THE UNIVERSITY OF CHICAGO

STUDIES OF THE ACCURACY, CAPACITY, AND EFFICIENCY OF MODERN HOPFIELD NETWORKS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
SHARBA BHATTACHARJEE

CHICAGO, ILLINOIS
AUGUST 2025

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Ivar Martin, for his support and advice. I also thank Profs. Michael Levin, Peter Littlewood, and William Irvine for agreeing to be in my dissertation committee and for their support and feedback. I also thank Prof. Kartiek Agarwal for his helpful insight and suggestions, in particular when working on quantum Hopfield networks. I also thank the staff of the Department of Physics at The University of Chicago for their help over the years. Finally, I would like to thank my parents and grandparents for their continued support and motivation, and my friends for making my time in grad school so much more enjoyable. Part this work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division.

ABSTRACT

The Hopfield network is an Ising spin system with long-range binary interactions. The interaction coefficients are defined such that a particular set of possible states of the system are the stable states, and the system tends to evolve into one of these states, exhibiting associative memory. The model can be generalized to have multi-spin interactions, resulting in systems called modern Hopfield networks, which have much larger memory capacities scaling as higher powers of the system size. In this thesis, I have studied how the retrieval accuracy and the capacity of modern Hopfield networks are affected by noise in the interaction coefficients mediating the multi-spin interactions. I also study the effect of clipping the interaction coefficients to all have the magnitude and only differ in sign, which can be a more convenient or efficient way of designing or simulating Hopfield networks as the interactions can be stored with smaller precision. I find that in most of these cases the capacity is reduced but its scaling behavior with the system size remains unchanged. While most of my work focuses on one step of synchronous Hopfield dynamics at zero temperature, I also investigate the equilibrium behavior of the system with noise-free fourth-order interactions using replica theory and obtain the phases of this system. Finally, I consider a possible quantum generalization of the network and compare its behavior with the classical network.

CHAPTER 1

INTRODUCTION

1.1 Hopfield network

Hopfield networks, introduced by John J. Hopfield in 1982 [1], are a class of artificial neural networks with binary neurons, which can be treated as Ising spins. The neurons interact via pairwise symmetric long-range interactions, and the interaction coefficients are determined by a subset of possible states of the network (called patterns). The patterns are the local minima of the energy of the system, and states close to one of the patterns evolve into that pattern. Hopfield networks thus function as models of associative (or content-addressable) memory, which are capable of recovering complete patterns from partial or noisy inputs. The interaction coefficients corresponding to a particular set of patterns are usually defined by the Hebb rule [2].

The number of patterns which can be embedded in a Hopfield network and retrieved successfully is limited, as for every pattern being retrieved, the other patterns effectively behave as noise in the interactions. Using replica theory, Amit, Gutfreund, and Sompolinsky [3] analyzed the memory capacity of the Hopfield model using replica theory, showing that the theoretical maximum number of patterns that can be reliably stored is approximately 0.138 times the number of neurons in the network. Using techniques from coding theory, McEliece et al. found the capacity of the network to be $N/2\ln(N)$ [4]. Subsequent studies have sought to improve the memory capacity by approaches such as using learning rules different the Hebb rule [5–7], though capacity still scales linearly with the system size in such models.

1.2 Modern Hopfield network

Modern Hopfield network, also called dense associative memory, is a class of neural models designed to extend the principles of classic associative memory systems to achieve enhanced storage capacity and robustness. Unlike classic Hopfield networks which utilize pairwise interactions between its component neurons, modern Hopfield networks generalize the energy function framework by incorporating higher-order interactions. This extension greatly enhances the memory capacity of the system.

One of the pioneering works on dense associative memories was by Krotov and Hopfield [8], who introduced Hopfield-like Ising systems with higher-order energy functions (as opposed to quadratic as in the classic model). This may be interpreted as arising from multi-spin interactions in the network. This allows the network to store a number of memories proportional to a higher power of the number of spins (specifically $\mathcal{O}(N^{n-1})$, for a network with N spins and nth order interactions) as opposed to the linear scaling in traditional Hopfield models.

Modern Hopfield networks have inspired further developments in both neural network theory and practical applications. Recent works [9, 10] have investigated connections between modern Hopfield networks and modern energy-based learning models, revealing that such networks exhibit trainable, powerful representations useful in both supervised and unsupervised learning.

1.3 Hopfield networks with synaptic noise

Any physical or biological realization of an associative memory model is expected to have noise in its interactions. The study of Hopfield networks with noisy interactions explores the effects of randomness or uncertainty in synaptic weights on the properties of memory storage, retrieval, and network dynamics.

While the exact form of the noise in the interactions will depend on the specifics of the physical system, it is often useful to consider additive and multiplicative noise. The latter can also correspond to networks where some of the interactions have been randomly deleted, if the multiplicative noise factors follow the Bernoulli distribution. Another network of interest is that with “quantized” or “clipped” interactions, where the interaction coefficients all have the same magnitude and only the sign is determined by Hebb rule. This allows the interaction coefficients to be stored with lower precision, which is of interest when designing potential physical realizations of such models as the parameters of the physical system may be tunable with finite precision. Storing the parameters with lower precision may also be an efficient way to implement artificial neural networks on computer hardware, and quantized parameters are sometimes used in deep learning models [11]. McEliece et al. showed that the capacity for the Hopfield network with clipped binary interactions would be related to that of the Hebbian model by a factor of $2/\pi$ [4], and Sompolinsky later verified it using mean field theory when he found the capacities of a variety of Hopfield networks with non-Hebbian couplings (including the clipped network as well as couplings with additive noise) [12]. Networks with multiplicative noise factors have been considered in Refs. [13–15]. Agliari et al. studied the capacity of modern Hopfield networks with multi-spin interactions in presence of additive noise in the coupling constants [16]. They found that the capacity of the network still scales with the system size as N^{n-1} in most cases, except when the additive noise scales with the number of patterns as well as a power of the system size.

1.4 Quantum Hopfield network

In recent years, much work has been done to extend classical neural networks to the quantum domain [17]. In particular, quantum generalizations to the Hopfield network have been considered. Rebentrost et al. [18] encoded N-component classical Hopfield networks using a system of $\lceil \log_2(N) \rceil$ qubits; however, most other approaches define the quantum Hopfield network as a system of N qubits, with dissipative dynamics defined by a Lindblad master equation. A common approach is to study Hopfield networks (with Ising interactions along σ_z) in presence of a transverse field along σ_x; such systems have been investigated by[19–27]. The transverse field term is found to effectively introduce noise in the system and reduces its memory capacity. Kimura and Kato [28] extended this approach to consider modern Hopfield networks (with higher order interactions along σ_z) in presence of a transverse field, whereas Seki and Nishimori [29] investigated modern Hopfield networks in presence of a transverse field as well as transverse antiferromagnetic interactions.

1.5 Description of chapters

In chapter 2, I present our work on the accuracy and capacity of modern Hopfield networks with noisy and clipped interactions, measured by their retrieval accuracy after one step of synchronous update. In chapter 3, I describe a technique for finding the phase diagram of the quartic modern Hopfield network (corresponding to its equilibrium behavior after many update steps at finite temperature) using replica theory. In chapter 4, I present some preliminary results on the behavior of a quantum generalization of the Hopfield network where the patterns are not all σ_z eigenstates. In chapter 5 I summarize my conclusions and suggest future research directions.

CHAPTER 2

ACCURACY AND CAPACITY OF MODERN HOPFIELD NETWORKS WITH SYNAPTIC NOISE

This chapter was co-authored by me and my dissertation advisor Ivar Martin. An earlier version of this material appeared in `arXiv.org` *as [30]. A later version has been submitted to a journal for publication.*

2.1 Introduction

The Hopfield network, proposed in 1982 [1], is a system of binary degrees of freedom (usually termed Ising spins or binary neurons) with long-range interactions, which exhibits a form of behavior known as associative memory. Specifically, we can choose a set of spin configurations (hereby called patterns), and define the coupling constants for the spin-spin interactions such that those patterns are the locally stable states of the system (with a spin configuration sufficiently close to any of the specified patterns relaxing to that pattern). This behavior is similar to a human brain reconstructing a memory when provided with incomplete information, and so the neural networks such as the Hopfield network have been used to model this process [31]. The coupling constants of the network are analogous to the strengths of the synaptic connections between neurons which encode biological memory; most commonly, the Hebb's learning rule is used to define them [2]. Hopfield networks have been physically realized in digital architecture [32] and confocal cavity QED systems [33].

The number of distinct patterns K that can be retrieved without significant error (network capacity) is determined by the number of spins N in the system. From probabilistic considerations and Monte Carlo simulations, Hopfield estimated the capacity of a network of size N to be $0.15N$ [1]. Using techniques from equilibrium statistical physics such as replica theory, Amit et al. [34] found that the overlap of a state of the system with the pattern closest to it at equilibrium can be nonzero for sufficiently few patterns (indicating small error), but jumps to zero when the number of patterns is approximately $0.138N$, exhibiting a first order phase transition. This critical number of patterns can thus be considered to be the capacity of the network as long as some retrieval errors were allowed. The above result assumed replica symmetry; later work which allowed for replica symmetry breaking [35] showed a somewhat higher capacity of $0.144N$. (This is the notion of capacity considered in most subsequent work including ours, though there have been other ways to define the retrieval phase of the system, such as in terms of the absence of spurious energy minima beyond the embedded patterns as in Ref. [36].) Using techniques from coding theory, McEliece et al. showed that if we require that for every pattern, every state within a Hamming distance of δN from that pattern (for $\delta < 1/2$) is retrieved perfectly after one step of synchronous update (with probability approaching 1 as $N \to \infty$), then the capacity of the Hopfield network is $(1-2\delta)^2 N/4\ln(N)$ [4].

The capacity of the Hopfield network has also been estimated in models where the couplings are not defined by the Hebb rule. Abu-Mostafa et al. showed that the capacity for any system with two-spin interactions is bounded above by N irrespective of how the couplings are defined [37]. An important example of alternatively defined couplings is the model with "quantized" or "clipped" couplings, where the coupling constants all have the same magnitude and only the sign is determined by Hebb rule. McEliece et al. showed that the capacity for this model would be related to that of the Hebbian model by a factor of $2/\pi$ [4], which was verified by Sompolinsky using mean field theory when he found the capacities of a class of models with non-Hebbian couplings (including the clipped network as well as couplings with additive noise) [12]. Networks with multiplicative noise factors (in particular where the multiplicative factor is 0 or 1 indicating the deletion of coupling terms) have been considered in Refs. [13–15].

The above models can be naturally extended to involve simultaneous interactions between more than two spins. Spin-glass models with multi-spin interactions have been studied before in refs. [38–41], as well as in the context of error correcting codes in Ref. [42]. Arenzon and de Almeida [43] and Lemke et al. [44] considered the equilibrium and dynamical properties of generalizations of the Hopfield network with n-spin interactions; Bovier and Niederhauser [45] and Baldi and Venkatesh [46] showed that the memory capacity of such systems scales with the system size N as N^{n-1}. Recently, Krotov and Hopfield generalized the Hopfield network to include a more general class of Hamiltonians, giving rise to networks with much higher capacities, called dense associative memory models or modern Hopfield networks [47]. Krotov and Hopfield had estimated the capacity of such models, and it was extended to include more general notions of capacity by Bao et al. [48], which accounted for finite basins of attraction around the patterns. Ramsauer et al. [49] introduced a modern Hopfield network whose capacity scales exponentially as system size (as shown for more general ensembles of patterns in [50]), and showed this network to be equivalent to the attention mechanism used in transformer models, a kind of neural network [51]. Such models have been analyzed in the context of Boltzmann machines [52], utilized to perform tasks such as in-context denoising [53] and storing a sequence of patterns [54], and refinements to their dynamics have also been proposed to further improve their memory capacity [55, 56] as well as to propose novel architectures such as energy transformers [57]. A system of Ising spins with n-spin interactions is one such class of dense associative memory ($n = 2$ being the conventional Hopfield network). Agliari et al. studied the capacity of modern Hopfield networks with multi-spin interactions in presence of additive noise in the coupling constants (e.g., resulting from errors when learning or storing the Hebbian interaction coefficients) [16]. They found that the capacity of the network still scales with the system size as N^{n-1} in most cases, except when the additive noise scales with the number of patterns as well as a power of the system size.

In this paper, we consider modern Hopfield networks with n-spin interactions, where the interaction coefficients deviate from the Hebb rule in a more general way. In particular, we consider models with additive and multiplicative noise in the interactions, as well as models where the interaction coefficients have been clipped to all have the same magnitude and only differ in sign. While the exact kind of noise in a physical or biological realization of a Hopfield network will depend on the details of the system, we considered additive noise as it is straightforward to analyze but can still be expected to be a reasonable approximation for the noise that may appear in realistic systems. While much of our analysis holds for general forms of multiplicative noise, we were mainly interested in the case where the noise factors were Bernoulli random variables (taking values 0 or 1), as this corresponds to a network where some interactions have been randomly deleted. We find that for the case of $n = 2$, our results agree with those found previously for the non-Hebbian Hopfield network by McEliece and Sompolinsky. For general n, we find that the capacity scales with the system size N as N^{n-1} in the presence of all these noise channels, significantly generalizing the result of Ref. [16] (which considered various kinds of additive noise in the networks but not multiplicative noise or clipping).

A complementary perspective to the memory storage/retrieval robustness described above is the one of the efficiency in terms of memory and computational requirements of a network. Storage of clipped synaptic weights requires only single bit per bond. Similarly, dropping a bond completely further reduces the storage requirements. Binary weights also allow to perform the retrieval computation faster than general weights. Therefore, the networks with clipped couplings can potentially outperform the “perfect" networks for the same memory storage tasks. We demonstrate that this is indeed the case.

We note that more sophisticated schemes exist for deleting less important couplings based on their information content, as a way to maximize the storage efficiency [58]. We do not consider this approach here; all interactions are modified randomly and independently with the same probability.

The rest of the paper is organized as follows. In section 2.2, we define the model of dense associative memory considered here. In section 2.3, we provide the definitions we have used for the capacity, following refs. [4] and [48]. In section 2.4, we estimate the storage capacity of modern Hopfield networks using statistical techniques. We find the capacity for networks with Hebbian interactions as well as those with noise or with clipped and/or deleted interactions, and compare their storage efficiencies. In section 2.5, we numerically simulate modern Hopfield networks and compare the results to our analytical estimates from the previous section. In section 2.6, we summarize our findings and mention possible applications and outlook.

2.2 Definition of the modern Hopfield network with synaptic noise

Consider a system of N Ising spins used to store K patterns $\boldsymbol{\xi}^\mu$, where we suppose that $N, K \gg 1$. The pattern components ξ_i^μ have been chosen randomly and are independent and identically distributed Rademacher random variables, with probability distribution $\mathrm{P}\big(\xi_i^\mu = 1\big) = \mathrm{P}\big(\xi_i^\mu = 1\big) = 1/2$, and moments $\mathrm{E}\big(\xi_i^\mu\big) = 0$, $\mathrm{E}\Big(\xi_i^{\mu 2}\Big) = 1$. Given set of memory patterns, we can define the Hamiltonian for an arbitrary state $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_N) \in \{1, -1\}^N$,

$$
\begin{aligned}
H_n^0(\boldsymbol{\sigma}) &= -\frac{1}{N^{n-1}} \sum_{1 \le i_1 < ... < i_n \le N} \sum_{\mu=1}^{K} \xi_{i_1}^\mu \cdots \xi_{i_n}^\mu \sigma_{i_1} \cdots \sigma_{i_n} \\
&= -\frac{1}{n! N^{n-1}} \sum_{i_1 \neq ... \neq i_n} \sum_{\mu=1}^{K} \xi_{i_1}^\mu \cdots \xi_{i_n}^\mu \sigma_{i_1} \cdots \sigma_{i_n},
\end{aligned} \tag{2.1}
$$

which is the model with n-spin interaction studied by Agliari et al. [16]. For $n > 2$, this is a dense associative memory model as defined by Krotov [47], which is made apparent by the fact that the above Hamiltonian can be written as a sum over n-th order polynomial functions of the overlaps $m_\mu = \sum_{i=1}^N \xi_i^\mu \sigma_i / N$ of the system with the patterns. In particular,

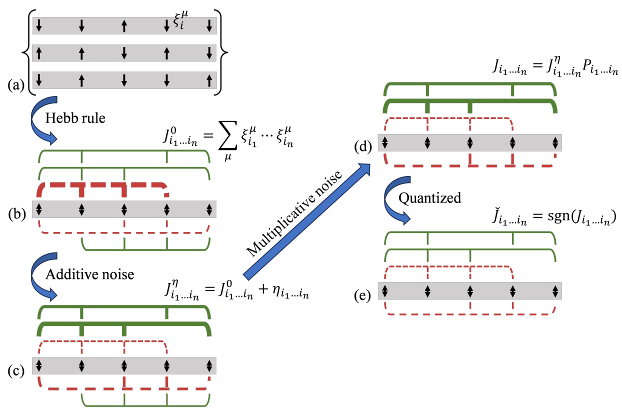

Figure 2.1: Schematic description of a modern Hopfield network with $N = 5$ spins, $K = 3$ patterns, and $n = 4$ spin interactions. (a) The set of patterns $\{\boldsymbol{\xi}^\mu\}$ used to define the interaction coefficients, with components $\xi_i^\mu \in \{\pm 1\}$ (shown as upward and downward arrows). (b) The n-spin interaction coefficients $J^0_{i_1 \ldots i_n}$ defined by the generalized Hebb rule. In this figure, each interaction connects $n = 4$ spins, shown by a branching line segment connecting the spins. The thickness of the line segment indicates the strength $\left| J^0_{i_1 \ldots i_n} \right|$ of the interaction. Positive interaction coefficients are shown using solid green lines, and negative ones using dashed red lines. (c) The interactions with additive noise $\eta_{i_1 \ldots i_n}$; in this example, they are independent Gaussian $N(\mu = 0, \sigma^2 = 4)$ random variables. (d) The interactions with additive as well as multiplicative noise $P_{i_1 \ldots i_n}$. In this example the multiplicative factors are Bernoulli random variables with parameter $p = 0.8$, which effectively "deletes" a fifth of the interactions. (e) Description of the alternate model considered in this paper, where the interactions are clipped to all have the same strength and differ only in sign.

as shown in appendix 2.9, in the limit of large N, we have

$$H_n^0(\boldsymbol{\sigma}) = -N \sum_{\mu=1}^{K} \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \frac{m_\mu^{n-2k}}{k!(n-2k)!(2N)^k}, \tag{2.2}$$

where the lower order corrections (corresponding to $k > 0$) appear because we have excluded self-interactions which effectively reduce the order of interaction. We note that for an arbitrary state $\boldsymbol{\sigma}$ and pattern $\boldsymbol{\xi}^\mu$, the overlap is of the order $m_\mu \sim 1/\sqrt{N}$, in which case every term in the above expression is of order $N^{1-n/2}$; however, if the state is close to a particular pattern (say $\boldsymbol{\xi}^1$), then $m_1 \sim 1$, and the terms in the sum over k have orders of N and lower.

The Hamiltonian H_n^0 in eq. (2.1) describes an Ising spin system system with n-spin interactions, mediated by the interaction constants $J^0_{i_1 \ldots i_n} = \sum_\mu \xi^\mu_{i_1} \cdots \xi^\mu_{i_n}$; we assume that $N \gg n$. This is the extension of the Hebb rule to higher order interactions. We can now generalize the model to introduce some synaptic noise in the interaction constants. Specifically, we consider the situation where the Hamiltonian is

$$H_n(\boldsymbol{\sigma}) = -\frac{1}{N^{n-1}} \sum_{1 \le i_1 < \ldots < i_n \le N} J_{i_1 \ldots i_n} \sigma_{i_1} \cdots \sigma_{i_n}, \tag{2.3}$$

which corresponds to n-body interactions mediated by the interaction coefficients

$$J_{i_1 \ldots i_n} = \left(\sum_\mu \xi^\mu_{i_1} \cdots \xi^\mu_{i_n} + \eta_{i_1 \ldots i_n} \right) P_{i_1 \ldots i_n}, \tag{2.4}$$

where the random variables $\eta_{i_1 \ldots i_n}$ and $P_{i_1 \ldots i_n}$ correspond to additive and multiplicative noise respectively. We assume that all of the random variables are independent. The variables $\eta_{i_1 \ldots i_n}$ are taken to have identical probability distributions, and thus have the same moments $\mathrm{E}\left(\eta_{i_1 \ldots i_n}\right) = \mu_\eta$ and $\mathrm{E}\left(\eta^2_{i_1 \ldots i_n}\right) = \mu_{\eta^2}$. In particular, we can consider the case where $\eta_{i_1 \ldots i_n} \sim N(0, \mu_{\eta^2})$, corresponding to Gaussian additive noise. Such additive noise changes the interaction strengths and may also flip the sign of some weaker interactions. Similarly, the variables $P_{i_1 \dots i_n}$ corresponding to multiplicative noise also have identical probability distributions, with moments $\mathrm{E}\big(P_{i_1 \dots i_n}\big) = \mu_P$ (which we will assume to be positive) and $\mathrm{E}\Big(P_{i_1 \dots i_n}^2\Big) = \mu_{P^2}$. Such multiplicative noise can also affect the interaction strengths, and it can also randomly delete some interactions (if $P_{i_1 \dots i_n}$ is zero) or flip their sign (if $P_{i_1 \dots i_n}$ is negative).

The mapping $\mathbf{T}_n$ corresponding to one step of synchronous update of the system, according to the Hamiltonian with interaction constants defined above, is

$$\begin{aligned}(\mathbf{T}_n \boldsymbol{\sigma})_i &= \operatorname{sgn}\left(\sum_{\substack{1 \le i_2 < \cdots < i_n \le N, \\ i_a \neq i \forall a}} J_{i i_2 \dots i_n} \sigma_{i_2} \cdots \sigma_{i_n} \right) \\ &= \operatorname{sgn}\left(\sum_{\mu, \{i_a\}} \xi_i^\mu \xi_{i_2}^\mu \cdots \xi_{i_n}^\mu \sigma_{i_2} \cdots \sigma_{i_n} P_{i i_2 \dots i_n} + \sum_{\{i_a\}} \sigma_{i_2} \cdots \sigma_{i_n} \eta_{i i_2 \dots i_n} P_{i i_2 \dots i_n} \right) \end{aligned} \tag{2.5}$$

2.3 Retrieval and capacity

The Hopfield network is a model for associative memory, where the patterns $\boldsymbol{\xi}^\mu$ are the stable states satisfying $\mathbf{T}_n(\boldsymbol{\xi}^\mu) = \boldsymbol{\xi}^\mu$ for all μ, and states that are close to a pattern (in Hamming distance) are mapped to the corresponding pattern ("retrieved") under the Hopfield update $\mathbf{T}$. To quantify this, we follow the work of McEliece et al. [4] and Bao et al. [48], and consider the following setup (as described schematically in figure 2.2):

1. Start with an arbitrary stored memory pattern, say $\boldsymbol{\xi}^1$.

2. Perturb the pattern by flipping each of its component spins independently with probability δ ($0 \le \delta < 1/2$) to obtain the state $\tilde{\boldsymbol{\xi}}^1$, where $\tilde{\xi}_i^1 = s_i \xi_i^1$ for i.i.d. random variables s_i taking the value 1 (-1) with probability $1 - \delta$ (δ). This is equivalent to retaining $1 - 2\delta$ of the original pattern components and assigning the rest randomly to be ± 1 with equal probability. $\tilde{\boldsymbol{\xi}}^\mu$ has a Hamming distance of $N\delta$ from $\boldsymbol{\xi}^\mu$ on average,

which can be used to define the basin of attraction around each pattern that we would want to be retrieved. $\delta = 0$ corresponds to the case where we only care if the patterns themselves are stable.

3. Perform one step of the Hopfield update on $\tilde{\boldsymbol{\xi}}^1$ to obtain the state $\mathbf{T}_n\tilde{\boldsymbol{\xi}}^1$, where

$$(\mathbf{T}_n\tilde{\boldsymbol{\xi}}^1)_i = \operatorname{sgn}\left(\sum_{\substack{1\leq i_2<\cdots<i_n\leq N,\\ i_a\neq i\forall a}} J_{ii_2\ldots i_n}\xi^1_{i_2}\ldots\xi^1_{i_n}s_{i_2}\ldots s_{i_n} \right). \tag{2.6}$$

4. Find the overlap $m = \sum_i \xi_i^1(\mathbf{T}_n\tilde{\boldsymbol{\xi}}^1)_i/N$ of the obtained state with the original pattern $\boldsymbol{\xi}^1$ to quantify how close it is to the original pattern. $m = 1$ corresponds to perfect retrieval, $\mathbf{T}_n\tilde{\boldsymbol{\xi}}^1 = \boldsymbol{\xi}^1$.

The overlap m defined as above is itself a random variable, where $m = 1$ corresponds to perfect retrieval. Therefore, we can use m to quantify how accurately the patterns are retrieved. One measure of retrieval quality is the expected overlap, $\mathrm{E}(m)$. This is a convenient measure, as it is easy to estimate using the specified probability distributions of ξ_i^μ, $\eta_{i_1\ldots i_n}$, $P_{i_1\ldots i_n}$, and s_i, and invoking central limit theorem where appropriate. Defining $m_i = \operatorname{sgn}\Big(\sum_{i_2,\ldots,i_n\neq i} J_{ii_2\ldots i_n}\xi_i^1\xi_{i_2}^1\cdots\xi_{i_n}^1 s_{i_2}\cdots s_{i_n}\Big)$, we note that $\mathrm{E}(m) = \mathrm{E}(m_i)$, so to find $\mathrm{E}(m)$ we just need to find the probability of one of the spins being retrieved correctly. Also, if we were instead estimating the capacity of the network using the techniques of equilibrium statistical physics (as done by Amit et al. [3]), we would be quantifying retrieval using an order parameter which is analogous to $\mathrm{E}(m)$.

An alternate measure for estimating retrieval would be to use probabilities instead of expected values. The probability $\mathrm{P}(m_i = 1)$ of any particular spin being retrieved correctly is linearly related to $\mathrm{E}(m)$. We can also consider the probability $\mathrm{P}(m \geq m_0)$, of the return overlap being greater than some threshold value m_0. In particular, $\mathrm{P}(m = 1) = \mathrm{P}(m_i = 1\forall i)$ is the probability of perfect retrieval. This is a useful measure as in some contexts, we may

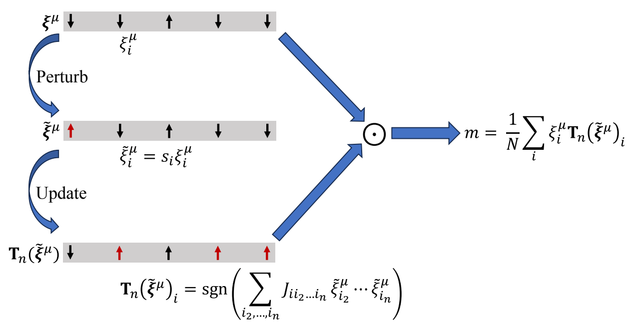

Figure 2.2: Schematic description of the procedure used to describe the retrieval accuracy for a modern Hopfield network with $N = 5$ spins. We consider a pattern, say $\boldsymbol{\xi}^\mu$ (where $\mu \in \{1, \dots, K\}$), and perturb it by randomly and independently flipping each spin ξ_i^μ with probability δ. That is, the perturbed state is $\tilde{\boldsymbol{\xi}}^\mu$, with components $\tilde{\xi}_i^\mu = s_i \xi_i^\mu$, where s_i can independently take values 1 and -1 with probabilities $1-\delta$ and δ respectively. In this figure, $\delta = 0.2$. One step of the Hopfield update $\mathbf{T}_n$ is then applied on the perturbed pattern, where n is the order of the interaction ($n = 2$ corresponds to the usual Hopfield network). We find the overlap m between the updated state $\mathbf{T}_n(\tilde{\boldsymbol{\xi}^\mu})$ and the original pattern $\boldsymbol{\xi}^\mu$, which is a measure of how accurately the pattern is retrieved after one step of the Hopfield update. m is a random variable, so we can study its expected value or probability distribution. In this figure, we have $m = -0.2$ as only 2 of the 5 components of $\mathbf{T}_n(\tilde{\boldsymbol{\xi}^\mu})$ match with $\boldsymbol{\xi}^\mu$. This is because the update protocol was based on the 4-spin interaction model with synaptic noise described in Fig. 2.1, for which the retrieval accuracy is strongly affected by the high perturbation rate $\delta = 0.2$ in this example.

care more whether the pattern is retrieved perfectly (or more accurately than some fixed threshold) or not, and in such cases it is useful to know the probability of “successful” retrieval under these criteria. This is, however, more difficult to estimate, as the states of the different spins after the Hopfield update are not independent, so the probability of perfect retrieval can not be simply written as $\mathrm{P}(m = 1) = \prod_i \mathrm{P}(m_i = 1)$. There is also no order parameter analogous to this probability when analyzing the behavior of the network using mean field theory.

We expect both $\mathrm{E}(m)$ and the probability of perfect retrieval $\mathrm{P}(m = 1)$ to be close to 1 for small number of patterns K and decrease for larger K. We can then define the capacity to be the maximum K for which the chosen measure of retrieval error remains below a particular tolerance threshold. It is common to choose the threshold to be $1/N$ (so it vanishes in the limit $N \to \infty$), so a commonly used definition of capacity is the number of patterns K for which the probability of *im*perfect retrieval is $\mathrm{P}\Big(\mathbf{T}_n\tilde{\boldsymbol{\xi}}^1 \neq \boldsymbol{\xi}^1\Big) = 1/N$.

2.4 Estimation of the retrieval accuracy and capacity

Following the prescription of the previous section, here we compute the retrieval accuracy of the system after a single step of Hopfield update in a system with n-spin interaction for different noise models. In appendix 2.7 we provide further algebraic details behind the analysis, and in appendix 2.10 we work through the derivation of the results for the usual Hopfield network with two-spin couplings (in which case we obtain the same results as previous work, e.g., Ref. [4])

It is convenient to define the random variable $X_i = \xi_i^1 \sum_{i_2 \ldots i_n} J_{ii_2 \ldots i_n} \tilde{\xi}_{i_2}^1 \ldots \tilde{\xi}_{i_n}^1$, which is positive if the ith spin is retrieved correctly and negative otherwise. Its mean and variance can be found to be

$$\mathrm{E}(X_i) = \binom{N-1}{n-1}\mu_P(1-2\delta)^{n-1}, \tag{2.7}$$

$$\mathrm{V}(X_i) = \binom{N-1}{n-1}K\mu_{P^2} + \binom{N-1}{n-1}\mu_{\eta^2}\mu_{P^2} - \binom{N-1}{n-1}\mu_P^2(1-2\delta)^{2(n-1)}$$
$$+ \mu_P^2(1-2\delta)^{2(n-1)}\binom{N-1}{n-1}\sum_{c=1}^{n-2}\binom{N-n}{n-1-c}\binom{n-1}{c}\Big((1-2\delta)^{-2c}-1\Big). \tag{2.8}$$

(These have been derived in detail in appendix 2.7.) To find the distribution of the $\mathrm{sgn}(X_i)$, we assume that the distribution of X_i is Gaussian (comparison with numerical simulations reveals this to be a good approximation, even though the simplest form of the central limit theorem is not applicable here as X_i is the sum of random variables which are not all independent and identically distributed). Then we find that

$$\mathrm{P}(X_i > 0) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\mathrm{E}(X_i)}{\sqrt{2\,\mathrm{V}(X_i)}}\right), \tag{2.9}$$

and the overlap with the pattern has expected value

$$\begin{aligned}\mathrm{E}(m) &= \mathrm{E}(\mathrm{sgn}(X_i))\\ &= \mathrm{erf}\left(\sqrt{\frac{1}{2}\binom{N-1}{n-1}}\left[(K+\mu_{\eta^2})(1-2\delta)^{-2(n-1)}\frac{\mu_{P^2}}{\mu_P^2} - 1\right.\right.\\ &\qquad \left.\left. + \sum_{c=1}^{n-2}\binom{N-n}{n-1-c}\binom{n-1}{c}\Big((1-2\delta)^{-2c}-1\Big)\right]^{-1/2}\right).\end{aligned} \tag{2.10}$$

Since we are considering the case where $N \gg 1$, and if we assume $K \sim N^{n-1}$ (which is the capacity of the network as we show below), this can be approximated as

$$\mathrm{E}(m) = \mathrm{erf}\left((1-2\delta)^{n-1}\sqrt{\frac{\binom{N-1}{n-1}\mu_P^2}{2\mu_{P^2}(K+\mu_{\eta^2})}}\right). \tag{2.11}$$

2.4.1 Noise-free Hebbian interactions

Let us first consider the special case where the interactions are Hebbian with no noise (i.e., $\eta_{i_1 \ldots i_n} = 0$ and $P_{i_1 \ldots i_n} = 1$ with probability 1), in which case we have

$$\mathrm{E}(m) = \mathrm{erf}\left((1-2\delta)^{n-1} \sqrt{\frac{\binom{N-1}{n-1}}{2K}} \right) = \mathrm{erf}\left(\frac{(1-2\delta)^{n-1}}{\sqrt{2(n-1)!\alpha_n}} \right), \tag{2.12}$$

where $\alpha_n = K/[(N-1)\cdots(N-n+1)] \approx K/N^{n-1}$ (if $N \gg n^2$). Perfect retrieval $\mathbf{T}_n \tilde{\boldsymbol{\xi}}^1 = \boldsymbol{\xi}^1$ corresponds to $m = 1$. Therefore, if we define the capacity of the model as the maximum number of patterns for which the expected overlap is above some threshold value of $m_c = 1 - \epsilon$, then the capacity is found to be

$$K_c = \frac{(1-2\delta)^{2(n-1)}}{2\,\mathrm{erf}^{-1}(m_c)^2} \binom{N-1}{n-1} \approx \frac{(1-2\delta)^{2(n-1)}}{\ln\bigl(2/\pi\epsilon^2\bigr)} N^{n-1}, \tag{2.13}$$

which will scale with the system size as N^{n-1}. If we choose $\epsilon = 1/N$ (which is the case when $\mathbf{T}_n \tilde{\boldsymbol{\xi}}^1$ differs from $\boldsymbol{\xi}^1$ only at a finite number of spins), then the corresponding capacity is

$$K_c \approx (1-2\delta)^{2(n-1)} \frac{N^{n-1}}{2\ln(N)}. \tag{2.14}$$

We know that the error function for $x \ll 1$ is approximately linear, $\mathrm{erf}(x) \approx 2x/\sqrt{\pi}$, and is close to 1 and almost flat for $x \gg 1$, $\mathrm{erf}(x) \approx 1 - e^{-x^2}/x\sqrt{\pi}$. Therefore, the expected overlap $\mathrm{E}(m)$ approaches 1 when $K = 1$ (assuming δ is small enough), and decays as $K^{-1/2}$ when K is of the order of N^{n-1} or larger. The factor of $(1-2\delta)^{n-1}$ in the argument of the error function implies that that the retrieval accuracy depends strongly on deviation δ of the perturbed initial state from the perfect pattern, in particular for higher interaction orders n. For $K \ll [(1-2\delta)^2 N]^{n-1}$ the expected overlap has a plateau at 1, before falling as K is increased. We verify this numerically in section 2.5.

2.4.2 Noisy interactions

Including the additive and multiplicative noise terms $\eta_{i_1 \dots i_n}$ and $P_{i_1 \dots i_n}$ affects the expected overlap $\mathrm{E}(m)$ by changing the argument of the error function in eq. (2.11) compared to eq. (2.12). The capacity of the noisy model (corresponding to an expected overlap of at least $m_c = 1 - \epsilon$) is

$$K_c \approx \frac{\mu_P^2}{\mu_{P^2}} \frac{(1-2\delta)^{2(n-1)}}{\ln\left(2/\pi\epsilon^2\right)} N^{n-1} - \mu_{\eta^2} \tag{2.15}$$

$$\approx \frac{\mu_P^2}{\mu_{P^2}} (1-2\delta)^{2(n-1)} \frac{N^{n-1}}{2\ln(N)} - \mu_{\eta^2} \quad \left(\text{if } \epsilon = \frac{1}{N}\right). \tag{2.16}$$

The additive noise effectively replaces the number of patterns K in eq. (2.12) with $K+\mu_{\eta^2}$ in eq. (2.11), where μ_{η^2} is nonnegative, and zero if and only if there is no additive noise. This affects the retrieval accuracy the same way as adding more patterns would, and the capacity is reduced by μ_{η^2}. Therefore, the additive noise significantly affects the capacity only if μ_{η^2} is of the order of N^{n-1} or larger. Thus not only can the modern Hopfield networks store many more patterns than the original $n = 2$ model, but they are also more robust to additive noise in their interactions. We also note that, at least as long as N and K are large enough that eq. (2.11) holds, the retrieval accuracy only depends on its second order raw moment (mean squared noise) μ_{η^2}. Therefore, a constant shift, a Gaussian noise centered at 0, and a Gaussian noise with an offset will affect the retrieval accuracy identically as long as they have the same μ_{η^2}.

Similar to the capacity defined above, we can also define a critical noise level $\mu_{\eta^2,c}$, which is the mean squared additive noise in the interaction coefficients of a network (with given number of patterns K) for which the expected overlap is at least $m_c = 1 - \epsilon$. This has the expression

$$\mu_{\eta^2,c} \approx \frac{\mu_P^2}{\mu_{P^2}} \frac{(1-2\delta)^{2(n-1)}}{\ln\left(2/\pi\epsilon^2\right)} N^{n-1} - K \tag{2.17}$$

$$\approx \frac{\mu_P^2}{\mu_{P^2}} (1-2\delta)^{2(n-1)} \frac{N^{n-1}}{2\ln(N)} - K, \quad \left(\text{if } \epsilon = \frac{1}{N}\right). \tag{2.18}$$

and thus if the network is not loaded to capacity (i.e. $K \ll N^{n-1}$, but still large enough that the approximations we made to obtain eq. (2.11) remain valid), the critical noise level is $\mu_{\eta^2,c} \sim N^{n-1}$. Therefore, for the same number of patterns, the modern Hopfield networks with higher interaction orders n can retrieve the patterns accurately in presence of much larger additive noise.

The multiplicative noise, on the other hand, multiplies the argument of the error function in eq. (2.11) by a factor of μ_P^2/μ_{P^2}, which is 1 if $P_{i_1 \ldots i_n}$ is a degenerate random variable (i.e., multiplying the interaction coefficients by a constant does not affect the retrieval accuracy) and less than 1 otherwise. This results in a reduction of the capacity by a factor of μ_P^2/μ_{P^2} irrespective of the interaction order n, as seen in eq. (2.15). As a special case, we can consider the situation where $P_{i_1 \ldots i_n}$ are i.i.d. Bernoulli random variables with probability p of being 1 and $1-p$ being 0; this corresponds to the scenario where some interaction terms have been randomly and independently deleted with probability $1-p$. In this case, we see that $\mu_P = \mu_{P^2} = p$, the capacity of the system is reduced by a factor of p. That is, the capacity is proportional to the fraction of retained coupling constants, which could be expected qualitatively.

2.4.3 Clipped couplings

In this section study the model where the interaction constants have all been constrained to be ± 1 using a sign function,

$$\check{J}_{i_1 \dots i_n} = \operatorname{sgn}\left[\left(\sum_\mu \xi^\mu_{i_1} \cdots \xi^\mu_{i_n} + \eta_{i_1 \dots i_n}\right) P_{i_1 \dots i_n}\right] = \operatorname{sgn}\left(\sum_\mu \xi^\mu_{i_1} \cdots \xi^\mu_{i_n} + \eta_{i_1 \dots i_n}\right) \check{P}_{i_1 \dots i_n}. \tag{2.19}$$

The projection to binary values does not itself introduce extra noise. However, the way noises affect the capacity is different from the Hebbian case. We take the sign of the interaction term *after* introducing the additive and multiplicative noise to ensure that the interaction coefficients retain the magnitude 1 (or 0) even with noise. Therefore, the additive noise $\eta_{i_1 \dots i_n}$ in this case effectively flips the sign of some of the interaction coefficients, with weaker interactions being more likely to be flipped. Only the sign $\check{P}_{i_1 \dots i_n} = \operatorname{sgn}\big(P_{i_1 \dots i_n}\big)$ of the multiplicative noise factor matters, so we will mainly consider the cases where the only possible values of $P_{i_1 \dots i_n}$ are 0 or ± 1, in which case $\check{P}_{i_1 \dots i_n} = P_{i_1 \dots i_n}$. In particular, if $P_{i_1 \dots i_n} \sim \text{Bernoulli}(p)$ are i.i.d. Bernoulli random variables, it corresponds to interactions being randomly deleted with probability $1-p$ as before. We can also consider the case where $P_{i_1 \dots i_n}$ takes the values 1 and -1 with probability q and $1-q$ respectively, which would be interpreted as the couplings being passed through a noisy channel which randomly flips the signs of some of the interactions.

We can similarly as before perturb a pattern, apply one step of the Hopfield update, and find its overlap with the original pattern. The variable determining whether the i-th spin is retrieved correctly in this case is (noting that $\operatorname{sgn}(s_j) = s_j$)

$$\check{X}_i = \sum_{\substack{1 \le i_2 < \cdots < i_n \le N, \\ i_a \ne i \forall a}} \operatorname{sgn}\left(\sum_{\mu=1}^{K} \xi_i^1 \xi_i^\mu \prod_{a=2}^{n} \left(\xi_{i_a}^1 \xi_{i_a}^\mu \right) + \xi_i^1 \eta_{ii_2\ldots i_n} \prod_{a=2}^{n} \xi_{i_a}^1 \right) \check{P}_{ii_2\ldots i_n} \prod_{a=2}^{n} s_{i_a}$$

$$\equiv \sum_{\{i_a\}} \operatorname{sgn}\left(W_{ii_2\ldots i_n}\right) s_{i_2} \cdots s_{i_n} \check{P}_{ii_2\ldots i_n} \equiv \sum_{\{i_a\}} \operatorname{sgn}\left(X_{ii_2\ldots i_n}\right). \tag{2.20}$$

The expectations and variances of the random variables appearing in the above expression can be found in the same manner as in the case where the interactions were not clipped, which is use to find that

$$\mathrm{E}\left(W_{ii_2\ldots i_n}\right) = 1, \tag{2.21}$$

$$\mathrm{V}\left(W_{ii_2\ldots i_n}\right) = (K + \mu_{\eta^2}) - 1. \tag{2.22}$$

Assuming $W_{ii_2\ldots i_n}$ is approximately Gaussian (which holds if K is large and $\eta_{ii_2\ldots i_n}$ is approximately Gaussian), we find the expected value of the sign of $X_{ii_2\ldots i_n}$ is

$$\mathrm{E}\left(\operatorname{sgn}\left(X_{ii_2\ldots i_n}\right)\right) = \operatorname{erf}\left(\frac{1}{\sqrt{2(K + \mu_{\eta^2}) - 2}} \right) (1 - 2\delta)^{n-1} \mu_{\check{P}} \equiv \check{e}, \tag{2.23}$$

where $\mu_{\check{P}} = \mathrm{P}\left(P_{ii_2\ldots i_n} > 0\right) - \mathrm{P}\left(P_{ii_2\ldots i_n} < 0\right)$, and its variance is

$$\mathrm{V}\left(\operatorname{sgn}\left(X_{ii_2\ldots i_n}\right)\right) = \mu_{\check{P}^2} - \check{e}^2, \tag{2.24}$$

where $\mu_{\check{P}^2} = \mathrm{P}\left(P_{ii_2\ldots i_n} \ne 0\right)$, and we assume $P(W_{ii_2\ldots i_n} = 0) = 0$ (which holds when K is large or when the additive noise is continuous). On the other hand, for two sets $\{i_a\} \ne \{j_b\}$, the sign of the product $X_{ii_2\dots i_n}X_{ij_2\dots j_n}$ is

$$\mathrm{E}\big(\mathrm{sgn}\big(X_{ii_2\dots i_n}X_{ij_2\dots j_n}\big)\big) = \mathrm{erf}\left(\frac{1}{\sqrt{2(K+\mu_{\eta^2})^2-2}}\right)\mu_P^2(1-2\delta)^{2(n-1-c)} \equiv \check{d}_c, \tag{2.25}$$

where $c = |\{i_1,\dots,i_n\}\cap\{j_1,\dots,j_n\}|$, and so their covariance is

$$\mathrm{Cov}\big(\mathrm{sgn}\big(X_{ii_2\dots i_n}\big),\mathrm{sgn}\big(X_{ij_2\dots j_n}\big)\big) = \check{d}_c - \check{e}^2. \tag{2.26}$$

Since $\mathrm{sgn}\big(X_{ii_2\dots i_n}\big)$ are not independent, we can not directly apply the central limit theorem. However, for large K, the pairwise covariances $\mathrm{Cov}\big(\mathrm{sgn}\big(X_{ii_2\dots i_n}\big),\mathrm{sgn}\big(X_{ij_2\dots j_n}\big)\big) \sim 1/K$ are negligible compared to their standard deviations, so it may be reasonable to ignore the correlations and also assume that their sum $\check{X}_i$ is approximately Gaussian; we later find that our results under these approximations agree with simulations. We can then estimate that

$$\mathrm{E}\big(\check{X}_i\big) = \binom{N-1}{n-1}\check{e}, \tag{2.27}$$

$$\mathrm{V}\big(\check{X}_i\big) = \binom{N-1}{n-1}(\mu_{\check{P}^2} - \check{e}^2). \tag{2.28}$$

Therefore, the expected overlap is

$$\begin{aligned}\mathrm{E}(\check{m}) = \mathrm{E}\big(\mathrm{sgn}\big(\check{X}_i\big)\big) &= \mathrm{erf}\left(\frac{\mathrm{E}\big(\check{X}_i\big)}{\sqrt{2\,\mathrm{V}\big(\check{X}_i\big)}}\right)\\ &= \mathrm{erf}\left(\sqrt{\frac{\binom{N-1}{n-1}\check{e}^2}{2(\mu_{\check{P}^2}-\check{e}^2)}}\right).\end{aligned} \tag{2.29}$$

For $K \gg 1$, we have $\check{e} \approx \sqrt{2\mu_{\check{P}}^2(1-2\delta)^{2(n-1)}/\pi(K+\mu_{\eta^2})} \ll 1$, which implies

$$\mathrm{E}(\check{m}) \approx \operatorname{erf}\left(\sqrt{\frac{(1-2\delta)^{2(n-1)}\mu_{\check{P}}^2\binom{N-1}{n-1}}{\pi\mu_{\check{P}^2}(K+\mu_{\eta^2})}}\right). \tag{2.30}$$

The capacity $\check{K}_c$ corresponding to $\mathrm{E}(\check{m}) \geq m_c = 1-\epsilon$ for the clipped network is

$$\check{K}_c \approx \frac{\mu_{\check{P}}^2}{\mu_{\check{P}^2}} \frac{2(1-2\delta)^{2(n-1)}}{\pi\ln\left(2/\pi\epsilon^2\right)} N^{n-1} - \mu_{\eta^2} \tag{2.31}$$

$$\approx \frac{2}{\pi}\frac{\mu_{\check{P}}^2}{\mu_{\check{P}^2}}(1-2\delta)^{2(n-1)}\frac{N^{n-1}}{2\ln(N)} - \mu_{\eta^2} \quad \left(\text{if } \epsilon = \frac{1}{N}\right). \tag{2.32}$$

Comparing this with the expression in eq. (2.15), where the interaction coefficients were not clipped, we note that (in the case where $P_{i_1\ldots i_n} = \check{P}_{i_1\ldots i_n}$) the capacities differ by a factor of $2/\pi$ in the first term. Therefore, the capacity of this model is less than that of the unclipped model by a factor of $2/\pi$ (for small additive noise), but otherwise scales with system size the same way.

If $P_{i_1\ldots i_n}$ can take values other than $0, \pm 1$ with nonzero probability, then the capacity is further modified as it now involves a factor of $\mu_{\check{P}}^2/\mu_{\check{P}^2}$ instead of μ_P^2/μ_{P^2}. For example, in a network with no additive noise and Gaussian multiplicative noise, $P_{i_1\ldots i_n} \sim N(\mu_P, \sigma_P^2)$ (i.i.d.), we have $\check{K}_c/K_c = 2(1+v^2)\operatorname{erf}\left(1/\sqrt{2}v\right)^2/\pi$, where $v = \sigma_P/\mu_P$. This ratio has the highest value 0.726 when $\sigma_P = 0.468\mu_P$, and approaches $(2/\pi)^2$ for large σ_P. If the multiplicative noise has large variation in its magnitude but small or no variation in its sign (e.g., it is positive with probability 1), then it is possible for the capacity of the clipped network to be larger than that of the unclipped network for certain probability distributions of the multiplicative noise (though such distributions may not be realistic in physical systems). For example, if $P_{i_1\ldots i_n}$ are i.i.d. exponential random variables with any rate parameter λ, then $\check{K}_c \geq K_c$, and in particular $\check{K}_c/K_c = 4/\pi$ if there is no additive noise.

Therefore, we find that the capacity in this case will scale as N^{n-1} (as found before in [16, 47, 48]). The capacity also depends strongly on the size of the required basin of attraction around each pattern; if we require patterns with a Hamming distance of up to $N\delta$ from a pattern to be retrieved, the capacity has a factor of $(1-2\delta)^{2(n-1)}$. Finally, if we use binary couplings, this introduces the same factor of $2/\pi$ in the capacity as had been previously obtained for the network with two-spin interactions [4]. Therefore, if we wish to design a network with a given capacity, the number of spins the network needs to have will only go up by a factor of $(\pi/2)^{1/(n-1)}$, which approaches 1 for large n.

2.5 Numerical simulations

We numerically simulated the dynamics of modern Hopfield networks using the NumPy library in Python3 [59] to compare against their predicted retrieval accuracy. The memory patterns ξ^μ were generated as random vectors of size N (with every element equal to ± 1), and the interaction coefficients $J_{i_1 \ldots i_n}$ were computed using the Hebb rule and stored as an n dimensional array. To determine the retrieval accuracy, we chose one of the patterns, perturbed it by flipping the sign of each of its components independently with probability δ, implemented one step of the Hopfield update on it using the stored interaction coefficients, and then found the overlap m of the resulting vector with the original pattern. We repeated the above procedure 100 times (for each network description and each number of patterns K) to obtain a sample large enough to estimate statistical parameters while computing in a reasonable amount of time. (For the plots in figs. 2.3 and 2.5, we generated 10 different network realizations, corresponding to 10 different sets of K patterns and each with a different noise profile, and for each network we chose 10 patterns randomly with replacement to perturb and update. For the plots in figs. 2.4 and 2.6, we generated 100 network realizations for each value of K, and for each network we chose one of the patterns to perturb and update.) We used the simulation data to estimate the average overlap $\mathrm{E}(m)$, and then repeated this whole procedure for varying numbers of patterns (keeping the system size N and perturbation probability δ fixed) to obtain a plot of the average overlap as a function of the number of patterns using the Python library Matplotlib [60].

The analytical expression obtained in eq. (2.12) was also calculated using the Python library Scipy [61] and plotted for comparison with the simulation results. Because this expression depends on K and N only via $\alpha_n = K/(N-1)\cdots(N-n+1)$, we expect plots of $\mathrm{E}(m)$ vs. α_n to agree with this expression for all sufficiently large values of N (with the possible exception when $\alpha_n \ll 1$, in which case the approximation used to obtain eqs. (2.11) and (2.12) (which assumed $K \sim N^{n-1}$) is not valid). In figure 2.3, we plot the average overlap $\mathrm{E}(m)$ vs. the scaled number of patterns $\alpha_n = K/(N-1)\cdots(N-n+1)$ (keeping N fixed and varying K for each plot) for two sets of modern Hopfield networks. The first plot shows the results for $N = 10, 20, 30$ spins forming a modern Hopfield network with $n =$ 3-spin interactions,, whereas the second plot is for $N = 10, 12, 15$ spins interacting vis $n =$ 4-spin interactions. In each case, the perturbation rate is $\delta = 0.1$. We find results for different N to be in agreement with each other and with the estimate given by eq. (2.12), except for smaller values $\alpha_3 < 0.3$ and $\alpha_4 < 0.1$. In the first plot, we note that the average overlap plateaus near 1 for very small α_3 and then falls off with α_3 as expected. For the second plot, because of the smaller system sizes, the overlap is less than 1 even for very few memories and falls off with K without a plateau.

Similarly, we simulated networks where the interaction coefficients had additive and/or multiplicative noise. In particular, we simulated the cases where the additive term on each interaction coefficient was a normal random variable centered at 0, and the multiplicative factor was a Bernoulli random variable taking the value 1 with probability p and 0 otherwise. Figure 2.4 plots the expected overlap vs the scaled number of patterns α_3 for a modern Hopfield network of size $N = 20$ with $n =$ 3-spin interactions, where the additive noise is Gaussian with mean 0 and variance 100, and the multiplicative noise follows a Bernoulli

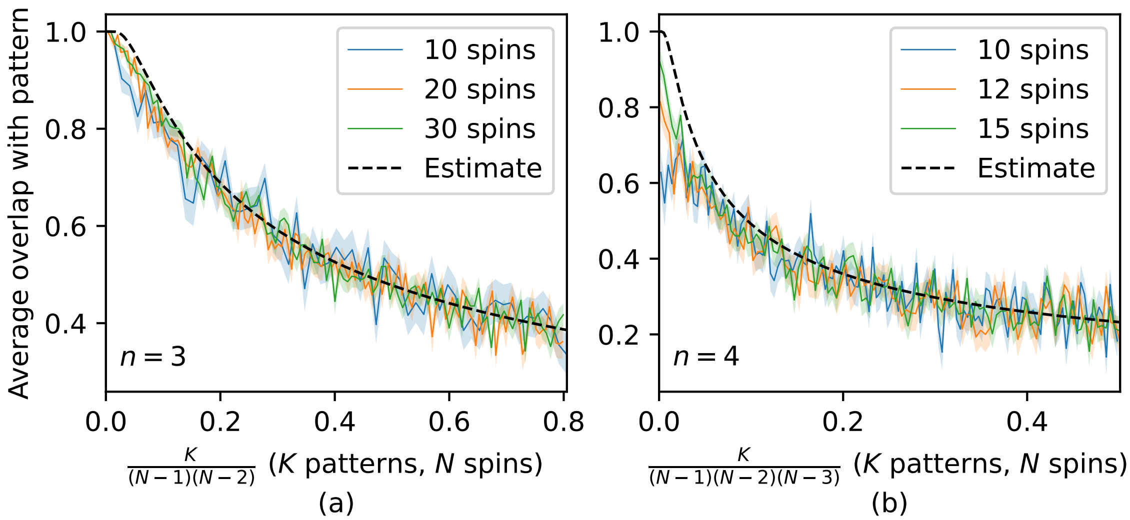

Figure 2.3: Pure Hebbian couplings. Average overlap of a pattern with itself after randomly "corrupting" $\delta = 0.1$ of the spins and doing one step of Hopfield update (in a modern Hopfield network with Hebbian n-body interactions) as a function of the scaled number of patterns $\alpha_n = K/(N-1)\cdots(N-n+1)$. The first plot (a) is for $n = 3$ for which the simulations were run for $N = 10$, 20, and 30 spins, and the second plot (b) is for $n = 4$, for which the simulations were run for system sizes of $N = 10$, 12, and 15. The results from the simulations have been shown, with the standard errors shown in lighter colors. The analytical estimates obtained using eq. (2.12) have also been plotted as black dashed lines. For $\delta \neq 0$, the expression in eq. (2.10) is a better approximation for large system size N and number of patterns K but deviates somewhat from the simulation results for smaller K, as can be seen in both plots and is more apparent in plot (b).

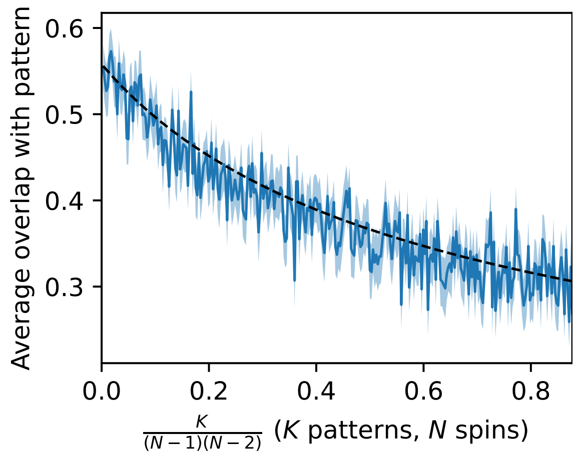

Figure 2.4: Noisy Hebbian couplings. Average overlap of a pattern with itself as a function of the number of stored patterns after randomly "corrupting" $\delta = 0.1$ of the spins doing one step of Hopfield update (in a modern Hopfield network with $N = 20$ spins interacting via $n = 3$-body interactions). The estimated standard error of the average is shown in light blue. The interaction coefficients are defined via the Hebb rule, but each coefficient also has an independent additive noise factor following the Gaussian distribution $N(\mu = 0, \sigma^2 = 100)$, and an independent multiplicative factor following the Bernoulli distribution $B(p = 0.9)$. The latter corresponds to 10% of the interaction terms being randomly deleted. The analytical estimate obtained in eq. (2.11) has also been plotted as a black dashed line.

distribution with parameter $p = 0.9$ (interpreted as retaining 90% the interaction terms and deleting the rest). The plots agree with the estimate given by eq. 2.11, also plotted in black. Because of the high noise, the expected overlap is low (around 0.6) even for very few stored patterns and falls off with the number of memories with no plateau. Figure 2.5 shows a color map of the average overlap as a function of both the scaled number of patterns $K/(N-1)(N-2)$ and the scaled mean square additive noise $\mu_{\eta^2}/(N-1)(N-2)$. We see that $E(m)$ is highest near $K = 1, \mu_{\eta^2} = 0$ and then decreases along the diagonal as $K + \mu_{\eta^2}$ increases, as expected from eq. (2.11).

Networks with clipped interactions (both in presence and absence of noise) were simulated by defining a new array of the clipped couplings $\check{J}_{i_1 \ldots i_n} = \operatorname{sgn}(J_{i_1 \ldots i_n})$. Figure 2.6 shows the plots of the expected overlap vs. α_3 in modern Hopfield networks of system size $N = 20$ with 3-spin interactions, where the initial perturbation is $\delta = 0.1$. Each interaction coefficient was perturbed with an additive Gaussian noise with mean 0 and variance 100 and then clipped to

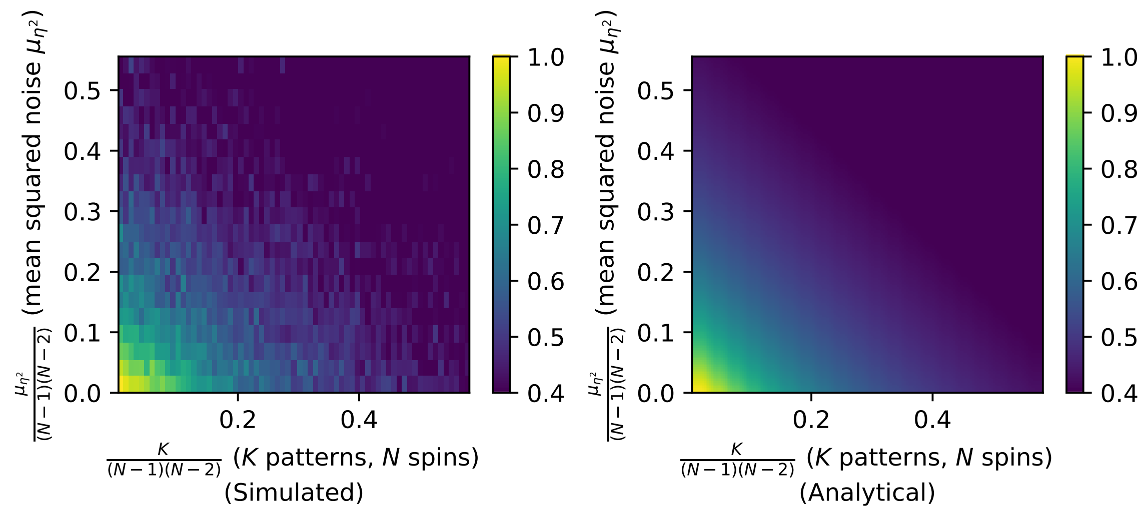

Figure 2.5: Color map of the average overlap of a pattern with itself as a function of the scaled number of stored patterns $K/(N-1)(N-2)$ and the scaled mean squared additive noise $\mu_{\eta^2}/(N-1)(N-2)$, after randomly "corrupting" $\delta = 0.1$ of the spins and doing one step of Hopfield update (in a modern Hopfield network with $N = 20$ spins interacting via $n = 3$-body interactions). The interaction coefficients are defined via the Hebb rule, but each coefficient also has an independent additive noise factor following the Gaussian distribution with zero mean and variance given by μ_{η^2}, and an independent multiplicative factor following the Bernoulli distribution Bernoulli($p = 0.9$). The latter corresponds to 10% of the interaction terms being randomly deleted. The analytical estimate obtained in eq. (2.11) has been plotted as a color map on the right for comparison.

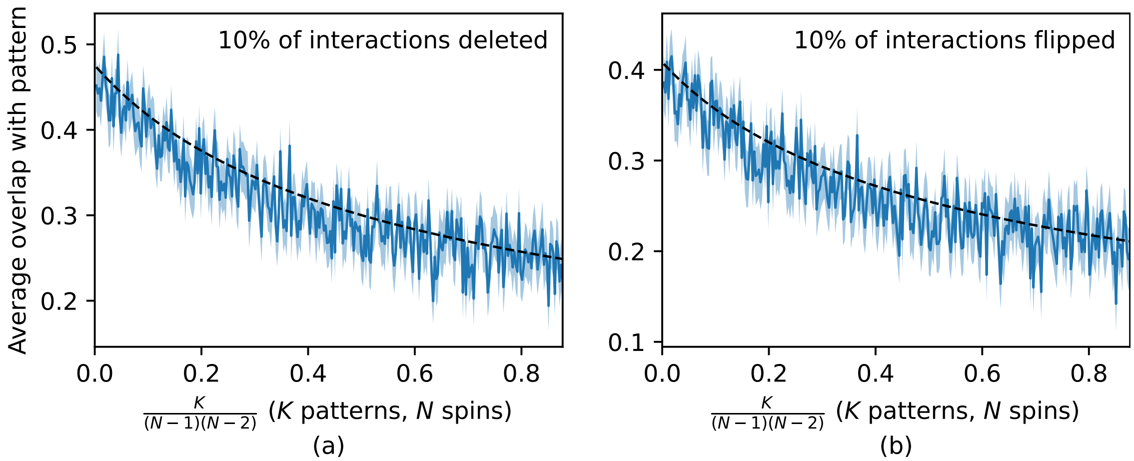

Figure 2.6: Clipped noisy Hebbian couplings. Average overlap of a pattern with itself as a function of the number of stored patterns after randomly "corrupting" $\delta = 0.1$ of the spins doing one step of Hopfield update (in a modern Hopfield network with $N = 20$ spins interacting via $n = 3$-body interactions). The estimated standard errors are shown in light blue. The interaction coefficients are initially defined via the Hebb rule, but each coefficient is also perturbed by an independent additive noise factor following the Gaussian distribution $N(\mu = 0, \sigma^2 = 100)$, and then clipped to have magnitude 1. The two plots correspond to systems where the interaction coefficients are (a) randomly deleted with probability 0.1, or (b) randomly flipped in sign with probability 0.1. The analytical estimate obtained in eq. (2.29) has also been plotted as a black dashed line in each case.

have magnitude 1. The first plot corresponds to the case where 10% of the interactions were randomly deleted, and the second is when 10% of the interactions were randomly flipped in sign. The plots are found to agree with the estimates given by eq. (2.29).

2.6 Discussions, applications and outlook

A model of associative memory capable of retrieving K patterns $\boldsymbol{\xi}^\mu$, each containing N binary neurons ξ_i^μ, stores NK bits of information. Equation (2.1) expresses the Hamiltonian of the modern Hopfield network in terms of the NK pattern components ξ_i^μ, in which case the pattern components themselves are the parameters of the model (and we need to introduce new parameters if more patterns need to be stored). While other ways to formulate and parametrize such models have been recently proposed (such as using random features [62]), a possible way to implement this dynamics in a physical or biological system is via interactions between the constituent neurons, similar to the interactions between spins in the Ising model. Such interactions $J_{i_1 \ldots i_n}$ involve two spins each ($n = 2$) for the standard Hopfield network and become many-body ($n > 2$) for the dense associative memory models considered here; typically they are constructed from the stored patterns using the Hebb rule. If we exclude self-interactions, the model is parametrized by $\binom{N}{n}$ interaction coefficients, each of which can take $K + 1$ possible values. It is of interest to investigate how resilient is the capacity of the network is to noise in these interaction coefficients, as well as to "coarsening" of the resolution in their values, as this determines the ability of such networks to successfully retrieve patterns in systems such as the brain.

We found that the effect of additive noise on the capacity is insignificant as long as the noise $\eta_{i_1 \ldots i_n}$ scales with the system size N at a power of less than $(n-1)/2$; specifically, if the mean squared noise satisfies $\mathrm{E}\left(\eta^2_{i_1 \ldots i_n}\right) \ll N^{n-1}$, then the capacity is unaffected. This constitutes the level of robustness of the network to extrinsic noise. Additive noise scaling as $\mathrm{E}\left(\eta^2_{i_1 \ldots i_n}\right) \sim N^{n-1}$ reduces the capacity by a constant factor without changing its dependence on system size, but for larger noise (such as when $\mathrm{E}\left(\eta^2_{i_1 \ldots i_n}\right) \sim N^b$ with $b > n - 1$) the system fails to retrieve in the limit of large system size irrespective of the number of patterns.

Another possibility to consider is when the additive noise scales with the number of patterns being stored. This may happen, for example, if, when training the interaction coefficients via he Hebb rule, an independent noise term gets introduced for every pattern being stored. In this case $\eta_{i_1 \ldots i_n} = \sum_\mu \eta^\mu_{i_1 \ldots i_n}$, where each of these terms can be thought to be i.i.d. with mean squared value $\tilde{\mu}_{\eta^2}$, which Agliari [16] calls noisy storing. In this case $\mu_{\eta^2} = K\tilde{\mu}_{\eta^2}$, and eq. (2.11) has the term $K + \mu_{\eta^2} = K(1 + \tilde{\mu}_{\eta^2})$. For $\tilde{\mu}_{\eta^2} << 1$ the effect of the additive noise is once again negligible, and for finite $\tilde{\mu}_{\eta^2}$ that does not depend on the system size the capacity is reduced by a factor of $1 + \tilde{\mu}_{\eta^2}$ but otherwise scales with system size as N^{n-1} as usual. If the additive noise per pattern diverges with system size as

$\tilde{\mu}_{\eta^2} \sim N^b$, then the scaling behavior of the capacity changes to $K_{\max} \sim N^{n-1-b}$, and thus the system can retrieve patterns in the thermodynamic limit only if $b \leq n-1$.

Multiplicative noise, on the other hand, is found to reduce the capacity without changing its scaling behavior with system size, irrespective of the magnitude of the noise (as long as its expectation is positive). In particular, if only a random subset of interactions is retained and the rest are set to zero (so each interaction has a probability p of being retained), then the capacity is reduced by a factor of p. Such a diluted network can thus still function as a dense associative memory, even if with a smaller capacity. A network where each interaction is retained with probability q and otherwise flipped in sign can still retrieve some patterns as long as $q > 1/2$, and in this case the capacity is reduced by a factor of $(2q-1)^2$.

We also considered networks where the interaction strengths had been clipped to all have the same magnitude and only the signs differ. We find that even with such drastic reduction in the precision of the interaction coefficients, the capacity is reduced by only a factor of $2/\pi$. This indicates that a physical system can be engineered to behave as a dense associative memory even if the interaction strengths are rounded to have lower "resolution" than the Hebbian values. If we simulate such a system digitally, each interaction coefficients can be stored in a single bit, more compactly than the Hebbian values which require $\log_2(K+1)$ bits. If we wish to retrieve a fixed number K of patterns using the smallest possible modern Hopfield network with n-spin interactions, and simulate the interactions digitally, then it turns out that for sufficiently large K, using clipped interactions would require storing fewer bits overall for the interaction coefficients, even taking into account the slightly larger number of spins required in the clipped network due to the reduced capacity. This is because the number of spins in the clipped model is larger by a factor of $(\pi/2)^{1/(n-1)}$, so the ratio of the required number of bits in the unclipped and clipped networks is $\log_2(K+1)/(\pi/2)^{n/(n-1)}$, which is greater than 1 for all interaction orders n as long as $K \geq 5$ (which always holds as we are assuming $N \gg 1$ and $K \sim N^{n-1}$). We did not consider cases where the interaction strengths are clipped less strongly and are allowed to have more than two different values, but we expect such networks to have capacities even closer to the unclipped Hebbian network.

We thus conclude that modern Hopfield networks defined via n-spin interactions can retrieve patterns in many cases even when the interactions have additive or multiplicative noise or are clipped to have lower precision compared to the Hebbian values. The capacity of the networks grows as the $(n-1)$-th power the system size in all cases, though they may be reduced by a factor depending on the noise or the level of clipping. The only cases where we found that the network may fail to retrieve are in the presence of additive noise growing with the system size faster than its $(n-1)$-th power, and in the presence of multiplicative noise whose expectation is nonpositive.

While we considered identically distributed additive and multiplicative noise terms affecting every interaction coefficient independently, further work can be done to explore the effects of other kinds of noise, perhaps physically motivated by the process by which the network learns and stores the interactions. For example, Agliari et al. have considered the effects of other forms of additive noise, which they associate with noisy patterns, learning, or storing [16]. We focused on the retrieval accuracy after one step of the synchronous Hopfield update, but it may be interesting to study the behavior of the network after multiple steps of stochastic dynamics given by a nonzero pseudotemperature, which may be analyzed using tools from equilibrium statistical physics such as replica theory, such as the analysis for the networks with pure Hebbian interactions carried out in [63], for very noisy networks with linear storage capacity in [64, 65] and for general classical and quantum neural networks in [66]. The effect of nonequilibrium dynamics on the capacity, as discussed for classic Hopfield networks in [67, 68], is also worth studying. It may also be of interest to consider networks with correlated patterns such as those considered in [69, 70], as well as systems with continuous components, such as the models discussed in [49, 71, 72], instead of the Ising systems studied here.

Acknowledgments

We thank Ilya Nemenman for useful discussions. This work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division.

2.7 Details for the derivation of the expected overlap

Here we describe in more detail the derivation of the expected overlap $\mathrm{E}(m)$ for the modern Hopfield network. We consider a network with N spins interacting via n-body interactions, and storing K patterns $\boldsymbol{\xi}^\mu$ whose components ξ_i^μ are i.i.d. Rademacher variables. Let us assume that we start with the μ-th pattern $\boldsymbol{\xi}^\mu$ and perturb it by flipping each of its components ξ_i^μ with probability δ. The resulting state $\tilde{\boldsymbol{\xi}}^\mu$ has components $\tilde{\xi}_i^\mu = \xi_i^\mu s_i$, where s_i are i.i.d. discrete random variables taking the values 1 and -1 with probabilities $1-\delta$ and δ respectively. (Therefore we have $\mathrm{E}(s_i) = 1 - 2\delta$, and $s_i^2 = 1$ with probability 1.) We perform one step of the Hopfield update defined in eq. (2.5) to obtain the state $\mathbf{T}_n \tilde{\boldsymbol{\xi}}^\mu$ with components

$$
\begin{aligned}
(\mathbf{T}_n \tilde{\boldsymbol{\xi}}^\mu)_i &= \operatorname{sgn}\left(\sum_{\substack{1 \le i_2 \le \cdots i_n \le N \\ i_a \neq i \forall a}} J_{i i_2 \ldots i_n} \tilde{\xi}_{i_2}^\mu \ldots \tilde{\xi}_{i_n}^\mu \right) \\
&= \operatorname{sgn}\left(\sum_{\substack{1 \le i_2 \le \cdots i_n \le N \\ i_a \neq i \forall a}} J_{i i_2 \ldots i_n} \xi_{i_2}^\mu \ldots \xi_{i_n}^\mu s_{i_2} \cdots s_{i_n} \right),
\end{aligned} \tag{2.33}
$$

where $J_{j_1\cdots j_n}$ are the n-body interaction coefficients. We assume that they follow the Hebb rule but may have additive or multiplicative noise, and thus are defined as

$$J_{j_1\ldots j_n} = P_{j_1\ldots j_n}\left(\sum_{\nu}\xi^{\nu}_{j_1}\cdots\xi^{\nu}_{j_n} + \eta_{j_1\ldots j_n}\right), \tag{2.34}$$

where the additive noise terms $\eta_{j_1\ldots j_n}$ are independent and identically distributed, and so are the multiplicative noise factors $P_{j_1\ldots j_n}$. We can then consider the variable

$$\begin{aligned} X_i &= \sum_{\substack{1\le i_2\le\cdots i_n\le N\\ i_a\neq i\forall a}} P_{ii_2\ldots i_n}\left(\sum_{\nu}\xi^{\nu}_{i}\xi^{\nu}_{i_2}\cdots\xi^{\nu}_{i_n} + \eta_{ii_2\ldots i_n}\right)\xi^{\mu}_{i}\xi^{\mu}_{i_2}\ldots\xi^{\mu}_{i_n}s_{i_2}\cdots s_{i_n} \\ &= \sum_{\nu,\{i_a\}} Y^{\nu}_{i\{i_a\}} + \sum_{\{i_a\}} Z_{i\{i_a\}}, \end{aligned} \tag{2.35}$$

where

$$Y^{\nu}_{i\{i_a\}} = \xi^{\mu}_{i}\xi^{\mu}_{i_2}\ldots\xi^{\mu}_{i_n}\xi^{\nu}_{i}\xi^{\nu}_{i_2}\cdots\xi^{\nu}_{i_n}\xi^{\mu}_{i}\xi^{\mu}_{i_2}\ldots\xi^{\mu}_{i_n}P_{ii_2\ldots i_n}s_{i_2}\cdots s_{i_n}, \tag{2.36}$$

$$Z_{i\{i_a\}} = \xi^{\mu}_{i}\xi^{\mu}_{i_2}\ldots\xi^{\mu}_{i_n}P_{ii_2\ldots i_n}\eta_{ii_2\ldots i_n}s_{i_2}\cdots s_{i_n}, \tag{2.37}$$

and $\{i_a\}$ is shorthand for $i_2,\ldots,i_n$. The sign of X_i will be $\mathrm{sgn}(X_i) = \xi^{\mu}_{i}(\mathbf{T}_n\boldsymbol{\xi}^{\mu})_i$, which determines whether the i-th spin is correctly retrieved, and the overlap of the state with the original pattern is $\langle\boldsymbol{\xi}^{\mu},\mathbf{T}_n\tilde{\boldsymbol{\xi}}^{\mu}\rangle = \sum_i \mathrm{sgn}(X_i)/N$. Therefore, the statistical properties of X_i can be used to estimate the retrieval accuracy.

Because ξ^{μ}_{i} are i.i.d. Rademacher variables, we have $\mathrm{E}\big(\xi^{\mu}_{i}\big) = 0$, and $\xi^{\mu 2}_{i} = 1$ with probability 1. Therefore, any term involving a product of ξ^{μ}_{i} factors will have nonzero expectation only if they "pair up". We then find that

$$\mathrm{E}\Big(Y^{\nu}_{i\{i_a\}}\Big) = \delta_{\mu\nu}\mu_P(1-2\delta)^{n-1}, \tag{2.38}$$

$$\mathrm{E}\Big(Z_{i\{i_a\}}\Big) = 0. \tag{2.39}$$

Since there are $\binom{N-1}{n-1}$ choices for the components $\{i_2, \ldots, i_n\} \subset \{1, \ldots, N\} \backslash \{i\}$, we conclude that $\mathrm{E}(X_i) = \binom{N-1}{n-1} \mu_P (1 - 2\delta)^{n-1}$.

To find the variance of X_i, we need to find the variances as well as pairwise covariances of the terms $Y^{\nu}_{i\{i_a\}}$ and $Z_{i\{i_a\}}$. We note that

$$\mathrm{Cov}\Big(Z_{i\{i_a\}}, Z_{i\{j_b\}}\Big) = \begin{cases} \mathrm{V}\Big(Z_{i\{i_a\}}\Big) = \mu_{P^2}\mu_{\eta^2}, & \text{if } \{i_a\} = \{j_b\}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.40}$$

$$\mathrm{Cov}\Big(Z_{i\{i_a\}}, Y^{\nu}_{i\{j_b\}}\Big) = 0. \tag{2.41}$$

For finding the covariance $\mathrm{Cov}\Big(Y^{\nu}_{i\{i_a\}}, Y^{\lambda}_{i\{j_b\}}\Big)$, let us first consider the case where $\{i_a\} = \{j_b\}$. In this case, the covariance $\mathrm{Cov}\Big(Y^{\nu}_{i\{i_a\}}, Y^{\lambda}_{i\{i_a\}}\Big)$ is μ_{P^2} if $\nu = \lambda \neq \mu$, $\mu_{P^2} - \mu_P^2(1 - 2\delta)^{2(n-1)}$ if $\nu = \lambda = \mu$, and 0 if $\nu \neq \lambda$. However, if $\{i_a\} \neq \{j_b\}$, the covariance can be nonzero only if $\nu = \lambda = \mu$. In this case, we have $\mathrm{Cov}\Big(Y^{\mu}_{i\{i_a\}}, Y^{\mu}_{i\{j_b\}}\Big) = \mu_P^2((1 - 2\delta)^{2(n-1-c)} - (1 - 2\delta)^{2(n-1)})$, where $0 \leq c = |\{i_a\} \cap \{j_b\}| \leq n - 2$ is the number of spins in common between the sets $\{i_a\}$ and $\{j_b\}$. (We note that this covariance is zero when $c = 0$, as well as for all c in the case of no initial perturbation $\delta = 0$.) The number of possible ways of sampling the sets $\{i_a\}, \{j_b\} \subset \{1, \ldots, N\} \setminus \{i\}$ with $|\{i_a\}| = |\{j_b\}| = n - 1$ and $|\{i_a\} \cap \{j_b\}| = c$ is $\binom{N-1}{n-1}\binom{n-1}{c}\binom{N-n}{n-1-c}$. Combining all the covariances, we find that

$$\begin{aligned} \mathrm{V}(X_i) = &\binom{N-1}{n-1}\Big[\mu_{P^2}(K + \mu_{\eta^2}) - \mu_P^2(1 - 2\delta)^{2(n-1)}\Big] \\ &+ \binom{N-1}{n-1}\mu_P^2(1 - 2\delta)^{2(n-1)}\sum_{c=1}^{n-2}\binom{n-1}{c}\binom{N-n}{n-1-c}\Big((1 - 2\delta)^{-2c} - 1\Big). \end{aligned} \tag{2.42}$$

As we shall show below, the capacity of this network is of the order of N^{n-1}; when the number of memories is of that order, the first term in the above expression of the variance scales with system size as $N^{2(n-1)}$. The term on the second line, however, is of order N^{2n-3} (corresponding to the term in the sum with $c = 1$) and thus can be ignored when $N \gg 1$

and $K \sim N^{n-1}$, which is the regime we will be interested in. (This term also vanishes in the case of no initial perturbation, $\delta = 0$.)

To make further progress, we assume that the probability distribution of X_i is approximately Gaussian. Although we do not prove if this approximation is valid, we provide some intuition here for why this may be a reasonable assumption (which is also found to agree with the results from numerical simulations). We define

$$X_i^{\setminus\mu} = \sum_{\{i_a\}} \left(\sum_{\nu\neq\mu} Y^{\nu}_{i\{i_a\}} + Z_{i\{i_a\}} \right) \equiv \sum_{\{i_a\}} X^{\setminus\mu}_{i\{i_a\}}. \tag{2.43}$$

The variables $X^{\setminus\mu}_{i\{i_a\}}$ for different $\{i_a\}$ are identically distributed, and they are pairwise independent (though not mutually independent for $n > 2$). Therefore, even though the conditions for central limit theorem are not fully met, we may expect $X_i^{\setminus\mu} = \sum_{\{i_a\}} X^{\setminus\mu}_{i\{i_a\}}$ to approximately follow a Gaussian distribution with mean 0 and variance $\binom{N-1}{n-1}\mu_{P^2}(K - 1 + \mu_{\eta^2})$. On the other hand, $Y^{\mu}_{i\{i_a\}} = P_{i\{i_a\}} \prod_a s_{i_a}$ are correlated for different sets $\{i_a\}$ if they are not disjoint (which is possible if $n > 2$) and $0 < \delta < 1/2$, and thus their sum $X_i^{\mu} = \sum_{\{i_a\}} Y^{\mu}_{i\{i_a\}}$ may not be normal. X_i^{μ} and $X_i^{\setminus\mu}$ are also not independent, even though they are uncorrelated. However, because $\mathrm{V}\big(X_i^{\setminus\mu}\big) \gg \mathrm{V}\big(X_i^{\mu}\big)$ for $N \gg 1$ and $K \sim N^{n-1}$, we expect $X_i = X_i^{\setminus\mu} + X_i^{\mu}$ to be approximately Gaussian for large enough system size. For any Gaussian random variable $G \sim N(\mu, \sigma^2)$, the probability of it being positive is $\mathrm{P}(G > 0) = (1 + \mathrm{erf}\big(\mu/\sqrt{2}\sigma\big))/2$. Therefore, the expected value of $\mathrm{sgn}(X_i)$ is estimated to be

$$\begin{aligned} \mathrm{E}(\mathrm{sgn}(X_i)) &= \mathrm{erf}\left(\frac{\mathrm{E}(X_i)}{\sqrt{2\,\mathrm{V}(X_i)}} \right) \\ &\approx \mathrm{erf}\left(\frac{\binom{N-1}{n-1}\mu_P(1-2\delta)^{n-1}}{\sqrt{2\Big[\binom{N-1}{n-1}\big(\mu_{P^2}(K+\mu_{\eta^2}) - \mu_P^2(1-2\delta)^{2(n-1)}\big)\Big]}} \right) \end{aligned} \tag{2.44}$$

While the cases where the interactions have additive and multiplicative noise can be treated together, we need to separately analyze the case where the interactions are also clipped. In this case we the interaction coefficients are defined as

$$\check{J}_{j_1 \dots j_n} = \operatorname{sgn}\Bigg(P_{j_1 \dots j_n} \Bigg(\sum_\nu \xi^\nu_{j_1} \cdots \xi^\nu_{j_n} + \eta_{j_1 \dots j_n} \Bigg) \Bigg), \tag{2.45}$$

i.e., similar to the previous case but with an additional sign function for the clipping. The analog of the variable X_i, determining whether the i-th spin is retrieved correctly, is now

$$\begin{aligned} \check{X}_i &= \sum_{\{i_a\}} \check{J}_{i\{i_a\}} \xi^\mu_i \Bigg(\prod_a \xi^\mu_{i_a} s_{i_a} \Bigg) \\ &= \sum_{\{i_a\}} \operatorname{sgn}\Bigg(\sum_\nu \xi^\mu_i \xi^\nu_i \prod_a \Big(\xi^\mu_{i_a} \xi^\nu_{i_a} \Big) + \eta_{i\{i_a\}} \xi^\mu_i \prod_a \Big(\xi^\mu_{i_a} \Big) \Bigg) \operatorname{sgn}\Big(P_{i\{i_a\}} \Big) \prod_a s_{i_a} \\ &\equiv \sum_{\{i_a\}} \operatorname{sgn}\Big(W_{i\{i_a\}} \Big) \operatorname{sgn}\Big(P_{i\{i_a\}} \Big) \prod_a s_{i_a} \equiv \sum_{\{i_a\}} \operatorname{sgn}\Big(X_{i\{i_a\}} \Big), \end{aligned} \tag{2.46}$$

where we now need to find the probability of each of the terms

$$\begin{aligned} W_{i\{i_a\}} &= \sum_\nu \xi^\mu_i \xi^\nu_i \prod_a \Big(\xi^\mu_{i_a} \xi^\nu_{i_a} \Big) + \eta_{i\{i_a\}} \xi^\mu_i \prod_a \Big(\xi^\mu_{i_a} \Big) \\ &= 1 + \sum_{\nu \neq \mu} \xi^\mu_i \xi^\nu_i \prod_a \Big(\xi^\mu_{i_a} \xi^\nu_{i_a} \Big) + \eta_{i\{i_a\}} \xi^\mu_i \prod_a \Big(\xi^\mu_{i_a} \Big) \end{aligned} \tag{2.47}$$

being positive. By similar arguments as above, we can find that

$$\mathrm{E}\Big(W_{i\{i_a\}} \Big) = 1, \tag{2.48}$$

since ξ_i^ν are independent with zero expectation. We also find that

$$\begin{aligned}
W_{i\{i_a\}}^2 &= \sum_\nu \left(\xi_i^\mu \xi_i^\nu \prod_a \left(\xi_{i_a}^\mu \xi_{i_a}^\nu\right)\right)^2 + \left(\eta_{i\{i_a\}} \xi_i^\mu \prod_a \left(\xi_{i_a}^\mu\right)\right)^2 \\
&\quad + 2\sum_{\nu<\lambda} \xi_i^\mu \xi_i^\nu \xi_i^\mu \xi_i^\lambda \prod_a \left(\xi_{i_a}^\mu \xi_{i_a}^\nu \xi_{i_a}^\mu \xi_{i_a}^\lambda\right) + 2\xi_i^\mu \xi_i^\nu \eta_{i\{i_a\}} \xi_i^\mu \prod_a \left(\xi_{i_a}^\mu \xi_{i_a}^\nu \xi_{i_a}^\mu\right) \\
&= K + \eta_{i\{i_a\}}^2 + 2\sum_{\nu<\lambda} \xi_i^\nu \xi_i^\lambda \prod_a \left(\xi_{i_a}^\nu \xi_{i_a}^\lambda\right) + 2\xi_i^\nu \eta_{i\{i_a\}} \prod_a \xi_{i_a}^\nu,
\end{aligned} \tag{2.49}$$

and thus

$$\mathrm{V}\left(W_{i\{i_a\}}\right) = K + \mu_{\eta^2} - 1. \tag{2.50}$$

Assuming $W_{i\{i_a\}}$ is approximately Gaussian, we find

$$\begin{aligned}
\mathrm{E}\left(\mathrm{sgn}\left(X_{i\{i_a\}}\right)\right) &= \mathrm{E}\left(W_{i\{i_a\}}\right) \mathrm{E}\left(\check{P}_{i\{i_a\}}\right) \prod_a \mathrm{E}(s_{i_a}) \\
&= \mathrm{erf}\left(\frac{1}{\sqrt{2(K + \mu_{\eta^2} - 1)}}\right) \mu_{\check{P}} (1 - 2\delta)^{n-1} \\
&\approx \frac{2\mu_{\check{P}} (1 - 2\delta)^{n-1}}{\sqrt{2\pi(K + \mu_{\eta^2})}} \quad (\text{for } K \gg 1),
\end{aligned} \tag{2.51}$$

where $\check{P}_{i\{i_a\}} = \mathrm{sgn}\left(P_{i\{i_a\}}\right)$. We know that $\mathrm{P}\left(s_j^2 = 1\right) = 1$, and if we ignore the probability of $W_{i\{i_a\}}$ being exactly zero, we obtain $\mathrm{V}\left(\mathrm{sgn}\left(X_{i\{i_a\}}\right)\right) = \mu_{\check{P}^2} - E(\mathrm{sgn}\left(X_{i\{i_a\}}\right))^2$. Then $\check{X}_i$ satisfies

$$\begin{aligned}
\mathrm{E}\left(\check{X}_i\right) &= \sum_{\{i_a\}} \mathrm{E}\left(\mathrm{sgn}\left(X_{i\{i_a\}}\right)\right) = \binom{N-1}{n-1} \mathrm{erf}\left(\frac{1}{\sqrt{2(K + \mu_{\eta^2} - 1)}}\right) \mu_{\check{P}} (1 - 2\delta)^{n-1} \\
&\approx \binom{N-1}{n-1} \frac{2\mu_{\check{P}} (1 - 2\delta)^{n-1}}{\sqrt{2\pi(K + \mu_{\eta^2})}},
\end{aligned} \tag{2.52}$$

and

$$\begin{aligned}
\mathrm{V}\big(\check{X}_i\big) &= \sum_{\{i_a\}} \mathrm{V}\Big(\mathrm{sgn}\Big(X_{i\{i_a\}}\Big)\Big) \\
&= \binom{N-1}{n-1}\left(\mu_{\check{P}^2} - \mathrm{erf}\left(\frac{1}{\sqrt{2(K+\mu_{\eta^2}-1)}}\right)^2 \mu_{\check{P}}^2 (1-2\delta)^{2(n-1)}\right) \\
&\approx \binom{N-1}{n-1}\left[\mu_{\check{P}^2} - \left(\frac{2\mu_{\check{P}}(1-2\delta)^{n-1}}{\sqrt{2\pi(K+\mu_{\eta^2})}}\right)^2\right],
\end{aligned} \tag{2.53}$$

where we have ignored the covariances $\mathrm{Cov}\Big(\mathrm{sgn}\Big(X_{i\{i_a\}}\Big), \mathrm{sgn}\Big(X_{i\{j_b\}}\Big)\Big)$ as they are much smaller than their individual variances when K is large. Therefore, the expected overlap is

$$\begin{aligned}
\mathrm{E}(\check{m}) = \mathrm{E}\big(\mathrm{sgn}\big(\check{X}_i\big)\big) &\approx \mathrm{erf}\left(\frac{\binom{N-1}{n-1}\frac{2\mu_{\check{P}}(1-2\delta)^{n-1}}{\sqrt{2\pi(K+\mu_{\eta^2})}}}{\sqrt{2\binom{N-1}{n-1}\left[\mu_{\check{P}^2} - \left(\frac{2\mu_{\check{P}}(1-2\delta)^{n-1}}{\sqrt{2\pi(K+\mu_{\eta^2})}}\right)^2\right]}}\right) \\
&= \mathrm{erf}\left((1-2\delta)^{n-1}\sqrt{\frac{\binom{N-1}{n-1}\mu_{\check{P}}^2}{\pi\mu_{\check{P}^2}(K+\mu_{\eta^2})}}\right).
\end{aligned} \tag{2.54}$$

2.8 Numerical results for retrieval accuracy after multiple update steps

In this work, we defined the retrieval accuracy and capacity in 2.3 using the overlap

$$m = \frac{1}{N}\sum_i \xi_i^\mu (\mathbf{T}_n \tilde{\boldsymbol{\xi}}^1)_i \tag{2.55}$$

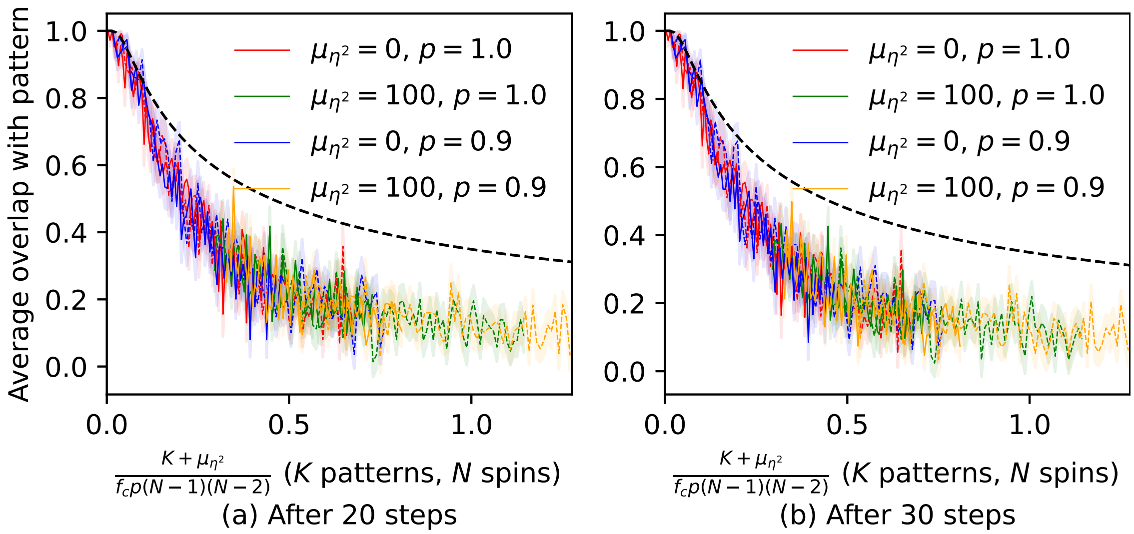

Figure 2.7: Average overlap of a pattern with itself after randomly "corrupting" $\delta = 0.1$ of the spins in a modern Hopfield network and doing (a) 20 and (b) 30 steps of Hopfield update, plotted in terms of the quantity $\gamma = (K + \mu_{\eta^2})/(f_C p(N-1)(N-2))$. We considered 8 types of networks in each plot, each with $N = 20$ spins and interaction order $n = 3$, but differing in the types of additive noise (none vs. Gaussian with mean squared noise $\mu_{\eta^2} = 100$), multiplicative noise (none vs. Bernoulli with parameter $p = 0.9$), and clipping (not clipped vs. clipped). The networks corresponding to different noise distributions have been plotted in different colors (with the standard errors shown in lighter colors), and the corresponding clipped networks have been plotted in the same colors as the unclipped networks but using dashed lines. The analytical estimate corresponding to one update step, obtained using eq. (2.57), has also been plotted as a black dashed line.

, which measures how close the perturbed pattern $\tilde{\boldsymbol{\xi}}^1$ is to the original pattern $\boldsymbol{\xi}^1$ after one step of the Hopfield update $\mathbf{T}_n$ (as defined in eq. (2.5)). However, it may be of interest to consider the retrieval accuracy of the network after multiple consecutive steps of $\mathbf{T}_n$, as after several steps we may expect the system to reach a fixed point (though this is not guaranteed for synchronous dynamics). We do not analyze the behavior of the network after multiple steps in this paper, but in this section we present some numerical results from simulated networks.

The numerical results were obtained from simulated networks using the methods similar to those described in section 2.5. We simulated 8 different types of modern Hopfield networks. Each had $N = 20$ spins and interaction order $n = 3$, but they differed in additive noise

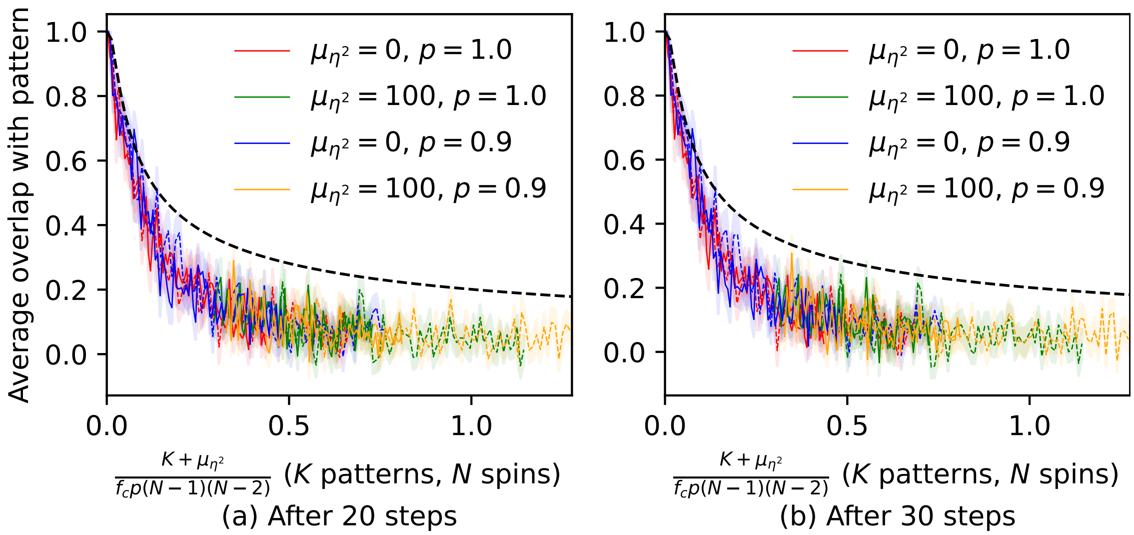

Figure 2.8: Average overlap of a pattern with itself after randomly "corrupting" $\delta = 0.2$ of the spins in a modern Hopfield network and doing (a) 20 and (b) 30 steps of Hopfield update, plotted in terms of the quantity $\gamma = (K + \mu_{\eta^2})/(f_c p(N-1)(N-2))$. We considered 8 types of networks in each plot, each with $N = 20$ spins and interaction order $n = 3$, but differing in the types of additive noise (none vs. Gaussian with mean 0 and variance $\mu_{\eta^2} = 100$), multiplicative noise (none vs. Bernoulli with parameter $p = 0.9$), and clipping (not clipped vs. clipped). The networks corresponding to different noise distributions have been plotted in different colors (with the standard errors shown in lighter colors), and the corresponding clipped networks have been plotted in the same colors as the unclipped networks but using dashed lines. The analytical estimate corresponding to one update step, obtained using eq. (2.57), has also been plotted as a black dashed line.

(none vs. Gaussian with mean squared noise 100), multiplicative noise (none vs. Bernoulli with parameter 0.9), and clipping (not clipped vs clipped). For each type of network and for each considered number of patterns K, we generated 10 different networks (corresponding to different sets of randomly generated patterns $\{\boldsymbol{\xi}^\mu\}$ and different realizations of noise), and for each network we sampled 5 data points by randomly choosing and perturbing one of the patterns (flipping each spin with probability δ) each time. We operated the Hopfield update on the perturbed pattern k times, and found the average overlap $m = \sum_i \xi_i^\mu \mathbf{T}_n^k \tilde{\boldsymbol{\xi}^\mu}_i / N$ of the updated pattern with the original pattern. The results are shown in figures 2.7 and 2.8.

From eqs. (2.11) and (2.29), we expect the expected overlap $\mathrm{E}(m)$ to depend on the number of patterns and the noise and clipping via the quantity

$$\gamma = \frac{K + \mu_{\eta^2}}{f_c(\mu_{P^2}/\mu_P^2)(N-1)(N-2)}, \tag{2.56}$$

where $\mu_{P^2}/\mu_P^2 = p$ if the multiplicative noise follows Bernoulli distribution with parameter p (which is the case we have considered here), and f_c is $2/\pi$ if the interactions were clipped and 1 otherwise. In particular, after one step of Hopfield update, we expect

$$\mathrm{E}(m) = \mathrm{erf}\Big((1-2\delta)^{n-1}/\sqrt{2\gamma(n-1)!}\Big), \tag{2.57}$$

, where δ is the initial perturbation and n the interaction order. In figures 2.7 and 2.8, we plot the average overlaps as functions of γ.

We note that figures 2.7 (a) and (b) correspond to the same networks after 20 and 30 update steps respectively, and the plots are seen to be almost identical. (The same is the case for figs. 2.8 (a) and (b).) We can therefore assume that the results after 30 steps exhibit the fixed point of the update dynamics. The simulated overlaps do not agree with the analytical estimate of the overlap after 1 step (also shown in the plot), but they are seen to agree with one another. Thus, even though we have not found the exact dependence of $\mathrm{E}(m)$ after multiple steps on the number of patterns or the noise, we expect it to be some function of γ. Therefore, if we define the capacity of the network in terms of its behavior after multiple steps of synchronous update at zero temperature, this capacity will be affected by noise or clipping the same way as the capacity we have defined in terms of one-step update.

Figure 2.7 corresponds to simulations where the starting pattern was initially perturbed by randomly flipping each of its spins with probability $\delta = 0.1$ (which would correspond to an initial overlap with the pattern of $m_0 = 0.8$). We note that the numerical values of $\mathrm{E}(m)$ remain close to the expected one-step results over the range of γ where $\mathrm{E}(m) > 0.8$, but get smaller than the one-step results for larger γ. Therefore, it appears that if the first step of synchronous Hopfield update brings the system closer to the pattern (i.e., the overlap with the pattern increases), then the subsequent steps make little difference and do not affect the overlap much. However, if the first step brings the state further away (decreasing the overlap), then the subsequent steps decrease it even further. Figure 2.8 corresponds to simulations with identical networks where the starting pattern was perturbed more strongly (given by $\delta = 0.2$ and initial overlap $m_0 = 0.6$). In this case, we see that the overlap after multiple steps is close to the one-step estimate as long as it's above 0.6 and falls below the one-step estimate for larger γ, which is consistent with our observations from fig. 2.7.

2.9 Expressing dense associative memory Hamiltonians involving higher powers of overlaps in terms of multi-spin interactions and vice versa

Much of the work on dense associative memories, including that by Krotov [47], define the Hamiltonians of such models in terms of the overlaps of the state with the patterns being stored. Therefore, if the system is in the state $\boldsymbol{\sigma}$ (with components σ_i) and we wish to store the patterns $\boldsymbol{\xi}^\mu$, then the Hamiltonian of the system can be expressed as

$$H = -N \sum_{\mu} F(m_\mu), \tag{2.58}$$

where $m_\mu = \sum_i \xi_i^\mu \sigma_i / N$ are the overlaps. For the quadratic function $F(x) = x^2$, this corresponds to the standard Hopfield network, whereas $F(x)$ involving higher powers of x indicates a dense associative memory.

It is also useful to interpret associative memories as arising from of interactions among their components. This is how Hopfield networks have been usually interpreted, where the components spins have pairwise interactions mediated by coupling constants usually defined by the Hebb rule. Modern Hopfield networks can be defined by allowing higher order interactions; this is the approach taken previously, for example, by Agliari et al. [16], and also by us in this work. If we define a modern Hopfield network whose component spins interact via n-body interactions, with the interaction coefficients defined by the generalized Hebb rule with no noise, then its Hamiltonian will be of the form

$$H_n^0 = -\frac{1}{N^{n-1}} \sum_{1 \leq i_i < \cdots < i_n \leq N} \sum_{\mu} \xi_{i_1}^\mu \cdots \xi_{i_n}^\mu \sigma_{i_1} \cdots \sigma_{i_n}, \tag{2.59}$$

where we have excluded self-interactions. If we had allowed self-interactions, it is clear the above Hamiltonian would have been identical to the one in eq. (2.58),with $F(x) = x^n$. Because $\sigma_i^2 = 1$, any n-body interaction term involving self-interactions can be written as an $(n - 2k)$-body interaction term excluding self-interactions for some nonnegative integer $k \leq n/2$. Thus, any Hamiltonian of the form in eq. (2.58), where F is an analytic function, can be written as a sum of interaction Hamiltonians of the form expressed in eq. (2.59). In particular, by expanding the powers and using combinatorics, we see that

$$-N\sum_{\mu} m_{\mu} = H_1^0, \tag{2.60}$$

$$-N\sum_{\mu} m_{\mu}^2 = 2H_2^0 - K, \tag{2.61}$$

$$-N\sum_{\mu} m_{\mu}^3 = 6H_3^0 + \left(\frac{3}{N} - \frac{2}{N^2}\right)H_1^0, \tag{2.62}$$

$$-N\sum_{\mu} m_{\mu}^4 = 24H_4^0 + \left(\frac{12}{N} - \frac{16}{N^2}\right)H_2^0 - K\left(\frac{3}{N} - \frac{2}{N^2}\right), \tag{2.63}$$

and so on. For general power n the expressions get complicated, but for large N we can keep the leading order terms to get the approximate expression

$$-N\sum_{\mu} m_{\mu}^n \approx n! \sum_{k=0}^{\lceil n/2 \rceil} \frac{1}{(2N)^k k!} H_{n-2k}^0 \text{ for } N \gg 1, \tag{2.64}$$

where $H_0^0 \equiv -NK$. We can invert the above to express the Hamiltonian in eq. (2.59) in terms of polynomial functions of the overlaps to obtain

$$H_1^0 = -N\sum_{\mu} m_{\mu}, \tag{2.65}$$

$$H_2^0 = -\frac{1}{2}N\sum_{\mu} m_{\mu}^2 + \frac{K}{2}, \tag{2.66}$$

$$H_3^0 = -N\sum_{\mu}\left(\frac{1}{6}m_{\mu}^3 - \left(\frac{1}{2N} - \frac{1}{3N^2}\right)m_{\mu}\right), \tag{2.67}$$

$$H_4^0 = -N\sum_{\mu}\left(\frac{1}{24}m_{\mu}^4 - \left(\frac{1}{4N} - \frac{1}{3N^2}\right)m_{\mu}^2 + \frac{1}{8N^2} - \frac{1}{4N^3}\right), \text{ etc.,} \tag{2.68}$$

and for $N \gg 1$ we have

$$H_n^0 = -N\sum_{\mu=1}^{K}\sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \frac{m_{\mu}^{n-2k}}{k!(n-2k)!(2N)^k}. \tag{2.69}$$

2.10 Recap of results for the classic Hopfield network

Here we derive the retrieval accuracy of a Hopfield network for the special case of two-body interactions (i.e., the network originally described by Hopfield [1]). While the analysis is very similar to the general derivation shown in section 2.4, it may be helpful to study the simpler and better-known case of the standard Hopfield network and and check that our results agree with previous work. As before, we consider a system of N Ising spins (binary neurons), where each spin can be in the state 1 or -1, and K patterns $\boldsymbol{\xi}^\mu \in \{1, -1\}^N$, $\mu \in \{1, \ldots, K\}$, where every component ξ_i^μ independently takes the value 1 or -1 with equal probability. Then the Hamiltonian for the Hopfield network is defined as an infinite-range Ising Hamiltonian with two-spin interactions,

$$H(\boldsymbol{\sigma}) = -\sum_{i<j} J_{ij} \sigma_i \sigma_j, \tag{2.70}$$

where the coupling constants J_{ij} can be defined by the Hebb rule (as described schematically in figure 2.9),

$$J_{ij}^0 = \sum_\mu \xi_i^\mu \xi_j^\mu. \tag{2.71}$$

We will generalize the coupling constants to include additive and multiplicative noise, in which case they have the form

$$J_{ij} = P_{ij} \left(\sum_\mu \xi_i^\mu \xi_j^\mu + \eta_{ij} \right), \tag{2.72}$$

where P_{ij} is a random multiplicative factor and η_{ij} is an additive noise term. In particular, we will mainly consider the cases where P_{ij} are i.i.d. Bernoulli random variables with parameter p (corresponding to p of the couplings being retained and the rest randomly deleted), and

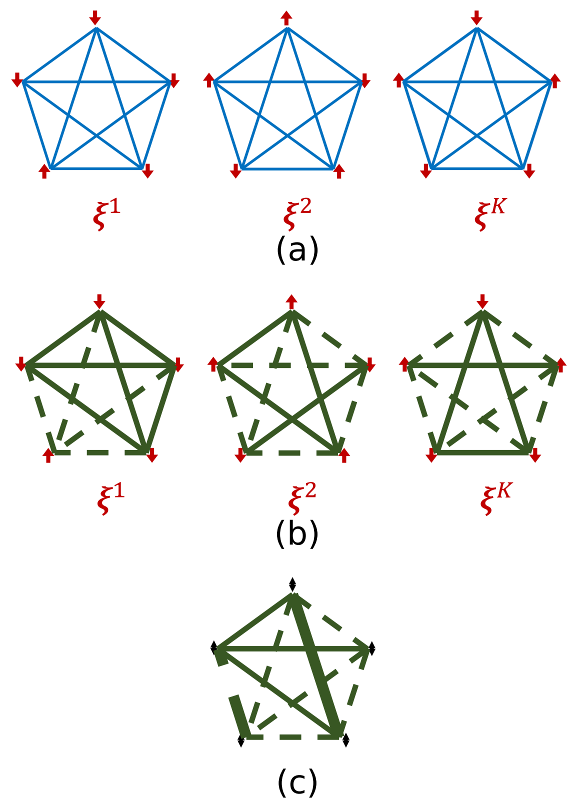

Figure 2.9: Schematic of a classic Hopfield network with $N = 5$ spins and $K = 3$ patterns. (a) Visual representation of the patterns, where upward and downward facing arrows correspond to a spin of 1 and -1 respectively. (b) Couplings associated with each pattern (defined as $J_{ij}^{\mu} = \xi_i^{\mu}\xi_j^{\mu}$ for pattern μ and spins i and j), shown as green lines connecting the arrows. Solid lines are ferromagnetic ($J_{ij}^{\mu} = 1$) and dotted lines are antiferromganetic ($J_{ij}^{\mu} = -1$). (c) Couplings defined using the Hebbian learning rule as $J_{ij} = \sum_{\mu} J_{ij}^{\mu}$, shown as before by lines connecting the arrows, where the thickness of the lines indicates the strength $\left|J_{ij}\right|$ of the coupling.

η_{ij} are i.i.d. Gaussian variables centered at 0.

We will also consider networks where all nonzero couplings $\check{J}_{ij}$ have been "clipped" to be ± 1, which we can define by taking just the sign of the coupling constant defined above (as shown in figure 2.11), i.e.,

$$\check{J}_{ij} = \operatorname{sgn}\left(P_{ij}\left(\sum_{\mu} \xi_i^{\mu}\xi_j^{\mu} + \eta_{ij}\right)\right). \tag{2.73}$$

Each spin σ_i can be thought of as experiencing a local field $B_i = \sum_{j\neq i} J_{ij}\sigma_j$, and at each time step the spins will orient themselves along their corresponding local fields. We consider synchronous updates at zero temperature, so the evolution of the system at each time step is given by the operator $\mathbf{T}$, defined such that

$$(\mathbf{T}\boldsymbol{\sigma})_i = \operatorname{sgn}\left(\sum_{j\neq i} J_{ij}\sigma_j\right) = \operatorname{sgn}\left(\sum_{\mu, j\neq i} \xi_i^{\mu}\xi_j^{\mu}\sigma_j\right). \tag{2.74}$$

We estimate the retrieval accuracy of the network in the manner described in section 2.3. We arbitrarily choose a pattern (say $\boldsymbol{\xi}^1$) and "corrupt" it by randomly flipping some of its component spins with probability δ. The corrupted state is $\tilde{\boldsymbol{\xi}}^1$ with components $\tilde{\xi}_i^1 = s_i\xi_i^1$, where s_i are independent and each takes the values 1 and -1 with probabilities $1-\delta$ and δ respectively. We then operate one step of the Hopfield update, and find the overlap $m = \sum_i \xi_i^1(\mathbf{T}\tilde{\boldsymbol{\xi}}^1)_i/N$ of the resulting state with the original pattern to quantify how accurately the original pattern was retrieved.

Let us first consider the Hopfield network with couplings defined by the Hebb rule but with noise as in eq. (2.72). Then the overlap m is

$$m = \frac{1}{N}\sum_i \xi_i^1\left(\mathbf{T}\tilde{\boldsymbol{\xi}}^1\right)_i = \frac{1}{N}\sum_i\left(\operatorname{sgn}\left(\sum_{\mu, j\neq i}\xi_i^{\mu}\xi_j^{\mu}\xi_i^1\xi_j^1 + \eta_{ij}\xi_i^1\xi_j^1\right)P_{ij}s_j\right). \tag{2.75}$$

We define the random variables $Y_{ij}^{\mu} = \xi_i^{\mu}\xi_j^{\mu}\xi_i^1\xi_j^1 P_{ij}s_j$, $Z_{ij} = \xi_i^1\xi_j^1\eta_{ij}P_{ij}s_j$, and $X_i = \sum_{j\neq i}(\sum_{\mu} Y_{ij}^{\mu} + Z_{ij})$. We also note that $\mathrm{E}\big(\xi_i^{\mu}\big) = 0$ for all μ and i, $\mathrm{E}(s_i) = 1 - 2\delta$ for all i, $\xi_i^{\mu 2} = 1$ almost everywhere, $\mathrm{E}\big(\eta_{ij}\big) = 0$, and the variables $\big\{\xi_i^{\mu}\big\}_{\mu,i}$, $\big\{\eta_{ij}\big\}_{i<j}$, $\big\{P_{ij}\big\}_{i<j}$, and $\{s_i\}_i$ are all independent. We define $\mu_{\eta^2} = \mathrm{E}\Big(\eta_{ij}^2\Big)$, $\mu_P = \mathrm{E}\big(P_{ij}\big)$, and $\mu_{P^2} = \mathrm{E}\Big(P_{ij}^2\Big)$. Then we can check that $\mathrm{E}\Big(Y_{ij}^{\mu}\Big) = \mu_P(1-2\delta)\delta_{\mu 1}$, $\mathrm{E}\big(Z_{ij}\big) = 0$, $\mathrm{V}\big(Z_{ij}\big) = \mu_{\eta^2}\mu_{P^2}$, $\mathrm{E}\Big(X_{ij}^{\mu\,2}\Big) = \mu_{P^2} - \mu_P^2(1-2\delta)^2\delta_{\mu 1}$, and all other covariances are zero. This implies $\mathrm{E}(X_i) = (N-1)\mu_P(1-2\delta)$ and $\mathrm{V}(X_i) = (N-1)\Big((K+\mu_{\eta^2})\mu_{P^2} - \mu_P^2(1-2\delta)^2\Big)$. Assuming that the probability distribution of X_i can be approximated by a Gaussian, the probabilities of the possible values of $\mathrm{sgn}(X_i)$ are given by $\mathrm{P}(\mathrm{sgn}(X_i) = \pm 1) = \Big[1 \pm \mathrm{erf}\Big(\mathrm{E}(X_i)/\sqrt{2\,\mathrm{V}(X_i)}\Big)\Big]/2$. Then the expected value of the overlap m is

$$\begin{aligned}\mathrm{E}(m) = \frac{1}{N}\sum_i \mathrm{E}(\mathrm{sgn}(X_i)) &= \mathrm{erf}\left(\sqrt{\frac{(N-1)\mu_P^2(1-2\delta)^2}{2((K+\mu_{\eta^2})\mu_{P^2} - \mu_P^2(1-2\delta)^2)}}\right)\\ &\approx \mathrm{erf}\left(\sqrt{\frac{N(1-2\delta)^2}{2(K+\mu_{\eta^2})\mu_{P^2}/\mu_P^2}}\right),\end{aligned} \tag{2.76}$$

where the approximation holds when $N, K \gg 1$. We compare this result against a numerically simulated Hopfield network in figure 2.10, and find them to be in agreement. We note that in absence of additive noise ($\eta_{ij} = 0$), N and K enter this expression only through K/N, so any definition of the capacity involving this expected overlap will scale linearly with the number of spins. For example, we may define the capacity as the number of patterns for which the average overlap is $m_0 = 1 - \epsilon$ for some $0 < \epsilon \ll 1$, in which case the capacity is found to be $K_c = N\mu_P^2(1-2\delta)^2/2\mu_{P^2}\,\mathrm{erf}^{-1}(m_0)^2 \approx N\mu_P^2(1-2\delta)^2/\mu_{P^2}\ln\big(2/\pi\epsilon^2\big)$. This is linearly related to the number of spins N as expected, and quadratically related to $1 - 2\delta$ quantifying how close the initial state was to the pattern being retrieved. If we assume that P_{ij} are Bernoulli variables with parameter p, corresponding to p of the couplings being retained and the rest deleted, then we have $\mu_P = \mu_{P^2} = p$. In that case, the capacity is also linearly related to the probability p with which the couplings are retained. The capacity goes

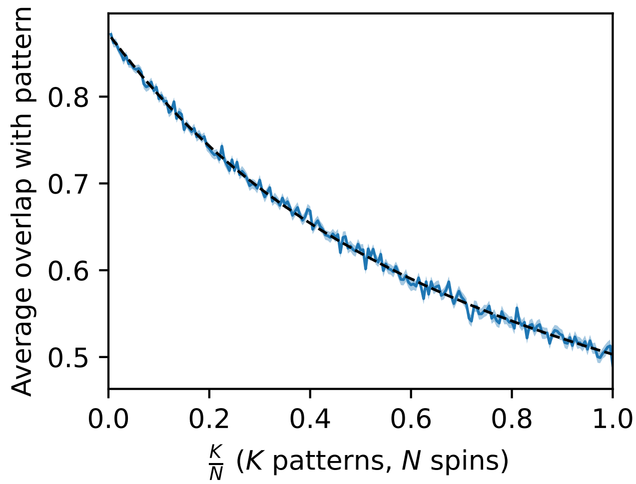

Figure 2.10: Average overlap of the state with the original pattern after flipping $\delta = 0.1$ of the spins and doing one step of Hopfield update, in a classic Hopfield network (with 200 spins and 2-spin interactions), as a function of the scaled number of patterns $\alpha = K/N$. The coupling constants are defined by the Hebb rule without clipping, but with Gaussian additive noise with mean 0 and variance 50. Also, $p = 0.8$ of the couplings have been retained and the rest deleted.The analytical estimate obtained in eq. (2.76) has also been plotted.

down if we discard more coupling constants or require a larger basin of attraction around each pattern, which agrees with our intuition.

We now consider the model where the couplings are clipped to be ± 1, following eq. (2.73). This "clipped" model has been discussed in refs. [4] and [12]; here we have also included the possibility of noise or network dilution by introducing η_{ij} and P_{ij}. The overlap $\check{m}$ of a perturbed pattern after one step of the Hopfield update (as described in section 2.3) is

$$\begin{aligned}\check{m} &= \frac{1}{N}\sum_i \operatorname{sgn}\left(\sum_{j\neq i} \operatorname{sgn}\left(P_{ij}\left(\sum_\mu \xi_i^\mu \xi_j^\mu + \eta_{ij}\right)\right)\xi_i^1\xi_j^1 s_j\right)\\ &= \frac{1}{N}\sum_i \operatorname{sgn}\left(\sum_{j\neq i} \operatorname{sgn}\left(\sum_\mu \xi_i^\mu \xi_j^\mu \xi_i^1\xi_j^1 + \eta_{ij}\xi_i^1\xi_j^1\right)\check{P}_{ij} s_j\right),\end{aligned} \tag{2.77}$$

where $\check{P}_{ij} = \operatorname{sgn}(P_{ij})$. To find the expected value of $\check{m}$, we define $W_{ij} = \sum_\mu \xi_i^\mu \xi_j^\mu \xi_i^1 \xi_j^1 + \eta_{ij}\xi_i^1\xi_j^1$, $X_{ij} = W_{ij}P_{ij}s_j$, and $\check{X}_i = \sum_{j\neq i}\operatorname{sgn}(X_{ij})$. Then for any fixed i, $X_i > 0$ if and only if the number of variables X_{ij} (for $j \neq i$) taking positive values is more than $(N-1)/2$.

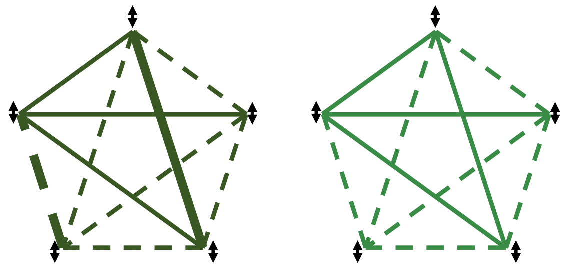

Figure 2.11: Schematic of a Hopfield network with $N = 5$ spins and couplings restricted to be binary. (a) corresponds to the model described in figure 2.9, with solid and dashed lines corresponding to ferromagnetic ($J_{ij} > 0$) and antiferromagnetic ($J_{ij} < 0$) interactions respectively and their widths corresponding to their magnitudes $\left|J_{ij}\right|$. (b) is the model with "clipped couplings", where all the lines have the same width (corresponding to the same magnitude of $\left|\check{J}_{ij}\right| = 1$) and can only differ in sign.

We see that $\mathrm{E}\big(W_{ij}\big) = 1$ and $\mathrm{V}\big(W_{ij}\big) = K + \mu_{\eta^2} - 1$. Then, assuming as before that the probability distribution is approximately Gaussian, the probability of any particular W_{ij} being positive is $\mathrm{P}\big(W_{ij} > 0\big) = (1 + \operatorname{erf}\Big(1/\sqrt{2(K + \mu_{\eta^2} - 1)}\Big))/2$. The expected sign of X_{ij} is $\mathrm{E}\big(\operatorname{sgn}\big(X_{ij}\big)\big) = \mu_{\check{P}}(1 - 2\delta)\operatorname{erf}\Big(1/\sqrt{2(K + \mu_{\eta^2} - 1)}\Big) \equiv \check{e}$, and its variance is $\mathrm{V}\big(\operatorname{sgn}\big(X_{ij}\big)\big) = \mu_{\check{P}^2} - \check{e}^2$. The variables X_{ij} for different j are independent, so their sum $\check{X}_i$ has mean $(N-1)\check{e}$ and variance $(N-1)(\mu_{\check{P}^2} - \check{e}^2)$. Therefore its expected sign is

$$
\begin{aligned}
\mathrm{E}\big(\operatorname{sgn}\big(\check{X}_i\big)\big) &= \operatorname{erf}\left(\frac{(N-1)\mu_{\check{P}}(1-2\delta)\operatorname{erf}\Big(1/\sqrt{2(K+\mu_{\eta^2}-1)}\Big)}{\sqrt{2(N-1)\left(\mu_{\check{P}^2} - \operatorname{erf}\Big(1/\sqrt{2(K+\mu_{\eta^2}-1)}\Big)^2\right)}}\right) \\
&\approx \operatorname{erf}\left((1-2\delta)\sqrt{\frac{N\mu_{\check{P}}^2}{\pi\mu_{\check{P}^2}(K+\mu_{\eta^2})}}\right),
\end{aligned}
\tag{2.78}
$$

where the approximation holds for $K \gg 1$. The expected value of the overlap is thus $\mathrm{E}(\check{m}) = \sum_i \mathrm{E}\big(\operatorname{sgn}\big(\check{X}_i\big)\big)/N = \operatorname{erf}\Big((1-2\delta)\mu_{\check{P}}\sqrt{N/\pi\mu_{\check{P}^2}(K+\mu_{\eta^2})}\Big)$, which (in the cases where the multiplicative noise only takes the values $0, \pm 1$) is identical to the expression of $\mathrm{E}(m)$ that we obtained for the usual Hopfield network if we replace N with $2N/\pi$. We plot the

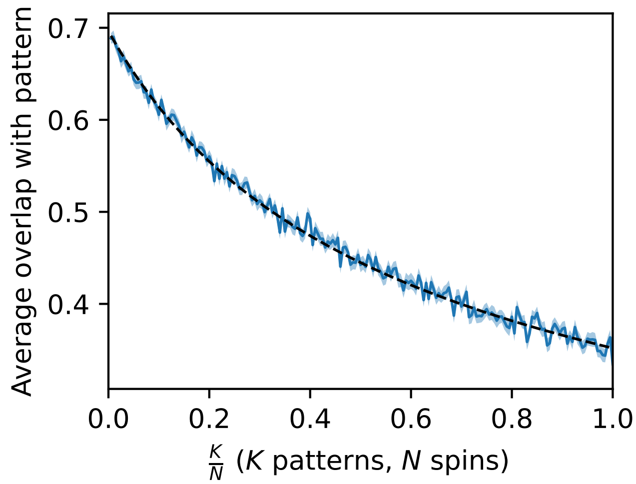

Figure 2.12: Average overlap of the state with the original pattern after flipping $\delta = 0.1$ of the spins and doing one step of Hopfield update, in a classic Hopfield network (with 200 spins and 2-spin interactions), as a function of the scaled number of patterns $\alpha = K/N$. The coupling constants are defined by the Hebb rule, but with Gaussian additive noise with mean 0 and variance 50. The coupling constants have also been clipped to have magnitude 1, and $p = 0.9$ of them have had their sign flipped. The analytical estimate obtained in eq. (2.78) has also been plotted.

expected overlaps for a numerically simulated Hopfield network with clipped interactions in figure 2.12, which are found to agree with the above result. Therefore, in absence of additive noise, any measure of capacity for this clipped Hopfield network will also scale linearly with the number of spins, but will be less than the capacity of the usual Hopfield network by a factor of $2/\pi$. This agrees with the results in [4] and [12]. The effects of clipping the couplings and removing them (with probability $1 - p$) affect the capacity independently, reducing it by factors of $2/\pi$ and p respectively.

CHAPTER 3

PHASE DIAGRAM OF QUARTIC DENSE ASSOCIATE MEMORY USING REPLICA THEORY

3.1 Introduction

In the previous chapter, we analyzed modern Hopfield networks with multi-spin interactions using the behavior of the network after one step of synchronous update of its component spins. However, it is difficult to extend this approach to study the network after long-time dynamics, in which case we can expect it to have reached an equilibrium. More progress can be made by instead using the tools from equilibrium statistical physics, by treating the Hopfield network as an Ising spin system at some temperature $1/\beta$ and finding the partition function and free energy of this system. The patterns $\{\boldsymbol{\xi}^\mu\}$ which define the interaction strengths and the energy of the model are chosen randomly, and we need to average over the possible realizations of the patterns when obtaining the free energy. This is similar to the configurational average over interaction coefficients required for finding the free energy of spin glass systems such as the Sherrington-Kirkpatrick model, and we can use similar techniques such as replica theory for the Hopfield model as well.

The replica symmetric free energy and the equations of state of the Hopfield network with two-spin interactions was obtained by Amit [3]. Solving these equations yielded the phase diagram of this network, showing distinct phases corresponding to retrieval (at sufficiently low temperatures and number of patterns), spin glass (for low temperatures but large numbers of patterns), and paramagnetic behavior (at high temperature). This was later extended to consider replica symmetry breaking [35].

Replica theory has also been used to find the phases of modern Hopfield networks with higher order interactions. For example, Arenzon and de Almeida [43] analyzed the network with both second and fourth order interactions, where the coefficients mediating the fourth order interactions had contributions from all possible pairs of patterns rather than the patterns themselves. Alemanno et al. [65], and Agliari et al. [64], on the other hand, analyzed the quartic model where the number of patterns embedded scales linearly as the system size (as opposed to cubically) but the interaction coefficients we allowed to have high levels of noise. Gardner [73] obtained the equations of state and the phase diagram of the network with p-spin interactions, by approximating the effect of the non-retrieved patterns as Gaussian and using cumulants. More recently, the phase diagrams of networks with multi-spin interactions have been obtained using replica symmetry [74] and under replica symmetry breaking [75], also by approximating the effective noise in the network due to the patterns not being retrieved as Gaussian random variables.

Here we explore another technique of obtaining the free energy and phase diagram of the Hopfield network with quartic interactions using replica theory. Our approach closely follows the techniques used by Amit et al. [3] for the classic HOpfield model, and does not directly invoke the approximations used in refs. [39, 74]. However, this approach can be used for networks with quartic interactions and can be potentially extended to networks whose interaction orders are higher powers of 2, but it can not be directly generalized to networks with other interaction orders.

In section 3.2 we describe the model studied here. In section 3.3, we study the network with a small number of embedded patterns, in which case we expect the patterns to be retrieved as long as the temperature is sufficiently low. In section 3.4 we find the replicated partition function of the network in the limit of a large number of patterns ($K \sim N^3$), and in section 3.5 we obtain its free energy assuming replica symmetry. Finally in section 3.6 we extremize the free energy to obtain the equations of state, which we solve to find the phases and the phase diagram of the network in 3.7.

3.2 Description of the model

We consider a set of K patterns $\boldsymbol{\xi}^\mu$ with componetns ξ_i^μ, $i \in \{1, \ldots, N\}$, $\mu \in \{1, \ldots, K\}$, in an Ising system with N spins. The pattern components are chosen randomly and independently, so each ξ_i^μ is an i.i.d. Rademacher random variable taking value 1 or -1 with equal probability. Following Krotov and Hopfield [8], we define the energy function

$$E(\sigma_i) = -\sum_\mu F\left(\sum_i \xi_i^\mu \sigma_i\right), \tag{3.1}$$

where F is a chosen function, often a power function $F(x) = x^n$. We note that for $n = 1$, the above Hamiltonian corresponds to a non-interacting system of spins, where each spin σ_i experiences a local field $h_i = \sum_\mu \xi_i^\mu$. Such a Hamiltonian can only learn one pattern; for $K > 1$ each spin will simply tend to orient in accordance with the spins at that site for the majority of patterns, and the system can not learn the pattern individually. For $F(x) = x^n$ with $n > 1$, the Hamiltonian can be rewritten as

$$\begin{aligned}
E(\sigma_i) &= -\sum_\mu \left(\sum_i \xi_i^\mu \sigma_i\right)^n \\
&= -\sum_\mu \left(\sum_{i_1=1}^N \xi_{i_1}^\mu \sigma_{i_1}\right) \cdots \left(\sum_{i_n=1}^N \xi_{i_n}^\mu \sigma_{i_n}\right) \\
&= \sum_{i_1,\ldots,i_m} \left(\sum_\mu \xi_{i_1}^\mu \cdots \xi_{i_n}^\mu\right) \sigma_{i_1} \cdots \sigma_{i_n} \\
&= \sum_{i_1,\ldots,i_n} J_{i_1,\ldots i_n} \sigma_{i_1} \cdots \sigma_{i_n},
\end{aligned} \tag{3.2}$$

Which is the Hamiltonian involving all possible (infinite-range) n-spin interactions, where the coupling coefficients are defined using the memories. For $n = 2$ this corresponds to the usually defined Hopfield network with two-spin interactions. Here we have allowed self-interactions as the indices $i_1 \ldots i_n$ in the above expression need not be all different; for Ising spins this means that the Hamiltonian also includes $(n-2p)$-spin interactions for $0 \le p \le n/2$.

If $F(x)$ is an increasing function of x, then $F(\sum_i \xi_i^\mu \sigma_i)$ has the largest value of $F(N)$ (and thus the smallest energy contribution) when $\boldsymbol{\sigma} = \boldsymbol{\xi}^\mu$. The memories are expected to be low energy states of the Hamiltonian (if we have only one memory then that state will be the ground state). If the rate of change of F is sufficiently large away from its maximum value, we expect that the energy minima will occur only when one of the terms in the energy expression is at its minimum (i.e. the memories will be the metastable states of this Hamiltonian).

In this work, we will focus on the case where $n = 4$ (the quartic model). This corresponds to the presence of 4-spin and 2-spin interactions between the component spins in the network.

3.3 Free energy of the quartic modern Hopfield network with small number of embedded patterns

For quadratic Hamiltonians, we can use the property of Gaussian integral to make the degree of freedom linear in the exponent of the partition (Hubbard-Stratonovich transformation):

$$e^{ax^2/2} = \sqrt{\frac{a}{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}y \, e^{-ay^2/2+ayx} \quad \text{for} \quad a > 0. \tag{3.3}$$

We also note that

$$e^{-ax^2/2} = \sqrt{\frac{a}{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}y \, e^{-ay^2/2+iayx} \quad \text{for} \quad a > 0. \tag{3.4}$$

For a quartic term in the exponent (corresponding to the modern Hopfield network with $n = 4$), we can introduce two integration variables to similarly obtain (for $a > 0$)

$$\begin{aligned}
e^{ax^4/2} &= \sqrt{\frac{a}{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}y\, e^{-ay^2/2+ayx^2} \\
&= \sqrt{\frac{a}{2\pi}} \int_{0}^{\infty} \mathrm{d}y\, e^{-ay^2/2}\Big(e^{ayx^2} + e^{-ayx^2}\Big) \\
&= \frac{a}{\pi\sqrt{2}} \int_{0}^{\infty} \mathrm{d}y\, \sqrt{y} e^{-ay^2/2} \int_{-\infty}^{\infty} \mathrm{d}z\, e^{-ayz^2}\Big(e^{2ayzx} + e^{2iayzx}\Big) \\
&= \frac{a\sqrt{2}}{3\pi} \int_{0}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}z\, e^{-au^{4/3}/2-au^{2/3}z^2}\Big(e^{2au^{2/3}zx} + e^{2iau^{2/3}zx}\Big) \quad \Big(u = y^{3/2}\Big) \\
&= \frac{2a\sqrt{2}}{3\pi} \int_{0}^{\infty} \mathrm{d}u \int_{0}^{\infty} \mathrm{d}z\, e^{-au^{4/3}/2-au^{2/3}z^2}\Big(\cosh\Big(2au^{2/3}zx\Big) + \cos\Big(2au^{2/3}zx\Big)\Big).
\end{aligned} \tag{3.5}$$

We can use this to find the partition function and free energy for quartic (and in principle higher orders which are powers of 2) Hopfield-like models, generalizing the analysis of [3]. We will ignore the imaginary exponent term corresponding to the $y < 0$ part of the integral, as for most values of u, $\cosh(u) \gg \cos(u)$.

Let us first consider the case with a finite number of memories ($K \ll N$) for a fourth order Hamiltonian ($n = 4$, $H = -\sum_\mu (\sum_i \xi_i^\mu \sigma_i)^4 / 2N^3$, with the power of N chosen to keep the energy extensive). Then the partition function can be found to be

$$Z = \mathrm{Tr}\exp\left(\frac{\beta}{2N^3}\left(\sum_i \xi_i^\mu \sigma_i\right)^4\right)$$
$$\sim \mathrm{Tr}\iint \prod_\mu \mathrm{d}u_\mu\,\mathrm{d}z_\mu \exp\left(\sum_\mu\left(-\frac{\beta}{2N^3}u_\mu^{4/3} - \frac{\beta}{N^3}u_\mu^{2/3}z_\mu^2 + \frac{2\beta}{N^3}{u_\mu}^{2/3}z_\mu\sum_i \xi_\mu^i\sigma_i\right)\right) \tag{3.6}$$

$$= \iint \prod_\mu \mathrm{d}u_\mu\,\mathrm{d}z_\mu \exp\Bigg(- \frac{\beta}{2N^3}\sum_\mu u_\mu^{4/3} - \frac{\beta}{N^3}\sum_\mu u_\mu^{2/3}z_\mu^2$$
$$+ \sum_i \ln\cosh\left(\frac{2\beta}{N^3}\sum_\mu {u_\mu}^{2/3}z_\mu\xi_\mu^i\right)\Bigg). \tag{3.7}$$

Approximating the above integral by Laplace's method, we can find the free energy,

$$F \sim \frac{1}{2N^3}\sum_\mu u_\mu^{4/3} + \frac{1}{N^3}\sum_\mu u_\mu^{2/3}z_\mu^2 - \frac{1}{\beta}\sum_i \ln\cosh\left(\frac{2\beta}{N^3}\sum_\mu {u_\mu}^{2/3}z_\mu\xi_\mu^i\right). \tag{3.8}$$

Extremizing this Free energy with respect to u_μ and z_μ, we get

$$u_\mu^{2/3} + z_\mu^2 = 2z_\mu \sum_i \xi_i^\mu \tanh\left(\frac{2\beta}{N^3}\sum_\nu u_\nu^{2/3}z_\nu\xi_i^\nu\right), \tag{3.9}$$
$$\text{and}\quad z_\mu = \sum_i \xi_i^\mu \tanh\left(\frac{2\beta}{N^3}\sum_\nu u_\nu^{2/3}z_\nu\xi_i^\nu\right). \tag{3.10}$$

We solve the above to obtain

$$z_\mu = \sum_i \xi_i^\mu \tanh\left(\frac{2\beta}{N^3}\sum_\nu z_\nu^3\xi_i^\nu\right), \qquad u_\mu = z_\mu^3. \tag{3.11}$$

To find the physical significance of z_μ, we can study the saddle-point condition of the partition function in eq. (3.6),

$$z_\mu = \sum_i \xi_i^\mu \sigma_i, \qquad\qquad u_\mu = z_\mu^3, \tag{3.12}$$

i.e. z_μ gives the overlap of the state with the μ^{th} state, and u_μ is simply its cube and does not give any additional information.

Here we have not averaged over the distribution of the memories ξ_i^μ (which would be analogous to averaging over the interactions J_{ij} in a usual spin-glass Hamiltonian and needs to be done using replicas). But for the case where we have few memories, we can consider the case where the state σ_i has high overlap with one of the memories (say $\boldsymbol{\xi}^1$) and negligible overlap with the rest of the memories (for a large number of memories, we can not assume that the overlap of the state with every single of the other states is negligible). In that case we have $z_1 = z$, $z_\mu \sim 0$ for $\mu \geq 2$, and the equation of state for z is

$$\frac{z}{N} = \frac{1}{N} \sum_i \xi_i^1 \tanh\left(\frac{2\beta}{N^3} z^3 \xi_i^1\right) = \tanh\left(2\beta\left(\frac{z}{N}\right)^3\right). \tag{3.13}$$

For an 8^{th} order Hamiltonian, we can similarly use the identity that

$$\begin{aligned} e^{ax^8/3} &= \frac{a}{\pi}\sqrt{\frac{2}{3}} \int \mathrm{d}u \int \mathrm{d}z\, e^{-au^{4/3}/2 - au^{2/3}z^2 + 2au^{2/3}zx^2} \\ &= \frac{1}{3}\left(\frac{a}{\pi}\right)^{3/2} \int \mathrm{d}u \int \mathrm{d}v \int \mathrm{d}w \exp\Big(-\frac{au}{2} - au^{1/2}v^{4/3} - 2au^{1/2}v^{2/3}w^2 \\ &\quad + 4au^{1/2}v^{2/3}wx\Big), \end{aligned} \tag{3.14}$$

(Ignoring the contributions from negative u and v in the integraks). By similar analysis as before, we obtain the free energy for the eighth order Hamiltonian,

$$F \sim \sum_{\mu} \left(\frac{1}{2N} u_{\mu} + \frac{1}{N} u_{\mu}^{1/2} v_{\mu}^{4/3} + \frac{2}{N} u_{\mu}^{1/2} v_{\mu}^{2/3} w_{\mu}^{2} \right) - \frac{1}{\beta} \sum_{i} \ln\cosh \left(\frac{4\beta}{N} \sum_{\mu} u_{\mu}^{1/2} v_{\mu}^{2/3} w_{\mu} \xi_{i}^{\mu} \right), \tag{3.15}$$

for which the extremization conditions are

$$w_{\mu} = \sum_{i} \xi_{i}^{\mu} \tanh \left(\frac{4\beta}{N} \sum_{\nu} w_{\nu}^{7} \xi_{i}^{\nu} \right), \qquad v_{\mu} = w_{\mu}^{3}, \qquad u_{\mu} = w_{\mu}^{8}. \tag{3.16}$$

3.4 Replicated partition function for the model with large number of patterns

Let us now consider the case where we let the number of patterns K be large, and find the configurational average (i.e. averaged over the distribution of memories, denoted here by $[\cdots]$) of the replicated partition function Z^p. For simplicity, we will only consider the retrieval of the first pattern $\{\boldsymbol{\xi}^1\}$. We have

$$
\begin{aligned}
\left[Z^p_{(4)}\right] &= \left[\operatorname{Tr}\exp\left(\frac{\beta}{N^3}\sum_{\mu=1}^{K}\sum_{\rho=1}^{p}\left(\xi_i^\mu\sigma_i^\rho\right)^4\right)\right] \\
&\sim \left[\operatorname{Tr}\prod_{\rho,\mu}\int \mathrm{d}u_\mu^\rho\int \mathrm{d}z_\mu^\rho\exp\left(\sum_{\mu,\rho}\left(-\frac{\beta}{N^3}(u_\mu^\rho)^{4/3}-\frac{2\beta}{N^3}(u_\mu^\rho)^{2/3}(z_\mu^\rho)^2\right.\right.\right. \\
&\quad\left.\left.\left.+\frac{4\beta}{N^3}(u_\mu^\rho)^{2/3}z_\mu^\rho\sum_i\xi_i^\mu\sigma_i^\rho\right)\right)\right] \\
&= \left[\operatorname{Tr}\prod_{\rho,\mu}\int \mathrm{d}u_\mu^\rho\int \mathrm{d}z_\mu^\rho\right. \\
&\quad\exp\left(\sum_{\mu\geq 2,\rho}\left(-\frac{\beta}{N^3}(u_\mu^\rho)^{4/3}-\frac{2\beta}{N^3}(u_\mu^\rho)^{2/3}(z_\mu^\rho)^2+\frac{4\beta}{N^3}(u_\mu^\rho)^{2/3}z_\mu^\rho\sum_i\xi_i^\mu\sigma_i^\rho\right)\right. \\
&\quad\left.\left.+\sum_\rho\left(-\frac{\beta}{N^3}(u_1^\rho)^{4/3}-\frac{2\beta}{N^3}(u_1^\rho)^{2/3}(z_1^\rho)^2+\frac{4\beta}{N^3}(u_1^\rho)^{2/3}z_1^\rho\sum_i\xi_i^1\sigma_i^\rho\right)\right)\right].
\end{aligned}
\tag{3.17}
$$

We perform the configurational average on the terms corresponding to $\mu\geq 2$ to obtain

$$
\begin{aligned}
&\left[\exp\left(\sum_{\mu\geq 2,\rho}\left(-\frac{\beta}{N^3}(u_\mu^\rho)^{4/3}-\frac{2\beta}{N^3}(u_\mu^\rho)^{2/3}(z_\mu^\rho)^2+\frac{4\beta}{N^3}(u_\mu^\rho)^{2/3}z_\mu^\rho\sum_i\xi_i^\mu\sigma_i^\rho\right)\right)\right] \\
&=\exp\left(\sum_{\mu\geq 2,\rho}\left(-\frac{\beta}{N^3}(u_\mu^\rho)^{4/3}-\frac{2\beta}{N^3}(u_\mu^\rho)^{2/3}(z_\mu^\rho)^2\right)+\sum_{\mu\geq 2,i}\ln\cosh\left(\frac{4\beta}{N^3}\sum_\rho(u_\mu^\rho)^{2/3}z_\mu^\rho\sigma_i^\rho\right)\right) \\
&=\exp\left(\sum_{\mu\geq 2,\rho}\left(-\frac{\beta}{N}(g_\mu^\rho)^{4/3}-\frac{2\beta}{N}(g_\mu^\rho)^{2/3}(m_\mu^\rho)^2\right)+\sum_{\mu\geq 2,i}\ln\cosh\left(\frac{4\beta}{N^{3/2}}\sum_\rho(g_\mu^\rho)^{2/3}m_\mu^\rho\sigma_i^\rho\right)\right) \\
&\left[z_\mu^\rho=\sqrt{N}m_\mu^\rho,\quad u_\mu^\rho=N^{3/2}g_\mu^\rho,\quad m_\mu^\rho\sim O(1)\text{ and }g_\mu^\rho\sim O(1)\text{ for }\mu>1\right] \\
&\approx\exp\left(\sum_{\mu\geq 2,\rho}\left(-\frac{\beta}{N}(g_\mu^\rho)^{4/3}-\frac{2\beta}{N}(g_\mu^\rho)^{2/3}(m_\mu^\rho)^2\right)+\sum_{\mu\geq 2,i}\frac{8\beta^2}{N^3}\sum_{\rho,\sigma}(g_\mu^\rho g_\mu^\sigma)^{2/3}m_\mu^\rho\sigma_i^\rho m_\mu^\sigma\sigma_i^\sigma\right) \\
&=\exp\left(-\frac{\beta}{N}\sum_{\mu,\rho}(g_\mu^\rho)^{4/3}-\frac{2\beta}{N}\sum_{\mu,\rho,\sigma}(g_\mu^\rho)^{1/3}m_\mu^\rho K^\mu_{\rho\sigma}(g_\mu^\sigma)^{1/3}m_\mu^\sigma\right),
\end{aligned}
\tag{3.18}
$$

where, at the last step, the sum over μ being for $\mu \geq 2$ is implicit, and $K^{\mu}_{\rho\sigma}$ is defined as

$$K^{\mu}_{\rho\sigma} = \delta_{\rho\sigma} - \frac{4\beta}{N^2}(g^{\rho}_{\mu})^{1/3}(g^{\sigma}_{\mu})^{1/3}\sum_i \sigma^{\rho}_i \sigma^{\sigma}_i. \tag{3.19}$$

Unlike the case of the quadratic Hamiltonian discussed in [3], the matrix now depends on the memory index μ and involves the variables u^{ρ}_{μ} which may then need to be integrated over, which makes the problem more complicated.

Integrating the above over m^{ρ}_{μ} and g^{ρ}_{μ}, we get

$$\begin{aligned}
&\prod_{\mu\geq 2,\rho}\int \mathrm{d}g^{\rho}_{\mu}\int \mathrm{d}m^{\rho}_{\mu}\exp\left(-\frac{\beta}{N}\sum_{\mu,\rho}(g^{\rho}_{\mu})^{4/3} - \frac{2\beta}{N}\sum_{\mu,\rho,\sigma}(g^{\rho}_{\mu})^{1/3}m^{\rho}_{\mu}K^{\mu}_{\rho\sigma}(g^{\sigma}_{\mu})^{1/3}m^{\sigma}_{\mu}\right)\\
&= \prod_{\rho,\mu}\int \mathrm{d}g^{\rho}_{\mu}\exp\left(-\frac{\beta}{N}\sum_{\mu,\rho}(g^{\rho}_{\mu})^{4/3}\right)\int \mathrm{d}m^{\rho}_{\mu}\prod_{\mu}\exp\left(-\frac{2\beta}{N}\sum_{\rho,\sigma}m^{\rho}_{\mu}(g^{\rho}_{\mu})^{1/3}(g^{\sigma}_{\mu})^{1/3}K^{\mu}_{\rho\sigma}m^{\sigma}_{\mu}\right)\\
&= \prod_{\rho,\mu}\int \mathrm{d}g^{\rho}_{\mu}\exp\left(-\frac{\beta}{N}\sum_{\mu,\rho}(g^{\rho}_{\mu})^{4/3}\right)\prod_{\mu}(\det\tilde{\mathbf{K}}^{\mu})^{-1/2}\quad\left(\tilde{K}^{\mu}_{\rho\sigma} = (g^{\rho}_{\mu})^{1/3}(g^{\sigma}_{\mu})^{1/3}K^{\mu}_{\rho\sigma}\right)\\
&= \prod_{\rho,\mu}\int \mathrm{d}g^{\rho}_{\mu}\exp\left(-\frac{\beta}{N}\sum_{\mu,\rho}(g^{\rho}_{\mu})^{4/3}\right)\prod_{\mu}\left(\prod_{\rho}(g^{\rho}_{\mu})^{2/3}\det\mathbf{K}^{\mu}\right)^{-1/2}\\
&= \prod_{\rho,\mu}\int \mathrm{d}g^{\rho}_{\mu}\,(g^{\rho}_{\mu})^{-1/3}\exp\left(-\frac{\beta}{N}\sum_{\mu,\rho}(g^{\rho}_{\mu})^{4/3}\right)\prod_{\mu}(\det\mathbf{K}^{\mu})^{-1/2}\\
&= \left(\prod_{\rho}\int \mathrm{d}g^{\rho}\,(g^{\rho})^{-1/3}\exp\left(-\frac{\beta}{N}\sum_{\rho}(g^{\rho})^{4/3}\right)(\det\mathbf{K})^{-1/2}\right)^{K-1}\\
&\sim \left(\prod_{\rho}\int \mathrm{d}\gamma^{\rho}\exp\left(-\frac{\beta}{N}\sum_{\rho}(\gamma^{\rho})^{2}\right)(\det(\mathbf{I}+\mathbf{D}))^{-1/2}\right)^{K-1}\\
&\left(\gamma^{\rho} = (g^{\rho})^{2/3},\quad D_{\rho\sigma} = -\frac{4\beta}{N}\sqrt{\gamma^{\rho}\gamma^{\sigma}}q_{\rho\sigma},\quad q_{\rho\sigma} = \frac{1}{N}\sum_i \sigma^{\rho}_i\sigma^{\sigma}_i,\quad q_{\rho\rho} = 1\right)\\
&\approx \left(\prod_{\rho}\int \mathrm{d}\gamma^{\rho}\exp\left(-\frac{\beta}{N}\sum_{\rho}(\gamma^{\rho})^{2}\right)\exp\left(-\frac{1}{2}\operatorname{tr}\ln(\mathbf{I}+\mathbf{D})\right)\right)^{K}\\
&= \left(\prod_{\rho}\int \mathrm{d}\gamma^{\rho}\exp\left(-\frac{\beta}{N}\sum_{\rho}(\gamma^{\rho})^{2}\right)\exp\left(\frac{1}{2}\sum_{j=1}^{\infty}\frac{(-1)^j}{j}\operatorname{tr}\mathbf{D}^j\right)\right)^{K}
\end{aligned}$$

$$
\begin{aligned}
&\approx \left(\prod_\rho \int \mathrm{d}\gamma^\rho \exp\left(-\frac{\beta}{N}\sum_\rho (\gamma^\rho)^2\right) \exp\left(\frac{2\beta}{N}\sum_\rho \gamma^\rho + \frac{4\beta^2}{N^2}\sum_{\rho,\sigma}\gamma^\rho\gamma^\sigma q_{\rho\sigma}^2\right)\right)^K \\
&\sim \left(\prod_\rho \int \mathrm{d}\gamma^\rho \exp\left(-\sum_{\rho,\sigma}\left(\frac{\beta}{N}\delta_{\rho\sigma} - \frac{4\beta^2}{N^2}q_{\rho\sigma}^2\right)\gamma_\rho\gamma_\sigma + \frac{2\beta}{N}\sum_\rho \gamma^\rho\right)\right)^K \\
&= \left(\prod_\rho \int \mathrm{d}\gamma^\rho \exp\left(-\frac{\beta}{N}\sum_{\rho,\sigma} R_{\rho\sigma}\gamma_\rho\gamma_\sigma + \frac{2\beta}{N}\sum_\rho \gamma^\rho\right)\right)^K \\
&\left(\mathbf{R} = \left(1-\frac{4\beta}{N}\right)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2, \quad (\mathbf{Q}_2)_{\rho\rho} = 0, \quad (\mathbf{Q}_2)_{\rho\sigma} = q_{\rho\sigma}^2 \text{ for } \rho \neq \sigma\right) \\
&\sim (\det \mathbf{R})^{-K/2} \exp\left(\frac{\beta}{N}\sum_{\rho,\sigma}(\mathbf{R}^{-1})_{\rho\sigma}\right)^K \\
&= \exp\left(-\frac{K}{2}\operatorname{tr}\ln \mathbf{R} + \frac{\beta K}{N}\sum_{\rho,\sigma}(\mathbf{R}^{-1})_{\rho\sigma}\right) \\
&= \prod_{\rho\neq\sigma}\int \mathrm{d}q_{\rho\sigma}\,\delta\left(q_{\rho\sigma} - \frac{1}{N}\sum_i \sigma_i^\rho\sigma_i^\sigma\right) \\
&\quad \exp\left(-\frac{K}{2}\operatorname{tr}\ln\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\right) + \frac{\beta K}{N}\sum_{\rho,\sigma}\left(\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\right)^{-1}\right)_{\rho\sigma}\right) \\
&\sim \prod_{(\rho\sigma)}\int \mathrm{d}q_{\rho\sigma}\int \mathrm{d}r_{\rho\sigma} \exp\left(\sum_{(\rho\sigma)}\left(-iCr_{\rho\sigma}q_{\rho\sigma} + \frac{iCr_{\rho\sigma}}{N}\sum_i \sigma_i^\rho\sigma_i^\sigma\right)\right) \\
&\quad \exp\left(-\frac{K}{2}\operatorname{tr}\ln\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\right) + \frac{\beta K}{N}\sum_{\rho,\sigma}\left(\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\right)^{-1}\right)_{\rho\sigma}\right).
\end{aligned}
\tag{3.20}
$$

A lot of matrices were introduced in the last few steps, so I am summarising them here. Later I can add a map connecting their relations.

$\mathbf{K}^\mu$, with entries $K^\mu_{\rho\sigma} = \delta_{\rho\sigma} - (4\beta/N)(g_\mu^\rho)^{1/3}(g_\mu^\sigma)^{1/3}q_{\rho\sigma}$, is the first $p \times p$ matrix we encounter (or rather a set of K such matrices). After approximating the $\ln\cosh$ function appearing in the partition function as a quadratic function, the integrand becomes a product of multivariate Gaussian functions in the p-component variables $\mathbf{m}_\mu$ (the replica index ρ

in m_μ^ρ labelling the different components of the Gaussian variable, the memory index μ labelling independent Gaussian variables whose integrals are multiplied at the end) with inverse covariance matrices K^μ. (Strictly speaking the inverse covariance is $\tilde{\mathbf{K}}^\mu$ with entries $\tilde{K}_{\rho\sigma}^\mu = (g_\mu^\rho)^{1/3}(g_\mu^\sigma)^{1/3}K_{\rho\sigma}^\mu$, but their determinants are related by some factors of g_μ^ρ and it's easier to work with $\mathbf{K}^\mu$ as its zeroth order part is just $\mathbf{I}$.)

The second matrix we encounter is $\mathbf{D}$, which is defined from $\mathbf{K}^\mu$ as $\mathbf{K}^\mu = \mathbf{I} + \mathbf{D}$. The entries of $\mathbf{D}$ also involve g_μ^ρ and thus we could label them as $\mathbf{D}^\mu$, but it turns out that the integrals for different memory labels μ separate so we can treat it as the integral over one set of parameters g^ρ, raised to the K-th power. The entries of $\mathbf{D}$ are $D_{\rho\sigma} = -(4\beta/N)\sqrt{\gamma^\rho\gamma^\sigma}q_{\rho\sigma}$, where $\gamma^\rho = (g^\rho)^{2/3}$. The elements of $\mathbf{D}$ are thus functions of g^ρ and need to be integrated, and they're $O(1/N)$ so we can express $\det(\mathbf{I}+\mathbf{D})$ as a power series in $\mathbf{D}$.

After integrating with respect to m^ρ and expanding $\operatorname{tr}\ln(\mathbf{I}+\mathbf{D})$ to $O(1/N)$, we can collect the terms quadratic in γ^ρ to obtain the matrix $R = (1-4\beta/N)\mathbf{I} - (4\beta/N)\mathbf{Q}_2$, where $(1/4)\operatorname{tr}\mathbf{D}^2 = (\beta/N)\sum_{\rho,\sigma}(\delta_{\rho\sigma} - R_{\rho\sigma})\gamma^\rho\gamma^\sigma$. $\mathbf{Q}_2$ is the replica overlap matrix with diagonal elements zero and off-diagonal elements $(\mathbf{Q}_2)_{\rho\sigma} = q_{\rho\sigma}^2 = N^{-2}(\sum_i \sigma_i^\rho\sigma_i^\sigma)^2$. Interestingly, $\mathbf{Q}_2$ involves the squares of the replica overlaps rather than the overlaps themselves, and it is also suppressed by a factor of $1/N$. After the γ^ρ integral, we have a factor of $(\det\mathbf{R})^{-K/2}$ in the partition function. We also have a additional term involving the sum of all entries of $\mathbf{R}^{-1}$ which appears because of the first order $\operatorname{tr}\mathbf{D}$ term appearing from the expansion of $\det\mathbf{D}$. The matrices appearing in the above analysis, as well as the connections showing which matrix is defined from which, is shown schematically in figure 3.1.

The overall replicated partition function is

$$[Z^p] = \operatorname{Tr}\prod_\rho \int \mathrm{d}u_1^\rho \int \mathrm{d}z_1^\rho \prod_{(\rho\sigma)} \int \mathrm{d}q_{\rho\sigma}\,\mathrm{d}r_{\rho\sigma} \exp\left(\sum_{(\rho\sigma)}\left(-iCr_{\rho\sigma}q_{\rho\sigma} + \frac{iCr_{\rho\sigma}}{N}\sum_i \sigma_i^\rho\sigma_i^\sigma\right)\right)$$

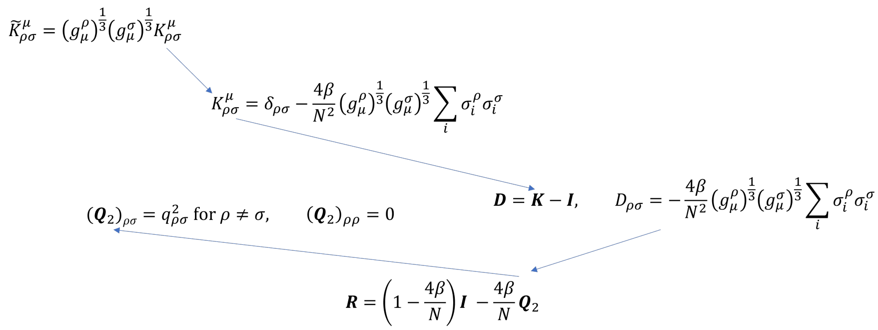

Figure 3.1: Schematic diagram showing the matrices appearing in the derivation of the replicated partition function of the quartic modern Hopfield network.

$$
\exp\left(-\frac{K}{2}\operatorname{tr}\ln\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I}-\frac{4\beta}{N}\mathbf{Q}_2\right)+\frac{\beta K}{N}\sum_{\rho,\sigma}\left(\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I}-\frac{4\beta}{N}\mathbf{Q}_2\right)^{-1}\right)_{\rho\sigma}\right)
$$

$$
\left[\exp\left(\sum_{\rho}\left(-\frac{\beta}{N^3}(u_1^{\rho})^{4/3}-\frac{2\beta}{N^3}(u_1^{\rho})^{2/3}(z_1^{\rho})^2+\frac{4\beta}{N^3}(u_1^{\rho})^{2/3}z_1^{\rho}\sum_i \xi_i^1\sigma_i^{\rho}\right)\right)\right]
$$

$$
\sim \operatorname{Tr}\prod_{\rho}\int \mathrm{d}g^{\rho}\int \mathrm{d}m^{\rho}\prod_{(\rho\sigma)}\int \mathrm{d}q_{\rho\sigma}\,\mathrm{d}r_{\rho\sigma}\exp\left(\sum_{(\rho\sigma)}\left(-iCr_{\rho\sigma}q_{\rho\sigma}+\frac{iCr_{\rho\sigma}}{N}\sum_i\sigma_i^{\rho}\sigma_i^{\sigma}\right)\right)
$$

$$
\exp\left(-\frac{K}{2}\operatorname{tr}\ln\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I}-\frac{4\beta}{N}\mathbf{Q}_2\right)+\frac{\beta K}{N}\sum_{\rho,\sigma}\left(\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I}-\frac{4\beta}{N}\mathbf{Q}_2\right)^{-1}\right)_{\rho\sigma}\right)
$$

$$
\left[\exp\left(\beta N\sum_{\rho}\left(-(g^{\rho})^{4/3}-2(g^{\rho})^{2/3}(m^{\rho})^2+\frac{4}{N}(g^{\rho})^{2/3}m^{\rho}\sum_i\xi_i^1\sigma_i^{\rho}\right)\right)\right]
$$

$$
\left(u_1^{\rho}=N^3g^{\rho},\quad z_1^{\rho}=Nm^{\rho}\right)
$$

$$
=\prod_{\rho}\int \mathrm{d}g^{\rho}\int \mathrm{d}m^{\rho}\prod_{(\rho\sigma)}\int \mathrm{d}q_{\rho\sigma}\,\mathrm{d}r_{\rho\sigma}
$$

$$
\exp\left(-i\sum_{(\rho\sigma)}Cr_{\rho\sigma}q_{\rho\sigma}-\beta N\sum_{\rho}\left((g^{\rho})^{4/3}+2(g^{\rho})^{2/3}(m^{\rho})^2\right)\right)
$$

$$
\exp\left(-\frac{K}{2}\operatorname{tr}\ln\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I}-\frac{4\beta}{N}\mathbf{Q}_2\right)+\frac{\beta K}{N}\sum_{\rho,\sigma}\left(\left(\left(1-\frac{4\beta}{N}\right)\mathbf{I}-\frac{4\beta}{N}\mathbf{Q}_2\right)^{-1}\right)_{\rho\sigma}\right)
$$

$$
\left[\operatorname{Tr}\exp\left(\frac{iC}{N}\sum_{(\rho\sigma)}r_{\rho\sigma}\sum_i\sigma_i^{\rho}\sigma_i^{\sigma}+4\beta\sum_{\rho}(g^{\rho})^{2/3}m^{\rho}\sum_i\xi_i^1\sigma_i^{\rho}\right)\right]
$$

$$= \prod_{\rho} \int \mathrm{d}g^{\rho} \int \mathrm{d}m^{\rho} \prod_{(\rho\sigma)} \int \mathrm{d}q_{\rho\sigma}\, \mathrm{d}r_{\rho\sigma}$$

$$\exp\left(-iC \sum_{(\rho\sigma)} r_{\rho\sigma} q_{\rho\sigma} - \beta N \sum_{\rho} \Big((g^{\rho})^{4/3} + 2(g^{\rho})^{2/3}(m^{\rho})^2\Big)\right)$$

$$\exp\left(-\frac{K}{2}\operatorname{tr}\ln\!\Big(\Big(1-\frac{4\beta}{N}\Big)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\Big) + \frac{\beta K}{N}\sum_{\rho,\sigma}\bigg(\Big(\Big(1-\frac{4\beta}{N}\Big)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\Big)^{-1}\bigg)_{\rho\sigma}\right)$$

$$\exp N\left[\log \operatorname{Tr}\exp\left(\sum_{(\rho\sigma)} \frac{iCr_{\rho\sigma}}{N}\sigma^{\rho}\sigma^{\sigma} + 4\beta\sum_{\rho}(g^{\rho})^{2/3} m^{\rho}\xi^{1}\sigma^{\rho}\right)\right]. \tag{3.21}$$

3.5 Replica symmetric free energy of the network

From the above partition function, the free energy in the thermodynamic limit is found to be

$$f(m^{\rho}, g^{\rho}, q_{\rho\sigma}, r_{\rho\sigma}) = \frac{iC}{\beta p N}\sum_{(\rho\sigma)} r_{\rho\sigma}q_{\rho\sigma} + \frac{1}{p}\sum_{\rho}\Big((g^{\rho})^{4/3} + 2(g^{\rho})^{2/3}(m^{\rho})^2\Big)$$
$$+ \frac{K}{2\beta N p}\operatorname{tr}\ln\!\Big(\Big(1-\frac{4\beta}{N}\Big)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\Big) - \frac{K}{N^2 p}\sum_{\rho,\sigma}\bigg(\Big(\Big(1-\frac{4\beta}{N}\Big)\mathbf{I} - \frac{4\beta}{N}\mathbf{Q}_2\Big)^{-1}\bigg)_{\rho\sigma}$$
$$- \frac{1}{\beta p}\left[\log \operatorname{Tr}\exp\left(\sum_{(\rho\sigma)} \frac{iCr_{\rho\sigma}}{N}\sigma^{\rho}\sigma^{\sigma} + 4\beta\sum_{\rho}(g^{\rho})^{2/3} m^{\rho}\xi^{1}\sigma^{\rho}\right)\right]. \tag{3.22}$$

Let us consider the replica symmetric solution, where $m^{\rho} = m$ and $g^{\rho} = g$ for all ρ, and $r_{\rho\sigma} = r$ and $q_{\rho\sigma} = q$ for all $\rho \neq \sigma$. Then $\mathbf{R} = (1 - 4\beta/N)\mathbf{I} - (4\beta/N)\mathbf{Q_2}$ has all diagonal entries $1 - 4\beta/N$ and all off-diagonal entries are $-4\beta q^2/N$, so its eigenvalues are $\lambda_0 = 1 - 4\beta(1 + (p-1)q^2)/N$ (corresponding to the eigenvector $\mathbf{v}_0$, $v_0^{\rho} = 1$), and $\lambda_k = 1 - 4\beta(1 - q^2)/N$ with multiplicity $p - 1$ (corresponding to the eigenvectors $\mathbf{v}_k$,

$v_k^\rho = \exp(2ik\rho\pi/p)$, $k \in 1, \ldots, p-1$). Then we have

$$\begin{aligned}
\lim_{p\to 0} \frac{1}{p} \operatorname{tr} \ln \mathbf{R} &= \lim_{p\to 0} \frac{1}{p} \left(\ln\left(1 - \frac{4\beta}{N}(1 + (p-1)q^2)\right) + (p-1)\ln\left(1 - \frac{4\beta}{N}(1-q^2)\right)\right) \\
&= \ln\left(1 - \frac{4\beta}{N}(1-q^2)\right) + \lim_{p\to 0}\frac{1}{p}\Bigg[\ln\left(1 - \frac{4\beta}{N}(1-q^2) - \frac{4\beta}{N}pq^2\right) \\
&\quad - \ln\left(1 - \frac{4\beta}{N}(1-q^2)\right)\Bigg] \\
&= \ln\left(1 - \frac{4\beta}{N}(1-q^2)\right) - \frac{4\beta q^2}{N}\, \frac{\mathrm{d}}{\mathrm{d}x} \ln\left(1 - \frac{4\beta}{N}(1-q^2) + x\right)\Bigg|_{x=0} \\
&= \ln\left(1 - \frac{4\beta}{N}(1-q^2)\right) - \frac{4\beta q^2}{N - 4\beta(1-q^2)}.
\end{aligned} \tag{3.23}$$

We can also check that $\mathbf{R}^{-1}$ has all diagonal entries $(1 - 4\beta(1 + (p-2)q^2)/N)/\eta$ and all off-diagonal entries $4\beta q^2/N\eta$, where $\eta = (1 - 4\beta/N)^2 - (p-2)(1-4\beta/N)4\beta q^2/N - (p-1)16\beta^2 q^4/N^2$. Then we have

$$\begin{aligned}
\frac{1}{p}\sum_{\rho,\sigma} (\mathbf{R}^{-1})_{\rho\sigma} &= \frac{N^2 - 4\beta N(1 + (p-2)q^2) + (p-1)4\beta N q^2}{(N-4\beta)^2 - (p-2)(N-4\beta)4\beta q^2 - (p-1)16\beta^2 q^4} \\
&= \frac{N^2 - 4\beta N(1-q^2)}{(N-4\beta)^2 - 4(p-2)\beta(N-4\beta)q^2 - 16(p-1)\beta^2 q^4} \\
\implies \lim_{p\to 0}\frac{1}{p}\sum_{\rho,\sigma} (\mathbf{R}^{-1})_{\rho\sigma} &= \frac{N^2 - 4\beta N(1-q^2)}{(N-4\beta)^2 + 8\beta(N-4\beta)q^2 + 16\beta^2 q^4} = \frac{1}{1 - 4\beta(1-q^2)/N}.
\end{aligned} \tag{3.24}$$

Then the free energy in the limit $p \to 0$ is

$$\begin{aligned}
f_{RS} &= -\frac{iC}{\beta N} rq + g^{4/3} + 2g^{2/3}m^2 + \frac{K}{2\beta N}\left(\ln\left(1 - \frac{4\beta}{N}(1-q^2)\right) - \frac{4\beta q^2}{N - 4\beta(1-q^2)}\right) \\
&\quad - \frac{K}{N(N - 4\beta(1-q^2))} \\
&\quad - \lim_{p\to 0} \frac{1}{\beta p}\left[\log \operatorname{Tr} \exp\left(\frac{iCr}{N}\left(\left(\sum_\rho \sigma^\rho\right)^2 - p\right) + 4\beta g^{2/3} m \xi^1 \sum_\rho \sigma^\rho\right)\right] \\
&= -\frac{iC}{\beta N} rq + g^{4/3} + 2g^{2/3}m^2 + \frac{K}{2\beta N}\left(\ln\left(1 - \frac{4\beta}{N}(1-q^2)\right) - \frac{4\beta q^2}{N - 4\beta(1-q^2)}\right) \\
&\quad - \frac{K}{N(N - 4\beta(1-q^2))} \\
&\quad + \frac{iCr}{\beta N} - \frac{1}{\beta p}\left[\log \operatorname{Tr} \exp\left(\frac{iCr}{N}\left(\sum_\rho \sigma^\rho\right)^2 + 4\beta g^{2/3} m \xi^1 \sum_\rho \sigma^\rho\right)\right] \\
&= \frac{iC}{\beta N} r(1-q) + g^{4/3} + 2g^{2/3}m^2 + \frac{K}{2\beta N}\ln\left(1 - \frac{4\beta}{N}(1-q^2)\right) - \frac{K(2q^2+1)}{N(N - 4\beta(1-q^2))} \\
&\quad - \frac{1}{\beta}\int \frac{\mathrm{d}w}{\sqrt{2\pi}} e^{-w^2/2}\left[\ln 2\cosh\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\xi\right)\right].
\end{aligned} \tag{3.25}$$

3.6 Equations of state

We can obtain the equations of state by extremizing the above free energy with respect to the parameters. The extremization condition for m gives

$$\begin{aligned}
m &= \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}}\left[\xi \tanh\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\xi\right)\right] \\
&= \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}} \tanh\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\right) \\
\implies m &= \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}} \tanh\left(w\sqrt{\frac{2iCr}{N}} + 4\beta m^3\right) \\
&= \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}} \tanh\left(\beta w\sqrt{\alpha_4 r} + 4\beta m^3\right),
\end{aligned} \tag{3.26}$$

Where we have substituted $g = m^3$ (as we have shown below). The equation for q gives

$$
\begin{aligned}
&-\frac{iC}{\beta N} r + \frac{K}{2\beta N}\left(\frac{32\beta^2 q^3}{(N - 4\beta(1-q^2))^2}\right) + \frac{16K\beta^2 q}{2\beta N(N - 4\beta(1-q^2))^2} = 0 \\
\implies r = &\frac{K}{2iC}\left(\frac{16\beta^2 q(2q^2+1)}{(N - 4\beta(1-q^2))^2}\right) = \frac{16q(2q^2+1)}{(1 - 4\beta(1-q^2)/N)^2} \approx 16q(2q^2+1). \qquad (3.27)
\end{aligned}
$$

The equation for g gives

$$
\begin{aligned}
\frac{4}{3}g^{1/3} + \frac{4}{3}g^{-1/3}m^2 &= \frac{8}{3}g^{-1/3}m \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}}\left[\xi \tanh\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\xi\right)\right] \\
\implies g &= \left(2m \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}} \tanh\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\right) - m^2\right)^{3/2} = m^3.
\end{aligned}
\qquad (3.28)
$$

Finally, extremizing with respect to r gives

$$
\begin{aligned}
\frac{iC}{\beta N}(1-q) &= \frac{1}{2\beta}\sqrt{\frac{2iC}{Nr}} \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}}\left[w \tanh\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\xi\right)\right] \\
&= \frac{1}{2\beta}\frac{2iC}{N}\left(1 - \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}}\left[\tanh^2\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\xi\right)\right]\right) \\
\implies q &= \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}}\left[\tanh^2\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\xi\right)\right] \\
&= \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}} \tanh^2\left(w\sqrt{\frac{2iCr}{N}} + 4\beta g^{2/3} m\right) \\
&= \int \frac{e^{-w^2/2}\,\mathrm{d}w}{\sqrt{2\pi}} \tanh^2\left(\beta w\sqrt{\alpha_4 r} + 4\beta m^3\right).
\end{aligned}
\qquad (3.29)
$$

Here we have defined $\alpha_4 = K/N^3$ and chosen $2iC = K\beta^2/N^2$.

The equations of state for m and q here are qualitatively similar to that of the quadratic Hopfield network [3]; the main differences are the presence of m^3 rather than m inside the tanh, as well as $\alpha = K/N$ being replaced by $\alpha_4 = K/N^3$, implying that the phase transitions will occur at $K \sim N^3$. g is a new parameter but it turns out to be simply m^3, at least when extremizing the free energy. The equation connecting q and r is qualitatively different from the corresponding equation for the quadratic model, as it has no dependence on temperature.

We can also compare these equations of state with those obtained by Gardner [73] using slightly different techniques. We have obtained a different equation connection q and r ($r = 36q(2q^2 + 1)$ as opposed to $r = 8q^3$ as obtained by Gardner). Further work needs to be done to find the origin of this discrepancy, but we note that both our and Gardner's equations imply that r depends purely on q via a cubic polynomial with no temperature dependence.

3.7 Phases of the quartic Hopfield network

Finally, we solve the equations of state to find the phases of the model under replica symmetry. The solutions can be classified by whether m and q are zero or non-zero; r is then immediately obtained from q using eq. (3.27). We note that substituting $q = 0$ (and thus $r = 0$) into eq. (3.26) gives $m = 0$, so no solutions can exist with $m \neq 0$ but $q = 0$. The other three possible classes of solutions are:

$m = q = 0$: This corresponds to a **paramagnetic** phase with no order. We note that $m = q = 0$ is a solution for the equations of state for all values of α and β, though they may or may not correspond to the phase of the system (minimum of the free energy) depending on α and β.

$m = 0$, $q \neq 0$: This corresponds to a **spin glass** phase, where the patterns are not being retrieved but there is still spin-glass order indicated by nonzero replica overlap. Solving eq. (3.29) with $m = 0$, we find that a solution with $m = 0$, $q \neq 0$ exists when $T \equiv 1/\beta < \alpha_4/\sqrt{0.251}$.

$m \neq 0$, $q \neq 0$: This is the case where the patterns are being **retrieved**, even if with nonzero

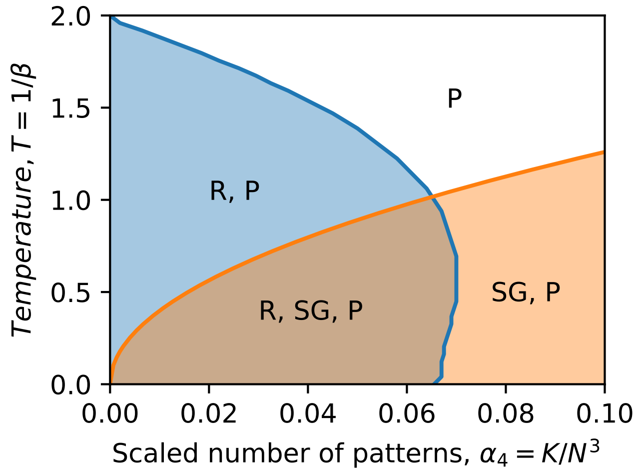

Figure 3.2: Phase diagram of the quartic modern Hopfield network showing the possible solutions to the equations of state, in terms of the parameters temperature $T = 1/\beta$ and scaled number of embedded patterns $\alpha_4 = K/N^3$. The paramagnetic (P) solution with $m = q = 0$ always satisfies the equations of state and is the only possible solution (and must correspond to the phase of the system) in the white region (above the orange line and to the right of the blue line). The spin glass (SG) solution with $m = 0$, $q \neq 0$ exists in the orange region $T \leq \alpha_4/\sqrt{0.251}$, below the orange line. the retrieval (R) solution with $m \neq 0, q \neq 0$ exists in the blue region to the left of the blue line, as obtained numerically. The overlap of the orange and blue regions is the part of the parameter space where all three types of solutions to the equations of state exist, and we need to find the physical phases of the system by checking which solutions correspond to global or local minima of the free energy.

error. To find the parameters for which there exists this retrieval phase $m \neq 0$, we need to solve eqs. (3.26) and (3.29) simultaneously. Numerically solving the equations, we obtain the line separating the retrieval and other phases. In particular, for $T \to 0$, we find that the transition from retrieval to spin glass phase occurs at $\alpha_4 = 0.066$.

The phase diagram indicating the regions of the parameter phase where each the three classes of solutions can exist is shown in figure 3.2. We notice that in the region with sufficiently small α_3 and $1/\beta < \alpha_4/\sqrt{0.251}$, both the spin glass solution $m = 0, q \neq 0$ and and retrieval solution $m \neq 0$ satisfy the equations of state. We expect that in this region, the phase of the network is the retrieval phase $m \neq 0$, and outside this region the $1/\beta = \alpha_4/\sqrt{0.251}$ line separates the spin glass and paramagnetic phases.

The phase diagram we obtained looks qualitatively similar to the phase diagrams obtained previously in [73, 74] by slightly different approaches. However, it is different from the phase diagram of the classic Hopfield network with binary interactions, where the two phase separation lines meet at $\alpha = 0$ but do not intersect. In the quartic network, because of the intersections of the phase separation lines, there are regions of the parameter space where multiple solutions to the equations of state exist, and we need to do further analysis of the free energy to determine which solution corresponds to the physical phase of the system.

3.8 Discussions and Outlook

We analyzed the equilibrium behavior of the modern Hopfield network with a quartic energy function using replica theory, and obtained its free energy and phase diagram assuming replica symmetry. The phase diagram we obtained agrees qualitatively with previous work done on similar models also using replica theory but different approximation techniques. However, more work needs to be done to interpret the phase diagram, in particular because of the multiplicity of solutions to the equations of state. Our methods may be applicable to other spin systems with quartic interactions (such as modern Hopfield networks with noisy or non-Hebbian interactions), and can be potentially extended to consider other interaction orders which are powers of 2. However, they may have limited applicability to general interaction orders.

CHAPTER 4

HOPFIELD NETWORKS WITH QUANTUM PATTERNS

4.1 Introduction

Hopfield networks hold a prominent position in the study of artificial neural networks, enabling associative memory through energy-minimization dynamics in classical Ising systems. With the recent advances in quantum computing, there is growing interest in exploring how quantum principles can be leveraged to extend and enhance classical neural network models [17]. In particular, quantum Hopfield networks have the possibility of having enhanced retrieval accuracy and memory capacity, as well as providing insight into the properties of other quantum many-body systems.

Several possible quantum extensions of Hopfield networks have been proposed. Rebentrost et al. [18] encoded N-component classical Hopfield networks using a system of $\lceil \log_2(N) \rceil$ qubits. Meinhardt et al. [76] considered feed-forward networks where the unitary operators acting on the qubits were trained by a data set. Liu et al. [77] encoded the states of the Hopfield network in the phase angles of qubits and used this model for image recognition. Hopfield networks (with Ising interactions along σ_z) in presence of a transverse field along σ_x have been analyzed by[19–27], and Kimura and Kato [28] extended this to consider modern Hopfield networks (with higher order interactions) in presence of a transverse field. Seki and Nishimori [29] investigate modern Hopfield networks in presence of a transverse field as well as transverse antiferromagnetic interactions. A quantum Hopfield network has also been implemented on hardware [78].

In this chapter, we investigate a different model of the quantum modern Hopfield network. We consider a system of N qubits with K states of the qubits embedded as patterns in the Hamiltonian of the system. Unlike the classical network, however, in this case the patterns are not all σ_z eigenstates, and the terms in the Hamiltonian do not commute with one another in general. This gives rise to the quantum behavior on the network, even in the absence of a transverse field.

In section 4.2, we describe the system being studied, in particular the protocol for choosing the patterns and the Hamiltonian of the system. In ref 4.3, we define the metric we have used to quantify the retrieval behavior of the model. In ref 4.4, we present some preliminary results from simulations, where the metrics defined in the previous section are presented graphically for networks with different numbers of patterns and different levels of "quantumness" In section 4.5, I discuss the preliminary results and discuss further work that can be done to analyze the behavior of these networks.

4.2 Description of the model

We consider a system of N qubits, corresponding to a 2^N dimensional Hilbert space. The patterns are K states of this system $|\boldsymbol{\xi}^\mu\rangle$ $(1 \leq \mu \leq K)$, which we choose to be separable states. Therefore, each pattern can be expressed as

$$|\boldsymbol{\xi}^\mu\rangle = \bigotimes_i \left|\vec{\xi}_i^\mu\right\rangle , \tag{4.1}$$

where $\vec{\xi}_i^\mu$ (with $\left|\vec{\xi}_i^\mu\right| = 1$) is the Bloch vector of the i-th qubit in the state $\left|\boldsymbol{\xi}_i^\mu\right\rangle$.

If we require the patterns $|\boldsymbol{\xi}^\mu\rangle$ to be the low-energy states of the system, then a reasonable choice of the Hamiltonian is

$$\begin{aligned} H &= -\sum_\mu |\boldsymbol{\xi}^\mu\rangle\langle\boldsymbol{\xi}^\mu| = -\sum_\mu \bigotimes_i \left|\vec{\xi}_i^\mu\right\rangle\left\langle\vec{\xi}_i^\mu\right| = -\frac{1}{2^N}\sum_\mu \bigotimes_i \left(\mathbb{1} + \vec{\xi}_i^\mu \cdot \vec{\sigma}_i\right) \\ &= -\frac{1}{2^N}\left(K\mathbb{1} + \sum_{\mu,i} \vec{\xi}_i^\mu \cdot \vec{\sigma}_i + \sum_{\mu,i<j}\left(\vec{\xi}_i^\mu \cdot \vec{\sigma}_i\right)\left(\vec{\xi}_j^\mu \cdot \vec{\sigma}_j\right) + \ldots + \sum_\mu \prod_i \left(\vec{\xi}_i^\mu \cdot \vec{\sigma}_i\right)\right), \end{aligned} \tag{4.2}$$

in which case the patterns span a low-energy subspace with energy -1 and the other states have energy 0. This Hamiltonian includes interactions at all orders, and we can pick out the terms corresponding to n-qubit interactions to define the Hamiltonian

$$H_n = -\sum_{\mu} \sum_{1\le i_1<\cdots<i_n\le N} \prod_{a=1}^{n} \left(\vec{\xi}^{\mu}_{i_a} \cdot \vec{\sigma}_{i_a}\right) = \sum_{1\le i_1<\cdots<i_n\le N} \sum_{\alpha_1,\ldots,\alpha_n} J^{\alpha_1\ldots\alpha_n}_{i_1\ldots i_n} \prod_a \sigma^{\alpha_a}_{i_a}, \quad (4.3)$$

where $\alpha_a \in \{1,2,3\}$ label the Cartesian components, and the n-qubit interaction coefficients are

$$J^{\alpha_1\ldots\alpha_n}_{i_1\ldots i_n} = \prod_a \xi^{\alpha_a}_{i_a}. \quad (4.4)$$

If the pattern Bloch vectors $\vec{\xi}^{\mu}_i$ are all chosen to be $\pm\hat{z}$, then the above Hamiltonian is identical to that of the classical modern Hopfield network with n-spin interactions, and all of its eigenstates are σ^z_i eigenstates which can be treated classically. However, of $\vec{\xi}^{\mu}_i$ are chosen arbitrarily from the Bloch sphere, then the terms in the Hamiltonian H_n do not commute in one another in general, and we expect its eigenstates to be entangled and potentially behave differently from the classical models.

To compare the behaviors of the classical and quantum networks, we choose the following protocol for choosing the Bloch vectors $\vec{\xi}^{\mu}_i$ for the patterns. We choose a polar angle $0 \le \theta_c \le \pi/2$, and then choose the Bloch vectors $\vec{\xi}^{\mu}_i$ independently and uniformly from the part of the Bloch sphere consisting of the two spherical caps $0 \le \theta \le \theta_c$ and $\pi - \theta_c \le \theta \le \pi$. Then $\theta_c \to 0$ corresponds to the case where the patterns are all σ^z eigenstates and the system is effectively classical, and $\theta_c = \pi/2$ is when the Bloch vectors are being chosen uniformly from the entire Bloch sphere. We can thus consider θ_c to be a measure of the "quantumness" of the system. In this work, I mostly focus on the cases where $\theta_c = \pi/2$ (which is the "quantum" model), and $\theta_c = \pi/16$ (where we expect the behavior of the model to be close to that of the classical network). I chose $\theta_c = \pi/16$ instead of 0 to probe the classical behavior as in the latter (completely classical) case, there is significant probability of some of the patterns

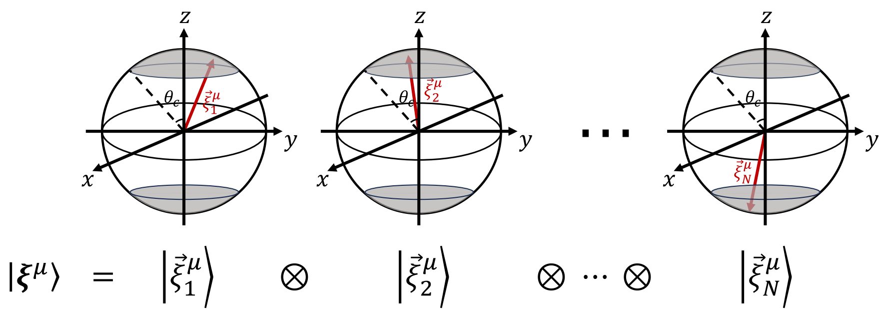

Figure 4.1: Schematic representation of the protocol by which the patterns are generated in the simulated quantum Hopfield network. Each pattern $|\boldsymbol{\xi}^\mu\rangle$ is a separable state of N qubits, where the individual qubits are described by Bloch vectors $\vec{\xi}_i^\mu$. The Bloch vectors are chosen uniformly from the subset of the Bloch sphere consisting of the two regions $0 \leq \theta \leq \theta_c$ and $\pi - \theta_c \leq \theta \leq \pi$.

being identical (and thus the memory subspace having dimension smaller than K) when $K \sim 2^{N/2}$, which is true for the small system sizes we have considered in our simulations and would complicate our analysis.

4.3 Metrics

To study the behavior of the above model, we need to define some reasonable metrics to quantify the behavior. The following are the quantities we define for this purpose.

Energy gap: For a classical Hopfield network which is able to successfully retrieve the patterns, we expect the K patterns to be the states with the lowest energy, with an energy gap between those states and the other states of the system. Generalizing this for the quantum network being considered here, we may expect the K lowest energy eigenstates to be associated with the patterns and have significantly lower energy than the rest of the spectrum. Therefore, it is worthwhile to study the spectrum of eigenvalues of the Hamiltonian, and in particular the energy gap $G_K = E_{K+1} - E_K$. The gap may be larger when we have a small number of patterns (and the system retrieves them successfully), and decrease for larger K.

Projection onto memory subspace: We expect the lowest K eigenstates of the Hamiltonian to have a large projection onto the subspace spanned by the patterns (i.e., they can be expressed as linear combinations of the patterns even if not the patterns themselves). To measure this, we define the quantity

$$p[|\psi\rangle] = \langle\psi|P_{\text{patterns}}|\psi\rangle\,, \tag{4.5}$$

where P_{patterns} is the projection operator onto the subspace spanned by the patterns $|\boldsymbol{\xi}^{\mu}\rangle$, and $p[\psi]$ is the norm of this projection for any state vector $|\psi\rangle$. If we define the column vector $\mathbf{c}[|\psi\rangle]$ of the overlaps of the state $|\psi\rangle$ with the patterns, $c_{\mu} = \langle\boldsymbol{\xi}^{\mu}|\psi\rangle$, and the matrix $\mathbf{O}$ of the pairwise overlaps of the patterns, $O_{\mu\nu} = \langle\boldsymbol{\xi}^{\mu}|\boldsymbol{\xi}^{\nu}\rangle$, then the projection can be calculated as

$$p[|\psi\rangle] = \mathbf{c}^{\dagger}\mathbf{O}^{-1}\mathbf{c}, \tag{4.6}$$

which is well-defined as long as $\mathbf{O}$ is invertible. If the patterns are orthonormal (which is usually not the case), $O_{\mu\nu} = \delta_{\mu\nu}$ and the above reduces to $\sum_{\mu}|c_{\mu}|^2$.

If we find the norm of the projection for the energy eigenvalues, we expect it to be large for the lowest K eigenvectors (which correspond to the patterns) and drop for the higher eigenvectors. The average of this quantity for the lowest K eigenvectors is expected to be large for sufficiently small K, and get smaller for larger K (when the number of patterns exceeds the “capacity” of thid quantum network)

Local stability: The dynamics of the classical Hopfield network are usually modeled by the flips of individual Ising spins σ_i due to their local field h_i (arising from the interactions between that spin and the other spins). The patterns are expected to be stable under such dynamics. For the model being considered here, the Hamiltonian is

$$H_n = -\sum_{\mu} \sum_{1 \leq i_1 < \cdots < i_n \leq N} \prod_{a=1}^{n} \left(\vec{\xi}_{i_a}^{\mu} \cdot \vec{\sigma}_{i_a} \right) = -\frac{1}{n} \sum_{i} \vec{h}_i \cdot \vec{\sigma}_i, \tag{4.7}$$

where the local field $\vec{h}_i$ acting on the qubit $\vec{\sigma}_i$ is

$$\vec{h}_i = \sum_{\mu} \sum_{\substack{1 \leq i_2 < \cdots < i_{n-1} \leq N \\ i_a \neq i \forall i}} \prod_{a=1}^{n-1} \left(\vec{\xi}_{i_a}^{\mu} \cdot \vec{\sigma}_{i_a} \right) \vec{\xi}_i^{\mu}. \tag{4.8}$$

If a states is stable to local spin flips, then we expect each of its component qubit Bloch vectors to point along their corresponding local fields. To quantify this, we define

$$s[|\psi\rangle] = \frac{1}{N} \sum_{i} \frac{\langle \psi | \vec{h}_i \cdot \vec{\sigma}_i | \psi \rangle}{\sqrt{\langle \psi | \left| \vec{h}_i \right|^2 | \psi \rangle}}. \tag{4.9}$$

We expect this quantity to be larger for the eigenstates with smaller energy, and in particular it may differ in behavior between the K states with the lowest energy (which may correspond to the patterns) and the remaining states.

4.4 Results from simulations

We considered networks with $N = 10$ qubits and $n = 3$-body interactions. For each of these cases, we numerically simulated networks with $K = 5$ and $K = 25$ patterns to observe the behavior of the networks for small and large numbers of patterns, and for $\theta_c = \pi/16$ and $\theta_c = \pi/2$ to analyze the behavior close to the classical limit and for the fully quantum model. For each network, we plotted the metrics defined in the above section for eigenvectors corresponding to the lowest subset of the eigenvalue spectrum.

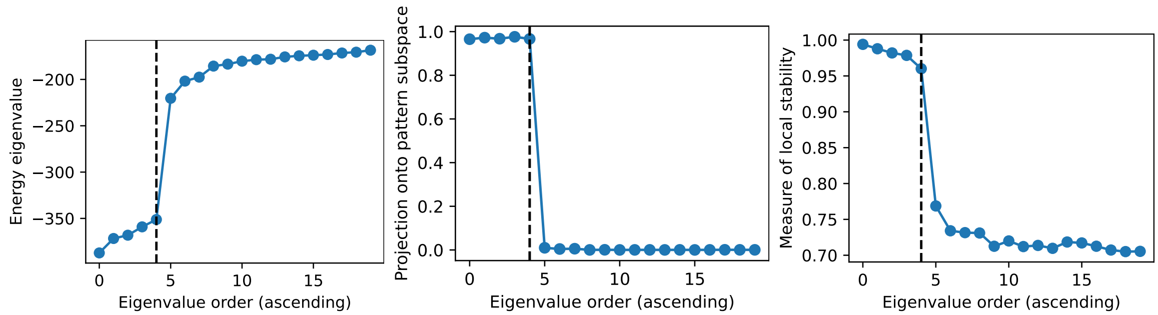

Figure 4.2: Plots of the lowest 20 eigenvalues, as well as the projections on to the pattern subspace and the local stability measures of the corresponding eigenstates, for quantum modern Hopfield networks with $N = 10$, $n = 3$, $K = 5$, and $\theta_c = \pi/2$. The dashed vertical line corresponds to the K-th eigenstate

4.4.1 Quantum models ($\theta_c = \pi/2$)

Figure 4.2 shows the metrics for a network with $K = 5$ patterns with $\theta_c = \pi/2$. In this case, we observe that the lowest K eigenstates have significantly lower energy and higher values of the other metrics compared to the remaining eigenstates, exhibiting a gap.

Figure 4.2 shows the metrics for a network with $K = 25$ patterns with $\theta_c = \pi/2$. The norm of the projection onto the pattern subspace still shows a gap between the lowest K eigenstates and the remaining eigenstates. However, the gap is smaller compared to the network with 5 memories, as the low energy eigenstates have high but not perfect projections onto the subspace, and some of the slightly higher eigenstates have small but nonzero projections onto the subspace. The Hamiltonian eigenvalues and the local stability measures have very small gapes, though the latter appears to have a change in slope at eigenvalue order K.

Figure 4.2 shows the metrics for a network with $K = 45$ patterns with $\theta_c = \pi/2$. In this case the number of patterns is large enough that none of the metrics exhibit a gap.

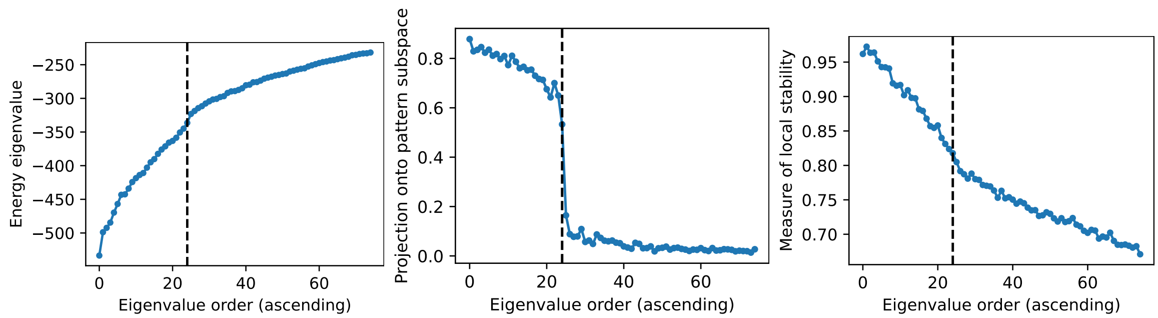

Figure 4.3: Plots of the lowest 75 eigenvalues, as well as the projections on to the pattern subspace and the local stability measures of the corresponding eigenstates, for quantum modern Hopfield networks with $N = 10$, $n = 3$, $K = 25$, and $\theta_c = \pi/2$. The dashed vertical line corresponds to the K-th eigenstate

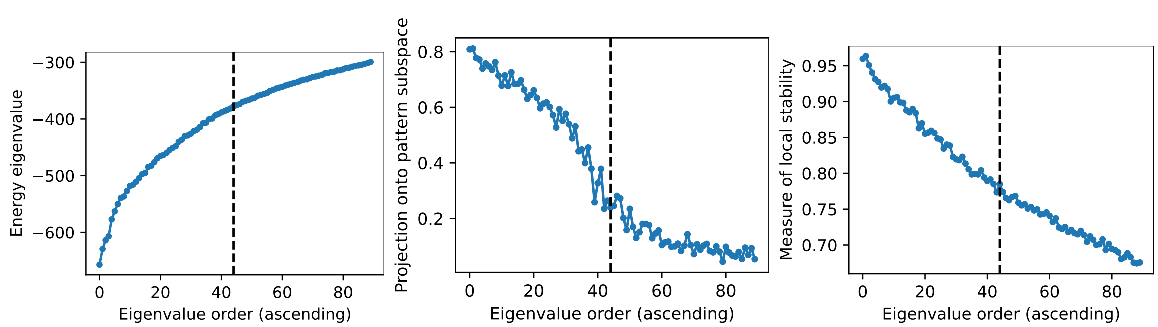

Figure 4.4: Plots of the lowest 90 eigenvalues, as well as the projections on to the pattern subspace and the local stability measures of the corresponding eigenstates, for quantum modern Hopfield networks with $N = 10$, $n = 3$, $K = 45$, and $\theta_c = \pi/2$. The dashed vertical line corresponds to the K-th eigenstate

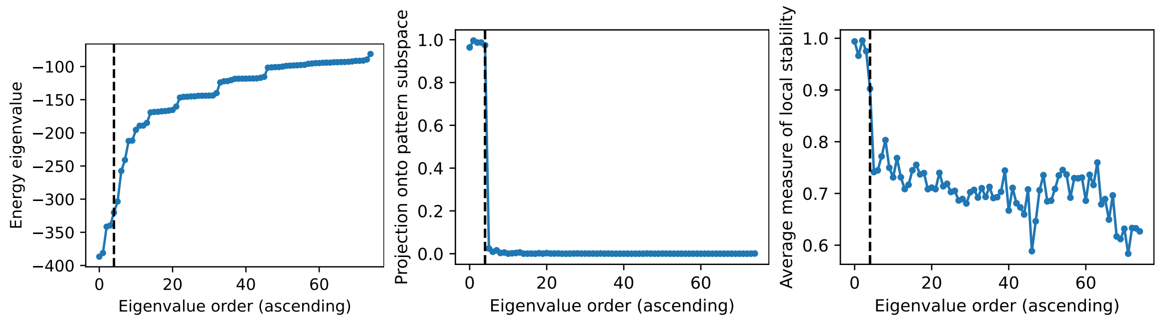

Figure 4.5: Plots of the lowest 75 eigenvalues, as well as the projections on to the pattern subspace and the local stability measures of the corresponding eigenstates, for quantum modern Hopfield networks with $N = 10$, $n = 3$, $K = 5$, and $\theta_c = \pi/16$. The dashed vertical line corresponds to the K-th eigenstate

4.4.2 Approximately classical models ($\theta_c = \pi/16$)

Figure 4.2 shows the metrics for a network with $K = 5$ patterns with $\theta_c = \pi/16$, corresponding to an approximately classical model. In this case, even for the small value of K, there is no significant energy. However, the norm of the projection onto the pattern subspace shows a significant gap between the lowest K eigenstates and the remaining eigenstates. The local stability is also high for the lowest K eigenstates and has a sharp decrease afterwards, which implies that the lowest eigenstates are linear combinations of the pattern states and are stable to rotations of individual spins. The local stability measure also exhibits much higher fluctuation in this model; the plot shown here was obtained by averaging over 10 randomly generated realizations of the network (unlike the other plots which are for one network realization each with no averaging).

Figure 4.3 shows the metrics for a network with $K = 25$ patterns with $\theta_c = \pi/16$. The eigenvalue spectrum appears to have a different distribution compared to that of the model with 5 embedded patterns. Both the projection and the local stability have high fluctuations for this network. Averaging over 10 spin realizations, we find that neither of those metrics shows a significant gap between the lowest K and the remaining eigenstates. This is different from the corresponding network with $\theta_c = \pi/16$, where gaps were observed. This indicates that, at least in terms of these metrics, the quantum model with $\theta_c = \pi/2$ outperforms the

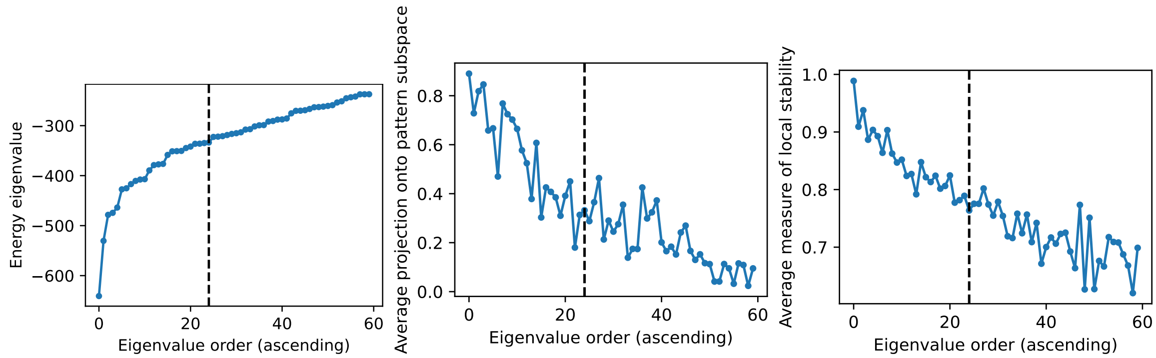

Figure 4.6: Plots of the lowest 60 eigenvalues, as well as the projections on to the pattern subspace and the local stability measures of the corresponding eigenstates, for quantum modern Hopfield networks with $N = 10$, $n = 3$, $K = 25$, and $\theta_c = \pi/16$. The dashed vertical line corresponds to the K-th eigenstate

approximately classical model.

4.5 Discussions and outlook

We explored the behavior of the quantum generalization of the modern Hopfield network, described as a system of N interacting qubits. The metrics we measured in simulated models show that the low-energy states of this network are linear combinations of the embedded patterns and stable to local spin rotations (indicating the patterns being "retrieved") for sufficiently small number of patterns K but less clearly so when K is large, which agrees with the behavior of classical Hopfield networks. This behavior also depends on how the pattern Bloch vectors are chosen, with "quantum" models (where all Bloch vectors are uniformly probably) performing better (according to the metrics shown above) compared to approximately "classical" models where the pattern Bloch vectors are close to σ_z eigenstates.

Further work needs to be done to quantify these results. In particular, we need to study the metrics (such as the energy gap and the average norm of the projection onto the pattern subspace) as functions of the number of patterns K and ascertain how they are related. This needs to be done for multiple system sizes N to check the scaling behavior of K; by analogy with classical networks, we may expect the retrieval metrics (for n-spin interactions) to depend on the number of patterns as K/N^{n-1}. It will also be interesting to consider the dynamics of such networks under a suitable master equation.

CHAPTER 5

CONCLUSIONS, DISCUSSIONS, AND OUTLOOK

In this thesis, I presented my work on the modern Hopfield network, which is a system of Ising spins with many-body interactions. I investigated the network with n-spin interactions, in particular in presence of noise and/or clipping of the interaction coefficients. We found that the memory capacity of such networks is robust to additive noise, and its scaling behavior with system size is not affected by multiplicative noise or by random deletion or clipping of interactions. We carried out our analysis for a single step of synchronous update of the component spins and obtained some numerical results for multi-step dynamics. A future direction of this research can be to investigate the network under other possibly more physically realistic kinds of dynamics, such as asynchronous dynamics in presence of fast noise.

I described a technique for obtaining the phase diagram of the modern Hopfield network with a quartic Hamiltonian using replica theory. The qualitative behavior of the phase diagram agrees with those of similar networks found previously (also using replica theory but with different approximation methods). It is currently unclear if this technique has any advantages over the other methods for investigating modern Hopfield networks found in the literature, but it may be worth exploring this method further, such as to investigate other systems.

I also presented our current work on quantum Hopfield networks, which is a system of qubits rather than classical Ising spins, and where the patterns are general separable states. I presented some preliminary results related to these models, where the quantum network appears to be able to retrieve patterns for larger numbers of embedded patterns compared to the corresponding classical network, at least based on some metrics. Much more work needs to be done, however, before we can make a conclusive statement. This includes analyzing the metrics over a large distribution of system sizes and numbers of patterns, as well as obtaining some analytical estimates.

REFERENCES

[1] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–2558, April 1982. ISSN 0027-8424. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC346238/`.

[2] D. O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Psychology Press, New York, April 2005. ISBN 978-1-4106-1240-3. doi:10.4324/9781410612403.

[3] Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky. Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks. *Physical Review Letters*, 55(14): 1530–1533, September 1985. ISSN 0031-9007. doi:10.1103/PhysRevLett.55.1530. URL `https://link.aps.org/doi/10.1103/PhysRevLett.55.1530`.

[4] R. McEliece, E. Posner, E. Rodemich, and S. Venkatesh. The capacity of the Hopfield associative memory. *IEEE Transactions on Information Theory*, 33(4):461–482, July 1987. ISSN 1557-9654. doi:10.1109/TIT.1987.1057328. URL `https://ieeexplore.ieee.org/document/1057328`. Conference Name: IEEE Transactions on Information Theory.

[5] L. Personnaz, I. Guyon, and G. Dreyfus. Collective computational properties of neural networks: New learning mechanisms. *Physical Review A*, 34(5):4217–4228, November 1986. doi:10.1103/PhysRevA.34.4217. URL `https://link.aps.org/doi/10.1103/PhysRevA.34.4217`. Publisher: American Physical Society.

[6] I. Kanter and H. Sompolinsky. Associative recall of memory without errors. *Physical Review A*, 35(1):380–392, January 1987. doi:10.1103/PhysRevA.35.380. URL `https://link.aps.org/doi/10.1103/PhysRevA.35.380`. Publisher: American Physical Society.

[7] Amos Storkey. Increasing the capacity of a hopfield network without sacrificing functionality. In Wulfram Gerstner, Alain Germond, Martin Hasler, and Jean-Daniel Nicoud, editors, *Artificial Neural Networks — ICANN'97*, pages 451–456, Berlin, Heidelberg, 1997. Springer. ISBN 978-3-540-69620-9. doi:10.1007/BFb0020196.

[8] Dmitry Krotov and John J. Hopfield. Dense Associative Memory for Pattern Recognition, September 2016. URL `http://arxiv.org/abs/1606.01164`. arXiv:1606.01164 [cond-mat, q-bio, stat].

[9] Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, Victor Greiff, David Kreil, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield Networks is All You Need, April 2021. URL `http://arxiv.org/abs/2008.02217`. arXiv:2008.02217 [cs, stat].

[10] Mete Demircigil, Judith Heusel, Matthias Löwe, Sven Upgang, and Franck Vermet. On a model of associative memory with huge storage capacity. *Journal of Statistical Physics*, 168(2):288–299, July 2017. ISSN 0022-4715, 1572-9613. doi:10.1007/s10955-017-1806-y. URL `http://arxiv.org/abs/1702.01929`. arXiv:1702.01929 [math].

[11] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference, December 2017. URL `http://arxiv.org/abs/1712.05877`. arXiv:1712.05877 [cs].

[12] H. Sompolinsky. The theory of neural networks: The Hebb rule and beyond. In J. L. van Hemmen and I. Morgenstern, editors, *Heidelberg Colloquium on Glassy Dynamics*, pages 485–527, Berlin, Heidelberg, 1987. Springer. ISBN 978-3-540-47819-5. doi:10.1007/BFb0057531.

[13] J. A. Hertz, G. Grinstein, and S. A. Solla. Irreversible spin glasses and neural networks. In J. L. van Hemmen and I. Morgenstern, editors, *Heidelberg Colloquium on Glassy Dynamics*, pages 538–546, Berlin, Heidelberg, 1987. Springer. ISBN 978-3-540-47819-5. doi:10.1007/BFb0057533.

[14] B. Derrida, E. Gardner, and A. Zippelius. An Exactly Solvable Asymmetric Neural Network Model. *Europhysics Letters*, 4(2):167, July 1987. ISSN 0295-5075. doi:10.1209/0295-5075/4/2/007. URL `https://iopscience-iop-org.proxy.uchicago.edu/article/10.1209/0295-5075/4/2/007/meta`. Publisher: IOP Publishing.

[15] A. Treves and D. J. Amit. Metastable states in asymmetrically diluted Hopfield networks. *Journal of Physics A: Mathematical and General*, 21(14):3155, July 1988. ISSN 0305-4470. doi:10.1088/0305-4470/21/14/016. URL `https://iopscience-iop-org.proxy.uchicago.edu/article/10.1088/0305-4470/21/14/016/meta`. Publisher: IOP Publishing.

[16] Elena Agliari and Giordano De Marzo. Tolerance versus synaptic noise in dense associative memories. *The European Physical Journal Plus*, 135(11):883, November 2020. ISSN 2190-5444. doi:10.1140/epjp/s13360-020-00894-8. URL `https://doi.org/10.1140/epjp/s13360-020-00894-8`.

[17] Vedran Dunjko and Hans J Briegel. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, 81(7):074001, June 2018. ISSN 0034-4885. doi:10.1088/1361-6633/aab406. URL `https://dx.doi.org/10.1088/1361-6633/aab406`. Publisher: IOP Publishing.

[18] Patrick Rebentrost, Thomas R. Bromley, Christian Weedbrook, and Seth Lloyd. Quantum Hopfield neural network. *Physical Review A*, 98(4):042308, October 2018. ISSN 2469-9926, 2469-9934. doi:10.1103/PhysRevA.98.042308. URL `http://arxiv.org/abs/1710.03599`. arXiv:1710.03599 [quant-ph].

[19] Yu-Qiang Ma, Yue-Ming Zhang, and C. D. Gong. A Hopfield Neural Network Model in a Transverse Field. *Communications in Theoretical Physics*, 21(2):253, March 1994. ISSN 0253-6102. doi:10.1088/0253-6102/21/2/253. URL https://iopscience-iop-org.proxy.uchicago.edu/article/10.1088/0253-6102/21/2/253/meta. Publisher: IOP Publishing.

[20] Yu-qiang Ma and Chang-de Gong. Statistical mechanics of a multiconnected Hopfield neural-network model in a transverse field. *Physical Review E*, 51(2):1573–1575, February 1995. doi:10.1103/PhysRevE.51.1573. URL https://link.aps.org/doi/10.1103/PhysRevE.51.1573. Publisher: American Physical Society.

[21] Hidetoshi Nishimori and Yoshihiko Nonomura. Quantum Effects in Neural Networks. *Journal of the Physical Society of Japan*, 65(12):3780–3796, December 1996. ISSN 0031-9015. doi:10.1143/JPSJ.65.3780. URL https://journals.jps.jp/doi/10.1143/JPSJ.65.3780. Publisher: The Physical Society of Japan.

[22] P Rotondo, M Marcuzzi, J P Garrahan, I Lesanovsky, and M Müller. Open quantum generalisation of Hopfield neural networks. *Journal of Physics A: Mathematical and Theoretical*, 51(11):115301, March 2018. ISSN 1751-8113, 1751-8121. doi:10.1088/1751-8121/aaabcb. URL https://iopscience.iop.org/article/10.1088/1751-8121/aaabcb.

[23] Masha Shcherbina, Brunello Tirozzi, and Camillo Tassi. Quantum Hopfield Model. *Physics*, 2(2):184–196, June 2020. ISSN 2624-8174. doi:10.3390/physics2020012. URL https://www.mdpi.com/2624-8174/2/2/12. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[24] Lukas Bödeker, Eliana Fiorelli, and Markus Müller. Optimal storage capacity of quantum Hopfield neural networks. *Physical Review Research*, 5(2):023074, May 2023. ISSN 2643-1564. doi:10.1103/PhysRevResearch.5.023074. URL https://link.aps.org/doi/10.1103/PhysRevResearch.5.023074.

[25] Joaquín J. Torres and Daniel Manzano. Dissipative Quantum Hopfield Network: A numerical analysis. *New Journal of Physics*, 26(10):103018, October 2024. ISSN 1367-2630. doi:10.1088/1367-2630/ad5e15. URL http://arxiv.org/abs/2305.02681. arXiv:2305.02681 [quant-ph].

[26] Koki Okajima and Yoshiyuki Kabashima. Exact Replica Symmetric solution for transverse field Hopfield model under finite Trotter size. *Journal of the Physical Society of Japan*, 94(4):044005, April 2025. ISSN 0031-9015, 1347-4073. doi:10.7566/JPSJ.94.044005. URL http://arxiv.org/abs/2411.02012. arXiv:2411.02012 [cond-mat].

[27] Rongfeng Xie and Alex Kamenev. Quantum Hopfield model with dilute memories. *Physical Review A*, 110(3):032418, September 2024. doi:10.1103/PhysRevA.110.032418. URL https://link.aps.org/doi/10.1103/PhysRevA.110.032418. Publisher: American Physical Society.

[28] Takeshi Kimura and Kohtaro Kato. Analysis of Discrete Modern Hopfield Networks in Open Quantum System, November 2024. URL http://arxiv.org/abs/2411.02883. arXiv:2411.02883 [quant-ph].

[29] Yuya Seki and Hidetoshi Nishimori. Quantum annealing with antiferromagnetic transverse interactions for the Hopfield model. *Journal of Physics A: Mathematical and Theoretical*, 48(33):335301, July 2015. ISSN 1751-8121. doi:10.1088/1751-8113/48/33/335301. URL https://dx.doi.org/10.1088/1751-8113/48/33/335301. Publisher: IOP Publishing.

[30] Sharba Bhattacharjee and Ivar Martin. Accuracy and capacity of Modern Hopfield networks with synaptic noise, February 2025. URL http://arxiv.org/abs/2503.00241. arXiv:2503.00241 [cond-mat].

[31] Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky. Spin-glass models of neural networks. *Physical Review A*, 32(2):1007–1018, August 1985. ISSN 0556-2791. doi:10.1103/PhysRevA.32.1007. URL https://link.aps.org/doi/10.1103/PhysRevA.32.1007.

[32] N. Hendrich. A scalable architecture for binary couplings attractor neural networks. In *Proceedings of Fifth International Conference on Microelectronics for Neural Networks*, pages 213–220, February 1996. doi:10.1109/MNNFS.1996.493793. URL https://ieeexplore-ieee-org.proxy.uchicago.edu/document/493793/?arnumber=493793. ISSN: 1086-1947.

[33] Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, and Benjamin L. Lev. Enhancing associative memory recall and storage capacity using confocal cavity QED. *Physical Review X*, 11(2): 021048, June 2021. ISSN 2160-3308. doi:10.1103/PhysRevX.11.021048. URL http://arxiv.org/abs/2009.01227. arXiv:2009.01227 [cond-mat, physics:quant-ph].

[34] Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky. Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5):2293–2303, March 1987. ISSN 0556-2791. doi:10.1103/PhysRevA.35.2293. URL https://link.aps.org/doi/10.1103/PhysRevA.35.2293.

[35] A. Crisanti, D. J. Amit, and H. Gutfreund. Saturation Level of the Hopfield Model for Neural Network. *Europhysics Letters*, 2(4):337, August 1986. ISSN 0295-5075. doi:10.1209/0295-5075/2/4/012. URL https://dx.doi.org/10.1209/0295-5075/2/4/012.

[36] M. V. Feigelman and L. B. Ioffe. The Statistical Properties of the Hopfield Model of Memory. *Europhysics Letters*, 1(4):197, February 1986. ISSN 0295-5075. doi:10.1209/0295-5075/1/4/007. URL https://iopscience-iop-org.proxy.uchicago.edu/article/10.1209/0295-5075/1/4/007/meta. Publisher: IOP Publishing.

[37] Y. Abu-Mostafa and J. St. Jacques. Information capacity of the Hopfield model. *IEEE Transactions on Information Theory*, 31(4):461–464, July 1985. ISSN 0018-9448. doi:10.1109/TIT.1985.1057069. URL http://ieeexplore.ieee.org/document/1057069/.

[38] T. R. Kirkpatrick and D. Thirumalai. p -spin-interaction spin-glass models: Connections with the structural glass problem. *Physical Review B*, 36(10):5388–5397, October 1987. ISSN 0163-1829. doi:10.1103/PhysRevB.36.5388. URL https://link.aps.org/doi/10.1103/PhysRevB.36.5388.

[39] E. Gardner. Spin glasses with p-spin interactions. *Nuclear Physics B*, 257:747–765, January 1985. ISSN 0550-3213. doi:10.1016/0550-3213(85)90374-8. URL https://www.sciencedirect.com/science/article/pii/0550321385903748.

[40] A. Crisanti and H. J. Sommers. The sphericalp-spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter*, 87(3):341–354, October 1992. ISSN 1431-584X. doi:10.1007/BF01309287. URL https://doi.org/10.1007/BF01309287.

[41] A. Crisanti, H. Horner, and H. J. Sommers. The sphericalp-spin interaction spin-glass model. *Zeitschrift für Physik B Condensed Matter*, 92(2):257–271, June 1993. ISSN 1431-584X. doi:10.1007/BF01312184. URL https://doi.org/10.1007/BF01312184.

[42] Nicolas Sourlas. Spin-glass models as error-correcting codes. *Nature*, 339(6227):693–695, June 1989. ISSN 1476-4687. doi:10.1038/339693a0. URL https://www.nature.com/articles/339693a0. Publisher: Nature Publishing Group.

[43] Jeferson J. Arenzon and Rita M. C. De Almeida. Neural networks with high-order connections. *Physical Review E*, 48(5):4060–4069, November 1993. ISSN 1063-651X, 1095-3787. doi:10.1103/PhysRevE.48.4060. URL https://link.aps.org/doi/10.1103/PhysRevE.48.4060.

[44] Ney Lemke, Jeferson J. Arenzon, and Francisco A. Tamarit. Chaotic dynamics of high-order neural networks. *Journal of Statistical Physics*, 79(1-2):415–427, April 1995. ISSN 0022-4715, 1572-9613. doi:10.1007/BF02179396. URL http://link.springer.com/10.1007/BF02179396.

[45] Anton Bovier and Beat Niederhauser. The spin-glass phase-transition in the Hopfield model with \$p\$-spin interactions. *Advances in Theoretical and Mathematical Physics*, 5 (6):1001–1046, 2001. ISSN 1095-0753. doi:10.4310/ATMP.2001.v5.n6.a2. URL https://link.intlpress.com/JDetail/1805562733108076546. Publisher: International Press of Boston.

[46] Pierre Baldi and Santosh S. Venkatesh. Number of stable points for spin-glasses and neural networks of higher orders. *Physical Review Letters*, 58(9):913–916, March 1987. ISSN 0031-9007. doi:10.1103/PhysRevLett.58.913. URL https://link.aps.org/doi/10.1103/PhysRevLett.58.913.

[47] Dmitry Krotov and John J. Hopfield. Dense Associative Memory for Pattern Recognition. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/hash/eaae339c4d89fc102edd9dbdb6a28915-Abstract.html.

[48] Han Bao, Richong Zhang, and Yongyi Mao. The capacity of the dense associative memory networks. *Neurocomputing*, 469:198–208, January 2022. ISSN 0925-2312. doi:10.1016/j.neucom.2021.10.058. URL https://www.sciencedirect.com/science/article/pii/S0925231221015447.

[49] Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Thomas Adler, David Kreil, Michael K. Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield Networks is All You Need. In *Advances in Neural Information Processing Systems*, October 2020. URL https://openreview.net/forum?id=tL89RnzIiCd.

[50] Carlo Lucibello and Marc Mézard. Exponential Capacity of Dense Associative Memories. *Physical Review Letters*, 132(7):077301, February 2024. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.132.077301. URL https://link.aps.org/doi/10.1103/PhysRevLett.132.077301.

[51] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[52] Toshihiro Ota and Ryo Karakida. Attention in a family of Boltzmann machines emerging from modern Hopfield networks, March 2023. URL http://arxiv.org/abs/2212.04692. arXiv:2212.04692 [cs].

[53] Matthew Smart, Alberto Bietti, and Anirvan M. Sengupta. In-context denoising with one-layer transformers: connections between attention and associative memory retrieval, February 2025. URL http://arxiv.org/abs/2502.05164. arXiv:2502.05164 [cs].

[54] Hamza Tahir Chaudhry, Jacob A Zavatone-Veth, Dmitry Krotov, and Cengiz Pehlevan. Long sequence Hopfield memory. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(10):104024, October 2024. ISSN 1742-5468. doi:10.1088/1742-5468/ad6427. URL https://dx.doi.org/10.1088/1742-5468/ad6427. Publisher: IOP Publishing.

[55] Dennis Wu, Jerry Yao-Chieh Hu, Teng-Yun Hsiao, and Han Liu. Uniform Memory Retrieval with Larger Capacity for Modern Hopfield Models. In *Proceedings of the 41st International Conference on Machine Learning*, pages 53471–53514. PMLR, July 2024. URL https://proceedings.mlr.press/v235/wu24i.html. ISSN: 2640-3498.

[56] Jerry Yao-Chieh Hu, Dennis Wu, and Han Liu. Provably Optimal Memory Capacity for Modern Hopfield Models: Transformer-Compatible Dense Associative Memories as

Spherical Codes. In *Advances in Neural Information Processing Systems*, volume 37, pages 70693–70729. Curran Associates, Inc., November 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/82846e19e6d42ebfd4ace4361def29ae-Paper-Conference.pdf.

[57] Benjamin Hoover, Yuchen Liang, Bao Pham, Rameswar Panda, Hendrik Strobelt, Duen Horng Chau, Mohammed Zaki, and Dmitry Krotov. Energy Transformer. *Advances in Neural Information Processing Systems*, 36:27532–27559, December 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/57a9b97477b67936298489e3c1417b0a-Abstract-Conference.html.

[58] Yann LeCun, John Denker, and Sara Solla. Optimal Brain Damage. In *Advances in Neural Information Processing Systems*, volume 2. Morgan-Kaufmann, 1989. URL https://proceedings.neurips.cc/paper_files/paper/1989/hash/6c9882bbac1c7093bd25041881277658-Abstract.html.

[59] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. ISSN 1476-4687. doi:10.1038/s41586-020-2649-2. URL https://www.nature.com/articles/s41586-020-2649-2. Publisher: Nature Publishing Group.

[60] John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, May 2007. ISSN 1558-366X. doi:10.1109/MCSE.2007.55. URL https://ieeexplore.ieee.org/document/4160265. Conference Name: Computing in Science & Engineering.

[61] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3): 261–272, March 2020. ISSN 1548-7105. doi:10.1038/s41592-019-0686-2. URL https://www.nature.com/articles/s41592-019-0686-2. Publisher: Nature Publishing Group.

[62] Benjamin Hoover, Duen Horng Chau, Hendrik Strobelt, Parikshit Ram, and Dmitry Krotov. Dense Associative Memory Through the Lens of Random Features. *Advances in*

Neural Information Processing Systems, 37:23549–23576, December 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/29ff36c8fbed10819b2e50267862a52a-Abstract-Conference.html.

[63] Linda Albanese, Francesco Alemanno, Andrea Alessandrelli, and Adriano Barra. Replica Symmetry Breaking in Dense Hebbian Neural Networks. *Journal of Statistical Physics*, 189:24, November 2022. ISSN 0022-4715. doi:10.1007/s10955-022-02966-8. URL https://ui.adsabs.harvard.edu/abs/2022JSP...189...24A. ADS Bibcode: 2022JSP...189...24A.

[64] Elena Agliari, Francesco Alemanno, Adriano Barra, Martino Centonze, and Alberto Fachechi. Neural networks with redundant representation: detecting the undetectable. *Physical Review Letters*, 124(2):028301, January 2020. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.124.028301. URL http://arxiv.org/abs/1911.12689. arXiv:1911.12689 [cond-mat].

[65] Francesco Alemanno, Martino Centonze, and Alberto Fachechi. Interpolating between boolean and extremely high noisy patterns through minimal dense associative memories. *Journal of Physics A: Mathematical and Theoretical*, 53(7):074001, February 2020. ISSN 1751-8113, 1751-8121. doi:10.1088/1751-8121/ab6943. URL https://iopscience.iop.org/article/10.1088/1751-8121/ab6943.

[66] Richard Barney, Michael Winer, and Victor Galitski. Neural Networks as Spin Models: From Glass to Hidden Order Through Training, August 2024. URL http://arxiv.org/abs/2408.06421. arXiv:2408.06421 [cond-mat, physics:nlin].

[67] Agnish Kumar Behera, Madan Rao, Srikanth Sastry, and Suriyanarayanan Vaikuntanathan. Enhanced Associative Memory, Classification, and Learning with Active Dynamics. *Physical Review X*, 13(4):041043, December 2023. ISSN 2160-3308. doi:10.1103/PhysRevX.13.041043. URL https://link.aps.org/doi/10.1103/PhysRevX.13.041043.

[68] Agnish Kumar Behera, Matthew Du, Uday Jagadisan, Srikanth Sastry, Madan Rao, and Suriyanarayanan Vaikuntanathan. A correspondence between Hebbian unlearning and steady states generated by nonequilibrium dynamics, October 2024. URL http://arxiv.org/abs/2410.06269. arXiv:2410.06269 [cond-mat].

[69] Giordano De Marzo and Giulio Iannelli. Effect of spatial correlations on Hopfield Neural Network and Dense Associative Memories, July 2022. URL http://arxiv.org/abs/2207.05218. arXiv:2207.05218.

[70] Thomas F. Burns. Semantically-correlated memories in a dense associative model, June 2024. URL http://arxiv.org/abs/2404.07123. arXiv:2404.07123 [cs].

[71] Omer Karin. EnhancerNet: a predictive model of cell identity dynamics through enhancer selection. *Development*, 151(19):dev202997, October 2024. ISSN 0950-1991. doi:10.1242/dev.202997. URL https://doi.org/10.1242/dev.202997.

[72] Anton Grishechkin, Abhirup Mukherjee, and Omer Karin. Hierarchical Control of State Transitions in Dense Associative Memories, January 2025. URL `http://arxiv.org/abs/2412.11336`. arXiv:2412.11336 [q-bio].

[73] E. Gardner. Multiconnected neural network models. *Journal of Physics A: Mathematical and General*, 20(11):3453, August 1987. ISSN 0305-4470. doi:10.1088/0305-4470/20/11/046. URL `https://dx.doi.org/10.1088/0305-4470/20/11/046`.

[74] Elena Agliari, Linda Albanese, Francesco Alemanno, Andrea Alessandrelli, Adriano Barra, Fosca Giannotti, Daniele Lotito, and Dino Pedreschi. Dense Hebbian neural networks: A replica symmetric picture of unsupervised learning. *Physica A: Statistical Mechanics and its Applications*, 627:129143, October 2023. ISSN 0378-4371. doi:10.1016/j.physa.2023.129143. URL `https://www.sciencedirect.com/science/article/pii/S0378437123006982`.

[75] Linda Albanese, Andrea Alessandrelli, Alessia Annibale, and Adriano Barra. Unsupervised and Supervised learning by Dense Associative Memory under replica symmetry breaking, December 2023. URL `http://arxiv.org/abs/2312.09638`. arXiv:2312.09638 [cond-mat].

[76] Nicholas Meinhardt, Niels M. P. Neumann, and Frank Phillipson. Quantum Hopfield Neural Networks: A New Approach and Its Storage Capacity. *Computational Science – ICCS 2020*, 12142:576–590, May 2020. doi:10.1007/978-3-030-50433-5_44. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7304743/`.

[77] Ge Liu, Wen-Ping Ma, Hao Cao, and Liang-Dong Lyu. A quantum Hopfield neural network model and image recognition. *Laser Physics Letters*, 17(4):045201, February 2020. ISSN 1612-202X. doi:10.1088/1612-202X/ab7347. URL `https://dx.doi.org/10.1088/1612-202X/ab7347`. Publisher: IOP Publishing.

[78] Nathan Eli Miller and Saibal Mukhopadhyay. A quantum Hopfield associative memory implemented on an actual quantum processor. *Scientific Reports*, 11(1):23391, December 2021. ISSN 2045-2322. doi:10.1038/s41598-021-02866-z. URL `https://www.nature.com/articles/s41598-021-02866-z`. Publisher: Nature Publishing Group.