpha(\alpha) \rho_\beta(\beta) [\Gamma^2(\alpha, \beta, \theta)\bar{C} - 2\Gamma(\alpha, \beta, \theta)\bar{D}]. \quad (22)$$

Так как

$\rho_\alpha(\alpha) \rho_\beta(\beta) \geq 0$, если выражение в квадратных скобках минимизируется в каждой точке, тогда весь интеграл в (22) минимизируется. Таким образом, мы можем продолжить следующим образом

$$\frac{\partial}{\partial \Gamma} [\Gamma^2(\alpha, \beta, \theta)\bar{C} - 2\Gamma(\alpha, \beta, \theta)\bar{D}] = 2\Gamma(\alpha, \beta, \theta)\bar{C} - 2\bar{D} = 0.$$

Где оптимум $\Gamma(\alpha, \beta, \theta)$ дается следующим выражением

$$\Gamma_{\text{opt}}(\alpha, \beta, \theta) = \frac{\bar{D}}{\bar{C}} = \frac{\frac{\alpha\beta}{\theta}}{\frac{\alpha\beta}{\theta} + \frac{(1-\alpha)(1-\beta)}{1-\theta}}.$$

Глава 2.1.9 Среднее расстояние между предложенной аппроксимацией функции $P(A/x_1, x_2)$ - $\Gamma(\alpha, \beta, \theta)$ и реальной функцией $P(A/x_1, x_2)$

Среднее расстояние из (18) следующее

$$\text{DIS} = E[|\Gamma(\alpha, \beta, \theta) - P(A/x_1, x_2)|] = \int_0^1 \int_0^1 \rho_\alpha(\alpha) \rho_\beta(\beta) d\alpha d\beta [\Gamma^2(\alpha, \beta, \theta) \bar{C} - 2\Gamma(\alpha, \beta, \theta) \bar{D}] + \text{Const},$$

где Const в этом уравнении определена следующим выражением

$$\text{Const} = E\left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{X_1, X_2}(x_1, x_2) [P(A/x_1, x_2)]^2 dx_1 dx_2\right].$$

Из этого уравнения мы можем найти границы Const. Из $0 \leq P(A/x_1, x_2) \leq 1$ мы можем заключить:

$$\text{Const} \leq E\left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{X_1, X_2}(x_1, x_2) P(A/x_1, x_2) dx_1 dx_2\right] = E[\theta] = \theta.$$

Второе условие следующее

$$\begin{aligned} 0 &\leq E\left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{X_1, X_2}(x_1, x_2) [P(A/x_1, x_2) - \theta]^2 dx_1 dx_2\right] = \\ &E\left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{X_1, X_2}(x_1, x_2) [P(A/x_1, x_2)^2 + \theta^2 - 2P(A/x_1, x_2)\theta] dx_1 dx_2\right] = \\ &E\left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{X_1, X_2}(x_1, x_2) [P(A/x_1, x_2)]^2 dx_1 dx_2\right] - \theta^2. \end{aligned}$$

Таким образом, из этих двух уравнений мы можем заключить

$$\theta^2 \leq \text{Const} \leq \theta.$$

На следующем шаге мы хотели бы найти функцию $\rho_\alpha(\alpha)$ ($\rho_\beta(\beta)$) в уравнении для DIS.

Ограничения на функцию $\alpha(F_1), 0 \leq F_1 \leq 1$ следующие:

$$(i) \int_0^1 \alpha(F_1) dF_1 = \theta,$$

$$(ii) 0 \leq \alpha(F_1) \leq 1.$$

In discrete form (for $N \rightarrow \infty$) we can rewrite $\alpha_{\text{set}} = \{\alpha_1, \alpha_2, \dots, \alpha_N\}$

$$(i) \frac{1}{N} \sum_{i=1}^N \alpha_i = \theta,$$

$$(ii) 0 \leq \alpha_i \leq 1, i = 1, 2, \dots, N.$$

Определим функцию $U(\alpha_{\text{set}})$ следующим образом

$$U(\alpha_{\text{set}}) = \begin{cases} \sum_{i=1}^N \alpha_i & \text{for } 0 \leq \alpha_i \leq 1, i = 1, 2, \dots, N, \\ +\infty & \text{otherwise} \end{cases},$$

$$U(\alpha_{\text{set}}) = \sum_{i=1}^N U_i(\alpha_i),$$

$$U_i(\alpha_i) = \begin{cases} \alpha_i & \text{for } 0 \leq \alpha_i \leq 1 \\ +\infty & \text{otherwise} \end{cases}.$$

Тогда функция, которая является, однородной функцией плотности, учитывая ограничения (i),(ii), будет следующая

$$\rho_{\alpha_{\text{set}}}(\alpha_{\text{set}}) = \frac{1}{C} \delta(U(\alpha_{\text{set}}) - N\theta), \quad (23)$$

где δ -делта-функция Дирака.

Константа C определяется таким образом:

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho_{\alpha_{\text{set}}}(\alpha_{\text{set}}) d\alpha_1 \dots d\alpha_N = 1.$$

Можно доказать, (см. курс "Statistical mechanics" [183]; преобразование от микроканонического к каноническому распределению) что для $N \mapsto \infty$ распределение (23) равно следующему распределению:

$$\rho_{\alpha_{\text{set}}}(\alpha_{\text{set}}) = \frac{1}{Z} e^{-K U(\alpha_{\text{set}})},$$

где Z и K могут быть найдены из уравнений:

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho_{\alpha_{\text{set}}}(\alpha_{\text{set}}) d\alpha_1 \dots d\alpha_N = 1, \quad (24)$$

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} U(\alpha_{\text{set}}) \rho_{\alpha_{\text{set}}}(\alpha_{\text{set}}) d\alpha_1 \dots d\alpha_N = N\theta. \quad (25)$$

Функция $\rho_{\alpha}(\alpha)$ может быть найдена следующим образом

$$\rho_{\alpha}(\alpha) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \rho_{\alpha_{\text{set}}}(\alpha_1, \dots, \alpha_j = \alpha, \dots, \alpha_N) \prod_{i=1, i \neq j}^N d\alpha_i = \frac{1}{D} e^{-K U_j(\alpha_j = \alpha)}, \quad (26)$$

где

$$D^N = Z. \quad (27)$$

Из уравнений (24),(25) мы можем найти

$$\frac{1}{Z} = \left(\frac{K}{1 - e^{-K}} \right)^N, \quad (28)$$

$$\theta = \Lambda(K), \quad (29)$$

где $\Lambda(K)$ -убывающая функция

$$\Lambda(K) = \begin{cases} 1 & \text{for } K = -\infty \\ 0 & \text{for } K = +\infty \\ 1/2 & \text{for } K = 0 \\ \frac{1}{K} - \frac{1}{e^K - 1} & \text{otherwise} \end{cases} .$$

Если K является корнем уравнения (29) мы можем записать из уравнений (26),(27),(28),(29) для функции $\rho_\alpha(\alpha)$:

$$\rho_\alpha(\alpha) = \begin{cases} \begin{cases} \text{For } K = 0 \\ \begin{cases} 1 & \text{for } 0 \leq \alpha \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{cases} \\ \begin{cases} \text{For } K = +\infty \\ \begin{cases} 2\delta(\alpha) & 0 \leq \alpha \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{cases} \\ \begin{cases} \text{For } K = -\infty \\ \begin{cases} 2\delta(\alpha - 1) & 0 \leq \alpha \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{cases} \\ \begin{cases} \text{For } \text{otherwise } K \\ \begin{cases} \frac{1}{D} e^{-K\alpha} & 0 \leq \alpha \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{cases} \end{cases} ,$$

где $2 \int_0^1 \delta(\alpha - 1) = 2 \int_0^1 \delta(\alpha) = 1$ и

$$\frac{1}{D} = \frac{K}{1 - e^{-K}} .$$

Глава 2.1.10 Случай более чем двух переменных A и X .

Пусть A - случайная величина, со значениями на множестве $0, 1, \dots, L$. Предположим, что априорная вероятность $P(A = i)$ известна и обозначим ее θ_i , здесь $i = 1, \dots, L$. Пусть X_1, \dots, X_K случайные величины, со значениями на некотором множестве, например $]-\infty; +\infty[$. Мы имеем следующую информацию : $X_1 = x_1, \dots, X_K = x_K$ (полученную из измерений). Кроме того, у нас есть системы - "классификаторы", которые дают x_1, \dots, x_K : $P(A = i / X_j = x_j) \doteq \alpha_{ij}$.

Мы хотим оценить вероятности $P(A = i / X_1 = x_1, \dots, X_K = x_K)$ в терминах α_{ij} и θ_i . Более определенно мы желаем найти функцию $\Gamma_{\text{opt}, M}(\alpha_{ij}, \theta_i)$, которая в среднем является наилучшим приближением для $P(A = M / x_1, \dots, x_K)$. Тем же самым путем, что в случае двух переменных мы можем найти что $\Gamma_{\text{opt}, M}(\alpha_{ij}, \theta_i)$ определенным уравнением:

$$\Gamma_{\text{opt},M}(\alpha_{ij}, \theta_i) = \frac{(\prod_{j=1}^K \alpha_{Mj}) / \theta_M^{K-1}}{\sum_{i=1}^L (\prod_{j=1}^K \alpha_{ij}) / \theta_i^{K-1}}.$$

Мы имеем очевидное ограничение на α_{ij}, θ_i

$$0 \leq \alpha_{ij} \leq 1$$

$$\sum_{i=1}^L \alpha_{ij} = 1,$$

$$0 \leq \theta_i \leq 1$$

$$\sum_{i=1}^L \theta_i = 1.$$

Глава 2.1.11 Выводы

Мы доказали, что Наивная Байесовская модель дает минимальную среднюю ошибку при равномерном распределении всех возможных корреляций между характерными переменными. Этот результат может объяснить описанную выше мистическую оптимальности Наивной Байесовской модели. Мы также нашли среднюю ошибку, которую Наивная Байесовская модель дает для равномерного распределения всех возможных корреляций.

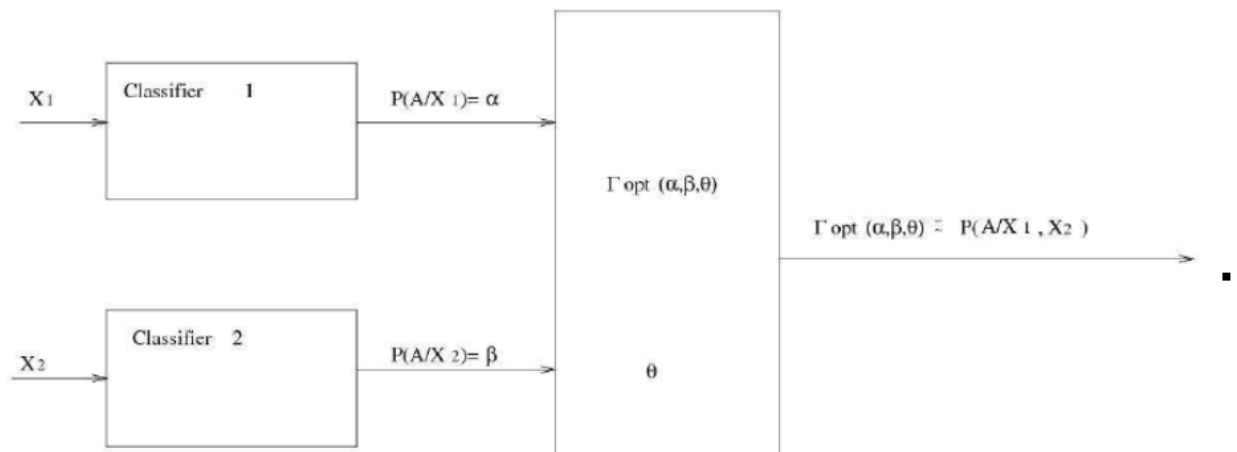


Рис 2 Функция $\Gamma(\alpha, \beta, \theta) : [0, 1]^3 \mapsto [0, 1]$

Глава 2.2 Четвертая группа проблем непредсказуемой динамики, где учет внешнего шума/наблюдателя происходит за счет интуиции наблюдателя.

Для четвертой группы проблем непредсказуемой динамики, где учет внешнего шума от окружения /наблюдателя происходит за счет интуиции наблюдателя, определены границы применимости научных методов предсказания систем и роль интуиции наблюдателя для непредсказуемых систем. Приведены примеры непредсказуемых систем.

Рассмотрено практическое применение непредсказуемых систем в области искусственного интеллекта.

Глава 2.2.1 Введение

Дадим определения *наблюдаемой и идеальной динамик* [6,184], а также объясним необходимость введения наблюдаемой динамики. Идеальной динамикой мы будем называть точные законы квантовой или классической механики. Почему мы назвали их идеальными? Потому что для большинства реальных систем выполняется закон возрастания энтропии или редукция волнового пакета в квантовом случае, противоречащие законом идеальной динамики. Идеальная динамика обратима и в ней происходят возвраты Пуанкаре, чего не наблюдается в необратимой наблюдаемой динамике. Откуда происходит это противоречие между динамиками?

Реальный наблюдатель – это всегда макроскопическая, далекая от термодинамического равновесия система. Он обладает собственной термодинамической стрелой времени, которая существует ограниченное время (до достижения равновесия) и может менять свое направление. Кроме того, существует малое взаимодействие наблюдателя с наблюдаемой системой, которое приводит к синхронизации их термодинамических стрел времени и, в случае квантовой механики редукции волнового пакета.

Наблюдатель описывает наблюдаемую систему в терминах макропараметров и относительно собственной стрелы времени. Именно это и ведет к различию наблюдаемой и идеальной динамики, которая формулируется относительно абстрактного координатного времени в терминах микропараметров.

Нарушения идеальной динамики связаны или с незамкнутостью измеряемых систем (т.е. объясняется влиянием внешней среды или наблюдателя), или невозможностью полного само-измерения и самоанализа для замкнутых и полных физических систем, включающих как внешнюю среду, так и наблюдателя. Что же делать в таких случаях? Реальная система или незамкнута или неполна, т.е. мы не можем использовать физику для предсказания динамики системы? Отнюдь нет!

Очень многие такие системы могут быть описаны уравнениями точной (или вероятностной) динамики, несмотря на незамкнутость или неполноту описания. Мы называем её наблюдаемой динамикой. Большинство уравнений физики - master equations статистической механики, уравнения гидродинамики вязкой жидкости, уравнение Больцмана в термодинамике, закон роста энтропии в изолированных системах являются уравнениями наблюдаемой динамики.

Для того чтобы обладать указанным выше свойством наблюдаемая динамика должна отвечать определенным условиям. Она не может оперировать полным набором микропеременных. В наблюдаемой динамике мы определяем лишь много меньшее число макропеременных, которые являются некими функциями микропеременных. Это делает ее намного устойчивее по отношению к ошибкам в задании начальных условий и шуму. Действительно, изменение микросостояния не приводит неизбежно к изменению макросостояния, поскольку одному макросостоянию отвечает большой набор микросостояний. Для газа макропеременными являются, например, плотность, давление, температура и энтропия. Микропеременными же являются скорости и координаты всех его молекул.

Как из идеальной динамики строится наблюдаемая динамика? Они получаются или введением в идеальные уравнения малого, но конечного внешнего шума, или же введением погрешностей начального состояния. Погрешности и/или шумы должны быть достаточно большими, чтобы нарушить ненаблюдаемую реально обратимость движения

или возвраты Пуанкаре. С другой стороны они должны быть достаточно малы, чтобы не влиять на протекание реально наблюдаемых процессов с ростом энтропии.

Для полной физической системы, включающей наблюдателя, наблюдаемую систему и окружающую среду наблюдаемая динамика не фальсифицируема в смысле Поппера [185] (при условии верности идеальной динамики). Т.е. разницу между Идеальной и Наблюдаемой Динамикой в этом случае невозможно наблюдать в эксперименте.

Однако, вполне возможны и случаи, когда введение наблюдаемой динамики невозможно и система остается все-таки непредсказуемой, вследствие незамкнутости системы или неполноты описания. Это случай *непредсказуемой динамики* [10, 186-192], обсуждаемой здесь.

Глава 2.2.2 Непредсказуемая динамика.

Введем понятие *синергетические модели* [193]. Будем называть таковыми простые физические или математические системы, иллюстрирующие в простой форме некие действительные или предполагаемые свойства непредсказуемых и сложных (в том числе живых) систем.

Непредсказуемые системы, именно вследствие причин своей непредсказуемости крайне неустойчивы к внешнему наблюдению и тепловому шуму. Чтобы их поведение не превратилось в полностью хаотическое и случайное, они должны иметь механизмы защиты от внешнего влияния.

Поэтому для нас важно создание в первую очередь синергетических моделей систем, способных противостоять шуму (декогеренции в квантовой механике). Они сохраняют свои внутренние корреляции (квантовые или классические), приводящие к обратимости движения или возвратам Пуанкаре. Также они могут сохранять свои корреляции с окружающим миром.

Существуют три метода такой защиты:

- 1) Пассивный метод – создание неких «стенок» или панцирей непроницаемых для шума. Можно также держать такие системы при очень низких температурах. Примером могут служить многие модели современных квантовых компьютеров.
- 2) Активный метод, обратный пассивному – подобно диссипативным или живым системам, они сохраняют своё неравновесное состояние благодаря активному взаимодействию и обмену энергией и веществом с окружением (метаболизмом). Думается, что будущие модели квантовых компьютеров должны браться из этой области.
- 3) Когда корреляции охватывают ВСЮ Вселенную. Внешний источник шума здесь просто отсутствует. Источник корреляций Вселенной в том, что Вселенная произошла из малой области и низкоэнтропийного состояния путем Большого Взрыва. Назовем это явление глобальными корреляциями. Иногда это образно называют «голографическая модель Вселенной»

Следует отметить следующие обстоятельства:

- 1) Многие сложные системы в своем развитии проходят динамические точки бифуркации – когда существуют несколько альтернативных путей развития и выбор конкретного из них зависит от малейших изменений состояния системы в точке динамической бифуркации [194,195]. Тут даже слабые (и сохраненные, указанными выше путями) корреляции могут оказать огромное влияние. Наличие подобных корреляций ограничивает предсказательную силу науки, но отнюдь не

- ограничивает нашу личную интуицию. Поскольку мы являемся неотделимой частью нашего мира, то мы вполне способны на субъективном уровне «ощущать» эти корреляции, недоступные научному предсказанию (Но, ни в кое мере не противоречащие самим законам физики!)
- 2) В описываемых ненаблюдаемых системах часто наблюдается уменьшение энтропии или поддерживается очень низкоэнтропийное состояние. Это не противоречит второму закону возрастания энтропии. Действительно, как их пассивная, так и активная защита требуют огромных затрат негэнтропии, которая черпается из окружения, поэтому суммарная энтропия системы и окружения только растет. Закон возрастания энтропии остается незыблемым для «большой» системы (наблюдаемая система + окружение + наблюдатель), хотя неверен для самой наблюдаемой системы. Уменьшения энтропии в большой системе согласно уравнениям идеальной динамики происходят (например, возвраты Пуанкаре в замкнутой системе с ограниченным объемом), но являются ненаблюдаемыми [6,184]. Поэтому они могут просто игнорироваться.
 - 3) Существование многих непредсказуемых систем сопровождается уменьшением энтропии (Это не противоречит росту энтропии согласно второму закону термодинамики как это объяснено выше в третьем пункте). Таким образом, существование таких систем подчиняется обобщенному принципу Ле-Шателье — Брауна: система препятствует любому изменению своего состояния, вызванному как внешним воздействием, так и внутренними процессами, или, иными словами, — любое изменение состояния системы, вызванное как внешними, так и внутренними причинами, порождает в системе процессы, направленные на то, чтобы уменьшить это изменение. В данном случае рост энтропии порождает системы, ведущие к ее уменьшению.
 - 4) Часто находит подтверждение принцип максимума производства энтропии (Maximum entropy production principle - MaxEPP) [196]. Согласно этому принципу неравновесная система стремится к состоянию, при котором рост энтропии в системе будет максимальным. Несмотря на кажущееся противоречие, MaxEPP не противоречит открытому Пригожину для линейных неравновесных систем принципу минимума производства энтропии (MinEPP) [196]. Это абсолютно разные вариационные принципы, в которых хотя и ищется экстремум одной и той же функции – производства энтропии, но при этом используются различные ограничения и различные параметры варьирования. Эти принципы не нужно противопоставлять, так как они применимы к различным этапам эволюции неравновесной системы. MaxEPP означает, что диссипативные непредсказуемые системы (в том числе живые системы), находясь в замкнутых системах с ограниченным объемом, приближают наступление их термодинамического равновесия. Это значит, что они сокращают и время возврата Пуанкаре, т.е. способствуют более быстрому возврату в низкоэнтропийное состояние. Это опять соответствует обобщенному принципу Ле-Шателье — Брауна: рост энтропии порождает системы, ведущие к ее уменьшению. Из всего вышеизложенного можно сделать очень интересный вывод: *глобальной «целью» диссипативных систем (в том числе и живых систем) является (а) минимизация их собственной энтропии (б) стимуляция глобальной полной системы к скорейшему возврату Пуанкаре в исходное низкоэнтропийное состояние.*
 - 5) Глобальные корреляции в общем случае «растекаются» по замкнутой системе с ограниченным объемом и ведут лишь к ненаблюдаемому возврату Пуанкаре [6,184]. Однако при наличии объектов с локальными корреляциями, глобальные корреляции могут проявляться в корреляции между такими объектами друг с другом и окружающим миром. Таким образом, наличие локальных корреляций

позволяет сделать глобальные корреляции наблюдаемыми, предотвращая их полное «растекание» по системе.

- 6) Верное определение термодинамической макроскопической энтропии сама по себе очень трудная задача для сложных физических систем в отсутствие локального равновесия [197]
- 7) Следует отметить очень важное обстоятельство. Неустойчивые корреляции существуют не только в квантовой, но и в классической механике. Следовательно, подобные модели не должны носить только квантовый характер. Они могут быть и классическими! Очень часто ошибочно считается, что только квантовая механика может описывать подобные явления [198,199]. Это не так [200-202]. Введение «руками» малого, но конечного взаимодействия при классическом измерении и малой погрешностей начального состояния стирает разницу между свойствами квантовой и классической механики (при наличии неустойчивых корреляций микросостояний).

Глава 2.2.3 Синергетические модели локальных корреляций

Приведем примеры синергетических моделей непредсказуемых систем, использующих пассивный или активный метод защиты от шума.

- 1) Имеются исключительные случаи, для которых не происходит синхронизация стрел времени [203,204].
- 2) Точки фазовых переходов или точки бифуркаций. В этих точках макроскопическая система, описываемая наблюдаемой динамикой, в процессе эволюции во времени или в процессе изменения какого-либо внешнего параметра может перейти не в одно, а в несколько различных макроскопических состояний. То есть, в этих точках наблюдаемая динамика теряет свою однозначность. В этих точках возникают огромные макроскопические флуктуации, и использование макропараметров не ведет к предсказуемости системы. Эволюция становится непредсказуемой, т.е. возникает непредсказуемая динамика.
- 3) Возьмем квантовую микроскопическую или мезоскопическую систему, описываемую идеальной динамикой, изолированную от декогеренции. Ее динамика зависит от неконтролируемых микроскопических *квантовых корреляций*. Эти корреляции очень неустойчивы и вследствие декогеренции (т.е. запутывания с окружением или наблюдателем) исчезают. Пусть некий первый наблюдатель фиксирует лишь начальное и конечное состояние системы. В промежутке времени между ними система полностью или почти изолирована от окружения или этого наблюдателя. В таком случае эти микроскопические корреляции не исчезают и влияют на динамику. Рассмотрим другого внешнего наблюдателя, не знающего начального состояния системы. В отличие от первого наблюдателя, знающего начального состояния системы, поведение системы для второго наблюдателя становится непредсказуемым! Т.е. с точки зрения такого наблюдателя возникает непредсказуемая динамика. В квантовой области примерами таких систем являются *квантовые компьютеры* и *квантовые криптографические передающие системы* [205,206].

Квантовые компьютеры имеют не только свойство непредсказуемости для наблюдателя, не информированного об их состоянии при запуске вычислений. Другим важным свойством является их высокая параллельность вычисления. Оно достигается за счет того, что начальное состояние является суперпозицией многих возможных начальных состояний «квантовых битов информации». За счет линейности уравнений квантовой механики эта суперпозиция сохраняется и «обработка» всех состояний, входящих в суперпозицию, происходит одновременно (параллельно). Эта

параллельность приводит к тому, что многие задачи, которые обычный компьютер решает очень медленно из-за того, что рассматривает все случаи последовательно, квантовый решает очень быстро. С этим свойством и связаны надежды на практическую пользу квантовых компьютеров.

Квантовые криптографические передающие системы используют в первую очередь свойство своей ненаблюдаемости «передаваемых квантовых сообщений» для внешнего наблюдателя, не информированного об их состоянии при начале передачи. Любая попытка прочесть передаваемое сообщение приводит к его взаимодействию с этим наблюдателем и, следовательно, «разрушению» передаваемого сообщения и невозможности прочесть это сообщение. Таким образом, перехват сообщений оказывается **в принципе невозможным** по законам физики.

- 4) Следует особо отметить, что, вопреки широко распространенному заблуждению, как квантовые компьютеры, так и квантовая криптография [205,206] имеют классические аналоги. Действительно, в классических системах в отличие от квантовых систем измерение можно провести абсолютно точно, не искажая измеряемое состояние. Однако и в классических хаотических системах имеются неконтролируемые и неустойчивые микроскопические дополнительные корреляции, обеспечивающие обратимость и возвраты Пуанкаре системы. Введём «руками» конечное, но малое взаимодействие в классическое измерение или конечную погрешность в начальные условия, которые в реальных ситуациях, и на самом деле, всегда существуют. Они стирают разницу между классической и квантовой системой. В реальных системах всегда присутствует малый внешний шум, выполняющий эту роль. Изолируя хаотическую классическую систему от этого шума, мы получаем классические аналоги изолированных квантовых устройств с квантовыми корреляциями.

Существуют синергетические модели классических компьютеров, обеспечивающих, подобно квантовым компьютерам, невероятную параллельность вычислений [200].

Рассмотрим классический аналог квантового компьютера в виде фрактального компьютера [207] (Рис. 1):

«Компьютеры конструируются таким образом, чтобы исключить влияние любого физического процесса, протекающего на уровне, масштабы которого меньше определенного порога. Алгоритм должен быть защищен от «ошибок». Возможно, наш мозг структурирован так же, но, возможно, что и не так, считает Пенроуз. Не исключено, что физические явления, протекающие на атомарном уровне, играют важную роль в процессе нашего мышления. Что касается молекулярного уровня, то это определено так. Достаточно лишь вспомнить о влиянии молекул нейротрансмиттеров на поведение нервных клеток. Более того, хорошо известно, что природа обладает свойством использовать физические возможности, ставя их на службу для реализации биологических операций. Если физические структуры простираются до определенного уровня, то имеем ли мы какие-то априорные основания полагать, что мозг не может воспользоваться явлениями, протекающими на этом уровне?

А что, если мозг способен пользоваться всеми уровнями структуры в бесконечно структурированной Вселенной? Чтобы продемонстрировать в самой грубой форме потенциальные возможности бесконечного мозга, я построил схему бесконечного компьютера, который пользуется в своей работе структурами всех уровней. В целях удобства демонстрации я буду условно считать, что во всех масштабах физические структуры обладают классическими свойствами.

Мой бесконечный компьютер по существу представляет собой квадрат, содержащий два прямоугольника и два других квадрата меньшего размера. Входная линия подведена слева к большому квадрату, и войдя в него, она сразу идет к первому прямоугольнику.

Последний представляет собой устройство обработки сигналов, которое я назову модулем подстанции. От модуля подстанции отходит по одной линии к каждому из меньших квадратов, а также к другому прямоугольнику, который мы будем ниже называть модулем сообщения.

Структура этого компьютера обладает свойством бесконечной регрессии. Каждый из меньших квадратов является точной копией большего квадрата, но имеет вдвое меньшие размеры. Когда сигнал распространяется по проводникам и модулям вдвое меньшего размера, ему требуется вдвое меньше времени для прохождения соответствующего пути, поэтому модули подстанции и сообщения срабатывают в два раза быстрее соответствующих модулей внешнего уровня.

Бесконечный компьютер решает знаменитую задачу о преобразовании слов, изобретенную математиком Акселем Туэ. В этой задаче даны два слова и словарь допустимых подстановок. Можно ли, пользуясь только допустимыми подстановками, получить второе слово из первого? Рассмотрим задачу на следующем примере: предположим, первое «слово» представлено цепочкой 100101110, а второе - 01011101110. Возможно ли прийти от первого слова ко второму, пользуясь подстановками 010 вместо 110, 10 вместо 111 и 100 вместо 001? Пример был выбран совершенно произвольно, и я специально не буду пытаться решить его.

Может случиться так, что ни одна последовательность подстановок не приведет к преобразованию первой цепочки во вторую. В то же время такая последовательность существует. В процессе применения подстановок могут возникнуть очень длинные промежуточные слова. В этом и заключается проблема. Так же, как и в отношении некоторых точек множества Мандельброта, мы по существу не можем прийти к ответу. Не существует алгоритма решения этой задачи, поскольку любой алгоритм (по определению) должен когда-то остановиться. Опасность состоит в том, что алгоритм может остановиться прежде, чем будет получен ответ. По этой причине о задаче Туэ говорят как о неразрешимой. Ни одна компьютерная программа, даже в принципе, не сможет решить всех частных случаев этой задачи.

Введем условие задачи в бесконечный компьютер. Слово, к которому нужно прийти, вводится в компьютер через его главную входную линию. За четверть секунды оно попадает в первый подстановочный модуль. Затем слово передается этим подстановочным модулем двум другим подстановочным модулям следующего уровня.

Однако теперь передача данных потребует уже только 1/8 секунды. Пересылки на следующие уровни потребуют соответственно 1/16, затем 1/32 с и т.д. Время, которое потребуются, чтобы загрузить все подстановочные модули вторым словом, составит, таким образом, половину секунды.

Потом в ходе аналогичного процесса, протекающего с той же скоростью, в компьютер вводятся три (или сколько угодно) подстановочные формулы. Однако на этот раз различные подстановочные модули на разных уровнях запрограммированы так, чтобы воспринимать только определенные подстановки из вводимой последовательности. Каждый подстановочный модуль, согласно содержащейся в нем программе, должен всегда пытаться проводить определенную подстановку в определенной позиции слова, которое поступает в него из внешнего уровня. Подробное описание схемы распределения подстановок и позиций в слове по различным модулям, наверное, утомило бы читателя, поэтому я опускаю его. Однако это не должно помешать тем, кто любит анализировать бесконечные процессы, представить себе, каким образом можно все устроить.

Вычисление начинается, когда мы посылаем первое слово в компьютер. Первый подстановочный модуль пытается провести свою подстановку на отведенной для нее позиции введенного слова. Если подстановка не может быть выполнена в требуемом месте, слово передается следующему подстановочному модулю, принадлежащему более низкому уровню машины; если ему удастся провести подстановку, он посылает

преобразованное слово в квадрат более высокого уровня. Если подстановка оказывается успешной и полученное в результате слово совпадает со вторым заданным словом, хранящимся в памяти подстановочного модуля, то последний посылает сигнал модулю сообщения: «успешно».

Каждый квадрат каждого уровня работает в соответствии с только что описанным алгоритмом. Как уже говорилось выше, всегда можно распределить подстановки (и позиции, в которых их следует применять) по бесконечному компьютеру таким образом, чтобы задача преобразования слов успешно решалась. Ответ будет выдан не позднее чем через одну секунду: половина секунды потребуется, чтобы провести вычисления на всем пути до бесконечно малых модулей, и половина секунды на то, чтобы сообщение «успешно» достигло главной выходной линии компьютера. Если подходящей последовательности подстановок не существует, то отсутствие сообщения по истечении одной секунды можно принять как ответ «нет». Читателям, возможно, будет интересно поразмыслить над бесконечным компьютером, воспользовавшись при этом многочисленными (а может быть, и бесконечным числом) структурами своего собственного мозга. »

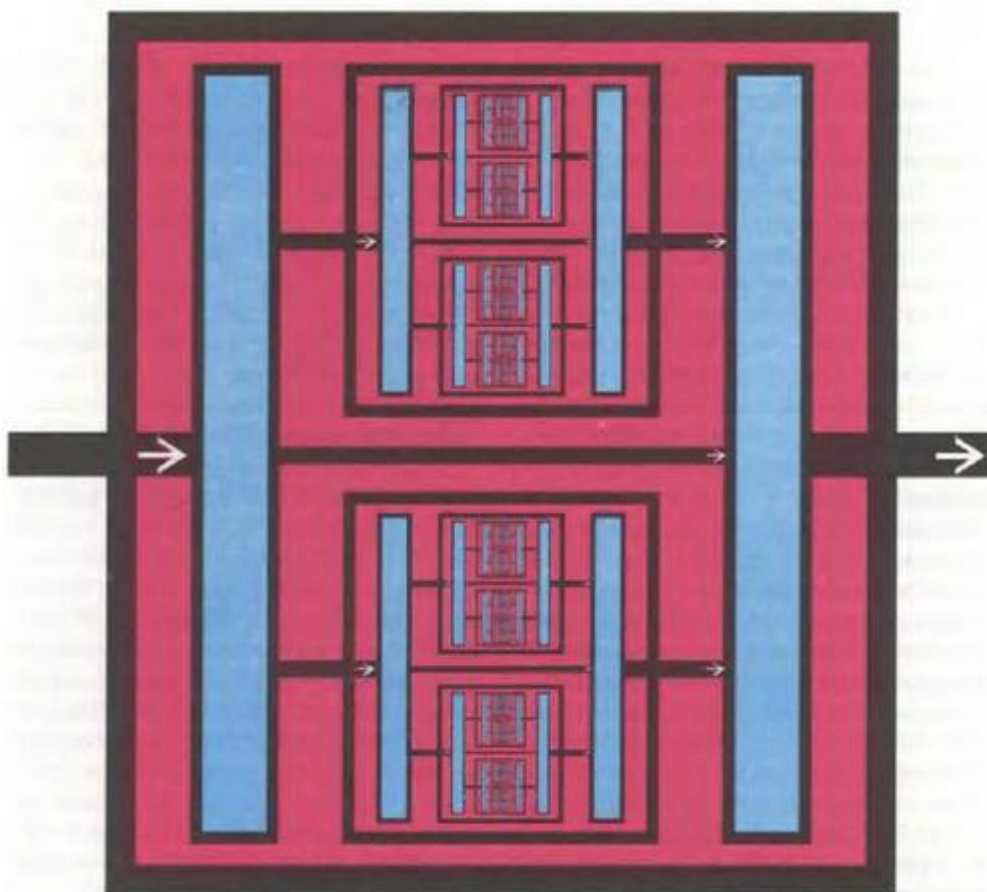


Рис. 1 Фрактальный компьютер решает задачу Туэ о преобразовании слов

Аналогом квантовых компьютеров являются и *молекулярные компьютеры* [202]. Большое количество молекул обеспечивает параллельность вычислений. Неконтролируемые и неустойчивые микроскопические дополнительные корреляции, обеспечивающие обратимость и возвраты системы, делают динамику неопределенной для наблюдателя, не информированного о состоянии компьютера в момент старта. Малое, но конечное взаимодействия при наблюдении приводит к тому, что

наблюдатель нарушит нормальный запланированный ход вычислений при попытке чересчур точно померить координаты и скорости молекул, чтобы предсказать результат работы компьютера.

Аналогичные аргументы могут быть использованы для создания классических криптографических передающих систем, использующих явление классических неустойчивых микроскопических дополнительных корреляции. Неустранимое малое взаимодействия с перехватчиком сообщений разрушает эти корреляции. Тем самым оно делает ненаблюдаемый перехват принципиально невозможным также и в классическом случае.

- 5) Сохранение неустойчивых микроскопических корреляций может быть обеспечено не только за счет пассивной изоляции от внешней среды и наблюдателя, но и за счет динамического, компенсирующего помехи механизма. Это происходит в так называемых физических стационарных системах, в которых равновесие системы поддерживается за счет непрерывного потока энергии или вещества через систему. Примером могут служить микромазеры [208] – маленькие и хорошо проводящие полости с электромагнитным излучением внутри. Размер полостей настолько мал, что излучение уже необходимо описывать квантово. Оно постепенно затухает из-за взаимодействия со стенками. Эту систему оптимально описывать матрицей плотности в базисе состояний, соответствующих различным собственным энергиям системы. Этот базис наиболее устойчив к внешним шумам для любой системы близкой к термодинамическому равновесию и, следовательно, наиболее подходит для наблюдаемой динамики. Микроскопические корреляции соответствуют недиагональным элементам матрицы плотности и стремятся к нулю много быстрее, чем диагональные элементы при затухании излучения. (Иными словами, время декогеренции много меньше времени релаксации.) Однако пропускание через микромазер пучка возбужденных частиц приводит к сильному замедлению затухания недиагональных элементов матрицы плотности (иными словами микрокорреляций) и отличному от нуля стационарному излучению.

Также в теории квантовых компьютеров разработаны методы активной защиты квантовых корреляций от декогеренции, способные поддерживать их сколь угодно долго, повторяя циклы активной квантовой коррекции ошибок (QUANTUM ERROR CORRECTION). Повторение кода в квантовой информации не возможно из-за теоремы о невозможности клонирования. Петер Шор первый нашел метод квантовой корректировки ошибок, перенося информацию с одного кубита на сильно-перепутанное состояние девяти кубитов [209].

- б) В физике обычно макросостояние рассматривается как некая пассивная функция микросостояния. Однако можно рассмотреть случай, когда система сама наблюдает (измеряет) как свое макросостояние, так и макросостояние окружения, записывая результат наблюдения (измерения) в микроскопическую «память». Таким путем образуется обратная связь через макросостояния на микросостояние.

Примером таких очень сложных стационарных систем являются *живые системы*. Они находятся в состоянии очень далеко от термодинамического равновесия и крайне сложны. Они упорядочены, хоть эта упорядоченность сильно отличается от периодичности неживого кристалла. Низкоэнтропийное неравновесное состояние живого поддерживается за счет роста энтропии в окружении¹. Неравновесное состояние поддерживается за счет метаболизма – непрерывного потока вещества и энергии через живой организм. С другой стороны само это неравновесное состояние является катализатором метаболического процесса, т.е. создает и поддерживает его на необходимом уровне. Поскольку состояние живых систем является сильно неравновесным, оно может поддерживать и существующие неустойчивые корреляции,

¹ Так, например, растет энтропия Солнца, служащего источником энергии для жизни на Земле.

препятствуя процессу декогеренции и внешнего шума. Эти корреляции могут быть как между частями самой живой системы, так и между живой системой и другими (живыми или неживыми) системами. Если это происходит, то динамику живой системы можно отнести к непредсказуемой динамике. Несомненные успехи молекулярной биологии позволяют предсказать и описать многие черты динамики живых систем. Но нет никаких фактов, свидетельствующих, что она будет способна полностью описать всю сложность процессов в живой системе, даже с учетом ее дальнейших достижений.

Довольно трудно проанализировать реальные живые системы в рамках концепций идеальной, наблюдаемой и непредсказуемой динамик из-за их огромной сложности. Но возможно построить гораздо менее сложные математические модели. Это, например, неравновесные стационарные системы с метаболизмом. Это позволит понять возможную роль всех трех динамик для таких систем. Эти модели могут быть как квантовыми [198,199, 210-213], так и классическими [200-202].

7) Описанными выше случаями не описывается все многообразие непредсказуемых динамик. Нахождение точных условий, при которых идеальная динамика переходит в наблюдаемую и непредсказуемую динамику – еще полностью не решенная задача для математики и физики. Также такой еще полностью не решенной проблемой (и, по-видимому, связанной с предыдущей задачей) является роль этих трех динамик в сложных стационарных системах. Решение этих проблем позволит глубже понять физические принципы, лежащие в основе жизни.

Глава 2.2.4 Синергетические модели глобальных корреляций, охватывающих всю вселенную.

С помощью синергетических «игрушечных» моделей можно понять синхронистичность² (одновременность) причинно не связанных процессов [214], а также явление глобальных корреляций.

Глобальные корреляции Вселенной и определение живых систем, как систем, способствующих сохранению корреляций в противовес внешнему шуму, хорошо объясняет загадочное молчание КОСМОСА, т.е. отсутствие сигналов от других разумных миров. Вселенная произошла из единого центра (Большой Взрыв) и все ее части коррелированы, жизнь лишь поддерживает эти корреляции в локальном масштабе и существует на их основе. Поэтому процессы возникновения жизни в различных частях скоррелированы и находятся на одном уровне развития, т.е. сверхцивилизаций, способных достичь Земли, пока просто нет.

2.2.4.1 Системы с «обострением» (blow up)

Примером являются нестационарные системы с «обострением» (blow up) [195,215-218], рассмотренные школой Курдюмова. В этих процессах определяется некая функция на

² Валерий Исаков, кандидат механико-математических наук и лидер небольшой группы исследователей аномальных явлений рассказал газете "Комсомольская правда" о существовании некой статистической аномалии. После того как Исакову и его группе не удалось получить данные по отказам от полетов у российских авиакомпаний, ученые воспользовались западной статистикой. Как выяснилось, за последние 20 лет от рейсов, закончившихся катастрофами, отказывались на 18% больше пассажиров, чем от благополучных.

<http://kp.ru/daily/25707/908213/>; http://newsru.co.il/world/23jun2011/isakov_606.html

плоскости. Ее динамика описывается нелинейными уравнениями, подобными уравнению горения.

$$\partial \rho / \partial t = f(\rho) + \partial / \partial r (H(\rho) \partial \rho / \partial r), \quad (\text{II})$$

где ρ - плотность, $N = \int \rho dr$, r – пространственная координата, t – временная координата, $f(\rho)$, $H(\rho)$ – нелинейные связи.

$$f(\rho) \rightarrow \rho^\beta, H(\rho) \rightarrow \rho^\sigma,$$

Эти уравнения имеют набор динамических решений, называемых решения с «обострением». Доказано существование явления локализации процессов в виде структур (при $\beta > \sigma + 1$), образование дискретного их спектра с разным числом простых структур (с одиночными максимумами разной интенсивности), объединенных в несколько типов сложных структур, которые имеют различные пространственные формы и несколько максимумов. Показано, что нелинейная диссипативная среда потенциально содержит в себе спектр таких различных структур-аттракторов. Пусть (r, φ) – полярные координаты.

$$\rho(r, \varphi, t) = g(t) \Theta_i(\xi, \varphi), \quad \xi = \frac{r}{\psi(t)}, \quad 1 < i < N$$

$$g(t) = \left(1 - \frac{t}{\tau}\right)^{\frac{1}{\beta-1}}, \quad \psi(t) = \left(1 - \frac{t}{\tau}\right)^{\frac{\beta-\sigma-1}{\beta-1}}$$

Число собственных функций:

$$N = \frac{\beta-1}{\beta-\sigma-1}$$

Для этих решений (Рис.2) значение функции может стремиться к бесконечности за конечное время τ . Интересно, что функция достигает бесконечности в максимумах в один и тот же момент времени, то есть синхронно. По мере приближения ко времени τ решение «сжимается», максимумы «обостряются» и движутся к общему центру. В момент 0.9τ система становится неустойчивой и разрушается флуктуациями начальных условий. При высокой корреляции в начальном условии можно уменьшать эти флуктуации до сколь угодно малой величины.

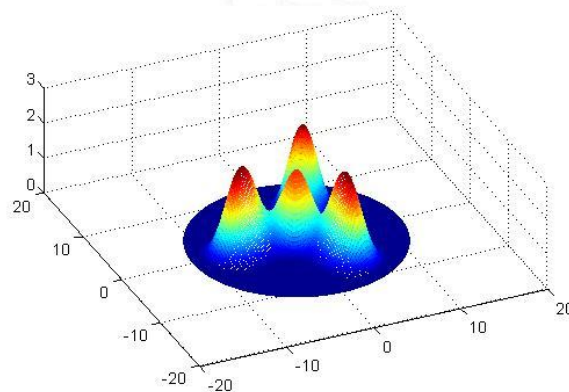


Рис. 2 из [219]. Один из аттракторов уравнения горения в виде решения с «обострением».

С помощью таких моделей иллюстрируют рост населения (или уровня технического развития цивилизаций) в мегаполисах нашей планеты [218]. Точки максимума функции – это мегаполисы, а плотность населения – это значение самой функции.

Можно распространить эту модель на всю Вселенную. Тогда точки бесконечного роста – это цивилизации, а плотность населения цивилизаций (или уровня технического развития цивилизаций) – это значение самой функции. Для этого усложним модель. Пусть в момент, когда процесс начинает выходить на растущее асимптотическое решение происходит очень быстрое расширение («инфляция») плоскости, в которой протекает процесс с «обострением». Тем не менее, процессы достижения бесконечности остаются синхронными и описываются уравнением того же типа (лишь с измененным масштабом), несмотря на то, что максимумы уже разделены большим расстоянием.

Этой более сложной моделью можно качественно объяснить синхронность развития процессов в очень далеких частях нашей резко расширившейся Вселенной в результате «инфляция» после Большого Взрыва. Высокая степень глобальных корреляций уменьшает флуктуации, ведущие к распаду структуры решения. Эти глобальные корреляции моделируют взаимосвязанность частей нашей Вселенной.

Процессы с «обострением» появляются с необходимой полнотой и сложностью лишь при некотором узком наборе коэффициентов уравнения горения ($N \gg 1$, $\beta > \sigma + 1$, $\beta \approx \sigma + 1$ – это необходимо для возникновения структуры с большим числом максимумов и их медленному сближению к центру). Это позволяет провести аналогию с «антропным принципом» [220]. Антропный принцип утверждает, что фундаментальные постоянные Вселенной имеют именно такие значения, чтобы в итоге могла возникнуть именно наша наблюдаемая Вселенная с «антропными» существами, способными ее наблюдать.

Следует обратить внимание ещё на одно обстоятельство. Чтобы упорядоченное состояние в модели не распадалось при $t=0.9\tau$, а прожило как можно дольше, требуется тонкая настройка *не только параметров модели, но и начального состояния*. Это нужно, чтобы возникающие из него флуктуации не разрушали упорядоченность как можно дольше. И это наличие этого редкого эксклюзивного состояния также может быть объяснено антропным принципом.

2.2.4.2 «Клеточная» модель Вселенной.

Также интересно проиллюстрировать сложные процессы с помощью «клеточной» модели. Хорошей базой служат дискретная модель Хопфилда [221,222]. Эта модель может интерпретироваться как нейронная сеть с обратной связью или как спиновая решетка (спиновое стекло) с неодинаковыми взаимодействиями между спинами. Подобная система используется для целей распознавания образов.

Эту систему можно описать как квадратную двумерную решетку ячеек $N \times N$, которые могут быть либо черными, либо белыми ($S_i = \pm 1$). Коэффициенты линейного взаимодействия между ячейками J_{ij} неравны для разных пар ячеек. Их можно выбрать так, что в процессе дискретной эволюции подавляющее большинство начальных состояний переходит в одно из возможных конечных состояний, из заранее заданного набора состояний (аттракторов).

$$S_i(t+1) = \text{sign} \left[\sum_{j=1}^N J_{ij} S_j(t) \right], \quad 1 \leq i \leq N$$

$$J_{ij} = J_{ji}, \quad J_{kk} = 0 \quad 1 \leq i, j, k \leq N$$

Аттракторы соответствуют минимуму энергии:

$$E = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} S_i S_j$$

Пусть аттракторами решетки выбраны буквы А или В.

Существуют такие начальные неустойчивые состояния, которые отличаются лишь на одну ячейку (*критический элемент*). При этом одно из них имеет в качестве аттрактора состояние А, а другое – состояние В. Подобные неустойчивые начальные состояния хорошо иллюстрирует свойство *глобальной неустойчивости* сложных систем. Эта неустойчивость присуща всей системе в целом, а не какой-то ее части. Лишь некий внешний наблюдатель может привести к изменению значения критического элемента и изменить эволюцию системы. Внутренняя динамика самой системы сделать это не может. *Глобальная корреляция* между ячейками неустойчивого начального состояния определяет к какому именно аттрактору эволюционирует эта решетка (либо А, либо В).

Можно несколько усложнить модель. Пусть каждая ячейка в описанной выше решетке сама является аналогичной подрешеткой. Определим эволюцию такой составной решетки, идущей в два этапа.

На первом этапе крупные ячейки не взаимодействуют, взаимодействие есть лишь в подрешетках, которое идет по тому же образцу, что и в описанной выше простой одноэтапной модели. Коэффициенты линейного взаимодействия между ячейками выбраны так, чтобы аттракторами, как и ранее, были буква А или В. Начальные состояния всех подрешеток можно выбрать неустойчивыми, содержащими критический элемент. Итоговое состояние А подрешетки будем воспринимать как черную ячейку для крупной решетки, а состояние В подрешетки будем воспринимать как белую ячейку.

Второй этап эволюции определяется как эволюция уже этой крупной решетки по тому же образцу, что и в описанной выше простой одноэтапной модели, с получившимся выше начальным состоянием. Это начальное состояние, возникающее на первом этапе, тоже является неустойчивым, содержащее критический элемент. В конце эволюции каждой черной ячейке присвоим состояние А подрешетки, каждой белой ячейке присвоим состояние В подрешетки.

Начальное состояние решетки до начала двухэтапного процесса всегда можно выбрать так, чтобы после него итоговым состоянием крупнозернистой решетки была буква А. Состоянию каждой ее крупной ячейки тоже соответствует буква А. Назовем это состояние «А-А». Тогда наличие именно такого, а не иного финального состояния можно объяснить:

- а) глобальными корреляциями неустойчивого начального состояния
- б) конкретным выбором всех коэффициентов взаимодействия между ячейками.

Усложним модель еще более. По аналогии с вышеописанным, сделаем ее решетку не двухуровневой, а трехуровневой, а процесс вместо двухэтапного - трехэтапным. Итоговым состоянием будет «А-А-А».

Будем считать, что до начала описанного выше трехэтапного процесса, наша крупнозернистая решетка занимала очень малую область физического пространства, но в результате расширения («инфляции») расширилась до больших размеров, после чего и начался описанный выше трехэтапный процесс. Тогда наличие коррелированного неустойчивого начального состояния составной решетки, приводящего именно к итоговому состоянию «А-А-А» можно объяснить тем, что до «инфляции» все ячейки находились близко друг от друга.

Всю эту крупнозернистую решетку в целом можно сравнить с нашей «Вселенной». Ее самые мелкие подрешетки можно сравнить с «разумными организмами». Отсутствие их взаимодействия с окружением до формирования итогового состояния «А» –

эквивалентно их защите от внешнего шума (активно или пассивно) своих внутренних корреляций. Решетки второго уровня в состоянии «А-А» соответствуют «цивилизациям», которые формируют образовавшиеся «разумные организмы» на втором этапе. На третьем этапе из «цивилизаций» формируется «сверхцивилизация» «А-А-А».

Тогда глобальные корреляции неустойчивых начальных состояний решеток могут служить аналогами возможных глобальных корреляций неустойчивого начального состояния нашей Вселенной, возникшего до ее инфляции. Коэффициенты взаимодействия ячеек соответствуют фундаментальным константам. Начальный процесс расширения решетки, до её трехэтапной эволюции, соответствует Большому Взрыву. Специфический выбор коэффициентов взаимодействия ячеек, приводящий к итоговой асимптотике (состоянию «А-А-А»), и начальные корреляции можно объяснить по аналогии с «антропным принципом». Антропный принцип утверждает, что фундаментальные постоянные Вселенной имеют именно такие значения, чтобы в итоге могла возникнуть именно наша наблюдаемая Вселенная с «антропными» существами, способными ее наблюдать.

Глава 2.2.5 Практическое применение непредсказуемых систем для искусственного интеллекта

Рассматриваются перспективы развития искусственного интеллекта (ИИ). Показывается, что разработка ИИ в будущем будет ближе к искусству, чем науке. Основой систем ИИ будут сложные диссипативные системы, поведение которых будет невозможно до конца понять даже в принципе. Тем не менее, это не будет препятствием для их практического использования.

2.2.5.1 Введение

Сейчас во всем мире активно развиваются технологии, связанные с построением систем искусственного интеллекта (ИИ). В этой статье хотелось бы обсудить не тактические, а стратегические задачи этого процесса. Сейчас не так много интересных работ на эту тему, хотя они и имеются [223]. Это связано с тем, что большинство серьезных специалистов занято решением именно тактических задач и часто не задумываются о более далеких перспективах. Однако не такова была ситуация на заре зарождения кибернетики. Тогда эти вопросы активно обсуждались. Поэтому мы построим нашу статью на обзоре задач кибернетики, как они виделись участниками симпозиума в 1961 году [224]. Мы постараемся дать обзор этих перспектив с точки зрения современной физической и кибернетической науки и ее последних достижений.

2.2.5.2 Анализ проблем.

Главное стратегическое направление в 1961 было задано лекцией Бира «На пути к кибернетическому предприятию». В ней он видит систему управления как некий черный ящик с огромным количеством внутренних состояний. В зависимости от внутреннего состояния черный ящик осуществляет разные функции, связывающие его вход и выход.

Среди всех этих функций ищется некая функция, оптимально реализующая его работу, согласно некоторым критериям оптимальности. Организуется обратная связь между выходом предприятия и внутренним состоянием черного ящика, обеспечивающая оптимальность поиска (рис. 3).

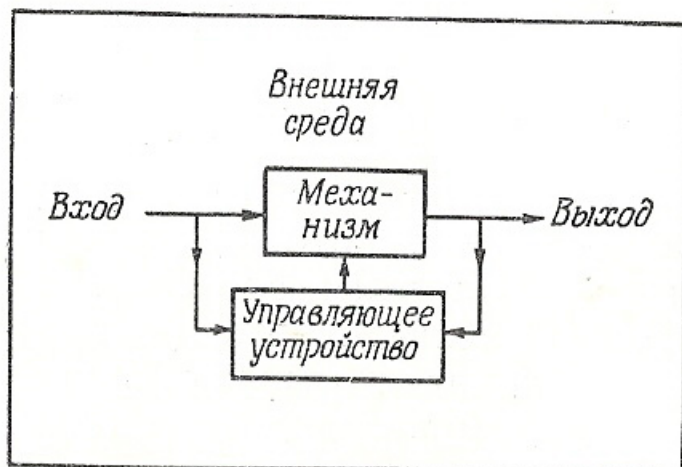


Рис. 3 Схематическое изображение управления механизмом (предприятием).

Тут возникают три трудности:

- 1) Понятно, что число внутренних состояний такого черного ящика должно огромно, чтобы обеспечить реализацию всех возможных функций. Для этого автор предлагает использовать некая глыба вещества, обладающий огромным числом внутренних состояний на атомарном уровне. Это нечто, вроде, например, коллоидной системы Гордона Паса, осуществляющее обращение матриц астрономического порядка.
- 2) Пространство поиска такого ящика огромно и перебор всех возможных внутренних состояний за разумное время не реален. Поэтому необходима стратегия, позволяющая находить пусть не самые оптимальные решения, но, по крайней мере, «хорошие». Такой стратегией в настоящее время считается «генетический алгоритм» [225], снабженный *случайным генератором*. Также используются метод эвристик. [226] Это набор эмпирических рецептов поиска оптимального внутреннего состояния. Они либо находятся из предыдущего опыта, либо заранее заданы внешним экспертом.
- 3) Критерии оптимальности не всегда можно четко сформулировать. Поэтому «целью» такого ящика можно сделать просто физическое «выживание». Тогда подобные критерии он будет искать сам. Либо, его действиям будет давать оценку некий внешний эксперт.

В указанных решениях проблем есть одна очень принципиальная трудность. Пусть наш черный ящик имеет n бинарных входов и один бинарный выход. Тогда число всех возможных внутренних состояний ящика 2^{2^n} . Насколько велико это число? Ответ дает Виллис «Область реализуемых функций для сложных систем» Проведенный им физический расчет, показывающий, что всех молекул Земли достаточно лишь для

реализации черного ящика с максимум $n=155$. Здесь не имеет смысла воспроизводить его расчет. Современная физика дает точный метод подсчета для верхней границы плотности хранения информации через энтропию черной дыры соответствующей массы [227]. (Ее, правда, проблематично извлечь из-за информационного парадокса.) Ответ, однако, вряд ли будет более утешительным. Понятно, что такого количества входов не достаточно для управления сложными системами. Отсюда следует, что количество возможных функций, реализуемых ящиком, должно быть неким подмножеством всех возможных функций. Как же выбрать это подмножество?

Сейчас активно развиваются методы, основанные на нейронных сетях [228] или нечеткой логике [229]. Они позволяют легко реализовать многие «интуитивные» алгоритмы, которые использует человек. Кроме того, для них существуют хорошо разработанные методы обучения или самообучения. Однако для обоих методов показано, что любая возможная функция реализуема этими методами. С одной стороны это хорошо, поскольку доказывает их универсальность. С другой стороны это плохо, поскольку эта избыточность не позволяют нам снизить пространство поиска черного ящика.

В своей лекции Виллис предлагает решение, которое актуально и поныне (рис. 4). Он предлагает использовать подмножество всех функций n переменных, которое реализуется комбинацией p функций k переменных, где

$$p \ll 2^n \tag{1}$$

$$k \ll n \tag{2}$$

Этот класс достаточно мал, чтобы его можно было реализовать.

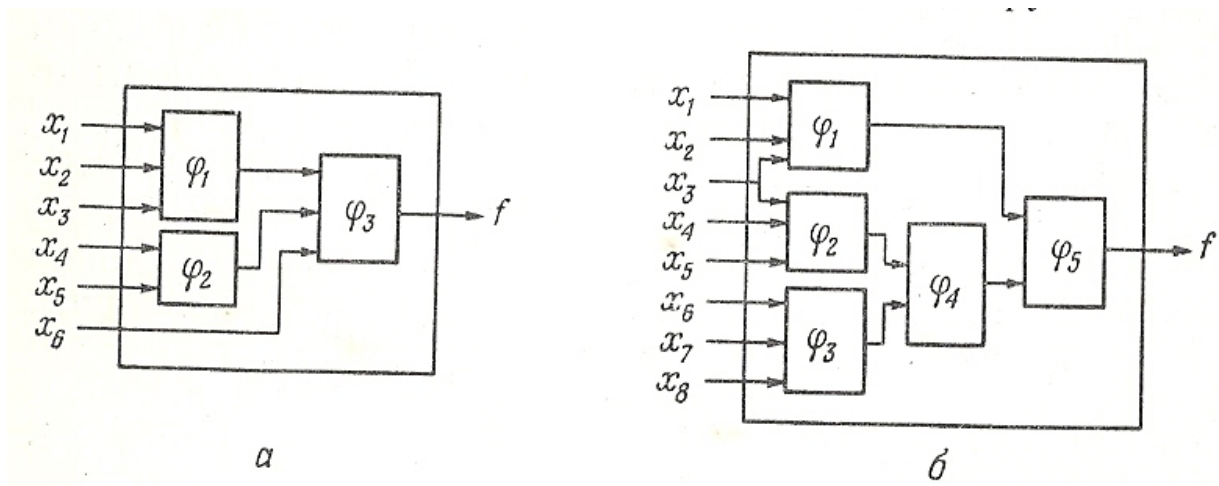


Рис. 4 Точное разложение переключательных функций на функции с меньшим числом переменных:

а) $n=6, p=3, k=3$

б) $n=8, p=5, k=3$

Для широкого класса задач это решение приемлемо. Например, нейронная сеть использовалась для распознавания цифр, высвечиваемых на экране [230]. Экран разбивался на ячейки (пиксели). Ячейка могла быть черной или белой. При этом ячейки разбивались на группы близлежащих ячеек (k -ячеек). Каждая группа поступала на вход сети с одним выходом. Эти выходы группировались также в k ближайших групп, которые подавались на входы сетей и т.д. В итоге имелись лишь 10 выходов, которые и давали результат классификации. Указанная сеть учитывает ограничения связанные с «локальностью» нашего мира.

Но можно вводить и иные критерии ограничивающие пространство поиска того же типа, но менее жесткие. Например, сохранив условие (1), не использовать условие (2), а создать ограничение на тип используемых функций, создать некую «библиотеку» полезных функций.

Например, для существующей области распознавания изображений такой набор функций уже существует – это программные пакеты функций для обработки изображений сосредоточенные в таких пакетах, как Matlab [231]. Комбинируя эти функции, можно создать массу полезных признаков для распознавания. Причем подбирать эти суперпозиции функций можно не случайным перебором генетического алгоритма, а используя человеческую интуицию: человек может комбинировать эти функции так, чтобы они воспроизводили некий интуитивно ощущаемый признак объекта, который человек сам не может математически точно определить. Это человеко-машинные системы поиска.

Следует отметить, что как создание подобных «библиотек», так и человеко-машинные поиск – это не алгоритмизуемые процессы. Они опираются на человеческую интуицию.

Именно поэтому мы считаем, что искусственный интеллект ближе к искусству, чем к науке.

Разберем проблемы, которыми страдает этот подход:

- 1) Те ограничения («библиотеки»), которые мы задаем на внутренние состояния ящика, создаются человеком. Это делает этот процесс трудоемким и ограниченным интуицией человека.
- 2) и человеко-машинный поиск, более эффективный, чем генетический алгоритм, но страдает теми же недостатками, что и описано выше.

Перейдем к следующему докладу, который, кажется, наиболее пророческий и даёт путь к решению этих проблем: Цопф «Отношение и контекст».

Главная его мысль заключается в том, что для поиска эффективных систем искусственного интеллекта мы должны обратиться не к математическим научным абстракциям, а к свойствам подобных систем в окружающем мире – живым адаптивным системам. Какие их свойства позволяют им преодолеть указанные выше ограничения и проблемы?

Самое главное их свойство заключается в том, что подобные системы не являются, как черный ящик, неким внешним объектом по отношению к окружающему миру. Они неразрывно связаны с ним. (Так, Цопф подчеркивает, что признаки, используемые для распознавания объекта, или даже сам «код» нейронов мозга (сознание) являются контекстно-зависимыми. Это значит, что они зависят не только от внутреннего состояния объекта или мозга, но также и его внешнего окружения.) Это объясняет эффективность ограничений на реализуемые внутренние состояния адаптивных систем. Им не нужно придумывать свою «библиотеку» функций – она дана им во многом от рождения. Поскольку эти системы произошли из окружающего мира и связаны с ним уже при рождении множеством незримых связей, эта «библиотека» весьма эффективна и оптимальна. То же самое относится и к алгоритмам адаптации – в отличие от «генетических алгоритмов» они уже оптимально подстроены под окружающий мир, что избавляет от перебора массы ненужных вариантов. Более того сами «цели» адаптивных систем не задаются кем-то извне. Они во многом уже подстроены под их алгоритмы поиска и ограничения окружающего мира.

Мы часто воспринимаем события в окружающем нас мире как набор независимых, случайных явлений. На самом деле, этот мир скорее напоминает самый сложный механизм, пронизанный множеством сложных связей. («Случайности не бывают случайными») Мы не можем наблюдать всю полноту этих связей. Во-первых, поскольку мы являемся лишь малой частью этого мира, то наших внутренних состояний не достаточно, чтобы отобразить всю его сложность. Во-вторых, мы неизбежно взаимодействуем с окружающим миром и влияем на него в процессе наблюдения. Современная физика утверждает, что это взаимодействие не может быть в принципе сведено к нулю

[6,184,187,191,192]. Чтобы промоделировать и учесть это влияние нам нужно отображать внутри себя не только внешний мир, но и самих себя! Такое самонаблюдение невозможно *провести в полной мере в принципе*, при любой нашей степени внутренней сложности. Введение физических макропеременных лишь снижает остроту проблемы, но не решает ее.

Тем не менее, как уже говорилось выше, за счет того, что мы являемся частью этого мира, связаны с ним множеством связей, мы способны на столь эффективное поведение, как будто способны эффективно всё предвидеть и рассчитать. Это свойство адаптивных живых систем можно назвать *сверхинтуицией*³ [232]. Оно значительно превышает адаптивные свойства любого черного ящика, разработанного чисто научными методами.

Следовательно, нам стоит строить наши будущие системы ИИ тоже на основе некоторых подобных «физических» адаптивных систем, обладающих *сверхинтуицией*. Дадим здесь список свойств таких систем [6,10,186,187].

- 1) Случайный генератор подобных систем (делающий выбор внутреннего состояния) не должен генерировать просто случайные числа. Подобные числа должны находиться в сильной связи (корреляции) как с окружающим миром, так и с внутренним состоянием системы ИИ, обеспечивая *сверхинтуицию*.
- 2) Внутренне состояние системы должно быть сложным. Оно должно быть не равновесным, а стационарным. Т.е. оно должно соответствовать динамическому равновесию, подобно водяной стене, падающего водопада. Оно должно быть или коррелированным и слабоустойчивым (даже с элементами хаоса) для систем классической механики, или когерентным квантовым для квантовой механики. Подобные системы способны поддерживать долгое время сложные корреляции между своими частями и между собой и внешним миром.
- 3) Внутреннее состояние система должно быть закрыто от внешнего наблюдения. Это достигается, во-первых, за счет высокой внутренней сложности системы. Во-вторых, система должна сильно менять свое внутренне состояние и поведение при попытке внешнего наблюдения. Этим свойством обладают как слабоустойчивые классические системы (близкие к хаосу), так и квантовые когерентные системы.

³"Еще в 1958 году американский социолог Джеймс Стаунтон проанализировал более 200 железнодорожных аварий за предшествовавшие 30 лет. Оказалось, что поезда, закончившие свой путь трагически, в среднем были заполнены на 61% от максимально возможного числа пассажиров, тогда как в благополучные поездки отправлялись не менее 76%" [Stephen King, "The Stand" \(1990\)](#)

Валерий Исаков, кандидат механико-математических наук и лидер небольшой группы исследователей аномальных явлений рассказал газете "Комсомольская правда" о существовании некой статистической аномалии. После того как Исакову и его группе не удалось получить данные по отказам от полетов у российских авиакомпаний, ученые воспользовались западной статистикой. Как выяснилось, за последние 20 лет от рейсов, закончившихся катастрофами, отказывались на 18% больше пассажиров, чем от благополучных.

<http://kp.ru/daily/25707/908213/>; http://newsru.co.il/world/23jun2011/isakov_606.html

- 4) Система должна быть сильно защищена от внешнего теплового шума (декогеренции).
- 5) Система должна поддерживать свое классическое неустойчивое или когерентное квантовое состояние и защищаться от внешнего теплового шума не столько пассивно, сколько активно. Т.е. это не должен быть твердый панцирь или низкие температуры. Скорее это должен быть активный метаболический процесс. Система должна находиться в стационарном динамическом равновесии, а не термодинамическом равновесии. Так вертикальная стена воды в водопаде поддерживается за счет постоянного её притока извне.
- 6) Главной целью подобных систем должно быть их «выживание».

Для того чтобы использовать подобные системы, нам не нужно детально знать их внутреннее состояние и алгоритмы работы, которые они установят при взаимодействии с окружающим миром. Более того, пытаясь сделать это, мы будем сильно рисковать нарушить их нормальную работу. Мы должны лишь озаботиться, чтобы цели, которые они преследуют для своего «выживания», совпадали с нужными нам задачами.

Мы видим, что в создании таких систем физика становится необходимой для создания кибернетических систем ИИ. Имеются ли сейчас прообразы подобных систем? Многие описанные черты присущи квантовым компьютерам [205,206,209] и их классическим аналогам – классическим неустойчивым компьютерам [200] и молекулярным компьютерам [202]. Кроме того, имеется много литературы, где «на бумаге» строятся синергетические системы, моделирующие указанные выше свойства живых систем. В квантовой области это [198,199,210-213], а для классических неустойчивых систем [201].

Тут перед нами возникает две проблемы:

- 1) Какие объекты, из описанных выше, будут наилучшим образом подходить для создания таких систем ИИ?
- 2) Какие цели, необходимые для «выживания» этих систем, нам нужно поставить, чтобы они совпадали с нашими задачами?

Решение обеих этих задач является неалгоритмизуемым творческим процессом, что опять сближает ИИ скорее с искусством, чем наукой. Действительно, зачастую мы не сможем даже знать, как устроены подобные системы внутри. Мы сможем наложить на них лишь нужные нам ограничения. Да и собственные цели и задачи мы часто не способны сами понять и четко сформулировать. Без всех этих знаний наука бессильна и создание подобных систем будет скорее сродни написанию музыки или рисованию картин. Лишь «кисточка» и «холст» даст нам наука.

Смогут ли всю эту работу выполнить за нас те же системы ИИ? Но если относительно первой задачи такие шансы есть, то вторая из этих целей вообще не может быть выполнена без нас. Ибо кто лучше нас знает, что мы хотим? Кроме того, обе эти задачи взаимосвязаны. Поэтому человеку всегда будет, чем заняться, как бы ни мощны были наши «умные помощники».

2.2.5.3 Выводы.

Рассмотрены перспективы развития искусственного интеллекта (ИИ). Показано, что разработка ИИ в будущем будет ближе к искусству, чем науке. Основой систем ИИ будут сложные диссипативные системы, поведение которых будет невозможно до конца понять даже в принципе. Тем не менее, это не будет препятствием для их практического использования. Но за человеком неизбежно сохраниться важная роль. Полностью исключить его из процесса невозможно

Глава 3 Разрешение основных парадоксов физики на основе учета влияния окружающей среды и наблюдателя

В этом разделе развиты основы термодинамики, позволяющие решить базовые парадоксы современной физики на базе новой кибернетики:

- (a) Противоречие между ростом энтропии и обратимым характером физических законов (парадоксы Пуанкаре и Лошмидта)
- (b) Парадокс Шредингеровского Кота (или парадокс редукции волнового пакета)
- (c) Информационный парадокс черных дыр
- (d) Парадокс дедушки для временных кротовых нор

Для разрешения парадоксов используются два метода новой кибернетики. Первый метод опирается на влияние внешнего наблюдателя или окружающей среды, которое приводит к разрушению корреляций в системе и приводит к синхронизации стрел времени. Второе приближение основывается на ограничениях в возможности самопознания системы, когда мы включаем как наблюдателя, так и окружение в рассматриваемую систему.

Ранее, когда мы имели дело с чисто микроскопическими или макроскопическими системами эти парадоксы можно было просто игнорировать, относя их к «философии». Сейчас, когда активно разрабатывается физика мезоскопических систем, которая имеет не только теоретическое, но и большее практическое значение, игнорирование этих парадоксов приводит к непониманию фундаментальных процессов протекающих в мезосистемах и грубейшим ошибкам, наносящих реальный практический вред. Поэтому последняя глава данной диссертации посвящена анализу и разрешению этих парадоксов.

Глава 3.1 Введение

Эта глава посвящена парадоксам, возникающим при сравнении описания систем на макро и микроуровне [6,184,187-190,233,234]. Статистическая классическая механика, квантовая механика и теория гравитации - это две разработанные и хорошо известные теории. Они представляют собой основу современной физики. Статистическая классическая механика получает свойства больших тел, исследуя движения мельчайших атомов и молекул, из которых эти тела состоят, используя классические законы Ньютона. Квантовая механика определяет законы движения мельчайших частиц на малых атомных расстояниях, рассматривая их как волны вероятности. Законы квантовой механики описываются уравнением Шредингера. Законы такого движения очень отличаются от

законов движения крупных тел, таких как планеты или камни. Описанные две теории давно известны и хорошо изучены. Тем не менее, они содержат ряд парадоксов. Это заставляют многих ученых усомниться во внутренней непротиворечивости этих теорий. Однако данные парадоксы могут быть разрешены в рамках существующей физики, без введения каких либо новых законов. Дано разрешение парадоксов с точки зрения влияния внешнего наблюдателя (окружающей среды), разрушающего корреляции между системы, или системы (в смысле полного описания ее состояния и предсказания ее будущего) в случае, когда и наблюдатель, и окружающая среда включены в систему. Введены понятия идеальной динамики и наблюдаемой динамики. Полученная наблюдаемая динамика может быть получена введением малых внешних шумов, источником которых является внешний наблюдатель или окружение. Это возможно, поскольку, для широкого интервала величин внешних шумов свойства этой динамики не зависят от величины и вида этих шумов, а определяются лишь свойствами самой системы.

Глава 3.2 Макроскопическая энтропия, закон увеличения энтропии и аргумент о стирании памяти наблюдателя

Глава 3.2.1 Релевантность аргумента о стирании памяти наблюдателя для классической механики

Недавняя статья Маккони, “Quantum solution to the arrow-of-time dilemma” [235,236] представляет решение дилеммы стрелы времени, основанное на существующих законах квантовой механики. В статье утверждается, что все явления, где энтропия уменьшается, не должны оставлять информацию в памяти наблюдателя о том, что они случались. При этом наблюдатель является частью всей системы. В статье также полагается, что квантовая механика необходима для этого аргумента, который, как полагает автор, не работает в классической механике (КМ). Однако легко может быть показано, что тот же самый аргумент имеет силу как для квантовой, так и для классической механики. Там используются мысленные эксперименты как Лошмидта (парадокс обращения времени), так и Пуанкаре (теорема возвращения), чтобы проиллюстрировать дилемму стрелы времени. В то же время, Маккони использует только один Лошмидта парадокс, хотя в дальнейшем дает математическое доказательство для общего случая уменьшения энтропии.

Аргументы, приводящие к решению обоих парадоксов в классической механике, являются следующими. Классическая механика позволяет, по крайней мере в принципе, исключить любой эффект воздействия наблюдателя на наблюдаемую систему. Однако большинство реальных систем являются *хаотическими*. Это значит, что слабое возмущение может привести к экспоненциальной расходимости траекторий. Также, между системами всегда есть незначительное взаимодействие. Позвольте привести простой пример газа, расширяющегося из небольшой области пространства в большой объем. В этом процессе с ростом энтропии эволюция макроскопических параметров устойчива к небольшим внешним возмущениям. Если после некоторого времени все скорости будут обращены, то газ вернется в небольшой стартовый объем; причем это верно только в отсутствие внешнего возмущения. Этот процесс с уменьшением энтропии явно неустойчив, и небольшое внешнее возмущение вызвало бы непрерывный рост энтропии. Таким образом, процессы с ростом энтропии устойчивы, а с уменьшением энтропии - нет. Естественное следствие сказанного - то, что стрелы времени наблюдателя и наблюдаемой системы (чьи направления определены ростом энтропии) синхронизированы из-за неизбежного незначительного взаимодействия между ними. Они могут возвратиться к начальному состоянию только вместе (как целая система) как в парадоксе Лошмидта, так и в парадоксе Пуанкаре. Таким образом, память наблюдателя оказывается стертой в конце.

При приближении к этой точке, стрела времени наблюдателя и наблюдаемой системы направлена противоположно координатному (абсолютному) времени. Значит, рост энтропии происходит относительно собственной стрелы времени наблюдателя в полной системе, так же как в ее двух частях, хотя энтропия уменьшается в абсолютном времени.

Закон роста энтропии – как ВВПЦ (во всех практических целях) закон как в классической, так и в квантовой механике. Важно отметить, что ненаблюдательность уменьшения энтропии правильна только для *практических* экспериментов в квантовой механике. Она связана с возмущением измеряемой системы, возникающем в эксперименте. Для идеального невозмущенного наблюдения и термодинамически правильного определения энтропии системы, уменьшение энтропии может *в принципе* наблюдаться в рамках квантовой механики.

Позвольте нам, во-первых, определить невозмущенное наблюдение в квантовой механике. Предположим, у нас есть некоторая квантовая система в известном начальном состоянии. Это начальное состояние может быть либо результатом некоторой подготовки (например, атом переходит в основное электронное состояние в вакууме в течение долгого времени), либо результатом измерительного эксперимента (квантовая система после измерения находится в чистом состоянии, соответствующем собственной функции измеренной переменной). Мы можем предсказать дальнейшую эволюцию начальной волновой функции. Поэтому, *в принципе* мы можем сделать дальнейшие измерения, выбирая измеренные переменные таким способом, чтобы одна из ее собственных функций соответствовала текущей волновой функции наблюдаемой системы. Такой измерительный процесс может позволить нам непрерывное наблюдение без возмущения наблюдаемой квантовой системы. Это невозмущенное наблюдение может быть легко обобщено на случай известного *смешанного* начального состояния.

Например, позвольте нам рассматривать некоторый квантовый компьютер (КК). У него есть некоторое четкое начальное состояние. Наблюдатель, которому известно это начальное состояние может, *в принципе*, произвести невозмущенное наблюдение за любым промежуточным состоянием КК. Однако, наблюдатель, который не знает начальное состояние, не может сделать такое наблюдение, потому что он не может предсказать промежуточные состояния КК.

Мы можем заключить, что закон об увеличении энтропии - *ВВПЦ (во всех практических целях) закон*. Это правильно для *возмущающего* наблюдения за макроскопическими квантовыми системами и классическими макроскопическими хаотическими системами из-за стирания памяти наблюдателя. Такое небольшое возмущение существует в любой реальной ситуации. Однако, в общем случае это не правильно.

Глава 3.2.2 Правильное определение термодинамической энтропии в парадоксе Лошмидта

Цель статьи Маккони состояла в том, чтобы разрешить парадокс Лошмидта, возникающим из-за противоречия между вторым законом термодинамики и обратимостью движения. Поэтому, должно быть выбрано термодинамически правильное определение энтропии, которое действительно используется в формулировке второго закона. Позвольте нам дать такое определение энтропии. Два различных определения энтропии могут быть даны: *макроскопический* энтропии и энтропии *ансамбля*. Макроскопическая энтропия - функция макроскопических параметров, которую мы находим, беря *все* микросостояния, которые соответствуют текущему макросостоянию. С другой стороны, энтропия ансамбля вычисляется для *некоторого* множества микросостояний, эволюционирующих во времени из начального состояния ансамбля. Наконец, стандартная формула для энтропии (Неймановская или для классической механики) по полученному распределению

используется. Вторым законом термодинамики (об увеличении энтропии) используется определение макроскопической энтропии.

Маккони определяет энтропию системы как сумму энтропий ансамбля для наблюдателя и наблюдаемой системы ($S(A \text{ и } C) \equiv S(\rho_A) + S(\rho_C)$). Для случая возмущающего квантового наблюдения за макроскопической системой его определение энтропии эквивалентно макроскопической энтропии, потому что и наблюдатель и наблюдаемая система находятся в смешанных состояниях и коррелируют через макроскопические переменные. Классические аналоги, которые использовал Маккони, чтобы доказать необходимость квантовой механики для доказательства роста энтропии, являются неправильными. В его формуле для смешанной энтропии $S(A:C) \equiv S(\rho_A) + S(\rho_C) - S(\rho_{AC})$ должна была использоваться *макроскопическая* энтропия подсистем как для квантовой, так и для классической механики, а не энтропии ансамбля. В противоположность классической механике, в квантовой механике как макроскопическая энтропия, так и энтропии множества имеют то же самое числовое значение для возмущающего квантового наблюдения, хотя их определения отличаются. Для классических макроскопических хаотических систем (наблюдатель и наблюдаемая система), даже при небольшом взаимодействии между ними, также могут использоваться энтропия ансамбля. Однако, по крайней мере, *начальные состояния* этих двух систем должны быть вычислены от макроскопических параметров, беря *все* микросостояния, им соответствующие.

Однако, в общем случае как в классической, так и в квантовой механике (например, невозмущающее наблюдение), такое определение термодинамической энтропии системы не верно. Действительно, позвольте нам взять простой пример газа, расширяющегося из небольшой области пространства в большой объем. Это - макроскопический процесс с увеличением энтропии. Мы должны использовать термодинамически корректную *макроскопическую* энтропию идеального газа: $S = kN \ln V + \text{const}$ ($T = \text{const}$). Из-за теоремы Пуанкаре о возвратах газ должен вернуться очень близко к стартовому небольшому объему после долгого времени в отсутствие внешнего возмущения. Это - макроскопический процесс уменьшением энтропии. Энтропия ансамбля газа не изменяется во время этой эволюции в отличие от макроскопической энтропии. Предположим, что мы знаем квантовое начальное состояние этого газа, таким образом, *в принципе* мы можем сделать невозмущенное наблюдение описанным выше способом. Поэтому, мы будем в состоянии наблюдать как и увеличение энтропии в начале, так и уменьшение энтропии в конце. Это противоречит главному заключению статьи Маккони. Однако, рассмотрение Маккони и его заключения верны для *практического случая возмущающего квантового* наблюдения за макроскопической системой. В этом случае, мы используем *фиксированное* множество макроскопических переменных для наблюдения. Это множество не зависит от начального состояния (в противоположность случаю невозмущенного наблюдения, описанного выше).

Мы не находим ошибок в рассмотрении энтропии Маккони в рамках квантовой механики для возмущенного наблюдения за макроскопическими системами. В противоположность этому, D. D. Jennings, T. Rudolph, "Comment on "Quantum Solution to the Arrow-of-Time Dilemma" of L. Maccone" [237,238] находят возражение для этого определения энтропии. Однако, возражение Дженнингса и Рудольфа не релевантно, поскольку мы рассматриваем макроскопические системы. В то время как примеры Дженнингса и Рудольфа соответствуют микроскопическим системам.

Глава 3.3 Парадокс Шредингеровского кота и спонтанная редукция.

Полное нарушение принципа суперпозиций волн (то есть полное обращение в нуль интерференционных членов) и редукция волновой функции возможны только во время взаимодействия квантовой системы с идеальным макроскопическим объектом или

устройством. Идеальный макроскопический объект или имеет бесконечный объем, или состоит из бесконечного числа частиц. Такой идеальный макроскопический объект может быть последовательно описан как квантовой, так и классической механикой.

Далее, (если иное не оговаривается) мы рассматриваем, как и в классическом случае, только системы с конечным объемом с конечным числом частиц. Такие устройства или объекты можно рассмотреть как макроскопические только приблизительно.

Однако реальной эксперимент показывает, что даже для таких неидеальных макроскопических объектов разрушение суперпозиции и соответствующая редукция волновой функции могут произойти. Мы определим такое приведение несовершенных макроскопических объектов как *спонтанная редукция*. Спонтанная редукция приводит к парадоксам, которые вызывают сомнения в замкнутости квантовой механики, несмотря на все ее огромные успехи. Мы приведем самый выразительный парадокс из этого ряда - парадокс Шредингерского Кота.

Это - мысленный эксперимент, который проясняет редукцию волновой функции и принцип суперпозиции. Кот помещен в ящик. За исключением Кота, есть капсула с ядовитым газом (или бомба) в ящике, которая может взорваться с 50-процентной вероятностью из-за радиоактивного распада атома плутония или случайно излученного светового кванта. Через некоторое время, ящик открывается, и становится ясным, жив Кот или нет.

Пока ящик не открыт (измерение не выполнено), Кот находится в очень странной суперпозиции двух состояний: "живой" и "мертвый" Кот. Для макрообъектов такая ситуация выглядит очень таинственной (В противоположность этому, для квантовых частиц суперпозиция двух различных состояний является очень естественной). Однако нет принципиального запрета на квантовую суперпозицию макросостояний.

Редукция этих состояний при открытии ящика внешним наблюдателем, не приводит ни к какому противоречию с квантовой механикой. Это легко объясняется взаимодействием внешнего наблюдателя с Котом во время измерения его состояния.

Однако парадокс возникает в закрытом ящике, когда наблюдатель - сам Кот непосредственно. Действительно, Кот обладает сознанием и способен наблюдать как непосредственно себя, так и свое окружение. При реальном самонаблюдении, Кот не может быть одновременно и жив и мертв, он находится только в одном из этих двух состояний. Опыт показывает, что любые существа, обладающие сознанием, ощущают себя либо живыми, либо мертвыми. Одновременно обе такие ситуации не существуют. Поэтому спонтанная редукция к одному из двух возможных состояний (живой и мертвый) действительно происходит. Кот, даже вместе со всем содержанием ящика, - не *идеальный* макроскопический объект. Поэтому, такая наблюдаемая необратимая спонтанная редукция противоречит к обратимой Шредингерской квантовой динамике. В описываемом случае это нельзя объяснить неким внешним влиянием, потому что система изолирована.

Противоречит ли редукция на самом деле Шредингерской квантовой динамике? Когда система становится достаточно макроскопической, чтобы обеспечить возможность для спонтанной редукции? Должна ли такая почти макроскопическая система иметь сознание подобно Коту?

Мультимировая интерпретация также не объясняет парадокс Шредингерского Кота, эта интерпретация только переформулирует и маскирует парадокс. Действительно, Кот наблюдает только один из существующих миров. Однако результаты будущих измерений зависят и от корреляций между мирами. Однако, ни эти миры, ни эти корреляции не наблюдаемы. «Параллельные миры», о которых мы не знаем ничего, могут всегда существовать. И эти миры могут действительно влиять на результаты некоторых будущих экспериментов. То есть знание текущего состояния только (в нашем "мире") и квантовые законы механики не позволяют нам

предсказывать будущее даже вероятно! Однако квантовая механика была создана именно для таких предсказаний! Только на основании спонтанной редукции, которая уничтожает квантовые корреляции между мирами, мы можем предсказать будущее, используя знание только настоящего (и действительно наблюдаемого) состояния нашего "мира". Парадокс Шредингеровского Кота возвращается, но только меняет свою форму.

Позвольте напоминать, что парадокс Кота Шредингера состоит в несогласованности между спонтанной редукцией, наблюдаемой Котом и эволюцией Шредингера, запрещающей такую редукцию. Необходимо рассмотреть парадокс Кота Шредингера с точки зрения двух наблюдателей, чтобы понять его правильно: это или внешний наблюдатель - экспериментатор, или сам Кот (то есть, *самонаблюдение*)

В случае внешнего наблюдателя-экспериментатора, парадокса не возникает. Если экспериментатор попытается увидеть, жив ли Кот или нет, это повлияет неизбежно на наблюдаемую систему (в согласии с квантовой механикой), что приведет к редукции. Система не изолирована и, следовательно, не может быть описана уравнением Шредингера. Редуцирующая роль наблюдателя может также играть окружающая среда вместо наблюдателя. Этот случай определен как декогеренция. Здесь роль наблюдателя более естественна и сводится только к регистрации редукции, полученной из-за декогеренцией. В обоих случаях, существует запутывание измеряемой системы с окружающей средой и/или наблюдателем, то есть, возникают корреляции измеряемой системы с окружающей средой и/или наблюдателем.

Что случится, если мы рассматриваем замкнутую полную физическую систему, включающую наблюдателя, наблюдаемую систему и окружающую среду? Это - случай самонаблюдения Кота. Система включает Кота и его окружающую среду в ящике. Следует отметить, что *полное самонаблюдение (полное в смысле квантовой механики) и полная проверка квантовых законов механики невозможны в изолированной системе, включающей самого наблюдателя.* Действительно, мы можем точно измерить и проанализировать состояние внешней системы, по крайней мере, в принципе. Однако если мы включаем себя также в рассмотрение, возникают естественные ограничения. Это имело отношение с возможностью держать в памяти и анализировать состояния молекул посредством этих же самих молекул. Такое предположение ведет к несоответствию. Поэтому, возможность найти экспериментально несоответствие между эволюцией Шредингера и спонтанной редукцией с помощью самонаблюдения в изолированной системе также ограничена.

Тем не менее, позвольте нам пробовать найти некоторые мысленные эксперименты, ведущие к несогласованности между эволюцией Шредингера и спонтанной редукцией.

- 1) Первый пример связан с обратимостью квантовой эволюции. Предположим, что мы ввели функцию Гамильтона, способную полностью обратить квантовую эволюцию полной системы - ящика и Кота. Хотя фактически это почти невозможно, теоретически никакой проблемы не существует. Если бы спонтанная редукция происходила, процесс был бы необратим. Если спонтанная редукция не происходит, система "ящик и Кот" возвращается к начальному чистому состоянию. Однако только внешний наблюдатель может делать такую проверку. Кот не может сделать ее самонаблюдением, потому что память Кота будет стерта после возвращения в начальном состоянии. С точки зрения внешнего наблюдателя, никакого парадокса и не существует, потому что он не наблюдает спонтанную редукцию, которая действительно может вести к парадоксу.
- 2) Второй пример связан с неизбежными возвратам Пуанкаре квантовой системы к начальному состоянию. Предположим, что начальное состояние было чистое. Если спонтанная редукция действительно существует в случае самонаблюдения Кота, она

приводит к смешанному состоянию. Тогда возврат становится уже невозможен - смешанное состояние не может перейти в чистое состояние согласно уравнениям Шредингера. Таким образом, если бы Кот смог детектировать возврат, это привело бы к несогласованности со спонтанной редукцией. Однако Кот не может детектировать возврат (в случае верности квантовой механики), потому что возвращение сотрет память Кота. Поэтому, нет и никакого парадокса. Внешний наблюдатель фактически может наблюдать это возврат, измеряя начальное и конечное состояние этой системы. Однако в этом случае также никакого парадокса уже не существует, потому что он не наблюдает никакой спонтанной редукции, которая действительно может вести к парадоксу. Стоит обратить внимание, что несогласованность между спонтанной редукцией и эволюцией Шредингера может быть экспериментально зарегистрирована только, когда спонтанная редукция сохранена в памяти наблюдателя, и эта память не стерта и не повреждена. Никакие эксперименты, описанные выше, не отвечают этим требованиям. Таким образом, эти примеры ясно показывают, что, хотя спонтанная редукция действительно может вести к нарушению эволюции Шредингера, это редукция экспериментально не наблюдаема (в случае верности квантовой механики).

- 3) Третий пример следующий. Квантовая механика дает суперпозицию живого и мертвого Кота в ящике. Теоретически, внешний наблюдатель мог всегда измерять эту суперпозицию *точно*, если она является одной из собственных функций измеряемой величины. Такое измерение не уничтожило бы суперпозицию, вопреки случаю, когда состояния живой или мертвый Кот являются собственными функциями измеряемой величины. Сообщив Коту о результате такого измерения, мы получим несогласованность со спонтанной редукцией, наблюдаемой Котом. Такая аргументация содержит двойную ошибку. *Во-первых*, этот эксперимент используется для *проверки* спонтанной редукции Кота, когда наблюдатель - сам Кот. Внешний наблюдатель не влияет на память Кота только, когда спонтанная редукция не присутствует, и состояние Кота - суперпозицию живого и мертвого состояний. Однако это действительно влияет и может уничтожить память Кота в случае спонтанной редукции. Поэтому, такой эксперимент не законен для проверки спонтанной редукции в *прошлом*. *Во-вторых*, данные, переданные Коту, сохраняются в его памяти. Таким образом, эта передача изменяет и его состояние, и эволюцию Кота нельзя рассматривать уже как изолированную после измерения. Поэтому, никакого противоречия с *будущим* также не возникает.

Внешний наблюдатель не видит спонтанной редукции и, следовательно, не наблюдает парадокс. Таким образом, с точки зрения внешнего наблюдателя, проверка помощью непрерывного невозмущающего наблюдения, описанного в пункте 3, является весьма возможным и законным. Это не влияет на память внешнего наблюдателя. Кроме того, такая проверка, которая не нарушает динамику наблюдаемой системы, позволяет измерять не только начальное и заключительное состояние системы, но также и всех промежуточные состояния. То есть, наблюдатель осуществляет непрерывное невозмущающее наблюдение!

Следует отметить, что внешний наблюдатель может наблюдать суперпозицию живого и мертвого Кота только теоретически. Фактически это почти невозможно. Однако для маленьких квантовых систем, суперпозицию легко наблюдаема. Это объясняет тот факт, что квантовую механику рассматривают обычно как теорию маленьких систем. Однако, для маленькой макроскопической (мезоскопической) системы такое наблюдение возможно также. Набор частиц при низких температурах или некоторые состояния фотонов могут быть примером. Позвольте сделать важное замечание: за последние годы были изданы очень интересные статьи в направлении строительства мезоскопических "синергетических" систем, возможно подобных в некоторых их свойствах живым

организмам. Следует упомянуть, что создание таких моделей - проблема физики и математики, а не философии.

Глава 3.4 Синхронизация стрел времени/декогеренция

Следующий вопрос может появиться. Позвольте предположить, что происходит некоторый процесс, идущий с уменьшениями энтропии. Для определенности, позвольте выбрать в качестве этого процесса самореконструкцию дома, предварительно разрушенного землетрясением.

Позвольте также взять простой пример газа, расширяющегося из маленькой области пространства в большой объем. Если после некоторого времени все скорости будут полностью обращены, то газ вернется в стартовом маленький объем.

Если мы включаем камеру, чтобы сделать ряд снимков во время обратных процессов, мы ожидаем, что камера будет делать запись процесса самореконструкции дома / (сокращение газа). Почему же камера не будет способна делать запись этого? Что может помешать камере сделать эти снимки?

Ответ эти вопросы следующий: даже очень небольшое взаимодействие между камерой и наблюдаемой системой разрушает обратный процесс, идущий с уменьшением энтропии, и приводит к синхронизации направлений стрел времени наблюдателя и наблюдаемой системы. (Направление стрелы времени определено в направлении увеличения энтропии). Это очень небольшое взаимодействие появляется из-за света, излученного наблюдаемым объектом и отраженного камерой (и также из-за света, излученного камерой). В отсутствие камеры роль наблюдателя может играть окружающая среда, излучая и отражая свет. У любого процесса без наблюдателя нет физического смысла. Наблюдатель должен появиться на некотором этапе любого процесса. Однако его влияние намного меньше, чем влияние окружающей среды. Внешний шум (взаимодействие) от окружающей среды/наблюдателя разрушает корреляции между молекулами наблюдаемой системы. Это приведет к предотвращению обратного процесса с уменьшением энтропии. В квантовой механике такой механизм определен как "декогеренция". Самореконструкция дома / (сжатие газа) будет остановлена, то есть, дом не восстановится / (газ не сожмется). Напротив, процессы с увеличением энтропии устойчивы.

Следующий пример - цитата из статьи Маккони, "Quantum solution to the arrow-of-time dilemma": «Однако наблюдатель является макроскопическим по определению, и все удаленно взаимодействующие макроскопические системы становятся коррелированными очень быстро (например, Борель замечательно вычислил, что, перемещение грамма материала по звезде Сирус на 1 м может влиять на траектории частиц в газе на Земле на временных рамках 1 s)»

Однако никакой проблемы не существует, чтобы полностью обратить вместе наблюдателя (камеру) и наблюдаемую систему. Из-за теоремы о возвратах Пуанкаре для замкнутой системы (включая наблюдателя и наблюдаемую систему), это должно случиться автоматически после очень большого промежутка времени. Однако стирание памяти наблюдателя не позволяет ему зарегистрировать этот процесс.

Большинство реальных системы являются *хаотическими* - слабое возмущение может привести к экспоненциальной расходимости траекторий, и также всегда есть незначительное взаимодействие между наблюдаемой системой и наблюдателем/окружающей средой. Однако *в принципе*, как в квантовой механике, так и в классической механике, мы можем осуществить невозмущенное наблюдение за процессом уменьшения энтропии. Хороший пример такого мезоскопического устройства - квантовый компьютер: никакой закон об увеличении энтропии не существует для такой системы. Это устройство очень хорошо изолировано от окружающей среды и наблюдателя. Однако, на *практике* невозмущенное наблюдение почти невозможно для

микроскопических систем. Мы можем заключить, что закон об увеличении энтропии - *ВВПЦ* (во всех практических целях) *закон*.

Следует упомянуть, что декогеренция (синхронизация стрел времени и "запутывание") и релаксация (во время релаксации система достигает своего равновесия) являются абсолютно различными процессами. Во время релаксации, макроскопические переменные (энтропия, температура, давление) значительно изменяются, достигают равновесных значений, и невидимые микроскопические корреляции между частями системы увеличиваются. Во время декогеренции макроскопические переменные (энтропия, температура, давление) являются почти постоянными. Невидимые микроскопические корреляции в подсистемах (окружающая среда, наблюдатель, и наблюдаемая система) сильно разрушены, но новые корреляции появляются между подсистемами. Это называют "запутыванием" в квантовой механике. Во время этого процесса также происходит синхронизация стрел времени. Время релаксации намного больше, чем времени декогеренции.

Позвольте рассмотреть синхронизацию стрел времени для двух не взаимодействующих (до некоторого начального момента) систем. Следует упомянуть, что это описание сделано в абсолютной (координатной) системе. Однако у обеих систем есть собственные первоначально противоположные стрелы времени также, которые определены как направление роста энтропии в каждой системе.

Это означает, что существуют две не взаимодействующие системы, такие, что в одной из них поток времени (то есть, направление увеличения энтропии) направлен в одном направлении, в то время как в другой системе время течет в противоположном направлении. Однако когда они входят во взаимодействие друг с другом, тогда одна из них ("более сильная") будет перетягивать другую ("более слабую") в его "более сильное" направление. Таким образом, в конечном счете, обе системы будут иметь время, текущее в том же самом направлении.

"Быть более сильным" - что это значит? Это что-то, что увеличивается с числом степеней свободы системы? нет, это не верно. "Более сильный" или "более слабый" не зависит значительно от числа степеней свободы систем (маленькие флуктуации являются исключением). Описанное выше взаимодействие асимметрично для абсолютного (координатного) времени. Для первой системы, взаимодействие появляется в ее *будущем*, начиная с начального момента (В начальный момент системы имеют противоположные стрелы времени) согласно собственному времени этой системы. Для второй же системы взаимодействие было в ее *прошлом* согласно ее собственному времени. Поэтому, ситуация не является *симметричной во времени*, и первая система всегда "более сильна". Это случается из-за неустойчивости процессов с уменьшением энтропии и стабильности процессов с увеличением энтропии, которые были описаны выше.

Действительно, позвольте нам снова рассмотреть два первоначально изолированных сосуда с газом. В первом, один газ расширяется (увеличение энтропии). Во втором, один газ сжимается (уменьшение энтропии).

В первом сосуде, газ расширяется из маленького объема в центре сосуда. Скорости молекул направлены от центра судна к его границе. Физически ясно, что маленькое возмущение скоростей не может остановить расширение газа. Действительно, скорости после случайного маленького возмущения продолжают быть направленными от центра судна к его границе. Шум может даже увеличить расширение. Поэтому, процесс расширения устойчив.

Во втором сосуде, газ сжимается от полного объема сосуда к его центру. Скорости всех молекул направлены к центру сосуда. Физически ясно, что маленькое случайное возмущение скоростей может легко остановить сжатие газа. Действительно, скорости даже после маленького возмущения не будут направлены к центру сосуда. Таким образом, процесс сжатия будет остановлен. Поэтому, мы можем заключить, что процесс

сжатия неустойчив. Этот процесс сжатия может быть получен путем обращения во времени газового расширения. Если мы обращаем скорости молекул *перед* столкновениями молекул газа друг с другом и с границами сосуда, то такая неустойчивость будет линейна и не сильна. Однако если сделать обращения скоростей *после* столкновений эта неустойчивость становится экспоненциальной и намного более сильной.

Оба направления времени равноправны. Однако маленькое случайное шумовое взаимодействие нарушает эту симметрию для описанных выше двух систем из-за неустойчивости процессов с уменьшением энтропии. Симметрия времени существует только для *полной* системы, включающей эти две, определенные выше подсистемы. Однако собственные стрелы времени взаимодействующих подсистем должны стать сонаправленными.

Вместо взаимодействия с бесконечным временем $[0, +\infty]$ мы можем рассмотреть взаимодействие с большим, но конечным временем T : $[0, T]$. Позвольте нам выбирать на сей раз T так, чтобы оно было намного меньшим, чем время возвращения Пуанкаре. Тогда в первой системе мы имеем взаимодействие в течение $[0, T]$ согласно ее собственному времени и во втором в течение $[-T, 0]$ согласно ее собственному времени. Можем ли мы все еще применить наш аргумент? Вместо асимметрии сил в этом случае мы получаем асимметрию начальных условий: В начальный момент 0 для первой системы ($[0, T]$) эти два сосуда имеют различные собственные стрелы времени. Однако, в начальный момент $-T$ для второй системы $[-T, 0]$ эти два сосуда имеют одинаковые собственные стрелы времени в отрицательном направлении.

Только если T точно равен времени возвращения Пуанкаре, ситуация будет действительно симметричная. Для такой ситуации, две собственных стрелы времени также отличны в момент T , но каждая теперь противоположна ее первоначальному направления в момент времени 0 . Снова, "более сильная" система имеет силы взаимодействия в ее будущем относительно ее собственной стрелы времени.

Эта теория может объяснить одинаковое направление роста энтропии во всех частях Вселенной. Однако она не может объяснить низкую энтропию в начале существования Вселенной. Это - вероятно результат антропного принципа.

Глава 3.5 Закон увеличения энтропии и синхронизации стрел времени/ декогеренция в гравитационной теории.

Глава 3.5.1 Черные Дыры

В общей теории относительности Эйнштейна движение обратимо подобно к классической механике. Однако также существует важное различие между общей теорией относительности и классической механикой. Общая теория относительности - неоднозначная теория. Действительно, в общей теории относительности два различных начальных состояния могут дать бесконечно мало близкие состояния после *конечного* временного интервала. Это случается, например, во время коллапса черной дыры. Позвольте нам рассмотреть обратный процесс, описывающий белую дыру. В этом процессе, бесконечно близкие начальные состояния за *конечный* временной интервал могут дать различные заключительные состояния. Это означает, что наблюдатель/окружающая среда может воздействовать значительно на ее развитие в течение *конечного* временного интервала, даже когда наблюдатель/окружающая среда бесконечно слабо взаимодействует с белой дырой. Следует упомянуть, что в гравитационной теории, произвольное маленькое, но конечное взаимодействие *всегда* существует в противоположность к обычной классической и квантовой механике.

Силы гравитации всегда существуют между произвольными двумя объектами с массами, отличными от нуля.

Из-за этих двух свойств, закон увеличения энтропии становится точным законом, а не ВВПЦ (во всех практических целях) в общей теории относительности. Поэтому, энтропия становится фундаментальным понятием. Действительно, есть такое фундаментальное понятие, как энтропия черной дыры. Кроме того, возможно объяснить существование этой энтропии возмущением, созданным наблюдателем. Это возмущение может быть теперь даже бесконечно слабым в отличие от классической механики. В процессе формирования черной дыры происходит увеличений энтропии. Обращение времени ведет к появлению белой дыры и к уменьшению энтропии.

Белая дыра не может существовать в реальности из-за уменьшения энтропии. Уменьшение энтропии запрещено в общей теории относительности по тем же самым причинам, что это запрещено в классической механике. Это - неустойчивость процессов уменьшения энтропии, которая намного более сильная в общей теории относительности, чем в классической механике. Эта неустойчивость приводит к синхронизации собственных стрел времени белой дыры и наблюдателя/окружающей среды. Направление собственной стрелы времени белой дыры изменяется на противоположное, совпадающее с собственной стрелой времени наблюдателя/окружающей среды. Белая дыра преобразовывается к черной дыре.

Рассмотри известный информационный парадокс черной дыры [239]: информация (которая в классической и квантовой механика сохраняется), исчезает в черной дыре навсегда. Казалось бы, нет никакой проблемы: вероятно информация сохраняется в черной дыре в некоторой форме. Однако хаотическое излучение Хокинга делает явным этот процесс информационных потерь: черная дыра испаряется, но информация не возвращается.

Излучение Хокинга относится к полуклассической теории тяготения. Однако парадокс может быть сформулирован также в рамках общей теории относительности. Сферическая черная дыра может быть обращена в белую дыру в некоторый момент времени. (Это кажется, что невозможным, но физически подобная ситуация с "кротовой норой", соединяющей черную и белую дыры в различных Вселенных рассматривается в [240]). Таким образом, процесс оказывается обращен. Однако, информация не может быть возвращена из-за неоднозначности (или, иначе говоря, бесконечно сильной неустойчивости) эволюции белой дыры. Обычно только два решения к этой проблеме рассматривают. Либо информация действительно исчезает, либо из-за внутренних корреляций излучения Хокинга (или точного обращения черной дыры после ее превращения в белую) информация сохраняется. Однако, наиболее вероятно, третье решение. Из-за неизбежного влияния наблюдателя/окружающей среды просто невозможно различить эти две ситуации экспериментально! Однако если это невозможно проверить экспериментально, то это - уже не предмет науки

Как для общей теории относительности, так и для полуклассического тяготения парадокс может быть решен посредством влияния наблюдателя/окружающей среды. Действительно, позвольте нам предположить, что излучение Хокинга является коррелированным, а не хаотическим (или белая дыра является точной инверсией черной дыры). Таким образом, бесконечно малое влияние наблюдателя/окружающей среды ведет к неизбежным потерям этих корреляций (и соответствующей информации) в течение конечного временного интервала. Бессмысленно включать наблюдателя в описываемую систему: полное самописание и самоанализ не невозможны. Информационный закон сохранения не может быть проверен экспериментально для такого случая, даже если он действительно верен.

На данный момент отсутствует общепризнанная теория квантового тяготения. Однако, для специального случая 5-мерного анти-де-Ситтеровского пространства этот парадокс, как полагают многие ученые, решен. Информация, как предполагается, сохраняется, потому что гипотеза о AdS/CFT дуальности, то есть гипотеза, что квантовое тяготение в 5-мерном анти-де-Ситтеровском пространстве (которое имеет отрицательный космологический член) математически эквивалентно конформной полевой теории на 4 поверхностях этого мира. Это было проверено в некоторых специальных случаях, но не доказало еще в общем случае.

Предположим, что эта гипотеза действительно верна. На первый взгляд, это автоматически решает информационную проблему. По сути дела конформная полевая теория является унитарной. Если она является действительно дуальной квантовому тяготению, тогда и соответствующая квантовая теория тяготения унитарна также. Поэтому, информация в этом случае не теряется. Однако, мы предполагаем, что это не верно. Процесс формирования черной дыры и ее последующего испарения случается на *всей поверхности* анти-де-Ситтеровского пространства (описанного в соответствии с конформной квантовой теорией). Это пространство включает также наблюдателя/окружающую среду. Однако наблюдатель не может точно знать свое начальное состояние и не может анализировать поведение системы и проверить унитарность, потому что он - часть этой системы! Так что его влиянием на систему нельзя пренебречь. Таким образом, экспериментальная проверка информационного парадокса снова становится невозможной!

Глава 3.5.2 Кротовые норы

Позвольте нам рассмотреть с точки зрения закона роста энтропии такой парадоксальный объект общей теории относительности, как кротовая нора [241]. Мы рассмотрим кротовую нору Морриса - Торна [242]. В соответствии с очень простой процедурой (мы помещаем один из входов кротовой норы на космический корабль, затем космический корабль с релятивистской скоростью движется по замкнутой петле, и возвращает вход в его первоначальное место) пространственная кротовая нора может быть преобразована во временную кротовую нору. После этого преобразования, кротовая нора может использоваться как машина времени, ведя к известному парадоксу дедушки. Как же этот парадокс может быть решен?

Для макроскопических кротовых нор, решение может быть найдено посредством закона роста энтропии. Реализация этого закона обеспечена неустойчивостью процессов, идущих с уменьшения энтропии, которая ведет к синхронизации стрел времени.

Действительно, пространственная кротовая нора не ведет к парадоксу. Если объект входит в один конец в некоторый момент времени, то он выходит из другого конца в некоторый более поздний момент времени. Таким образом, объект путешествует из начального низко-энтропийного окружения в будущее высоко-энтропийное окружение. В течение движения по кротовой норе, энтропия объекта также увеличивается. Таким образом, направления стрел времени объекта и окружающей среды одинаковы. Те же самые заключения верны для путешествия от прошлого к будущему в кротовой норе, пересекающей время.

Однако, при путешествии из будущего в прошлое, направление стрел времени объекта и окружающей среды будут уже противоположны. Действительно, объект путешествует начального высоко-энтропийного окружения в будущее низко-энтропийное окружение. Однако его собственная энтропия и не уменьшается, а увеличивается! Поскольку мы говорили ранее, такой процесс неустойчив и будет предотвращен процессом синхронизации стрел времени.

Изначальная синхронизация кротовой норы с ее окружением происходит при возвращении конца кротовой норы к ее начальному состоянию при ее формировании.

Откуда берется окружающая среда внутри кротовой норы? Окружающая среда путешественника внутри кротовой норы - это радиация, полученная термодинамическим излучением массивных концов кротовой норы

"Свободная воля" позволяет нам инициировать только необратимые процессы с увеличением энтропии, но не с ее уменьшением. Таким образом, мы не можем послать объект из будущего в прошлое. Процесс синхронизации стрел времени (и соответствующий закон роста энтропии) запрещают начальные условия, которые являются необходимыми для путешествия макроскопического объекта в прошлое (и реализацию условий для парадокса дедушки).

В статье [243] демонстрируется, что термодинамическая стрела времени не может всегда иметь ориентацию идентичную с координатной стрелой времени вдоль замкнутой временно-подобной кривой из-за закона роста энтропии. Описанный здесь процесс синхронизации стрел времени (связанный с бесконечно большой неустойчивостью и неоднозначностью процессов с уменьшением энтропии) является тем *физический механизм*, который фактически гарантирует и эту невозможность и реализует закон роста энтропии вдоль этой же термодинамической стрелы времени.

Для микроскопических кротовых нор, ситуация абсолютно другая. Если начальные условия совместимы с путешествием в прошлое по кротовой норе, нет никаких причин, которые могут предотвратить это. Если некоторое малое (даже бесконечно малое) возмущение начальных условий ведет к несогласованности с существованием кротовой норы, она может всегда легко разрушаться [244]. Действительно, вспомним упомянутое выше свойство общей теории относительности - это ее бесконечно большая неустойчивость (неоднозначность). Это означает, что бесконечно-малое возмущение начальных условий может привести к конечному изменению итогового состояния в течение конечного времени!

Однако это не может быть решением парадокса дедушки, который является макроскопическим, а не микроскопическим явлением.

Действительно, позвольте предположить, что есть два процесса с противоположными направлениями стрелы времени: космонавт и окружающая его Вселенная. Космонавт путешествует по кротовой норе из будущей Вселенной к прошлой Вселенной. Однако для собственного направления стрелы времени космонавта это будет путешествие из его прошлого в его будущее.

Для общей теории относительности ситуация, описанная выше невозможна даже в принципе в отличие от классической механики: даже бесконечно малое взаимодействие ведет к бесконечно большой неустойчивости (неоднозначности) процессов, идущих с уменьшением энтропии (в данном случае "процесс с уменьшением энтропии" является космонавтом, путешествующим из будущего в прошлое).

В общем случае, описанная выше несогласованность между макроскопическими начальными условиями может сопровождаться как уничтожением кротовой норы [244], так и сохранением кротовой норы, обращением стрелы времени космонавта и стиранием его памяти [243].

Фактически, с *очень высокой вероятностью* закон роста энтропии приводит к синхронизации стрел времени, соответствующей инверсия стрелы времени космонавта и стиранию его памяти. Это приводит с *очень высокой вероятностью* к стабильности первоначальной макроскопической пространственно-временной топологии (включающей набор кротовых нор) [243] и очень маленькой вероятности макроскопического разрушения кротовых нор.

Однако с очень маленькой вероятностью синхронизация стрел времени может не произойти. Это - очень редкая крупномасштабная флуктуация. В этом случае, может произойти разрушение кротовой норы.

Позвольте сформулировать заключение: для макроскопических процессов неустойчивость процессов с уменьшением энтропии, гравитационное взаимодействие и соответствующая синхронизация стрел времени делают почти невозможным возникновение начальных условий, несовместимых с существованием макроскопических кротовых нор. Эта неустойчивость предотвращает как разрушение макроскопических кротовых нор, так и путешествие макроскопического объекта в прошлое, ведущее к "парадоксу дедушки" [234].

Наконец, мы видим замечательную ситуацию. Те же самые причины, который позволили нам решать парадокс редукции, парадоксы Лошмидта и Пуанкаре, позволяют также решить информационный парадокс черных дыр и парадокс дедушки для кротовых нор. Это замечательная универсальность!

Заключение

Автором были получены следующие достижения:

- 1) Для систем, рассчитываемые на основе известной физики подсистем разработаны следующие практические применения указанной теории: решены математические и компьютерные модели с высокой универсальностью, позволившие решить следующие задачи:
 - (a) Для первой группы проблем, (где учет влияния внешней среды/наблюдателя, рассчитывается на основе точных законов физики - идеальной динамики) разработана континуальная модель среды для полимерных молекул и их комплексов (нанокompозитные материалы, коллоидные частицы, растворы полимеров): Был развит существующий алгоритм построения гладкой поверхности и усовершенствована программа TAGSS; разработан адаптивный механизм для триангуляции; доведены до конца методы расчета аналитических градиентов параметров поверхностных элементов при сдвиге атомов; проведен на этой основе расчет энергии сольватации и аналитических градиентов энергии сольватации при сдвиге атомов в методах COSMO, SGB, PCM; написана и запатентована программа DISOLV, проводящая эти расчеты; проведен расчет диагональных элементов в методе COSMO; доведена до конца точная аналитическая интегральная модель растворителя
 - (b) Для второй группы проблем (где учет влияния внешнего шума/наблюдателя осуществляется на основе наблюдаемой динамики макропараметров, рассчитываемой из идеальной динамики и независимой от амплитуды шума) детально проанализированы полюсные решения в теории горения смеси газов в прямоугольном канале; проанализировано влияние малых шумов на распространение фронта горения и на его основании объяснено ускорение фронта пламени в случае газовой смеси для радиальной и сферической геометрии; объяснено увеличение скорости фронта при увеличении ширины прямоугольного канала; объяснена ячеистая структура фронта пламени; показаны естественные методы регуляризации полюсного решения в теории фильтрационного горения, позволяющие сохранить полученное точное аналитическое полюсное решение, объяснена асимптотика решения в виде одного пальца, равного половине ширины канала.
- 2) Для эмерджентных систем, т.е. систем, не рассчитываемых аналитически на основе известной физики подсистем, получены следующие результаты:
 - (a) Для третьей группы проблем (где учет внешнего шума/наблюдателя за счет модели «черного ящика» и построение его наблюдаемой динамики происходит из реального/численного эксперимента) доказана оптимальность наивной байесовской модели, используемой при поиске количественных соотношений структура-свойство в молекулярной физике (Quantitative Structure-Activity Relationship (QSAR))
 - (b) Для четвертой группы проблем непредсказуемой динамики, где учет внешнего шума/наблюдателя происходит за счет интуиции наблюдателя, определены границы применимости научных методов предсказания систем и роль интуиции наблюдателя для непредсказуемых систем; приведены примеры непредсказуемых систем
- 3) Развита основы термодинамики, позволяющие решить базовые парадоксы современной физики на базе новой кибернетики, учитывающей влияния внешнего шума со стороны внешней среды/наблюдателя:

- (a) Противоречие между ростом энтропии и обратимым характером физических законов (парадоксы Пуанкаре и Лошмидта)
- (b) Парадокс Шредингеровского Кота (или парадокс редукции волнового пакета)
- (c) Информационный парадокс черных дыр
- (d) Парадокс дедушки для временных кротовых нор

Библиография

- [1] Norbert Wiener, "Cybernetics Or Control and Communication in the Animal and the Machine", USA, MIT Press, 1965, p. 212
- [2] Heinz von Foerster (2003), Understanding Understanding: Essays on Cybernetics and Cognition, New York : Springer-Verlag, 362pp
- [3] Thomas Marlowe, "Systemics and Requirements: A Missing Dimension?" , WMSCI 2013, 9-12 July, Orlando, Florida, USA
- [4] Louis H. Kauffman, "Circularity, Topology and Cybernetics: Second Order Science", WMSCI 2013, 9-12 July, Orlando, Florida, USA
- [5] Karl H. Muller, "Unfolding and Expanding Science with the Help of the New Science of Cybernetics (NSC)", WMSCI 2013, 9-12 July, Orlando, Florida, USA
- [6] Oleg Kupervasser, "The Universal Arrow of Time is a Key for the Solution of the Basic Physical Paradoxes", EJTP, Vol 10, No 29, July 2013
- [7] Einstein A. (1916 (translation 1920)), Relativity: The Special and General Theory, New York: H. Holt and Company
- [8] Waldrop, M. Mitchell (1992). Complexity: The Emerging Science at the Edge of Order and Chaos. New York: Simon & Schuster.
- [9] Paul Glendinning, "Stability, Instability and Chaos. An Introduction to the Theory of Nonlinear Differential Equations", Cambridge University Press, 1994
- [10] Licata, I., "Emergence and Computation at the Edge of Classical and Quantum Systems". In: Licata, I. , Sakaji, A. Eds. Physics of Emergence and Organization, World Scientific, 2008:1-25
- [11] Susskind, Leonard (1995). "The World as a Hologram". Journal of Mathematical Physics 36 (11): 6377–6396. arXiv:hep-th/9409089
- [12] Karl Pribram, "Holonomic Brain Theory," Scholarpedia, 2(5):2735, 2007, http://www.scholarpedia.org/article/Holonomic_brain_theory
- [13] Abdi. H., & Williams, L.J. (2010). "Principal component analysis.". Wiley Interdisciplinary Reviews: Computational Statistics, 2: 433–459
- [14] Robinson-Riegler, Bridget Robinson-Riegler, Gregory. "Cognitive psychology : applying the science of the mind", (3rd ed. ed.). Boston: Pearson Allyn & Bacon
- [15] T. Grandon Gill, "Interdisciplinary Research, Education, and Communication through Case Studies and Methodologies", WMSCI 2013, 9-12 July, Orlando, Florida, USA
- [16] С.Л. Баженов, А.А. Берлин, А.А. Кульков, В.Г. Ошмян, Полимерные композиционные материалы. Прочность и технология, Издательский дом ИНТЕЛЛЕКТ, Долгопрудный, 347 стр., 2010

- [17] Kupervasser Oleg, Natalia Wanner « Continuum Solution Models for Computer Aided Drug Design», Aracne, 2015
<http://www.aracneeditrice.it/aracneweb/index.php/pubblicazione.html?item=9788854884922>
- [18] Kupervasser Oleg, «Pole Solutions for Flame Front Propagation», Springer (Mathematical and Analytical Techniques with Applications to Engineering), Springer 2015
<http://www.springer.com/us/book/9783319188447>
- [19] Kupervasser Oleg, «Universal Arrow of Time and Basic Paradoxes of Physics», Aracne, 2014,
<http://www.aracneeditrice.it/aracneweb/index.php/pubblicazione.html?item=9788854879027>
- [20] Купервассер О.Ю., Жабин С.Н., Мартынов Я.Б., Федулов К.М., Офёркин И.В., Сулимов А.В., Сулимов В.Б.// “Континуальная модель растворителя: программа DISOLV – алгоритмы, реализация и валидация”, Вычислительные методы и программирование. 2011. Т.12. С. 246-261
- [21] Kupervasser O.Yu., Wanner N.E, Continual Model of Medium I: an Algorithm for the Formation of a Smooth. Molecular Surface, Журнал «Наука и бизнес: пути развития» (Journal "Science and business: development ways"), №11(17), 2012
- [22] Yu S., Geng W., Wei G.W. //J Chem Phys. 2007. V.126 P.244108
- [23] Yu S, Geng W, Wei G. W. //J Chem Phys. 2007. V.127 P.114106
- [24] Hou S. , Wang W. , Wang L. // Journal of Computational Physics. 2010. V.229. P.7162
- [25] Connolly M.L.//J. Mol. Graphics. 1993. V. 11. P.139
- [26] Sanner M. F., Olson A. J.,Spehner J. C.// Biopolymers. 1996 V.38, P.305-
- [27] Totrov M., Abagyan R. //Journal of structural Biology. 1996. V.116. P.138
- [28] Vorobjev Y. N.,Hermans J. //Biophysical Journal. 1997. V. 73 P.722
- [29] Жабин С.Н., Сулимов В.Б. TAGSS: Свидетельство о государственной регистрации программ для ЭВМ № 2006613753// Зарегистрировано в реестре программ для ЭВМ Федеральной службы по интеллектуальной собственности, патентам и товарным знакам 27 Октября 2006
- [30] S. Zhabin, V. Sulimov, “Construction of the adaptively triangulated smooth molecular surface: program TAGSS”, (MEPhI) Electronic Journal “Scientific Visualization”, V. 3, № 2, pp. 27-53,
<http://sv-journal.org/2011-2/03.php?lang=en>
- [31] Жабин С.Н., Сулимов В.Б. Программа построения доступной растворителю поверхности для произвольных органических молекул и интерактивный просмотр положений лигандов в активных центрах белков// Сборник материалов XIII российского национального конгресса «Человек и лекарство», 3-5 апреля 2006. С.15.
- [32] Жабин С.Н., Сулимов В.Б. Программа для визуализации и редактирования молекул «MOLRED»//Сборник материалов II всероссийской конференции «Многомаштабное моделирование процессов и структур в нанотехнологиях» , 27-29 мая 2009. С. 166

- [33] Купервассер О.Ю., Жабин С.Н., Сулимов В.Б. DISOLV: Свидетельство о государственной регистрации программ на ЭВМ № 2010612994// Зарегистрировано в реестре программ для ЭВМ Федеральной службы по интеллектуальной собственности, патентам и товарным знакам 6 мая 2010 года.
- [34] Pascual-Ahuir J.L., Silla E. // J. Comput. Chem. 1990. V.11. P.1047
- [35] Silla E., Tunon I., Pascual-Ahuir J.L. // J. Comput. Chem. 1991. V.12. P.1077
- [36] Pascual-Ahuir J.L., Tunon I., Silla E. // J. Comput. Chem. 1994. V.15. P.1127
- [37] Cossi M., Mennucci B., Gammi R. // Journal of Computational Chemistry. 1996. V.17. P.57
- [38] Vorobjev Y. N., Hermans J. // Biophysical Journal. 1997. V. 73 P.722
- [39] Bordner A.J., Cavasotto C.N., Abagyan R.A. // J. Phys. Chem. B. 2002. V.106 P.11009
- [40] Bondi A. // J. Phys. Chem. 1964. V.68 P.441
- [41] Cheng A., Best S.A., Merz Jr K.M., Reynolds C.H. // Journal of Molecular Graphics and Modelling. 2000. V.18. P.273
- [42] Scharlin P., Battino R., Silla E., Tunon I., Pascual-Ahuir J.L. // Pure and App. Chem. 1998. V. 70 P. 1895
- [43] Hermann R.B. // J. Phys. Chem. 1972. V.76. P.2754
- [44] Connolly M.L. // Science. 1983. V.221. P.709
- [45] Tomasi J., Persico M. // Chem. Rev. 1984 V.94 P.2027
- [46] Rinaldi D. , Rivail J.L. // Theor. Chim. Acta. 1973. V.32 P.57
- [47] Rivail J.L., Rinaldi D. // Chem Phys. 1976. V.18. P.233
- [48] Harrison S. W., Nolte N.J., Beveridge D. // J. Phys. Chem. 1976 V.80. P.2580
- [49] Felder C.E. // J. Chem. Phys. 1981. V.75 P.4679
- [50] [Rivail J.L., Terryn B. // Journal de Chimie Physique et de Physico-Chimie Biologique. 1982. V.79 P.1
- [51] Rinaldi, D. , Ruiz-Lopez M.F., Rivail J.L. // J. Chem. Phys. 1983. V.78. P.834
- [52] Kirkwood J.G. // J. Chem. Phys. 1934. V.2. P.351
- [53] Giard J., Macq B. // Int. J. Biomed. Imaging. 2010. V.2010 P.1
- [54] Varshney A., Brooks F.P. // IEEE Proceedings of the 4th conference on Visualization, 1993. P.300
- [55] Pan Q., Tai X.-C. // Commun. Comput. Phys. 2009. V.6 P.777
- [56] Grant J., Pickup B. // Journal of Phys Chem. 1995. V.99 P.3503

- [57] Can T., Chen C., Wang Y. //J. Mol. Graph. Model. 2006. V.24. P.442
- [58] Pan Q., Tai X.-C. // Commun. Comput. Phys. 2009. V.6 P.777
- [59] Richards E.M.// Ann. Rev. Biophys. Bioeng. 1977. V.6 P.151
- [60] Connolly M.L. //J. Appl. Crystallogr. 1983. V.16 P.548
- [61] Greer J., Bush B.L.// In Proceedings of the National Academy of Sciences USA. 1978 V. 75. P.303
- [62] Liang J., Edelsbrunner H., Fu P., Sudhakar P.V., Subramaniam S.// Proteins: Structure, Function, and Genetics. 1998. V.33. P.1
- [63] Ryu J., Park R., Cho C.-H., Kim D.-S. // KSIAM Meetings 2005, http://www.ksiam.org/conference/annual052/upfile/abstract_2.pdf
- [64] Bhat S., Purisima E.O. //Proteins: Structure, Function, and Bioinformatics, 2006. V.62. P.244
- [65] Купервассер О.Ю., Жабин С.Н., Сулимов В.Б., Континуальные модели растворителя. Подсчет свободной энергии растворения//Сборник материалов II всероссийской конференции: Многомаштабное моделирование процессов и структур в нанотехнологиях , 27-29 мая 2009. С. 230.
- [66] Edelsbrunner H. // Discrete and Computational Geometry. 1999. V.21. P. 87
- [67] Bates, P. W., Wei, G. W. and Zhao, S.//J. Comput. Chem. 2008. V.29. P.380
- [68] Kupervasser O.Yu., Wanner N.E, Continual Model of Medium II: a Universal Adaptive Algorithm for the Triangulation of a Smooth Molecular Surface, Журнал «Глобальный научный потенциал» (Journal "Global Scientific Potential"), №11(20),2012
- [69] Cheng H.-L., X. Shi X. //In Proceedings of the conference on IEEE Visualization, 2004. P.481
- [70] Laug P., Borouchaki H.// Engineering with Computers. 2002. V. 18. P.199
- [71] Connolly M.L.// J. Appl. Cryst. 1985. V.18. P.499
- [72] Akkiraju N., Edelsbrunner H.// Journal Discrete Applied Mathematics - Special volume on computational molecular biology. 1996. V. 71. P.5
- [73] Akkiraju N. ,Advisers: Edelsbrunner N. Doctoral Dissertation, Molecule surface triangulation from alpha shapes, USA:University of Illinois at Urbana-Champaign Champaign, 1996
- <http://portal.acm.org/citation.cfm?id=924371>
- [74] Ryu J., Park R., Seo J., Kim C., Lee H.C., Kim D.-S.// ICCSA'07 Proceedings of the international conference on computational science and its applications, 2007. V.Part I. P.55-67
- [75] Cheng H.-L., Dey T.K., Edelsbrunner H., Sullivan J.//SODA '01 Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms: Society for Industrial and Applied Mathematics Philadelphia,2001. P.47

- [76] Ju T., Losasso F., Schaefer S., Warren J. //Proc. ACM Int.Conf. Computer Graphics and Interactive Techniques (SIGGRAPH '02), 2002. V.21. P.339
- [77] Lorensen W., Cline H.// Proc. ACM Int.Conf. Computer Graphics and Interactive Techniques (SIGGRAPH '87), 1987. V.21 P.163
- [78] Zhang Y., Xu G., Bajaj C. //Comput. Aided Geom. Des, 2003. Vol. 23. P.510
- [79] Frey P.J., Geotge P.-L. Mesh Generation. Application to Finite Elements. Wiley-ISTE , 2008. P.850
- [80] Lohner R. // Communications in Numerical Methods in Engineering 1996. V.12. P.683
- [81] Moller P., Hansbo P.// International Journal for Numerical Methods in Engineering 1995. V.38. P.3551
- [82] Cuiliere J. C.// Computer Aided Design 1998. V.30. P.139
- [83] [101]Tristano J.R., Owen S.J., Canann S.A. // 7th International Meshing Roundtable: Sandia National Labs, 1998. P.429
- [84] Guan Z., Shan J., Zheng Y., Gu Y.// Int. J. Numer. Methods Eng. 2008. V.74. P.642
- [85] Hilton A., Stoddart A.J., Illingworth J., Windeatt T.// IEEE International Conference on Image Processing, 1996. P.381
<ftp://ftp.ee.surrey.ac.uk/pub/vision/papers/hilton-icip96.ps.Z>
- [86] Hartmann E.// Vis. Comput. 1998. V.14. P.95
- [87] Karkanis T., Stewart A.J.// IEEE Computer Graphics and Applications. 2001. v.21. P.60
- [88] Minxin Chen M., Lu B. // J. Chem. Theory Comput. 2011. V.7. P.203
- [89] Isenburg M., Lindstrom P., Gumhold S., Snoeyink J. // Proceedings of IEEE Visualization: Seattle, Washington, 2003 P.465
- [90] Blacker T., Stephenson M. // Int. J. Numerical Methods in Eng. 1991. V.32. P.811
- [91] Kupervasser O.Yu., Wanner N.E, Continual Model of Medium III: Calculation of analytical gradients of parameters of surface meshes on the molecular surfaces over atomic coordinates, Журнал «Перспективы науки» (Journal "Science Prospects"), №11(38), 2012
- [92] Kupervasser O.Yu., Wanner N.E, Continual Model of Medium IV: Calculation of analytical gradients for solvation energy over atomic coordinates, «Перспективы науки» (Journal "Science Prospects"), №12(39), 2012
- [93] Basilevsky M.V., Leontyev I.V., Luschekina S.V., Kondakova O.A., Sulimov V.B. // J. Comp. Chem. 2006. V. 27. P. 552
- [94] Klamt A., Schuurmann G. // J. Chem. Soc. Perkin Trans 2. 1993. P.799-805.

- [95] Truong T.N., V. Stefanovich E.V.// J. Chem. Phys. 1995. V.103. P.3709
- [96] Cammi R., J. Tomasi J.//J. Chem. Phys. 1994.V.100. P.7495
- [97] Cammi R., J. Tomasi J.// J. Chem. Phys. 1994. V.101. P.5900
- [98] Cammi R., J. Tomasi J.//Journal of Computational Chemistry 1995. V.16. P.1449
- [99] Totrov M., Abagyan R. // Biopolymers (Peptide Science). 2001. Vol. 60. P.124
- [100] Pomelli S. , Tomasi J.// Theor Chem Acc. 1997. V.96 P.39
- [101] Romanov A.N., Jabin S.N., Martynov Y.B., Sulimov A.V., Grigoriev F.V., Sulimov V.B. // J. Phys. Chem. A. 2004. V.108. N. 43. P. 9323
- [102] Totrov M.// J. Comp. Chem. 2004. V.25 P.609
- [103] Halgren T.A. // J. Comp. Chem. 1996. V.17 P.490
- [104] Halgren T.A. // J. Comp. Chem. 1996. V.17 P.520
- [105] Halgren T.A. // J. Comp. Chem. 1996. V.17 P.553
- [106] О. Ю. Купервассер, И. П. Кикоть «Укрупненные поверхностные элементы и условия нормировки для столбцов и строк матриц метода COSMO» 2011, Т. 30, N. 10, С. 86-93
- [107] Shewchuk J.R. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. Technical report. Carnegie Mellon University: School of Computer Science, 1994.
- [108] Беклемишев Д.В. Курс аналитической геометрии и линейной алгебры. М.: Наука, 1984.
- [109] Ф. В. Григорьев, О.Ю. Купервассер, И. П. Кикоть «Итерационный метод расчета полярной составляющей энергии Гиббса растворения молекул при условии плавного изменения диэлектрической проницаемости раствора», Химическая физика, 2011, Т. 30, N. 12, С. 76-82
- [110] Basilevsky M.V., Grigoriev F.V., Kupervasser O.Yu. // J. Phys. Chem. B. 2010. V.114. P.16427
- [111] Basilevsky M.V., Grigoriev F.V., Nikitina E.A., Leszczynski J.// J. Phys. Chem. B. 2010. V.114. P.2457.
- [112] Im W., Beglov D., Roux B. // Computer Physics Communications, 1998. V.111.P.59
- [113] Воротинцев М.А. , Корнышев А.А. Электростатика сред с пространственной дисперсией. М.: Наука, 1993.
- [114] Kharkats Y., Kornyshev A.A., Vorotyntsev M.A.// J. Chem. Soc., Faraday Trans. 2. 1976. V. 72. P. 361.
- [115] Frisch M.J., Trucks G.W., Schlegel H.B. et al. Gaussian 03 (Revision B.03); Pittsburgh PA: Gaussian Inc, 2003.

- [116] Chudinov G.E., Napolov D.V., Basilevsky M.V. // Chem. Phys. 1992. V.160. P.41.
- [117] Jorgensen W.L., Ulmschneider J.P., Tirado-Rivers J. // J. Phys. Chem. B. 2004. V.108. P.16264.
- [118] A. Ben-Naim, Solvation Thermodynamics (Plenum Press, NY, 1987).
- [119] О. Купервассер, З. Олами Случайный шум и динамика полюсов при неустойчивом распространении фронта, «Физика горения и взрыва», 2013, Т. 49, № 2, С. 20-33
- [120] Liberman M.A., Ivanov M.F., Peil O.E., Valiev D.M., Eriksson L.-E. Self-acceleration and fractal structure of outward freely propagating flames // Physics of Fluids. - 2004. - V. 16, N 7. - PP. 2476-2482
- [121] Gostintsev Yu.A., Istratov A.G., Shulenin Yu.V. Self-similar propagation of a free turbulent flame in mixed gas mixtures // Combust. Expl. Shock Waves. - 1988.- V. 24, N 5. - PP. 63-70
- [122] Wu F., Jomaas G., Law C.K. On Self-Acceleration of Cellular Spherical Flames // Fall Technical Meeting of the Eastern States Section of the Combustion Institute Hosted by the University of Connecticut, Storrs, CT - 2011. - <http://lcg.princeton.edu/publications/conference-publications/wu-on-self-acceleration-of-cellular-spherical-flames.aspx>
- [123] Sivashinsky G.I. Nonlinear analysis of hydrodynamics instability in laminar flames I. Derivations of basic equations Nonlinear analysis of hydrodynamics instability in laminar flames I. Derivations of basic equations // Acta Astronautica - 1977. - V. 4 - PP. 1177-1206
- [124] Filyand L., Sivashinsky G.I., Frankel M.L. On self-acceleration of outward propagating wrinkled flames // Physica D - 1994. - V. 72 - PP. 110-118
- [125] Olami Z., Galanti B., Kupervasser O., Procaccia I. Random Noise and Pole-Dynamics in Unstable Front Propagation // Phys. Rev. E. - 1997. - V. 55, N 3 - PP. 2649-2663
- [126] Kupervasser O., Olami Z., Procaccia I. Geometry of Developing Flame Fronts: Analysis with Pole Decomposition // Phys. Rev. Lett. - 1996. - V. 76 - PP. 146-149
- [127] B. Galanti, Kupervasser O., Olami Z., Procaccia I. Dynamics and Wrinkling of Radially Propagating Fronts Inferred from Scaling Law in Channel Geometries // Phys. Rev. Lett. - 1998. - V. 80, N 11 - PP. 2477-2480
- [128] Kupervasser O., Olami Z., Procaccia I. Stability Analysis of Flame Fronts: Dynamical Systems Approach in the Complex Plane // Phys. Rev. E - 1999. - V. 59, N 3 - PP. 2587-2593
- [129] Kupervasser O. Pole solutions in the case of problems of flame front propagation and Saffman-Taylor "finger" formation without surface tension: open problems and possible ways of their solutions - 2003. - arXiv:nlin/0306038
- [130] Kupervasser O. Random Noise and Pole-Dynamics in Unstable Front Propagation - 2003. - arXiv:nlin/0302017
- [131] Karlin V., Sivashinsky G. The Rate of Expansion of Spherical Flames // Combustion Theory and Modelling - 2006. - V. 10, N 4 - PP. 625-637

- [132] Karlin V., Sivashinsky G. Asymptotic modelling of self-acceleration of spherical flames // Proceedings of the Combustion Institute - 2007. - V. 31, N 1 - PP. 1023-1030
- [133] Lee Y.C., Chen H.H. Nonlinear dynamical models of plasma turbulence // Phys. Scr. (Sweden) - 1982.- T2A - PP. 41-47
- [134] O. Thual, U.Frisch and M. Henon Application of pole decomposition to an equation governing the dynamics of wrinkled flame fronts // J. Physique - 1985. - V. 46, PP. 1485-1494
- [135] Joulin G. On the hydrodynamic stability of curved premixed flames // J. Phys. France - 1989. - V. 50 - PP. 1069–1082
- [136] Joulin G. On the Zhdanov-Trubnikov equation for premixed flame instability // J. Exp.Theor. Phys. -1990.-V. 100, Is.2(8) - PP. 428-432
- [137] Minaev S.S. Set of steady solutions describing a cellular flame in the case of hydrodynamic instability // Combustion, Explosion, and Shock Waves - 1992. - V. 28, N 1 - PP. 35–39
- [138] Minaev S.S., Pirogov E.A., Sharypov O.V. A Nonlinear model of hydrodynamic instability of outward propagating flame // Combustion, Explosion, and Shock Waves - 1996. - V. 32, N 5 - PP. 481-488
- [139] Fursenko R.V., Pan K.L., Minaev S.S. Noise influence on pole solutions of the Sivashinsky equation for planar and outward propagating flames // PHYS. REV.E - 2008. - V. 78, N 5 - P. 056301
- [140] Pan K.L., Fursenko R.V. Characteristics of cylindrical flame acceleration in outward expansion // PHYSICS OF FLUIDS - 2008. - V. 20, N 9 - P. 094107
- [141] Fursenko R.V., Minaev S.S., Pan K.L. Hydrodynamic Instability of Inward-Propagating Flames // Combustion, Explosion, and Shock Waves - 2009. - V. 45, N 5, PP. 511-517
- [142] Gutman S., Sivashinsky G.I. The cellular nature of hydrodynamic instability// Physica D - 1990. - V. 43 PP. 129-139
- [143] Risken H. The Fokker -Planck Equation. - Berlin: Springer, 1984 (P.124 Eq.(5.111))
- [144] Cambray P., Joulin G. Length-Scales of Wrinkling of Weakly-Forced, Unstable Premixed Flames // Combust. Sci. Tech. - 1994. - V. 97 PP. 405–428
- [145] Rahibe M., Aubry N., Sivashinsky G.I., Lima R. Formation of wrinkles in outwardly propagating flames // Phys. Rev. E - 1995. - V. 52, N 4 - PP. 3675–3686
- [146] Rahibe M., Aubry N., Sivashinsky G.I. Stability of pole solution for planar propagating flames // Phys. Rev. E - 1996. - V. 54, N 5 - PP. 4958-4972
- [147] Rahibe M., Aubry N., Sivashinsky G.I. Instability of pole solutions for planar propagating flames in sufficiently large domains // Combust. Theory Modelling - 1998. - V. 2, N 1 - PP. 19 – 41
- [148] Ashurst Wm.T. Darrieus-Landau instability, growing cycloids and expanding flame acceleration // Combust. Theory Modelling - 1997. - V. 1 - PP. 405–428

- [149] Kortsarts Y., Brailovsky I., Sivashinsky G.I. On Hydrodynamic Instability of Stretched Flames // *Combust. Sci. Tech.* - 1997. - V. 123 - PP. 207–225
- [150] O. Kupervasser, “Laplacian growth without surface tension in filtration combustion: Analytical pole solution”, *Complexity*, Early View (Online Version of Record published before inclusion in an issue),
<http://onlinelibrary.wiley.com/doi/10.1002/cplx.21627/abstract;jsessionid=C41E33196B794A993BDF306828272628.f03t01>
- [151] Pelce P. *Dynamics of Curved Fronts.* - Boston: Academic press, 1988.
- [152] Aldushin A.P., Matkowsky B.J. Instabilities, Fingering and the Saffman - Taylor Problem in Filtration Combustion // *Combust. Sci. and Tech.* – 1998. - V. 133. – P. 293-341.
- [153] Aldushin A.P., Matkowsky B.J. Selection in the Saffman-Taylor Finger Problem and the Taylor-Saffman Bubble Problem Without Surface Tension // *Appl. Math. Lett.* – 1998. - V. 11, N 6. - P. 57-62.
- [154] Aldushin A.P., Matkowsky B.J. Extremum principles for selection in the Saffman–Taylor finger and Taylor–Saffman bubble problems // *Phys. Fluids.* – 1999. - V. 11, N 6. - P. 1287-1296.
- [155] Chapman S.J. On the role of Stokes lines in the selection of Saffman-Taylor fingers with small surface tension // *Eur. J. Appl. Math.* – 1999. - V. 10, N 6. - P. 513-534.
- [156] Combescot R., Dombre T., Hakim V., Pomeau Y. Shape selection of Saffman-Taylor fingers // *Phys. Rev. Lett.* – 1986. - V. 56, N19.- P. 2036–2039.
- [157] Combescot R., Hakim V., Dombre T., Pomeau Y., Pumir A. Analytic theory of the Saffman-Taylor fingers // *Phys. Rev. A* – 1988. - V.37, N 4. - P. 1270–1283.
- [158] Hong D.C., Langer J.S. Analytic theory of the selection mechanism in the Saffman- Taylor problem // *Phys. Rev. Lett.* – 1986. - V.56. - P. 2032–2035.
- [159] Shraiman B.I. Velocity selection and the Saffman-Taylor problem // *Phys. Rev. Lett.* – 1986. - V.56. - P. 2028–2031.
- [160] Chapman S.J., King J. R. The selection of Saffman-Taylor fingers by kinetic undercooling // *Journal of Engineering Mathematics* –2003. - V.46, N 1 - P. 1-32.
- [161] Tanveer S. Surprises in viscous fingering // *J. Fluid Mech.* –2000. - V.409. - P. 273–308.
- [162] Abanov A., Mineev-Weinstein M., Zabrodin A. Multi-cut solutions of Laplacian growth // *Physica D: Nonlinear Phenomena* – 2009 - V. 238 (17) – P.1787–1796
- [163] Mineev-Weinstein M., Kupervasser O. Formation of a Single Saffman-Taylor Finger after Fingers Competition An Exact Result in the Absence of Surface Tension // *On 82nd Statistical Mechanics Meeting, Rutgers University.* - 1999 December 10-12; the annotation is published in *Journal of Statistical Physics* – 2000. - V.99, N3-4. - P. 1057-1065, <http://arxiv.org/abs/patt-sol/9902007>

- [164] Feigenbaum M., Procaccia I., and Davidovich B. Dynamics of Finger Formation in Laplacian Growth without Surface Tension // J. Stat. Physics. –2001. - V.103. - P. 973 -1007.
- [165] Dawson P. S., Mineev-Weinstein M. Long-Time Behavior of the N-finger Solution of the Laplacian Growth Equation // Physica D. –1994. - V.73. - P. 373-387
- [166] Siegel M., Tanveer S. Singular Perturbation of Smoothly Evolving Hele-Shaw Solutions // Phys. Rev. Lett. –1996. - V.76. - P. 419-422
- [167] Mineev-Weinstein M. Selection of the Saffman-Taylor Finger Width in the Absence of Surface Tension: An Exact Result // Phys. Rev. Lett. –1998. - V.80. - P. 2113-2116.
- [168] Sarkissian A., Levine H. Comment on “Selection of the Saffman-Taylor finger width in the absence of surface tension: An exact result” // Phys. Rev. Lett. –1998. - V.81. - P. 4528-4528.
- [169] Mineev-Weinstein M. A Reply to the Comment by Armand Sarkissian and Herbert Levine // Phys. Rev. Lett. –1998. - V.81. - P. 4529-4529
- [170] Casademunt J., Magdaleno F. X. Comment on “Selection of the Saffman-Taylor Finger Width in the Absence of Surface Tension:An Exact Result” // Phys. Rev. Lett. –1998. - V.81. - P. 5950-5950.
- [171] Almgren R.F. Comment on “Selection of the Saffman-Taylor Finger Width in the Absence of Surface Tension:An Exact Result” // Phys. Rev. Lett. –1998. - V.81. - P. 5951-5951.
- [172] Mineev-Weinstein M. A Reply to the Comment by J. Casademunt and F. X. Magdaleno, and also R.F. Almgren // Phys. Rev. Lett. –1998. - V.81. - P. 5952-5952
- [173] Kessler D. A., Levine H. Microscopic Selection of Fluid Fingering Patterns // Phys. Rev. Lett. –2001. - V.86. - P. 4532-4535
- [174] E. Paun' e, F.X. Magdaleno and J. Casademunt Dynamical Systems approach to Saffman-Taylor fingering. A Dynamical Solvability Scenario // Physical Review E – 2002 - V. 65 (5) - P. 056213
- [175] Mark Mineev-Weinstein, Gary D. Doolen, John E. Pearson, Silvina Ponce Dawson Formation and Pinch-off of Viscous Droplets in the Absence of Surface Tension: an Exact Result // <http://arxiv.org/abs/patt-sol/9912006>
- [176] Kupervasser O., The mysterious optimality of Naive Bayes: Estimation of the probability in the system of “classifiers”, Pattern Recognition and Image Analysis, 24 (1), pp. 1-10 (2014), arXiv:cs/0202020v1, (2002)
<http://arxiv.org/abs/cs/0202020v1>
- [177] V. Ramensky, A. Sobol, N. Zaitseva, A. Rubinov, V. Zosimov, A novel approach to local similarity of protein binding sites substantially improves computational drug design results, Proteins: Structure, Function, and Bioinformatics, , 69(2), pp 349–357 (2007)
- [178] S. Nikitin¹, N. Zaitseva, O. Demina, V. Solovieva, E. Mazin, S. Mikhalev, M. Smolov, A. Rubinov, P. Vlasov, D. Lepikhin, D. Khachko, V. Fokin, C. Queen, V. Zosimov, A very large diversity space of synthetically accessible compounds for use with drug design programs, Journal of Computer-Aided Molecular Design, 19, pp 47–63 (2005)

- [179] Raymer M. L., Doom T. E., Kuhn L. A., Punch W. F., "Knowledge discovery in medical and biological datasets using a hybrid bayes classifier/evolutionary algorithm," IEEE Transactions on Systems, Man, and Cybernetics, 33B, 802 (2003)
- [180] Domingos, P., and Pazzani, M., On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning 29, 103 (1997)
- [181] Zhang H., The Optimality of Naive Bayes, In FLAIRS Conference (2004) <http://www.cs.unb.ca/profs/hzhang/publications/FLAIRS04ZhangH.pdf>
- [182] Kuncheva L.I., On the optimality of Naive Bayes with dependent binary features, Pattern Recognition Letters, 27,830 (2006)
- [183] Landau L.D., Lifshitz E.M., Statistical Physics, 5,Elsevier Science Technology, United Kingdom, (1996)
- [184] Oleg Kupervasser, Hrvoje Nikolic, Vinko Zlatic "The Universal Arrow of Time", Foundations of Physics 42, 1165-1185 (2012) , arXiv:1011.4173 <http://www.springerlink.com/content/v4h2535hh14uh084/>
- [185] Karl Popper Logik der Forschung, 1934 (The Logic of Scientific Discovery, English translation 1959)
- [186] Hrvoje Nikolic, "Closed timelike curves, superluminal signals, and "free will" in universal quantum mechanics", arXiv:1006.0338
- [187] Oleg Kupervasser "The basic paradoxes of statistical classical physics and the quantum mechanics", "Universal Journal of Physics and Application" Vol 1, No 3,Nov 2013,pp. 299 - 349 <http://www.hrpub.org/download/20131107/UJPA11-18400957.pdf>
- [188] Oleg Kupervasser, Comment on "Quantum Solution to the Arrow-of-Time Dilemma" of L. Maccone, Proceedings of the 17th World Multi-Conference on Systemics, Cybernetics and Informatics, July 9-12, 2013 - Orlando, Florida, USA WMSCI 2013
- [189] Oleg Kupervasser, New Cybernetics and the Application of its Principles in Physics, Journal of Systemics, Cybernetics and Informatics, Volume 11 - Number 7 - Year 2013, Pages: 74-83
- [190] Oleg Kupervasser, Comment on "Entanglement and the Thermodynamic Arrow of Time" and Correct Reply on "Comment on "Quantum Solution to the Arrow-of-Time Dilemma"" of David Jennings and Terry Rudolph
- ["Frontiers in Science",Vol.2, No.6, December 2012](#)
- [191] O. Kupervasser, "Foundations of Nonlinear Dynamics or Real Dynamics, Ideal Dynamics, Unpredictable Dynamics and the "Schroedinger's Cat Paradox"" (2005)
- <http://arxiv.org/abs/nlin/0508025>
- [192] O. Kupervasser, "Base of nonlinear dynamics or Real Dynamics, Ideal Dynamics, Unpredictable Dynamics and "Schrodinger cat""
- <http://arxiv.org/abs/nlin/0407033>

- [193] Malinetskii, G.G. Mathematical basis of synergetics, LKI, Moscow, 2007 (in Russian)
- [194] Getling, A.V. Rayleigh-Benard Convection: Structures and Dynamics, World Scientific Publishing Company, Library Binding, Incorporated, 1997, 250 pages
- [195] Samarskii, A.A.; Galaktionov, V.A.; Kurdyumov, S.P.; Mikhailov, A.P. Blow-up in Quasilinear Parabolic Equations, Walter de Gruyter, Berlin, 1995.
- [196] L.M. Martyushev L.M., Seleznev V.D. Maximum entropy production principle in physics, chemistry and biology. Physics Reports. 2006. Vol.426, (1). P.1-45.
- [197] J. Miguel Rubi, "Does Nature Break the Second Law of Thermodynamics?", Scientific American Magazine, Oct 28, 2008, P. 66
<http://www.scientificamerican.com/article.cfm?id=how-nature-breaks-the-second-law>
- [198] Roger Penrose, The Emperor's New Mind, Oxford University Press, New York, NY, USA 1989
- [199] Roger Penrose, Shadows of the Mind, Oxford University Press, New York, NY, USA 1994
- [200] Siegelmann, H.T. Neural Network and Analog Computation: Beyond the Turing Limit, Birkhauser, 1998
- [201] Calude, C.S., Paun, G. Bio-steps beyond Turing, BioSystems, 2004, v 77, 175-194
- [202] Nicolas H. Vöelcker; Kevin M. Guckian; Alan Saghatelian; M. Reza Ghadiri Sequence-addressable DNA Logic, Small, 2008, Volume 4, Issue 4, Pages 427 – 431
- [203] Schulman, L.S., Phys. Rev. Lett. 83, 5419 (1999).
- [204] Schulman, L.S., Entropy 7[4], 208 (2005)
- [205] Valiev K.A., Kokin A.A., Quantum computers: Expectations and Reality, Izhevsk, RKhD, 2004
- [206] Introduction to quantum computation and information, eds. Hoi-Kwong Lo, Sando Popescu, Tim Spiller, Word Scientific Publishing (1998)
- [207] A.K. Dewdney, "A Pandora's box of minds, machines and metaphysics", Scientific American, Volume 261, Issue 6, December 1989
- [208] The micromaser spectrum ,Marlan O.Scully, H. Walther, Phys. Rev. A 44, 5992–5996 (1991)
- [209] Peter W. Shor, "Scheme for reducing decoherence in quantum computer memory", Phys. Rev. A 52, R2493–R2496 (1995)
- [210] George Musser, Easy Go, Easy Come. (How Noise Can Help Quantum Entanglement), Scientific American Magazine, 2009, November
<http://www.scientificamerican.com/sciammag/?contents=2009-11>
- [211] Michael Moyer, Chlorophyll Power. (Quantum Entanglement, Photosynthesis and Better Solar Cells), Scientific American Magazine, 2009, September

<http://www.scientificamerican.com/article.cfm?id=quantum-entanglement-and-photo>

[212] Jianming Cai; Sandu Popescu; Hans J. Briegel Dynamic entanglement in oscillating molecules and potential biological implications, *Phys. Rev. E* 82, 021921 (2010)

<http://arxiv.org/abs/0809.4906>

[213] V Capek and T Mancal, «Phonon mode cooperating with a particle serving as a Maxwell gate and rectifier», *J. Phys. A: Math. Gen.*, V.35, N. 9 (2002)

[214] Jung, On Synchronicity (in *Man and Time*), *Papers from the Eranos Yearbooks.3*, NY and London, 1957

[215] Helena N. Knyazeva, Kurdyumov S.P., “ Foundations of synergetics ”, part1, Moscow, “KomKniga”,2005 in Russian

[216] Helena N. Knyazeva; Kurdyumov, S.P. Foundations of synergetics, part2, KomKniga, Moscow, 2006-2007 (in Russian)

[217] Samarskii, A.A.; Galaktionov, V.A.; Kurdyumov, S.P.; Mikhailov, A.P. Blow-up in Quasilinear Parabolic Equations, Walter de Gruyter, Berlin, 1995

[218] Kapitza, S.P.; Kurdyumov, S.P.; Malinetskii, G.G. Synergetics and Prognoses of the Future, Nauka Publishers, Moscow,1997 (in Russian).

[219] Nikolsky I.M. Investigation of a spectrum of the many-dimensional dissipative structures developing in a regime with blow-up.//the Proceedings of the international conference Lomonosov-2005, Moscow, 2005, P.45-46

[220] Hogan, J. Why the Universe is Just So, *Rev.Mod.Phys.*, 2000, 72 (arxiv: astro-ph/9909295)

[221] Malinetskii, G.G. Mathematical basis of synergetics, LKI, Moscow, 2007 (in Russian)

[222] Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities, *Proceedings of National Academy of Sciences*, 1982, April, vol. 79, no. 8, pp. 2554–2558. PNAS Reprint (Abstract) PNAS Reprint (PDF)

[223] Nick Bostrom «HOW LONG BEFORE SUPERINTELLIGENCE? » [Originally published in *Int. Jour. of Future Studies*, 1998, vol. 2] , [Reprinted in *Linguistic and Philosophical Investigations*, 2006, Vol. 5, No. 1, pp. 11-30.] <http://www.nickbostrom.com/superintelligence.html>

[224] Principles of self-organisation: Transactions of the University of Illinois, Symposium on Self-Organization, June 8–9, 1961, eds. Foerster H.V., Zopf G.W. Oxford: Pergamon Press (1962)

[225] Fogel, David B. (2000). *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence*. New York: IEEE Press. pp. 140.

[226] D.A. Pospelov, SEMIOTIC MODELS IN ARTIFICIAL INTELLIGENCE PROBLEMS, International Joint Conference on Artificial Intelligence, TBILISI GEORGIA, USSR, V. 1&2,P.65 (1975)

<http://ijcai.org/Past%20Proceedings/IJCAI-75-VOL-1&2/PDF/010.pdf>

- [227] Seth Lloyd , “Ultimate physical limits to computation” , NATURE,VOL 406, P.1047-1054, (2000)
- [228] Lawrence, Jeanette “Introduction to Neural Networks”, California Scientific Software Press (1994)
- [229] Von Altrock, Constantin “Fuzzy logic and NeuroFuzzy applications explained”. Upper Saddle River, NJ: Prentice Hall PTR, (1995).
- [230] Yuchun Lee, “Handwritten digit recognition using k nearest-neighbor, radial-basis function, and backpropagation neural networks”, Journal Neural Computation, Volume 3 Issue 3, (1991)
- [231] Gonzalez, “Digital Image Processing Using MATLAB”, Woods, and Eddins Prentice Hall (2004)
- [232] Michael B. Mensky, Consciousness and Quantum Mechanics: Life in Parallel Worlds. Miracles of Consciousness from Quantum Reality, Imperial college press, P. 250 (2010)
- [233] ILASOL Conference 2006, at the Weizmann Institute of Science, What is life?,4 January (2007), http://www.hrpub.org/journals/article_info.php?aid=815
- [234] Kupervasser, O.Yu., “Grandfather paradox in non-quantum and quantum gravitation theories”, Natural Science, Vol.6, No.11, 2014 (http://www.scirp.org/journal/PaperInformation.aspx?PaperID=47897#U-n_RqNUo9k)
- [235] L. Maccone, Phys. Rev. Lett. 103 (2009) 080401, <http://arxiv.org/abs/0802.0438>.
- [236] L. Maccone L. “A quantum solution to the arrow-of-time dilemma: reply”, 2009. <http://arxiv.org/abs/0912.5394>
- [237] D. Jennings, T. Rudolph, Phys. Rev. Lett. 104 (2010) 148901.
- [238] D. Jennings, T. Rudolph, Phys. Rev. E 81 (2010) 061130.
- [239] Black hole information paradox - Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Black_hole_information_paradox
- [240] Nikodem J. Popławski “Radial motion into an Einstein–Rosen bridge”, Physics Letters B, V. 687, pp.110–113 (2010)
- [241] Joaquin P. Noyola, “Relativity and Wormholes” (2006) http://www.uta.edu/physics/main/resources/ug_seminars/papers/RelativityandWormholes.doc
- [242] M. Morris and K. Thorne. “Wormholes in spacetime and their use for interstellar travel: A tool for teaching general relativity.” Am. J. Phys., V. 56, pp. 395-412 (1988)
- [243] [Hrvoje Nikolic, “Causal paradoxes: a conflict between relativity and the arrow of time”, Foundations of Physics Letters, V.19 (3), pp. 259-267 (2006)
- [244] S. V. Krasnikov, “The time travel paradox”, Phys.Rev. D, V. 65 (2002) <http://arxiv.org/abs/gr-qc/0109029>