



Известия высших учебных заведений. Прикладная нелинейная динамика. 2023. Т. 31, № 1  
Izvestiya Vysshikh Uchebnykh Zavedeniy. Applied Nonlinear Dynamics. 2023;31(1)

Научная статья  
УДК 530.182

DOI: 10.18500/0869-6632-003028  
EDN: CFEFMZ

## Пороги устойчивости аттракторов сети Хопфилда

И. А. Соловьев<sup>1,2</sup>, В. В. Клиньшов<sup>1,2</sup>✉

<sup>1</sup>Федеральный исследовательский центр  
Институт прикладной физики РАН, Нижний Новгород, Россия

<sup>2</sup>Национальный исследовательский  
Нижегородский государственный университет им. Н. И. Лобачевского, Россия  
E-mail: solocool46@gmail.com, ✉vladimir.klinshov@ipfran.ru

Поступила в редакцию 24.11.2022, принята к публикации 26.12.2022,  
опубликована онлайн 18.01.2023, опубликована 31.01.2023

**Аннотация.** Цель работы заключается в детальном изучении аттракторов сети Хопфилда и бассейнов их притяжения в зависимости от параметров системы — размера сети и числа запомненных образов. Для характеристики бассейнов притяжения аттракторов использовался метод вычисления порога устойчивости — минимального расстояния от аттрактора до границы его бассейна притяжения. Для полезных аттракторов данная величина соответствует минимальному искажению запомненного образа, после которого система не в состоянии его распознать. В результате исследования показано, что зависимость среднего порога устойчивости полезных аттракторов от числа запомненных образов может быть немонотонной, за счет чего устойчивость сети может возрастать при запоминании новых образов. Анализ порогов устойчивости позволил оценить максимальное число образов, которые может хранить сеть без фатальных ошибок в их распознавании. При этом порог устойчивости полезных аттракторов оказывается близким к минимально возможному, то есть к единице. В *Заключении* работы сделан вывод о том, что вычисление порогов устойчивости дает важную информацию о бассейнах притяжения аттракторов сети.

**Ключевые слова:** динамические сети, коллективная динамика, ассоциативная память.

**Благодарности.** Работа выполнена при поддержке Российского Фонда Фундаментальных Исследований, грант № 20-52-12021.

**Для цитирования:** Соловьев И. А., Клиньшов В. В. Пороги устойчивости аттракторов сети Хопфилда // Известия вузов. ПНД. 2023. Т. 31, № 1. С. 75–85. DOI: 10.18500/0869-6632-003028. EDN: CFEFMZ

Статья опубликована на условиях Creative Commons Attribution License (CC-BY 4.0).

## Stability thresholds of attractors of the Hopfield network

I. A. Soloviev<sup>1,2</sup>, V. V. Klinshov<sup>1,2</sup>✉

<sup>1</sup>Institute of Applied Physics of the RAS, Nizhny Novgorod, Russia

<sup>2</sup>National Research Lobachevsky State University of Nizhny Novgorod, Russia

E-mail: solocool46@gmail.com, ✉vladimir.klinshov@ipfran.ru

Received 24.11.2022, accepted 26.12.2022, available online 18.01.2023, published 31.01.2023

**Abstract.** Purpose of the work is the detailed study of the attractors of the Hopfield network and their basins of attraction depending on the parameters of the system, the size of the network and the number of stored images. To characterize the basins of attraction we used the *method* of the so-called stability threshold, i.e., the minimum distance from an attractor to the boundary of its basin of attraction. For useful attractors, this value corresponds to the minimum distortion of the stored image, after which the system is unable to recognize it. In the *result* of the study it is shown that the dependence of the average stability threshold of useful attractors on the number of stored images can be nonmonotonic, due to which the stability of the network can improve when new images are memorized. An analysis of the stability thresholds allowed to estimate the maximum number of images that the network can store without fatal errors in their recognition. In this case, the stability threshold of useful attractors turns out to be close to the minimum possible value, that is, to unity. To *conclude*, calculation of the stability thresholds provides important information about the attraction basins of the network attractors.

**Keywords:** dynamical networks, collective dynamics, associative memory.

**Acknowledgements.** This work was supported by Russian Foundation for Basic Research, grant No 20-52-12021.

**For citation:** Soloviev IA, Klinshov VV. Stability thresholds of attractors of the Hopfield network. Izvestiya VUZ. Applied Nonlinear Dynamics. 2023;31(1):75–85. DOI: 10.18500/0869-6632-003028

This is an open access article distributed under the terms of Creative Commons Attribution License (CC-BY 4.0).

### Введение

Сеть Хопфилда [1] является классической моделью ассоциативной памяти или памяти с адресацией по содержимому. Функция памяти в этой сети реализуется за счет ее коллективной динамики, которая обеспечивает восстановление сохраненной в ней информации по предоставленному фрагменту. Для этого связи в сети выбираются по определенному правилу, что обеспечивает формирование в ее фазовом пространстве аттракторов, соответствующих сохраненным в ней данным.

Предложенная более сорока лет назад модель Хопфилда по сей день привлекает пристальное внимание исследователей в области нейронауки и информационных систем. Рассматривались варианты сети как в классическом дискретном, так и в непрерывном исполнении [2]. Были предложены реализации сети Хопфилда в виде физических систем, например оптических систем [3] или цепей фазовой автоподстройки частоты [4]. Было показано, что помимо ассоциативной памяти, сеть Хопфилда может решать ряд других задач, например оптимизации [5] и линейного программирования [6]. Большое внимание при исследовании сетей Хопфилда уделялось исследованию емкости памяти, то есть максимального количества данных, которое система способна запомнить без существенного ухудшения производительности [7, 8]. Последние годы ознаменовались появлением так называемых новых сетей Хопфилда (modern Hopfield networks, [9]) или сетей плотной ассоциативной памяти (dense associative memories, [10]), в которых использование более сильной нелинейности приводит к существенному росту емкости памяти.

С точки зрения нелинейной динамики емкость памяти системы равна максимальному числу кодирующих данные аттракторов, которые могут быть созданы в фазовом пространстве одновременно. Однако оказывается, что помимо таких «полезных» аттракторов могут возникать и другие, «паразитные» аттракторы, не соответствующие никаким сохраненным данным. Важной характеристикой как полезных, так и паразитных аттракторов является их бассейн притяжения — множество всех начальных состояний сети, из которых она эволюционирует к данному аттрактору.

В контексте ассоциативной памяти бассейн притяжения аттрактора представляет собой совокупность различных начальных данных, которые сеть распознает как соответствующий образ из ее памяти. В многомерном фазовом пространстве сети бассейны притяжения ее аттракторов могут представлять собой чрезвычайно сложные множества, и их определение и описание является весьма нетривиальной задачей.

Настоящая работа посвящена исследованию как полезных, так и паразитных аттракторов сети Хопфилда и их бассейнов притяжения. Мы показываем, что при сохранении в сети достаточно большого числа образов в ней также появляются многочисленные паразитные аттракторы, которые располагаются в областях фазового пространства, удаленных от полезных аттракторов. Для характеристики бассейнов притяжения аттракторов мы используем численную меру — порог устойчивости, равный минимальному расстоянию от аттрактора до границы его бассейна [11]. Мы показываем, что использование данной характеристики позволяет получить новую важную информацию о структуре и свойствах бассейнов притяжения.

## 1. Модель сети

В настоящей работе мы рассматриваем дискретный вариант сети Хопфилда, в которой каждый элемент (нейрон) описывается переменной состояния  $V_i$ , принимающей значение 0 или 1. Вектор  $\mathbf{V} = (V_1, \dots, V_N)$ , определяющий состояние сети, будем также иногда называть «образом». Динамика сети разворачивается в дискретном времени, и на каждом временном шаге состояние всех нейронов одновременно или последовательно изменяется по следующему правилу:

$$V_i \rightarrow 0, \text{ если } \sum_{j=1}^N T_{ij} V_j < U, \quad (1a)$$

$$V_i \rightarrow 1, \text{ если } \sum_{j=1}^N T_{ij} V_j > U. \quad (1b)$$

Здесь  $N$  — размер сети,  $T_{ij}$  — коэффициенты связи между нейронами. Матрица связи при этом выбирается в следующем виде. Пусть заданы  $S$  образов  $\mathbf{V}^1, \dots, \mathbf{V}^S$ , тогда

$$T_{ij} = \sum_{k=1}^S (2V_i^k - 1) (2V_j^k - 1). \quad (2)$$

Можно показать [1], что при такой матрице связи каждый из образов  $\mathbf{V}^k$  становится аттрактором сети, то есть образы «записываются» в ее память. Теперь при выборе начальных условий вблизи некоторого из записанных образов система перейдет к нему, то есть осуществит восстановление запомненного образа. В дальнейшем мы также будем называть образы  $\mathbf{V}^k$  «полезными» аттракторами системы.

## 2. Паразитные аттракторы

Оказывается однако, что наряду с полезными аттракторами в сети могут возникать и «паразитные», не соответствующие ни одному из запомненных образов. Выбор начального состояния сети вблизи паразитного аттрактора приводит к неспособности сети распознать предъявленный ей образ. Таким образом, наличие паразитных аттракторов оказывает существенное влияние на функционирование сети, поэтому вопрос об их наличии и количестве является важным. Кроме того, важно понять, являются ли паразитные аттракторы слегка искаженными версиями полезных или представляют собой совершенно отличные, непохожие на них образы.

Для исследования количества и расположения паразитных аттракторов было проведено масштабное численное моделирование в зависимости от размера  $N$  и числа запомненных образов  $S$ . Для каждого набора параметров рассматривалось 360 различных реализаций сети, в каждой из которой запоминались  $S$  случайных образов. Для каждой реализации сети выбиралось 100 начальных условий и моделировалась динамика сети до ее прихода к аттрактору, в результате чего составлялся список уникальных паразитных аттракторов. Важно отметить, что для любого аттрактора сети инвертированный образ также является аттрактором, что легко показать на основе правила (1). Поэтому аттракторы, являющиеся инвертированными запомненными образами, мы не считали паразитными. Также два паразитных аттрактора, являющиеся инвертированными версиями друг друга, считались за один.

На рис. 1 и 2 приведены результаты исследования, усредненные по всем реализациям сети. На рис. 1,  $a$  представлена зависимость количества уникальных паразитных аттракторов от размера сети  $N$ , а на рис. 1,  $b$  — от числа сохраненных образов  $S$ . Зависимость числа паразитных аттракторов от размера сети является монотонно возрастающей и носит насыщающийся характер. Гораздо более интересна зависимость числа паразитных аттракторов от числа запомненных образов  $S$ . Эта зависимость является немонотонной с максимумом при  $S = 7-8$ , причем это значение практически не меняется при варьировании  $N$  в широких пределах от 20 до 600.

На рис. 2 приведены данные о среднем расстоянии от паразитного аттрактора до ближайшего полезного аттрактора, а также до ближайшего другого паразитного аттрактора. При этом расстояние между двумя образами рассчитывалось как расстояние Хэмминга, то есть определялось число несовпадающих элементов:

$$L(\mathbf{V}^1, \mathbf{V}^2) = \sum_{i=1}^N |V_i^1 - V_i^2|. \quad (3)$$

Так, рассчитанное расстояние связано с коэффициентом корреляции  $\rho$  между двумя образами как  $L = N(1 - \rho)/2$ . Заметим, что на графиках расстояние нормировалось на размер сети  $N$ . Видно, что зависимость среднего расстояния от паразитных аттракторов до другого ближайшего аттрактора близка к линейной, особенно при больших  $N$ . Таким образом, паразитные аттракторы

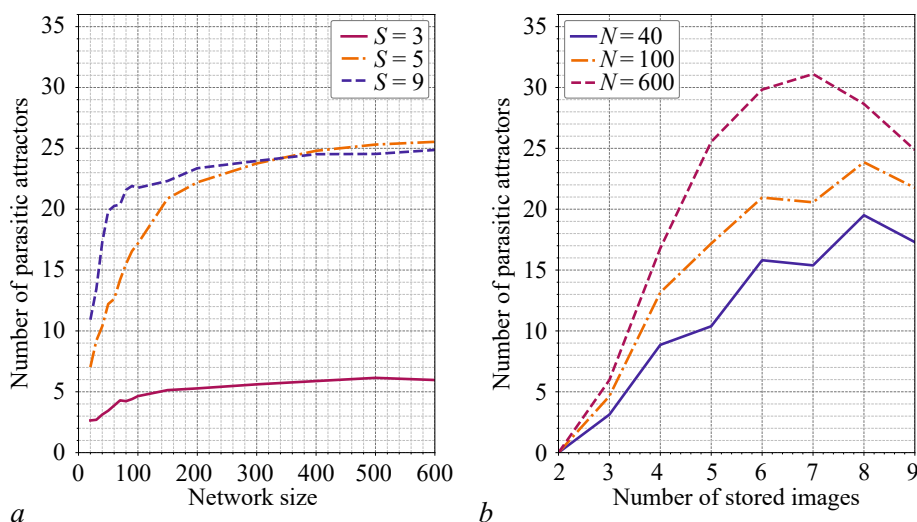


Рис. 1. Среднее количество паразитных аттракторов в сети Хопфилда в зависимости от ее размера ( $a$ ) и от числа запомненных образов ( $b$ )

Fig. 1. Average number of parasitic attractors in the hopfield network depending on the network size ( $a$ ) and the number of stored images ( $b$ )

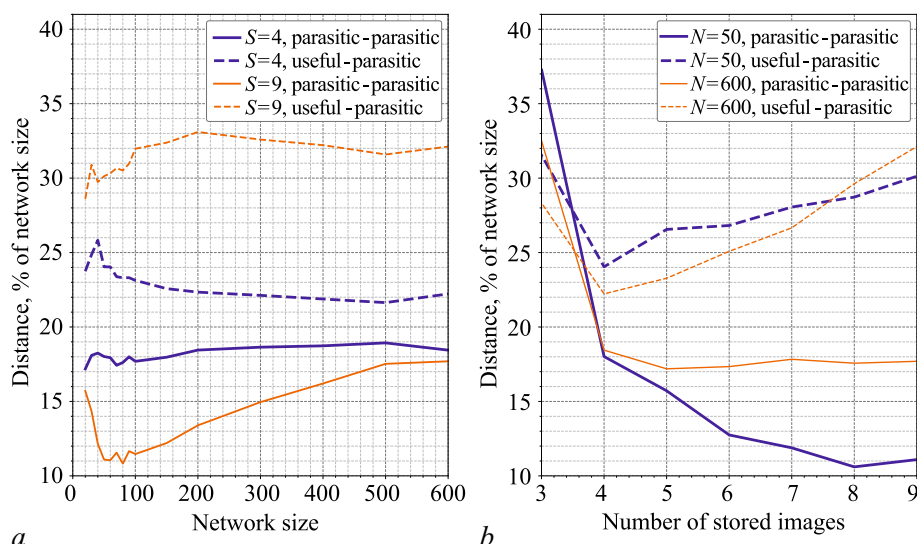


Рис. 2. Среднее расстояние от паразитных аттракторов до ближайшего полезного и ближайшего паразитного аттрактора в зависимости от размера сети (а) и от числа запомненных образов (б) (цвет онлайн)

Fig. 2. Average distance between parasitic attractors and the nearest useful attractor, average distance between parasitic attractors and the nearest parasitic attractor depending on the network size (a) and the number of stored images (b) (color online)

существенно отличаются от полезных и друг от друга. Однако следует заметить, что для паразитных аттракторов характерна значительная корреляция с некоторым из запомненных образов, так как расстояние до ближайшего из них всегда существенно меньше чем  $N/2$ .

Интересна зависимость от числа запомненных образов, которая существенно различна для расстояния до ближайшего паразитного и ближайшего полезного аттрактора. В первом случае зависимость является монотонно убывающей, а во втором демонстрирует минимум при  $S = 4$ , причем это значение практически одинаково для всех  $N$  в пределах от 20 и до 600.

### 3. Бассейны притяжения и их характеристики

Важной характеристикой аттрактора является его бассейн притяжения — множество начальных состояний сети, которые за счет ее динамики сходятся к данному аттрактору. В контексте ассоциативной памяти бассейн притяжения полезного аттрактора представляет собой множество образов, распознаваемых системой. Бассейны притяжения паразитных аттракторов, в свою очередь, определяют множество образов, предъявление которых системе приводит к сбою в ее работе. Для хорошей работоспособности системы памяти бассейны притяжения полезных аттракторов должны быть в некотором смысле большими, а бассейны паразитных аттракторов — маленькими. Для формализации этого интуитивного соображения необходимо ввести некоторые способы количественной оценки бассейнов притяжения.

В многомерном фазовом пространстве бассейны притяжения аттракторов представляют собой чрезвычайно сложные множества, и их описание является нетривиальной задачей. Было предложено несколько различных численных мер для характеристики размеров бассейнов сложных динамических систем. Например, бассейновая устойчивость [12] характеризует объем бассейна в фазовом пространстве. Порог устойчивости [11] представляет собой кратчайшее расстояние от аттрактора до границы его бассейна притяжения. Для системы ассоциативной памяти порог устойчивости полезных аттракторов представляется наиболее информативной характеристикой, так как он задает максимальную степень искажения образа, который гарантированно

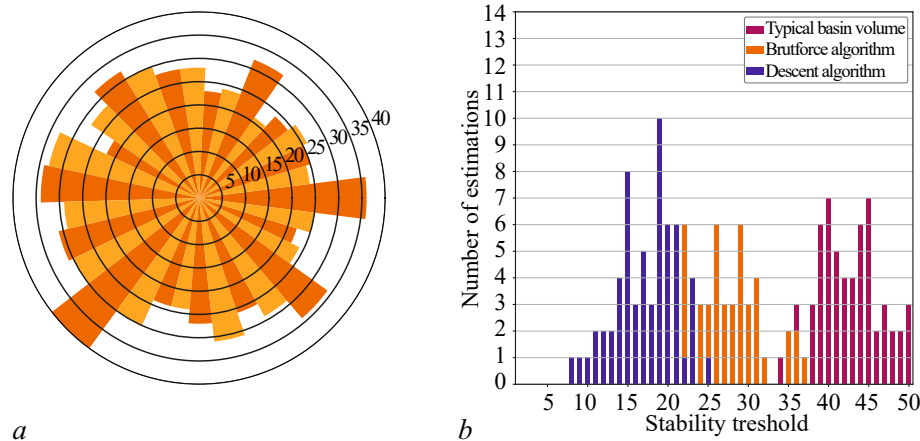


Рис. 3. *a* — Результаты определения порога устойчивости одного аттрактора при запуске алгоритма с различных случайных условий. *b* — Результаты определения порогов устойчивости различных аттракторов методом спуска и методом «грубой силы» в сравнении с характерными размерами бассейнов (цвет онлайн)

Fig. 3. *a* — Stability threshold estimations of one attractor for several runs of the algorithm with different initial conditions. *b* — Distributions of stability threshold estimation for descent and brutforce algorithms compared to the typical basin volume (color online)

будет распознан сеть. При этом возмущение, соответствующее порогу устойчивости, является минимальным искажением образа, которое приводит к сбою работы системы. Нахождение такого возмущения позволяет определить «слабое место» системы, уязвимое для возможных атак [13].

Для определения порога устойчивости сети Хопфилда мы использовали вариант градиентного спуска, предложенного в [11], модифицированный для дискретных систем. Алгоритм стартует со случайного образа  $\mathbf{U}$ , находящегося вне бассейна притяжения интересующего нас аттрактора  $\mathbf{V}$ , и пытается приблизиться к аттрактору, делая шаги в его направлении. Шаг делается следующим образом: из множества элементов образа  $\mathbf{U}$ , отличающихся от соответствующих элементов вектора  $\mathbf{V}$ , случайным образом выбирается один элемент, и его значение инвертируется, что приводит к уменьшению расстояния между  $\mathbf{U}$  и  $\mathbf{V}$ . В случае попадания образа  $\mathbf{U}$  в бассейн притяжения аттрактора  $\mathbf{V}$  сделанный шаг отвергается, в противоположном случае он принимается. Описанные шаги повторяются до тех пор, пока не возникает ситуация, при которой никакой шаг не возможен. В этом случае образ  $\mathbf{U}$  находится на границе бассейна притяжения аттрактора  $\mathbf{V}$ , причем достигнут локальный минимум расстояния между аттрактором и границей его бассейна притяжения. Стартуя с различных начальных образов  $\mathbf{U}$ , алгоритм находит множество локальных минимумов, и минимальный из них принимается в качестве оценки порога устойчивости.

Рис. 3 иллюстрирует типичные результаты поиска порога устойчивости одного из аттракторов сети. Алгоритм стартовал 100 раз с различных случайных образов, и круговая диаграмма на рис. 3, *a* отображает расстояния до найденных локальных минимумов. Отметим несколько важных моментов. Во-первых, локальных минимумов оказывается очень много: практически каждый раз алгоритм находит новый минимум. Во-вторых, расстояние от аттрактора до найденных минимумов существенно меньше характерного размера бассейна аттрактора\*, которое также отмечено на рисунке. Указанные особенности говорят о чрезвычайно сложной, изрезанной форме бассейна притяжения, для которого характерно наличие множества глубоких и узких «впадин» (ср. [14]). Определение порога устойчивости для бассейна такой формы чрезвычайно важно, так как позволяет определить очень специфические слабые возмущения, приводящие в ошибкам в работе системы.

\*Характерный размер бассейна аттрактора рассчитывался как радиус гипершара с объемом, равным объему данного бассейна.

Также полученные результаты свидетельствуют об эффективности нашего алгоритма для получения оценки порога устойчивости. На рис. 3, *b* приведено распределение порогов устойчивости, полученных с помощью предложенного алгоритма спуска для аттракторов 60 различных сетей с одинаковыми параметрами. Для сравнения на том же рисунке представлены оценки, полученные с помощью метода «грубой силы», то есть путем случайного перебора различных начальных условий и проверки их принадлежности бассейну. Параметры обоих алгоритмов выбирались таким образом, чтобы обеспечить одинаковое вычислительное время. При этом алгоритм спуска дает существенно лучшие результаты.

#### 4. Исследование порогов устойчивости системы

С помощью предложенного алгоритма было проведено систематическое исследование бассейнов притяжения как полезных, так и паразитных аттракторов сети Хопфилда в зависимости от размера сети  $N$  и числа запомненных образов  $S$ . Для каждого набора параметров  $N$  и  $S$  рассматривалось 120 различных реализаций сети, в каждой из которых запоминаемые образы выбирались случайно. Для каждой реализации сначала находились все паразитные аттракторы, как это описано выше, а затем определялись пороги устойчивости полезных и паразитных аттракторов. Полученные результаты усреднялись по всем аттракторам и по всем реализациям сети.

Результаты исследования представлены на рис. 4, *a, b*, на которых показана зависимость порогов устойчивости от размера сети и числа запомненных образов. Заметим, что на графиках порог нормируется на размер сети  $N$ . Зависимость нормированного порога устойчивости от размера сети насыщается при больших  $N$  как для полезных, так и для паразитных аттракторов. Таким образом, аттракторы в больших сетях  $N \gg S$  характеризуются достаточно сильной устойчивостью: для вывода системы из их бассейна притяжения необходимо приложить возмущение размером порядка 0.15–0.25, то есть изменить состояние 15–25% элементов.

Большой интерес представляет зависимость порогов устойчивости от числа запомненных в сети образов. Остановимся сначала на порогах устойчивости полезных аттракторов.

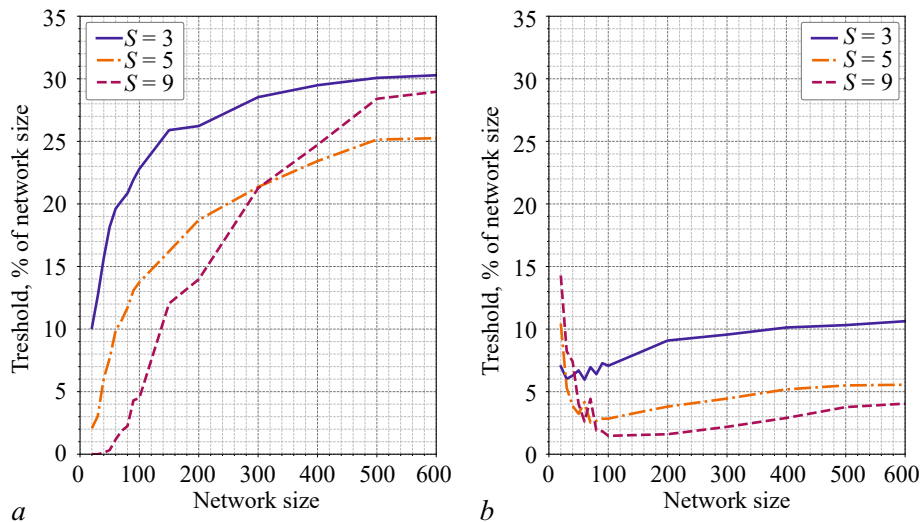


Рис. 4. Зависимость среднего порога устойчивости от размера сети  $N$  при различном числе запомненных образов  $S$ , указанном на рисунке, для полезных аттракторов (*a*) и паразитных аттракторов (*b*)

Fig. 4. Average stability threshold depending on the network size  $N$  with different numbers of stored images  $S$  for useful (*a*) and parasitic (*b*) attractors

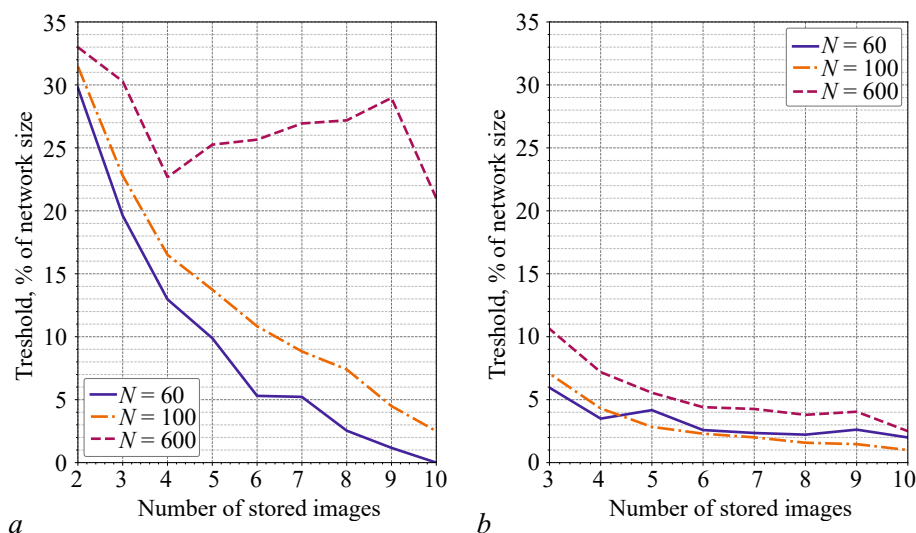


Рис. 5. Зависимость среднего порога устойчивости от числа запомненных образов  $S$  при различном размере сети  $N$ , указанном на рисунке, для полезных аттракторов (a) и паразитных аттракторов (b)

Fig. 5. Average stability threshold depending on the number of stored images  $S$  with different network sizes  $N$  for useful (a) and parasitic (b) attractors

Из рис. 5, a видно, что зависимость различна для малых ( $N \leq 100$ ) и больших ( $N > 100$ ) сетей. Если для малых сетей порог устойчивости монотонно убывает с ростом числа запомненных образов, то для больших сетей эта зависимость носит немонотонный характер: порог устойчивости сначала убывает, достигает локального минимума (при  $S = 4$  для  $N = 600$ ), затем возрастает, и снова начинает убывать при достаточно больших  $S$ . Наличие возрастающего участка на данной зависимости противоречит интуиции, так как при добавлении в систему новых аттракторов средний размер их бассейнов притяжения уменьшается, и естественно ожидать также уменьшения порогов устойчивости. Однако оказывается, что форма бассейнов «сглаживается», и наиболее глубокие «впадины» становятся мельче, что приводит к увеличению порога устойчивости. Данное наблюдение приводит к неожиданным рекомендациям по увеличению устойчивости системы к внешним атакам: иногда для этого достаточно добавить в систему новые образы, что приведет к увеличению порогов устойчивости.

Заметим однако, что аномальная возрастающая зависимость порога устойчивости от числа запомненных образов может наблюдаться только при небольшом (по сравнению с  $N$ ) числе образов. Дальнейший рост  $S$  всегда приводит к уменьшению порога устойчивости. Для паразитных аттракторов порог устойчивости также уменьшается с ростом  $S$ , как показано на 5, b. Однако при больших  $S$  порог устойчивости паразитных аттракторов уменьшается медленнее, чем порог устойчивости полезных аттракторов, и при определенном числе запомненных образов их значения становятся равными, как показано на рис. 6. Число образов, при котором достигается такое равенство, является важной характеристикой системы, определяющей предельное число образов, которые система может запомнить и корректно распознать. При попытке запомнить большее число образов порог устойчивости полезных аттракторов быстро падает до нуля, что означает потерю ими устойчивости и отказ системы.

Определим критическое число образов  $S^*$  как число образов, при запоминании которого в системе средний порог устойчивости полезных аттракторов равен среднему порогу устойчивости паразитных аттракторов. На рис. 7, a представлена зависимость критического числа образов от размера системы, и эта зависимость близка к линейной. Интересно, что таким образом определенное критическое число образов близко к оценке максимально допустимого числа образов



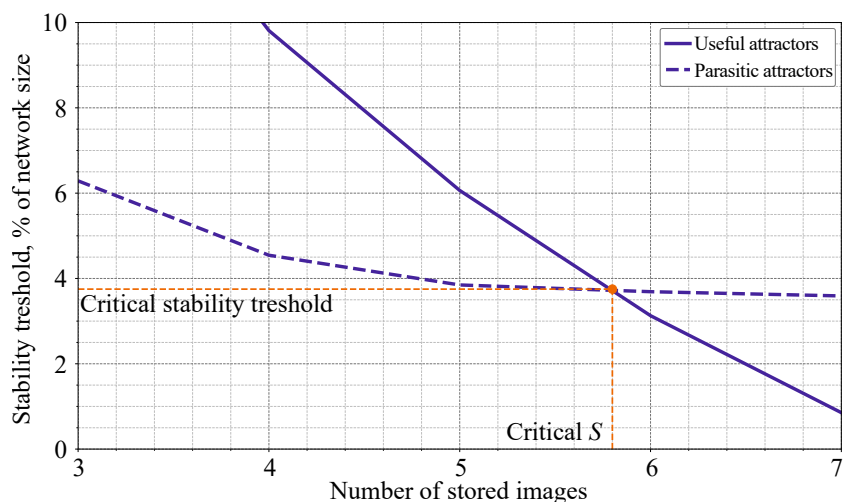


Рис. 6. Зависимость среднего порога устойчивости полезных и паразитных аттракторов от числа запомненных образов  $S$  при  $N = 40$ . Точке пересечения графиков соответствуют критическое число образов и критический порог

Fig. 6. Average stability threshold depending on the number of stored images  $S$  with network size  $N = 40$  for useful and parasitic attractors. The point of intersection determines the critical number of stored images and the critical stability threshold

$S \approx 0.1N$ , данной в оригинальной работе [1]. Таким образом, сравнение порогов устойчивости полезных и паразитных аттракторов позволяет получить независимую оценку емкости памяти сети Хопфилда, которая хорошо согласуется с традиционной. Интересно при этом проанализировать величину порога устойчивости при критическом числе образов. Зависимость этого критического порога от размера сети представлена на рис. 7,  $b$ , причем следует заметить, что в отличие от остальных графиков здесь представлена абсолютная величина порога, не нормированная на размер сети. Из графика видно, что критический порог устойчивости в абсолютном выражении близок к единице, то есть возмущение всего лишь одного элемента запомненного образа приводит к невозможности его распознавания, то есть сбою системы.

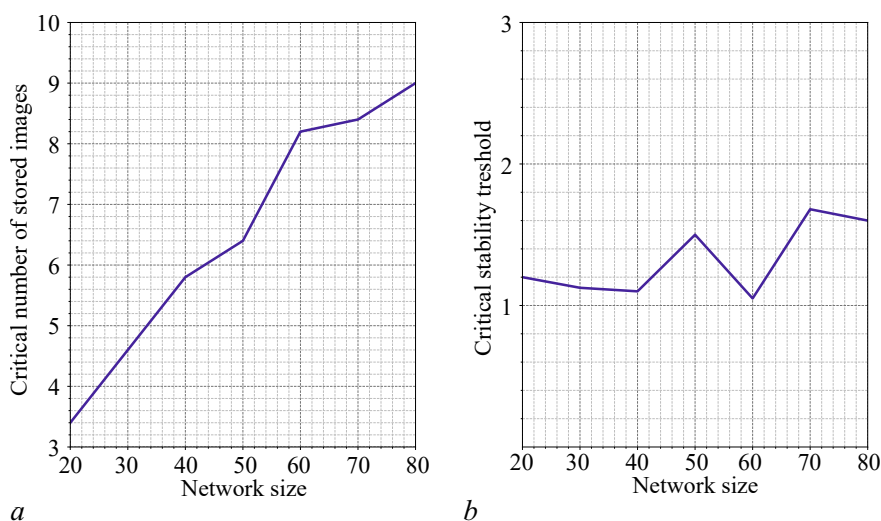


Рис. 7. Зависимость от размера сети критического числа образов ( $a$ ) и критического порога устойчивости ( $b$ )

Fig. 7. Critical number of stored images ( $a$ ) and critical stability threshold ( $b$ ) depending on the network size

## Заключение

В настоящей работе были исследованы аттракторы сети Хопфилда и их бассейны притяжения. Показано, что при записи в сеть более двух образов, помимо полезных аттракторов, соответствующих этим образам, в системе возникают также и паразитные аттракторы, не соответствующие и не близкие ни к каким из записанных образов. Число паразитных аттракторов демонстрирует насыщающуюся зависимость от размера сети и немонотонную зависимость с выраженным максимумом от числа запомненных образов. Паразитные аттракторы располагаются достаточно далеко от полезных аттракторов и друг от друга.

Для изучения бассейнов притяжения аттракторов был использован метод на основе порога устойчивости, предложенный в работе [11]. Данный метод позволяет найти минимальные по амплитуде возмущения аттрактора, приводящие к выходу из его бассейна притяжения. В контексте ассоциативной памяти, реализуемой в сети Хопфилда, порог устойчивости соответствует минимальному искажению образа, приводящему к его неправильному распознаванию. Определение таких возмущений важно с точки зрения уязвимости для возможных атак.

Был предложен алгоритм вычисления порога устойчивости в дискретных системах, и на его основе было проведено детальное исследование бассейнов в сети Хопфилда в зависимости от ее параметров. Показано, что типичный бассейн притяжения аттрактора характеризуется сложной формой, в которой присутствуют многочисленные узкие и глубокие впадины. Были изучены зависимости порогов устойчивости как полезных, так и паразитных аттракторов сети от ее параметров. Наиболее интересной является зависимость среднего порога устойчивости от количества запомненных образов при больших размерах сети  $N$ . Данная зависимость демонстрирует выраженный минимум, откуда следует неожиданная возможность увеличения устойчивости сети к внешним атакам путем добавления в ее память дополнительных образов.

На основе исследования предложен новый критерий определения максимального числа образов, которое способна хранить система без существенного ухудшения качества их распознавания: это такое число образов, при котором порог устойчивости полезных аттракторов становится равным (в среднем) порогу устойчивости паразитных аттракторов. Показано, что таким способом определенное критическое число образов близко к классической оценке емкости системы  $0.1N$ , а критический порог близок к единице. Таким образом, метод порогов устойчивости позволил получить новую важную информацию о свойствах сети Хопфилда.

## References

1. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. U. S. A. 1982;79(8):2554–2558. DOI: 10.1073/pnas.79.8.2554.
2. Hopfield JJ. Neurons with graded response have collective computational properties like those of two-state neurons. Proc. Natl. Acad. Sci. U. S. A. 1984;81(10):3088–3092. DOI: 10.1073/pnas.81.10.3088.
3. Farhat NH, Psaltis D, Prata A, Paek E. Optical implementation of the Hopfield model. Applied Optics. 1985;24(10):1469–1475. DOI: 10.1364/AO.24.001469.
4. Hoppensteadt FC, Izhikevich EM. Pattern recognition via synchronization in phase-locked loop neural networks. IEEE Transactions on Neural Networks. 2000;11(3):734–738. DOI: 10.1109/72.846744.
5. Joya G, Atencia MA, Sandoval F. Hopfield neural networks for optimization: study of the different dynamics. Neurocomputing. 2002;43(1–4):219–237. DOI: 10.1016/S0925-2312(01)00337-X.
6. Wen UP, Lan KM, Shih HS. A review of Hopfield neural networks for solving mathematical programming problems. European Journal of Operational Research. 2009;198(3):675–687. DOI: 10.1016/j.ejor.2008.11.002.

7. McEliece R, Posner E, Rodemich E, Venkatesh S. The capacity of the Hopfield associative memory. *IEEE Transactions on Information Theory*. 1987;33(4):461–482. DOI: 10.1109/TIT.1987.1057328.
8. Storkey A. Increasing the capacity of a hopfield network without sacrificing functionality. In: Gerstner W, Germond A, Hasler M, Nicoud JD. editors. *Artificial Neural Networks — ICANN'97*. ICANN 1997. Vol. 1327 of *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer; 1997. P. 451–456. DOI: 10.1007/BFb0020196.
9. Krotov D, Hopfield JJ. Dense associative memory for pattern recognition. In: *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*. 5–10 December 2016, Barcelona, Spain. New York: Curran Associates Inc.; 2016. P. 1180–1188. DOI: 10.5555/3157096.3157228.
10. Ramsauer H, Schäfl B, Lehner J, Seidl P, Widrich M, Adler T, Gruber L, Holzleitner M, Pavlović M, Sandve GK, Greiff V, Kreil D, Kopp M, Klambauer G, Brandstetter J, Hochreiter S. Hopfield networks is all you need [Electronic resource]. arXiv:2008.02217. arXiv Preprint; 2020. 94 p. Available from: <https://arxiv.org/abs/2008.02217>.
11. Klinshov VV, Nekorkin VI, Kurths J. Stability threshold approach for complex dynamical systems. *New Journal of Physics*. 2016;18(1):013004. DOI: 10.1088/1367-2630/18/1/013004.
12. Menck PJ, Heitzig J, Marwan N, Kurths J. How basin stability complements the linear-stability paradigm. *Nature Physics*. 2013;9(2):89–92. DOI: 10.1038/nphys2516.
13. Chakraborty A, Alam M, Dey V, Chattopadhyay A, Mukhopadhyay D. Adversarial attacks and defences: A survey [Electronic resource]. arXiv:1810.00069. arXiv Preprint; 2018. 31 p. Available from: <https://arxiv.org/abs/1810.00069>.
14. Amari SI, Maginu K. Statistical neurodynamics of associative memory. *Neural Networks*. 1988;1(1):63–73. DOI: 10.1016/0893-6080(88)90022-6.



*Соловьев Игорь Александрович* — родился в Нижнем Новгороде (2000). Окончил бакалавриат радиофизического факультета Нижегородского государственного университета им. Н. И. Лобачевского (2022). С 2021 года работает в Институте прикладной физики РАН. Область научных интересов — нелинейная динамика, динамика нейронных сетей.

Россия, Нижний Новгород 603950, ул. Ульянова, 46  
 Институт прикладной физики РАН  
 Россия, 603950 Нижний Новгород, пр-т Гагарина, 23  
 Нижегородский государственный университет им. Н. И. Лобачевского  
 E-mail: solocool46@gmail.com



*Клиньшов Владимир Викторович* — родился в Нижнем Новгороде (1984). Окончил радиофизический факультет Нижегородского государственного университета им. Н. И. Лобачевского (2006). Защитил диссертацию на соискание учёной степени кандидата физико-математических наук (2009, ИПФ РАН). С 2003 года работает в Институте прикладной физики РАН, в настоящее время — в должности старшего научного сотрудника. Область научных интересов — нелинейная динамика, динамические сети, синхронизация, системы с запаздыванием. Имеет более 60 научных публикаций по указанным направлениям. В 2017 году награжден медалью Российской академии наук с премией для молодых ученых РАН.

Россия, Нижний Новгород 603950, ул. Ульянова, 46  
 Институт прикладной физики РАН  
 Россия, 603950 Нижний Новгород, пр-т Гагарина, 23  
 Нижегородский государственный университет им. Н. И. Лобачевского  
 E-mail: vladimir.klinshov@ipfran.ru  
 ORCID: 0000-0003-4733-1352  
 AuthorID (eLibrary.Ru): 247610