

DOI: 10.51790/2712-9942-2021-2-4-4

СТАТФИЗИЧЕСКИЕ СВОЙСТВА ОПТИЧЕСКОЙ НЕЙРОСЕТИ, ОСНОВАННОЙ НА ПРОЦЕССАХ ЧЕТЫРЕХВОЛНОВОГО СМЕШЕНИЯ

Б. В. Крыжановский, Л. Б. Литинский

*Федеральное государственное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», г. Москва, Российская Федерация
ORCID: <https://orcid.org/0000-0002-0901-6370>, kryzhanov@niisi.ras.ru*

Аннотация: исследованы статфизические свойства оптической нейросети. Получены условия, при которых возможно обучение нейросети алгоритмом максимального правдоподобия. Исследование проведено на примере трехмерной модели Изинга, в которой последовательно добавляется дальное действие так, что в пределе модель можно описывать теорией среднего поля. Получены аналитические оценки для критической температуры нейросети при учете взаимодействия со вторыми и третьими соседями. Данные оценки на всем интервале значений параметров взаимодействия хорошо согласуются с результатами, полученными методами Монте-Карло. Установлено, что с ростом числа положительных межсвязей величина критической температуры падает и алгоритм максимального правдоподобия может применяться практически без ограничений.

Ключевые слова: оптическая нейросеть, модель Изинга, дальное действие, метод n-окрестности.

Благодарности: работа выполнена в рамках госпрограммы ФГУ ФНЦ НИИСИ РАН, проект 0065-2019-0003.

Для цитирования: Крыжановский Б. В., Литинский Л. Б. Статфизические свойства оптической нейросети, основанной на процессах четырехволнового смешения. *Успехи кибернетики*. 2021;2(4):42–48. DOI: 10.51790/2712-9942-2021-2-4-4.

STATISTICAL PHYSICAL PROPERTIES OF FOUR-WAVE MIXING OPTICAL NEURAL NETWORK

B. V. Kryzhanovsky, L. B. Litinsky

*Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, Moscow, Russian Federation
ORCID: <https://orcid.org/0000-0002-0901-6370>, kryzhanov@niisi.ras.ru*

Abstract: the paper investigates the statistical physical properties of an optical neural network. The conditions for training a neural network by the maximum likelihood algorithm are identified. The study uses a three-dimensional Ising model, to which a long-range action is sequentially added so that in the limit the model can be described by the mean-field theory. Analytical estimates of the critical neural network temperature were obtained considering the interaction with the second and third-order neighbors. The estimates for the entire interval of the interaction parameters are in good agreement with the results obtained by Monte Carlo methods. It is found that as the number of positive interconnections increase, the critical temperature value decreases and the maximum likelihood algorithm can be applied virtually without any restrictions.

Keywords: optical neural network, Ising model, long-range interaction, n-coherence method.

Acknowledgements: the work was financially supported by SRISA under the government program No. 0065-2019-0003.

Cite this article: Kryzhanovsky B. V., Litinsky L. B. Statistical Physical Properties of Four-Wave Mixing Optical Neural Network. *Russian Journal of Cybernetics*. 2021;2(4):42–48. DOI: 10.51790/2712-9942-2021-2-4-4.

Введение

Цель проводимого исследования состоит в том, чтобы изучить возможность использования алгоритма максимального правдоподобия [1, 2] (ML-algorithm (Maximum-Likelihood Algorithm)) для глубокого обучения нейросети, адаптированной к распознаванию оптических паттернов. Поскольку этот

алгоритм можно использовать только в области температур, где нет фазового перехода, основной упор сделан на анализ того, как критическая температура зависит от параметров матрицы связи. Описание функционирования такой сети дается далее. Формально оно сводится к описанию эволюции спиновой системы. Соответственно, для моделирования термодинамических свойств нейросети мы используем трехмерную модель Изинга, в которой учитывается взаимодействие не только с ближайшими, но и с последующими соседями.

Если при рассмотрении 3D-модели Изинга учитывать взаимодействие только между ближайшими соседями, система демонстрирует фазовый переход второго рода, характеризующийся бесконечной особенностью теплоемкости. Мы не знаем наверняка, является ли это логарифмической сингулярностью или теплоемкость является степенной функцией с очень малым показателем степени, но важно, чтобы при стремлении температуры к критическому значению теплоемкость увеличивалась неограниченно. Теоретические оценки [3–6], а также результаты многих компьютерных симуляций [7, 8] свидетельствуют в пользу такого поведения.

Нас интересует, что произойдет, если мы также учтем взаимодействия со следующими ближайшими и следующими ближайшими соседями. Влияет ли такое изменение модели на характер особенности в критической точке? Поясним причину нашего интереса. В принципе, мы также можем включить в наши расчеты взаимодействия с соседями «следующий-следующий-следующий-ближайший» и так далее. Таким образом, в конечном итоге становится возможным описать систему с помощью модели среднего поля. В рамках приближения среднего поля фазовый переход в модели Изинга тоже второго рода, но теплоемкость имеет конечный скачок. Следовательно, если учесть все больше и больше удаленных соседей, где-то изменится тип особенности теплоемкости. Другими словами, мы должны получить конечный скачок вместо бесконечной особенности.

Одной из наших задач было определить: изменяет ли тип сингулярности включение взаимодействия с соседями из последующих координационных сфер? Мы не получаем однозначного ответа, однако мы показываем, что увеличение числа взаимодействующих соседей изменяет поведение теплоемкости вблизи критической точки.

Другой задачей было выяснить: как включение в рассмотрение вторых и третьих соседей отразится на значении критической температуры? Если взаимодействия с ними положительны — будет ли критическое значение возрастать или падать? Это важно, если иметь в виду аппаратную реализацию оптической нейросети.

Работа устроена следующим образом. Сначала дается сжатое описание нейросети, адаптированной для распознавания оптических паттернов, описываются ограничения на использование ML-алгоритма. Далее излагаются результаты численного эксперимента для 3D-модели Изинга с включением дальнего действия. Найденные экспериментальные значения критической температуры сравниваются с теоретическими оценками по методу n -окрестностей [9, 10]. В следующем разделе суммированы экспериментальные результаты для спонтанной намагниченности, внутренней энергии и теплоемкости, а далее изложены результаты, относящиеся к учету взаимодействия с более дальними (третьими) соседями. Заключительные замечания отнесены в последний раздел.

Обучение оптической нейросети

В работах [11, 12] изучены распознающие возможности оптико-электронной нейросети, способной хранить и обрабатывать информацию, закодированную в виде частотно-фазовой модуляции. Информативные сигналы в такой сети передаются по межсвязям в виде квазимонохроматических импульсов на r частотах. Однако формальное описание сети удобнее всего вести на языке векторного формализма, описывая r различных состояний нейронов ортами r -мерного пространства $\{e_l\}_{l=1}^r$. Тогда состояние сети как целого описывается набором векторов:

$$\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N), \quad (1)$$

где вектор-нейрон \mathbf{s}_i принимает одно из базисных значений \mathbf{e}_l , N — число нейронов, а динамика сети описывается гамильтонианом

$$E = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{s}_i^+ \mathbf{J}_{ij} \mathbf{s}_j.$$

Здесь $r \times r$ -матрица \mathbf{J}_{ij} задает связь между i -м и j -м вектор-нейронами, причем $\mathbf{J}_{ii} \equiv \mathbf{0}$, поскольку самодействие в системе отсутствует. Эволюция системы подчиняется следующим правилам. Суммарный сигнал, поступающий в момент времени τ на i -й нейрон, описывается r -мерным вектором, который вычисляется как линейная комбинация

$$\mathbf{h}_i(\tau) = \sum_{j=1}^N \mathbf{J}_{ij} \mathbf{s}_j(\tau). \quad (2)$$

Далее отыскивают проекции вектора $\mathbf{h}_i(\tau)$ на орты $\{\mathbf{e}_l\}_{l=1}^r$ и определяют максимальную по модулю проекцию — пусть это будет проекция на \mathbf{e}_k . Тогда в следующий момент времени i -му нейрону приписывают состояние $\mathbf{s}_i(\tau + 1) = \mathbf{e}_k \cdot \text{sign}(\mathbf{e}_k \cdot \mathbf{h}_i(\tau))$ и переходят к анализу следующего нейрона. Иными словами, из входящего на нейрон сигнала (2) выделяется частота с максимальной амплитудой и нейрон испускает единичный импульс на этой частоте. Формализм (1)–(2) совпадает с описанием спиновой модели Поттса, в которой состояния спинов описываются r -мерными векторами. В дальнейшем мы будем использовать спиновую терминологию: спин \mathbf{s}_i под действием локального поля \mathbf{h}_i разворачивается вдоль ближайшей к \mathbf{h}_i оси координат \mathbf{e}_k .

Детальный анализ свойств подобной нейросети, матрица связей которой построена по обобщенному правилу Хебба, проделан в [11, 12], в которых показано, что сеть способна хранить и эффективно распознавать Мискаженных паттернов, где $M = Nr(r - 1)(1 - 2a)^2/2$. Здесь a — уровень шума, т.е. вероятность искажения компонент поступающего на вход сети сигнала. В простейшем случае $r = 2$ такая сеть сводится к модели Хопфилда, обрабатывающей бинарные сигналы: $s_i = \pm 1$.

Для конволюционных нейросетей нет возможности построения матрицы связей по правилу Хебба (отсутствует понятие «паттерн»). Такую сеть обучают с помощью ML-алгоритма [1], который основан на методе контрастной дивергенции (Contrastive Divergence Method) [2]. Приведем основные выражения из работы [1], кратко описывающие суть этого алгоритма. Пусть распределение по состояниям \mathbf{S} задается вероятностью:

$$P(\mathbf{S}, \mathbf{J}) = \frac{1}{Z(\mathbf{J})} e^{-E(\mathbf{S}, \mathbf{J})}, \quad (3)$$

где $E(\mathbf{S}, \mathbf{J})$ — энергия в состоянии \mathbf{S} при заданной матрице связей $\mathbf{J} = \{\mathbf{J}_{ij}\}$, а $Z(\mathbf{J}) = \sum_{\mathbf{S}} e^{-E(\mathbf{S}, \mathbf{J})}$ — нормировочная константа (то же, что статистическая сумма). Тогда обучение ML-методом при заданной обучающей выборке X сводится к градиентному спуску:

$$\mathbf{J}^{(\tau+1)} = \mathbf{J}^{(\tau)} + \eta \left. \frac{\partial L(\mathbf{J}; X)}{\partial \mathbf{J}} \right|_{\mathbf{J}^{(\tau)}}, \quad X = \{\mathbf{S}_i\}_{i=1}^M, \quad (4)$$

где η — параметр обучения, а функция правдоподобия задана в виде:

$$L(\mathbf{J}, X) = \frac{1}{M} \sum_{i=1}^M \log P(\mathbf{S}_i, \mathbf{J}). \quad (5)$$

Как показывает большой опыт, применение этого подхода дает отличные результаты: нейросеть быстро и качественно обучается [17–20]. Однако такой успех в обучении имеет место, когда элементы матрицы связей знакопеременны с нулевым средним. В этом случае у системы нет фазового перехода и градиентный спуск (4) позволяет найти оптимальные значения матричных элементов.

Отметим, что реализация оптической нейросети со знакопеременными матричными элементами технически весьма сложная задача. Гораздо проще реализовать оптическую нейросеть с положительными матричными элементами. Но в этом случае у спиновой системы может существовать фазовый переход, которого при ML-обучении следует избегать. Действительно, если в системе имеется фазовый переход, то обучение по формулам (4)–(5) может привести систему в окрестность критической точки, в которой имеется сингулярность, и выражение (4) становится неприменимым. Анализу ситуации с фазовым переходом, когда элементы матрицы связи положительны, посвящены следующие разделы.

Компьютерное моделирование

Методами Монте-Карло (алгоритм Метрополиса и алгоритм Ванга-Ландау) исследовались термодинамические свойства модели Изинга с периодическими граничными условиями и ферромагнитным взаимодействием спинов в узлах кубической решетки $N = 20 \times 20 \times 20$. Величина связи с шестью ближайшими соседями принималась за единицу ($J_1 = J_2 = J_3 = 1$), а взаимодействие с 12 соседями из следующей координационной варьировалось от $J_4 = 0$ до $J_4 = 1$.

Для каждого значения обратной температуры $\beta \sim 1/kT$ вычислялась внутренняя энергия U , дисперсия энергии σ^2 и намагниченность m . Вследствие конечного размера решетки сингулярности функции $\sigma = \sigma(\beta)$ не наблюдалось. Однако она имеет максимум при $\beta = \beta_{max}$ и мы принимали, что величиной β_{max} аппроксимируется критическая температура: известно, что при $N \rightarrow \infty$ имеем $\beta_{max} \rightarrow \beta_c$. При $N \sim 10^4$ наши оценки показали:

$$\beta_{max} \approx 1.11 \cdot \beta_c. \quad (6)$$

Результаты численного эксперимента сведены в таблице 1, где представлены значения исследуемых параметров в критической точке $\beta = \beta_{max}$. На рис. 1 показаны нормированные кривые $\sigma^2(\beta)/\sigma_0^2$ при разных значениях связи J_4 . Нормировка на $\sigma_0^2 = 3(1 + 2 \cdot J_4^2)$ позволила показать все кривые на одном рисунке (пиковые значения кривых σ_{max}^2 при различных J_4 отличаются более чем на порядок, см. таблицу 1).

Таблица 1

Критические значения β_{max} , U_{max} и σ_{max}^2

J_4	σ_0^2	β_{max}	U_{max}	σ_{max}^2
0.0000	3.000	0.2230	-1.0921	53.35
0.0812	3.040	0.1850	-1.1845	78.15
0.1623	3.158	0.1575	-1.1884	89.68
0.2500	3.375	0.1223	-1.3707	124.17
0.3543	3.753	0.1190	-1.5077	173.01
0.5000	4.500	0.0995	-1.5229	230.23
0.7000	5.940	0.0740	-1.8049	366.10
0.8500	7.335	0.0656	-1.8957	465.91
1.0000	9.000	0.0653	-2.1212	602.85

Как и ожидалось, с увеличением J_4 величина β_{max} уменьшается и поведение кривых вблизи критической точки изменяется: с ростом J_4 кривая $\sigma^2 = \sigma^2(\beta)$ сдвигается влево по оси абсцисс, вблизи которой образует своеобразную ступеньку. При достаточно больших значениях связи J_4 поведение $\sigma^2 = \sigma^2(\beta)$ в критической точке больше похоже на скачок, чем на бесконечную сингулярность. Для прояснения этой неопределенности необходимо провести ряд экспериментов со значительно большими размерами решетки.

На рис. 3 и рис. 4 показаны соответственно зависимости спонтанной намагниченности $m = m(\beta)$ и внутренней энергии $U = U(\beta)$ для различных значений параметра J_4 . Как видим, величина $U(\beta)$ монотонно убывает, а в критической точке выпуклость кривой $U = U(\beta)$ сменяется вогнутостью. Смена выпуклости соответствует тому, что в критической точке производная $\partial U / \partial \beta \sim -\sigma^2(\beta)$ имеет бесконечную сингулярность. Отметим, что с ростом β внутренняя энергия асимптотически стремится к энергии основного состояния $E_0 = -3(1 + 2J_4)$. Для однообразия на рис. 4 приведена нормированная величина $U(\beta) / |E_0|$, поэтому при больших β все кривые стремятся к -1.

В заключение приведем простые оценочные зависимости для пиковых значений, вытекающие из данных эксперимента:

$$\sigma_{max}^2 = 370 \cdot (J_4 + 0.25)^2 + 33, \quad U_{max} = -1.08 - 1.03J_4, \quad (7)$$

достаточно точно описывающие зависимость σ_{max}^2 , U_{max} от J_4 : относительная ошибка аппроксимации σ_{max}^2 меньше 3 %, а для U_{max} меньше 2 %. Сравнение (7) с результатами эксперимента приведено на рис. 5 и рис. 6.

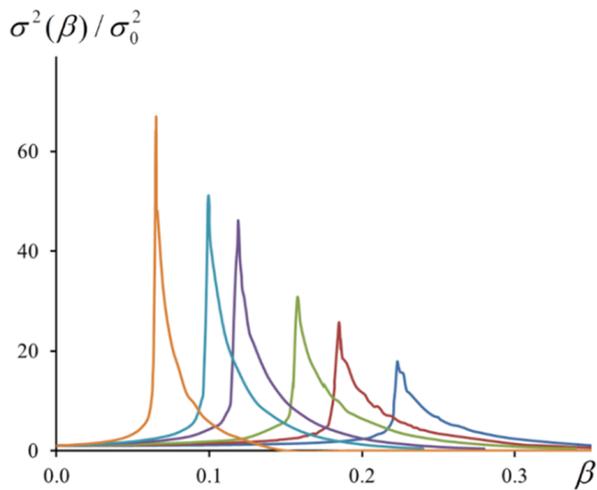


Рис. 1. Зависимость $\sigma^2 = \sigma^2(\beta)$ при $J_4 = 0, 0.0815, 0.16227, 0.35425, 0.5, 1$ (справа налево)

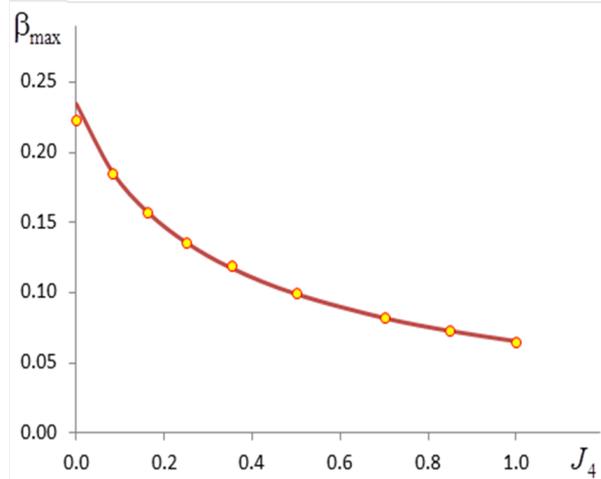


Рис. 2. Зависимость β_{\max} от J_4 : кружки — результат численного эксперимента; сплошная кривая построена по формулам (6) и (9)

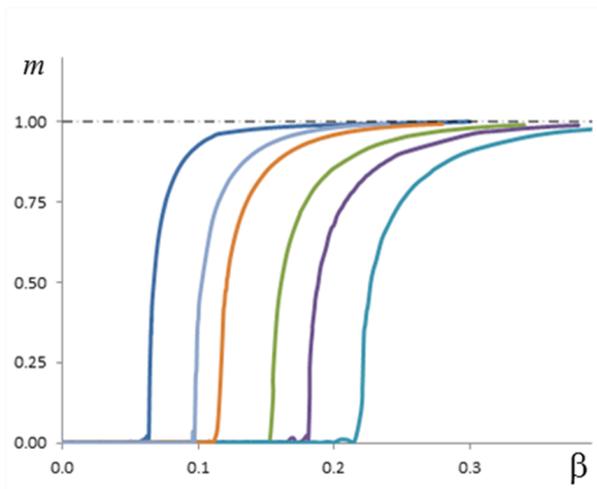


Рис. 3. Зависимость спонтанной намагниченности $m = m(\beta)$ при $J_4 = 0, 0.0815, 0.16227, 0.35425, 0.5, 1$ (справа налево)

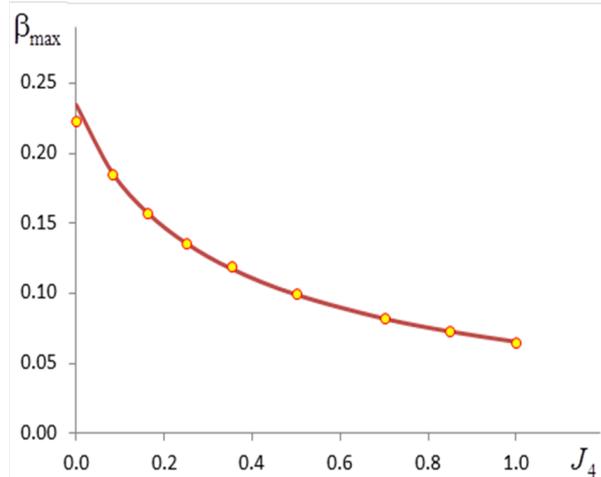


Рис. 4. Зависимость внутренней энергии $U = U(\beta)$ при $J_4 = 0, 0.0815, 0.16227, 0.35425, 0.5, 1$ (справа налево)

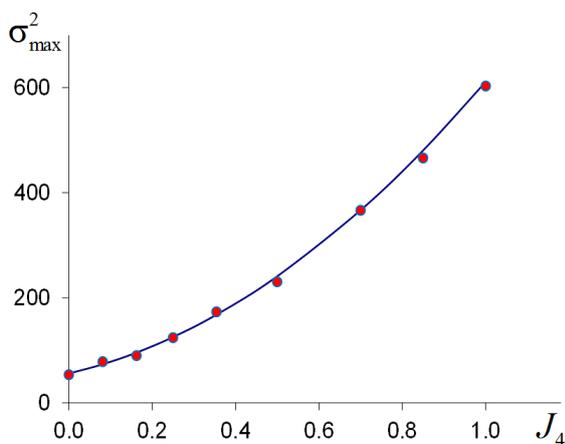


Рис. 5. Зависимость σ_{\max}^2 от величины связи J_4 : кружки — эксперимент, сплошная линия — формула (7)

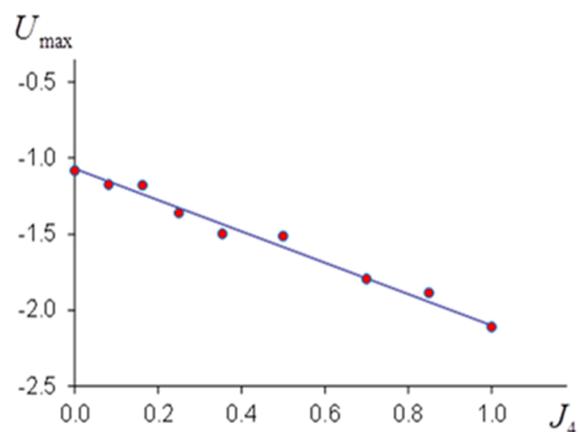


Рис. 6. Зависимость U_{\max} от величины связи J_4 : кружки — эксперимент, сплошная линия — формула (7)

Аналитические выражения для β_{\max}

В работах [7–9] нами развит метод n -окрестности, позволяющий приближенно вычислять свободную энергию. Метод основан на гауссовом приближении плотности состояний с точно вычисляемыми моментами истинной плотности. Приложение этого метода к рассматриваемой здесь задаче позволяет получить аналитическое выражение для критической температуры β_c (см. [9, 10]). Из результатов [9, 10] следует, что чем больше размерность гиперкуба d , тем лучше согласие между теоретической оценкой β_c и результатами компьютерного моделирования. Например, для кубической решетки ($d = 3$) теория предсказывает величину β_c с относительной ошибкой порядка 2.4 %; при возрастании размерности с $d = 3$ до $d = 7$ ошибка быстро уменьшается до 0.02 %.

При использовании метода n -окрестности для модели Изинга на гиперкубе мы имеем дело с единственным параметром

$$\gamma = \left(\sum J_{ij} \right)^2 / N \sum J_{ij}^2, \quad (8)$$

аккумулирующим всю необходимую информацию о связях в решетке. В частности, в случае модели Изинга на d -мерном гиперкубе при учете только ближайших соседей параметр γ равен в точности числу ближайших соседей: $\gamma = 2d$. Когда мы учитываем взаимодействие с более дальними соседями, то величина γ перестает быть целочисленной.

Как показано в [9, 10], метод n -окрестности можно применять только при $\gamma > 16/3$: в этом случае имеет место фазовый переход второго рода, сопровождающийся конечным скачком теплоемкости. Более того, в этом случае величина β_c определяется как наименьший корень уравнения

$$N = \beta \cdot \sum J_{ij} - \beta^2 \cdot \sum J_{ij}^2. \quad (9)$$

В рассматриваемом здесь случае величина $\gamma = 6 \cdot (1 + 2J_4)^2 / (1 + 2J_4^2)$ удовлетворяет неравенству $\gamma > 16/3$ и мы можем использовать (9), откуда следует:

$$\beta_c = \frac{1 + 2J_4}{2(1 + 2J_4^2)} \left(1 - \sqrt{1 - \frac{2(1 + 2J_4^2)}{3(1 + 2J_4)^2}} \right). \quad (10)$$

Выражения (6) и (10) позволяют дать теоретическую оценку зависимости $\beta_{\max} = \beta_{\max}(J_4)$. На рис. 2 сплошной линией показана величина $\beta_{\max} = 1.11 \cdot \beta_c$, а кружками — данные эксперимента (третья колонка в таблице 1). Как видим, наблюдается хорошее согласие теории с экспериментом.

Обсуждение результатов

Мы использовали метод n -окрестности для анализа модели Изинга на гиперкубе для случая $\gamma > 16/3$. В области применимости метода наблюдается хорошее согласие предсказаний развитой теории с известными результатами и данными нашего эксперимента. Сказанное верно не только для оценки критической температуры (10). Например, для решеток большой размерности ($d = 5, 6$ и 7) мы имеем хорошее согласие в предсказании скачка теплоемкости в критической точке. Кроме того, мы провели ряд экспериментов, в которых учитывалось более сильное дальное действие, т.е. учитывалось взаимодействие с соседями последующих координационных сфер. Показано, что при достаточном увеличении дальнего действия (при увеличении параметра γ) результаты асимптотически стремятся к пределу теории среднего поля.

Из проведенного анализа следует, что с ростом числа связей (с ростом эффективного числа соседей γ) критическая величина β_c быстро уменьшается. Действительно, при целочисленных значениях межсвязей ($J_1 = J_2 = J_3 = J_4 = \dots = 1$) выражение (10) принимает вид $\beta_c = \frac{1}{2}(1 - \sqrt{1 - 4/\gamma})$. Но тогда, при большом числе соседей ($\gamma \gg 1$), критическая температура убывает как $\beta_c \approx 1/\gamma$, переходя в предел, описываемый теорией среднего поля. Это означает, что в случае нейронной сети с очень большим числом связей (аналог полного графа, в котором $\gamma \rightarrow N$ и $\beta_c \rightarrow 0$) ML-algorithm может применяться на интервале температур $\beta_c \leq \beta \leq \infty$, т.е. практически без всяких ограничений.

ЛИТЕРАТУРА

1. Carreira-Perpiñán M. Á., Hinton G. On Contrastive Divergence Learning. *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, PMLR R5:33–40, 2005.

2. Hinton G. E. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*. 2002;14(9):1771–1800.
3. Доценко В. С. Физика спин-стекольного состояния. *УФН*. 1993;163:1.
4. Patashinskii A. Z., Pokrovskii V. L. *Fluctuation Theory of Phase Transitions*. Oxford: Pergamon Pr.; 1979.
5. Butera P., Comi M. Critical Universality and Hyperscaling Revisited for Ising Models of General Spin Using Extended High-Temperature Series. *Phys. Rev. B*. 2002;65:144431.
6. Morozov O. G., Sakhabutdinov A. J. Addressed Fiber Bragg Structures in Quasi-Distributed Microwave-Photonic Sensor Systems. *Computer Optics*. 2019;43:535–543.
7. Муртазаев А. К., Рамазанов М. К., Касан-Оглы Ф. А., Курбанова Д. Р. Фазовый переход в антиферромагнитной модели Изинга. *ЖЭТФ*. 2015;147:127.
8. Häggkvist R. et al. On the Ising Model for the Simple Cubic Lattice. *Advances in Physics*. 2007;56:653–755.
9. Крыжановский Б. В., Литинский Л. Б. Обобщенное уравнение Брегга-Вильямса для систем с произвольным дальним действием. *ДАН*. 2014;459(6):680–684.
10. Kryzhanovsky B., Litinskii L. Applicability of n-vicinity Method for Calculation of Free Energy of Ising Model. *Physica A*. 2017;468:493–507.
11. Kryzhanovsky B. V., Kryzhanovsky V. M., Mikaelian A. L., Fonarev A. Parametric Dynamic Neural Network Recognition Power. *Optical Memory & Neural Network*. 2001;10(4):211–218.
12. Крыжановский Б. В., Микаэлян А. Л. О распознающей способности нейросети на нейронах с параметрическим преобразованием частот. *ДАН*. 2002;383:318–321.