

Проект №19455 "Решетчатые модели макромолекул"

Пчелинцев Илья Игоревич, БПМ-192

Содержание

1	Геометрические свойства модели Ising-ISAW с точки зрения числа соседей в узлах	2
1.1	Введение	2
1.2	Особенности ранних результатов на квадратной решётке	2
1.3	Сравнение модели Изинга и полимерной цепочки в решетках с 2-6 возможными соседями номеров	2
1.4	Алгоритм исследования характера зависимости значения долей узлов от длины при $J=0$	6
1.5	Сравнение геометрических свойств модели Изинга на треугольной решётке с квадратной в $J=0$	8
1.6	Сравнение геометрических свойств модели Изинга на решётках с большим числом соседей в $J=0$	10
1.7	Число соседей и атмосферы блужданий	10
2	Геометрические свойства простого случайного блуждания	13
2.1	Алгоритм генерации блужданий	13
2.2	Результаты симуляций	14
2.3	Погрешности результатов	16
2.4	Шкалирование результатов	18
2.5	Зависимость доли уникальных узлов от количества шагов	20
2.6	Итоговое сравнение функций долей узлов n_i	23
3	Исследование поведения концов блужданий модели RW	25
3.1	Зависимость атмосфер от количества шагов блуждания и уникальных узлов	26
3.2	Общее сравнение поведений атмосфер блужданий и долей узлов простого случайного блуждания	29
3.3	Планируемая деятельность	30
4	Литературный обзор	31
4.1	Livne, Meirovich: Polymers Adsorbed on a surface	31
4.1.1	Особенности модели блуждания	31
4.1.2	Подробнее о статсумме и методе Сканирования	31
4.1.3	Результаты работы	31
4.2	Madras, Sokal: The Pivot Algorithm	32
4.2.1	Основные принципы алгоритма	32
4.3	Спицер: Теория случайных блужданий	32
5	Программно-техническое приложение	33
5.1	Применение jit-компиляции при программировании на языке Python	33
5.2	Итерации программного комплекса Rand-Walk	33

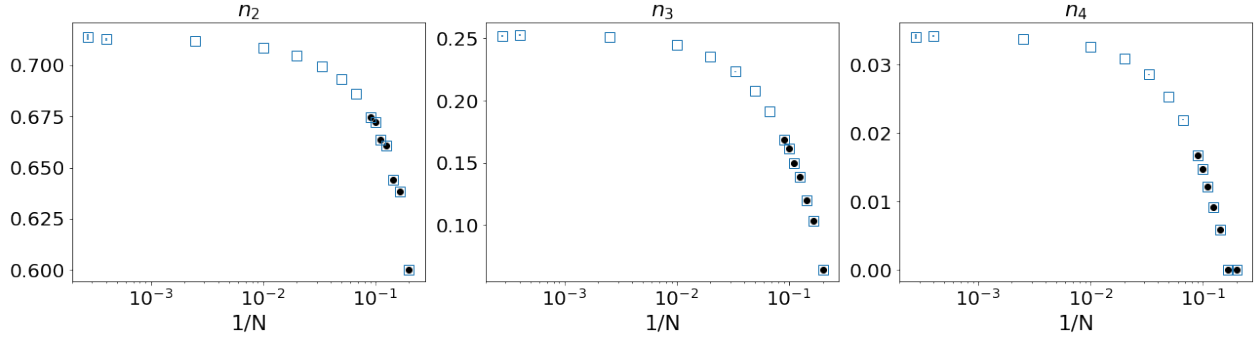


Рис. 2: Зависимостей средних долей узлов конформации с фиксированным числом соседей (от 2 до 4) модели ISAW при $J=0$ от обратной длины $1/N$ при длинах конформации $N=5-3600$. Пустые квадраты - результаты симуляций Монте-Карло, черные точки - расчёты полученные путём полного перебора возможных конформаций [2]

1 Геометрические свойства модели Ising-ISAW с точки зрения числа соседей в узлах

1.1 Введение

В данном разделе мы изучаем такое геометрическое свойство модели, как доли узлов с фиксированным числом соседей. У каждого узла можно определить число соседей или количество близких связей на смежных ячейках исследуемой решётки (см. левый рисунок 4). Рассмотрим пример конформации на квадратной решётке на рисунке 1. Чёрные точки соответствуют узлам с 2-мя соседями, а последовательность таких узлов подряд в конформации можно интерпретировать как "одномерный" участок. Узлы с тремя соседями расположены, как правило, на границах кластеров, и отображены на примере синими треугольниками, в то время как узлы с четырьмя соседями (красные квадраты) типичны для узлов в глубине кластера.

Сначала, чтобы определить правильность алгоритма расчёта долей искомым узлов, были проведены симуляции Монте-Карло модели ISAW при $J=0$ на длинах N от 5 до 3600, а так же произведены расчёты вручную для цепочек малых длин - от 5 до 11. Результаты изображены на рисунке 2 - разные типы расчётов полностью совпали, что говорит о правильности используемого алгоритма.

1.2 Особенности ранних результатов на квадратной решётке

Мы провели симуляции Монте-Карло для долей узлов с фиксированным числом соседей для моделей Ising-ISAW и ISAW с зависимостью от значения константы взаимодействия J для длин $N=1000, 2500, 3600, 4900$. Результаты изображены на рисунке 3, а также опубликованы в работе [5].

На графиках 3 примечательны значения в точке $J=0$ у графиков узлов с 2-мя (левый) и 3-мя (средний) соседями: было первоначальное предположение, что в пределе бесконечной длины конформации они будут равны $3/4$ и $1/4$ соответственно. Так же интересен вопрос универсальности данного свойства на других решётках: будут ли эти значения долей n_2 и n_3 при тех же условиях равны или хотя бы похожи в других решётках.

1.3 Сравнение модели Изинга и полимерной цепочки в решетках с 2-6 возможными соседями мономеров

Рассмотрим средние доли узлов с фиксированным числом соседей в решётках, которые имеют от 2-х до 6-ти возможных соседей: в кубической, у которой 5-й и 6-й соседи мономера расположены в соседних плоскостях,

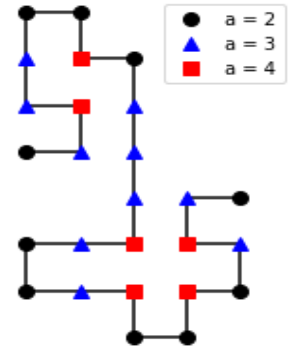


Рис. 1: Пример конформации на квадратной решётке с подсчётом соседей

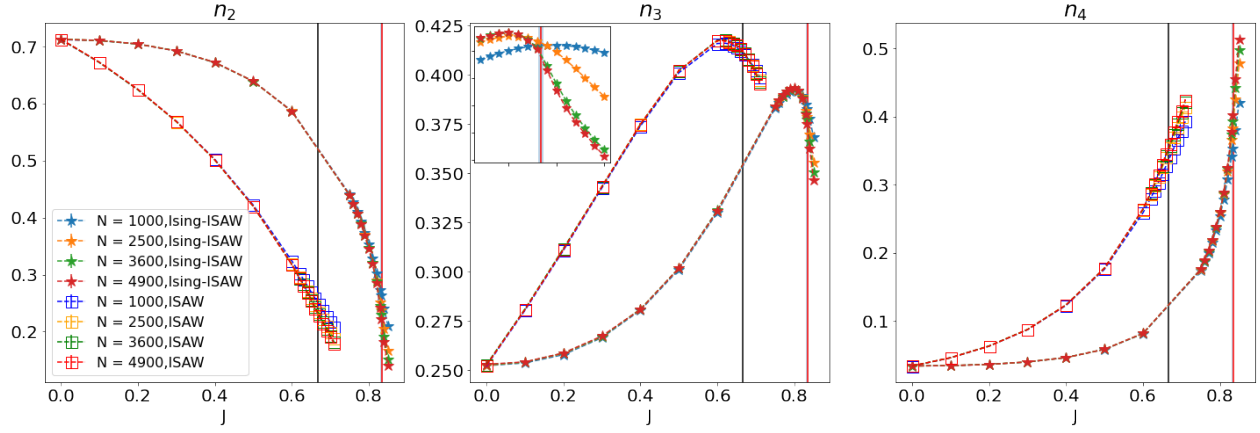


Рис. 3: Зависимость доли узлов конформации с двумя (слева), тремя (по центру) и четырьмя соседями (справа) у моделей Ising-ISAW (звезды) и ISAW на квадратной решётке от J . Черной линией обозначена точка фазового перехода модели ISAW, красной - Ising-ISAW, на квадратной решётке (см. таблицу 1). График взят из работы [5]

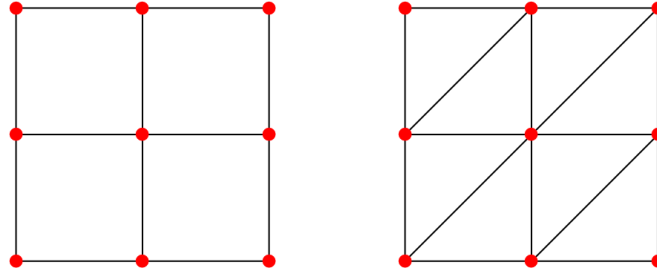


Рис. 4: Связи узлов в квадратной (слева) и треугольной решётках (справа)

и треугольной, где 5-й и 6-й сосед мономера лежат на диагонали, проходящей через данный узел (в данной решётке лишь одна плоскость, см. правый рисунок 4).

График зависимости долей от константы взаимодействия J (используется в гамильтониане конформации по формуле ??, однако в отличие от одномерного случая, где считаются связи между соседними по индексу узлами конформации, здесь считаются связи между узлами, лежащими на соседних ячейках исследуемой решётки) изображен на рисунке 5 - слева показаны результаты симуляций Монте-Карло на кубической решётке, справа - на треугольной решётке. Цвета графиков соответствуют длинам цепочек - $N=100$ зелёные, 300 синие, 600 красные и 1200 фиолетовые. Число шагов симуляций - от 10^{10} вдали от пиков до 10^{12} в районе пиков графиков. Вертикальными линиями отмечены точки критического перехода:

lattice	Ising-ISAW	ISAW
square	0.8340(5) [5]	0.6673(5) [4]
triangular	Unknown	0.41(7) [10]
cubic	0.5263 ± 0.055 [6]	0.2779 ± 0.0041 [11]

Таблица 1: Значения J критических точек фазового перехода модели Изинга на случайном блуждании (Ising-ISAW) и гомополимера (ISAW) на квадратной, треугольной и кубической решётке соответственно (в порядке строк)

Результаты симуляций модели ISAW отмечены пустыми квадратами, а модели Ising-ISAW - звёздами. Примечательно, что графики зависимости долей от J данной модели значительно плавнее, чем у модели Изинга на случайном блуждании, а так же процессы уплотнения конформаций (когда доли n_2 и n_3 уменьшаются, а доли узлов с большим числом соседей увеличивается) начинаются раньше, пропорционально значению

точки перехода J_c . Последнее, скорее всего, связано с тем, что точка перехода модели ISAW меньше, чем у Ising-ISAW (для кубической это известно, для треугольной просто предположение). Возможно, что при масштабировании левой части графиков кубической решетки относительно J_c (то есть, от $0 * J_c$ до $1 * J_c$), мы бы получили примерно одинаковые графики.

В то же время, предельные значения у данных моделей совпадают - графики одинаковых длин и решёток разных моделей исходят из одной точки при $J=0$ (что логично, ведь при $J=0$ поведение Ising-ISAW соответствует ISAW) и приходят в одну точку при $J=1$.

Данные модели Ising-ISAW в свою очередь отмечены на графике 5 звездочками. Стоит отметить, что при прохождении точки перехода в кубической решётке, графики долей узлов с любым числом соседей словно претерпевают скачок, усиливающийся с ростом длины цепочки, в отличие от треугольной решётки, где процесс непрерывен.

Говоря о свойствах Ising-ISAW кубической решётки, необходимо подчеркнуть, что в на графике $\langle n_3 \rangle$ мы видим похожее поведение в $J=0$ - значение довольно близко к 0.25, стоит проверить предел значения доли узлов с 3-мя соседями в $J=0$ при бесконечной длине и характер приближения к нему, если таковой имеется. Значение $\langle n_2 \rangle$ при $J=0$ визуально отличается от предполагаемого $3/4$. В следующих разделах мы рассмотрим развитие значения долей $\langle n_{2-6} \rangle$ в точке $J=0$ (где модели ISAW и Ising-ISAW ведут себя идентично с обычным невзаимодействующим блужданием SAW) на разных решётках на пределе бесконечной длины.

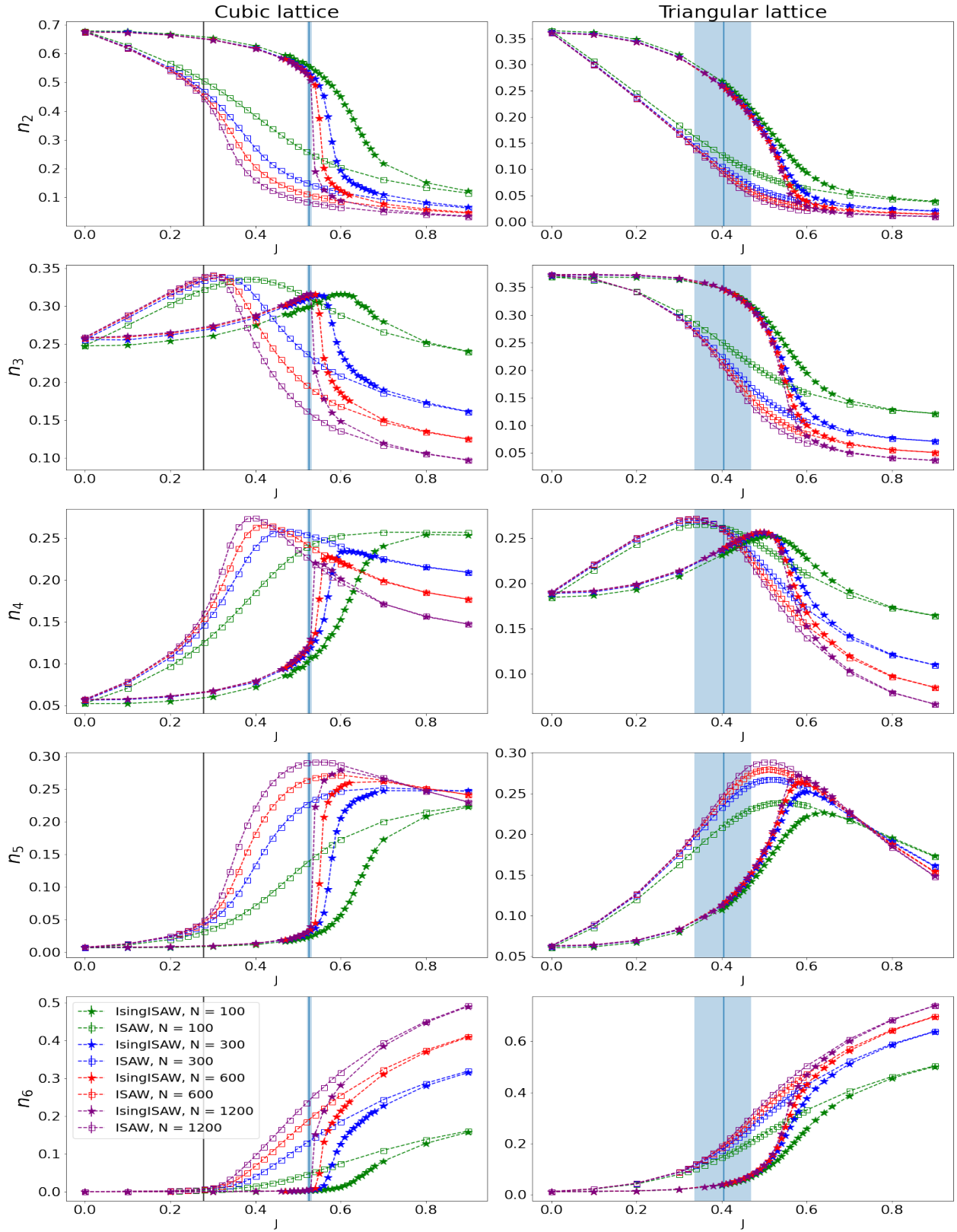


Рис. 5: Зависимость доли узлов моделей Ising-ISA (звезды) и ISA (квадраты) на кубической (слева) и треугольной решётках (справа) с 2-6 соседями (сверху вниз) от J с длинами $N = 100$ (зеленые), 300 (красные), 600 (синие) и 1200 (фиолетовые). Вертикальные линии отмечают точки фазового перехода моделей 1

1.4 Алгоритм исследования характера зависимости значения долей узлов от длины при $J=0$

Здесь рассматривается способ определения характера зависимости у графиков долей узлов с фиксированным числом соседей при $J=0$. Для примера взят случай n_2 у квадратной решётки модели Ising-ISAW. Первоначально рассматривается три возможных способа аппроксимации результатов, варьирующихся зависимостью от обратной длины конформации $x = 1/N$:

1. Линейная аппроксимация

$$y = ax + b \quad (1.1)$$

2. Лог-линейная или экспоненциальная аппроксимация

$$y = b \exp(ax) + c \quad (1.2)$$

3. Степенная или лог-лог аппроксимация

$$y = bx^a + c \quad (1.3)$$

Чтобы гарантировано получить результат использовалась функция `linregress` из пакета `scipy.stats`, поэтому на данном этапе погрешностью результатов симуляций мы временно пренебрегаем. Так же, чтобы показать нагляднее характер аппроксимации, графики соответствующих способов фитирования будут рассмотрены в том же масштабе - линейный в линейном, экспоненциальный в лог-линейном, степенном в лог-лог-масштабе - таким образом графики фитов будут линейными. Результаты аппроксимаций в порядке, изложенном в списке выше, изображены на рисунках 6a и 6b - в левом столбце аппроксимации записаны для данных цепочек с длинами от 100 до 4900, в правом - длины от 250 до 4900, чтобы оценить поведение модели на больших длинах, следовательно, ближе к нулю.

Графики на рисунках 6a и 6b показывают, что в данном случае экспоненциальная аппроксимация ведёт себя как линейная (что логично вблизи нуля), поэтому можно рассматривать вместо первых двух только линейную. С другой стороны, степенная функция совсем не совпадает с графиком результатов. Более того, значение степени функции-фита настолько мало, что итоговая функция больше похожа на константную прямую.

Таким образом, в данном случае определён линейный характер зависимости. Теперь, чтобы оценить качество приближения при рассмотрении точек всё ближе и ближе к нулю, оценим ошибку фитирования - теперь мы можем использовать функцию `curve-fit` из пакета `scipy.optimize`.

N	a	b
100-4900	-0.44(1)	0.71292(4)
250-4900	-0.473(6)	0.71299(2)
400-4900	-0.47(1)	0.71298(2)
1000-4900	-0.48(6)	0.71299(4)

Таблица 2: Значения и погрешности коэффициентов линейного фитирования (1.1) зависимости долей узлов с 2-мя соседями на квадратной решётке модели Ising-ISAW при $J=0$ от исследуемого интервала длин

Результаты использования других диапазонов точек на таблице 2 показывают, что наиболее оптимальный фит (с наименьшей ошибкой) достигается при выборе точек от 250 до 4900. Это можно объяснить тем, что при выборе точек большего диапазона линейный характер будет выражен слабее, а при выборе точек меньшего диапазона количество рассматриваемых данных уменьшается, что приводит к росту ошибки (недостаточно статистики). Подобная операция была выполнена и для других чисел соседей и решёток (более подробные графики см. в Bulk2-6.ipynb в разделе "Расчёты .ipynb" [1]), результаты представлены в следующем разделе в виде графиков для узлов с 2-мя и 3-мя соседями и в виде таблицы коэффициентов линейного фитирования (1.1) без графиков.

Результаты линейного фитирования при выборе разной наименьшей рассматриваемой длины можно увидеть на таблицах 3 и 4. По погрешностям на первых строках обеих таблиц понятно, что оптимальным диапазоном будет 250-4900. Для 3-4D-гиперкубических решёток так же заметно (по погрешностям соответствующих строк), что отбрасывание длины $N=100$ из рассматриваемых улучшило точность результатов. Единственное

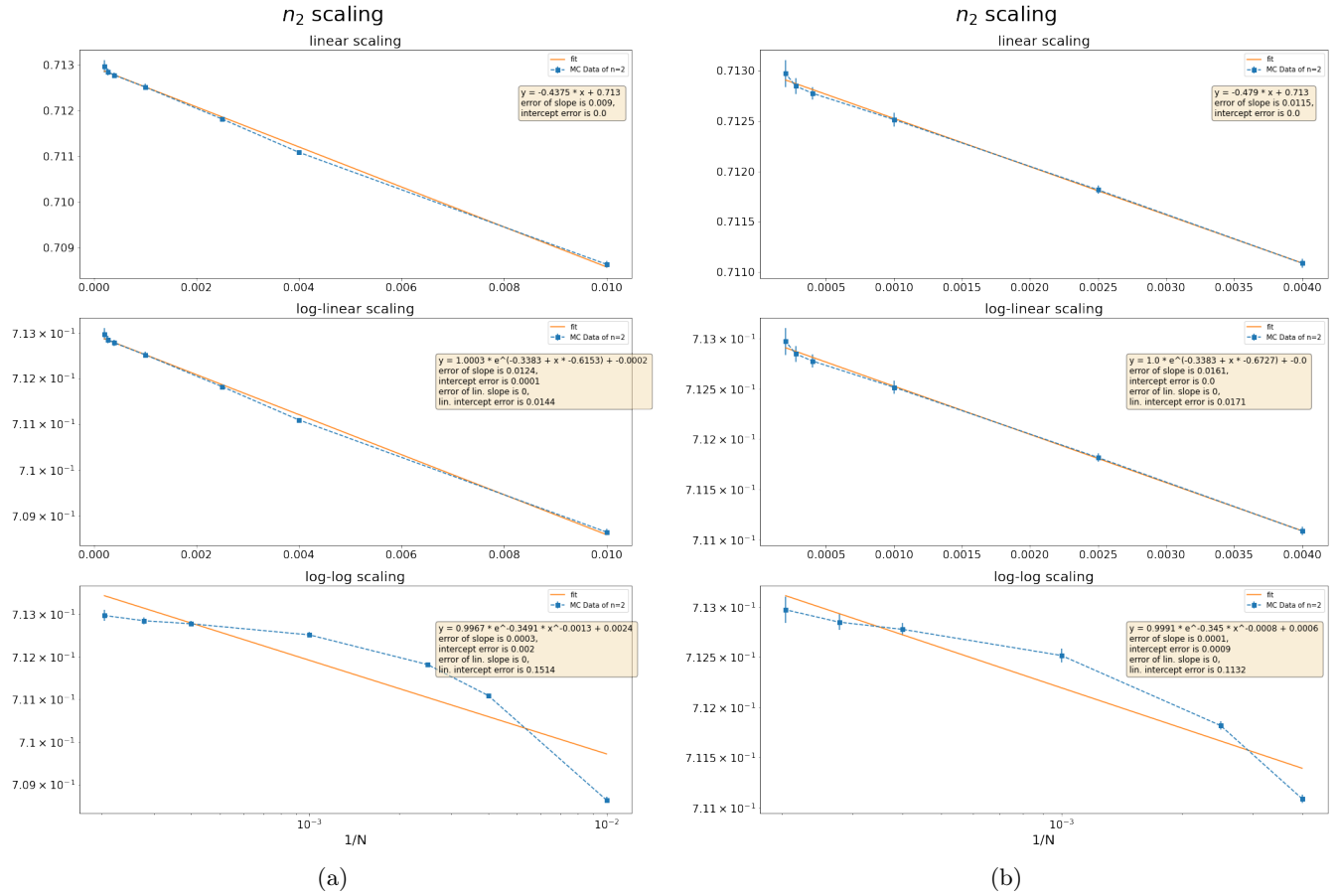


Рис. 6: Результаты аппроксимации (оранжевая линия) данных Монте-Карло о долях узлов с двумя соседями n_2 модели Ising-ISAW на квадратной решётке (синие точки) различными способами на диапазоне длин 100-4900 (a) и 250-4900 (b)

исключение - треугольная решётка: на ней линейный характер результатов настолько заметен, что при отбрасывании наименьшей длины $N=100$ ошибка увеличивается (недостаточность статистики стала сильнее, а "линейность" не изменилась).

1.5 Сравнение геометрических свойств модели Изинга на треугольной решётке с квадратной в $J=0$

На графике 7 наглядно показано сравнение приближений долей "одномерных"участков (то есть, долей мономеров с двумя соседями) и узлов с тремя соседями в цепочках на квадратной, треугольной, кубической и гиперкубической решётках. Для расчётов долей на треугольной решётке были использованы длины 100-1200, для квадратной - 250-4900, для кубической и гиперкубической - 200-1200. Фитирование долей треугольной решётки имеет отчётливый линейный характер, даже в приближении на короткие длины. Линейность долей прямоугольных решёток всех размерностей также подтверждается (с учётом погрешности расчётов с наибольшей длиной).

Lattice	$\langle n_2 \rangle$		$\langle n_3 \rangle$		$\langle n_4 \rangle$	
	a	b	a	b	a	b
Square	-0.44(1)	0.71291(4)	-0.843(8)	0.25297(3)	-0.154(3)	0.03412(1)
Triangular	0.492(2)	0.35989(1)	-0.519(3)	0.37410(1)	-0.609(4)	0.19080(1)
Cubic	0.37(2)	0.67440(7)	-1.24(1)	0.26005(5)	-0.525(5)	0.05758(1)
Hypercubic	0.15(2)	0.71978(9)	-1.20(1)	0.22080(6)	-0.468(5)	0.04589(2)

Таблица 3: Коэффициенты прямых, полученные линейным фитированием (1.1) данных симуляций Монте-карло по долям узлов с 2-4 соседями из рисунков 7 для длин N от 100 до 4900 (для квадратной) и 1200 (для остальных решёток)

Lattice	$\langle n_2 \rangle$		$\langle n_3 \rangle$		$\langle n_4 \rangle$	
	a	b	a	b	a	b
Square	-0.473(6)	0.71299(1)	-0.809(3)	0.25291(1)	-0.145(4)	0.03410(1)
Triangular	0.491(3)	0.35989(1)	-0.523(6)	0.37411(1)	-0.603(8)	0.19079(2)
Cubic	0.418(1)	0.67429(3)	-1.27(1)	0.26012(2)	-0.538(4)	0.05761(1)
Hypercubic	0.26(1)	0.71958(3)	-1.27(1)	0.22720(2)	-0.494(6)	0.04596(1)

Таблица 4: Коэффициенты прямых, полученные линейным фитированием (1.1) данных симуляций Монте-карло по долям узлов с 2-4 соседями из рисунков 7 для длин N от 250 до 4900 (для квадратной) и от 200 до 1200 (для остальных решёток)

Lattice	$\langle n_2 \rangle$			$\langle n_3 \rangle$			$\langle n_4 \rangle$		
	a	b	N	a	b	N	a	b	N
Square	-0.473(6)	0.71299(2)	250-4900	-0.809(4)	0.25291(1)	250-4900	-0.145(4)	0.03410(1)	250-4900
Triangular	0.492(2)	0.35989(1)	100-1200	-0.519(3)	0.37410(1)	100-1200	-0.609(4)	0.19080(1)	100-1200
Cubic	0.42(1)	0.67429(3)	200-1200	-1.270(7)	0.26012(2)	200-1200	-0.538(4)	0.05671(1)	200-1200
Hypercubic	0.26(1)	0.71957(3)	200-1200	-1.27(1)	0.22721(2)	200-1200	-0.494(6)	0.04595(1)	200-1200

Таблица 5: Коэффициенты прямых, полученные линейным фитированием (1.1) данных симуляций Монте-карло по долям узлов с 2-4 соседями из рисунков 7 - наилучшие приближения с подбором диапазона длин для каждого графика (в столбце N)

Из таблицы 5 по первым двум строкам, отображающим данные о прямых-фитов квадратной и треугольной решётки соответственно, сходства между одномерием треугольной и квадратной решётки с точки зрения коэффициентов фитирования a и b (1.1) почти не наблюдается - они имеют как разные значения свободных членов, так и значения и даже (в случае 2-х соседей) знаки коэффициента наклона, разница который значительно превышает погрешность фита.

Значение свободного члена b для $\langle n_2 \rangle$, то есть предела значения долей при бесконечной длине цепочки, у квадратной и треугольной решётки (первый блок первых двух строк таблицы 5) отличается почти в два раза: 0.71299(2) и 0.35989(1) (что логично, ведь в треугольной решётке диагональные ячейки так же считаются соседними, поэтому половина поворотов конформации добавит соседей).

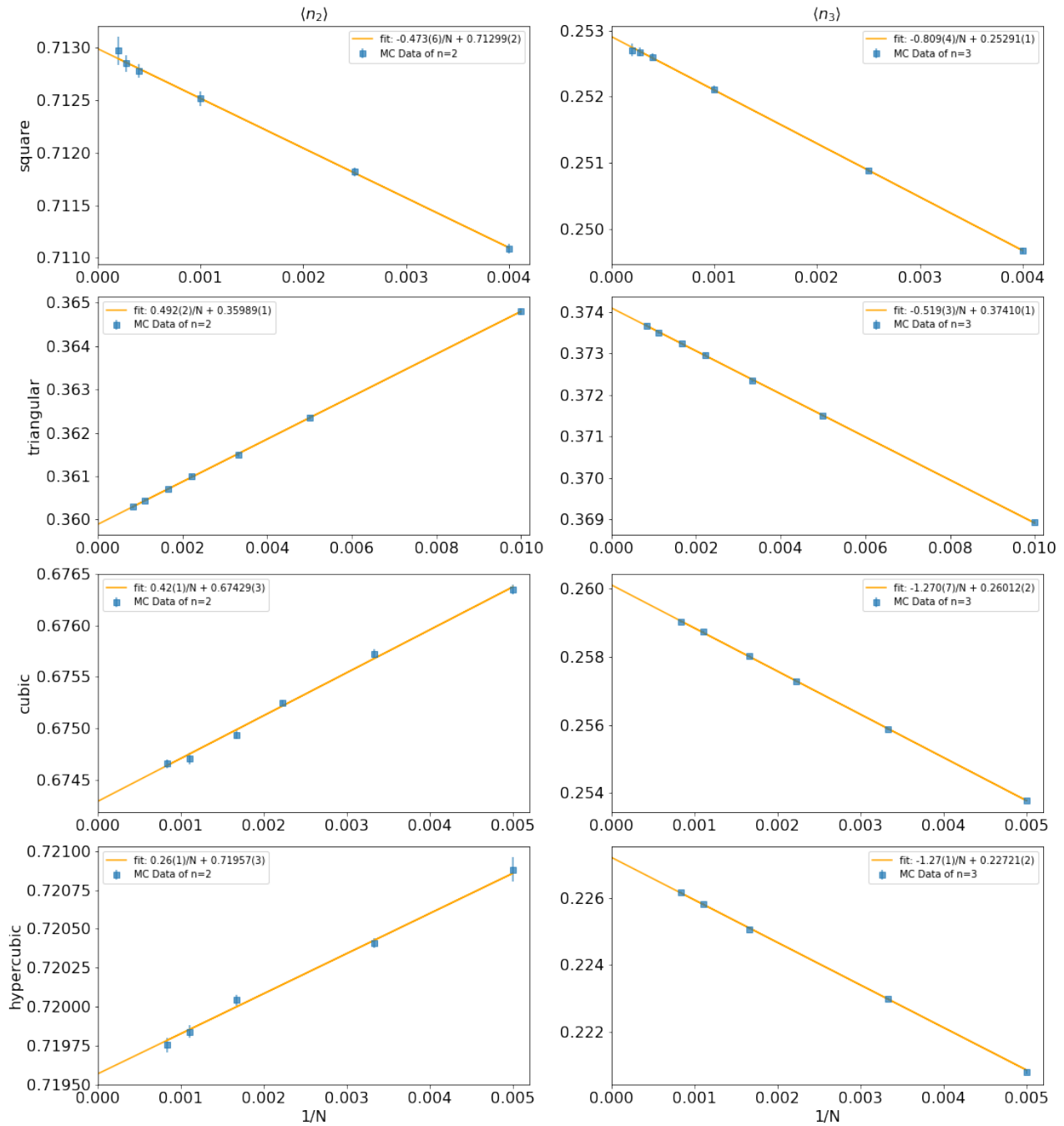


Рис. 7: Зависимость средней доли узлов с 2-мя соседями (слева) и 3-мя (справа) от обратной длины $1/N$ в модели Изинга на случайном блуждании на квадратной, треугольной, кубической и гиперкубической (сверху вниз). Синие точки описывают результаты симуляций Монте-Карло, оранжевая линия - график линейной аппроксимации результатов, ошибки рассчитаны с учётом погрешностей полученных данных. Коэффициенты и диапазоны длин рассматриваемых данных записаны в таблице 5

	$\langle n_5 \rangle$			$\langle n_6 \rangle$		
Lattice	a	b	N	a	b	N
Triangular	-0.274(2)	0.063145(6)	100-1200	-0.055(1)	0.012081(2)	100-1200
Cubic	-0.100(2)	0.007536(4)	200-1200	-0.0074(2)	0.000452(1)	200-1200
Hypercubic	-0.102(2)	0.00658(1)	200-1200	-0.0140(3)	0.000659(1)	200-1200

Таблица 6: Коэффициенты прямых, полученные линейным фитированием (1.1) данных симуляций Монте-Карло по долям узлов с 5-6 соседями

	$\langle n_7 \rangle$			$\langle n_8 \rangle$		
Lattice	a	b	N	a	b	N
Hypercubic	-0.0011(1)	0.0000420(3)	200-1200	-0.000024(35)	0.0000010(1)	200-1200

Таблица 7: Коэффициенты прямых, полученные линейным фитированием (1.1) данных симуляций Монте-Карло по долям узлов с 7-8 соседями

1.6 Сравнение геометрических свойств модели Изинга на решётках с большим числом соседей в $J=0$

Здесь мы сравниваем линейное фитирование результатов симуляций Монте-Карло треугольной решётки с кубической, имеющей такое же количество возможных соседей, а так же результаты для гиперкубической решётки в $J=0$. Коэффициенты линейного фитирования (1.1) отображены в таблицах 5 и 6: поскольку в таких условиях плотность конформаций минимальна, доля узлов с 7 и 8 соседей в конформациях на гиперкубической решётке почти нулевая, что видно по таблице 7, поэтому мы рассматриваем число соседей лишь от 2 до 6.

Рассматривая средние строки таблицы 5, где записаны коэффициенты прямых фитирования для n_2 и n_3 треугольной и кубической решётки соответственно, а так же средние графики на рисунке 7, мы видим примерно ту же ситуацию как и в случае сравнения треугольной с квадратной - кубическая решётка на графике 7 показывает почти чёткий линейный характер приближения в пределах погрешности наибольших длин (для $n=3$ линейно видна значительно лучше), но ни коэффициенты наклона a , ни значения свободных членов b не имеют никакого сходства. Единственное отличие от сравнения с квадратной решёткой - графики соответствующих долей треугольной, кубической и гиперкубической решёток имеют одинаковое поведение с точки зрения знака наклона, что действительно и для долей узлов с большим числом соседей. Можно утверждать, что треугольная решётка с точки зрения поведения доли одномерных участков больше похожа на кубическую решётку, нежели квадратную, однако точной численной универсальности (например, почти равных в пределах погрешности коэффициентов) поведения доли "одномерных"участков между ними при бесконечно больших длинах конформации не обнаружена.

Единственная пара коэффициентов, которая оказалась равна в пределах погрешности, являются коэффициенты наклона у линейного фитирования a (1.1) для долей узлов с 3-мя соседями $\langle n_3 \rangle$ у кубической и гиперкубической решёток (см. таблицу 5).

1.7 Число соседей и атмосферы блужданий

В статье [9] в пространстве невзаимодействующих случайных блужданий без самопересечений было рассмотрено так свойство конформации, как "атмосфера количество возможных направлений для удлинения цепочки длины N или количество возможных $N+1$ -х узлов.

Мы предполагаем, что данное свойство имеет связь с числом соседей при рассмотрении процесса удлинения цепочки и такие величины, как доля узлов цепочки $\langle n_i \rangle$ с фиксированным числом соседей и вероятность конформации иметь атмосферу k - $p^{(k)}$ - по-разному описывают одно и то же поведение цепочек с точки зрения их плотности.

Рассмотрим верхний рисунок 8: если конец цепочки длины N (назовём его " N -ым узлом") имеет атмосферу три (три оранжевые точки вокруг правого конца), то при добавлении нового $N+1$ -го узла N -й будет иметь два соседа: $N-1$ -й и $N+1$ -й узлы (бывший правый конец стал черной точкой).

Так же при атмосфере 2 - как на среднем рисунке 8 - когда, уже имея два соседа (черная конечная точка)

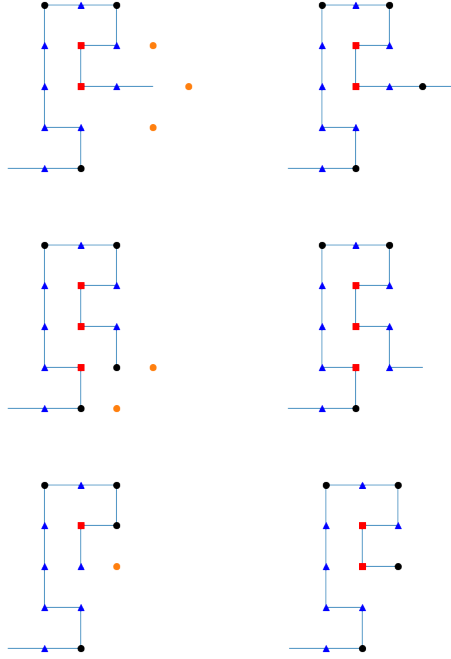


Рис. 8: Пример удлинения цепочки на квадратной решётке с атмосферой 3,2,1 (сверху вниз): слева изображена конформация до удлинения, справа - после, возможные способы добавить новый узел отмечены оранжевым, разметка узлов по количеству соседей соответствует рисунку 1

и две возможности для удлинения, N-ый узел при удлинении будет иметь 3 соседа (треугольник в том же месте на правой половине).

И наконец, при атмосфере 1 (последний рисунок 8) удлинение цепочки единственным возможным способом (одна оранжевая точка) приведёт к тому, что старый конец цепочки будет иметь 4 соседа (красный квадрат вместо треугольника). Примеры таких явлений можно увидеть на рисунке 8. Очевидно, что случай удлинения при атмосфере 0 рассмотреть невозможно, и провести аналогию с соседями нельзя.

Подобная интерпретация данных свойств в контексте удлинения цепочки показывает, что событие "цепочка длины N имеет атмосферу 3/2/1" при удлинении однозначно переходит к состоянию "N-й узел цепочки (теперь предпоследний) имеет 2/3/4 соседа" соответственно.

С другой стороны, подобная интерпретация атмосферы Преллберга не учитывает перерасчёт соседей у других узлов после удлинения цепочки - так, на примере атмосферы 1 (на нижнем рисунке 8) видно, что у одного из узлов, кроме конечного (бывшая черная точка справа), так же увеличилось число соседей (с 2-х до 3-х), тем самым она стала поверхностным узлом (синим треугольником в том же месте на правой половине).

Проведём сравнение долей узлов в фикс. числом соседей в модели Ising-ISAW при $J=0$ и вероятность конформации модели не взаимодействующего блуждания иметь атмосферу k в пределе на бесконечно большую длину на квадратной решётке.

k	$p^{(k)}$	i	$b(\langle n_i \rangle)$
3	0.711 14(3)	2	0.71299(2)
2	0.225 00(2)	3	0.25291(1)
1	0.054 76(1)	4	0.03410(1)
0	0.009 096(4)	-	-

Таблица 8: Сравнение свободных членов линейных приближений вероятностей у конформации иметь n -ю атмосферу (слева) и долей мономеров с i соседями (справа) в зависимости от обратной длины конформации $1/N$

На таблице 8 слева изображены значения свободных членов графика зависимости вероятности гомополимерной цепочки иметь атмосферу k в статье [9], то есть вероятность, что второй конец цепочки бесконечно большой длины N имеет k возможных направления для удлинения и следовательно, k возможных узлов, которые могут стать новым узлом в цепочке. Справа изображены значения свободных членов приближений графиков долей узлов с i соседями. Хотя все значения отличаются больше чем на погрешность расчётов, однако нельзя не заметить довольно близкое сходство $p^{(3)}$ и свободного члена $\langle n_2 \rangle$, хотя сами приближения имеют противоположные по знаку наклоны.

В частной переписке с автором статьи была предложена следующая коррекция результатов [3]: поскольку мы рассматриваем состояние при котором удлинения точно произойдёт, то сравнивать необходимо именно условные вероятности вида $P(\text{цепочка имеет атмосферу } k \mid \text{удлинение возможно}) = P(\text{цепочка имеет атмосферу } k) / P(\text{цепочка имеет положительную атмосферу})$:

$$p^{(1/2/3)'} = p^{(1/2/3)} / (p^{(1)} + p^{(2)} + p^{(3)})$$

Рассмотрим такую "приведённую" вероятность атмосфер и сравним с результатами для долей соседей.

k	$p^{(k)'}$	i	$b(\langle n_i \rangle)$
3	0.7177	2	0.71299(2)
2	0.2271	3	0.25291(1)
1	0.0553	4	0.03410(1)

Таблица 9: Вероятности у конформации иметь k -ю атмосферу (слева) и долей мономеров с i соседями (справа) в пределе бесконечной длины в случае гарантированно возможного удлинения

Разница между $p^{(3)'}$ и $\langle n_2 \rangle$ увеличилась. Остальные величины так же не удалось приравнять в пределах погрешности, что говорит о том, что величины обозначают несколько разные поведения модели.

2 Геометрические свойства простого случайного блуждания

В качестве завершения исследования поведения долей узлов с фиксированным числом соседей рассмотрим модель простого случайного блуждания (далее Random_Walk или RW) на квадратной решётке. В модели RW отсутствует ограничение самопересечений, и, следовательно, есть возможность попадания в ранее занятые узлы решётки.

Определим два геометрических свойства конформаций модели RW: из семейства SAW-моделей взято *количество шагов блуждания* N . Оно является параметром модели, и при генерации блужданий все конформации имеют фиксированное количество шагов. Добавляется новая наблюдаемая величина – *доля уникальных узлов блуждания* n_{unique} – отношение количества занятых блужданием узлов N_{unique} к количеству шагов N .

$$n_{\text{unique}} = \frac{N_{\text{unique}}}{N} \quad (2.1)$$

Как и в предыдущих разделах, исследуемыми свойствами будут величины $n_i, i \in \{1..4\}$ – *доли узлов с фиксированным числом соседей*. Так же будет рассмотрена атмосфера блужданий модели RW – *число занятых узлов решётки вокруг конца блуждания*. В частности, исследуется поведение вероятности $p_N^{(k)}$ – доли конформаций из N шагов с атмосферой $k \in \{0..3\}$ среди сгенерированных – с увеличением количества шагов N . Доли n_i и, конечно, атмосфера k , считаются сразу среди уникальных узлов сгенерированной конформации, для чистоты результатов и возможности сравнения с данными исследования случайного блуждания без самопересечений. Все величины будут рассмотрены в виде двух функций:

- Доли узлов как функция количества шагов простого случайного блуждания N :

$$\langle n_i \rangle = f_i(N), \quad i \in \{1,2,3,4, \text{unique}\} \quad (2.2)$$

- Доли узлов как функция количества уникальных узлов блуждания $N_{\text{unique}} = N n_{\text{unique}}$:

$$\langle n_i \rangle = g_i(N_{\text{unique}}), \quad i \in \{1,2,3,4\} \quad (2.3)$$

Основными целями раздела будут:

1. Определение характера шкалирования наблюдаемых величин при бесконечно большом блуждании ($N \rightarrow \infty, N_{\text{unique}} \rightarrow \infty$)
2. Оценка коэффициентов фитирующих функций f_i, g_i , в особенности - асимптотического предела наблюдаемых
3. Для проверки результатов: численное сравнение фитирующих функций при прямой зависимости от кол-ва шагов $f_i(N)$ и сложной зависимости от N_{unique} , которое, в свою очередь, зависит от N $g_i(f_{\text{unique}}(N))$

2.1 Алгоритм генерации блужданий

Блуждания $\{\omega_i\}$ генерируются в виде последовательности индексов направлений $\{d_i\}$ в любой решётке, что ускоряет процесс моделирования. Первая точка блуждания на решётке ω_0 лежит в начале координат $(0,0)$, далее блуждание определяется как последовательность узлов по формуле (2.4):

$$\{\omega_i\} = \begin{cases} \omega_0 = (0,0) & \text{steps} = [(0,1), (0,-1), (1,0), (-1,0)], \\ \omega_i = \omega_{i-1} + \text{steps}[d_i] & d_i = \text{Random}(\{0,1,2,3\}), i = 1..N, \end{cases} \quad (2.4)$$

где steps - массив фиксированных смещений из точки, определяемые законами решётки. Точность подсчёта наблюдаемых определяется лишь количеством повторов эксперимента.

С другой стороны, отсутствие требования непересекаемости блуждания вызывает ряд осложнений для сравнения результатов с классом блужданий без самопересечений. Например, возможны случаи, когда два идущих подряд направления противоположны друг другу - то есть, на i -м шаге блуждание смещается из точки ω_{i-1} , а $i+1$ -м - возвращается в него, то есть $\omega_{i-1} = \omega_{i+1}$. В таком случае на графике блуждания возможны "шипы", концы которых будут узлами с всего одним соседом - основанием "шипа".

Алгоритм обработки каждого модерируемого блуждания описан на картинках 9a, 9b и 9c:

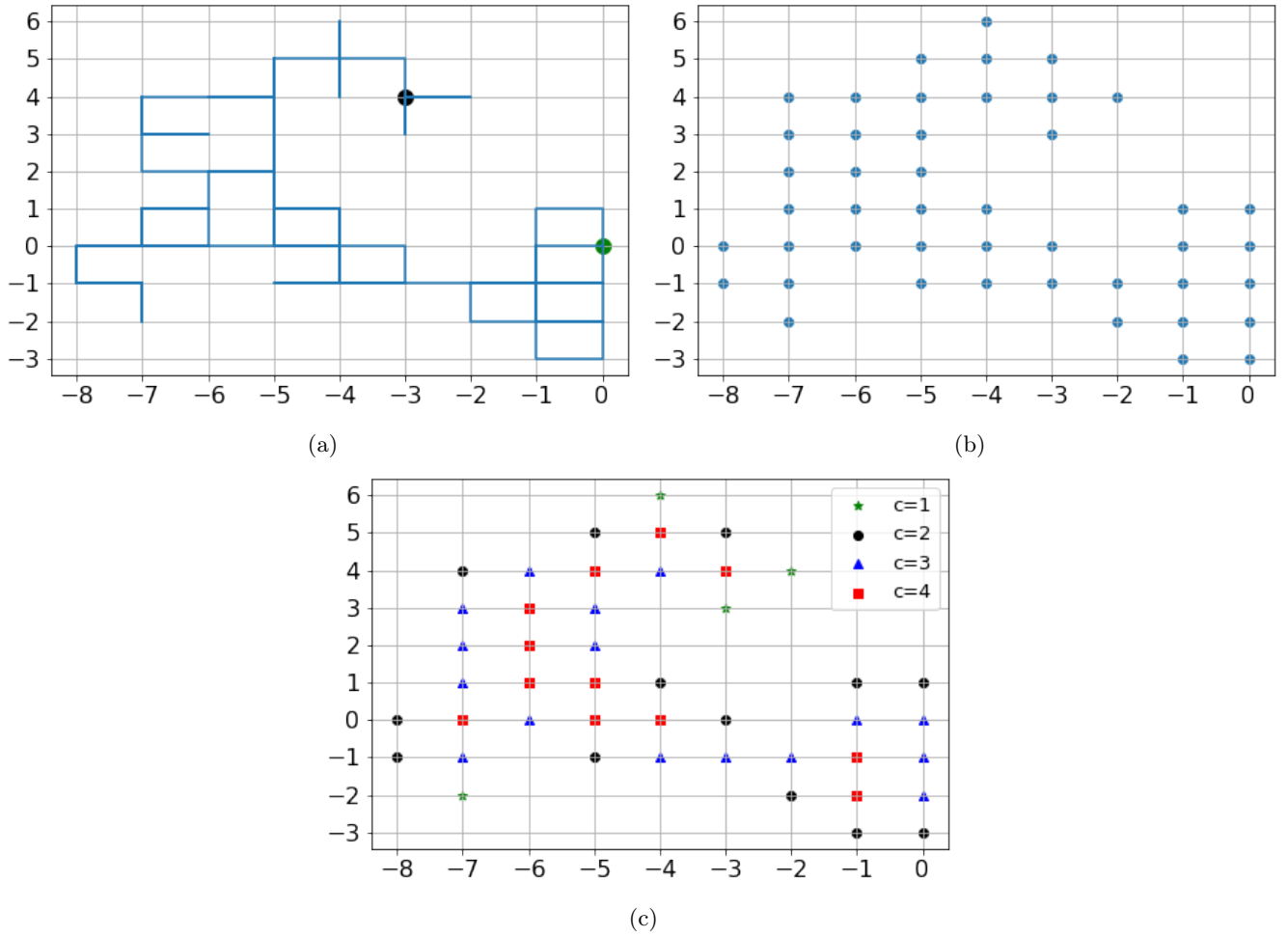


Рис. 9: а) Сгенерированное блуждание модели RW $\{\omega_i\}$ (2.4) из N шагов. Концы блуждания отмечены жирными точками, ходы блуждания – линией. Началом блуждания является зелёная точка $(0,0)$, концом блуждания – черная, в которой так же рассчитывается атмосфера блуждания k . б) Набор уникальных точек $\{\omega_i^u\}$ (2.5), принадлежащих блужданию модели RW, количество которых – N_{unique} . в) Подсчёт соседей у каждого узла блуждания - доля узлов с k соседями считается по формуле (2.6).

- Из сгенерированного блуждания (рисунок 9а) отбираются все уникальные точки узлов (2.5) - образуется набор точек решётки $\{\omega_i^u\}, i = \{0, N_{\text{unique}} - 1\}$ (рисунок 9б). Так же считается доля уникальных узлов блуждания (2.1).

$$\begin{aligned} \{\omega_i^u\} &= \text{Unique}(\{\omega_i\}) \\ N_{\text{unique}} &= |\{\omega_i^u\}| \end{aligned} \quad (2.5)$$

- Для каждого уникального узла рассчитывается кол-во его соседей $c_i \in \{1, 2, 3, 4\}$ (рисунок 9с).
- Доля узлов с k соседями считается как отношение количества уникальных узлов с k соседями к общему количеству уникальных узлов.

$$n_k = \frac{\sum_{i=0}^{N_{\text{unique}}-1} [c_i = k]}{N_{\text{unique}}} \quad (2.6)$$

2.2 Результаты симуляций

Была проведена генерация модели RW с количеством шагов $N = 10^2 - 10^4$. Доли уникальных узлов n_{unique} так же брались во внимание при симуляциях. Результаты симуляций, а так же количество итераций для

каждой длины, описаны в таблице 10 и изображены на графиках 10а, 10b и 10с.¹.

В следующих подразделах будет проведён анализ полученных результатов - сначала будет проведено обсуждение погрешностей данных, затем - оценка и сравнение шкалирующих функций (2.2), (2.3) для наблюдаемых величин.

N	M	n_{unique}	n_1	n_2	n_3	n_4
100	96430000	0.490868(8)	0.067676(3)	0.33516(1)	0.357310(7)	0.239851(9)
150	69360000	0.462622(9)	0.057825(3)	0.30787(1)	0.356280(7)	0.27802(1)
200	36140000	0.44436(1)	0.052236(4)	0.29043(1)	0.353706(9)	0.30362(2)
350	17070000	0.41251(2)	0.043761(4)	0.26070(1)	0.34590(1)	0.34963(2)
500	7720000	0.39439(2)	0.039590(5)	0.24424(2)	0.33965(2)	0.37652(3)
750	4810000	0.37559(2)	0.035672(5)	0.22759(2)	0.33188(2)	0.40487(4)
1000	2480000	0.36325(3)	0.033333(6)	0.21696(3)	0.32610(2)	0.42361(5)
3000	420000	0.32265(6)	0.02650(1)	0.18336(5)	0.30351(5)	0.4865(1)
5000	140000	0.30679(9)	0.02434(1)	0.17086(8)	0.29322(8)	0.5116(2)
6500	100000	0.2992(1)	0.02330(2)	0.16518(9)	0.28816(9)	0.5234(2)
7000	305000	0.29610(6)	0.02302(1)	0.16340(5)	0.28657(5)	0.5270(1)
8000	240000	0.29338(6)	0.02253(1)	0.16070(5)	0.28409(6)	0.5327(1)
9000	195000	0.29022(7)	0.02215(1)	0.15830(6)	0.28185(6)	0.5377(1)
10000	160000	0.28751(8)	0.02179(1)	0.15628(6)	0.27991(7)	0.5420(1)

Таблица 10: Средние доли узлов с 1-4-мя (столбцы n_1 - n_4) соседями, а так же доля уникальных узлов (столбец n_{unique}) в конформациях модели Random-Walk длин 10^2 - 10^4 (столбец N). Также изображены на графиках 10а, 10b и 10с. В столбце M выписано количество итераций алгоритма Монте-Карло.

¹Процесс симуляций был запрограммирован на языке Python и проводился с использованием суперкомпьютера НИУ ВШЭ. Оптимизация требовала дополнительного изменения окружения - см. технический раздел 5.1

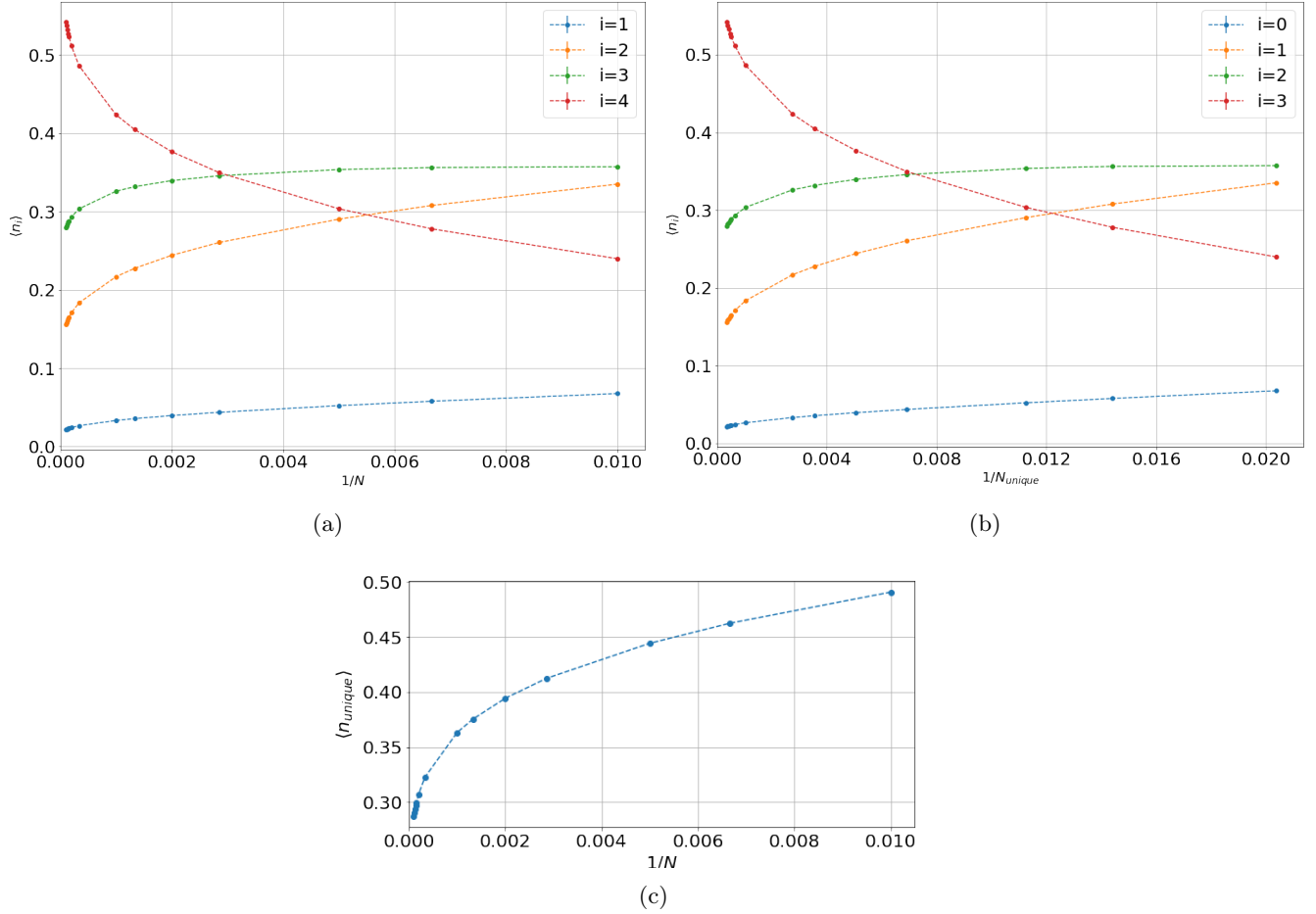


Рис. 10: Доли узлов с фиксированным числом соседей (столбцы n_1 - n_4 из таблицы 10): а) от обратного количества шагов блуждания $1/N$; б) обратного количества уникальных узлов $1/N_{\text{unique}}$; в) Доли уникальных узлов (столбец n_{unique} из таблицы 10) от обратного количества шагов блуждания $1/N$.

2.3 Погрешности результатов

Полученные в данной подсекции результаты имели ранее необоснованно большие погрешности, что потребовало более тщательного исследования. Необходимо проверить распределение результатов со временем, а так же сходимость средних наблюдаемых величин и их ошибок. В качестве примера рассмотрим первую исследуемую длину $N = 100$, т.к. именно её симуляции протекают быстрее всех.

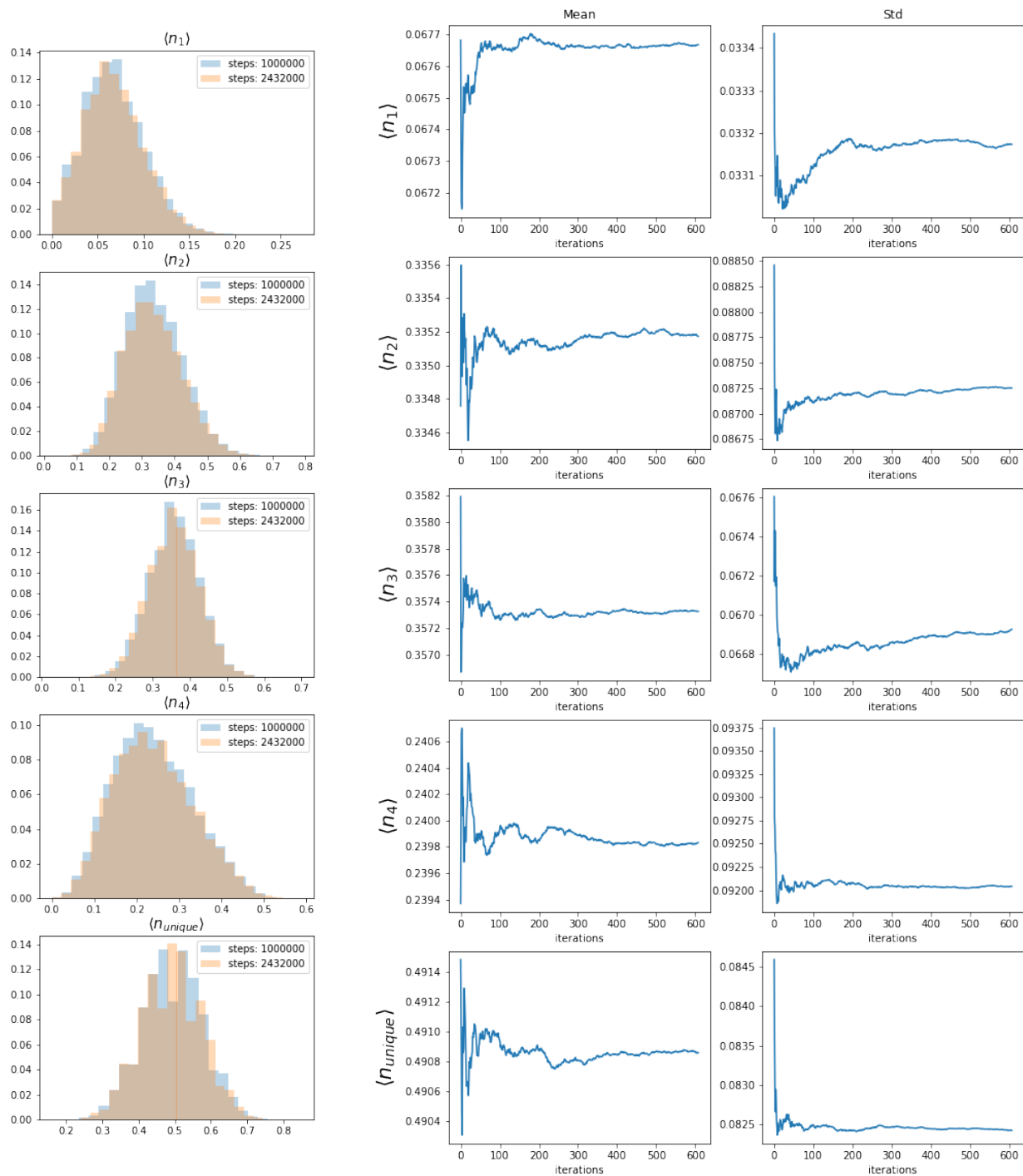
Распределение наблюдаемых долей узлов с фиксированным числом соседей 1-4, а так же доли уникальных узлов рассмотрены на гистограммах на левом графике рисунка 11 в двух моментах времени: после 10^6 шагов и после $2.5 \cdot 10^6$ шагов. На рисунке видно, что данные всех величин имеют нормальное или близко к нормальному распределению, а несимметричные склоны некоторых величин (n_1 и n_2) объясняются близостью соответствующего края к нулю.

Сходимость наблюдаемых величин можно увидеть на правом графике 11, где замеры средних проводились через каждые 4000 шагов. На графике средних заметна сходимость средней величины и уменьшение колебаний. С другой стороны, график среднего квадратического отклонения не стремится к нулю как ожидалось, а так же сходится с уменьшением колебаний к ненулевому значению. Это показывает противоречивость результатов (по крайней мере замеров ошибки - среднее явно сходится), причину чему следует искать в коде.

Для удостоверения, что причина не лежит в jit-компиляции, был проведён запуск некомпелированного с помощью numba кода. Результаты оказались идентичны с jit-компиляцией, и следовательно проблема в другом месте.

Причиной столь больших погрешностей результатов была неверная интерпретация понятия "ошибка сред-

Рис. 11: Слева: Распределение долей узлов с 1-4 соседями и уникальных узлов блуждания длины 100 в два момента времени. Справа: История результатов (Столбец mean - средняя величина, столбец std - значение ошибки на i-м замере) долей узлов с 1-4 соседями и уникальных узлов блуждания длины 100 с интервалом замеров в 4000 шагов.



него толковавшаяся ранее, как выборочное среднее квадратическое отклонение $\sigma(x)$ результатов симуляций x - на деле ошибкой среднего является формула вида:

$$\Delta\langle x \rangle = \sigma(x)/\sqrt{N},$$

где N - объём выборки или количество экспериментов.

2.4 Шкалирование результатов

Применим к результатам из таблицы 10 те же методы анализа, что и ранее для доли узлов с фиксированным числом соседей в СБС – определим характер шкалирования долей при стремлении длины конформации модели RW к бесконечности. Рассмотрим данные в трёх предполагаемых масштабах: в линейной, лог-линейной и лог-логарифмической масштабностях от обратной длины $1/N$ и кол-ва уникальных узлов $1/N_{\text{unique}}$. Пример исследуемых данных показан на графике 12. На нём видно, что в случае лог-лог-шкалирования (или степенного) график обретает наилучшую среди трёх масштабностей линейность. Оно же оказалось наиболее подходящим в графиках всех долей узлов n_{1-4} в обеих функциях $f_i(N), g_i(N_{\text{unique}})$, а так же для зависимости n_{unique} от кол-ва шагов N .

Тогда, шкалирующая функция $f_i(N)$ рассматривается в виде:

$$f_i(N) = k_i(1/N)^{a_i} + b_i, \quad i \in \{1,2,3,4, \text{unique}\} \quad (2.7)$$

Полученные коэффициенты с погрешностями выписаны на таблице 11.

	k	a	b	N
n_1	0.3425(8)	0.417(2)	0.014(1)	3000-10000
n_2	0.573(4)	0.171(1)	0.037(2)	3000-10000
n_3	0.588(3)	0.219(3)	0.202(3)	3000-10000
n_4	-1.239(9)	0.189(3)	0.759(5)	500-10000
n_{unique}	0.831(1)	0.2049(2)	0.1616(4)	500-10000

Таблица 11: Коэффициенты степенной шкалирующей функции доли узлов от обратного количества шагов блуждания $1/N$ (2.7).

Аналогичный анализ проведён для зависимости долей узлов $n_1 - n_4$ от количества уникальных узлов N_{unique} :

$$g_i(N_{\text{unique}}) = q_i(1/N_{\text{unique}})^{s_i} + d_i, \quad i \in \{1,2,3,4\} \quad (2.8)$$

	q	s	d	N_{unique}
n_1	0.313(1)	0.479(2)	0.015(1)	967-2875
n_2	0.567(3)	0.214(1)	0.053(2)	967-2875
n_3	0.542(5)	0.244(2)	0.203(2)	967-2875
n_4	-1.20(1)	0.225(4)	0.741(5)	197-2875

Таблица 12: Коэффициенты степенной шкалирующей функции доли узлов от обратного количества уникальных узлов блуждания $1/N_{\text{unique}}$ (2.8).

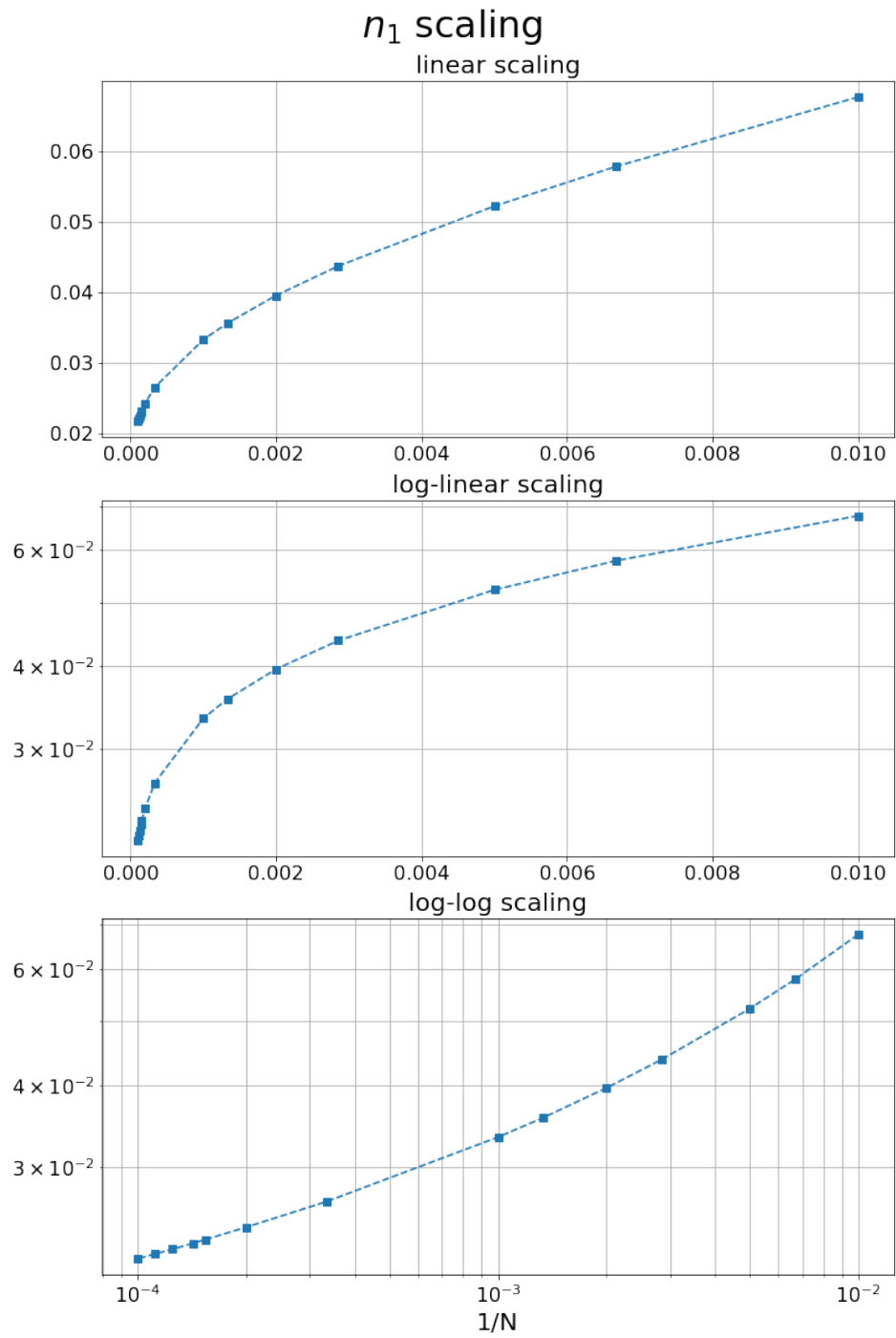


Рис. 12: Зависимость n_1 от $1/N$ в линейной, лог-линейной и лог-логарифмической масштабностях (сверху-вниз), по данным из таблицы 10.

2.5 Зависимость доли уникальных узлов от количества шагов

В данном подразделе проверяется численная эквивалентность фитирующих функций долей узлов $n_1 - n_4$: f_i (2.7), имеющей прямую зависимость от числа шагов N и $g_i(N_{\text{unique}})$ (2.8), со сложной зависимостью от N .

$$\langle n_i \rangle = g_i(N_{\text{unique}}) = q_i(1/N_{\text{unique}})^{s_i} + d_i \quad (2.9)$$

Из результатов прошлого подраздела были получены коэффициенты фитирующей функции n_{unique} (строка n_{unique} таблицы 11). Определим их как k_u, a_u, b_u соответственно и раскроем их в функции аргумента:

$$N_{\text{unique}} = N n_{\text{unique}}(N) = N(k_u(1/N)^{a_u} + b_u) \quad (2.10)$$

Рассмотрим аппроксимацию произвольной функции $g_i(N_{\text{unique}})$ с выражением её аргумента через N . Подставим (2.10) в (2.9) и проведём линеаризацию - сначала $1/N_{\text{unique}}(N)$ при $1/N \rightarrow 0$, а затем $(1/N_{\text{unique}}(N))^{s_i}$ при $1/N \rightarrow 0$:

$$\begin{aligned} 1) \quad & (N(b_u + k_u(1/N)^{a_u})^{-1} = (Nb_u)^{-1}(1 + \frac{k_u}{b_u} \frac{1}{N^{a_u}})^{-1} = \frac{1}{b_u N} (1 - \frac{k_u}{b_u} \frac{1}{N^{a_u}} + O(\frac{1}{N^{2a_u}})) \\ 2) \quad & (-// -)^{s_i} = \frac{1}{(b_u N)^{s_i}} (1 - \frac{k_u}{b_u} \frac{1}{N^{a_u}} + O(\frac{1}{N^{2a_u}}))^{s_i} = \frac{1}{(b_u N)^{s_i}} (1 - \frac{k_u s_i}{b_u} \frac{1}{N^{a_u}} + O(\frac{1}{N^{2a_u}})) \end{aligned}$$

Итоговое выражение примет следующий вид:

$$g_i(N) = \frac{k_i}{b_u^{s_i}} \frac{1}{N^{s_i}} - \frac{k_i s_i k_u}{b_u^{a_i+1}} \frac{1}{N^{a_u+s_i}} + d_i + O(\frac{1}{N^{2a_u+s_i}}), \quad N \rightarrow \infty \quad (2.11)$$

Таким образом, мы свели функцию $g_i(N_{\text{unique}})$ (2.9) к функции вида (2.7), сохранив дополнительные степенные поправки. Очевидно, линеаризация повлияет на поведение в функции области небольших длин блуждания, поэтому оценивать теорически ожидаемые линейный и степенной коэффициент по полученной функции (2.5) невозможно. Это объясняет различие коэффициентов k_i, s_i .

С другой стороны, проведенные преобразования не дали никакой поправки для асимптотического предела d_i - следовательно, вне зависимости от взятого аргумента, N или N_{unique} , функции соответствующих долей узлов с фиксированным числом соседей f_i и g_i должны сходиться на бесконечности в одной точке, а столбцы b и d таблиц 11 12 соответственно - равными в пределах погрешностей. Это так же подтверждается тем, что если $N \rightarrow \infty$, то и, очевидно $N_{\text{unique}} \rightarrow \infty$, поскольку $b_u > 0$.

Рассмотрим графики трёх функций на каждую долю $n_1 - n_4$: как функцию $f_i(N)$, как функцию $g_i(N n_{\text{unique}}(N))$, а так же аппроксимацию второй функции (2.5).

Графики функций в линейном масштабе изображены на рисунке 13. По ним видно, что функция $f(N)$ и $g(N_{\text{unique}}(N))$ почти не имеют отличий, что говорит о полном взаимозаменяемости аргументов и правильности полученных результатов на небольших длинах. Зелёная линия соответствует аппроксимирующему виду $g(N_{\text{unique}}(N))$ и имеет поправку, уменьшающуюся при стремлении N к бесконечности, но так же визуально сливается с первыми двумя функциями в области больших N .

Логарифмический масштаб графиков представлен на рисунке 14. Здесь ситуация выглядит совершенно иначе: на всех графиках наблюдается расхождение f_i и g_i по мере сближения с нулём. Причём теперь g_i и её аппроксимация сливаются в одну кривую (что говорит о правильности полученной линеаризацией функции (2.5)). Из прошлого раздела мы узнали, что асимптотические пределы n_1 и n_3 равны в пределах погрешности между выбранными зависимостями.

Однако пределы двух других функций как численно, так и графически расходятся между f_i и g_i , что противоречит предположениям о связи зависимостей в пределе бесконечного числа шагов блуждания.

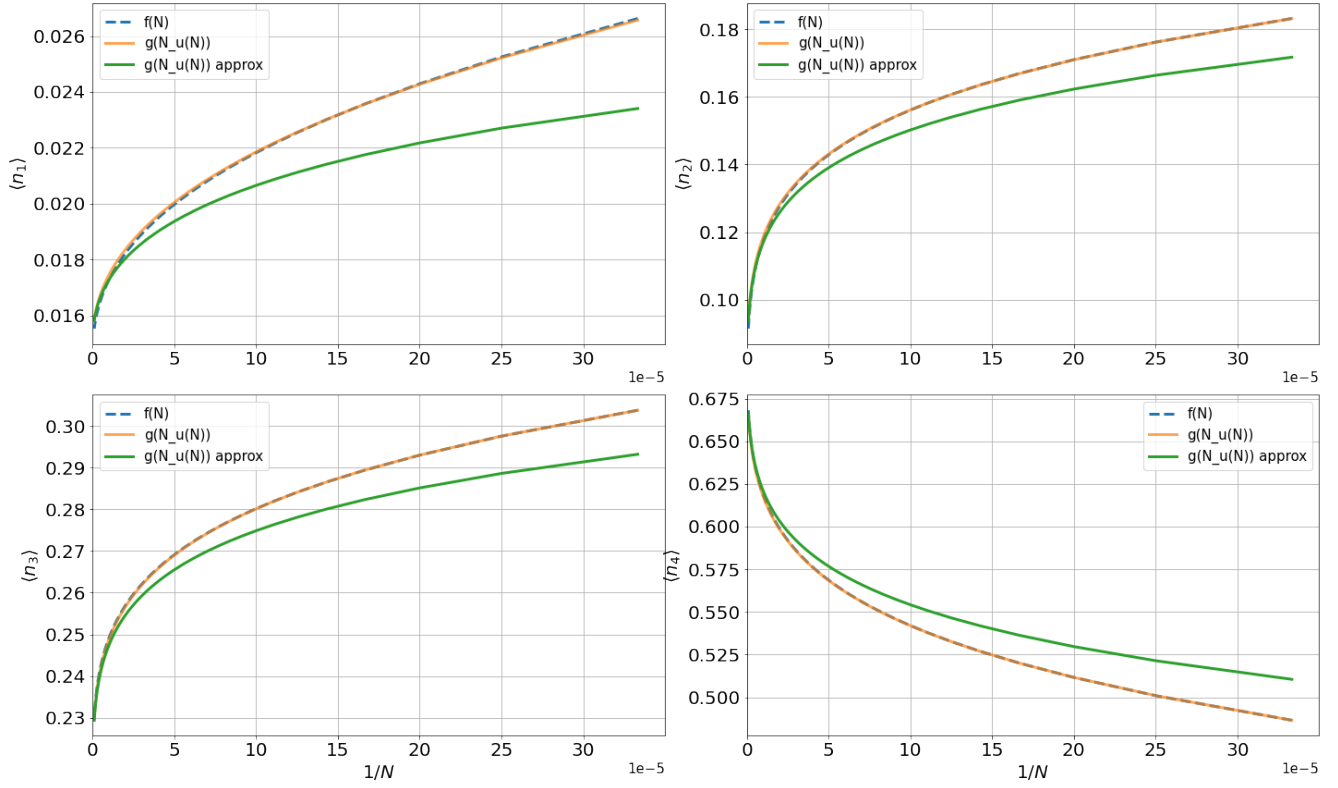


Рис. 13: Доли узлов $n_1 - n_4$ в линейном масштабе как функции от количества шагов (синий пунктир) (2.7), а так же сложные функции от количества уникальных узлов от количества шагов (оранжевая линия – прямая подстановка функций (2.10) в (2.9), зелёная - аппроксимация (2.5)), по горизонтали – обратное количество шагов блуждания $1/N$. Коэффициенты взяты из таблиц 11 и 12.

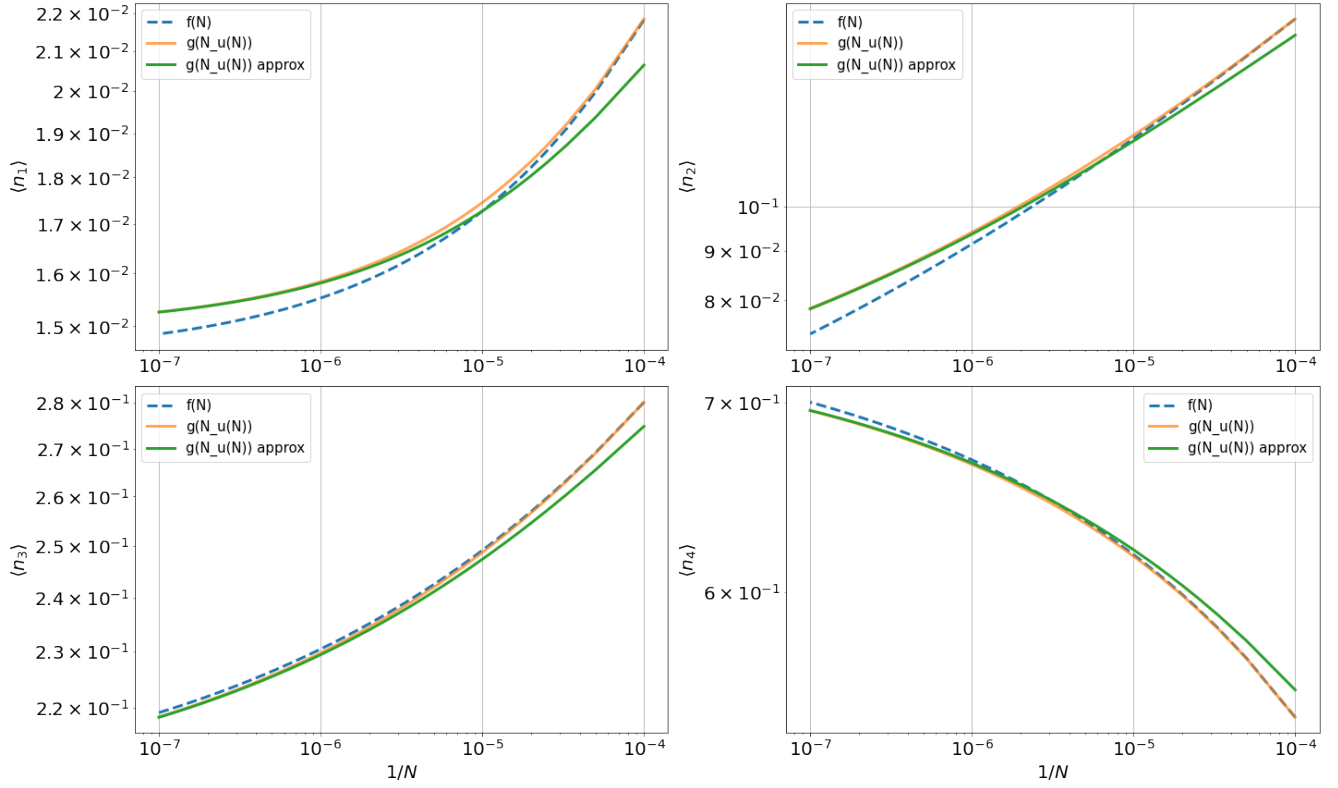


Рис. 14: Доли узлов n_1 - n_4 в лог-лог масштабе как функции от количества шагов (синий пунктир) (2.7), а так же сложные функции от количества уникальных узлов от количества шагов (оранжевая линия – прямая подстановка функций (2.10) в (2.9), зелёная – аппроксимация (2.5)), по горизонтали – обратное количество шагов блуждания $1/N$. Коэффициенты взяты из таблиц 11 и 12.

2.6 Итоговое сравнение функций долей узлов n_i

Данный подраздел завершает сравнение функций f_i (2.7) и g_i (2.8), на этот раз напрямую оценивая их сходство по линейным, степенным и свободным коэффициентам соответственно.

В первую очередь сравним свободные коэффициенты шкалирующих функций (2.7), (2.8) (столбцы b и d таблиц 11 и 12 соответственно) построчно. Очевидно, несмотря на разные структуры функций (f_i - функция от N , g_i - фактически сложная функция от N), пределы функций f_1 и g_1 , f_2 и g_2 и т.д. должны полностью совпадать, так как $N_{\text{unique}} \rightarrow \infty$ при $N \rightarrow \infty$. Так же, в виду обнаруженного степенного характера шкалирования величин при бесконечном количестве шагов, интересно различие между ближайшими к пределу результатами симуляций Монте-Карло (последняя строка таблицы 10) и полученными пределами.

	b	d	$n_i(N = 10000)$
n_1	0.014(1)	0.015(1)	0.02179(1)
n_2	0.037(2)	0.053(2)	0.15628(6)
n_3	0.202(3)	0.203(2)	0.27991(7)
n_4	0.759(5)	0.741(5)	0.5420(1)

Таблица 13: Пределы шкалирующих функций долей узлов (с фиксированным числом соседей, построчно) от N (2.7) (левый столбец), от N_{unique} (2.8) (центральный столбец) и ближайшие к пределу результаты Монте-Карло (правый столбец).

Наглядное сравнение проведено на таблице 13. Видно, что лишь у двух величин из четырех свободные коэффициенты шкалирующих функций равны между собой в пределах погрешности - это пределы долей n_1 (b_1 и d_1) и n_3 (b_3 и d_3). Другие два предела значительно отличаются - их разница многократно превышает погрешность. Однако между функциями сохраняется соразмерность пределов: как $b_1 < b_2 < b_3 < b_4$, так и $d_1 < d_2 < d_3 < d_4$.

Так же особо примечательна огромная разница между результатами симуляций Монте-Карло и соответствующими пределами у всех величин. Это показывает значительное влияние степенной составляющей функций вблизи $1/N \rightarrow 0$.

	a	s
n_1	0.417(2)	0.479(2)
n_2	0.171(1)	0.214(1)
n_3	0.219(3)	0.244(2)
n_4	0.189(3)	0.225(4)

Таблица 14: Степенные коэффициенты шкалирующих функций долей узлов (с фиксированным числом соседей, построчно) от N (2.7) (левый столбец) и от N_{unique} (2.8) (правый столбец).

Проведём аналогичное сравнение степенных коэффициентов a и s шкалирующих функций (2.7), (2.8) относительно соответствующей наблюдаемой величины на таблице 14. Численная схожесть степенных коэффициентов функций не наблюдается ни при каких величинах n_i . Заметно, что коэффициенты функций g_i несколько больше чем у функций f_i . Из этого можно справедливо предположить, что число уникальных узлов N_{unique} обладает более сильными геометрическими свойствами, чем количество шагов N . Это так же подтверждается результатами шкалирования доли уникальных узлов (строка n_{unique} таблицы 11) - доля уникальных узлов случайного бесконечно долгого блуждания очень небольшая, чуть больше 16%. Выходит, в бесконечно отдалённый момент времени блужданию нужно совершить в среднем больше 6 шагов, чтобы найти не посещённый ранее узел решётки.

Сохраняется соразмерность степенных коэффициентов: как $a_1 > a_3 > a_4 > a_2$, так и $s_1 > s_3 > s_4 > s_2$. Примечательно так же то, что у величин с наименьшими степенными коэффициентами среди всех четырех (n_2, n_4) имеют разные пределы функций f_i и g_i - можно предположить, что это вызвано техническими особенностями метода наименьших квадратов, использованного для оценки коэффициентов шкалирующих функций. Малое значение степенного коэффициента шкалирующей функции означает слабое движение антиградиента функции ошибок. Это даёт понимание, что фактически погрешность пределов функций f_i и g_i больше, чем была подсчитана методом ранее.

	k	q
n_1	0.3425(8)	0.313(1)
n_2	0.573(4)	0.567(3)
n_3	0.588(3)	0.542(5)
n_4	-1.239(9)	-1.20(1)

Таблица 15: Линейные коэффициенты шкалирующих функций долей узлов (с фиксированным числом соседей, построчно) от N (2.7) (левый столбец) и от N_{unique} (2.8) (правый столбец).

Сравнение линейных коэффициентов изображено на таблице 15. В данном случае между функциями не наблюдается ни численного, ни соразмерного сходства - несмотря на равенство всех коэффициентов до первого знака после запятой, в пределах погрешности равенства нет.

Имеется лишь знаковое сходство – все коэффициенты функций, кроме k_4 и q_4 , положительны, что говорит о сохранении схожести функций по поведению.

В итоге, с некоторым допущением относительно используемых методов аппроксимации, мы получили примерное равенство пределов шкалирующих функций от N и от N_{unique} , что говорит о корректности полученных функций шкалирования.

3 Исследование поведения концов блужданий модели RW

Следующие разделы посвящены исследованию поведения частного случая локального координационного числа на конце блуждания, или "атмосфере" блуждания. В качестве основной наблюдаемой берётся $p_N^{(k)}$ – вероятность существования блуждания с количеством шагов N с атмосферой k . Атмосферу блуждания можно легко посчитать как разность координационного числа решётки (для квадратной - 4) и локального координационного числа (число соседей вокруг узла) на конце блуждания. В рамках симуляций Монте-Карло данная наблюдаемая считается среди сгенерированной выборки:

$$\begin{aligned} a_t &= 4 - c_{\text{end}}(t), \\ p_N^{(k)} &= \frac{\sum_{t=0}^M [a_t=k]}{\text{steps}}, \end{aligned} \quad (3.1)$$

где $c_{\text{end}}(t)$ – число соседей на конце t -го сгенерированного блуждания, M – объём сгенерированной выборки, а a_t – атмосфера t -го сгенерированного блуждания. Пример подсчёта атмосферы блуждания изображен на рисунке 15а.

Ранее исследование данного свойства проводилось в работе [9], на невзаимодействующих случайных блужданиях без самопересечений. Так же подобная задача рассматривалась в книге [12], на странице 206 под номером 9, для *возвратного простого случайного блуждания*. Задача формулируется следующим образом:

- Случайное блуждание на квадратной решётке начинается из некоторой точки $x_0 = \chi$, не лежащей в начале координат.
- Процесс случайного блуждания длится не фиксированное количество шагов, а до фиксированной *точки остановки* - до достижения блужданием начала координат $x_{\text{end}} = 0$
- До достижения точки остановки блуждание может посетить одну или несколько соседних с ним точек - $(0,1)$, $(1,0)$, $(0,-1)$, $(-1,0)$. Пусть число уникальных посещенных блужданием соседних точек $N \in \{1, 2, 3, 4\}$
- Задачей является вычислить вероятности блуждания посетить каждое возможное количество уникальных соседних точек для бесконечно удаленной от начала координат начальной точки блуждания χ :

$$p_n = \lim_{|\chi| \rightarrow \infty} P_\chi[N = n] = ?, \quad n = 1, 2, 3, 4$$

Так же в качестве подсказки было указано, что отношение $p_1 : p_2 : p_3 : p_4$ почти равно $4 : 3 : 2 : 1$. Пример возвратного блуждания с точкой остановки в начале координат изображен на рисунке 15b.

В рамках летней производственной практики (см. Отчёт о практике) задача из работы Спитцера была теоретически и экспериментально решена. Результаты расчётов можно увидеть в таблице 16:

p_1	p_2	p_3	p_4
0.393566	0.314680	0.190025	0.101729

Таблица 16: Аналитическое решение задачи из работы [12]

В разделе 2.4 были найдены пределы долей узлов с фиксированным числом соседей n_i (2.6). Задачей следующих разделов, аналогично исследованию долей n_i , будет определение характера шкалирования $p_N^{(k)}$ (3.1), оценка коэффициентов шкалирующих функций, а так же сравнение асимптотических пределов вероятностей $p_N^{(k)}$ при стремлении числа шагов блуждания к бесконечности ($N \rightarrow \infty$):

1. с ответами на ранее описанную задачу [12]
2. с найденными в разделе 2.4 пределами долей n_i (2.6) (таблицы 11 и 12)

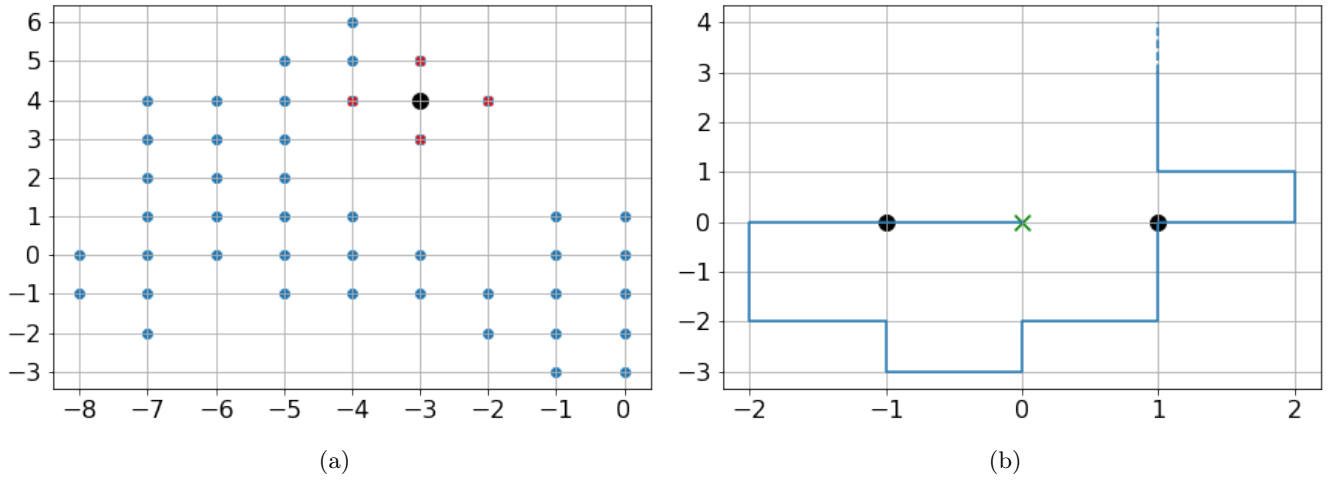


Рис. 15: а) Подсчёт атмосферы блуждания модели RW (продолжение рисунков 9). Для данного блуждания все соседние узлы его конца уже посещены ранее, атмосфера блуждания $a = 0$. б) Пример исследуемого блуждания в задаче [12]: из бесконечно удаленной точки (движение отмечено пунктиром) блуждание посетило две соседние началу координат точки (отмечены чёрным), прежде чем остановилась в нём (точка остановки отмечена зелёным крестиком).

3.1 Зависимость атмосфер от количества шагов блуждания и уникальных узлов

Была рассчитана вероятность блуждания модели RW фиксированной длины N иметь атмосферу k . Под длиной блуждания N здесь имеется в виду количество случайных шагов, проделанных блужданием, без учёта, сколько уникальных узлов оно занимает. Результаты можно увидеть в таблице 17 и на графиках 16а, как функции от N , и 16б, как функции от N_{unique} .

N	M	n_{unique}	$p^{(0)}$	$p^{(1)}$	$p^{(2)}$	$p^{(3)}$
100	96430000	0.490868(8)	0.182831	0.246855	0.290593	0.279720
150	69360000	0.462622(9)	0.212044	0.250342	0.277737	0.259877
200	36140000	0.44436(1)	0.231971	0.251413	0.269204	0.247413
350	17070000	0.41251(2)	0.268341	0.251656	0.253724	0.226279
500	7720000	0.39439(2)	0.290471	0.250914	0.244515	0.214100
750	4810000	0.37559(2)	0.313906	0.249196	0.234730	0.202167
1000	2480000	0.36325(3)	0.329962	0.247547	0.228218	0.194273
3000	420000	0.32265(6)	0.385626	0.239993	0.205155	0.169226
5000	140000	0.30679(9)	0.409736	0.235407	0.195493	0.159364
6500	100000	0.2992(1)	0.420400	0.232740	0.191620	0.155240
7000	305000	0.29610(6)	0.421036	0.233311	0.192348	0.153305
8000	240000	0.29338(6)	0.427387	0.230854	0.189233	0.152525
9000	195000	0.29022(7)	0.431959	0.232795	0.187205	0.148041
10000	160000	0.28751(8)	0.436369	0.229075	0.185300	0.149256

Таблица 17: Результаты экспериментов, описанных на графиках 16а и 16б: столбцы $p^{(0)}-p^{(3)}$ показывают среднюю вероятность блуждания модели RW с количеством шагов N иметь атмосферу 0-3 соответственно, объём сгенерированной выборки блужданий для каждой длины записан в столбце M .

Первичное рассмотрение графиков зависимости вероятностей от обратной длины $1/N$ в линейной, лог-линейной и лог-логарифмической масштабностях показало, что график лучше всего выпрямляется в третьем случае. Аналогичный результат показали графики остальных вероятностей, как функций от N так и от N_{unique} . Поэтому для всех четырех вероятностей аппроксимирующая функция при $N \rightarrow \infty$ ищется так же, как и для долей узлов, в двух видах - сначала как функция от количества шагов блуждания N , по формуле

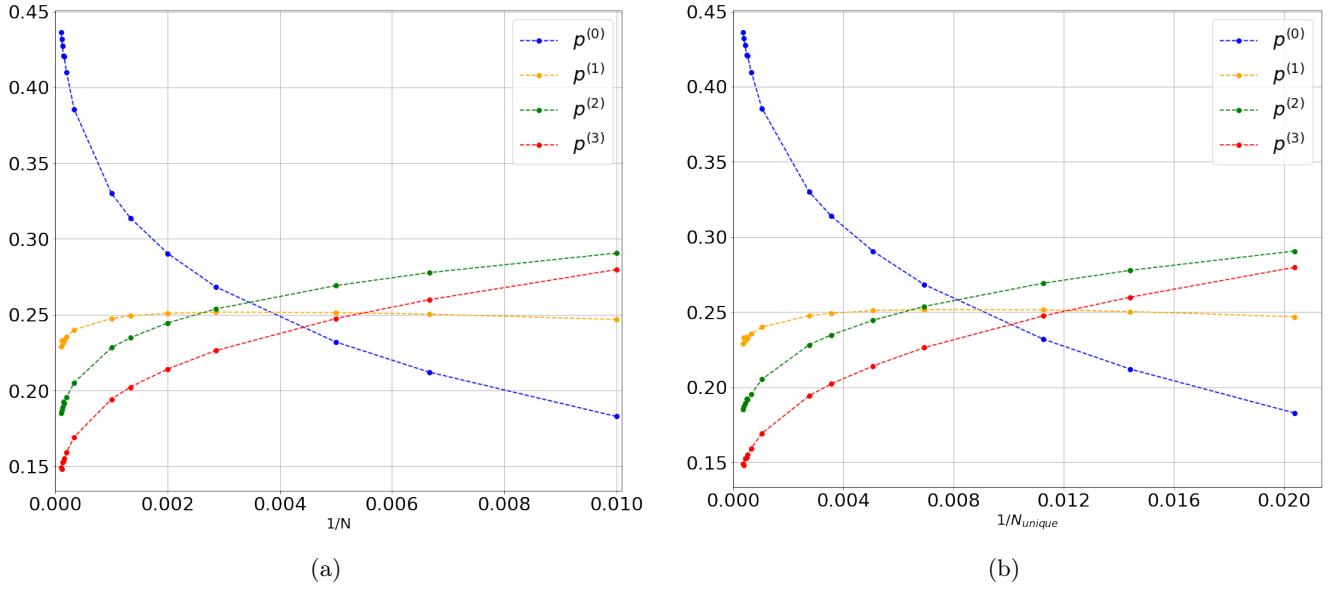


Рис. 16: Вероятность конформации модели RW иметь атмосферу $k=0,1,2,3$ (столбцы $p^{(0)}-p^{(3)}$ из таблицы 17): от а) от обратного количества шагов $1/N$ (столбец N); б) от обратного количества уникальных узлов $N_{\text{unique}} = Nn_{\text{unique}}$.

(3.2).

$$p^{(i)}(N) = k_i * (1/N)^{a_i} + b_i, \quad i \in \{0,1,2,3\} \quad (3.2)$$

где k_i - линейный наклонный коэффициент, a_i - степенной коэффициент, а b_i - свободный коэффициент. Оно же является пределом вероятности $p^{(i)}$ при $N \rightarrow \infty$. Для поиска коэффициентов использовался метод наименьших квадратов. Результаты аппроксимации графиков вблизи $1/N = 0$, диапазон выбранных длин цепочек для подбора функции, а так же стартовое положение описаны в таблицах 18 и 19.

	k_i	a_i	b_i	N	start
$p^{(0)}$	-1.17(1)	0.202(7)	0.62(1)	3000-10000	-1, 1, 0.4
$p^{(1)}$	0.54(1)	0.37(3)	0.213(6)	3000-10000	0.5, 0.5, 0.245
$p^{(2)}$	0.596(4)	0.272(6)	0.137(4)	1000-10000	0.5, 0.5, 0.16
$p^{(3)}$	0.613(5)	0.259(6)	0.092(4)	750-10000	0.5, 0.5, 0.15

Таблица 18: Коэффициенты степенных функций, аппроксимирующих вероятности $p^{(0)}-p^{(3)}$ от $1/N$ по формуле (3.2), описанных на графиках 17.

$$p^{(i)}(N_{\text{unique}}) = q_i * (1/N_{\text{unique}})^{s_i} + d_i, \quad i \in \{0,1,2,3\} \quad (3.3)$$

	q_i	s_i	d_i	N_{unique}	start
$p^{(0)}$	-1.142(9)	0.25(1)	0.59(2)	1533-2875	-1, 1, 0.7
$p^{(1)}$	0.52(1)	0.44(4)	0.214(6)	967-2875	0.5, 0.5, 0.23
$p^{(2)}$	0.585(5)	0.323(7)	0.141(3)	363-2875	0.5, 0.5, 0.16
$p^{(3)}$	0.604(5)	0.310(6)	0.097(3)	281-2875	0.5, 0.5, 0.15

Таблица 19: Коэффициенты шкалирующих функций вероятности $p^{(0)}-p^{(3)}$ от $1/N_{\text{unique}}$ по формуле (3.3).

Графики зависимости от $1/N$ представлены на рисунках 17.

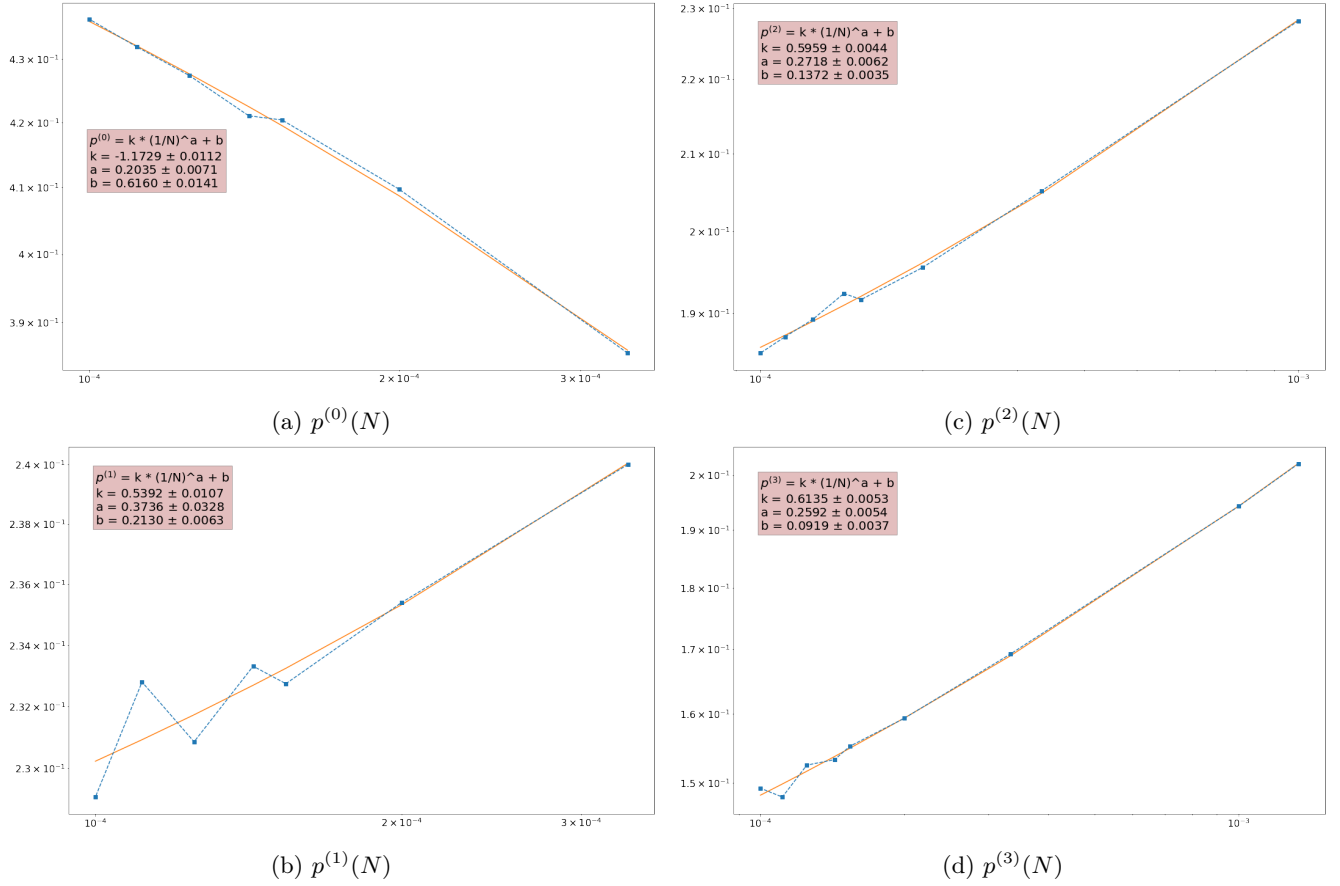


Рис. 17: Вероятности блуждания иметь атмосферу 0-3 от обратной длины конформации в степенном масштабе.

Небольшие отклонения графиков аппроксимирующей функции от прямолинейного вида обусловлены наличием ненулевого свободного линейного члена, не входящего в классическую лог-лог регрессию $y = b * x^a$. Больше всего сомнений вызывает график $p^{(1)}$ в виду сильных колебаний долей блужданий с атмосферой 1 при больших длинах. Остальные графики $p^{(0)}$, $p^{(2)}$ и $p^{(3)}$ всё же подтверждают степенной (и что наиболее важно, с сильно отличными от нуля степенными коэффициентами) характер сходимости вблизи области бесконечно большой длины.

Интересно, что в данном случае свободные коэффициенты b_i функций от N (таблица 18) и N_{unique} (таблица 19) численно очень похожи, в пределах погрешностей. Линейные и степенные коэффициенты, в свою очередь, имеют большие различия, далеко за пределами соотв. погрешностей.

3.2 Общее сравнение поведений атмосфер блужданий и долей узлов простого случайного блуждания

Для простого случайного блуждания можно отметить сильное по смыслу родство понятий "атмосферы к"блуждания и "доли узлов с i соседями". В данном случае очевидно, что если у конца блуждания некоторое число v соседей, то количество незанятых вокруг него узлов всегда равно $4 - v$. Связь этих свойств гораздо сильнее, чем в случайном блуждании без самопересечений, где для попытки их сопоставления требовалось доп. условие, что блуждание не замкнуто и всегда имеет возможность добавить к себе доп. узел.

Рассмотрим таблицы 11 и 18, 12 и 19 на предмет сходства коэффицентов между функциями $\langle n_v \rangle$ и $\langle p^{(4-v)} \rangle$. Сравнение показывает, что между таблицами отсутствует явная корреляция, за исключением идентичности знаков линейного коэффицента фугирующих функций как от N , так и от N_{unique} .

Можно предположить, что связь между ними существует - об этом говорит как подтверждённый степенной характер сходимости, так и схожесть знаков линейных коэффицентов - однако, она крайне слаба ввиду разной статистической мощности наблюдаемых величин - очевидно, что $\langle n_v \rangle$ охватывает геометрическое поведение всего блуждания, а $\langle p^{(v)} \rangle$ описывает поведение лишь на его концах, характер которых с увеличением длины блуждания становится некоррелируемым с поведением внутренних узлов.

3.3 Планируемая деятельность

- 3-я итерация программного кода для симуляции модели Rand-Walk - добавление в модель аналога квадратной решётки с целью упрощения расчётов уникальных узлов и их соседей.

4 Литературный обзор

С целью поиска информации о локальном координационном числе (что в случае блужданий может также быть названо числом соседей узла), был проведён обзор литературы, возможно имеющей отношение к рассматриваемым в рамках проекта моделям.

4.1 Livne, Meirovich: Polymers Adsorbed on a surface

4.1.1 Особенности модели блуждания

В работе [7] исследуется поведение адсорбирующего случайного блуждания без самопересечений на кубической решётке со следующими особенностями симуляции

- Случайное блуждание длины $N+1$ строится пошагово ($N+1$ мономеров в цепочке или N шагов), из начала координат ($x=0, y=0, z=0$) с ограничением на верхнее полупространство (то есть, $z \geq 0$ и плоскость $z=0$ имеет открытые граничные условия).
- Энергия конформации считается как число мономеров, лежащих на поверхности (у которых $z_i = 0$), умноженное на константу взаимодействия полимера и поверхности ϵ
- Вероятность i -й конформации считается последовательно: вводится новая статсумма, суммирующая для заданного направления текущей недостроенной цепочки всевозможные хвосты остаточной длины (10) [7].

4.1.2 Подробнее о статсумме и методе Сканирования

В данном подразделе вольным образом объясняется действие статсуммы, созданное методом сканирования. Так как при симуляции строится новое блуждание "с нуля" требуется оценка вероятности как каждого шага (точнее, направления v_k) так и всего блуждания.

Поэтому для k -го шага вероятность рассчитывается следующим образом:

1. Считается статсумма куска будущего блуждания из b ($\leq N - k + 1$) шагов, начинающая с направления v на высоте z_{k-1} :

$$Z_k(v, b, z_{k-1}, v_{k-1}) = \sum_j \exp(-\epsilon m_j(0)/k_b T) \quad (4.1)$$

2. Затем проводится расчёт вероятности выбрать направление v из всех возможных на k -м шаге:

$$p_k(v|b, z_{k-1}, v_{k-1}) = Z_k(v, b, z_{k-1}, v_{k-1}) / \sum_v Z_k(v, b, z_{k-1}, v_{k-1}) \quad (4.2)$$

3. Итоговой вероятностью всего построения будет произведение всех вероятностей каждого шага по выбранным направлениям:

$$P_i(b) = \prod_{k=1}^N p_k(v_k|b, z_{k-1}, v_{k-1}) \quad (4.3)$$

4.1.3 Результаты работы

Основными итогами работы являлось подтверждение эффективности метода "сканирования" для работы с длинными цепочками в модели адсорбирующего блуждания, определено критическое шкалирование перпендикулярного радиуса инерции (радиуса инерции проекции блуждания на ось z), а также профиля мономерной концентрации $p(z)$ (средняя доля узлов конформации длины $N+1$ на фиксированной высоте z от поверхности).

Информации о локальном координационном числе в статье найдено не было.

4.2 Madras, Sokal: The Pivot Algorithm

Работа [8] повествует о работе и эффективности алгоритма Пивота в изучении модели случайного блуждания без самопересечений (СБС).

4.2.1 Основные принципы алгоритма

Каждый шаг алгоритма проводит следующие действия над уже сгенерированной цепочкой длины $N+1$:

- Случайно выбирается с равномерным распределением для рассматриваемых узлов $p_k = 1/N$ k -й узел цепочки ($0 \leq k \leq N-1$, хотя начальную точку $k=0$ на практике не используют)
- Последующую половину цепочки ($\omega_{k+1}, \omega_{k+2}, \dots, \omega_N$ заменяют элементов группы симметрии (проще говоря, отражают, поворачивают или проводят комбинацию этих действий)
- В случае, если полученная операцией цепочка осталась без самопересечений, шаг принимается - в противном случае, шаг производится заново

В статье так же была доказана эргодичность алгоритма, а так же средние вероятности принятия каждого из возможных преобразований.

Для симуляций в качестве стартовой позиции использовалось два варианта: прямые цепочки "rods", при которых проводилось некоторое кол-во шагов до достижения термального равновесия системы (в таком состоянии процесс из следующих состояний цепочки становится близким по распределению к стационарному стохастическому), или же "димеризованные цепочки", состояние которых уже считается равновесным. Вторым методом становится крайне времязатратным при большой длине цепочки, поэтому при $N \geq 2400$ чаще применялась термализация прямых цепочек.

Пристальное внимание в статье было обращено к среднему радиусу инерции S_N^2 и квадрату расстояния между концами ω_N^2 , а так же к оценке метрической экспоненты ν , характеризующей обе величины в крит. области модели:

$$\begin{aligned}\langle \omega_N^2 \rangle &\sim N^{2\nu} \\ \langle S_N^2 \rangle &\sim N^{2\nu}\end{aligned}$$

В оценке будущей работы было так же отмечено, что алгоритм Пивота не подходит для расчёта связующей μ и критической γ экспонент (связующую константу так же называют *эффективным координационным числом*), так как алгоритм работает лишь в случае канонического ансамбля (при фиксированной длине цепочки) и требуется алгоритм, работающий уже в большом каноническом ансамбле (с цепочками изменяемой длины).

В статье не рассматривалось как таковое "число соседей узлов".

4.3 Спизер: Теория случайных блужданий

Основным объектом исследования в книге [12] является класс случайных блужданий, допускающих пересечение. В их число входит интересовавшее в рамках летней производственной практики простое случайное блуждание на двумерной решётке.

Первые две главы являлись ознакомительными, как с самим классом случайных блужданий и их свойств, так и с аппаратом исследования их поведения способом гармонического анализа. В наиболее привлекавшей внимание главе №3 косвенным образом рассматривалось свойство локального координационного числа одного из концов блуждания бесконечной длины - оно было представлено в задаче 3.9.

Полный обзор книги и в особенности главы №3 можно прочесть в отчёте о проведении летней производственной практики.

5 Программно-техническое приложение

В данном разделе будут описаны особенности работы с суперкомпьютером НИУ ВШЭ, которые могут быть важными дополнением к основной инструкции пользователя.

5.1 Применение jit-компиляции при программировании на языке Python

Симуляции случайного блуждания с самопересечениями (для кода см. папку *Random_Walk* [1]) были запрограммированы на языке Python с компиляцией с помощью пакета numba метод jit. В качестве окружения была использована стандартная библиотека *Python/Anaconda_v11.2021* встроенная в стандартное ПО суперкомпьютера.

Выполнение первых экспериментов по симуляциям шло крайне медленно - результаты за семь дней можно увидеть на таблице 20

N	steps	unique	n_1	n_2	n_3	n_4
100	7450000	0.49(8)	0.07(3)	0.33(9)	0.36(7)	0.24(9)
200	5684000	0.44(7)	0.05(2)	0.29(7)	0.35(5)	0.30(9)
500	2045000	0.39(6)	0.04(1)	0.24(5)	0.34(4)	0.38(8)
1000	654000	0.36(5)	0.03(1)	0.22(4)	0.33(4)	0.42(7)
2500	132000	0.33(4)	0.027(7)	0.19(3)	0.31(3)	0.48(6)
5000	37000	0.31(4)	0.024(5)	0.17(3)	0.29(3)	0.51(6)
10000	10000	0.29(3)	0.021(4)	0.16(2)	0.28(3)	0.54(5)

Таблица 20: Средние доли узлов с 1-4-мя соседями в конформациях модели Random-Walk длин $10^2 - 10^4$

Для сравнения с другими платформами, в случае длины цепочки $N = 10000$, процесс из 10000 шагов на Google Colab занимал не более 7 часов.



Решением проблемы оказалось создание собственного окружения с другими версиями используемых пакетов numru и numba (полный список так же есть в репозитории с кодом [1]). Новые результаты за 7 дней описаны в продолжении основного раздела.

При обсуждении столь значительного различия во времени выполнения между окружениями поддержкой было выдвинуто предположение, что окружения отличаются сторонними библиотеками линейной алгебры, используемой пакетом numru: наиболее распространенными считаются OpenBLAS и Intel MKL. Основным фактором преимущества той или иной библиотеки является именно процессор (Intel или non-Intel).

В новом окружении пакетом numru использовалась именно библиотека OpenBLAS, в то время как в Anaconda - Intel MKL. Это следовало из применения в данных окружениях следующего:

```
import numpy
print(numpy.show_config())
```

Подробнее об определении какая библиотека линейной алгебры используется в пакете numru можно найти [здесь](#).

5.2 Итерации программного комплекса Rand-Walk

Подраздел посвящён описанию версий программного комплекса для симуляций модели простого случайного блуждания фиксированной длины N на квадратной решётке. (для кода см. папку *Random_Walk* [1])

1. **Drunken_Sailor_def.py** - базовый алгоритм симуляций, предназначенный для проверки работы основных функций:

- *experiment* - генерация цепочки и подсчёт наблюдаемых (доли узлов с числом соседей 1-4, а так же доля уникальных узлов цепочки)
- *complex_experiment* - запись результирующего массива для одной цепочки (*experiment*) и набора цепочек (шаг - кол-во опытов между анализом данных)
- *write_results* - запись текущих результатов (средних наблюдаемых по всем экспериментам) в текстовый файл
- *save_distr* - распределение значений наблюдаемых по всем экспериментам
- *save_history* - сохранение истории средних значений для анализа сходимости результатов симуляций

Цепочка генерируется как двумерный массив точек, потому наиболее его медленной частью является поиск уникальных узлов цепочки через *np.unique*, не поддерживающий njit-компиляцию при обработке двумерного массива.

2. **Drunken_Sailor.py** - первая версия симуляционного комплекса с jit-компилируемой частью. Алгоритмически не отличается от **Drunken_Sailor_def.py**, но значительно быстрее базовой версии

3. **Drunken_Sailor_v2.py** - оптимизированная версия **Drunken_Sailor.py** с расширенной njit-компиляцией:

- *create_walk* - генерация цепочки как массива поворотов блуждания начиная с начальной точки (0,0), затем - как массив всех точек блуждания
- *calc_fractions* - основная функция подсчёта наблюдаемых. Так же модифицирована над подсчёт атмосферы каждого блуждания
- В *complex_experiment* добавлено распараллеливание проведение набора экспериментов за шаг между выводом данных, что позволило значительно ускорить работу комплекса.
- *stats* - подсчёт текущего результата для наблюдаемых долей
- *atm_bins* - подсчёт долей блужданий с атмосферой 0-3

Список литературы

- [1] Github repository: Chpel/projectmagnet.
- [2] Github repository: kamilla0503/saw.
- [3] Prellberg private communication 2021.
- [4] Sergio Caracciolo, Marco Gherardi, Mauro Papinutto, and Andrea Pelissetto. Geometrical properties of two-dimensional interacting self-avoiding walks at the θ -point. *Journal of Physics A: Mathematical and Theoretical*, 44(11):115004, 2011.
- [5] Kamilla Faizullina, Ilya Pchelintsev, and Evgeni Burovski. Critical and geometric properties of magnetic polymers across the globule-coil transition. *Phys. Rev. E*, 104:054501, Nov 2021.
- [6] Damien Paul Foster and Debjyoti Majumdar. Critical behavior of magnetic polymers in two and three dimensions. *Physical Review E*, 104(2):024122, 2021.
- [7] Shelly Livne and Hagai Meirovitch. Computer simulation of long polymers adsorbed on a surface. i. corrections to scaling in an ideal chain. *The Journal of Chemical Physics*, 88(7):4498–4506, 1988.
- [8] Neal Madras and Alan D Sokal. The pivot algorithm: a highly efficient monte carlo method for the self-avoiding walk. *Journal of Statistical Physics*, 50(1):109–186, 1988.
- [9] AL Owczarek and Thomas Prellberg. Scaling of the atmosphere of self-avoiding walks. *Journal of Physics A: Mathematical and Theoretical*, 41(37):375004, 2008.
- [10] V Privman. Study of the θ point by enumeration of self-avoiding walks on the triangular lattice. 19(16):3287–3297, nov 1986.
- [11] M C Tesi, E J Janse van Rensburg, E Orlandini, and S G Whittington. Interacting self-avoiding walks and polygons in three dimensions. 29(10):2451–2463, may 1996.
- [12] Ф Спицер. *Принципы случайного блуждания*. М.: Мир, 1969.