Investigations of Decoherence in an Ensemble of Two-Level Quantum Systems Coupled to a Boson Bath in Vacuum

Ben Lemberger

A dissertation subsmitted in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Physics)

at the UNIVERSITY OF WISCONSIN, MADISON
2019

Date of final oral examination: 8/5/2019

The dissertation is approved by the following members of the Final Oral Committee:
Deniz Yavuz, Professor, Physics
Mark Saffman, Professor, Physics
Thad Walker, Professor, Physics
Mikhail Kats, Professor, Electrical Engineering

Acknowledgements

First I'd like to thank Deniz, who had the original ideas behind all of the work presented here. It was under his instruction and guidance that I was able to follow through with these investigations, and he taught me a great deal of physics through our conversations. He demonstrated by example how to think about problems and determine good questions to ask, and I am thankful to have learned from someone with such a keen sense for this.

I would also like to thank Jeff Schmidt and Jim Reardon on staff at UW Madison, who both greatly influenced my time here. Their appreciation for physics is contagious, and their perspectives are ones that I feel lucky to have encountered.

To my committee as well, I am grateful for the time they took to read this thesis and the valuable expertise they will bring in evaluating my work.

I must acknowledge my fellow grad students, whose camaraderie made grad school possible and fun. I learned a lot from them during our classes and in our many discussions in and out of the physics building. There are some additional members of the UW community outside of the physics department who greatly impacted by time here in a positive way, most notably Lauren, Jason, Aaron, and Kai.

There are many people at Oberlin College, my undergraduate institution, to whom I owe my understanding of physics such as it is. The faculty there gave me a firm foundation without which I would not have made it very far. They nurtured my interests and gave me the idea that I might pursue them in graduate school. They prepared me well for it and in this moment in particular, as I complete my dissertation work, I am grateful for that.

The students at Oberlin also formed a large part of my physics education as we did problem sets together, studied for tests, and argued about confusing new things like

quantum mechanics. Their continued contact and support is a constant motivation as a physicist and a person, and they are owed much credit for this work.

Finally I would like to thank my family; they support me in whatever I do, including this, and that support is always invaluable. My cat Ragnar helped out a lot too, even thought he doesn't know it and I'm sure he would have preferred I spent all that time playing with him instead.

Madison, August 2019

Ben Lemberger

Abstract

We will study an ensemble of two-level quantum systems coupled to a common boson bath. We find the spectrum in the large N limit of the exchange Hamiltonian which arises from the boson bath and allows the subsystems to exchange a quanta of energy. We study the implications of this spectrum in quantum error correction, decoherence, and the measurement problem. Numerical studies support the analytic approximations and are used to approximately simulate the collective system.

Contents

List of Figures

1 Motivation From Quantum Error Correction

1.1 The Exciting Possibility of a Quantum Computer

Quantum Computers are built out of two-level quantum systems, or qubits, instead of classical bits. As shown in Fig. 1.1, we physically model a Quantum Computer as N two-level atoms, each with levels $|0\rangle$ and $|1\rangle$, in a two- or three-dimensional geometry.

They are an exciting possibility because they could potentially vastly outperform classical computers at certain tasks which are of interest both academically and practically [1–8]. For example, the best known quantum algorithms for simulating any quantum system, or finding the prime factors of large integers, is far more efficient than the best known classical algorithm for the same problem [9] [10]. The former problem would be of interest to scientists probing quantum phenomena which are difficult to access experimentally and theoretically, and the latter problem would be of interest to anyone wishing to decrypt electronic communications since most communications today are encrypted using the RSA algorithm, which relies on the difficulty of factoring large integers.

In this context, the efficiency of an algorithm means the relationship between the size of the input and how long the algorithm takes to complete for a typical input of that size. In our two example applications, the input size would be the dimension of the Hilbert space to be simulated and number of bits in the composite integer to be factored, respectively. When comparing algorithms to decide which scales more favorably with input size (or more briefly, which is more efficient) it is well accepted that the comparison should be done by taking the limit of arbitrarily large inputs. For example if the input size is characterized by the number N, and the time it

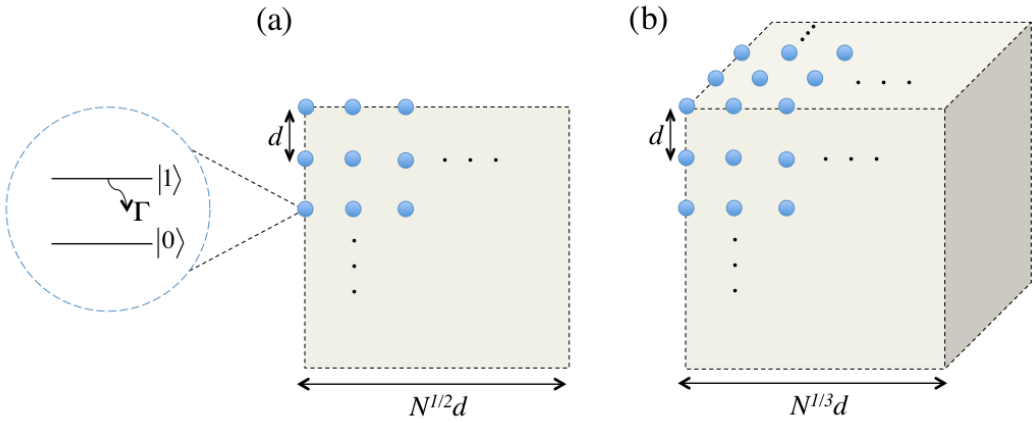

Figure 1.1: An N qubit quantum computer in (a) two and (b) three dimensional geometry with the qubits coupled to a common bosonic bath. For a single qubit present, the interaction with the bath causes an independent decay rate of Γ between the qubit levels. When the whole ensemble of qubits is present, correlated decay causes correlated errors to build up across the whole computer. Since we are primarily interested in the $N \to \infty$ limit, we do not impose restrictions on the size of the sample, i.e. the spacing between the qubits may be much larger than the radiation wavelength, $d >> \lambda_a$. For concreteness, we focus on a square and a cube arrangement of atoms with regularly spaced qubits, but the results are not sensitive to the precise shape and arrangement structure of the array.

takes an algorithm to complete goes as $aN + bN^2$, one would simply say that the algorithm scales as N^2, which is the only important term for large N. There is no guarantee that such an attitude will lead to the most efficient algorithm in practice - any real implementation of the algorithm will have a finite value of N at which the asymptotic scaling will not necessarily have taken over, especially if $a >> b$. Despite this flaw, using the large N scaling for algorithms has proven meaningful, even if no one is exactly sure why (as is demonstrated by the foundational and still-open problem of $P = NP$ [11]).

An even cruder classification which appears to be deeply meaningful is the polynomial-exponential divide. An algorithm which requires an amount of time which scales for large N as N^p for some power p is considered polynomial, and when the required time scales as e^{N^p} the algorithm is considered to be exponential. One of the main results which inspired the field of quantum computing was that of Peter Shor in 1994, which showed that a polynomial time factoring algorithm existed

for a quantum computer [12, 13]. Despite decades of searching both before and since this quantum result, no such improvement has been made upon long-known exponential-time classical algorithms. This established the idea that computers built from bits whose state lived in a quantum state space might be characteristically better than their classical counterparts, at least for some problems.

1.2 Quantum Error Correction and the Threshold Theorem

There is a complication: the classical and quantum algorithms discussed above are ideal in the sense that they assume perfect control over the bits, with no influence from its physical environment besides that of the person or machine implementing the desired algorithm. However any physical implementation of an algoithm will always have some unwanted interactions or imprecisions, even if they can be made very small. For classical computers, error correction can be implemented in such a way as to not affect the polynomial-exponential categorization of ideal algorithms; if an ideal algorithm is polynomial (exponential), then implementations of that algorithm which includes error correction can also be polynomial (exponential) [14]. This important property rests on the fact that error correction is implemented actively during the computation, as opposed to trying to make the errors so unlikely that not too many will occur during the length of the entire computation. If this latter approach were used, algorithms would become increasingly inaccurate for longer calculations. Since their scaling properties are understood for arbitrarily large inputs and therefore arbitrarily long computations, this would call into question the validity of such a limit and require algorithm efficiency to be compared at specific input sizes. We are spared such an unfortunate situation because of the ease at which active error correction can be performed classically. Although more sophisticated techniques are used, the idea can be simply demonstrated as follows.

If the ideal algorithm requires a bit in the state 1, the real algorithm encodes the state of this conceptual, or logical, bit in the state of, say, 10 different real bits

all in the state 1 (and similarly encode one logical 0 in 10 physicals 0s). The bits might randomly flip between 0 and 1 at some error rate, but if we repeatedly check the 10 real bits with a period much smaller than the inverse error rate and find, for example, all but one classical bits are in the 1 state, it is much more likely that one bit flipped than that 9 of them did. Accordingly, the error correction algorithm repairs the state by flipping the single 0 back into a 1. There is no guarantee that the error rate can be made low enough that it is possible to perform such diagnoses and corrections fast enough, but it can be shown that a threshold relationship exists between the error rate and error correction rate which guarantees that, when met, error correction can be performed such that the computation's result remains arbitrarily reliable for arbitrarily large N while only requiring the duration of the algorithm to increase by a factor polynomial in input size. Crucially, this threshold between the rate of errors and corrections is also independent of how long the computation is or the input size.

Quantum error correction is not as simple as its classical analogue due to the delicacy of quantum superpositions and the inability to copy a quantum state. Even once it was determined to be possible there was doubt as to whether it would be possible to do in a scalable way. Errors can be mitigated through correction, but could they be mitigated in such a way that they did not compound during a long computation, therefore ruining the desireable scaling properties of the ideal algorithms? The Threshold Theorem [15–19] showed that it was indeed possible, so long as the interaction Hamiltonian which describes the errors obeyed a stringent set of assumptions. The assumptions were initially unrealistic, but it was celebrated because it established for the first time that quantum error correction could, in principle, be implemented in a scalable way. The Threshold Theorem has been improved upon many times to apply it to a wider and wider class of interactions or errors which now include many realistic error models. This has led to optimism that scalable, fault-toleranct quantum computers are possible, which in turn led to the ongoing expenditure of great resources in order to realize such a possibility [20–26]. Appendix C contains more details on the Threshold Theorem

1.3 The Possibility of Fundamental Failure

We have identified a source of error which does not conform to the assumptions of the most general Threshold Theorem currently known to me [27]. This is the effective interaction that arises when the qubits are all coupled to a common boson bath, the same interaction which can give rise to the well-known but little-understood effects of super- and sub-radiance. It is rarely the dominant form of error in the small quantum computers being built today - however we will argue that this could change as the number of qubits in the computer grows, and in particular that the scaling properties of the error ruins any possible error correction scalability when that scaling is taken arbitrarily far. We will discuss precisely how this is different than the classical case. In this Chapter we will show why the current Threshold Theorem does not apply to this interaction, a result which by itself is cause for only mild concern, since the Threshold Theorem has been continuously generalized and therefore a future version may apply to this interaction as well. In the rest of this thesis we will examine this question by studying the interaction in greater detail, and argue based on the results that there is no conceivable Threshold Theorem which could apply to this interaction. This would mean no quantum computation would be scalable in the way its classical counterpart is. This does not preclude the possibility of quantum computers - perhaps this scaling issue only comes into play when the number of qubits is much larger than would be needed for useful quantum computations. We do not address the question here of for what computer size this effect is important, in part because it depends sensitively on the specifics of the technology and architecture used to build the quantum computer, but also because it is much more difficult than analyzing only the scaling itself. Therefore I will argue at least that these results imply a previously unrecognized difference between quantum and classical computing, and possibly that errors of this sort will prevent the dream of scalable fault-tolerant quantum computation from being realized.

There are already many reasons to suspect that this dream is too good to be true [19, 28–30]. For example, if the entire observable universe were filled with the best

possible classical computer (with the fundamental constraint on the amount of information stored in a volume implied by the Bekenstein bound for black holes [31]), it would only be able to store the state of around 400 qubits [32]. To store the state of a 500 qubit quantum state, then, would require about a volume 2^{100} the size of our observable universe, each filled with computers storing information in the densest way possible without collapsing into a black hole. Of course quantum physics is quite different than classical physics, and there is no reason a priori that the universe can't have that much information hidden in an almost classical-looking universe, but this is no mere difference in orders of magnitude. However if quantum computing is fundamentally impossible, new physics will be needed to explain why.

1.4 Proof of the Inapplicability of the Threshold Theorem for the exchange interaction

The exchange interaction arises whenever a collection of qubits is coupled to a common boson; this is always the case, although not always relevant for quantum error correction since other types of error may be significantly more prominent. The exchange interaction is described by sum of pairwise Hamiltonians, one for each pair of qubits and which affects only those two qubits: $H_{\text{exchange}} = \sum_{(jk)} H_{jk}$. One of the assumptions of the current Threshold Theorem [27] is that, for such a pairwise interaction, for any fixed qubit j,

$$\sum_{k \neq j} \mid H_{jk} \mid < \eta \tag{1.1}$$

where the inequality is supposed to hold as $N \to \infty$ for some fixed η. The value of the threshold η depends on the specifics of the computer, including the error rate, gate time, and the error correction code used. It is often shown to be around 10^{-4}, but for surface codes can be as high as 10^{-2} [33]. Assuming that the sum converges, the concern is over whether or not it converges to a small enough value to meet the threshold. So long as it converges, its value can be changed by implementing using

better error correcting codes, improving the technology, etc. Thus the fact that such a threshold exists implies that implementing scalable fault-tolerant quantum computing is simply a matter of developing better error correcting codes and technology. In contrast, the converge of this sum does not depend on any of these specifics but only on the structure of the coupling.

The convergence of 1.1 is not an unreasonable possibility, since as N increases the new qubits are added at the edge of the computer, and the interaction strength between two qubits j and k (quantified by the magnitude of H_{jk}) decreases with distance. In the same way that $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6 < \infty$, we might hope that even as N goes to infinity, and therefore also the number of terms in the sum $\sum_{k\neq j}\left|H_{jk}\right|$ goes to infinity, that it still remains finite. However just as $\sum_{k=1}^{\infty} k^{-1}$ does not converge, there may be kinds of interactions for which 1.1 does not converge even though the strength decays with distance. Recall that this condition for scalable fault-tolerance is sufficient but not necessary; it must be so, for it does not take into the account the relative phases of the different interactions. The overall effect on a single qubit would be much greater if the effects of all others added constructively in phase, but this crude upper bound would not discern any difference if this were the case. For large samples, where the phases of the interactions $e^{ik(\left|\vec{r}_j - \vec{r}_k\right|)}$ become random, it seems plausible that these phases might be the savior since they do not describe constructive interference. However no one has to my knowledge proven an inequality like 1.1 where the phases of the interactions are included - it is merely remains a possibility.

The dipole interaction between atoms decays like r^3 and is a common source of noise for atomic qubits [34]. Even disregarding the phases, for many computer geometries and error types, including the dipole interaction, the Threshold Theorem condition (1.1) holds. However the effective interaction arising due to the interaction with a common boson bath decays more slowly like r^{-1}, and we now show that in this case the threshold condition is not met. For a more complete discussion on this effective interaction, and a derivation of it from the more fundamental interaction between the atoms and the boson bath, see Appendix E.

For a collection of N qubits coupled to a common bath, the effective interaction Hamiltonian is $H = \sum_{(jk)} H_{jk}$ with

$$H_{jk} = F \frac{e^{-ik(|\vec{r}_j - \vec{r}_k|)}}{k \mid \vec{r}_j - \vec{r}_k \mid} \left(\sigma_j^+ \sigma_k^- + \sigma_j^- \sigma_k^+ \right) \tag{1.2}$$

where $\sigma_{j,k}^{+,-}$ are the raising and lowering operators for qubits j and k, F is a coupling constant, and $\vec{r}_{j,k}$ are their spatial positions. Here we have kept only leading orders in $1/r$ since the limit of large N also means the the limit of large separations (since the spatial density of qubits is taken to be fixed). The norm of an operator, in the sense used in the Threshold Theorem, is bounded below by the magnitude of any of its matrix elements in any normalized basis. If this lower bound diverges, then so does the actual sum; this will be our strategy.

The only non-zero matrix elements in H_{jk} in the normalized computational basis are $Fe^{-ik|\vec{r}_j - \vec{r}_k|}/k|\vec{r}_j - \vec{r}_k|$ which has magnitude F/kr_{jk} where $r_{jk} = |\vec{r}_j - \vec{r}_k|$ for short. Since we are only looking at whether or not this sum converges, we can ignore the coupling constant F, and say that the sum (1.1) is bounded below by

$$\sum_{k \neq j} 1/r_{jk} \tag{1.3}$$

This sum depends on the geometry of the computer - however one thing its convergence or divergence does not depend on is the overall scale of the computer i.e. if this sum diverges for some spatial qubit arrangement then if we move all the qubits away from each other uniformly, it will still diverge no matter how far we separate them; this is because if $r_{jk} \to \lambda r_{jk}$ for all pairs, then the sum only changes by a multiplicative constant, $\sum_{k \neq j} H_{jk} \to \lambda^{-1} \sum_{k \neq j} H_{jk}$. Thus we need not worry about the spatial scale of the computer, only its shape. Quantum computer realizations efforts are restricted to 2 and 3 spatial dimensions, so we examine only those cases here. In simple square or box geometries in either 2 or 3 dimensions, this sum diverges as $N^{1/2}$ and $N^{2/3}$, respectively, as shown in Appendix A. Thus, regardless of the error correcting code, the technology used, and the rate of errors

(so long as it is not zero which of course it can never be), the Threshold Theorem in its current form cannot guarantee that scalable fault-tolerant quantum computing is possible in the face of this error. As mentioned, this does not mean it is impossible, only that more work must be done to ensure that it is. In order to answer the question of whether or not this error source is scalably error correctable, in Chapter 2 we will examine this interaction in greater detail. The results of that Chapter will lead us to the conclusion that it is definitively not error correctable in a scalable way, as well as to several other conclusions and applications which are addressed in the following Chapters.

The idea that the effect of a boson bath on a quantum ensemble decreases in time as the number of subsystems grows is not new. Since the seminal paper by Dicke in 1954 ([35]), it has been known that the decay rate of an ensemble can experience increases or decreases, known as super- and sub-radiance respectively, in a manner that depends on the total number of subsystems. For infinitely small systems, where all interactions are in phase, these rates are changed by up to a factor of N, which is an enormous change even for modest N. When the physical size of the sample is taken into account, such that the interactions have a more complex phase relationships, much less is known except that the effect is lessened but still present. Although we will not specifically focus on the phenomona of super- or sub-radiance, they show how dynamics with strong N-dependence arise in such systems, and may be useful tools for conceptually understanding the mathematical results.

2 The Spectrum of the Exchange Hamiltonian

2.1 Motivation

Large collections of two-level systems are common in nature and physics labs; such systems include ferromagnets [36], black holes [37–39], and quantum computers [1], and spin glasses [40]. They exhibit a wide variety of behaviors which are under study both theoretically and experimentally; these behaviors include sub- and super-radiance [41], many-body localization [42], and the entanglement entropy area law of low lying energy states [43]. Despite decades of study, many properties of these systems remain mysterious. This is because they can't be understood exactly, at least in practice and possibly even in principle. For example, as mentioned in Chapter 1, if the entire observable universe were filled with the best conceivable classical computing hardware, it would only be able to store the quantum state of spin systems with around 400 spins [32], whereas the systems under consideration can contain millions or even $\sim 10^{26}$ spins. This difficulty is due to the exponential scaling of the size of Hilbert space where the abstract physical representation of such a system's state lives. The dimension of this space is 2^N where N is the number of subsystems, and this becomes intractable to simulate for $N > 30$ or so. Even if computing power were increased by 3 orders of magnitude, this maximum would only increase to about 40. Therefore approximate methods are needed, which can be tailored to the regime or phenomenon being studied. In this thesis, I present a new tool to try to understand these systems; we will find the eigenvalue distribution of the interaction Hamiltonian for physically motivated couplings F_{jk} in the large N limit, giving new insight into the timescales of the dynamics such systems undergo.

2.2 The Hamiltonian for Many-Spin Interactions

Consider a collection of 2-level systems indexed by the integers $j = 1, ..., N$ with the ground and excited state of the jth system denoted by $|0\rangle_j$ and $|1\rangle_j$, respectively, and an energy difference of $\hbar\omega = \omega$ (setting $\hbar = 1$). If the subsystems are non-interacting, the time evolution of the whole system is determined by the Hamiltonian $H = \omega \sum_j \sigma_j^z$ via the Schrödinger Equation $d|\psi\rangle/dt = iH|\psi\rangle$ where $\sigma_j^z = |1\rangle_j \langle 1|_j$ is the z Pauli matrix with a constant added to the energy to allow for a shifted definition, deviating from the more typical $\sigma^z = \frac{1}{2}(|1\rangle\langle 1| - |0\rangle\langle 0|)$.

Any state of the whole system can be written $|\psi\rangle = \sum_q c_q |q\rangle$ where $\{|q\rangle\}$ is the computational basis in which, for each element, each subsystem has a definite state, and the c_q are complex numbers (normalized so that $\langle\psi \mid \psi\rangle = \sum_q \left|c_q\right|^2 = 1$). For the example of $N = 3$, $\{|q\rangle\} = \{|000\rangle, |001\rangle, |010\rangle, |100\rangle, |011\rangle, |101\rangle, |110\rangle, |111\rangle\}$. Note that the set $\{|q\rangle\}$ has 2^N elements, which we denote as $\left|\{|q\rangle\}\right| = 2^N$. The computational basis diagonalizes the Hamiltonian, and thus its time dependence is fully and easily understood; each component c_q rotates in the complex plane with an angular frequency given by ω times the number of excited subsystems in it, i.e. $c_q(t) = c_q(0)e^{-i\omega M_q t}$ where M_q is the number of 1s in the arrangement q. However, once interactions between the subsystems is added, the diagonal basis is not generally known and the dynamics can be much more difficult to calculate exactly or even approximate.

In this thesis, we will consider interactions of the form

$$H_{\text{int}} = \sum_{(jk)} H_{jk} = \sum_{(jk)} F_{jk}(\sigma_j^+\sigma_k^- + \sigma_j^+\sigma_k^-) \tag{2.1}$$

where $\sum_{(jk)}$ denotes the sum over all pairs of subsystems (i.e. $\sum_{(jk)} = \frac{1}{2}\sum_j \sum_{k\neq j}$) and H_{jk} acts non-trivially only on subsystems j and k. The $\sigma_j^\pm$ operators raise or lower the state of subsystem j so that $\sigma_j^+ |0\rangle_j = |1\rangle_j$, $\sigma_j^+ |1\rangle_j = 0$, $\sigma_j^- |0\rangle_j = 0$, and $\sigma_j^- |1\rangle_j = |0\rangle_j$. We will study F_{jk} of the form $F_{jk} = F(\vec{r}_{jk}) = f\sin(\theta_{jk})(kr_{jk})^{-\alpha}$ where $\vec{r}_{jk}$ is the separation vector between the two subsystems, k is the wavelength

associated with the energy difference ω ($k = 2\pi/\lambda = 2\pi\omega/c$), f is a complex constant with units of energy or frequency, and θ_{jk} are phases randomly drawn uniformly from $[0, 2\pi]$ for each pair of subsystems. When the notion of 'averaging' is used later, it means with respect to randomly drawn positions of the subsystems from some spatial distribution $\rho(\vec{r})$, such as a box, sphere, or gaussian cloud with a fixed shaped and a characteristic size R which grows with N as $R \sim N^{1/D}$, where $D = 2$ or 3 is the number of spatial dimensions of the distribution of subsystems. This scaling is such that the spatial subsystem density remains constant on average with respect to N. We restrict our attention to isotropically shaped ensembles, as opposed to the cigar or other shapes often analyzed in the context of super- or sub-radiance. Here and throughout, $\sim$ means 'scales with N as'.

The random phase assumption is physically reasonable because typically this phase is given by kr_{jk}, and in the large N limit, for almost all pairs j and k, kr_{jk} becomes much larger than 2π and the distribution of the phases $kr_{jk} \mod 2\pi$ approaches uniform in the interval $[0, 2\pi]$.

We have taken F_{jk} to be both complex and symmetric, meaning H_{int} is not Hermitian; this is to be expected, since this kind of interaction arises when the subsystems are coupled to a common boson bath and it is well known that the dynamics of an open quantum system can be approximated by a non-Hermitian Hamiltonian acting only on the open quantum system, as opposed to the larger quantum system which includes the external system and is still governed by a hermitian Hamiltonian.

Using all of these assumptions throughout will allow for significant simplification, although a similar analysis could likely be applied in the absence of some of them. Interactions which conform to these assumptions arise physically in many situations, such as in atomic interactions where there the dipole interaction is of the Exchange type and has couplings F_{jk} which decays as distance cubed [34], and the boson exchange interaction in which decays as distance to the first power [44]. In order to gain new insight into such systems, we will analytically find the spectrum of these kinds of Hamiltonians when N is arbitrarily large.

2.3 The Moments of the Spectrum Distribution

Rather than compute the eigenvalues directly, we will find the moments of their distribution, or rather the smooth distribution which the discrete spectrum asymptotes to for large N. The distribution $f(\lambda)$ is constructed from its moments by taking the fourier transform of

$$F(\xi) = \sum_{p=0}^{\infty} \frac{(i\xi)^p}{p!} \sigma^{(p)} \tag{2.2}$$

where the pth moment of the distribution is defined as

$$\sigma^{(p)} = 2^{-N} \sum_{q} \langle q| H^p |q\rangle = 2^{-N} \sum_{\lambda} \lambda^p = \int_{-\infty}^{\infty} \lambda^p f(\lambda) d\lambda \tag{2.3}$$

where $\{\lambda\}$ are the eigenvalues of H. The p-th spectrum moment can be thought of as the average of the eigenvalues raised to the p-th power. The fourier transform of 2.2 is the distribution because it is the expectation value of $e^{i\lambda\xi}$ where the expectation value is with respect to λ i.e.

$$F(\xi) = \int_{\infty}^{\infty} e^{i\xi\lambda} f(\lambda) d\lambda = \sum_{k=0}^{\infty} \frac{(-i\xi)^k}{k!} E[\lambda^k] = \sum_{k=0}^{\infty} \frac{(i\xi)^k}{k!} \int_{-\infty}^{\infty} \lambda^k f(\lambda) d\lambda = \sum_{k=0}^{\infty} \frac{(i\xi)^k}{k!} \sigma^{(k)} \tag{2.4}$$

In order to estimate the distribution after finding the moments, we will truncate the sum over p in the above expression. In order to accomplish this we will need to utilize the large N limit to find the moments analytically. Since N is arbitrarily large and we truncate the sum over p at some fixed $p_{\max}$, we can always take $p \leq p_{\max} << N$.

Using only the interaction Hamiltonian (2.1), the moments can be written

$$\sigma^p = 2^{-N} \sum_{q} \langle q| \prod_{w=1}^{p} \sum_{(j_w k_w)} H_{j_w k_w} |q\rangle = 2^{-N} \sum_{q} \langle q| \left(\prod_{w=1}^{p} \sum_{(j_w k_w)} F_{j_w k_w} (\sigma_{j_w}^{+} \sigma_{k_w}^{-} + \sigma_{j_w}^{-} \sigma_{k_w}^{+}) \right) |q\rangle \tag{2.5}$$

The sum and product $\prod_{w=1}^{p} \sum_{(j_w k_w)}$ can be understood as having one term for

each unique selections of p pairs, and thus there are $\binom{N}{2}^p$ such terms. For one such choice of p pairs, $\{\{j_1k_1\}...\{j_pk_p\}\}$, we have the associated term $H_{j_1k_1}...H_{j_pk_p}$ which contributes to $\sigma^{(p)}$ an amount equal to $2^{-N}\sum_q \langle q| H_{j_1k_1}...H_{j_pk_p} |q\rangle$. In total, $\sigma^{(p)}$ is the sum of all such contributions from each of the $\binom{N}{2}^p$ ways to choose the p pairs $\{\{j_1k_1\}...\{j_pk_p\}\}$.

2.3.1 Triplets

We are going to associate each term $H_{j_1k_1}...H_{j_pk_p}$ one or more triplets $(S, \{j\}, \pi)$, composed of a pattern S, a subsystem selection $\{j\}$, and a subsystem assignment π. After defining them, we will see there is at least one triplet for each term $H_{j_1k_1}...H_{j_pk_p}$, and that every triplet which obeys the Compatibility Conditions (to be specified) constructs exactly one sequence of pairs, or equivalently one term $H_{j_1k_1}...H_{j_pk_p}$. This means we can rewrite $\prod_w \sum_{(j_wk_w)}$ in (2.5) with a sum over all patterns, subsystem selections, and subsystem assignments, as long as we use only triplets which obey the Compatibility Conditions, and divide by a degeneracy factor $D(S, \{j\}, \pi)$ equal to the number of different triplets which construct the same term. We denote the term constructed from a triplet $(S, \{j\}, \pi)$ as $\prod H(S, \{j\}, \pi) = H_{j_1k_1}...H_{j_pk_p}$ where $\{\{j_1k_1\}...\{j_pk_p\}\}$ are the p ordered pairs constructed from the triplet in a way yet to be specified. We similarly define $\prod F(S, \{j\}, \pi) = F_{j_1k_1}...F_{j_pk_p}$ for future use.

Achieving this, we can write the moments as

$$\sigma^{(p)} = 2^{-N}\sum_q \langle q| \left(\sum_S \sum_{\{j\}} \sum_\pi \prod H(S, \{j\}, \pi)/D(S, \{j\}, \pi) \right) |q\rangle \tag{2.6}$$

where on the right hand side, the p dependence is hidden inside the sum over patterns - for each p a different set of patterns is allowed, and for a given pattern a certain set of subsystem selections and assignments $\{j\}$ and π are allowed, according to the Compatibility Conditions defined below. Re-arranging the sum in this way is desireable because it will allow us to utilize the large N limit in a way that is not manifest in the expression (2.5). Later, we will calculate the degeneracy factor

$D(S, \{j\}, \pi)$ and observe that it, as well as the number of different patterns S and the number of different compatible subsystem assignments π, do not depend on N. The N dependence of sum over q can accounted for in a straightforward manner, leaving all the remaining N dependence in the sum over subsystem selection $\sum_{\{j\}}$ and the F_{jk} factors, due to the fact that the total system size grows with N so as to keep subsystem density fixed, meaning that as we increase N the added subsystems are at the edge of the system and therefore the additional F_{jk} will typically be smaller than the ones that already existed for smaller N (given the assumed form of F_{jk} whose magnitude shrinks with the distance between subsystems j and k). Once all sources of N-dependence have been taken into account we will be in a position to identify a very special subset of patterns, $\{S_t\}$, for which in the large N limit the associated term's contribution to $\sigma^{(p)}$ becomes arbitrarily larger than the contribution to the moment from all patterns not in $\{S_t\}$. Since we will see that the number of patterns, and therefore the number of such terms, does not depend on N, this will allow us to conclude that for large N, only the patterns in $\{S_t\}$ need to be included.

Calculating the contribution to the moments for all patterns is a difficult task, but for the special patterns S_t this task can be much more readily accomplished. Rather than depend on complicated products of F_{jk} raised to different powers and having possibly overlapping subscripts, we will find that for the S_t we can express their contribution to the moments, and therefore the distribution $f(\lambda)$, in terms of only the average of F_{jk}^2 (averaged over all pairs (jk)). Then for any given coupling F_{jk} corresponding to some system of interest, its energy spectrum density can be approximated for large but finite N, potentially leading to new physical insight.

2.3.2 Pattern and Subsystem Selection, Assignment Definitions

We will now define a pattern, a subsystem selection, and a subsystem assignment, and then demonstrate how to construct a term from a pattern or find a pattern

associated with a given term. We establish along the way that every term can be associated with at least one triplet, and that under certain compatibility conditions, every triplet constructs one and only one term. Once established, these facts justify the proposed sum re-arrangement 2.6, provided the sum over triplets is taken to mean only the triplets which obey the Compatibility Conditions. These are summarized below for ease of reference.

2.3.3 Compatibility Conditions

-Every subset in S has an even number of elements which are integers between 1 and p (inclusive) and each integer between 1 and p (inclusive) appears in exactly two different subsets in S

-Given a pattern S in a triplet, the number of subsystems referenced in a $\{j\}$ which appears in the same triplet as S must be the same as the number of subsets in S

-Given a pattern S and subsystem selection $\{j\}$, any π which appears in the same triplet must be a permutation of a number of objects equal to the number of elements in S (and therefore also the number of subsystems in $\{j\}$)

2.3.4 Pattern Definition

A pattern is a list with collections of locations of Hamiltonians where each collection contains the integers associated with all of the locations at which a given subsystem appears. For example, the pattern associated with $H_{12}H_{24}H_{23}$ is $\{\{1\},\{1,2,3\},\{3\},\{2\}\}$ because the first subsystem appears only in the first pairwise Hamiltonian, the second appears in all three, and the last two only appear in the third and second, respectively. For visual clarity, we are using an ordering convention that the subsystem labelled 1 appears in the locations in the first collection and so forth (left to right), but it is important to remember going forward that mathematically we treat the pattern as an un-ordered set. Note that every integer between 1 and p, inclusive, appears exactly twice among the collections in

S, following the Compatibility Conditions. This ensures that exactly two subscripts are assigned to each pairwise H_{jk}.

The total number of elements in all the subsets of S is $2p$. Note also that which subsystems appear is immaterial to the pattern, which only contains information about the coincidences of subsystem appearances. For example $H_{56}H_{67}H_{68}$ corresponds to the same pattern as the example above. Thus in $H_{12}H_{24}H_{23}$ it is useful to think of the integers not as subsystem indexes but rather 'slots' in which some subsytem will be placed - which subsystem is placed in which slot is determined by the subsystem selection and subsystem assignment together. This makes it clear why, in addition to the pattern, we require both the subsystem selection and subsystem assignment to construct the term $H_{j_1k_1}...H_{j_pk_p}$ uniquely.

2.3.5 Subsystem Selection Definition

The subsystem selection associated with a term $\{\{j_1k_1\}...\{j_pk_p\}\}$ or $H_{j_1k_1}....H_{j_pk_p}$ is the set of all unique subsystems $\{j\}$ which appear in it. The size of this set is equal to the number of subsets in S because there is one subset in S for each subsystem. If the number of collections appearing in S is specified to be u, we denote this by writing $S(u)$, and use the notation $\left|\{j\}\right| = u$ to mean that the size of the set $\{j\}$ is u, i.e., it has u subsystems listed in it. A general subsystem selection with u subsystems is denoted $\{j\}(u)$, just as a general pattern with u collections is denoted $S(u)$. One of the Compatibility Conditions is that the subsystem selection in a triplet has the same number of elements as collections in S. In this notation, this Compatibility Condition can be stated as follows: for any triplet $(S(u), \{j\}(u'), \pi)$, we require $u' = u$. Since the sum over patterns, subsystem selections, and subsystem assignments in (2.6) is done in that order, we can enforce this Compatibility Condition by making the sum over subsystem selections for a given pattern be only over subsystem selections with the same value of u as S.

2.3.6 Subsystem Assignment Definition

The subsystem assignment π contains information about which subsystem in $\{j\}$ appears in each collection of subscript locations in S. It must link up u different subsystems in $\{j\}(u)$ with u subsets in $S(u)$; therefore each π compatible with some $S(u)$ and $\{j\}(u)$ must be a permutation of u objects, or the different ways to pair up the u elements in each of those objects. Denoting such a permutation of u elements as $\pi(u)$, another Compatibility Condition is that for a triplet $(S(u), \{j\}(u), \pi(u'))$, $u' = u$. This means there are $u!$ different subsystem assignments $\pi(u)$ compatible with any two $S(u)$ and $\{j\}(u)$.

Equipped now with the definition of a triplet, we establish that for every term $H_{j_1k_1}...H_{j_pk_p}$ there is at least one triplet associated with it, that each triplet that obeys the Compatibility Conditions is associated with exactly one term, and compute the degeneracy factor $D(S, \{j\}, \pi)$ by counting how many other triplet $\{S', \{j\}', \pi'\}$ construct the same term as $(S, \{j\}, \pi)$.

2.3.7 Identifying the Term Associated with a Triplet

Here we demonstrate how to construct a unique ordered set of p pairs $\{\{j_1k_1\}...\{j_pk_p\}\}$, or equivalently a term $H_{j_1k_1}...H_{j_pk_p}$, from a given triplet $(S, \{j\}, \pi)$ which obeys the Compatibility Conditions. We begin with a 'blank' term $H_{\_\_}...H_{\_\_}$ and fill in the subscripts with subsystem references (which are indexed by the integers) according to the triplet as follows: for each subsystem in $\{j\}$, first use π to determine which collection in S that subset is associated with, and then insert that subset's index in the subscript of all the $H_{\_\_}$ at the locations in its associated collection. Since $H_{jk} = H_{kj}$ we need not specify which subscript, although if there is already a subscript filled in then of course it must be placed in the other. Given this recipe, the only conditions under which the same term might be constructed from two or more different triplets is if there is at least one pair of subsets who only appear as subscripts of the same pairwise Hamiltonians; in this case the degeneracy factor

$D(S, \{j\}, \pi)$ is 2 raised to the number of such pairs. We will use this calculation much later, but for now, the salient feature is that it is independent of N (since the number of pairs is at most half of p which is considered fixed with respect to N).

We also note the fact that, as is evident from its calculation, D only depends on S and not $\{j\}$ or π. Therefore from now on write it simply as $D(S)$.

2.3.8 Associating a Triplet with Any Term

Now we will show how to identify a compatible triplet associated with any term $H_{j_1k_1}...H_{j_pk_p}$, thus establishing that all are accounted for when summing over Compatible triplets in 2.6. Arbitrary decisions will need to be made, due to the fact that some triplets have $D(S, \{j\}, \pi) > 1$.

We will deduce the three constituents of a triplet from a sequence of pairs $\{\{j_1k_1\}...\{j_pk_p\}\}$ or a term $H_{j_1k_1}...H_{j_pk_p}$ by selecting one by one, in any order, each subsystem which appears: for each such subsystem, first add it to the subsystem selection list, then locate every individual pairwise Hamiltonian $H_{j_wk_w}$ at which it appears in the term and collect the integer identifiers of these locations into a collection (again using the convention that 1 is the left-most pair and incrementing by 1 to the right), add that collection to the pattern S, and finally add an element to π matching that subsystem with that collection. The Compatibility Conditions will automatically be met by a triplet constructed from a term in this way; thus, it is clear that the sum over all compatible triplets will include each term at least once. By dividing by $D(S)$, we also ensure none of them are counted more than once, justifying 2.6.

Something to keep in mind going forward: despite the notation, $S(u), D(u)$, and $\{j\}(u)$ are not functions of u. Rather, the purpose of the notation is to specify the value of u for the given object. For example, in the same way that S should be thought of as a pattern for some integer p, $S(u)$ should be thought of some pattern for the integer p which contains u collections of locations. Similarly, while $\{j\}$ is any list of any number of different subsystems, $\{j\}(u)$ is any list of u different

subsystems.

2.3.9 Back to $\sigma^{(p)}$

We are now equipped to make some fruitful manipulations to the expression for σ^p. First, note that $2 \leq u \leq 2p$ since there are only $2p$ empty slots in the term to be filled with u different subsystems; u is maximized when there is a different subsystem in each of the $2p$ slots and minimized when the same two subsytems appear at each location (recall that the original sum is over p pairs of subsystems, meaning the same pairwise Hamiltonian can't have a subsystem as both of its subscripts, leading to a minimum of 2 instead of 1).

Recall that we denote the term $\prod_w H_{j_w k_w} = H_{j_1 k_1} ... H_{j_p k_p}$ constructed from a triplet $(S, \{j\}, \pi)$ as $\prod H(S, \{j\}, \pi)$. For all q and all $\prod H(S, \{j\}, \pi)$, when the triplet $(S, \{j\}, \pi)$ is fixed then $\langle q| \prod H(S, \{j\}, \pi) |q\rangle$ has only two possible values: 0, or some value which depends on the triplet and the coupling F_{jk}, but not on which q it is (among the q that have the property $\langle q| \prod H(S, \{j\}, \pi) |q\rangle \neq 0$). In other words, whenever q is such that $\langle q| \prod H(S, \{j\}, \pi) |q\rangle \neq 0$, it is always equal to the same thing for all other q for which the same quantity is also non-zero. When it is nonzero, we define it to be equal to $\prod F(S, \{j\}, \pi)/\Omega(S)$, where $\Omega(S)$ is the number of $|q\rangle$ such that the quantity $\langle q| \prod H(S, \{j\}, \pi) |q\rangle$ is not zero. Defined this way, the new quantity $\prod F(S, \{j\}, \pi)$ relates to $\prod H(S, \{j\}, \pi)$ via

$$\Omega(S, \{j\}, \pi) \prod F(S, \{j\}, \pi) = \sum_q \langle q| \prod_w H_{j_w k_w} |q\rangle \tag{2.7}$$

This is consistent with the previous definition $\prod F(S, \{j\}, \pi) = \prod_{w=1}^{p} F_{j_w k_w}$. Due to the invariance of the problem under the relabelling of subsystems, it must be the case that $\Omega(S, \{j\}, \pi)$ is the same for any (compatible) $\{j\}$ and π placed into a triplet with that S, since any modification to $\{j\}$ or π can be effectively undone by relabelling the subsystems. Therefore, like $D(S)$, Ω also depends only on S and we similarly write it simply as $\Omega(S)$.

Some patterns construct terms for which there is no $|q\rangle$ with non-zero $\langle q| \prod H(S, \{j\}, \pi) |q\rangle$; for these we define $\Omega(S) = 0$, allowing us to write

$$\begin{aligned} \sigma^{(p)} &= 2^{-N} \sum_q \langle q| \left(\sum_S D(S)^{-1} \sum_{\{j\}} \sum_\pi \prod H(S, \{j\}, \pi) \right) |q\rangle \\ &= 2^{-N} \sum_S D(S)^{-1} \Omega(S) \sum_{\{j\}} \sum_\pi \prod F(S, \{j\}, \pi) \\ &= 2^{-N} \sum_{u=2}^{2p} \sum_{S(u)} \frac{\Omega(S(u))}{D(S(u))} \sum_{\{j\}(u)} \sum_{\pi(u)} \prod F(S(u), \{j\}(u), \pi(u)) \end{aligned} \tag{2.8}$$

2.3.10 Extracting the N dependence of $\Omega(S)$

$\Omega(S)$ depends on N, but we will now extract the N dependence of $\Omega(S)$ by factoring that dependence out of a related, N-independent function $\Omega'(S)$. In order to do so, we must establish the notion of comparing two different computational basis elements $|q\rangle$ and $|q'\rangle$; we say that they 'agree on', or 'are the same for' a subsystem if that subsystem has the same status in both of them, and similarly that they 'disagree on', or 'are different for', those subsystems that are opposite. Every subsystem will either disagree or agree for each pair of computational basis states because by definition every subsystem has a definite state in each of the computational basis states. For example, with $N = 5$ the two computational basis states $|01101\rangle$ and $|01011\rangle$ agree on subsystems 1, 2, and 5, and are different on subsystems 3 and 4 (where have again employed the convention of indexing the subsystems left to right, starting at 1 and incrementing by 1).

Observe that the quantity $\langle q| \left(\prod_{w=1}^{p} \sum_{(j_w k_w)} H_{j_w k_w} \right) |q\rangle$ is invariant upon altering the status in q of any subsystem not among the j_w or k_w, or equivalently any subsystem not in $\{j\}$ for the quantity $\langle q| \prod H(S, \{j\}, \pi) |q\rangle$. Using the sense of sameness of subsystems between pairs of states defined above, we can say that $\langle q| \prod H(S, \{j\}, \pi) |q\rangle = \langle q'| \prod H(S, \{j\}, \pi) |q'\rangle$ if q and q' agree on all subsystems in $\{j\}$. Now consider a new quantity $\Omega'(S, \{j\}, \pi)$ which is similar to $\Omega(S, \{j\}, \pi)$ except instead of counting all arrangements of all subsystems for which

$\langle q| \prod H(S, \{j\}, \pi) |q\rangle \neq 0$, we count only arrangements of the subsystems in $\{j\}$. Ω' also depends only on S, and similarly we write it as $\Omega'(S)$. The maximum of $\Omega'(S(u))$ is 2^u since that is the total number of arrangements of any $\{j\}(u)$ compatible with $S(u)$, whereas the maximum of $\Omega(S(u))$ is 2^N since that is the total number of arrangements of all N subsystems. $\Omega'(S(u))$ is only counting arrangements of u subsystems instead of N and therefore does not depend on N (since $u < 2p$ and p is fixed with respect to N). There is an exception when $\Omega(S)$ or $\Omega'(S)$ is zero for some pattern S - this can be relevant for N dependence since the N scaling of that pattern's associated term is then merely hypothetical, not corresponding to an actual contribution to the moments. We will use the N dependence in order to discard contributions which have a relatively vanishing contribution for large N, regardless of the constant (with respect to N) out front - however this logic clearly only applies when that constant is not zero. Noting that $\Omega'(S) \neq 0$ implies $\Omega(S) \neq 0$ and vice versa, this exception can be accounted for by assuring that $\Omega'(S_t) \neq 0$ for the final patterns S_t whose contributions have an N-scaling greater than the contribution from all other patterns for which $\Omega'(S) \neq 0$.

For each of the 2^u arrangements of the subsystems in $\{j\}(u)$ which could count towards $\Omega'(S(u), \{j\}(u), \pi(u))$, there are 2^{N-u} arrangements of all subsystems which counts towards $\Omega(S(u), \{j\}(u), \pi(u))$: one for each possible arrangement of the other $N - u$ subsystems. Thus $\Omega(S(u)) = 2^{N-u}\Omega'(S(u))$. Applying this to the spectrum moments sum, we have

$$\sigma^{(p)} = \sum_{u=2}^{2p} 2^{-u} \sum_{S(u)} \frac{\Omega'(S(u))}{D(S(u))} \sum_{\{j\}(u)} \sum_{\pi_u} \prod F(S(u), \{j\}(u), \pi(u)) \tag{2.9}$$

Now all N dependence is contained in the sum over subsystem selections and the values of the F_{jk} (we will count them later, but it follows from the definition of a pattern that the number of them for a given p is independent of N). The number of terms in the sum over subsystem selections $\{j\}(u)$ is $\binom{N}{u}$ since this is the number of different ways to select u subsystems from a list of N of them. The N scaling of this quantity is N^u. The N dependence of the couplings F_{jk} is trickier to quantify;

in order to get a handle on it we define Expectation Values in the next section.

2.3.11 Expectation Values

Define the expectation value $E[\prod F(S)]$ of $\prod F(S,\{j\},\pi)$ for some fixed S by averaging over all subsystem selections $\{j\}$ and assignments π:

$$E\Big[\prod F(S)\Big] = A_S^{-1}\sum_{\{j\}}\sum_{\pi}\prod F(S,\{j\},\pi) \tag{2.10}$$

where A_S is a normalization constant equal to the total number of terms in the sum to its right, consistent with the idea of $E[\ ]$ as an average over $\{j\}$ and π. A_S is equal to the number of different $\{j\}$ compatible with S times the number of π compatible with S, denoted as $A_S = \big|\{j\}\big| \times |\pi|$.

For a pattern $S(2)$ with $u = 2$, both of the subsystems appear in the subscript of each pairwise H_{jk} and accordingly $\prod F(S(2),\{j,k\},\pi)) = F_{jk}^p$ (all π are equivalent in this case since all subsystems are assigned to the same location). Define $E[F^B]$ for any integer B to be the average of $\Big|F_{jk}^B\Big|$ over all pairs,

$$E[F^B] = \binom{N}{2}^{-1}\sum_{(jk)} \mid F_{jk} \mid^B \tag{2.11}$$

so that $E[\prod F(S(2))] = E[F^p]$.

The absolute value has been added because for odd B, $\sum_{(jk)} F_{jk}^B \to 0$ due to the random phase assumption. For a given large but finite N and fixed subsystem positions, $\sum_{jk} F_{(jk)}^B$ will not generally be zero; although it's just as likely to be negative as positive, its typical magnitude is nonzero and given by the central limit theorem. When comparing the N scaling of the contributions from different patterns, or estimating for the purpose of approximation, the typical magnitude is the relevant quantity - whether the contribution from a pattern adds positive or negative value to $\sigma^{(p)}$, when determining which are the dominant contributions for large N the quantity of interest is the overall size of the contribution. Thus we will

define the expectation values as capturing the typical size of the contribution from a pattern.

The addition of the absolute value does not alter the quantity for even B and makes it for a more meaningful definition for odd B.

What is the value of $E[F^B]$ for odd B, and how does its N scaling relate to the case of even B? Odd $E[F^B] = \sum_{(jk)} F_{jk}^B$ are sums of $\binom{N}{2}$ summands of the same kind as in the even case, the difference is that they have a randomly signed factor $\sin(\theta_1)\sin(\theta_2)$ instead of the always positive geometric factor $\sin(\theta_1)^2$ of the even B case (here $\theta_{1,2}$ is a random phase). Thus their average is zero, but by the central limit theorem they will approach a gaussian distribution a variance equal to the variance of the summands, $E[F^{2B}]$, divided by $\sqrt{\binom{N}{2}}$. This means that for odd B the typical magnitude of the average over the F_{jk}^B, whose definition was substituted for 0 in $E[F^B]$, scales as $E[F^{2B}]^{1/2}/N$. For the purposes of actual spectrum approximation, the odd expectation values can be generated by drawing from a gaussian distribution centered at 0 and with width $E[F^{2B}]^{1/2}/N$.

Define a set of integers $B(S)$ for a given pattern S which is the unordered set of integers to which power each unique F_{jk} appearing in $\prod F(S)$ is raised. For example, for the pattern $\{\{1,2,4\}\{1,2,3\}\{3,4\}\}$ corresponding to the term $F_{12}F_{12}F_{23}F_{13}$, $B(S) = \{2,1,1\}$. Note that the elements of $B(S)$ will always sum to p.

Suppose we have two different patterns S and S' which correspond to the same integers $B(S) = B(S')$, for example a pattern corresponding to $F_{12}^{B_1}F_{23}^{B_2}$ and a pattern corresponding to $F_{12}^{B_1}F_{34}^{B2}$. The sum over subsystem selections in $E[\prod F(S)]$ can be thought of as a sum over the positions corresponding to the locations of the subsystems, and in the large N limit this sum becomes well approximated by a continuous integral which is well defined because the assumed form of $F_{jk} = f\sin(\theta_{jk})(kr_{jk})^{-\alpha}$ is defined for all positions, not just the specific locations of the subsystems. Using this property we can define the closely related function $F(\vec{r}_1, \vec{r}_2) = f\sin(k \mid \vec{r}_1 - \vec{r}_2 \mid)(k \mid \vec{r}_1 - \vec{r}_2 \mid)^{-\alpha}$ where the angle will approach a uniform distribution when N is large and therefore the system is large. Defining

$F(S, \{\vec{r}\}, \pi)$ to be the value of the term $F(S, \{j\}, \pi)$ where the subsystems $\{j\}$ are at the locations $\{\vec{r}\}$, the integral limit of the $E[\prod F(S)]$ is

$$\begin{aligned} E\Big[\prod F(S(u))\Big] &= A_{S(u)}^{-1} \sum_{\{j\}} \sum_{\pi} \prod F(S(u), \{j\}(u), \pi(u)) \\ &\to \frac{1}{u!} \int \left(d^D r\right)^u \sum_{\pi} F(S(u), \{\vec{r}\}, \pi) \Big(\prod \rho(\vec{r}) \Big) - ... \end{aligned} \tag{2.12}$$

where D is the number of spatial dimensions, $\rho(\vec{r})$ is the subsystem density function normalized so that $\int d^D r \rho(\vec{r}) = 1$ (which accounts for part of $A_{S(u)}$), and the product over ρ is the product over the density at the location of each of the u subsystems. Because the subsystems must be unique, we must disallow $\vec{r}_1 = \vec{r}_2$, etc., which is represented by the subtraction of an unspecified quantity in the expression above. However, note that the quantity to remove is a set of a measure zero on the integration domain, so by the continuity and finiteness of $f(\vec{r})$ everywhere, this subtraction is negligible. This means that

$$E[\prod F(S)] \to \prod_{B_i \in B(S)} E[F^{B_i}] \tag{2.13}$$

This means that if S and S' share the same integers $B(S) = B(S')$, then $E[\prod F(S)] = E[\prod F(S')]$ in the large N limit, a fact which we will use often.

As noted, for odd B, F_{jk}^B averages to zero due to the random phase assumption, and the present analysis makes it clear that similarly the expectation value $E[\prod F(S)]$ of a pattern S with at least one odd element of $B(S)$ is also zero. Thus we adopt a similar adjustment for the definition of $E[\prod F(S)]$, such that it is their typical magnitude, or the width of their distribution around zero, rather than their true average. Due to 2.13, this redefinition is automatic from the definition of $E[F^B]$ which includes the absolute value.

We can now rewrite the spectrum moments in terms of these expectation values:

$$\sigma^{(p)} = \sum_{u=2}^{2p} 2^{-u} \sum_{S(u)} \frac{\Omega'(S(u))}{D(S(u))} A_{S(u)} E\Big[\prod F(S(u))\Big] = \sum_{u=2}^{2p} 2^{-u} \sum_{S(u)} \frac{\Omega'(S(u))}{D(S(u))} A_{S(u)} \prod_{B_i \in B(S)} E[F^{B_i}] \tag{2.14}$$

2.3.12 Isolating and Quantifying N Dependence to Identify $\{S_t\}$

$\Omega'(S)$, $D(S)$, the number of terms in the sum over u, and the number of patterns i.e., the number of terms in the sum $\sum_S$ do not depend on N. Thus, the expression for $\sigma^{(p)}$ in 2.14 is an N-independent number of summands, each corresponding to a pattern and comprised of an N-independent prefactor multiplied by $A_S E[\prod F(S)]$. Therefore in the large N limit, if a pattern S corresponding to an $A_S E[\prod F(S)]$ with $\Omega'(S) \neq 0$ scales less quickly with N than $A_{S'} E[\prod F(S')]$ for some other pattern S' with $\Omega(S') \neq 0$, then the term contributing to $\sigma^{(p)}$ that corresponds to the pattern S may be safely discarded when N becomes arbitrarily large. We now identify the special set of patterns S_t which have contributions to $\sigma^{(p)}$ that have the largest scaling with N among all patterns, meaning that all other patterns can be ignored (provided $\Omega'(S_t) \neq 0$ for all such S_t, which we will confirm once they are identified).

The strategy to determine this final set of patterns $\{S_t\}$ will be to start with any pattern S and make a sequence of transformations $S \to S_1 \to S_2 ... S_t$, where each right arrow is a transformation which increases the N scaling of the associated terms which contributes to $\sigma^{(p)}$. We can only discard patterns with lesser N scaling contributions to $\sigma^{(p)}$ if, for the pattern with the greater N scaling of its contribution, the constant prefactor in front of its contribution is not zero, i.e., Ω' of that pattern is not zero. It is sufficient to make sure that this prefactor is non-zero only for the final sets, which we do because the transformations will be many and it would significantly complicate what follows to show this at every step. We will show that any initial pattern not in $\{S_t\}$ can be brought to an element of $\{S_t\}$ by a series of such transformations, establishing that this set contains the patterns corresponding

to all non-negligible contributions to $\sigma^{(p)}$, and we need only show for $\Omega'(S_t) \neq 0$ for all $S_t \in \{S_t\}$.

To evaluate the spectrum moments in this limit we will actually need to calculate $\Omega'(S_t)$, which is fortunately much easier than calculating $\Omega'(S)$ for any given pattern, and at that point it will be seen that it is not equal to zero. Also, note that since the existence of any term with a greater N scaling means the original term can be neglected, these transformations need not be unique - it needs only be established that there is at least one pattern with a contribution with a greater N scaling.

What we will find is that for the even moments, the final patterns $\{S_t\}$ will turn out to be the ones with $u = p$, and with each pairwise Hamiltonian paired up with exactly one other which has the same two subsystems as subscripts (and those subscripts appear nowhere else in the pattern). This pairing up is impossible for odd moments, for which there is an odd number of pairwise Hamiltonians; the important patterns for odd p will be the ones for which all but three pairwise Hamiltonians are paired up and the last three form a cycle of the form $H_{12}H_{23}H_{31}$ where the subsystems appearing in this triple do not appear anywhere else. We now derive these resuls.

2.3.13 Pattern Transformations

We now define the transformations that will together take any pattern S into S_t by increasing the N scaling of its associated contribution to the moments, $A_S E[\prod F(S)]$. We will define certain classes of patterns, denoted by $\{S_1\}$, $\{S_2\}$, etc., for which S_1 can be thought of as any pattern which conforms to the class $\{S_1\}$ in the same way S can be thought of as any pattern and $S(u)$ can be thought of any pattern with u elements. There will be several types of transformations $\rightarrow_i$ which, after repeated applications, take elements of $\{S_{i-1}\}$ into $\{S_i\}$ (where $\{S_0\}$ is simply the set of all patterns). Thus the sequence of patterns with increasing N scaling of $A_S E[\prod F(S)]$ is $S \rightarrow_1 S' \rightarrow_1 ... \rightarrow_1 S_1 \rightarrow_2 S_1' \rightarrow_2 ... \rightarrow_2 S_2 \rightarrow_3 S_2' \rightarrow_3 ... \rightarrow S_t$. We will describe the process which takes a general pattern not in $\{S_{i>0}\}$ all the way to $\{S_t\}$, but for

a starting pattern which happens to already conform to type $\{S_i\}$ for some i, we can simply begin at that point in the process (and of course for patterns which are already of type S_t, we let them be). There will only be six kinds of transformations, but each one in general will be applied many times.

The desired N-scaling properties of the transformations, that it increases the N scaling of the associated contribution to $\sigma^{(p)}$, is established if the following holds for every transformation:

$$\frac{A_{S_i} E[\prod F(S_i)]}{A_{S_i'} E[\prod F(S_i')]} \to 0 \tag{2.15}$$

where the arrow indicates the large N limit. After defining each type transformation, we will show that this limit holds. To do so, we will require certain conditions on the F_{jk} and their spatial distribution for large N; we call these the F-Conditions, and will tabulate them as we go along. Only one will not follow from the central limit theorem and insead may depend sensitively on the details of $F_{jk} = F(\vec{r}_{jk})$. We will analyze this condition after it has been established, but suffice it to note here that it is satisfied for $\alpha = 1$ and $D = 2$ or 3, the case corresponding to coupling to a common boson bath which is highly physical relevant.

We now define each of the $\to_i$ separately, and establish 2.15 for each.

2.3.14 $\to_1$

The first kind of transformation we will make to a generic initial pattern S, called $\to_1$, reduces an element of $B(S)$ by 2, increases u by 2, and adds a new element equal to 2 to $B(S)$. If there are no elements of $B(S)$ greater than 2 then it is already of type S_i for $i \geq 1$, and we begin the sequence of transformations at the appropriate, later place. If not, then this transformation will turn one pairwise factor in $\prod_{B_i \in B(S)} F_{jk}^{B_i}$ with $B_i > 3$ from $F_{jk}^{B_i}$ into F_{jk}^{B-2} as well as add a new pairwise Hamiltonian squared, $H_{j'k'}^2$, where j' and k' appear nowhere else in the pattern. For example, $\to_1$ takes $F_{12}F_{23}^2F_{14}^5 \to_1 F_{12}F_{23}^2F_{14}^3F_{56}^2$ (recall that the subscripts should be thought of as slots, not specific references; in the example, 1 could be any subsystem, 2 could be any other subsystem, and so on). If there is more than one $B \geq 3$, or

some B is still ≥ 3 after this manipulation, we repeat it on each such $B \geq 3$ until all none of the elements of B are greater than 2. When this is the case, we will have transformed the initial state S into an element of $\{S_1\}$; the defining characteristic of a pattern S_1 is that all elements of $B(S_1)$ are either 1 or 2.

Now we must show that $\rightarrow_1$ satisfies (2.15). Recall that $A_S = \mid \{j\} \mid \times \mid \pi \mid$, and that $\mid \pi(u) \mid = u!$ is independent of N. Thus the N-dependence of A_S is the same as the N-dependence of $\mid \{j\} \mid$, which for $\mid \{j\}(u) \mid$ is the number of ways to choose u subsystems from a set of N, $\binom{N}{u}$. For large N, $\binom{N}{u}$ scales with N as N^u; in our notation this is written $\left|\{j\}\right| \sim N^u$. Since $|\pi|$ is independent of N and $\sim$ means only the N scaling, then it is also true that $A_{S(u)} \sim N^u$. Thus for $S \rightarrow_1 S'$ the ratio of the contributions from S and S' (discarding constant prefactors which do not effect the N scaling), (2.15), is

$$\frac{A_{S(u)} E[\prod F(S(u))]}{A_{S'(u+2)} E[\prod F(S'(u+2))]} \rightarrow \frac{N^u \prod_{B_i \in B(S(u))} E[F^{B_i}]}{N^{u+2} \prod_{B_i \in B(S(u+2))} E[F^{B_i}]} = \frac{E[F^{B_i}]}{N^2 E[F^{B_i - 2}] E[F^2]} \rightarrow 0 \tag{2.16}$$

Thus an F-condition is that, for any integer $B \geq 3$, $E[F^B](N^2 E[F^{B-2}]E[F^2])^{-1} \rightarrow 0$.

After exhaustive (no more $B > 2$) application of $\rightarrow_1$ to a starting pattern S, the N scaling of the contribution to $\sigma^{(p)}$ will have increased. Thus if it can be shown that $\Omega'(S_1) \neq 0$ for the patterns in $\{S_1\}$, then in the large N limit the contribution to $\sigma^{(p)}$ from the terms in 2.14 corresponding to patterns not in $\{S_1\}$ can be neglected. As noted above, we will put off showing that $\Omega'(S_i) \neq 0$ until we arrive at the final set $\{S_t\}$, since showing that $\Omega'(S_t) \neq 0$ is sufficient to conclude that only the contributions from those patterns are important for large N. We now define the next transformation, $\rightarrow_2$, which will, after exhaustive application to all patterns in $\{S_1\}$, create a new class of patterns S_2.

2.3.15 $\rightarrow_2$

To apply $\rightarrow_2$ to a pattern S_1 in order to make a new pattern S_1', we choose a subsystem reference in a subscript of an F factor raised to the second power, then look to see if that same subsystem also appears in the subscript of any other F factor; if it does, we unlink these references via $F_{12}^2 F_{23}^b \rightarrow F_{12}^2 F_{34}^b$ where b can be 1 or 2. Note that the two F factors cannot share both subscripts, as otherwise the pattern would have an element of $B(S)$ greater than 2 and therefore not be in $\{S_1\}$. This transformation increases u by 1 and adds a new element to S which contains the all the locations (equal to b in number) in the term $\prod F$ where the new subsystem reference appears. It also modifies the existing element of S corresponding to 2 in the example above, removing the b locations where the new subsystem is to appear in S'.

To establish that this transformation satisfies 2.15, note that $B(S) = B(S')$, and thus $E[\prod F(S)] = E[\prod F(S')]$. Since u increases by 1, $A_{S_1}/A_{S_1'} = 1/N$. Together, then, in this case 2.15 becomes

$$\frac{A_{S_1} E[\prod F(S_1)]}{A_{S_1'} E[\prod F(S_1')]} \rightarrow 1/N \tag{2.17}$$

establishing that if $\Omega'(S_1') \neq 0$, the terms in 2.14 corresponding to elements of $\{S_1\}$ but not $\{S_2\}$ can be ignored for large N. If we apply this transformation repeatedly until it is impossible to apply again, then the resulting pattern S_2 will have no subsystem references in subscripts of any F_{jk} raised to the second power which also appear anywhere else in the pattern. This is the defining propery of $\{S_2\}$ (along with the inherited property of S_1, that all elements of $B(S_{1,2})$ equal to 2 or 1).

2.3.16 $\rightarrow_3$

Like the transformations before it, the next transformation $\rightarrow_3$ acts on a pattern $S_2 \in \{S_2\}$ and, after exhaustive application, produces a new pattern $S_3 \in \{S_3\}$. For one such transformation $\rightarrow_3$, we identify any subsystem reference which appears 4

or more times among the subscripts of the F_{jk} which are raised to the first power i.e., find an element of S_2 which has 4 or more elements (note that for every pattern S_2, no subsystem which appears in the subscript of a F_{jk} factor raised to the first power also appears in the subscript of an $F_{j'k'}$ factor raised to the second power, so a subsystem that appears in an F^1_{jk} can only appear in other $F^1_{j'k'}$). In its basic form, this transformation takes $H_{12}H_{23}H_{31}H_{14}H_{45}H_{51} \rightarrow_3 H_{12}H_{23}H_{31}H_{45}H_{56}H_{64}$. More generally, for the chosen set of 4 or more locations of pairwise Hamiltonians, pick any one location in it, and follow it to another pairwise Hamiltonian by arbitrarily picking any other pairwise Hamiltonian in which the other subsystem appearing as a subscript in the original pairwise Hamiltonian at the current location also appears. Because, as noted, no subscript appearing in an F_{jk} appears also in an $F^2_{j'k'}$, when moving between pairwise Hamiltonian locations this way we will never be led to an F^2. Since each subsystem must appear at least twice this hopping will repeat indefinitely. We add the restriction that it does not repeat locations, and choose to terminate this process when the same pairwise Hamiltonian at which originating subsystem reference occurred is landed on again. This must happen despite the non-repeatability condition because, like all subsystem references, the original subsystem reference must appear at least twice (actually at least 4 times, but this is not necessary for the conclusion at hand). Once this loop is completed, we take the last two locations at which the original subsystem which initiated the process appeared in the loop, remove those two locations from their associated element of S, and add a new element to S with these two locations in it.

Thus u is increased by 1, and accordingly the N dependence of A_S increases by N. It does not alter the set $B(S)$ and therefore also does not alter $E[\prod F(S)]$, and thus if $S_2 \rightarrow_3 S_2'$,

$$\frac{A_{S_2} \mid E[\prod F(S_2)] \mid}{A_{S_2'} \mid E[\prod F(S_2')] \mid} \rightarrow 1/N \tag{2.18}$$

This establishes that the contribution to $\sigma^{(p)}$ from all patterns in $\{S_2\}$ but not $\{S_3\}$ vanish compared to the contribution from the patterns in $\{S_3\}$, the set formed by exhaustively applying $\rightarrow_3$ to all elements of $\{S_2\}$. This transformation can no longer be applied once none of the subsystems appearing in an F^1_{jk} appear in more

than 2 locations; this property characterizes S_3 (along with the properties of its supersets $\{S_1\}$ and $\{S_2\}$). Since each subsystem appearing among the F_{jk}^2 also appears only twice, then all collections of locations in a pattern S_3 have exactly two elements i.e., each subsystem appears exactly twice amongst all of the subscripts in any associated pattern $F_{j_1k_1}...F_{j_pk_p}$.

2.3.17 $\rightarrow_4$

To set the foundation of the next transformation, $\rightarrow_4$, observe that the location hopping done according to the prescription above, which required arbitrary choices, is non arbitrary for patterns S_3 because each subsystem reference only has one pair it is attached to, meaning there is a unique location to hop to from any given pairwise Hamiltonian F_{jk} (provided we know which of its two subsystem references lead us there). Given an initial subsystem reference, there is still ambiguity as to which of its locations to begin at, but the one other location will inevitably be the destination, so this ambiguity amounts only to an arbitrary direction of the travel while the path itself is uniquely defined.

Choose a collection of locations in S_3 (now reduced to only pairs of locations) and find the path that can be traced through the pairwise Hamiltonian locations that begins and ends at the two locations in the chosen collection (according to the hopping prescription described above); if it is longer than 4 locations long, adjust it as follows: $F_{12}F_{23}F_{34}F_{45}F_{51} \rightarrow F_{12}F_{23}F_{31}F_{45}^2$. More precisely, we take the last two pairwise Hamiltonians in the path and replace both of their two subscripts with the same two subsystem references which do not appear anywhere else in the pattern, then adjust one of the references in the third-to-last pairwise Hamiltonian along the path so as to make sure each subsystem still appears twice.

This transformation does not alter u, and therefore also not $A_{S_3(u)}$. Thus all change in the N scaling of the contribution to the moments comes from the expectation value $E[\prod F(S_3(u))]$, which goes from $E[F]^l \rightarrow E[F]^{l-2}E[l^2]$ where l is the number

of pairwise Hamiltonians in the loop. As usual we must test the condition (2.15),

$$\frac{A_{S_3}E[\prod F(S_3)]}{A_{S_3'}E[\prod F(S_3')]} \to \frac{E[F]^l}{E[F]^{l-2}E[F^2]} \to \frac{E[F]^2}{E[F^2]} \tag{2.19}$$

As discussed when defining the expectation values $E[S]$, by the central limit theorem $E[F] \to \sqrt{E[F^2]/\binom{N}{2}}$ making the above ratio go like $1/N^2$ and therefore vanishing for large N. This establishes that $\to_4$ increases the N dependence of the pattern's associated contribution to $\sigma^{(p)}$.

Once $\to_4$ has been applied exhaustively to all elements of $\{S_3\}$, we arrive at the set $\{S_4\}$ which contains only patterns with such loops of length 4 or 3 among the F^1_{jk} (all longer ones have been removed by the transformations, a path of length 2 would mean that F factor is raised to the second power, and paths of length 1 are not allowed since the same subsystem cannot appear as both indices of a pairwise Hamiltonian).

2.3.18 $\to_5$

The next transformation is to take all paths of length 4 and pair them up via $F_{12}F_{23}F_{34}F_{41} \to F^2_{12}F^2_{34}$. Again the A_S factor is unaltered because u does not change, but the expectation value increases in N scaling:

$$\frac{A_{S_4}E[\prod F(S_4)]}{A_{S_4'}E[\prod F(S_4')]} \to \frac{E[F]^4}{E[F^2]^2} \to \frac{\left(\sqrt{E[F^2]/\binom{N}{2}}\right)^4}{E[F^2]^2} \to \frac{1}{N^4} \to 0 \tag{2.20}$$

Thus after exhaustive application of $\to_5$ to all elements of $\{S_4\}$, we have a new set $\{S_5\}$ in which every pattern has some number of F^2_{jk} all with different j and k, and some number of cycles of three $F_{j_1j_2}F_{j_2j_3}F_{j_3j_1}$ with $j_{1,2,3}$ all different from each other and from all of the j, k appearing in any F^2_{jk} in the same pattern.

2.3.19 $\rightarrow_6$

The final transformation, $\rightarrow_6$, pairs up these loops of length 3 via $F_{j_1j_2}F_{j_2j_3}F_{j_3j_1}F_{k_1k_2}F_{k_2k_3}F_{k_3k_1} \rightarrow F^2_{j_1k_1}F^2_{j_2k_2}F^2_{j_3k_3}$. Again u does not change, so for $\rightarrow_6$ the condition 2.15 is

$$\frac{A_{S_4}E[\prod F(S_4)]}{A_{S'_4}E[\prod F(S'_4)]} \rightarrow \frac{E[F]^6}{E[F^2]^3} \rightarrow \frac{\left(\sqrt{E[F^2]/\binom{N}{2}}\right)^6}{E[F^2]^3} \rightarrow \frac{1}{N^6} \rightarrow 0 \tag{2.21}$$

Once $\rightarrow_6$ is exhaustively applied, there will be 0 or 1 paths, or loops, of length 3 left. All other pairwise Hamiltonians will be paired up with a second pairwise Hamiltonian in the sense that they share the same two subscripts, and no other pairwise Hamiltonians in the pattern share any of these subscripts. These are the final patterns, denoted by $\{S_6\} = \{S_t\}$.

p is the total number of pairwise Hamiltonians, so if p is even then this final transformation process must be able to pair up every single triplet. If p is odd this transformation can no longer be applied when there is one length 3 cycle left, as there is not another length 3 cycle to pair it with.

Now, as promised, we make sure $\Omega'(S_t) \neq 0$ for the patterns in $\{S_t\}$ for the odd and even case, in order to establish that it is safe to neglect all other patterns in the large N limit.

2.3.20 $\Omega'(S_t) \neq 0$

Recall that $\Omega'(S_t)$ is the number of arrangements $|q\rangle$ of the u subsystems appearing in S_t for which $\langle q|\prod H(S_t)|q\rangle \neq 0$. Conceptually, this definition is best imagined for a triplet with a specific subsystem selection $\{j\}$ and assignment π in mind, which allows the related term $\prod H(S_t) = H_{j_1k_1}...H_{j_pk_p}$ to have specific subsystems in its subscripts, but since the value of $\Omega'(S_t)$ depends only on the pattern part of the triplet, it is written as a function of its pattern only. For patterns in $\{S_t\}$, the term $\prod H(S_t) = H_{j_1k_1}...H_{j_pk_p}$ takes the form $H^2_{j_1j_2}...H^2_{j_{p-2}j_p}$ for even p and

$H^2_{j_1 j_2}...H^2_{j_{p-4}j_{p-3}}H_{j_{p-2}j_{p-1}}H_{j_{p-1}j_p}H_{j_p j_{p-2}}$ for odd B. The different H_{jk} in $\prod H(S_t)$ can appear in any order, for example pairs need not occur consecutively; but since each subsystem appears in no more than 2 places for the patterns $\{S_t\}$, this ordering does not matter. For these patterns, the only restriction on $|q\rangle$ such that $\langle q|\prod H(S_t)|q\rangle \neq 0$ is that, for each pair of subsystems appearing together in the subscripts of two H factors in $\prod H(S_t)$, they are oppositely aligned in $|q\rangle$. For even p, all subsystems only ever appear in an H_{jk} with one other subsystem, so for every such pair there are two options: one up and the other down, or vice versa. Thus for even p, $\Omega'(p) = 2^{p/2}$ since there are $p/2$ pairs and there are two choices for each pair.

For odd p, there are $p-3$ pairs as in the even p case, which together contribute a factor of $2^{(p-3)/2}$ to $\Omega'(p)$. For the loop $F_{12}F_{23}F_{31}$ there are only two choices: up down up, or down up down. The loop then contributes a factor of 2, leading to $\Omega'(S_t) = 2 \times 2^{(p-3)/2} = 2^{(p-1)/2}$ for odd p. Since in neither case, for any integer $p \geq 1$, $\Omega'(S_t)$ is zero, then we have now proven that when F_{jk} obeys the F-conditions, only terms in 2.6 corresponding to the patterns in $\{S_t\}$ contribute meaningfully for large N. In other words, for large N, the sum over S can be replaced by a sum over S_t without significantly changing the value of the moment $\sigma^{(p)}$.

Using this fact, we can actually estimate the moments in terms of $E[F^2]$. It remains to calculate $D(S_t)$, $A(S_t)$, and $E[\prod F(S_t)]$, as up to now we've only calculated the N-scaling of the latter two quantities, and not at all discussed the value of $D(S_t)$. Then we will know the contribution to the moment from each S_t, and upon summing over them will have found the moments themselves.

2.4 Evaluating $\sigma^{(p)}$

For ease of notation, from here on out we write the relevant patterns S_t simply as S, which before stood for any pattern but now will stand for S_t, a pattern with the property that $B(S) = \{2, 2, ..., 2\}$ and $u = p$ (these two facts alone specify $\{S_t\}$ for even p), or $B(S)\{2, 2, ..., 2, 1, 1, 1\}$ and $u = p$ (for odd p). We have already

calculated $\Omega'(S_t = S)$ to be equal to $2^{(p-r)/2}$ where $r = 0$ for even p and 1 for odd p; in order to make numerical estimations of $\sigma^{(p)}$ for a specified coupling $F_{jk} = F(\vec{r_{jk}})$ and other parameters, we now evaluate the remaining necessary quantites in the special case of these final patterns $\{S_t\}$ (now referred to simply as S): $D(S)$, $A(S)$, and $E[\prod F(S)]$.

2.4.1 $D(S_t)$

$D(S)$ is the number of triplets S' there are that construct the term $\prod H(S)$. This compensates for overcounting in the sum over triplets, which there are more of than ways to choose p pairs $\{\{j_1k_1\}...\{j_pk_p\}\}$ for the terms $\sum_{(j_1k_1)} ... \sum_{(j_pk_p)} H_{j_1k_1}...H_{j_pk_p}$. In the final patterns, each subsystem appears twice and each subsystem appears in exactly the same locations as one other subsystem; thus if π were altered to switch the collections in S to which two such pairs of subsystems are assigned, the constructed term $\prod F(S)$ would not change. For example, for $F_{12}^2F_{34}^2$, $S = \{\{1,2\},\{1,2\},\{3,4\},\{3,4\}\}$ and $\{j\} = \{1,2,3,4\}$, π could either assign the first listed element of S to 1 and the second to 2, or vice versa; a similar consideration for subsystems 3 and 4 leads to the conclusion that $D = 4$, a factor of two for each F^2. For even p there are $p/2$ F^2 factors, and so $D(S) = 2^{p/2}$. For odd p there are $(p-3)/2$ such pairs, leading to $D(S) = 2^{(p-3)/2}$. There is no contribution from the cycle of 3 in the odd S since none of the subsystems in it are assigned to exactly the same locations of any of the other subsystems in it.

2.4.2 $A(S)$

$A(S)$ is the number of terms in $\sum_{\{j\}} \sum_{\pi}$, the sum over subsystem selections and assignments, which are compatible with S (according to the Compatibility Conditions). Since $\{j\}$ is unordered, there are $\binom{N}{u}$ of them compatible with a pattern $S(u)$. Being the enumeration of ways to pair up u objects, $\pi(u) = u!$. Thus $A_{S(u)} = \binom{N}{u}u! = N!/(N-u)!$, or for the final patterns with $u = p$, $A_S = N!/(N-p)!$. Since we are taking $N >> p$, we will approximate this as $A_S \to N^p$.

2.4.3 $E[\prod F(S)]$

As previously noted, $E[\prod F(S)] = \prod_{B_i \in B(S)} E[F^{B_i}]$. Given that $B(S_t) = \{2, 2, ..., 2\}$ for even p, $E[\prod F(S_t)] = E[F^2]^{p/2}$ in that case. For odd p, $B(S_t) = \{2, 2, ..., 2, 1, 1, 1\}$ and $E[\prod F(S_t)] = E[F^2]^{(p-3)/2} E[F]^3$.

2.4.4 The Number of Final Patterns

We can now finally conclude that the contribution to the spectrum moments from any final pattern S_t with an even value of p is

$$2^{-p} D(S)^{-1} \Omega'(S) A(S) E[\prod F(S)] = 2^{-p} 2^{-p/2} 2^{p/2} N^p E[F^2]^{p/2} = (N E[F^2]^{1/2}/2)^p \tag{2.22}$$

and for odd p the contribution from each of the patterns is

$$\begin{aligned} 2^{-p} D(S)^{-1} \Omega'(S) A(S) E[\prod F(S)] &= 2^{-p} 2^{-(p-3)/2} 2^{(p-1)/2} N^p E[F^2]^{(p-3)/2} E[F]^3 \\ &= 2 \frac{E[F]^3}{E[F^2]^{3/2}} (N E[F^2]^{1/2}/2)^p \end{aligned} \tag{2.23}$$

The expression for the spectrum moments contains a sum over S so we must determine how many such final patterns there are, each of which contributes the quantity above to the spectrum moments (with the alternate expressions used accordingly for even or odd moments). We have already determined that this number does not depend on N, but to estimate the spectrum moments we must know it exactly.

For even p, in a final pattern each pairwise Hamiltonian must be paired up with another so the number of patterns is simply number of ways to pair up p objects. There are $\binom{p}{2}$ choices for the first pair, $\binom{p-2}{2}$ for the second, and so forth, but we must divide by $(p/2)!$ since it doesn't matter which order the pairs are selected in. Thus the number of patterns is $\binom{p}{2}\binom{p-2}{2}...\binom{4}{2}/(p/2)! = p!/((p/2)!2^{p/2})$.

For odd p, first we pick which three pairwise Hamiltonian locations will be in the triplet; there are $\binom{p}{3}$ choices for this. We then have to pair up the remaining $p-3$ things, which as established in the even case there are $(p-3)!/(((p-3)/2)!2^{(p-3)/2})$ ways to do. Thus for odd moments there are $\binom{p}{3}((p-3)!/(((p-3)/2)!2^{(p-3)/2})) = \frac{p!}{3\times 2^{(p-1)/2}((p-3)/2)!}$ patterns.

Thus, finally, we can calculate the moments by multiplying the contribution from each pattern by the number of such patterns. For even p then,

$$\sigma^{(p)} \to \frac{p!}{(p/2)!2^{p/2}}(NE[F^2]^{1/2}/2)^p = \frac{p!}{(p/2)!}\big(N(E[F^2]/8)^{1/2}\big)^p \tag{2.24}$$

and for odd p the typical magnitude of the moment is

$$\sigma^{(p)} \to \frac{p!}{3\times 2^{(p-1)/2}((p-3)/2)!}\frac{2E[F]^3}{E[F^2]^{3/2}}(NE[F^2]^{1/2}/2)^p = \frac{2^{3/2}E[F]^3}{3E[F^2]^{3/2}}\frac{p!}{((p-3)/2)!}(N(E[F^2]/8)^{1/2})^p \tag{2.25}$$

2.5 The Non-trivial F-Condition

When showing that the transformations $\to_i$ increased the N-dependence of the contribution to the spectrum moments, we assumed certain conditions on the coupling function $F(\vec{r})$. Most required properties follow from the central limit theorem and therefore applies to any $F_{jk} = F(\vec{r}_{jk})$ as long as they have bounded variance (when considered to be sampling from a distribution of pairs of positions drawn from some distribution $\rho(\vec{r}_{jk})$). The F-Condition that depends more carefully on the functional form of $F(\vec{r})$, due to the transformation $\to_1$, is that

$$\frac{E[F^B]}{N^2E[F^2]E[F^{B-2}]}0 \tag{2.26}$$

for all $B \geq 3$. It is difficult to say for general $F(\vec{r})$ when this is satisfied and when it is not, but it can be determined in some cases through the use of inequalities. If we can find a quantity that is greater than $E[F^B](N^2E[F^2]E[F^{B-2}])^{-1}$ that vanishes in the large N limit, then so will $E[F^B](N^2E[F^2]E[F^{B-2}])^{-1}$. We will implement

this strategy for $F(\vec{r})$ of the form $\mid F(\vec{r}) \mid \sim r^{-\alpha}$ for some positive α. For such functions, the the minimum magnitude of F_{jk} of all pairs, which we denote by $F_{\min}$, is a constant i.e. $F_{\min} \sim N^0$. This is because the distribution function $\rho(\vec{r})$ is taken to scale with N such that the subsystem spatial density is constant. We similarly define $F_{\max} \sim N^{-\alpha/D}$ where D is the number of spatial dimensions. This is because $r_{\max} \sim N^{1/D}$, again because of the subsystem distribution scaling.

Using these observations, we make another:

$$E[F^B] = \binom{N}{2}^{-1} \sum_{(jk)} \left|F_{jk}\right|^B \leq \binom{N}{2}^{-1} \sum_{(jk)} F_{jk}^{B-2} F_{\max}^2 = E[F^{B-2}] F_{\max}^2 \tag{2.27}$$

and similarly

$$E[F^B] \geq E[F^{B-2}] F_{\min}^2 \tag{2.28}$$

Using the first inequality, we can re-express the non-trivial F-Condition 2.26 as

$$\frac{E[F^B]}{N^2 E[F^2] E[F^{B-2}]} \leq \frac{F_{\max}^2}{N^2 E[F^2]} \tag{2.29}$$

and thus a sufficient but not necessary condition for the satisfaction of this F-Condition is that $F_{\max}^2/(N^2 E[F^2]) \to 0$. We will explore the particular case of $\alpha = 1$ since it corresponds to the interaction of a common boson bath with the collective system, but these inequalities could be examed for other coupling constants, and if not satisfied then perhaps a tighter upper bound could be found.

As derived in Appendix A, for $\alpha = 1$ and in D spatial dimensions (for $D = 2$ or 3 which are the dimensions of interest), $E[F^2] \sim N^{-2/D} \log(N)^{3-D}$. Since $F_{\max} \sim N^{-\alpha/D} = N^{-1/D}$, then $F_{\max}^2/(N^2 E[F^2])$ scales as $\sim N^{-2/D} N^{-2} N^{2/D} \log(N)^{D-3} = N^{-2} \log(N)^{D-3} \to 0$ for the allowed values of D (2 and 3). Thus the inequality is established for $\alpha = 1$.

We make a quick sidenote for the sake of a future reference, when we find the spectrum of the single qubit Hamiltonian in Chapter 3: there we will meet the same F-Condition but with N^2 replaced by N. As can clearly be seen, in this case the

fraction still vanishes, thus establishing the F-Condition in that case as well.

2.6 The Energy Spectrum

We now use these spectra to calculate the spectral density function via (2.2). We approximate the spectra by truncating the sum at some cutoff $p_{\max}$ which well approximates the function near its center and with the validity extending into the tails for larger $p_{\max}$. For example, a gaussian distribution can be well approximated this way to around 3 times its width with $p_{\max} = 6$.

First, let us demonstrate that the odd moments become negligible for large N, making the spectrum asymptotically symmetric. We do this by grouping consecutive even and odd terms in (2.2) as so:

$$\begin{aligned} F(\xi) &= \sum_{p=0}^{\infty} \frac{(i\xi)^p}{p!}\sigma^{(p)} = \sum_{p=0}^{\infty} \Big(\frac{(i\xi)^{2p}}{(2p)!}\sigma^{(2p)} + \frac{(i\xi)^{2p+1}}{(2p+1)!}\sigma^{(2p+1)} \Big) \\ &= \sum_{p=0}^{\infty} \frac{(i\xi)^{2p}}{(2p)!} \Big(1 + \frac{i\xi}{2p+1}\frac{\sigma^{(2p+1)}}{\sigma^{(2p)}} \Big) \sigma^{(2p)} \end{aligned} \tag{2.30}$$

We now show that the term in parenthesis, $1 + i\xi\sigma^{(2p+1)}/(\sigma^{(2p)}(2p+1))$, asymptotes to 1 for large N, making the spectrum's fourier transform asymptote to

$$F(\xi) = \sum_{p=0}^{\infty} \frac{(i\xi)^{2p}}{(2p)!}\sigma^{(2p)} = \sum_{p=0}^{\infty} \frac{(-\xi^2)^p}{(2p)!}\sigma^{(2p)} \tag{2.31}$$

To come to this conclusion, let us recall the N scaling of the various quantities appearing in $i\xi\sigma^{(2p+1)}/(\sigma^{(2p)}(2p+1))$. Only the moments are not constant with respect to N. For even p, $\sigma^{(p)} \sim N^p E[F^2]^{p/2}$, and for odd p, $\sigma^{(p)} \sim N^p E[F^2]^{p/2} E[F]^3 / E[F^2]^{3/2}$. Thus the N scaling of the ratio of the even and odd moments is

$$\frac{\sigma^{(2p+1)}}{\sigma^{(2p)}} \sim \frac{N^{2p+1} E[F^2]^{p+1/2}}{N^{2p} E[F^2]^p} \frac{E[F]^3}{E[F^2]^{3/2}} = N \frac{E[F]^3}{E[F^2]} \tag{2.32}$$

By the central limit theorem (and the definition of $E[F]$ as its average magnitude,

not it's average value of 0), $E[F] \sim E[F^2]^{1/2}/N$ and therefore

$$\frac{\sigma^{(2p+1)}}{\sigma^{(2p)}} \sim N\frac{E[F^2]^{3/2}}{E[F^2]N^3} \sim \frac{E[F^2]^{1/2}}{N^2} \tag{2.33}$$

Since the coupling strength decreases with distance, and as we increase N we increase the total system size by adding new subsystems to the 'edge' of the system, this ratio indeed vanishes for large N - in fact, to conclude as such we require only that the average magnitude of F_{jk} does not grow with N. Since the ratio of the even and odd moments contained all the N-dependence of $i\xi\sigma^{(2p+1)}/(\sigma^{(2p)}(2\mathrm{p}+1))$, then this fraction vanishes for large N, justifying the exclusion of the odd moments in the same limit.

Using the value of $\sigma^{(p)}$ for even p, we finally have that

$$\begin{aligned} F(\xi) \to \sum_{p=0}^{\infty} \frac{(i\xi)^{2p}}{(2p)!}\sigma^{(2p)} &= \sum_{p=0}^{\infty} \frac{(i\xi)^{2p}}{(2p)!}\frac{(2p)!}{p!}\big(N(E[F^2]/8)^{1/2}\big)^{2p} = \sum_{p=0}^{\infty} \frac{\big(-\xi^2N^2E[F^2]/8\big)^p}{p!} \\ &= e^{-\xi^2N^2E[F^2]/8} \end{aligned} \tag{2.34}$$

which is a gaussian function of ξ. The fourier transformation of a gaussian is another gaussian, and so the eigenvalue distribution is a gaussian with width equal to $NE[F^2]^{1/2}/2$.

The specific magnitude of $E[F^2]$, and therefore $f(\lambda)$, will depend on the particulars a given coupling for some specific system. However, the N scaling of these quantities depends only on the exponent of the coupling α and the geometry of the computer. Thus we may speak generally about the N scaling of the moments, and therefore the size of the distribution, without specifying anything about the system other than the exponent on the coupling and whether it lives in 2 and 3 dimensions (assuming it has a regular shape, such as a box, sphere or gaussian cloud). As noted previously, and as shown in Appendix A, for $\alpha = 1$ and $D = 2$ or 3, $E[F^2] \sim N^{-2/D}\log(N)^{3-D}$. Therefore the $2p$-th moment scales as $N^{2p}E[F^2]^p \sim N^{2p}N^{-2p/D}\log(N)^{p(3-D)}$. We can compare the different moments to each other by raising them to the $1/(2p)$

power, making them all have the same units. Upon doing this, we see that they all scale uniformly with N as $N^{1-1/D}\log(N)^{(3-D)/2}$. Thus the entire distribution uniformly grows in size with N while preserving its shape, and the scaling of its size factor is $N^{1-1/D}\log(N)^{(3-D)/2}$. Since all moments scale in the same way, we can take this scaling to be the scaling of the timescale of the dynamics generated by the interaction Hamiltonian (once raised to the $1/2p$ power, the $2p$-th moment has units of inverse time, meaning that as it increases, the timescale of the dynamics becomes shorter). We will have more to say about what this final scaling result can tell us about such systems in the following chapters.

2.7 Conclusion

We have now estimated the eigenvalue spectrum density for a wide class of pairwise interaction Hamiltonians for an ensemble of two-level systems, and expressed them in terms of the typical value of F_{jk}^2 averaged over all pairs. The constraints specified, including the large N limit, are physically relevant. We now explore what can be learned from this approximate spectrum about the physical systems which obey these constraints. Because the approximation error decreases without bound with the system size, and systems with $N \sim 10^{26}$ and much larger are commonplace, we anticipate that observations made about such systems according to its spectrum might have relevance to physical systems which can be found or created.

3 Implications for Quantum Computing

3.1 Introduction

A quantum computer is a physical system which obeys the constraints under which the spectrum calculation of Chapter 2 applies. This can be said in generality because the two-level systems, or qubits, which comprise the computer must be able to interact with experimental instruments which implement the algorithm; if the qubits were completed isolated from another, no entanglement or multi-qubit gates could be performed, which are essential for quantum computation. This means the qubits must share an environment (in some sense any two pieces of the same universe share an environment, but the needs of quantum computation demands that they share a very immediate environment). Although pains may be taken to minimize the interaction with their environment, it is both an experimental fact and a theoretical necessity that some interaction always takes place. This interaction is due to the exchange of bosons, the force carriers of our universe. For an atomic system, for example, photons would be such a boson. It is well known that when two-level systems interact with a common boson field, the effective Hamiltonian includes the exchange type, defined in Chapter 1, with a coupling exponent $\alpha = 1$ i.e., the interaction strength between two qubits falls off like one over their separation to the first power. As observed there, all necessary conditions for the derivation of the asymptotic spectrum moments are met for $\alpha = 1$. Not all errors afflicting the quantum computer will have this form, but since this one grows without bound then either this error dominates for large N or there is an even greater error source. Either way, if this error alone is enough to cause the breakdown of scalable fault-tolerance, then it will fail no matter if there is some greater error or not.

For any specific implementation of a quantum algorithm, which takes a fixed and finite amount of time, if the error rate is low enough it will be likely to succeed. This means the output will be likely to be correct (given the random collapse during

the measurement of the final quantum state of the computer, most algorithms aim to make the correct answer likely, not guaranteed - this is particularly useful for problems such as integer factorization whose answer is difficult to identify but can be easily verified). Much effort is expended to reduce the coupling of the qubits to environmental degrees of freedom and each other in the pursuit of controlling the system. Error rates due to unwanted interactions can never be zero, but can be mitigated so that the decay or decoherence it causes is not fatal to the function the quantum control protocol.

As discussed in Chapter 1, making the errors so slow that their time scale is not relevant over the entire length of the computation has not proven to be a viable solution - instead quantum error correction techniques are implemented as the algorithm goes along. With this approach, the timescale of the errors need only be slow enough compared to the timescale of a single gate, not the entire computation. In addition to being a more practical way to approach error correction, this approach has the added benefit of mirroring the classical case.

In classical computing, algorithms are fundamentally classified based on how the time or space (memory) resources needed scale with the size of the problem or the size of the input. Although predicting precisely how long a given computation will take is difficult, discerning the scaling with input size for large inputs is much less so. The asymptotic scaling of the resources needed has proven to be a meaningful categorization; for example, the non-existence of an integer factorization algorithm which is polynomial in time is a fact which greatly concerns many scientists in many fields [11]. It's non-existence has not been proven but it also has not been found despite decades of intense search. Were such an algorithm found, it would have profound consequences - Shor partially solved this problem by identifying such an algorithm for a quantum computer [12], and it inspired the field of quantum computation. However the identification of such an algorithm would have, by itself, no implications for the actual factorization of integers, for who is to say that this polynomial algorithm is faster than our current exponential algorithms for any size of integer that any human ever has or will desire to factor? However, informed

by the historical trend of the discovery of algorithms, we expect this not to be the case. Although it need not be for the particular values of N which interest humans in their computations, the scaling of a process with N for large N has proven to be a good indicator of whether a computational problem is easy or hard even in practice. For example, although one could not in principle add any two given integers together using today's best computers, the limits of computation are not relevant for addition of the kinds of numbers humans wish to add. In the other extreme, the inability to factorize integers of a size which can easily be stored in modest computers is so reliably difficult that it forms the basis of the world's computer security, the RSA encryption scheme. Why the exponential-polynomial divide has proven such a good classification tool for the human computational scale, or more generally that the N scaling for arbitrarily large N has such meaning on our finite scale, remains somewhat a mystery. If this classification scheme were not so meaningful for quantum computers, that would be of interest. The N scaling arguments presented in this thesis imply that fault-tolerant algorithms are not scalable for arbitrarily large N, and therefore if quantum algorithms do prove useful in practice, a different classification scheme will be necessary. We will not attempt here to speculate on such a possible new classification scheme.

If one does hope to have fault-tolerant, scalable quantum computing, as many do, one must find a way to error correct during the computation such that errors do not accumulate, meaning that the same error correction protocol will work no matter how long the computation is. Of course, the error correction takes both time and space resources, but in the classical case it adds only a polynomial overhead, and therefore leaves algorithms unchanged with respect to the polynomial-exponential classification. As might be hoped, a similar result was found for quantum error correction in what is known as the Threshold Theorem [16, 17, 27]. According to this theorem, error correction with the desired properties (scalable with only polynomially many extra resources on top of the ideal algorithm) is possible under certain conditions. This result, and the generalizations that followed, is the source of much optimism in the field of quantum computing. Appendix C summarizes the

main ideas of the Threshold Theorem in greater detail.

Due to the inability to measure and therefore assess errors in a quantum system without destroying it, the possibility of active quantum error correction was long thought impossible; however, in the 1990s, it was demonstrated that it is possible to encode the state of one quantum two-level system into many quantum two-level systems in such a way that partial measurement can be made on the ensemble with enough information gleaned so as to assess and correct the error, but not so much as to have destroyed the quantum coherences which encode the state of the so-called logical qubit [45]. This task is the quantum analogue of determining whether a classical bit has been flipped on accident without looking to see whether it is 0 or 1.

The subsequent Threshold Theorem improved upon this result, showing that this kind of active error correction can be done successfully for an arbitrarily large and long computation, provided that the interaction obeys a certain set of assumptions that are independent of the length of the computation or the size of the computer [15–19]. Although in its first version these assumptions were strict and non-physical, subsequent generalizations have extended the essential result to a much wider class of interactions, including many of those actually encountered in quantum computer experiments today. As was discussed in the Chapter 1, the exchange interaction does not obey the conditions set forth by the most current version of the Threshold Theorem [27]. However, since the result has been extended multiple times before, this alone may not cause concern - surely some future Threshold Theorem might show that the boson exchange interaction, too, is correctable in a scalable way. As the theorem gives sufficient but not necessary conditions, its failure to apply to the exchange interaction does not mean that error is not error correctable, only that further work is needed to show that it is.

One of the main hopes for a quantum computer is that it can might solve certain problems much faster than classical computers. In particular, it is known that some problems for which the best known classical algorithm takes an amount of time that scales exponentially with the size of the input can be solved with only polynomial scaling time on an ideal quantum computer [12]. As discussed, this scaling is

determined asymptotically for arbitrarily large inputs. Thus, if this same kind of speed-up is hoped to be demonstrated on a physical, non-ideal quantum computer with errors, the Threshold Theorem must be proven for arbitrarily large computers, as this scaling distinction is only made in this limit. Indeed, the Threshold Theorem is proved for an arbitrarily large quantum computer, i.e., one which contains an arbitrarily large number of qubits. In this section, we argue that the spectrum of the exchange Hamiltonian, derived in Chapter 2, implies that the rate of errors it causes becomes arbitrarily fast when compared to the time it takes to perform single gate, which is considered fixed with respect to N. This would mean that error correction is not possible when no limit is placed on N, as the errors would eventually become much faster than the error correction scheme, no matter how fast the error correction scheme is. In this case, no conceivable Threshold Theorem could prove that errors due to this Hamiltonian are correctable in a scalable way unless the definition of scalable were altered.

So why is this worrisome? Just because there exists some computer size at which error correction fails, who's to say that that size is anywhere near what an experimentalist may ever attempt. After all, there are limits to the size of classical computers, imposed for example by the total amount of energy in the observable universe, and yet the conceptual framework which involves hypothetical arbitrary scaling still proves meaningful. However, error correction is considered an essential part of both classical and quantum computation while the finiteness of the observable universe is not, and thus restrictions placed based on it should be taken more seriously.

Practically speaking, if this error grew large for a physically realized N, it would simply be a matter of improving the technology so as to push that N limit larger until it is no longer an issue. Although it may be that this is does not significantly hinder the aims of quantum computers, it does point to a fundamental difference between classical and quantum computation that I am not aware has been observed before. The following scenario demonstrates this definition.

Consider a fixed system size and coupling magnitude (for example, $1/137$). One could

conceivably manufacture faster and faster gates so that the necessary conditions of the boson interaction is met, that it is slow enough that a single fault-tolerant quantum gate (which performs the logical operation as well as checks for and corrects errors) can be implemented at a success rate above some threshold which guarantees fault-tolerance. Since error correction occurs as the computation goes along, as long as a single gate can be performed with high (enough) fidelity, defined by the threshold, then a computation of any length can be performed without an overall failure rate that grows to 1 for long computations (this is oversimplifying a bit, but gets the idea across in the relevant way for the present discussion - showing that fault-tolerance of a single gate implies fault-tolerance of the entire computation is not so straight-forward). As we will see, for any fixed timescale of computation, as the system size increases, eventually the exchange interaction will decay or decohere the system so quickly that the computation's outcome is unhelpful. This is in contrast to the case of classical computing, where error rates are independent of computer size. For a classical computer, no such interaction exists because the interaction is described by a quantum Hamiltonian. To see the difference between this and the classical case, consider a physical classical machine, comprised of bits, that runs an algorithm to, say, add two numbers. Concurrently with the adding algorithm, it will also run some sort of error correction scheme since bits can be randomly flipped by the environment that the environment is located in. If one wished to improve the machine so as to double the number of digits allowed for the maximum input size, one would need to double the number of bits in the computer. Another copy of the original computer would work, so long as you could hook the two up in a way that allowed the two computers to act in concert on a single instance of the addition algorithm (augmented with error correction). The same algorithm executed on the same technology can produce reliable results in this manner, so long as one is willing and capable to expand the physical volume of the computer and wait longer for the computation to finish. If there are no changes to the technology or algorithm then at a minimum the volume would need to double, but due to heat considerations or other physical factors it may need to be increased by more. Of course, many technical challenges exist to

increase the speed at which computation occurs, as well as how to store it in a physically smaller volume, but in principle one can double the size of a problem that a classical computer can solve with a copy of that computer and sufficient space to store it in and sufficient time to wait for it to complete. This process can be greatly improved upon by reducing errors, writing better algorithms, and physically smaller bits and machinery that manipulates them, but even without these improvements the process can be repeated to add ever larger and larger numbers until one runs out of space or time. Past technological progress has allowed for computers to shrink drastically in size and increase drastically in speed, and many of the same technological difficulties have and will arise for quantum computers. However, if the error rate per qubit grows without bound with the total number of qubits, as we will demonstrate in this section using the derived spectrum, then there is an additional engineering problem on top of all of these already well-known challenges, which is unique to quantum systems and possibly not yet encountered in real quantum computers due to their currently small size. If, like in the case of the classical computer, we wish to increase the size of the problem that a quantum computer can solve, we cannot simply make it bigger. Because the error rate scales with system size, if the same technology and algorithm (including the error correction) are used on ever larger machines, eventually they will fail, unlike in the classical case where in principle they will work as long as you have enough time and space. Note that simply increasing the computer size to decrease the qubit-qubit coupling does not help, since the gate time is limited by causality to scale in the opposite way, meaning that the ratio t_d/t_g, with t_d being the decoherence time and t_g the gate time, cannot shrink indefinitely in this way (although it could shrink some, as current gate times are far above the minimum times dictated by causality) Thus if one has access to a certain kind of quantum technology, including a quantum algorithm and error correction, then even if they had enough time and space to build a bigger computer and wait for it to complete the algorithm, the outcome of the algorithm will, at some point in the scaling process, no longer be useful because it has been scrambled by errors. No such failure occurs for the analogous classical case, pointing to a fundamental difference between the two.

How much larger does the computer need to get before this happens? Estimating this threshold N at which the interaction becomes a problem is highly dependent on the specific algorithm and technology used, and could be the subject of much further study. In this thesis, we will stick to general statements about such systems in the large N limit, without reference to a specific implementation or value of N, since the asymptotic N scaling does not depend on implementation, only computer geometry and the coupling exponent α, which we take to be 1 from the boson coupling. Although there is always an interaction with a boson field, often a different error is dominate for the finite N encountered in real quantum computers; some of these errors fit the form of the exchange Hamiltonian with $\alpha \neq 1$ and so might also be investigated with this technique.

3.2 N-dependent Error Rate

I claimed above that the error rate in a quantum computer grows without bound with the number of qubits; in this section, we will show how this follows from the spectrum calculated in Chapter 2. We will go about this in two different ways: first by computing the decoherence time of a single subsystem which is part of a collection of N identical systems, and second by calculating the spectrum of the single qubit Hamiltonian, $H_j = \sum_{k \neq j} H_{jk}$, and using its width to estimate the timescale of single qubit dynamics. These two methods give the same result through different means, strengthening both arguments.

3.2.1 First Approach: Reduced Density Matrix

Suppose we have some initial state $|\psi(0)\rangle = \sum_q c_q |q\rangle$ which evolves according to the Schröedinger Equation $i\partial_t |\psi(t)\rangle = H |\psi(t)\rangle$ for some operator H conforming to the kind prescribed in Chapter 2. Define $c_q(t)$ such that $c_q(0) = c_q$ and $\langle q \mid \psi(t)\rangle = c_q(t)$ for all q. For the eigenbasis $\{|\lambda\rangle\}$ defined by $H |\lambda\rangle = \lambda |\lambda\rangle$, if $|\psi(0)\rangle = |\lambda\rangle$ then $\langle \lambda \mid \psi(t)\rangle = e^{-i\lambda t}$. By the linearity of the Schröedinger Equation, if $|\psi(0)\rangle =$

$\sum_\lambda f_\lambda |\lambda\rangle$ then $|\psi(t)\rangle = \sum_\lambda f_\lambda e^{-i\lambda t} |\lambda\rangle$. The initial state can also be expressed in the computational basis, $|\psi(0)\rangle = \sum_q c_q |q\rangle$, and the f_λ and c_q are related by $f_\lambda = \sum_q c_q \langle \lambda \mid q\rangle$ or $c_q = \sum_\lambda f_\lambda \langle q \mid \lambda\rangle$. Note that, due to normalization, the average magnitude of $\langle q \mid \lambda\rangle$ is $2^{-N/2}$ because for any basis x, $\sum_x |x\rangle \langle x|$ is the identity operator and therefore $1 = \langle q \mid q\rangle = \sum_\lambda \langle q \mid \lambda\rangle \langle \lambda \mid q\rangle$, so the average value of $\langle q \mid \lambda\rangle \langle \lambda \mid q\rangle = \left|\langle q \mid \lambda\rangle\right|^2$ is 2^{-N} since there are 2^N of them and their sum is 1.

Using this, we can write $c_q(t)$ as

$$\begin{aligned} c_q(t) = \langle q \mid \psi(t)\rangle &= \sum_\lambda f_\lambda e^{-i\lambda t} \langle q \mid \lambda\rangle = \sum_\lambda \sum_{q'} c_{q'}(0) \langle \lambda \mid q'\rangle \langle q \mid \lambda\rangle e^{-i\lambda t} \\ &= \sum_{q'} c_{q'}(0) \sum_\lambda \langle q \mid \lambda\rangle \langle \lambda \mid q'\rangle e^{-i\lambda t} \\ &= \sum_{q'} c_{q'}(0) \Lambda_{qq'}(t) \quad , \quad \Lambda_{qq'}(t) = \sum_\lambda \langle q \mid \lambda\rangle \langle \lambda \mid q'\rangle e^{-i\lambda t} \end{aligned} \tag{3.1}$$

Note that at $t = 0$, $\Lambda_{qq'}(0) = \langle q| \sum_\lambda |\lambda\rangle \langle\lambda| |q'\rangle = \delta_{qq'}$ as is required for consistency at $t = 0$:

$$c_q(t) = c_q(0)\Lambda_{qq}(t) + \sum_{q' \neq q} c_{q'}(0)\Lambda_{qq'}(t) \tag{3.2}$$

The coefficient of $c_q(0)$ in $c_q(t)$ is $\Lambda_{qq}(t) = \sum_\lambda \mid \langle q|\lambda\rangle \mid^2 e^{-i\lambda t}$. The size of this coefficient measures how much the value of $c_q(t)$ directly depends on its initial value $c_q(0)$. In the sum over λ in $\Lambda_{qq}(t)$, each summand is the phase $e^{-i\lambda t}$ multiplied by a positive real number. For any fixed distribution of the λ, once t is large enough the phases λt mod 2π will be distributed uniformly between 0 and 2π; the sum will have become dephased and average to zero. A sum with X randomly signed summands drawn from a distribution centered about a mean magnitude Y averages to zero and has typical magnitude $\sqrt{X}Y$. Brownian motion is a well known example of this in physics, where the direction is chosen randomly but the distance moved during each step is fixed or centered on some typical length scale; its typical displacement after some number of steps scales with the length scale to the first power and the square root of the number of steps. In the case of $\Lambda_{qq}(t)$, the sum over q has 2^N summands, and average value of the $\left|\langle q \mid \lambda\rangle\right|^2$ is 2^{-N} by normalization - therefore although its

average is zero, its typical magnitude is $\sim \sqrt{2^N}2^{-N} \sim 2^{-N/2}$. For long enough times $\Lambda_{qq}(t)$ becomes arbitrarily small for arbitrarily large N, but at $t = 0$ it is equal to 1 and for some interval after it will have a non-arbitrarily-small magnitude. Because this shrinks to zero for large N, that means if we wait long enough then $c_q(t)$ has almost no relation to $c_q(0)$. The value of $c_q(t)$ typically has, for long enough times, the same weight given to $c_q(0)$ as $c_{q'}(0)$ for all other q'. This means that for large N, where there are vastly more than one $q' \neq q$, the value of $c_q(t)$ has much more to do collectively with the initial states of the other subsystems than $c_q(0)$ itself. This kind of evolution is desirable in the case of a logical gate, but the interaction in question describes errors due to the exchange of bosons, something that cannot be controlled or predicted with certainty. Once the condition that $\left|\Lambda_{qq}(t)\right| \sim 2^{-N/2}$ is met, the system will have changed characteristically such that its relationship with its initial condition is almost entirely obscured. We call the timescale over which this occurs t_1, and next we will calculate its scaling with N as well as show that it is an upper limit on the decoherence time i.e., decoherence must have occurred by the time t_1. However, first, we repeat this analysis on the $\Lambda_{qq'}(t)$ for $q \neq q'$ for later use.

For $q \neq q'$, $\Lambda_{qq'}(t) = \sum_\lambda \langle q \mid \lambda\rangle \langle \lambda \mid q'\rangle e^{-i\lambda t}$ is typically vanishingly small for all times because the sum is always decoherent. $\Lambda_{qq'}(t)$ is an out of phase sum over 2^N summands, each with typical magnitude 2^{-N}. Since there are 2^N summands with random phases with typical magnitude 2^{-N}, the typical magnitude of the sum is $\sqrt{2^N}2^{-N} = 2^{-N/2}$. The product of two different $\langle \lambda \mid q\rangle$ and $\langle \lambda \mid q'\rangle$ for $q \neq q'$ will have a random phase, since there is no special relationship between the computational basis and the eigenbasis.

Now let $f(\lambda)$ be the distribution of eigenvalues and σ its width. When N is very large, the real part of the initially macroscopic sum $\sum_\lambda \mid \langle q|\lambda\rangle \mid^2 e^{-i\lambda t}$ is well approximated by the integral $\int_{-\infty}^{\infty} f(\lambda)\cos(\lambda t)d\lambda$. The imaginary part is always vanishingly small, so we focus on the real part since it must be responsible for the initially non-vanishing value. For example, when $f(\lambda)$ is gaussian i.e. $f(\lambda) = e^{-\lambda^2/2\sigma^2}/\sqrt{2\pi}\sigma$, then as t varies from 0 to ∞, this integral will monotonically decrease from 0 to 1.

We define t_1 as when this integral equals $2^{-N/2}$, its typical magnitude value when treated stochastically:

$$2^{-N/2} = \int_{-\infty}^{\infty} \frac{e^{-\lambda^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \cos(\lambda t_1) d\lambda = e^{-t_1^2\sigma^2/2} \implies t_1 = \frac{N^{1/2}}{\sigma}\sqrt{\log(2)} \tag{3.3}$$

Similar analysis could be performed for any given eigenvalue distribution $f(\lambda)$, and it is possible that the the time t_1 scales as $N^{1/2}/\sigma$ for a wide class of Hamiltonians.

Let us now observe that t_1 is an upper limit on the decoherence time. This will be done by examining the off-diagonal elements of the reduced density matrix in the computational basis, for some singled out qubit which we will call j. When these off-diagonal elements become vanishingly small, decoherence has occured. This is a common view of decoherence [46] and makes intuitive sense since $|\alpha|^2 |1\rangle\langle 1| + |\beta|^2 |0\rangle\langle 0|$ is the classical ensemble with probabilities due to observer ignorance associated with the quantum state $|\alpha|^2 |1\rangle\langle 1| + |\beta|^2 |0\rangle\langle 0| + \alpha\beta^* |1\rangle\langle 0| + \alpha^*\beta |0\rangle\langle 1|$ with intrinsically quantum probabilities. However, the definition of decoherence cannot precisely be 'when the reduced density matrix becomes diagonal' as of course there is always a basis in which this is true. There is nothing special about the computational basis, as the computation could be done in any basis (in other words, the 'up' and 'down' state could be mapped to any two orthogonal states). Having nothing special about it is precisely the desired property - for the following results will be true for any generic basis, the exception being the one in which the interaction is diagonal. Identifying this diagonal basis would be a computationally difficult task, and the ability to perform it would call into question the very need for the quantum computer. Since it would need to be completed before operating the quantum computer, it therefore must be done classically, and classically diagonalizing a matrix which grows exponentially in system size (the Hamiltonian is 2^N by 2^N) is well known to take an exponentially long amount of time. Thus to the extent that we hope for quantum computers to be able to surpass classical computers, we discount the possibility of performing such a diagonalization and using that basis as the computational basis, which would render the following results untrue. The decoherence time will apply to generic bases, where generic means that it does not

relate to the interaction eigenbasis in any special way; we will do the calculations in the computational basis which is an example of one such generic basis but use only its generic properties, allowing the results to apply to any basis except the special one described above. In summary, the decoherence time will be how long it takes the off-diagonal reduced density matrix elements in a generic basis with non-negligible initial off-diagonal density matrix elements to evolve to their typical, vanishingly small (with large N) magnitude.

Typicality is defined by what is overwhelmingly likely for a state chosen with uniform random components in the computational (or any other) basis. A typical state can be generated by choosing 2^N phases uniformly from $[0, 2\pi]$, choosing 2^N magnitudes uniformly from $[0, 1]$, dividing each magnitude by the squared sum of all magnitudes, multiplying the phases by the magnitudes, and assigning these products at random to the 2^N different c_q. The division of the magnitudes is necessary so that the state is normalized i.e., $\sum_q \left|c_q\right|^2 = 1$. After normalization, the average value of $\left|c_q\right|$ is exactly $2^{-N/2}$; for large N, if we skip the normalization step but instead choose the magnitudes from some distribution centered around $2^{-N/2}$, the resulting state will typically be very close to normalized.

Let the initial state of the ensemble be $|\psi\rangle = \left(a\,|0\rangle_j + b\,|1\rangle_j\right) \otimes \tilde{\sum}_r c_r\,|r\rangle$, where j is the subsystem in question and $\tilde{\sum}_r$ is the sum over the computational basis of the $N-1$ other sub subsystems. Such a product form ensures that the off-diagonal elements of the reduced density matrix are not vanishingly small. Let $c_{r0}(0) = ac_r$ and $c_{r1}(0) = bc_r$, where $|r0\rangle$ is the computational basis element of all N two-level system with the subsystem j in state 0 and the rest of the system in state r, and similarly for $|r1\rangle$. Using (3.1), the time evolution of the off-diagonal matrix element

$\rho_{01}(t)$ is

$$\begin{aligned}
\rho_{01}(t) &= \tilde{\sum_r} c_{r0}(t)c^*_{r1}(t) = \tilde{\sum_r}\sum_{q,q'} c_q(0)c^*_{q'}(0)\Lambda_{r0,q}(t)\Lambda^*_{r1,q'}(t) \\
&= \tilde{\sum_r}\sum_{q,q'} c_q(0)c^*_{q'}(0)\sum_{\lambda,\lambda'} e^{i(\lambda'-\lambda)t}\left\langle r0|\lambda\right\rangle\left\langle\lambda|q\right\rangle\left\langle q'|\lambda'\right\rangle\left\langle\lambda'|r1\right\rangle \\
&= \tilde{\sum_r}\Bigg[\tilde{\sum_{r',r''}}\Bigg(c_{r'0}(0)c^*_{r''0}(0)\Lambda_{r0,r'0}(t)\Lambda_{r1,r''0}(t) + c_{r'0}(0)c^*_{r''1}\Lambda_{r0,r'0}(t)\Lambda_{r1,r''1}(t) \\
&\qquad + c_{r'1}(0)c^*_{r''0}(0)\Lambda_{r0,r'1}(t)\Lambda_{r1,r''0}(t) + c_{r'1}(0)c^*_{r''1}(0)\Lambda_{r0,r'1}(t)\Lambda_{r1,r''1}(t)\Bigg)\Bigg] \\
&= \tilde{\sum_r}\Bigg[\tilde{\sum_{r',r''}}\Bigg(\mid a\mid^2 c_{r'}c^*_{r''}\Lambda_{r0,r'0}(t)\Lambda_{r1,r''0}(t) + ab^*c_{r'}c^*_{r''}\Lambda_{r0,r'0}(t)\Lambda_{r1,r''1}(t) \\
&\qquad + a^*bc_{r'}c^*_{r''}\Lambda_{r0,r'1}(t)\Lambda_{r1,r''0}(t) + \mid b\mid^2 c_{r'}c^*_{r''}\Lambda_{r0,r'1}(t)\Lambda_{r1,r''1}(t)\Bigg)\Bigg]
\end{aligned} \tag{3.4}$$

Let us examine more closely the four terms comprising $\rho_{01}(t)$ in the final line. Even though for a generic subsystem j in a generic state, $\rho_{01}(t)$ is arbitrarily small for large N, by assumption we have begun with initial conditions such that $\rho_{01}(t)$ is initially finite and by continuity of the Schrödinger Equation it must remain finitely large for some time interval afterwards. Thus, for the initial value of $\rho_{01}(t)$ to be nonvanishingly small with large N, there must be at least one of the four terms which are non-vanishingly small initially. Additionally, since each term is a continuous function of time, that quantity will remain finite for some interval of time afterwards. During this interval, the only non-vanishing contribution to its value is from terms in (3.4) with a coefficient $\Lambda_{qq}(t)$, which happens when $r = r' = r''$ and the a and b coefficients is ab^*, not $a^*b, |a|^2$ or $|b|^2$. The condition on the a and b coefficients make sense because $\rho_{01}(0) = ab^*$, so the coefficient of that term in (3.4) must be initially non-vanishing, and all the others must be initially vanishing. After t_1, even the $\Lambda_{qq'}(t)$ with $q = q'$ will become vanishingly small, meaning that $\rho_{01}(t > t_1)$ itself is vanishingly small. Again this makes sense because for almost all of Hilbert space this quantity is vanishingly small for any qubit j, and since the exchange Hamiltonian is highly non-diagonal in the computational basis, the

evolution it generates will not keep initial states for which this is not true in that tiny fraction of Hilbert. Once the off-diagonal element is vanishingly small in the computational basis, that qubit has decohered, as discussed above. Since t_1 is the time at which this occurs, we may say that by $t = t_1$ the $N-1$ two-level systems have decohered the singled-out qubit j. This establishes the claim above, that t_1 is an upper bound on the decoherence time. This means most of the information initially in the single-out two-level system has leaked into the other $N-1$ two-level systems. Past t_1, when all $\Lambda_{qq'}$ are vanishingly small for $q = q'$ and $q \neq q'$, we can see from (3.4) that the value of ρ_{01} for qubit j is due much more to the state of the other qubits than its own initial conditions; its initial information has almost leaked out entirely into the rest of the system.

3.2.2 Second Approach: The Single Qubit Hamiltonian

Here, we will examine the single qubit Hamiltonian instead of the entire Hamiltonian. The single qubit Hamiltonian H_j is defined by singling out a subsystem j and including only terms in the Hamiltonian which directly act on j, i.e., $H_j = \sum_{k \neq j} H_{jk}$. For short timescales, the time evolution according to this Hamiltonian only will approximate the dynamics of the reduced density matrix for subsystem (qubit) j. If the single qubit Hamiltonian commuted with the entire Hamiltonian, $[H_j, \sum_{(kk')} H_{(kk')}] = 0$, then the reduced density matrix for j would be correct for all times. However for $H_{jk} = F_{jk}(\sigma_j^+ \sigma_k^- + \sigma_j^- \sigma_k^+)$ this is not the case, and for times when $t^2 \big|[H_j, H_{jk}]\big|$ is on the order of $t\big|H_j\big|$ for some k, then the approximation will break down. This can be thought of as the inaccuracies of the state of qubit k, due to the lack of terms $H_{kk'}$ for $k' \neq j$, leaking back into the dynamics of qubit j and indirectly causing inaccuracies there. Thus, the decoherence time calculated according to the full Hamiltonian versus just the single qubit Hamiltonian might be different. Does the inclusion of all terms $H_{kk'}$ for $k, k' \neq j$ in the Hamiltonian tend to increase or decrease the decoherence time, or neither? This will come down to whether or not the collective effect of all other qubits is coherent or decoherent; if it is perfectly decoherent, it can't prefer one direction or another and so must not

affect the decoherence time on average. It is reasonable that the effects approach decoherent for large N because the phases of the interactions becomes uniformly distributed in $[0, 2\pi]$. But the effect under consideration here is not characterized by the interactions H_{jk} whose relative phases become completely random, but rather by the difference between including or excluding $H_{kk'}$ for $k, k' \neq j$. It is not so clear what the coherence or decoherence condition is; we will look for evidence that the different single qubits Hamiltonians have no special phase relations amongst themselves by comparing each of their spectra to the spectra of the full Hamiltonian. In order to do so we must first calculate the spectrum of the single qubit Hamiltonian.

This will be done in a similar manner than the full Hamiltonian, although many simplifications are possible and so we will define some quantities slightly differently than in that case. Be warned that when changes are made we do not change the names of these quantities so that the analogous steps in both processes are more easily identified. We begin in the same place:

$$\sigma_j^{(p)} = 2^{-N} \sum_q \langle q| H_j^p |q\rangle = 2^{-N} \sum_q \langle q| \prod_{w=1}^{p} \sum_{k_w \neq j} H_{jk_w} |q\rangle \tag{3.5}$$

First, observe that for odd p, $\sigma_j^{(p)}$ is identically zero since for all $\{k_1...k_p\}$ (all not equal to j) and for all $|q\rangle$, the state $H_{jk_1}...H_{jk_p}|q\rangle$ is orthogonal to $|q\rangle$ because j is oppositely oriented in the two states, having been flipped an odd number of times. Thus we will focus on the case of even p, which is assumed throughout the rest of this section.

For similar reasons, each of the $k_w \neq j$ in $\{k_1...k_p\}$ must appear an even number of times, otherwise $\langle q| H_{jk_1}...H_{jk_p}|q\rangle = 0$ for all $|q\rangle$. Like before, we will break the sum over choices of all ordered sets of p subsystems not equal to j into a pattern, subsystem selection, and subsystem assignment $(S, \{k\}, \pi)$. We denote a term $H_{jk_1}...H_{jk_p}$ associated with a triplet $(S, \{j\}, \pi)$ as $\prod H(S, \{j\}, \pi)$ and similarly for $\prod F(S, \{j\}, \pi) = F_{jk_1}...F_{jk_p}$.

The elements of S are collections of locations in the sequence $\{k_1...k_p\}$ where each

subsystem reference appears, $\{k\}$ is the set of all subsystems which appear, and π assigns each subsystem in $\{k\}$ to the appropriate set of locations in S. For example, $H_{jk_1}...H_{jk_p} = H_{j1}H_{j7}H_{j3}H_{j3}H_{j1}H_{j7}$ or $\{k_1...k_p\} = \{1,7,3,3,1,7\}$ has the pattern $S = \{\{15\}\{26\}\{34\}\}$ (using the labelling convention that the leftmost location in $H_{jk_1}...H_{jk_p}$ is labelled 1 and incrementing by 1 to the right), the subsystem selection is $\{1,7,3\}$, and the assignment π pairs up the elements of S with elements of $\{k\}$ in the order that they appear as written above (these sets are mathematically unordered, hence the need for the assignment π).

Now we can write

$$\sigma_j^{(p)} = 2^{-N} \sum_S \sum_{\{k\}} \sum_\pi D(S)^{-1} \sum_q \langle q| \prod H(S,\{k\},\pi) |q\rangle \tag{3.6}$$

where as before the sum over patterns S, subsystem selection $\{k\}$, and subsystem assignments π are assumed to be only over compatible such triplets (to be defined shortly), and $D(S)$ is the degeneracy factor for each triplet that is invariant with respect to specific choices of compatible $\{k\}$ or π for a given S and so written is as $D(S,\{k\},\pi) = D(S)$. Let us establish again the convention that a pattern S with u different collections in it is denoted $S(u)$, a subsystem selection of length u is denoted $\{k\}(u)$, and a subsystem assignment π which matches up u different things is denoted $\pi(u)$.

As before, the compatibility conditions are that for $(S(u),\{k\}(u'),\pi(u''))$, we require that $u = u' = u''$. Note that S has a slight different meaning here than in the full Hamiltonian case; since each pairwise Hamiltonian automatically has j in it, we need only specify which other subsystem is in it. Thus there are only up to p different collections in it, not $2p$. A related deviation is that, as before, there must be at least two different subsystems appearing in each pattern (since terms $H_{kk'}$ for $k = k'$ are not allowed), but since j is always taken to be one of the two subscripts in each pairwise Hamiltonian, $u+1$ is the total number of subsystems appearing in $\prod H(S,\{k\},\pi)$, not u. In particular, a pattern with only 2 different subsets in total corresponds to $u = 1$, not $u = 2$, and so u must be between 1 and p. Additionally,

instead of each location appearing exactly twice in two different collections, each location instead appears exactly once. The subsystem selection and assignment $\{j\}$ and π work in the same way as the full Hamiltonian case.

We define $\Omega'(S(u))$ to be the number of states $|q\rangle$ of the u subsystem slots filled by any compatible $\{k\}(u)$ and $\pi(u)$ for which $\langle q| \prod H(S(u)) |q\rangle \neq 0$. As discussed in the derivation of the full Hamiltonian spectrum, this number does not depend on which subsystems appear where in the pattern, and thus depends only on S and not $\{k\}$ or π. Defined this way, $\sum_q \langle q| \prod H(S(u), \{k\}(u), \pi(u)) |q\rangle = 2^{N-u}\Omega'(S(u)) \prod F(S(u), \{k\}(u), \pi(u))$. Additionally, it was observed that Ω' does not depend on N in the sense that it does not affect the N scaling argument, as long as the final patterns S_t determined by the N scaling argument have the property that $\Omega'(S_t) \neq 0$, which we will establish once identifying them.

Due to the appearance of j in every pairwise Hamiltonian, there is a restriction on any pattern for which there is any contribution to $\sigma^{(p)}$. Beginning with the leftmost H_{jk} in $H_{jk_1}...H_{jk_p}$, label the pairwise Hamiltonian factors starting with 1 and incrementing by 1 to the right, as in the convention used in the example above. We now have a sense of even or odd pairwise Hamiltonians: those with an even or odd label according to this scheme. As observed, all subsystems in $\{k\}$ must appear an even number of times in order for there to be any contribution to the moments. Additionally, consecutive appearances of the same subsystem must alternate between even and odd pairwise H_{jk} in the sense just defined. This is because if the state $|q\rangle$ is acted on by H_{jk} and j and k are not oppositely oriented in $|q\rangle$, then $H_{jk} |q\rangle = 0$. Thus when we go to act $H_{jk_{p-l-1}}$ on the state $H_{jk_{p-l}}...H_{jk_p} |q\rangle$ where one or more of the $\{k_{p-l-1}...k_p\}$ is equal to k_{p-l}, if there is an odd number of H_{jk} appearing between the leftmost $H_{jk_{p-l}}$ in $H_{jk_{p-l-1}}...H_{jk_p}$, then since j and k_{p-l} were oppositely aligned when that pairwise Hamiltonian acted (otherwise the whole term would already be zero), then they are now not oppositely aligned, making $H_{jk_{p-l}}H_{jk_{p-l-1}}...H_{jk_p} |q\rangle = 0$. So in the sum over patterns, we may consider it to be a sum over only patterns for which each subsystem appears an even number of times and at alternatively even and odd locations. The contribution from other

patterns is zero and so may be omitted without consequence.

Decomposing the sum over all patterns into a sum over u and a sum over all patterns with a specific value of u, we have

$$\sigma^{(p)} = 2^{-N} \sum_{u=1}^{p} \sum_{S(u)} 2^{N-u} \frac{\Omega'(S(u))}{D(S(u))} \sum_{\{k\}} \sum_{\pi} \prod F(S, \{k\}, \pi) \tag{3.7}$$

Defining $E[\prod F(S)]$ such that

$$A_{S(u)} E\big[\prod F(S(u))\big] = \sum_{\{k\}(u)} \sum_{\pi(u)} \prod F(S(u), \{k\}(u), \pi(u)) \tag{3.8}$$

we have as before the new expression of the spectrum moments:

$$\sigma_j^{(p)} = \sum_{u=1}^{p} 2^{-u} \sum_{S(u)} \frac{\Omega'(S(u))}{D(S(u))} A_{S(u)} E\big[\prod F(S(u))\big] \tag{3.9}$$

Note that the expectation value means something slightly different here, since j is fixed to appear as one of the two subscripts in each pairwise Hamiltonian. Thus while before, $E[F^B] = \binom{N}{2}^{-1} \sum_{(kk')} F^B_{kk'}$, for the single qubit case it is $E[F^b] = (N-1)^{-1} \sum_k F^b_{jk}$. In Appendix A we study these quantities in greater detail and note that while the prefactor can change between the two cases, single qubit and full, the asymptotic N scaling is not affected

As before, all N dependence is in A_S and the E values; neither its prefactor, nor the number of patterns or number of possible values of u, depends on N. We will again identify a transformation (only 1 kind is needed here) which increases the N scaling of the contribution of the pattern to $\sigma_j^{(p)}$, and after exhaustive applications of it we will be left with a final set of patterns $\{S_t\}$ which, once it is observed that $\Omega'(S_t) \neq 0$, will be the only important patterns for large N.

This transformation takes any $k \in \{k\}$ which appears 4 or more times and reduces its occurences by 2. It does so by choosing any 2 locations in the collection in S which k is assigned to by π, removing those 2 locations from their collection in S,

and adding a new element to S with just these two locations. This is analogous to $\rightarrow_1$ from the full case, but here u increases only by 1 since the new H^2_{jk} only requires one new subsystem to be put in it, since it is given that the other subsystem will be j. When we choose the two locations to remove, we make sure they are consecutive occurrences of k; that way both the old and the new subsystem references will alternate at even and odd locations, since by assumption the old subsystem reference which appeared 4 or more times appeared in alternatively even and odd locations.

As previously observed, the test of the necessary N-scaling properties for two patterns $S \rightarrow S'$ related via this transformation is that

$$\frac{A_S E[\prod F(S)]}{A_{S'} E[\prod F(S')]} \rightarrow 0 \tag{3.10}$$

for large N. $A_{S(u)} \sim N^u$ is the number of compatible subsystem selections and assignments. As previously shown, in the large N limit the expectation values depend only on the power to which each given factor F_{jk} appears. In particular, if the set of those integers for a pattern S is $B(S)$, then $E[\prod F(S)] = \prod_{B_i \in B(S)} E[F^{B_i}]$ where $E[F^B] = (N-1)^{-1} \sum_{k \neq j} F^B_{jk} \rightarrow N^{-1} \sum_{k \neq j} F^B_{jk}$. Note that there is no need to add absolute values, since for any sequence or pattern with a nonzero contribution to $\sigma^{(p)}$ has all even exponents in $B(S)$.

Using these facts, the ratio scales as

$$\frac{A_S E[\prod F(S)]}{A_{S'} E[\prod F(S')]} \sim \frac{E[F^{2b}]}{N E[F^{2b-2}] E[F^2]} \tag{3.11}$$

where $2b$ is the (necessarily even) number of times the chosen subset to be acted on appeared, which was 4 or greater by assumption. The vanishing of this fraction for all integers $b > 1$ is thus the condition under which the argument applies, which if valid would mean that only the final patterns arrived at after applying this transformation as many times as possible contribute meaningfully to the moments $\sigma^{(p)}$. This is similar but more stringent to the condition arising in the full case for $\rightarrow_1$; however, we noted there that the condition is still satisfied in this more stringent case and defer to that section for the proof.

Thus we can conclude that only the patterns arrived at after transforming each pattern as many times as possible need to be included in the sum over patterns which constitutes $\sigma^{(p)}$ in the large N limit. These patterns, called S_t, have the property that every subsystem besides j appears exactly twice, and thus the $p/2$ elements of any S_t are pairs of locations. $\Omega'(S_t)$ is the same for all patterns S^t and is equal to 2, since once the orientation of j is fixed in $|q\rangle$, the orientation of all other subsystems in $\{k\}$ is fixed when one requires that $\langle q|\prod H(S_t,\{k\},\pi)\,|q\rangle \neq 0$ for any $|q\rangle$. All S_t also share a value of $D(S_t)$ which is equal to 1 since both the subsystem selection and assignment are determined uniquely from a list $\{k_1...k_p\}$ and a pattern S which conforms to that list. This is because the subsystem selection is simply the set of all subsystems appearing in $\{k_1...k_p\}$, and since for S_t each subsystem appears in two locations at which no other subsystem appears (besides j, which is not part of the selection and assignment since it is automatically in each F_{jk}) then there is no ambiguity as to which element of S_t pairs with which subsystem in $\{k\}$, nor which subsystems are contained in $\{k\}$. Finally, A_{S_t} is also the same for all S_t, and is equal to the number of subsystem selections compatible with S_t, $\binom{N-1}{p/2}$, times the number of compatible subsystem assignments, $(p/2)!$. In the large N limit, A_{S_t} is then $(N-1)!/(N-1-p/2)! \to N^{p/2}$. The number of different patterns S_t is equal to the number of ways to pair up the $p/2$ even locations with the $p/2$ odd ones, which there are $(p/2)!$ ways to do.

Thus the even spectrum moments for the single qubit Hamiltonian H_j, using 3.9, are

$$\begin{aligned}\sigma_j^{(p)} = \sum_{u=1}^{p} 2^{-u} \sum_{S(u)} \frac{\Omega'(S(u))}{D(S(u))} A_{S(u)} E[\prod F(S(u))] \to 2^{-p/2} \sum_{S_t} \frac{\Omega'(S_t)}{D(S_t)} A_{S_t} E[\prod F(S_t)] \\ = 2^{-p/2}(p/2)! \frac{2}{1} N^{p/2} E[F^2]^{p/2} = 2\,(p/2)! \Big(N E[F^2]/2 \Big)^{p/2}\end{aligned} \tag{3.12}$$

Using again the construction of a distribution from its moments 2.2, the single qubit

Hamiltonian spectrum's fourier transform $F(\xi)$ asymptotes for large N to

$$\begin{aligned} F(\xi) &= \sum_{p=0}^{\infty} \frac{\big(i\xi\big)^p}{p!}\sigma^{(p)} \rightarrow \sum_{p=0}^{\infty} \frac{(i\xi)^{2p}}{(2p)!}\sigma^{(2p)} \rightarrow 2\sum_{p=0}^{\infty} \frac{\big(i\xi\big)^{2p}}{(2p)!}p!\Big(NE[F^2]/2\Big)^p \\ &= 2\sum_{p=0}^{\infty} \frac{p!}{(2p)!}\big(-\xi^2 NE[F^2]/2\big)^p \end{aligned} \tag{3.13}$$

Like the full Hamiltonian, these moments scale uniformly with N and so the shape is constant with respect to N with only a size factor that varies with N and the typical strength of the coupling $E[F^2]$ (which depends on N but also the overall coupling strength, the computer size, etc). We approximate this distribution near its center by truncating the sum over p and plot the shape for an arbitrary size in 3.1

In contrast, the width of the single qubit Hamiltonian for the independent case, $\omega\sigma_j^z$, has a constant width (its spectrum has only two values; 2^{N-1} eigenvalues are ω and the other 2^{N-1} are 0). This suggests that the error dominates over the more basic Hamiltonian for large N. This result is interesting on its own and reinforces the conclusions drawn from the full Hamiltonian. It also implies that the single qubit error rate might scale as the width of the single qubit Hamiltonian spectrum which grows without bound with N. This will be the case if the effect of the other single qubit Hamiltonians H_k for $k \neq j$ does not prefer to increase or decrease the decoherence time of qubit j on average (for large N). The only way one of these preferences could be chosen is if there is a special relationship between the single qubit Hamiltonians - however, the relationship between the full and single qubit Hamiltonian suggests there is not. This is because the full spectrum is exactly what one would expect if adding together N random operators together, each with the single qubit Hamiltonian spectrum. When adding together operators in such a manner, the eigenvalues of the composite operator will typically be $\sqrt{N}$ times larger, which is indeed implies the relationship between the eigenvalue density moments in the single and full, or composite, case. By random, we mean with respect to each other; the matrix elements themselves can be totally random or very regular

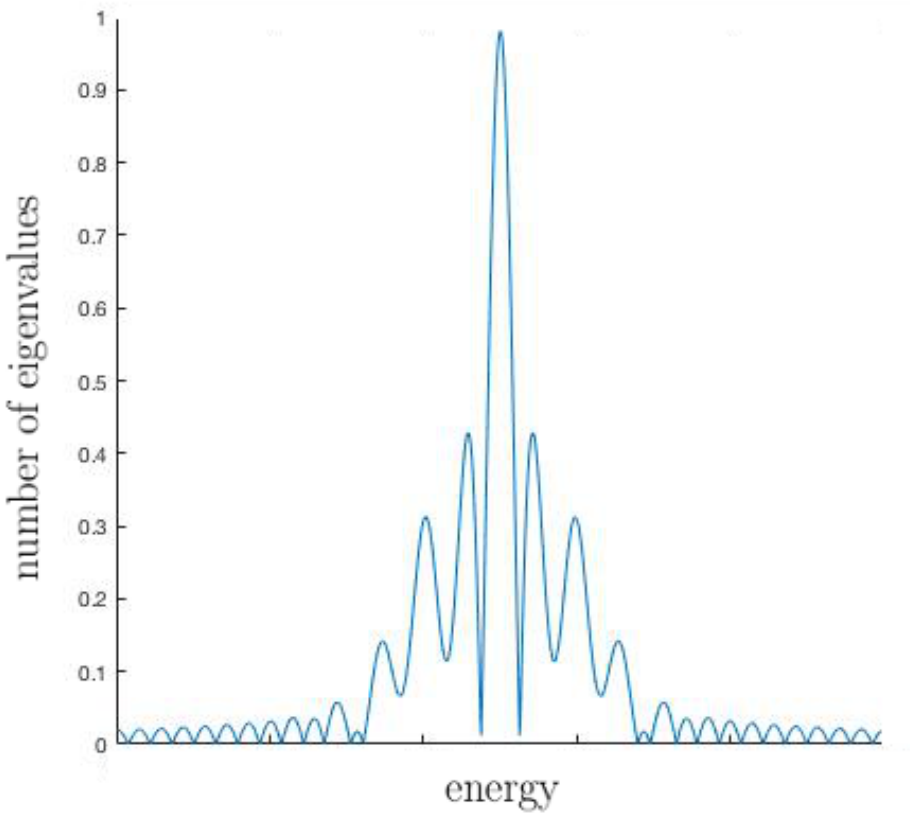

Figure 3.1: The single qubit Hamiltonian spectrum. The vertical axis is relative number of eigenenergies at that energy, and the horizontal axis has units of energy but with an arbitrary scaling that will depend sensitively on many parameters. In contrast, its shape plotted above will always be achieved for large enough N with only its overall size determined by other parameters. Its overall size does grow with N in an asymptotically simple way, as $N^{1/6}$ in 3 dimensions and $\log(N)$ in 2 dimensions. This means that, for any parameter values, the spectrum becomes arbitrarily wide for arbitrarily large N. The zero of the vertical axis is unlabled because it is immaterial, amounting only to a shift to all energies by a constant and therefore affecting only the overall complex phase of the wavefunction, an unphysical parameter.

looking depending on the basis chosen to write down the elements in. This implies that if we were to multiply each single qubit Hamiltonian by a random phase before adding them to make the full Hamiltonian, the final spectrum would not differ on average, meaning there is no special phase relationship between the different H_j that affects the time scale of the dynamics produced by the full Hamiltonian and that the inclusion of the other $H_{k\neq j}$ does not on average cause faster or slower decoherence of qubit j.

The decoherence time may then be safely calculated according to only H_j, and since the moments of its spectrum scale uniformly with N we may take this time to be $(\sigma^{(p)})^{-1/p}$ which has the same N scaling for all p. We will call this time t_1 for comparison with the first approach in which we found that $t_1 \sim N^{1/2}/\sigma$ where σ is the width of the full Hamiltonian. We can see that the scaling of the single qubit Hamiltonian moments are $N^{p/2}$ smaller than that of the full Hamiltonian, so that if σ is the width of the full spectrum then indeed $t_1 \sim (\sigma^{(p)})^{-1/p} \sim \left(\sigma/N^{1/2}\right)^{-1} \sim N^{1/2}/\sigma$ which agrees with the first approach. Since this has been arrived at within a different framework and using a different assumptions, these two results lend credence to each other.

Another way to understand this result is to look at the magnitude of the changes to state coefficients, which is deduced from the eigenvalues. For an eigenvector with eigenvalue λ, the magnitude of the change it experiences in a time step dt is equal to λdt. For a superposition of eigenvectors (which of course any state can be expressed as), each of its eigencomponents changes by a different amount according to its eigenvalue. For a random or typical vector, its total magnitude change will converge for large N to the average change among eigenvectors, which is dt times the average eigenvalue magnitude. The action of the full Hamiltonian upon a state is equal to (one half) the sum of the action of each single qubit Hamiltonian i.e.,Zombie Attack / Dueal / Bridge Explosion 5:12 Fernando Velazquez 0 1 MPEG audio file $H_{\text{full}} |\psi\rangle = \frac{1}{2}(\sum_j H_j) |\psi\rangle = \frac{1}{2} \sum_j (H_j |\psi\rangle)$. How much the full state changes depends both on the magnitude of the changes made by the various H_j, as well as the phase relations between these various changes. Since they each have typical magnitude

$\sqrt{\sigma_j^{(2)}}^{1/2} = \sigma$, the total change could be either $N\sigma$ if they are all perfectly in phase, or typically of magnitude $\sqrt{N}\sigma$ with a random sign if they all have random phases. If the moments are all multiplies by the same factor, it can be seen from 2.2 that this amounts to a re-scaling of time. Thus if the phase relations between the changes induced by the various H_j are randomly phased with respect to one another, we would expect that on average time evolution is $\sqrt{N}$ times faster than in the single qubit case, and therefore that the moments are all $\sqrt{N}$ times larger. As observed above this is indeed the case, again implying that the changes in evolution due to each H_j are randomly phased with respect to one another.

3.3 The Relative Error Rate

We have now established that the error rate increases without bound with N, but the only way to decide if these errors are 'fast' or 'slow' is to compare them to the rate of computation. Since active error correction is assumed, the relevant timescale for comparison is the fault-tolerant gate time t_g. Here 'fault-tolerant' means that, in addition to the time it takes to perform the logical gate, t_g also includes any time it takes to complete a quantum error correction cycle, typically done after each logical operation. We argued that this comparison demonstrated the impracticality of scalable quantum computing, since the errors become arbitrarily fast compared to the gate time, whose function depends on the rarity of errors during its execution. Note that it does not matter what the gate time is, for large enough N the errors will always become much faster. One could continually make the gate times shorter and shorter as N increases, but conceptually this would be a very different from classical computing, where such technological improvements are done for convenience and are not necessary in principle for arbitrary scaling of problem size and computer size. A simple way to decrease the error rate, effectively increasing the gate time, would be to move all of the qubits farther from each other. Since their couplings go as inverse separation, this would uniformly slow down the errors. We can imagine spreading the computer out as N grows so that the error

rate is always much less than the gate time. However this does not help, because the gate time cannot be held constant during this process; the gate time between two qubits is bounded below by r/c where r is their separation and c is the speed of light, so as the computer becomes more and more spread out the gate time will need to increase as well. As the size scaling increases the gate time linearly and decreases the coupling also linearly, we can see that there is no net impact on their ratio. Thus such a scaling can be useful in a limited way if the current gate time is above the lower limit imposed by relativity, but this approach cannot ultimately avoid the scaling conclusions arrived at in this thesis.

3.3.1 Is it even a Quantum Computer anymore?

If the full exchange Hamiltonian is compared to the full base Hamiltonian $\omega_a \sum_j \sigma_j^z$, an even more concerning implication arises. This base Hamiltonian that tells us that we have a collection of two level systems, each with energy difference ω_a. For large N, the spectrum of this Hamiltonian approaches a gaussian distribution with a width that scales as $\sqrt{N}$. Thus the width of the base Hamiltonian becomes negligible compared to the width of the interaction Hamiltonian for large N, since the interaction Hamiltonian's width scales as $N^{1-1/D}\log(N)^{(3-D)/2}$ for $D = 2$ or 3, which grows faster than $N^{1/2}$ in both cases. When perturbation theory is used, a common analysis tool for modelling quantum errors, it is assumed that the effect of the errors is small compared to the base or ideal Hamiltonian. This is analogous to a Taylor expansion, in which a small change is added to some base value, and the effect of this small change is estimated in a way that is accurate so long as it is small. Due to the global phase invariance of $|\psi\rangle$, the range of eigenvalues rather than their values characterize the timescales of the evolution generated by an operator. For regular distributions, such as Laplace and Gaussian, the spectrum width characterizes this range well. Unlike the case of real numbers, their operator character makes it less simple to compare the sizes of the base and error Hamiltonians. But one of the chief differences between operators and real numbers is that operators need not commute; however in this case, $[H_{\text{base}}, H_{\text{int}}] = 0$,

as the operator H_{int} leaves invariant the number of excited subsystems of the state that it acts on. Thus in at least one important way these operators are similar to real numbers, lending credence to the taylor series analogy and the idea that their timescales can be determined independently. Comparing the sizes of their widths, then, implies that perturbation theory is inappropriate in this case, or even that for large enough N the base Hamiltonian would become a perturbation off of the interaction Hamiltonian. The collective system at that point has lost almost all sense of being a collection of two level systems - this two-level structure is instead a perturbation off some much more prominent, unknown structure. This calls into question the very concept of a quantum computer as a collection of two-level systems. Although errors will have dominated long before this limit is reached, it is another way to view the unscalability implied by the interaction Hamiltonian spectrum found in Chapter 2.

4 Numerical Studies

4.1 Introduction

In this chapter, we will explore numerical ways to test and explore the system under consideration in this thesis, namely an ensemble of two-level quantum systems coupled to a common boson bath at zero temperature. First we present a novel semi-classical algorithm to approximate the partial dynamics of such a system. While many details of the system are lost in such a simulation, the observable that is focused on is the status of each system in its individual, non-interacting eigenbasis i.e., whether it is in the excited or ground state. The word 'on' is substituted for 'excited' in this code. In addition, we will refer to the subsystems as atoms, but as is the case throughout this thesis, they may be substituted with any two-level quantum system. While the assumptions and simplifications used are not rigorously justified, the results qualitatively match what is known to occur in such systems when there are initially many excitations. Because at all times each subsystem is either in the excited or ground state, this model is referred to as the 'Definite Model'.

4.2 Definite Model

In the Definite Model, each atom always has a definite state: excited or ground. This means it is an essentially classical model, but we insert the quantum mechanics by allowing for spontaneous and stimulated emission as well as absorption.

In this model, an initial state is specified by a total number of atoms and the fraction that are initially excited. This model is intended to work best when the number of excitations is some finite fraction of the whole, such as $N/2$ which can be physically achieved by blasting the sample with near-resonance photons for a long time.

During each timestep of the computation, each subsystem that is excited at the beginning of that step is allowed some small chance of decay equal to its decay rate Γ times the duration of the timestep, dt, which must be chosen so that $\Gamma dt << 1$. If such a spontaneous decay occurs, a photon is created in the system and accordingly each other subsystem is immediately given a chance to absorb or emit (depending in whether it itself is in the excited or ground state) due to the new photon. The probability of stimulated emission and absorption must be equal, but we must determine what this probability is.

The atomic coupling is $\left|F_{jk}\right| \approx \Gamma/5kr_{jk}$ which should define the timescale of their interaction, where r_{jk} is the separation between the pair of subsystems in the sample. We do not include the factor of Γ because this has already been accounted for in the spontaneous emission of the first atom; in order for the probability of one atom emitting and stimulating another to be $\Gamma dt/5kR$ as desired, we require the conditional probability of stimulated emission (conditioned on the original spontaneous emission) to be $1/5kR$. The rates of absorption are the same as that of stimulated emission, and which is process is possible is determined by if the other atom is excited (in which case it might undergo stimulated emission) or in its ground state (in which case it might absorb the photon). When stimulated emission does occur, the electric field amplitude is increased due to the presence of another (in phase) photon, and similarly after absorption occurs the electric field amplitude decreases as a photon is destroyed. Thus the probability $1/5kR$ should be multiplied by the total number of in-phase photons due to the original spontaneous emission (when there is no original emission, the number of photons is zero and there is no chance of stimulated emission or absorption). We track the number of photons, initializing the count to 1 when a spontaneous emission occurs, and increasing or decreasing the count by 1 whenever there stimulated emission or absorption, respectively. If at any point the number of photons becomes zero, the process of stimulated emission or absorption stops; no more atoms are given a chance to emit or absorb for that initial spontaneous emission. In this case, the remaining subsytems are given a chance to spontaneously emit during that time,

initiating the above process for each subsystem that does. When all have been given such a chance, and the ones that do emit have the emitted photon interact with each other subsytem, the algorithm moves onto the next time step. At each step we record the total number of excited subsystems, so that this quantity versus time may be graphed once the algorithm is finished.

In a reflective cavity each emitted photon bounces around until measurement, continually stimulating emission and absorption for every atom; however, the MOT is contained in absorptive walls, which effectively measure the photons when they strike the wall. This justifies the idea that each emitted photon has one chance to stimulate emission or be absorbed by each atom as it 'passes by'. If these different processes were happening continuously, as in the case of the reflective walls, their relative phases would become of great importance as time passes, when certain possibilities constructively or deconstructively interfere in complicated ways. Thus we can think of the computational time step to be the time it takes for a photon to radiate from the sample to the walls; as is required of the timestep, this is much smaller than the timescale of the dynamics of the Hamiltonian. There still could be interference between different spontaneous emission processes that occur in the same computational time step, but neglecting these yields a great increase in computational efficiency. We hope that the small photon travel time, coupled with the large sample size, make this approximation valid. Comparison with experiments might determine if these assumptions are valid or not.

The order that the other atoms are given a chance to spontaneously emit or absorb should not matter in the limit of small step sizes. Similarly, when N is large, it is a good approximation to simply choose random separations for each pair of atoms when needed during evolution. This is especially true because of the semi-classical approximation where each atom is only in either the excited or ground state, with no other information about its state being recorded or used. This means any sort of quantum correlations that build up more strongly between neighbors, and therefore may affect future evolution, is not accounted for. Thus it doesn't really matter which atom a photon comes from or which atom absorbs it. When their positions are

sampled randomly inside of a gaussian cloud a great many times, it will statistically converge to the exact case where each atom is assigned a static position relative to each other atom.

As a basic example of this model, see Fig. 4.1. There is a log plot of the number of excitations versus time, with the purely exponential curve that would occur if no interactions between the atoms were allowed plotted for comparison. As can be seen, the interactions allow for excitations to stay within the system for many lifetimes longer than without inter-atom interactions, a phenomenom known as subradiance. While much work has been done to characterize the timescales of subradiance, we are not aware of numerical or analytic models that predict the entire curve, beginning with small deviations from the independent case and growing to diverge by orders of magnitude for longer times. This model does not predict superradiance, in which at very early times the excitations leave the system at a rate much greater than the independent.

4.3 Decoherence Simulation

In this section, we will numerically compare the decoherence time predictions made in Chapter 2 with a simulation of the exact Schrödinger Equation. We will calculate a modified decoherence time, defined as the time at which the off-diagonal matrix elements of a single subsystem's reduced density matrix shrinks below a fixed threshold in the computational basis. This differs from the decoherence time, whose threshold shrinks with N because it is defined as its typical value when averaged over all of its Hilbert space. The value of this constant threshold is unimportant for the scaling of N as long as it is $<< 1$. For numerical convenience we choose a constant threshold equal to $1/2e$, instead of the N-dependent threshold proposed in the definition of decoherence in Chapter 3. These simulations are meant as a consistency check between the exact Schrödinger Equation and the approximate analytic method, so what is important is that we choose the same threshold in both cases. $1/2e$ is chosen because the off-diagonal matrix elements are taken to have an

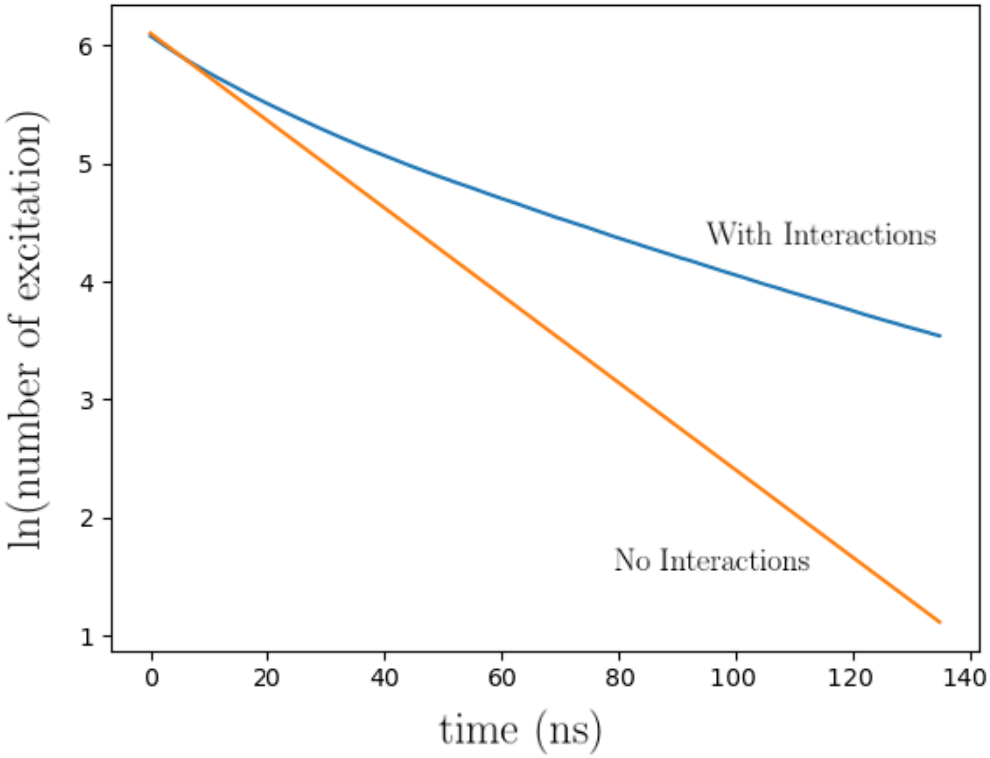

Figure 4.1: An example output of the Definite Model. This instance uses the physical parameters of an atomic MOT with 900 atoms, a lifetime of 27 ns, a transition wavelength of 780 nm and a gaussian spatial distribution with a $1/e^2$ waist of 5μm.

initial value of 1/2, their maximum value, and so the threshold represents the point at which the matrix element has dropped to $1/e$ times its original value.

For the exact simulation, we first choose random positions inside of a unit box for each of the N subsystems (a rescaling of the box amounts to a rescaling of the coupling parameter so there is no need for a second, independent parameter corresponding to the system size). If two subsystems happened to be chosen very near one another, then their interaction can be very large and drastically alter the properties of time evolution; to avoid this, a shell size was chosen such that no two subsystems could be nearer than that distance from one another. In the unit box, this shell was chosen to be of size .2, and similarly applies when choosing positions for the approximate comparison. We then randomly select one of the subsystems as the one whose reduced density matrix will be analyzed. According to this selection, we choose a random initial state for the system subject to the constraint that the off-diagonal matrix element for the chosen subsystem is 1/2. The state is then evolved in time according to the fourth order Runge-Kutta method until the threshold value of the off-diagonal reduced density matrix element is met. This evolution is performed hundreds of times for each value of N with different random positions and initial states, and the threshold times are averaged over. The standard deviation of the mean forms the error bars depicted in 4.2, where 'Full SE' refers to this exact simulation of the full Schrödinger Equation.

Since the Hilbert space dimension is 2^N, practical limitations meant that we could only simulate up to $N = 15$. Although we have not attempted to quantify when the large N behavior takes over, it is not hard to imagine that 15 is not large enough. Thus when calculating the spectrum, we will not only include patterns that matter for large N, as the analytic expression does, but actually all patterns. In this way, the smallness of N is made much less important. We will average over many random selections of subsystem positions, making the use of expectation values in calculating the contribution to the moment from each pattern justified in comparison. This large number of parallel computing jobs were performed on the High-Throughput Computing Cluster at UW-Madison.

To test the assumptions made by the approximate analytic methods used in the Chapters 2 and 3, we first calculate the average width of spectrum for a given value of N with respect to randomly chosen positions (subject to the same shell size). This can easily be done by directly calculating $\sum_q \langle q| \sum_{(jk)} H_{jk}^2 |q\rangle$. We then choose 450 eigenvalues from a gaussian with that width and numerically compute when the chosen off-diagonal matrix element decreases to the threshold of $1/2e$. This is done by taking the state to initially be comprised of 450 different components, each in phase and with phases rotating according to the randomly sampled eigenvalues. As the interference grows for $t > 0$, the magnitude of the matrix element decreases. This calculation is performed many times for each value of N with different eigenvalues sampled from the distribution according to the width for that N, using a range of values for this width according to its uncertainty. Since this is a much less computationally expensive task than the full Schrödinger Equation approach, it is not difficult to average over many more cases than in the full approach, making the error bars much smaller; for this reason they are not displayed.

We expect these two approaches to well approximate each other for large N, and in fact the results of this thesis depend on that correspondence. Indeed, such agreement is demonstrated in 4.2 where it can be seen that the agreement increases for larger N, as expected.

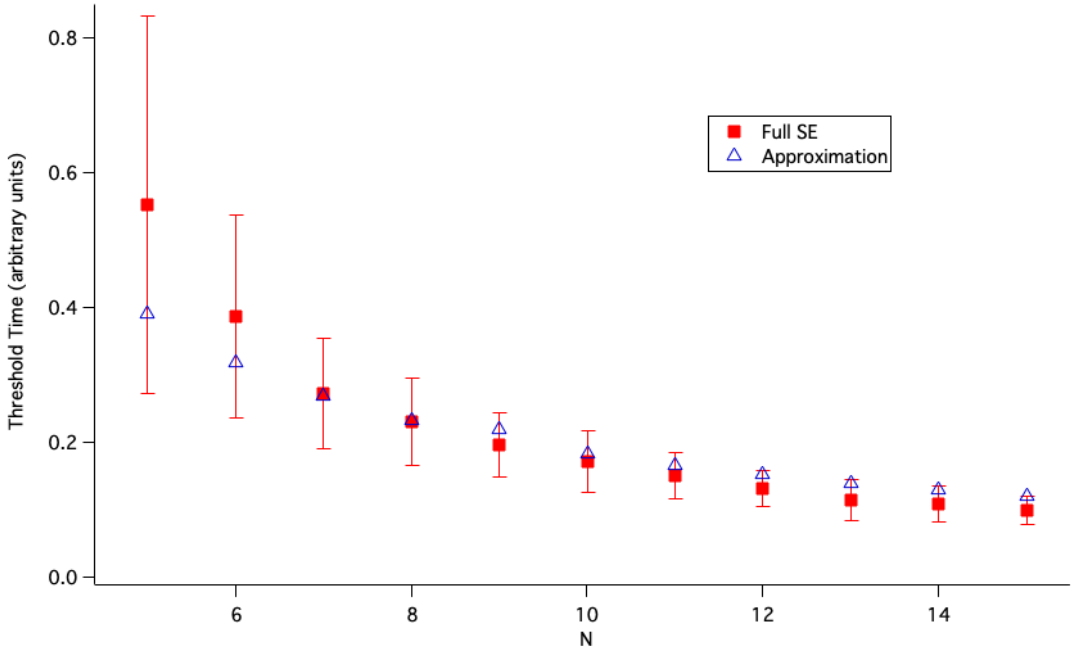

Figure 4.2: A comparison between how long it takes for an off-diagonal reduced density matrix element initialized to .5 to reach the threshold value of $1/2e$ for a simulation of the full Schrödinger Equation (labelled Full SE) and an approximate method described in the main text

5 System Induced Superselection

5.1 Introduction

The quantum measurement problem has long plagued any attempt to use quantum mechanics to describe how the universe really is, as opposed to simply using the math to predict the outcome of a measurement of a subsystem of the universe. The Copenhagen Interpretation, which stresses the measurement outcome paradigm, is widely taught. In this interpretation, an interaction with a classical observer takes place at certain times during the quantum evolution, projecting the quantum state onto one of its eigenvectors with probabilities according to the Born rule which is a postulate of the interpretation. Only the outcomes of these measurements can be determined through experiment, and so the underlying quantum states can be conceptually reduced to a mathematical device without saying in what sense they are real.

What is the difference between a quantum-quantum interaction and a classical-quantum measurement? In other words, what makes one system quantum and another classical? It is usually easy to tell since when doing an experiment the measuring device is always classical, allowing the theory to be useful in many cases. But efforts to make this classical-quantum distinction scientific have not been successful. Some success has been found in treating both the observer and observed system as quantum in the theories of decoherence and Environment Induced SuperSELECTION, or EINSELECTION. We will briefly summarize these ideas and their main flaws, some of which can be erased by switching from environment induced superslection to a new paradigm of system induced superselection. Abbreviated as sinselection, we will also discuss how this new idea is physically motivated by the results of Chapter 2.

5.2 Decoherence

Any quantum two-level quantum state $|\psi\rangle$ can be written $|\psi\rangle = a\,|1\rangle + b\,|2\rangle$ for two complex numbers a and b which obey $|a|^2 + b^2 = 1$. If a measurement is made on the $|1\rangle, |2\rangle$ basis then the probability of observing the system in state $|1\rangle$ ($|2\rangle$) is $|a|^2$ ($|b|^2$). Despite this fact, it is not possible to interpret this state as really being in $|1\rangle$ or $|2\rangle$ according to those probabilities because measurements made in a different basis can reveal the relative phase between a and b. In order to see quantitatively how such a quantum state compares with the state corresponding to $|1\rangle$ with probability $|a|^2$ and $|2\rangle$ with probability $|b|^2$, we must examine the density matrix $|\psi\rangle\langle\psi|$, another way of representing the state. Writing this state as a matrix in the $|1\rangle$, $|2\rangle$ basis, it is

$$\begin{bmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{bmatrix} \tag{5.1}$$

whereas a system that is in $|1\rangle$ ($|2\rangle$) with probability $|a|^2$ ($|b|^2$) is represented in the same basis by the density matrix

$$\begin{bmatrix} |a|^2 & 0 \\ 0 & |b|^2 \end{bmatrix} \tag{5.2}$$

It is therefore the non-vanishing value of the off-diagonal matrices that prevents us from interpreting the quantum system as really having a definite state, with uncertainties only coming about due to the ignorance of the observer. This is in stark contrast to the outcome of every measurement, in which the quantum state is always found to be in an eigenvector of the measurement operator, never a superposition. Since the quantum state just prior to measurement cannot be interpreted as actually in such a state, the Copenhagen Interpretation adds the measurement postulate which says that the system undergoes non-unitary evolution when interacting with a classical observer wherein that state collapses to one of the eigenvectors of the measurement operator. But this effect can also be partially

understood in a fully quantum treatment according to decoherence and einselection, as follows.

When a measurement in the $|1\rangle$, $|2\rangle$ basis is made, the system interacts with the environment (which includes the measuring instrument) in such a way that if the system were in $|1\rangle$ the environment goes into one set of states, for example the set of states with a dial pointing to '1', and a different set of states if the system is in $|2\rangle$. Classically, then, we may observe this piece of the environment, usually some kind of meter or display, to infer the state of the system at the time of measurement.

However, if we wish to model the environment as a large quantum system, then the interaction has the form

$$|\psi\rangle \otimes |\epsilon_0\rangle = \big(a\,|1\rangle + b\,|2\rangle\,\big) \otimes |\epsilon_0\rangle \rightarrow a\,|1\rangle \otimes |\epsilon_1\rangle + b\,|2\rangle \otimes |\epsilon_2\rangle \tag{5.3}$$

where $|\epsilon_0\rangle$ is the initial state of the environment, and $|\epsilon_{1(2)}\rangle$ is its final state if the system was in the state $|1\rangle$ ($|2\rangle$).

For what follows we will need to distinguish between the observer, which is a part of the environment, and the rest of the environment. We define the observer states analogously to $|\epsilon_{0,1,2}\rangle$ as $|\eta_{0,1,2}\rangle$ and let $|\epsilon_{0,1,2}\rangle$ now stand for the environment minus the observer/apparatus. This measurement process is now written as

$$|\psi\rangle \otimes |\eta_0\rangle \otimes |\epsilon_0\rangle \rightarrow a\,|1\rangle \otimes |\eta_1\rangle \otimes |\epsilon_1\rangle + b\,|2\rangle \otimes |\eta_2\rangle \otimes |\epsilon_2\rangle \tag{5.4}$$

The tensor product structure of the initial and final states is highly suspect, but a common simplification or assumption made in the literature. If the observer is ignorant of the state of the rest of the environment, their best predictions made about the system will not be represented by its state vector, since that vector is now entangled with an environment whose exact state they do not know. The mathematical object representing their best predictions in the face of this ignorance is the reduced density matrix, which is made by constructing the density matrix of the entire system and tracing out the unknown, environmental degrees of freedom.

For some environmental basis $|E_i\rangle$, this reduced density matrix ρ is expressed as

$$\rho = \sum_i \langle E_i| \Big[|a|^2 |1\eta_1\epsilon_1\rangle \langle 1\eta_1\epsilon_1| + |b|^2 |2\eta_2\epsilon_2\rangle \langle 2\eta_2\epsilon_2| + ab^* |1\eta_1\epsilon_1\rangle \langle 2\eta_2\epsilon_2| + a^*b |2\eta_2\epsilon_2\rangle \langle 1\eta_1\epsilon_1| \Big] |E_i\rangle \tag{5.5}$$

where the shorthand $|\psi\eta\epsilon\rangle$ is used for $|\psi\rangle \otimes |\eta\rangle \otimes |\epsilon\rangle$. The trace sum over $|E_i\rangle$ can be done in any basis, and we can always construct a basis with the two states $|\epsilon_{1,2}\rangle$ as elements and the rest of the basis states orthogonal to those two (as long as $|\epsilon_0\rangle$ and $|\epsilon_1\rangle$ are not the same state, which they are not by assumption - otherwise there is no interaction between the system and environment). Therefore this may be written as

$$\rho = \sum_{i=1,2} \langle \epsilon_i| \Big[|a|^2 |1\eta_1\epsilon_1\rangle \langle 1\eta_1\epsilon_1| + |b|^2 |2\eta_2\epsilon_2\rangle \langle 2\eta_2\epsilon_2| + ab^* |1\eta_1\epsilon_1\rangle \langle 2\eta_2\epsilon_2| + a^*b |2\eta_2\epsilon_2\rangle \langle 1\eta_1\epsilon_1| \Big] |\epsilon_i\rangle \tag{5.6}$$

For any individual environmental degree of freedom, it's not clear that the two possible outcomes of the system lead to states orthogonal to each other - however if the environment is made up of a huge number of degrees of freedom, then it is reasonable to assume that the possible overall states $|\epsilon_{1,2}\rangle$ are very close to orthogonal. Mathematically this can be stated as $\big|\langle \epsilon_1 \mid \epsilon_2\rangle\big| << 1$. This quantity is the overlap between the two states; if it is zero they are orthogonal, and if it is one they are parallel. For example, if the environment is made out of M subsystems $|l\rangle$ which each evolve into final states $|l_{1,2}\rangle$ which are almost parallel, $\big|\langle l_1 \mid l_2\rangle\big| = .9$, then the full overlap $\big|\langle \epsilon_1 \mid \epsilon_2\rangle\big| = .9^M$ is already as low as .0003 for $M = 100$. Real environments can easily have 10^{26} or more degrees of freedom, making the approximate that $\langle \epsilon_1 \mid \epsilon_2\rangle << 1$ a very good one. In that case, the reduced density matrix is approximately $\rho = |a|^2 |1\eta_1\rangle \langle 1\eta_1| + |b|^2 |2\eta_2\rangle \langle 2\eta_2|$. The off-diagonal elements have been suppressed by the environment, allowing us to think of this state as either 'the system is in state $|1\rangle$ and the observer measured that outcome and thus is in the state $|\eta_1\rangle$, or similarly for $|2\rangle$, which is consistent with how we experience experiments. Thus it can be understood entirely within a

quantum framework how the outcomes of experiments appear classical.

However, the off-diagonal suppression and quantum-to-classical transition cannot be thought of as actually happening - it only appears that way to an observer when they are ignorant of their environment. This is because we can consider a second observer measuring the first who can perform experiments which conclusively show that the first observer is still in a superposition of outcomes [46]. This is made consistent with experience by noting that, relative to each branch of the observer state $|\eta_{1,2}\rangle$, the system appears to be in the state $|1\rangle$ or $|2\rangle$, but it shows that the quantum superposition never disappears, only becomes obscured.

Experiments suggest that even large objects display quantum properties [47], calling into question the essential classicality of any observer. To take the Schrödinger Equation as universal and treat all physical systems as quantum leads to the controversial Many-Worlds picture first put forward by Everett, which contains no inconsistencies but is considered by many to be a dissatisfying theory of the physical world, largely because of its difficulty in explaining the Born rule [48]. Another approach is to adopt the point of view of QBism in which state functions and probabilities are considered purely subjective; this solves some formal issues but at a price [49, 50].

Here we will use the mathematics of einselection, combined with the results of Chapter 2, to motivate superselection rules induced by the system itself instead of its environment. We will argue that this effective collapse of a system's wave function due to itself allows for the collapse to be considered to be actually happening without running into the contradiction mentioned above. The exchange interaction is present in all composite quantum systems where the subsystems can exchange quanta of energy. This interaction arises from a coupling to a boson bath, but is different from an environment interaction in the sense of einselection. For the exchange interaction, the boson bath is taken to be in vacuum at all times - it is only the existence of the boson degrees of freedom that gives rise to the interaction between the system and itself. In particular, for the exchange interaction alone, correlations are not generated between the system and the boson bath; evolution

under its Hamiltonian leaves an initially product state in a product state. In einselection, the non-product state structure of the system and environment, which is typically formed after an initial product state, is central to all its results about superselection and measurement statistics. Another difference between the exchange interaction and the environmental interactions considered in einselection is that for the exchange interaction the boson bath is taken to be at zero temperature; in contrast, a common calculation in einselection shows that the reduced density matrix of a system in a superposition of two different spatial locations becomes diagonal when interacting with a large environment on a timescale which is exponentially small in both the temperature of the environment and the separation squared [51]. Thus for a zero temperature environment no such decoherence occurs, we will show that exchange interaction causes similar decoherence except with no energy in the environment and no energy exchange between the environment and system.

5.3 Einselection

In the theory of Einselection [51], a special basis is shown to be selected by the environment. Over timescales which are extremely short for realistic environments, through the framework of decoherence discussed above, the reduced density matrix of the system assumes a diagonal form in this special basis, replicating the statistics of a classical mixture of the special states. This basis is chosen by the environmental interaction, or more briefly we say that it is einselected; it can be identified in many ways but here we will focus on the predictability sieve. More precisely the einselected basis should be an eigenvector of the system's entire Hamiltonian, but during most measurements the interaction dominates. The idea behind the predictability sieve is that states can be assigned a predictability based on how quickly they are altered by the interaction Hamiltonian. The energy eigenvectors are not changed at all in time, while superpositions of eigenvectors can change extremely rapidly as the different eigencomponents become out of phase with one another. Since the global phase is unphysical, then even though an eigenstate's phase is rotating like $e^{-i\lambda t}$ if

it's eigenvalue is λ, this rotation can never be detected. Only when a superposition occurs is the dephasing physical (this is why we have focused on the width of the energy spectrum, as opposed to the magnitude of the eigenvalues themselves). The rate of this dephasing grows with the environment size and temperature and is extraordinarily short for realistic systems and environments.

Not only are energy eigenvectors stable, but they also can be said to have an objective existence in the sense that two different observers can both make repeated measurements of which eigenstate the system is in and agree every time (quantum states are altered through measurement by projection onto the energy eigenstates, which has no effect on a state that is already an eigenstate). The idea behind einselection is that the predictable or stable states are superselected by the environment and the unpredictable or unstable states are disallowed. If a non-eigenvector state of the system-environment composite system did exist at some moment in time, then a very short amount of time later it would be altered dramatically, disallowing it from being observed in the traditional sense.

Many interactions depend on position, meaning the Hamiltonian is made from position operators. In that case the position operator will commute with the Hamiltonian, making the position basis an energy eigenbasis. This is why states which are well localized in position are stable and objective in the sense that different observers can measure the position of the same object and agree on the result. If, however, we try to put a system in a superposition of position eigenstates, then it will be extremely unstable and not conform to how macroscopic humans experience states to behave.

Of course, quantum states which change in time are commonplace, but they can only be used for prediction when either the interaction is precisely known, or the timescales of prediction are short enough that the unknown part of the interaction does not completely scramble the state. In the case of a large environment, the interaction is both unknown and happens over very fast time scales, making the superselection rules a very good approximation.

To posit that this apparent collapse physically happens creates problems; if the physical collapse has to do with the interaction with the environment/observer, then some distinction must be made between fully quantum systems and decohering environments, and it is unclear how to do so in a conceptually and quantitatively satisfying way, leading back to the problem with the Copenhagen Interpretation. One could instead hypothesize that the physical collapse takes place in the system before the interaction with the environment, but the mathematical and physical motivation for such a collapse comes from the environmental interaction, making it unclear how or why such an intrinsic collapse would occur. We will argue here that the results from Chapter 2 actually do motivate such a collapse in the system itself before any measurement is made.

5.4 Sinselection

First, we will discuss how sinselection comes about in simple and arbitrarily large quantum systems, then discuss why it might occur in any finite system with a finite timescale.

5.4.1 Arbitrarily Large Systems

Consider a collection of N two level quantum subsystems where N is arbitrarily large. For simplicity, we assume each of the N two-level systems have the same energy difference, but allowing for detuning will not change the result as long as we assume that we cannot have infinitely large energy differences between any ground and excited state. Assuming this, as N grows without bound the number of other subsystems within some threshold of resonance of any given subsystem also grows without bound, allowing for the resonant case to apply.

For any such collection of two-level systems there will be some coupling to a common boson bath, even if it is very small, leading to the effective pairwise interactions between the subsystems of the form $H_{jk} = F_{jk}(\sigma_j^+\sigma_k^- + \sigma_j^-\sigma_k^+)$ where $\sigma_{j,k}^{\pm}$ is the

raising or lowering operator for the subsystem labelled by index j, k [52–54]. This same coupling also leads to dissipation effects like spontaneous and stimulated emission, but this thesis will focus on the effect on how the system interacts with itself, not the environment.

We will also assume the subsystems do not become arbitrarily dense, meaning the collective system has arbitrarily large spatial extent. This important assumption is physically motivated because, if a system becomes too dense, it will collapse into a black hole. Even if the interactions between a given pair of subsystems are more complex for non-localized, overlapping two-level systems, for almost all pairs of subsystems j and k the pairwise interactions will depend on their separation r_{jk} like $\sim 1/r_{jk}$ to leading order in $1/r_{jk}$ because for arbitrarily large N almost all pairs of subsystems are very far away from each other. This interaction is due to the environment but is crucially different from the type of interaction considered in einselection because it is intrinsic to all systems and the resulting dynamics can be written only in terms of the system itself; accordingly it preserves any initial tensor product structure between the system and environment. It does not generate correlations between the two systems - the boson bath stays in vacuum the whole time. Interaction with the environment necessarily increases dramatically when a measurement is made, but the kind of environmental effect which causes the exchange interaction is independent of measurement, distinguishing sinselection from einselection in a fundamental way.

As shown in Chapters 2 and 3, as N grows the timescale of the interaction Hamiltonian dynamics shrinks without bound. For large N the eigenvectors will be unknown to the observer; for example, for $N = 400$ it would require the best possible computing hardware filling the entire observable universe to simply write down a single eigenvector of such a system, much less find it in the first place or write down all 2^{400} of them [32]. When the eigenvectors are unknown, the effect of the interaction Hamiltonian cannot be predicted, and when N is arbitrarily large its effect happens arbitrarily quickly. Thus system states which are not eigenvectors are unpredictable in the sense of the predictability sieve; they suffer the

same instability that causes such superpositions to be disallowed in the einselection paradigm. If the superselection rule is valid there, which seems to be the case as such superpositions are never observed post measurement, it will also be valid internally to such a large system. Here the special basis is the eigenbasis of the system's own Hamiltonian, which as discussed is dominated by H_{int} due to boson coupling for large N. If a measurement is made in a different basis, the system's Hamiltonian will change during such a measurement and it will become an eigenvector of that new Hamiltonian, allowing for consistency between sinselection and einselection.

This can be thought of as the system losing information about its phases in the eigenbasis. If these phases are rotating arbitrarily quickly according to H_{int} then any finite time observation will necessarily see averaged phases, putting the outcome of experiments on such systems in accordance with classical uncertainty. As noted, even though the eigenvalue itself sets the rate of phase rotation, since the overall phase of the system is unobservable the relevant quantity is the spread of eigenvalues i.e., how fast their phases are changing with respect to each other. When these phases are unknown for any reason they must be averaged over, and this results in a diagonal reduced density matrix in the energy eigenbasis. For an arbitrarily large system, since the action of the unknown H_{int} is arbitrarily fast, we can say that this process takes place immediately. In other words, arbitrarily large systems cannot display quantum behavior. This notion seems consistent with our experience, but is not helpful until we discuss how it applies to finite systems.

5.4.2 Finite Systems

When N is finite, H_{int} will have some finite characteristic timescale which can be viewed as the timescale of sinselection. For large N it can be very fast, just like a large environment, such that, to humans and the measuring devices we make, it appears to happen almost instantaneously. For smaller N the decoherence caused by a system on itself might compete with quantum effects, allowing them to observed. For now the decoherence is only effective, just as in einselection - it is not until

the next section that we take the newly possible next step and posit its objective reality. Thus it is acceptable at this point for it to depend on the observer's knowledge, which is also as in einselection where it is assumed the observer does not have detailed knowledge of its large environment. If the observer does not know the eigenvectors of H_{int}, their predictions which rely on quantum interference will become increasingly vague or inaccurate as time goes on, on a timescale generically governed by the width of H_{int}. This is the timescale at which the density matrix (note this is not the reduced density matrix as there is no environment being traced over) becomes diagonal in its energy eigenbasis for that observer, in the sense that that observer's best guesses about the outcome of any experiment done on the system is summarized in a density matrix which is diagonal in the energy eigenbasis.

For $N < 400$, it is possible that the observer knows the eigenvectors of H_{int}. However there will always be some uncertainty in the eigenvalues as infinite precision is not possible. After an amount of time t, the uncertainty in the energy ΔE will become an uncertainty in the phase of that eigenvector component equal to $t\Delta E \bmod 2\pi$. When t is large enough, the phase will be entirely uncertain and the experimenter must average over them, resulting in a diagonal density matrix in the energy eigenbasis. Thus eventually only classical behavior can be predicted for such a system i.e., it can be considered to actually be in one of its eigenvectors with different classical probabilities for each eigenvector. In particular, because the timescales of the intrinsic H_{int} were shown to grow without bound with system size in, this implies that all quantum phenomena are necessarily relegated to either small systems or small timescales, explaining why it is rare to find on human scales.

Even if the eigenenergies were somehow known exactly, the duration between state preparation and measurement cannot be due to fundamental constraints on timekeeping [55]. In this case one must average the phases over some time window according to the range of durations the experiment at hand might represent, and for long enough times the smallest possible time uncertainty will be enough to entirely randomize the phases due to the modulo 2π, making the system again appear classical.

Another constraint on systems of any size is that their positions cannot be exactly known, leading to a fundamental uncertainty in the pairwise interaction energies and ultimately to the same averaging and diagonalization as above. Of course this applies to all systems of all sizes, even those dominated by H_{int} for large N, but were it not for H_{int} causing a much faster effective collapse, quantum interference effects could exist on large scales for arbitrarily long times, rendering the framework no better than Many Worlds. So we have observed both that the the destruction of quantum interference is ubiquitous, and that for a certain but general class of large systems this destruction will take place fast enough that for large N, quantum interference will be difficult or impossible to detect.

Although there is no physical mechanism within unitary quantum mechanics to actually cause this diagonalization, it will always look like it happens to any observer; this is exactly as in the einselection case. When something looks like it happens to every conceivable observer, we usually simply say that it happens - however, for einselection this is not possible because it causes a contradiction when one considers a second observer observing the first. In sinselection no such paradox occurs, allowing the collapse to be physical. In fact, since it will appear to happen to every conceivable observer, it seems unreasonable to not conclude that it is actually happening, absent any specific reason not to. In the next section we hypothesize a modification to the Schrödinger Equation to formalize this physical process, and explore the benefits of such a hypothesis.

5.4.3 Stochastic Wavefunction Collapse

What if the composite quantum system is not a collection of two level systems, either because its constituents are more complex or it has only one constituent? Many of the arguments made above are general and might apply to these as well. Since there is no longer a need for an appeal to the observer to explain the wave function collapse, we are free to posit that it is actually happening. The most natural hypothesis is that the state of any system stochastically approaches one

the eigenvectors of its Hamiltonian in addition to its normal unitary evolution. We call this the Delayed Definite Eigenvector Hypothesis. The probability of ending in any given eigenvector is equal to the state's overlap with that eigenvector, which is constant with respect to the unitary evolution.

We leave open the specifics such evolution, opting to instead merely establish that it has many desireable qualities: it does not lead to a measurement paradox, it does not draw a fundamental distinction between measurements and other types of quantum interactions, it is physically motivated to apply generically, it is empirically consistent with the theory of decoherence and einselection, and it can be applied to a universal function without interpretational confusion. Although there are many interpretations of quantum mechanics or wave function collapse which have some of these properties, I am not aware of any that have them all. In the next section we discuss this last property, about which nothing has yet been said. There is much work to be done to propose and evaluate specific stochastic mechanisms, but the main result here is that such a mechanism is possible without leading to contradiction or divisions between classical and quantum, or observer and observed, and also could plausibly be consistent with the experience of all possible observers.

5.4.4 Implications for Quantum Cosmology

When confronted with the universal wave function, einselection falters because there is no observer outside of it. In contrast, sinselection naturally extends to the universe as a whole, which is treated no differently than any quantum system. Evidently the universe is not currently, and therefore did not start in one of its energy eigenstates (according to the Delayed Definite Eigenvalue Hypothesis). This is demonstrated every time quantum interference is observed. But the universal wave function is slowly approaching one of these eigenstate, which is demonstrated every time quantum interference is destroyed. While by definition this destruction can only be detected through measurement, in the sinselection paradigm it is always happening to quantum systems whether they are observed or not.

Why, then, can quantum phenomena be generated, for example by manipulating systems to carefully create quantum interference? At first this appears to be in conflict with the idea that all quantum systems are approaching one of their energy eigenvectors. But these interferences can only be set up and measured through the careful manipulation by other systems, usually humans or human-made machines. Thus the wave function describing both the system and its manipulator is approaching one of its energy eigenstates, but the manipulation is done in such a way that the state of a subsystem moves away from one its sub-eigenstates. This is analogous to the explanation of why humans can create low entropy objects, such as computers or houses or our own bodies. Entropy can locally decrease, as long as in a wider perspective it is still increasing. It is well known that any entropy decrease caused by a human or machine is well accounted for by the entropy increase associated with heat being added to an environment, a necessary byproduct of doing work. Similarly, the local movement away from an energy eigenvalue is dwarfed by a global movement towards one.

Like in the case of entropy, the existence of any quantum phenomena implies something about the initial conditions of the universe - in particular, that it did not start in or too near to an energy eigenvector. However, unlike in the case of entropy, this initial condition is generic because almost all elements of Hilbert space are superpositions of many different energy eigenvectors.

This framework also could give insight into the conceptual problem implied by the Wheeler-DeWitt equation for the universal wavefunction, $H|\psi\rangle = 0$, which seems to say that nothing ever happens, contrary to our everyday and scientific experiences [56]. But if there were stochastic evolution towards an energy eigenstate, something would be happening to the universal wave function. It would seem that right now we are near enough to an eigenvector that quantum phenemona is suppressed on large scales, but still far enough from one that it is present on small scales. Once the universe does reach one of its energy eigenvectors, indeed nothing will happen anymore; this is a quantum heat death of the universe.

5.5 Conclusions

We have physically motivated a conceptually satisfying picture of quantum mechanics, similar to einselection but with greater ontological aspirations. In this picture, quantum effects are inherently transient, dissipating over timescales strictly limited by the system size. Much like how the entropy of the universe is always increasing, the quantumness of the universe (defined by how far a state is from one of its energy eigenstates) is always decreasing. And just like how local areas of low entropy can be found naturally or made intentionally, it is possible to find or create quantum phenomena on small scales. Thus the universe is essentially classical in nature - not that it lives in a classical configuration space but that it is driven towards its energy eigenstates which are, unlike the rest of Hilbert space, objective, definite, and stable.

6 Conclusion

Motivated by a shortcoming of the Threshold Theorem in Quantum Error Correction in Chapter 1, we found the spectrum of the exchange interaction $H_{\text{int}} = \sum_{(jk)} F_{jk}(\sigma_j^+\sigma_k^- + \sigma_j^-\sigma_k^+)$ for couplings of the form $F_{jk} = \alpha \sin(kr_{jk})/r_{jk}$ and in the large N limit in Chapter 2. Equipped with the spectrum, we analytically studied the decoherence of a single subsystem within the ensemble in Chapter 3. We verified some of our approximations numerically, and used numerical methods to predict the time evolution of systems evolving according to the exchange interaction in Chapter 4. Finally, in Chapter 5 we looked at the implications for the measurement problem, and postulated how the scaling features of the ubiquitous exchange Hamiltonian could explain the lack of observed macroscopic quantum phenomena.

More work could be done related to the derivation of the energy spectrum. What is the spectrum when there is an exponent of position other than -1 in the coupling F_{jk}, or even when it takes a different form altogether? In order to apply to a range of cases, there are many areas where the derivation might be improved or adjusted. The transformations used and final set $\{S_t\}$ may need to be altered, although if most patterns cannot be discarded, it seems unfeasible to calculate the necessary quantities such as $\Omega'(S)$, which can be extremely complicated for patterns not in the $\{S_t\}$ used here. In order to have a smaller $\{S_t\}$, different F-conditions could be used, and different techniques used to prove them. The upper bound we placed on the F-condition used here was weak, and although it was sufficient to prove what we needed here, it suggests that the methodology is flexible.

We made no attempt here to estimate how accurate the derived spectrum is for finite N, focusing on the simpler fact that it becomes arbitrarily accurate for arbitrarily large N. But any physical test of this scaling will involve finite N so in order to make predictions of relevance to experiments this would need to be done. As real quantum computers are being built and increasing in size, when will this effect be seen and how big will it be? As noted, this will depend sensitively on the specifics

of the quantum computer implementation. For architectures in which spontaneous emission is a very small error, the collective effect will likely also be very small until N is much larger than is currently possible, meaning its effect won't be noticed for a very long time if at all. But for architectures in which spontaneous emission is a mains source of error, it is possible that it's collective nature will become relevant in the not-too-distant future.

The single subsystem studies in Chapter 3 could be taken further. One direction of interest is in applying the same formalism to the scrambling time, which is how long it takes for initial information in a single subsystem to propogate to all the other subsystems [57]. The decoherence can be thought of similarly; it is how long it takes initial information in a single subsystem to leave that subsystems. Thus the scrambling time is always greater than the decoherence time. In between these two times, the information in the initial subsystem has leaked out but has not yet spread to all the other subsystems, instead staying localized in other subsystems near the original subsystem.

The numerical studies in Chapter 4 could be improved and extended. Are there more realistic models which are still fast enough to run for large N? Another outstanding question that arises when doing such simulations in relation to the spectrum is: when an eigenstate of the interaction decays exponentially according to the complex part of its eigenvalue, what state is it decaying into, or more generally how many subsystem excitations leave the system? Due to the collective nature of the interaction it seems that the answer is likely not 1. If there was a good way to estimate this quantity, numerical studies more closely linked to the spectrum would be possible.

Perhaps the area with the most room for growth, and the most challenging to make progress in, is Chapter 5. In that chapter we posited a framework called sinselection and discussed how its basic properties were physically motivated and could resolve the measurement paradox. However, the details are left unspecified, leaving many open questions. To name a few: in what manner does a system's state stochastically approach an energy eigenvector? How is the final eigenvector selected? What is the

general timescale of this process, even when H_{int} is not present?

Another intriguing possibility is to hypothesize that systems always evolve toward their lowest energy eigenstate. This is motivated by the universal dissipation felt by all subsytems of the universe; its implications or validity when applied to the universal wave function could be of great interest. It also could possibly explain the ubiquity of the area law for quantum many-body entanglement. This is a property that many low-lying energy many-body quantum states have, and means that the entanglement entropy between the subsystems contained in a spatial volume and the rest of the system scales as the surface area of that volume, not the volume itself [43]. This is surprising because, for almost all of Hilbert space, this entanglement should scale with volume. But the black hole area law, which states that the entropy of a black hole is proportional to its surface area, suggest that this law may hold for all systems, including quantum ones. Like in the quantum case, classically it is not immediately clear why the area law should hold; if I double the volume of a gas while holding its other properties constant, its entropy doubles - entropy scales with volume, not area. But it turns out that gravitational collapse prevents the formation of a classical state that violates the area information bound. This means that, if one keeps increasing the volume of the gas, for a while the entropy will scale with volume so that you're on track to beat the bound for a large enough volume, but before reaching that volume the gas will collapse into a black hole. Thus in the quantum case, since low-lying energy states are the ones that obey the area law, there may be some similar physical principle that keeps the state in that part of Hilbert space. It is suspected that this may have to do with the structure of spacetime and how it is built out of quantum entanglement, a new idea with much promise [58, 59]. Sinselection could provide this physical principle, if the Delayed Definite Eigenstate hypothesis were altered to include a tendency towards its lowest energy eigenvalue. However note that this is a separate timescale, as the outcome of every measurement is not necessarily the one with the lowest energy.

In short, there are many ways this work could be extended. There are many ways to use an operator's eigenvalues to understand its properties. Of course, a

complete understanding of an operator requires a knowledge of both its eigenvectors and eigenvalues, but unfortunately finding the eigenvectors of H_{int} appears to be intractable. However, partial information about a wide variety of many-body systems might be gleaned using this technique.

Appendices

A The distributions and expectation values of various powers of the exchange coupling constants

In this Appendix, we discuss the probability density functions and the expectation values of various powers of the exchange coupling constants $F_{jk} \propto \sin \kappa_a r_{jk}/(\kappa_a r_{jk})$ for both the 2D and 3D geometries. For large distances $\sin \kappa_a r_{jk}$ term essentially produces a random sign, whose square contributes a factor of $1/2$ to the expected value. We therefore focus on the remaining $1/(\kappa_a r_{jk})$ term in this section. We define a random variable $U \equiv 1/(\kappa_a^2 r_{jk}^2)$ (which are proportional to the square of the exchange coupling constant $U \sim F_{jk}^2$) and first discuss the probability density function and the expected value of this random variable. We will then discuss the distributions and the expected values of various powers of U.

A1 Two-dimensional geometry

Consider a two-dimensional array of subsystems in a square geometry, in $x - y$ dimensions. Note that $r_{jk}^2 = x_{jk}^2 + y_{jk}^2 = (x_j - x_k)^2 + (y_j - y_k)^2$ where (x_j, y_j) and (x_k, y_k) are the coordinates of the j'th and k'th subsystems respectively. We view each of these coordinates x_j, x_k, y_j, and y_k to be uniformly distributed random variables within the interval $[0\ N^{1/2}d]$. Taking these initial uniformly distributed random variables, the probability density function of the random variable $U \equiv 1/(\kappa_a^2 r_{ij}^2)$ can be found using the methods outlined in [60]. Defining $f_U(u)du \equiv$

$P\{u \leq U \leq u + du\}$, this probability density function is:

$$\begin{aligned} f_U(u) &= 0 \qquad \text{if} \quad u < 0 \\ &= 0 \qquad \text{if} \quad 0 \leq u \leq \frac{1}{2L^2} \\ &= \frac{1}{u^2}\left[\frac{2}{L^2}\arcsin\left(2L^2u - 1\right) - \frac{2}{L^2} - \frac{1}{uL^4} + \frac{4}{L^3}\sqrt{\frac{1}{u} - L^2}\right] \qquad \text{if} \quad \frac{1}{2L^2} \leq u \leq \frac{1}{L^2} \\ &= \frac{1}{u^2}\left[\frac{\pi}{L^2} + \frac{1}{uL^4} - \frac{4}{\sqrt{u}L^3}\right] \qquad \text{if} \quad \frac{1}{L^2} \leq u \leq \frac{1}{\kappa_a^2 d^2} \quad . \end{aligned} \tag{A.1}$$

Here, the quantity $L \equiv N^{1/2}\kappa_a d$ is the phase accumulation over the full length of the square. Using the density function of Eq. (A.1), the expectation value of U can be found, which gives, in the $N \to \infty$ limit:

$$E[U] = \int f_U(u) u du = \frac{\pi}{\kappa_a^2 d^2}\frac{\ln N}{N} \quad . \tag{A.2}$$

Using the probability density function of the random variable U, we can find the distributions of various powers of U (such as $\sqrt{U}$, $U^{3/2}$, and so on) again using the methods outlined in [60]. These distributions are then used to find the corresponding expected values various powers of the exchange coupling constants. These results are:

$$\begin{aligned} E[|F|] &= \frac{\Gamma}{\kappa_a d}\frac{2\pi - \frac{10}{3}}{N^{1/2}} \quad , \\ E[F^2] &= \pi\frac{\Gamma^2}{\kappa_a^2 d^2}\frac{\ln N}{N} \quad , \\ E[F^n] &\sim \frac{\Gamma^n}{\kappa_a^n d^n}\frac{1}{N} \text{ for } n > 2 \quad . \end{aligned} \tag{A.3}$$

A2 Three-dimensional geometry

Here, we consider a three-dimensional array of cubits in a cube geometry, in $x-y-z$ dimensions and we have $r_{jk}^2 = x_{jk}^2 + y_{jk}^2 + z_{jk}^2 = (x_j - x_k)^2 + (y_j - y_k)^2 + (z_j - z_k)^2$. Following the above discussion, taking all the coordinates to be uniformly distributed,

the probability density function of $U \equiv 1/(\kappa_a^2 r_{jk}^2)$ is:

$$
\begin{aligned}
f_U(u) &= 0 \qquad \text{if} \quad u < 0 \\
&= 0 \qquad \text{if} \quad 0 \le u \le \frac{1}{3L^2} \\
&= \frac{1}{u^2}\int_{1/u-L^2}^{2L^2}\left[\frac{2}{L^2}\arcsin\left(\frac{2L^2-t}{t}\right)-\frac{2}{L^2}-\frac{t}{L^4}+\frac{4\sqrt{t-L^2}}{L^3}\right]\left[\frac{1}{\sqrt{\frac{1}{e}-tL}}-\frac{1}{L^2}\right]dt \\
&\qquad \text{if} \quad \frac{1}{3L^2} \le u \le \frac{1}{2L^2} \\
&= \frac{1}{u^2}\int_{1/u-L^2}^{L^2}\left[\frac{\pi}{L^2}+\frac{t}{L^4}-\frac{4\sqrt{t}}{L^3}\right]\left[\frac{1}{\sqrt{\frac{1}{u}-tL}}-\frac{1}{L^2}\right]dt \\
&\quad +\frac{1}{u^2}\int_{L^2}^{1/u}\left[\frac{2}{L^2}\arcsin\left(\frac{2L^2-t}{t}\right)-\frac{2}{L^2}-\frac{t}{L^4}+\frac{4\sqrt{t-L^2}}{L^3}\right]\left[\frac{1}{\sqrt{\frac{1}{u}-tL}}-\frac{1}{L^2}\right]dt \\
&\qquad \text{if} \quad \frac{1}{2L^2} \le u \le \frac{1}{L^2} \\
&= \frac{1}{u^2}\left[\frac{2\pi}{\sqrt{u}L^3}-\frac{3\pi}{eL^4}+\frac{4}{u\sqrt{e}L^5}-\frac{1}{2u^2L^6}\right] \qquad \text{if} \quad \frac{1}{L^2} \le u \le \frac{1}{\kappa_a^2 d^2} \quad . \qquad \text{(A.4)}
\end{aligned}
$$

Here, the probability density function in some of the regions cannot be evaluated analytically and as a result they are left in the integral form (the quantity t is the integration variable). The quantity $L \equiv N^{1/3}\kappa_a d$ is again the phase accumulation over one length of the cube. We numerically find that for the calculation of the expectation value, the majority of the contribution comes from the $\frac{1}{L^2} \le u \le \frac{1}{\kappa_a^2 d^2}$ region which can be evaluated analytically:

$$E[U] = \int f_U(u) u du = (\pi + \frac{29}{12})\frac{1}{\kappa_a^2 d^2}\frac{1}{N^{2/3}} \quad . \qquad \text{(A.5)}$$

The exact numerical result that includes all the regions of the probability density function [given in Eq. (A.4)] differs from the analytical result of Eq. (A.5) only by 1.2%. Following the 2D discussion, with the probability density function $f_U(u)$ known, the distribution of the various powers of U can be evaluated and the corresponding expectation values for the coupling constants are (again with an

accuracy at the level of few percent):

$$\begin{aligned} E[|F|] &= \frac{\Gamma}{\kappa_a d}\frac{2\pi - \frac{9}{5}}{N^{1/3}} \quad , \\ E[F^2] &= (\pi + \frac{29}{12})\frac{\Gamma^2}{\kappa_a^2 d^2}\frac{1}{N^{2/3}} \quad , \\ E[F^3] &= \frac{4\pi}{3}\frac{\Gamma^3}{\kappa_a^3 d^3}\frac{\ln N}{N} \quad , \\ E[F^n] &\sim \frac{\Gamma^n}{\kappa_a^n d^n}\frac{1}{N} \text{ for } n > 3 \quad . \end{aligned} \tag{A.6}$$

In Fig. A2.1, we plot the probability density function $f_U(u)$ for both the 2D (dashed red line) and 3D (black solid line) geometries. The functions are plotted for the case when the length of each side is $L = 1$. For an arbitrary L, the horizontal axis of the plot is scaled by $1/L^2$ whereas the vertical axis is scaled by L^2.

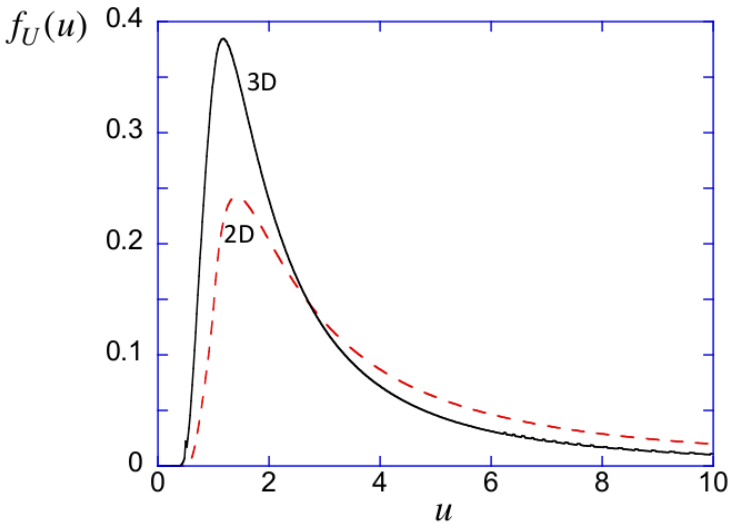

Figure A2.1: The probability density functions of the random variable $U \equiv 1/(k_a^2 r_{ij}^2)$ for an array of atoms in two (dashed red line) and three (solid black line) dimensions. The functions are plotted for the case when the length of each side is $L = 1$. For an arbitrary L, the horizontal axis of the plot is scaled by $1/L^2$ whereas the vertical axis is scaled by L^2

In the above discussion, we have taken selected subsystem j and k to be chosen randomly among all the subsystems and have calculated the distribution and the expected value of the exchange coupling constants F_{jk}. For the single-subsystem error Hamiltonian and subsequent discussion, one of the subsystems, subsystem j

is fixed, and the summations are over the remaining subsystems in the computer. For this case, the precise values of the distributions and the expected values will depend on the choice of the subsystem j, for example whether it is chosen to be at the edge of the array or at the center. However using the above formalism, as expected, one can show that these considerations do not change the N scalings of the expected values. Therefore all scaling results here can apply to expectation values in the single qubit Hamiltonian context as well as the full Hamiltonian.

B Pattern Decomposition Example

Here we will go through an example of calculating the moments in detail for $p = 4$ and $N = 5$ in order to make the various definitions and ideas more clear and concrete.

We begin with the expression

$$\sigma^{(4)} = 2^{-N} \sum_{q} \langle q| H^4 |q\rangle = 2^{-N} \sum_{(j_1 k_1)} \sum_{(j_2 k_2)} \sum_{(j_3 k_3)} \sum_{(j_4 k_4)} \sum_{q} \langle q| H_{j_1 k_1} H_{j_2 k_2} H_{j_3 k_3} H_{j_4 k_4} |q\rangle \tag{B.1}$$

which, per the analysis in Chapter 2, was broken into

$$\begin{aligned} \sigma^{(4)} &= 2^{-N} \sum_{S} D(S)^{-1} \sum_{\{j\}} \sum_{\pi} \sum_{q} \langle q| \prod H(S, \{j\}, \pi) |q\rangle \\ &= 2^{-N} \sum_{u=2}^{2p} 2^{N-u} \sum_{S(u)} \sum_{\{j\}(u)} \sum_{\pi(u)} \frac{\Omega'(S(u))}{D(S)} \prod F(S(u), \{j\}(u), \pi(u)) \end{aligned} \tag{B.2}$$

Before analyzing the quantities in the triplet breakdown, first we will use the knowledge of the final patterns to directly find the answer as a sum of sequences $H_{j_1 k_1} ... H_{j_4 k_4}$ in order to compare with the triplet method. Recall that the form of the final patterns is that every H_{jk} is paired up with one other, and no pairs of H_{jk} have any subsystems in common. Due to this property, the ordering of the H factors doesn't matter in the sense that there are the same number of $|q\rangle$ such that $\langle q| H_{12} H_{34} H_{12} H_{34} |q\rangle \neq 0$ as for $H_{12} H_{12} H_{34} H_{34}$. This means we can lump all sequences with the same value of $\prod F$ together and simply multiply by how many there are. Each $\prod F$ will be $F^2_{j_1 k_1} F^2_{j_2 k_2}$ for some four distinct j_1, k_1, j_2, k_3 (but we don't distinguish between j_1 and k_1 or j_2 and k_2, for example $j_1 = 1, k_1 = 5$, and j_2 and k_2 equal to anything other than 1, 5, or each other, is the same term as with $j_1 = 5$ and $k_1 = 1$. For each different value of $\prod F$, i.e. each choice of 4 distinct subsystems and each way to pair those 4 up, there is some number of times that $\sum_{(j_1 k_1)} ... \sum_{(j_4 k_4)} F_{j_1 k_1} ... F_{j_4 k_4}$ equals that product. This number is independent of the chosen subsystems and how they are paired, due to the structure of the final

patterns (this would not be true for all patterns). The number of times this happens for a given $prodF$ is equal to $\binom{4}{2}$ because once the two locations of, say $F_{j_1k_1}$ are chosen out of the 4 possible locations, the locations of the $F_{j_1k_2}$ are fixed. We will denote the sum over pairs of pairs of subsystems chosen from the $N = 5$ total subsystems as $\sum_{((j_1k_1)(j_2k_2))}$. As for the sum over q, this will contribute a total of $N^{N-4} \times 2^2 = 2^{N-2}$ since $\Omega'(S_t) = 2^{p/2} = 2^2$. Thus we have

$$\sigma^{(4)} \to \frac{1}{4}\binom{4}{2} \sum_{((j_1k_1)(j_2k_2))} F^2_{j_1k_1} F^2_{j_2k_2} \tag{B.3}$$

Since neither of j_1 or k_1 is equal to j_2 or k_2, the average of the different $F^2_{j_1k_1}F^2_{j_2k_2}$ is $E[F^2]^2$ (in the large N limit, which we are working in and only using a small N for illustrative purposes). Therefore, since there are $\binom{5}{2}\binom{3}{2}\frac{1}{2} = 15$ different pairs of pairs,

$$\sum_{((j_1k_1)(j_2k_2))} F^2_{j_1k_1} F^2_{j_2k_2} = 15E[F^2]^2 \tag{B.4}$$

and accordingly,

$$\sigma^{(4)} \to \frac{45}{2} E[F^2]^2 \tag{B.5}$$

To demonstrate consistency, we will now arrive at this same result using the triplet framework.

Since every subsystem appearing in $\prod F$ must appear an even number of times in order for there to be any $|q\rangle$ such that $\langle q| \prod F |q\rangle \neq 0$, u is capped at p and we now list all patterns which obey this constraint (labelling the leftmost $H_{j_1k_1}$ one and incrementing by 1 to the right):

$$\begin{array}{llll} \{\{1234\}\{1234\}\} & \{\{1234\}\{12\}\{34\}\} & \{\{1234\}\{13\}\{24\}\} & \{\{1234\}\{14\}\{23\}\} \\ \{\{12\}\{12\}\{34\}\{34\}\} & \{\{13\}\{13\}\{24\}\{24\}\} & \{\{14\}\{14\}\{23\}\{23\}\} & \\ \{\{12\}\{13\}\{24\}\{34\}\} & \{\{12\}\{14\}\{23\}\{34\}\} & \{\{13\}\{14\}\{23\}\{24\}\} & \end{array} \tag{B.6}$$

Letting integers now refer to subsystems 1 through 5, instead of pairwise Hamiltonian

locations 1 through 4 as above, we can enumerate all the terms associated with a given pattern. For example, for the pattern $\{\{12\}\{13\}\{24\}\{34\}\}$, some of the possible pair sequences are

$$\{\{12\}\{13\}\{24\}\{34\}\} \quad \{\{31\}\{34\}\{12\}\{24\}\} \quad \{\{45\}\{42\}\{41\}\{21\}\} \tag{B.7}$$

Since this pattern has no two identical elements, $D(S) = 1$ and there are 4! different π assigning the different elements of S to the 4 different elements of $\{j\}$, which there are 5 ways to construct (each $\{j\}$ has one subsystem missing, so one j for each subsystem). This means in total there are $5 \times 4! = 5! = 120$ different pairwise sequences associated with that one pattern, which is why we did not list them all.

According to the analysis of Chapter 2, only the middle row of the listed patterns contribute meaningfully for large N, which we will call the S_1, S_2, and S_3 in the order they are written (note that this contradicts with previous notations, wherein S_i meant a certain type of pattern in the context of the transformations $\rightarrow_i$). There are three of these patterns, which matches the derived number of final patterns equal to $p!/((p/2)!2^{p/2})$ which is 3 when $p = 4$. These patterns have $u = 4$, and so

$$\sigma^{(4)} \rightarrow 2^{-4} \sum_{i=1}^{3} \frac{\Omega'(S_i)}{D(S_i)} \sum_{\{j\}(4)} \sum_{\pi(4)} \prod F(S_i, \{j\}(4), \pi(4)) \tag{B.8}$$

For these patterns, since each has two pairs of elements of S_i which are identical, $D(S_i) = 2^2$. This matches the derived form $D(S_t) = 2^{p/2}$ since $p = 4$. For example, let us examine the sequence $\{\{12\}\{12\}\{34\}\{34\}$. It can be constructed from any triplet with $S = \{\{12\}\{12\}\{34\}\{34\}\}$ (even though this looks identical to the sequence, remember its interpretation is very different! - in the sequence the numbers refer to subsystems, or $H_{12}H_{12}H_{34}H_{34}$, while in the patter they refer to locations, so the pattern should be read that one subsystem appears in the first and second H_{jk}, another one also appears in those two, one appears at locations 3 and 4, and another also appears at 3 and 4). The triplet must have $\{j\} = \{1, 2, 3, 4\}$ since these are the four subsystems appearing in the sequence or term. There are 4

different subsystem assignments π that are compatible with that S and $\{j\}$, and construct the given sequence: temporarily using A, B, C, D to refer to the four elements of S (in the order listed so that $A, B = \{12\}$ and $C, D = \{34\}$), the four assignments are:

$1 \to A$, $2 \to B$, $3 \to C$, $4 \to D$

$1 \to B$, $2 \to A$, $3 \to C$, $4 \to D$

$1 \to A$, $2 \to B$, $3 \to D$, $4 \to C$

$1 \to B$, $2 \to A$, $3 \to D$, $4 \to C$

Thus there are $4 = 2^2 = D(S)$ triplets in total which construct that term, each with the same pattern S and subsystem selection $\{j\}$, and one of four possible subsystem assignments π enumerated above.

$\Omega'(S)$ is the number of arrangements $|q\rangle$ of the subsystems appearing in a triplet $(S, \{j\}, \pi)$ (i.e. the subsystems in $\{j\}$) constructed from S for which $\langle q| \prod H(S, \{j\}, \pi) |q\rangle \neq 0$ (recall that this number is invariant with respect to $\{j\}$ and π and so is written as a function of S only). For any of the three final patterns, $\Omega'(S) = 2^2 = 4$ which is consistent with the derived value $\Omega'(S_t) = 2^{p/2}$ with $p = 4$: using the notation 1 is up and 0 is down, the four such arrangements of the subsystems $\{1234\}$ are:

$|1010\rangle$

$|1001\rangle$

$|0110\rangle$

$|0101\rangle$

The only restriction is that 1 and 2 are opposite, because they appear in together in an H_{jk}, and similarly for 3 and 4. With an independent choice of 2 arrangements for each pair, and two pairs, we have $2 \times 2 = 4$ total arrangements.

Using these, we now have

$$\sigma^{(4)} \to 2^{-4} \frac{4}{4} \sum_{i=1}^{3} \sum_{\{j\}(4)} \sum_{\pi(4)} \prod F(S_i, \{j\}(4), \pi(4)) = 2^{-4} \sum_{i=1}^{3} A_{S_i} E[\prod F(S_i)] \quad \text{(B.9)}$$

A_S is the number of selections and assignments compatible with the pattern S.

In this case, that is the number of ways to choose 4 subsystems from all 5, times the number of ways to assign those 4 subsystems to the 4 elements of S. Since all S_i have 4 elements, for all of them $A_S = \binom{5}{4}4! = 5! = 120$. This agrees with the derived $A(S_t) = N!/(N-p)! = 5!$.

For all the final patterns S_i, $E[\prod F(S_i)] = E[F^2]^2$. Together, this means

$$\sigma^{(4)} \to \frac{120}{2^4} 3E[F^2]^2 = \frac{45}{2} E[F^2]^2 \tag{B.10}$$

as required.

Note that the final expression $\sigma^{(p)} \to \frac{p!}{(p/2)!}(N(E[F^2]/8)^{1/2})^p$ does not agree with this expression, because the limit $N!/(N-p)! \to N^p$ has been taken for that expression, which is only valid if $p << N$. Since that is not valid here, the formula is incorrect. However, substituting $N!/(N-p)!$ back in for N^p does recreate the correct prefactor:

$$\frac{p!}{(p/2)!}\frac{N^p}{8^{p/2}} \to \frac{p!}{(p/2)!}\frac{N!}{8^{p/2}(N-p)!} = \frac{4!}{2!}\frac{5!}{8^2\,1!} = \frac{45}{2} \tag{B.11}$$

C The Threshold Theorem

Here we will sketch the ideas behind the Threshold Theorem, mainly following the arguments in [27] and [18]. We will call the quantum computer the system, and denote its ideal Hamiltonian as H_S. This Hamiltonian exactly enacts the desired time evolution, presumably a quantum algorithm with error correction. Whatever environment the quantum computer lives in is called the bath and its Hamiltonian is H_B. The total Hilbert space is $\mathcal{H}_S \otimes \mathcal{H}_B$ where $\mathcal{H}_{S(B)}$ is the Hilbert space for the system (bath). $H_{S(B)}$ acts nontrivially only on the system (bath) degrees of freedom. The interaction between the two is described by H_{SB}. H_{SB} is not the fundamental interaction between the qubits and the quantum environment, but rather the effective coarse-grained interaction after integrating out high-frequency and short-distance degrees of freedom.

Without loss of generality, this interaction can be decomposed as

$$H_{SB} = \sum_{a=1}^{\infty} \sum_{(j_1 \ldots j_a)} H^{(a)}_{j_1 \ldots j_a} \tag{C.1}$$

where $H^{(a)}_{j_1 \ldots j_a}$ acts non-trivially on no system degrees of freedom except the qubits $\{j_1 \ldots j_a\}$. The superscript (a) denotes how many different qubits the Hamiltonian $H^{(a)}$ acts non-trivially on, and its subscript specifies which qubits, totalling a in number, it does act on. The sum over a runs up to infinity so that there is a term for the interaction of all qubits together in the limit that there are infinitely many qubits. It is done this way so that the fault-tolerance of the computation can be said to be scalable in the sense that it would in principle work for a computer of any size, an approach motivated by classical computing.

The unitary time evolution operator $U(t)$ generated by H_{SB} is $e^{-itH_{SB}}$ (assuming H_{SB} is time independent, which we do for simplicity and is the case for the boson exchange interaction). Using the property that $U(t_2)U(t_1) = U(t_1 + t_2)$, for any time interval Δ the operator $U(t)$ can be broken in $\prod U(\Delta) = e^{-i\Delta H_{SB}} \ldots e^{-i\Delta H_{SB}}$

where the number of terms in the product is t/Δ (we will constrain Δ such that this ratio is an integer). For a small enough Δ, $U(\Delta)$ is arbitrarily well approximated by $\big(1-i\Delta H_{SB}\big)$ and therefore $U(t)$ by $\big(1-i\Delta H_{SB}\big)...\big(1-i\Delta H_{SB}\big)$. Imagine expanding each H_{SB} as $\sum_{a=1}^{\infty}\sum_{(j_1...j_a)} H^{(a)}_{j_1...j_a}$ and then distributing each sum, so that from each $e^{-i\Delta H_{SB}}$ there is either a 1 or some $-i\Delta H^{(a)}_{j_1...j_a}$. The terms of this expansion would be the product of t/Δ such factors, and there would be one term for each choice of possible factor from each of the t/Δ different $\big(1-i\Delta\sum_{a=1}^{\infty}\sum_{(j_1...j_a)} H^{(a)}_{j_1...j_a}\big)$.

Define a location to be a single gate, such as a single qubit gate, multi-qubit gate, or idle (identity) gate. The computer is assumed to be able to perform gates in parallel, so that at any one time there can be many locations. For a single term in the expansion imagined above, a location is said to be faulty if, for any of the timesteps Δ during its duration t_g, there is a term $-i\Delta H^{(a)}_{j_1...j_a}$ which includes on one or more of the qubits involved in that gate.

Consider a set of r different locations, denoted by $\mathcal{I}_r$. Let $E(\mathcal{I}_r)$ be the sum of all terms in the expansion described above such that each of the locations in $\mathcal{I}_r$ are faulty. Therefore if $\mathcal{I}_0$ is the empty set, then $E(\mathcal{I}_0) = U(t)$ where t is the length of the entire computation, since there are no qubits in any location and therefore all terms are allowed. If we add one location to $\mathcal{I}$, then we would exclude any term such that, for each Δ slice during the added location, none of the qubits involved in that location (gate) are included in any $H^{(a)}_{j_1...j_a}$.

Since terms are excluded as one increases the number of locations in $\mathcal{I}_r$, one would expect in general for the magnitude of $E(\mathcal{I}_r)$ to decrease as r grows. The noise caused by H_{SB} is said to have effective strength ϵ if

$$\big\|E(\mathcal{I}_r)\big\| \leq \epsilon^r \tag{C.2}$$

for all r and any set of r locations. The norm of the operator $E(\mathcal{I}_r)$ is proportional to the probability of each of the locations in r being faulty. In order for the computation to be likely to succeed without fault, this number should be very small for all $\mathcal{I}_r$. While the value of ϵ can be somewhat decreased, for example, by reducing

the coupling of the qubits to their environment (recall that H_{SB} is the effective interaction, not the fundamental one), but this only gets one so far. The strategy will be to implement quantum error correcting codes which create a higher-level realization of the algorithm but with an effectively reduced noise ϵ.

For our purposes, we may consider a quantum error correction code to be the encoding of the state of one qubit onto the state of a collection of a small number of qubits. The encoded qubit is called the logical qubit, and each member of the collection which encodes the state of the logical qubit is called a physical qubit. A desired measurement of the logical qubit can be realized by a different measurement on the physical qubits. However, because there are extra pieces of information (the Hilbert space of the physical qubits is much larger than 2, the size of the Hilbert space for one qubit), that means there are some measurements that can be made that will not reveal any information about the state of the logical qubit and therefore leave it entirely undisturbed. This is the desired property, since those same measurements can diagnose errors that might have occurred so that they can protected. Of course if too many errors occur then this is not possible, but for a single qubit any error is fatal, whereas for the collection of physical qubits some errors can occur without dooming the computation. If a code can correct p errors, i.e. if no more than p errors occur among the physical qubits then the correct state of the logical qubit can be restored, then that code is called distance $2p + 1$. For example, if a code can correct one error, it is a distance 3 code. The details of the code are unimportant, because fault-tolerance is not attained through the use of an extremely high distance code, but rather by concatenating the code many times. For each concatenation, a physical qubit is turned into a logical qubit and replaced by a constellation of physical qubits according to the code. If the code encodes one logical qubit into 7 physical qubits, then after 2 concatenations of the code on a single qubit, there would be $7 \times 7 = 49$ physical qubits encoding the state of a single logical qubit; each of the 7 physical qubits after the first concatentation becomes 7 physical qubits after the second. When this encoding happens a single location of the logical qubits turns into a multitude of locations, since each gate acting on a

single logical qubit is realized as multiple gates acting on multiple physical qubits. A code is said to have depth k if it has been concatenated k times. We will call a collection of locations of this code corresponding to a single location of the original, logical location a k-location.

For concreteness, we will consider a code which can only fix one error - the essential result will not change in a different case. In order for a k-location to be faulty, at least two $k-1$-locations must be faulty. Thus the probability of a faulty 1-location is proportional to the probability of a faulty 0-location squared. Since 0-locations are governed directly by H_{SB}, this error rate is captured by ϵ. Thus, if (C.2) is satisfied for 0-locations, then (C.2) is also satisfied for 1-locations for some new noise $\epsilon^{(1)}$, related to the old noise ϵ by

$$\epsilon^{(1)} = \epsilon^2/\epsilon_0 \quad \text{or} \quad \epsilon^{(1)} = \epsilon_0(\epsilon/\epsilon_0)^2 \tag{C.3}$$

where ϵ_0 is some constant with units of ϵ. Using the latter expression above, by induction the effective noise strength for the k-th level is

$$\epsilon^{(k)} = \epsilon_0(\epsilon/\epsilon_0)^{2^k} \tag{C.4}$$

As long as $\epsilon < \epsilon_0$, this quantity quickly goes to 0 for large k. Thus, after performing k concatenations of the code, the effective noise strength can be made arbitrarily small, so long as the original noise ϵ is less than some threshold ϵ_0.

In this manuscript, we challenge the notion that the effective noise ϵ is finite at all, much less below some threshold ϵ_0. For $r = 1$, let us examine $\left\|E(\mathcal{I}_1)\right\|$ which is a lower bound on ϵ. The strategy employed to bound $\left\|E(\mathcal{I}_r)\right\|$ is to consider a subset of the terms which contribute to it which have a fixed time step Δ at which the first term from H_{SB} acts on a qubit in that location, and a fixed term $H^{(a)}_{j_1\ldots j_a}$ which acts at that time step and contains at least one qubit in the location. The key advantage to such a subset is that, for all time steps Δ after the first $H^{(a)}_{j_1\ldots j_a}$ acts on a qubit in the location, in all future time steps all terms in H_{SB} are allowed and therefore included. When all of them are included, the time evolution is unitary and therefore

has unit norm. Before this fixed time step, but still during the location, there is also unitary evolution according to a modified H_{SB} which has removed from it all terms which act nontrivially on any qubits in the location. Before the location, all terms are allowed and therefore the time evolution is unitary.

Thus the sum of all terms contributing to $E(\mathcal{I}_1)$ with a fixed timestep and term during the location in $\mathcal{I}_1$ which first causes a fault at that location is a unitary operator times $-i\Delta H_I$ times another unitary operator, where H_I is the fixed term in H_{SB} which affects a qubit in the location. According to the property that $\|AB\| \leq \|A\|\|B\|$ and the fact that unitary operators have norm 1, this contribution to $E(\mathcal{I}_1)$ is bounded above by $\Delta\|H_I\|$. Since all constituents of $E(\mathcal{I}_1)$ have a specific earliest timestep during the location in $\mathcal{I}_1$ at which a qubit in that location is hit, and a specific term in H_{SB} which hits a qubit in that location in that timestep, all constituents of $E(\mathcal{I}_1)$ belong to exactly one subset with a fixed timestep and term described above. In other words, $E(\mathcal{I}_1)$ is the sum of all such subsets. Using the property that $\|A+B\| \leq \|A\| + \|B\|$, then, $E(\mathcal{I}_1) \leq \sum_I \sum_{\Delta \text{ in } t_g} \Delta\|H_I\| = t_g \sum_I \|H_I\|$ where the sum over I is the sum over a from 1 to ∞ and the sum over all choices of a qubits which include at least one qubit involved in the location or gate in $\mathcal{I}_1$. For the case of a single-qubit gate and the effective boson interaction, it was showed in Chapter 1 that this sum does not converge, rendering this proof of the Threshold Theorem, in its current form, inapplicable to that source of noise.

This is clearly not a definitive result; just because a certain upper bound on $\left\|E(I_a)\right\|$ does not converge does not mean that $\left\|E(\mathcal{I}_a)\right\|$ diverges. In particular, as discussed in Chapter 1, the upper bounds used above neglect the phases between the different H_I and therefore is a drastic overestimate in the large N limit where the phases are random (assuming finite qubit spatial density). In Chapter 2 we go beyond these crude upper bounds in order to try to determine if a more conservative derivation of the Threshold Theorem could possibly include that noise. The results there imply that this is not possible. The theorem could be altered to not be proven in the $N \to \infty$ limit, but this would be a paradigm shift and create a rift between

classical and quantum computing, where the analogous error correction results can be proven for $N \to \infty$.

D Source Code for the Numerical Simulations

D1 Definite Model

In this section is presented the source code for the Definite Model described in Chapter 4. It is written in python, and outputs a graph of the average number total number of excitations versus time for a sample with the given parameters. It is also writes the average number of excitations at each time step, as well as a list of those time steps, to a file.

```
# import packages that will be needed
import numpy as np
import matplotlib.pyplot as plt
import copy

# included is the possibility to truncate the interactions at
# first order, but for the desired parameters this option
# is not used because it is not a good approximation
first_order = False

tau = .000000027  # 27 ns: the lifetime (in seconds) of
# the relevant transition
wavelength = .00000078  # 780 nm: the wavelength, in
# meters, of the relevant transition
k = 2*np.pi/wavelength  # calculating the wavenumber of
# the transition

N = 160000  # the total number of subsystems
coop_amp = .00035  # the  e-squared radius of the gaussian
```

```
20 # sample, in meters
21
22 num_samples = 1  # how many times to run the simulation; at
23 # the end, all trials will be averaged over to produce one graph
24
25 initial_excite_prob = .5  # choose what fraction of the
26 # subsystems begin excited; note that the initially excited
27 # fraction is not fixed to be exactly this, rather is
28 # the probability for each to be excited
29
30 # num_steps = 760
31 # total_time = 6 * tau
32
33 num_steps = 30  # total number of time steps
34 total_time = .3 * tau  # total duration of simulation
35
36 tstep = total_time/num_steps  # calculate the size of each time
37 # step from the specified parameters
38
39
40 # choose two random positions in the gaussian cloud (using
     spherical
41 # coordinates), calculate their separation and return the overall
42 # coupling factor given their separation which is equal to 1/5kr (
     with
43 # separation equal to r and k being the wavenumber of the
     transition)
44 def random_distanced_coupling():
45     r1 = np.random.normal(0, coop_amp/2, 1) * k
46     r2 = np.random.normal(0, coop_amp/2, 1) * k
47     th1 = 0
48     th2 = np.random.rand()*2*np.pi
49     phi1 = 0
50     phi2 = np.random.rand()*np.pi
51     x1 = r1 * np.sin(th1) * np.sin(phi1)
52     x2 = r2 * np.sin(th2) * np.sin(phi2)
53     y1 = r1 * np.cos(th1) * np.sin(phi1)
```

```
54      y2 = r2 * np.cos(th2) * np.sin(phi2)
55      z1 = r1 * np.cos(phi1)
56      z2 = r2 * np.cos(phi2)
57      random_distance = np.sqrt((x1-x2)**2+(y1-y2)**2+(z1-z2)**2)
58      return (1/(5*random_distance))**2
59
60
61  # make a class for a subsystem which has two possible states, up
62  # and down when one is created, it is randomly
63  # assigned the up or down state
64  class Emitter:
65      def __init__(self, lifetime):
66          self.lifetime = lifetime
67          if np.random.rand() < initial_excite_prob:
68              self.state = 1
69          else:
70              self.state = 0
71
72
73  # a class for a collection of Emitters
74  class Ensemble:
75      def __init__(self, amp=0):
76          self.emitters = []
77          self.m_tracker = []
78          self.evolution_step = 0
79          self.amp = amp
80          self.N = len(self.emitters)
81          self.times = []
82
83      # method to add one emitter to the ensemble
84      def add_emitter(self, emitter):
85          self.emitters.append(emitter)
86          self.N += 1
87
88      # method for the emission process in one timestep for one
        emitter
89      def single_emit_initial(self, period, emitter_index):
```

```
90          if self.emitters[emitter_index].state == 1:
91              lifetime = self.emitters[emitter_index].lifetime
92              if period / lifetime > 1:
93                  print('nonlinearity has become a problem')
94              if np.random.rand() < period / lifetime:
95                  self.emitters[emitter_index].state = 0
96                  photon_count = 1
97                  j = 0
98                  outcome = 0
99                  stim_prob = self.N * self.amp
100                 if first_order:
101                     if stim_prob > .7:
102                         print('stim prob too high for first order
    ')
103                     if np.random.rand() < stim_prob:
104                         emitted_index = np.random.randint(self.N)
105                         if self.emitters[emitted_index].state ==
    0:
106                             self.emitters[emitted_index].state = 1
107                         else:
108                             self.emitters[emitted_index].state = 0
109                 else:
110                     while (photon_count != 0)and(j < self.N):
111                         if j != emitter_index:
112                             if self.emitters[j].state == 1:
113                                 outcome = self.
    single_emit_stimulated(period, j, self.amp)
114                             else:
115                                 outcome = self.single_absorb(
    period, j, self.amp)
116                         photon_count += outcome
117                         j += 1
118
119     # a method used after each emission on every other emitter,
    allowing
120     # the possibility of stimulated emission
121     def single_emit_stimulated(self, period, emitter_index, amp):
```

```
122         if np.random.rand() < random_distanced_coupling():
123             self.emitters[emitter_index].state = 0
124             return 1
125         return 0
126 
127     # same as above, but for absorption
128     def single_absorb(self, period, emitter_index, amp):
129         if np.random.rand() < random_distanced_coupling():
130             self.emitters[emitter_index].state = 1
131             return -1
132         return 0
133 
134     # implement one time step of the evolution
135     def time_step(self, period):
136         for i in range(0, self.N):
137             if self.emitters[i].state == 1:
138                 self.single_emit_initial(period, i)
139 
140     # returns the total number of subsystems in the up or 'on'
    state
141     def total_on(self):
142         result = 0
143         for emitter in self.emitters:
144             if emitter.state == 1:
145                 result += 1
146         return result
147 
148     # calculate the total time evolution of the ensemble, and
    write the results to a file
149     def evolve(self, period, num_steps, file):
150         t_count = 0
151         while t_count < num_steps:
152             self.m_tracker.append(self.total_on())
153             self.time_step(period)
154             m = self.total_on()
155             m_str = f"{m}"
156             file.write(m_str)
```

```
157             file.write(", ")
158             data_list.append(m_str)
159             # print(str)
160             t_count += 1
161 
162     # creates the array with each time point and initiate the
    evolution
163     def evolve_m(self, period, num_steps, file):
164         self.times = [step * period for step in range(0, num_steps
    )]
165         self.evolve(period, num_steps, file)
166 
167 
168 # a class for averaging over many ensembles
169 class EnsembleAverage:
170     def __init__(self, ensemble, file):
171         self.initial_ensemble = ensemble
172         self.file = file
173 
174     # prepares a graph of the number of excitations versus time,
    averaging over
175     # many ensembles
176     def graph_m(self, period, num_steps, sample_size):
177         for i in range(0, sample_size):
178             print('progress:', (i+1)/sample_size)
179             ensemble_instance = copy.deepcopy(self.
    initial_ensemble)
180             ensemble_instance.evolve_m(period, num_steps, self.
    file)
181             if i == 0:
182                 times = ensemble_instance.times
183                 avg_m = [0 for t in times]
184             avg_m = [avg_m[j] + ensemble_instance.m_tracker[j] for
     j in range(0, len(times))]
185         print('samples completed: averaging and plotting
    completing or completed')
186 
```

```
187
188 # a function for averaging over a list of all data from many
      trials
189 def list_average(long_list):
190     avgd_list = []
191     segment_num = num_samples
192     segment_length = num_steps
193     for i in range(0, segment_length):
194         elem = 0
195         for j in range(0, segment_num):
196             elem += float(long_list[i+segment_length*j])
197         avgd_list.append(elem/segment_num)
198     return avgd_list
199
200
201 # prepares files to be written to so that the outcomes are saved
202 data_file = open("out_emitters.txt", "w")
203 data_list = []
204 times_file = open("t_emitters.txt", "w")
205 times_list = []
206
207 # make instances of the above classes
208 cloud = Ensemble(coop_amp)
209 for emitter_count in range(0, N):
210     cloud.add_emitter(Emitter(tau))
211
212 D = EnsembleAverage(cloud, data_file)
213
214 # check to make sure the timestep is small enough
215 if tstep/tau > .05:
216     print('tstep is too high')
217
218 # perform the evolution and graph preparation on ensemble D
219 D.graph_m(tstep, int(total_time/tstep), num_samples)
220
221 # prepare the time axis for graphing
222 times_table = np.arange(0, total_time, tstep)
```

```
223 times_str = f"{times_table}"
224 times_file.write(times_str)
225
226 # perform the average over all instances
227 avg_pop = list_average(data_list)
228
229 # open the averaged file for writing, then write to it
230 data_file_avg = open("avg_out_emitters.txt", "w")
231 for i in avg_pop:
232     wstring = f"{i}"
233     data_file_avg.write(wstring)
234     data_file_avg.write(", ")
235
236 # make the plot for displaying purposes, along with a plain
      exponential for comparison
237 times = np.arange(0, total_time-.5*(total_time/num_steps),
      total_time/num_steps)
238 plt.plot(times, np.log(avg_pop), times, np.log(N*
      initial_excite_prob*np.exp(-((times+tstep/2)/tau))))
239
240
241
242 # close the files that were being written to
243 data_file.close()
244 data_file_avg.close()
245 times_file.close()
246
247 # show the population versus time plot
248 plt.show()
```

D2 Approximate Decoherence Simulations

In Chapter 4, the decoherence of a single subsystem was compared between the exact evolution of the ensemble of two level systems and an approximate method which uses many of the assumptions necessary for the scaling arguments of Chapter

2 to hold. The source code for this approximate method is contained here. First, for the given parameter values, the width of the gaussian energy spectrum is to be calculated. That width will determine the gaussian from which eigenvalues are drawn to approximate the decoherence of the single subsystem due to decoherence.

Although only the second moment is needed in the large N limit, it is useful to also calculate higher order moments in order to see how well the large N limit applies. For example, one can compare what the fourth moment actually is to what it should be according to the second moment in the large N limit, where the spectrum becomes perfectly gaussian and so the fourth moment can be deduced simply from the second. Such calculations have shown that, at least in this sense, the large N limit is not a bad approximation for the parameter values used.

Below is the code used to calculate all moments, although it becomes very slow past the sixth order. It is written for python, and outputs the desired moment averaged over many specific instances of subystem locations, as well as the standard deviation and standard deviation of the mean, of that average.

```
# import the necessary packages
import numpy as np
import itertools as it
import copy
import numpy.linalg as lin
from sympy.utilities.iterables import multiset_permutations

# define parameters for the calculation
N = 5  # number of subsystems
moment = 2  # which moment to calculate
kL = 1098052  # unitless size of the sample - make >>100 to
    randomize interaction phases
F = 1 * kL  # set the coupling constant in the same units of
    sample size
dim = 2  # spatial dimensions of the subsystem positions
duplicate_check = False
exact_positions = True
random_phases = False
```

```
17 verbose = False
18 show_coupling = True
19 show_it_all = False
20 samples = 594940
21 shell = .2
22
23
24 total_shell_count = 0
25
26 result = []
27 secondmom_list = []
28 thirdmom_list = []
29 fourthmom_list = []
30 fourthmom_listapprox = []
31 fourth_correction_list = []
32
33
34 # makes all lists of numbers which add up to the target
35 def subset_sum(numbers, target, partial=[], sub_result=[]):
36     s = sum(partial)
37     if s == target:
38         sub_result.append(partial)
39     if s >= target:
40         return
41     for i in range(len(numbers)):
42         n = numbers[i]
43         remaining = numbers[i:]
44         subset_sum(remaining, target, partial + [n], sub_result)
45     return sub_result
46
47
48 # takes all combinations of 2 qubits in available qubits, along
      with partially constructed patterns, and
49 # and coincides them in all possible ways into new patterns
50 def assign(partial_pattern, qubits_for_assignment):
51     constructed_pattern = []
52     qubit_list = [qubit[0] for qubit in qubits_for_assignment]
```

```
53      for pair in it.combinations(qubit_list, 2):
54          new_partial_pattern = copy.deepcopy(partial_pattern)
55          new_partial_pattern.append(pair)
56          new_available_qubits = copy.deepcopy(qubits_for_assignment
    )
57          to_remove = []
58          for i in range(0, len(new_available_qubits)):
59              if new_available_qubits[i][0] in pair:
60                  if new_available_qubits[i][1] == 1:
61                      to_remove.append(i)
62                  else:
63                      new_available_qubits[i][1] += -1
64          shift_count = 0
65          for j in to_remove:
66              new_available_qubits.pop(j - shift_count)
67              shift_count += 1
68          constructed_pattern.append([new_partial_pattern,
    new_available_qubits])
69      return constructed_pattern
70
71
72 # get rid of extra formatting in pattern lists
73 def flatten(unflat_patterns):
74      flatpat, empty = zip(*unflat_patterns)
75      if len(flatpat) == 1:
76          return flatpat[0]
77      else:
78          return flatpat
79
80
81 # returns all qubits that appear in a given pattern
82 def get_qubits(pattern):
83      flat_qubit_index_list = [item for sublist in list(pattern) for
     item in sublist]
84      return set(flat_qubit_index_list)
85
86
```

```
87 #
88 def check_arrangement(arr, pattern):
89     contemp_arr = list(copy.deepcopy(arr))
90     for pair in pattern:
91         if contemp_arr[pair[0]] == contemp_arr[pair[1]]:
92             return 0
93         else:
94             contemp_arr[pair[0]] = (contemp_arr[pair[0]] + 1) % 2
95             contemp_arr[pair[1]] = (contemp_arr[pair[1]] + 1) % 2
96     if contemp_arr == list(arr):
97         return 1
98     else:
99         return 0
100
101
102 # calculate how many times a given pattern will contribute to the
      given moment
103 def pattern_mult(pattern):
104     multiplicity = 0
105     unique_qubit_num = len(get_qubits(pattern))
106     if show_it_all:
107         print('finding multiplicity for', pattern)
108     for num_excited in range(1, int(np.ceil(unique_qubit_num/2))):
109         for start_arr in multiset_permutations(list(np.ones(
      num_excited)) + list(np.zeros(N-num_excited))):
110             if check_arrangement(start_arr, pattern):
111                 multiplicity += 2
112     if unique_qubit_num % 2 == 0:
113         for later_arrs in multiset_permutations(list(np.ones(int(
      unique_qubit_num/2))) + list(np.zeros(int(unique_qubit_num/2 -
      1)))):
114             start_arr = [0.0] + later_arrs
115             if check_arrangement(start_arr, pattern):
116                 multiplicity += 2
117                 if show_it_all:
118                     print('found one!', start_arr)
119     multiplicity = multiplicity * np.power(2, N - unique_qubit_num
```

```
    )
120     if show_it_all:
121         print('total multiplicity with pattern:', multiplicity,
    pattern)
122     return multiplicity
123
124
125 # gets multiplicities for all patterns
126 def get_mults(flattened_patterns):
127     if len(flattened_patterns[0]) == 2:
128         flattened_patterns = [flattened_patterns]
129     multiplicities_list = []
130     for pattern in flattened_patterns:
131         multiplicities_list.append(pattern_mult(pattern))
132     return multiplicities_list
133
134
135 # removes patterns with zero multiplicity, then combine the
    remaining
136 # patterns with their nonzeor multiplicities
137 def combine_pat_mult(plain_patterns, plain_multiplicities, zeros=
    False):
138     mults_w_pat = []
139     if len(plain_patterns[0]) == 2:
140         plain_patterns = [plain_patterns]
141     for i in range(0, len(plain_patterns)):
142         this_mult = plain_multiplicities[i]
143         if zeros or (this_mult != 0):
144             mults_w_pat.append([this_mult, plain_patterns[i]])
145     return mults_w_pat
146
147
148 # returns the separation difference between two positions
149 def get_separation(pos1, pos2):
150     difference_vec = [pos1[x] - pos2[x] for x in range(0, len(pos1
    ))]
151     return lin.norm(difference_vec)
```

```
152
153
154 # returns the coupling given the sample size and the normalized
      distance
155 def coupling_term(distance, kL):
156     factor = distance * kL
157     if random_phases:
158         term = F * np.sin(np.random.rand()*2*np.pi)/factor
159     else:
160         term = F * np.sin(factor)/factor
161     return term
162
163
164 # find the total contribution to the moment from a given pattern
      for a given set of qubit positions
165 def pattern_contribution(pat, poslist):
166     global shell_count
167     contribution = 0
168     qubits = get_qubits(pat)
169     num_qubits = len(qubits)
170     all_qubits = [i for i in range(0, N)]
171     for qubit_choices in it.combinations(all_qubits, num_qubits):
172         term = 1
173         for pair in pat:
174             pseparation = get_separation(poslist[qubit_choices[
      pair[0]]], poslist[qubit_choices[pair[1]]])
175             if separation < shell:
176                 shell_count += 1
177                 print('bumped into shell!')
178             term = term * coupling_term(pseparation, kL)
179         contribution += term
180     return contribution
181
182
183 # calculates the moment for a given list of qubit positions
184 def get_moment(patterns_with_mults, poslist):
185     numerical_moment = 0
```

```
186     for p in patterns_with_mults:
187         numerical_moment += p[0] * pattern_contribution(p[1],
    poslist)
188     return numerical_moment/np.power(2, N)
189
190
191 # calculates the moment many times and prints the progress
192 print_counter = 0
193 for sample in range(0, samples):
194     # print out progress
195     if print_counter == 100:
196         print('fraction complete:', sample/samples)
197         print_counter = 0
198     else:
199         print_counter += 1
200
201     # for counting how many times the r shell is needed
202     shell_count = 0
203
204     # generate patterns of qubit coincidence in even numbers
205     even_list = np.arange(2, moment + 1, 2)
206     pattern_pairs = subset_sum(even_list, 2 * moment, [], [])
207
208     # make initial list of all qubits with the number of each for
    each coincidence pattern from above
209     available_qubits_list = []
210     for r in pattern_pairs:
211         available_qubits = []
212         for j in range(0, len(r)):
213             available_qubits.append([j, r[j]])
214         available_qubits_list.append([[], available_qubits])
215
216     # generate all patterns and multiplicities
217     all_assigned = False
218     while not all_assigned:
219         assigned_indicator = 0
220         for elem in available_qubits_list:
```

```
221             if elem[1] != []:
222                 available_qubits_list.remove(elem)
223                 available_qubits_list = available_qubits_list +
    assign(elem[0], elem[1])
224                 assigned_indicator += 1
225         if assigned_indicator == 0:
226             all_assigned = True
227 
228     patterns = flatten(available_qubits_list)
229     multiplicities = get_mults(patterns)
230     combined = combine_pat_mult(patterns, multiplicities)
231 
232     # make the positions list
233     positions, separations, coupling_list = [], [], []
234     while len(positions) < N:
235         next_position = np.random.rand(dim)
236         separation_pass = True
237         for j in range(0, len(positions)):
238             if get_separation(positions[j], next_position) < shell
    :
239                 separation_pass = False
240         if separation_pass:
241             positions.append(next_position)
242             for j in range(0, len(positions)-1):
243                 separation = get_separation(next_position,
    positions[j])
244                 separations.append(separation)
245                 coupling_list.append(coupling_term(separation, kL)
    )
246 
247     # either print the result, if there's only 1 sample, or add
    the result of each sample to a list
248     if samples == 1:
249         print(f'moment {moment}:', get_moment(combined, positions)
    )
250         print(f'parameters: N= {N}, kL={kL}, F={F}')
251     else:
```

```
252         mom = get_moment(combined, positions)
253         result.append([shell_count, mom])
254 
255 # if there is more than 1 sample, compile and print the final
    results
256 if samples != 1:
257     shell_counts, moment_instances = zip(*result)
258     sigma = np.std(moment_instances)
259     print('average:', np.average(moment_instances))
260     print('std, std of the mean:', sigma, sigma/np.sqrt(samples))
```

Next we will use the width, calculated from the code above, to approximate the decoherence time.

```
1 # import the necessary packages
2 import numpy as np
3 
4 
5 width = np.power(13.29, 1/2)  # set the width of the gaussian
    distribution
6 sample_num = 450  # number of times to sample from the
    distribution each time
7 sample_size = 500  # number of times to run the simulation
8 step_size = .001  # the step size for time evolution
9 
10 
11 # a function to evolve the off diagonal matrix element according
12 # to either real or imaginary eigenvalues
13 def offdagrho_evolution(samples, t, real):
14     if real:
15         exp_factor = 1
16     else:
17         exp_factor = 0 + 1j
18     result = 0
19     for s in samples:
20         result += np.exp(-exp_factor * s * t)
21     return np.abs(result)/len(samples)
22 
```

```
23
24 # a function to identify when the off diagonal matrix element
       reaches the threshold
25 def find_t1(samples, step_size, real):
26     result = 0
27     t = step_size
28     while result == 0:
29         if offdagrho_evolution(samples, t, real) < 1/np.e:
30             result = t
31         else:
32             t += step_size
33     return result
34
35
36 # find t1 many times and average over them
37 t1_avg = 0
38 for x in range(0, sample_size):
39     samples = np.random.normal(0, width, sample_num)
40     # print(find_t1(samples, step_size, False))
41     t1_avg += find_t1(samples, step_size, False)
42
43 # display the result
44 print('average:', t1_avg/sample_size)
```

E Derivation and Discussion of the Exchange Hamiltonian

Since the seminal paper by Dicke [35, 41], the problem of an ensemble of two level systems coupled to a common boson bath, often in the context of the specific phenomenon of super- or sub-radiance, has been analyzed by a large number of authors and this problem continues to be relevant for a wide range of physical systems [61–70].

Throughout this section we will take all the qubits in the computer to be in the "causality cone", that is sufficient time evolution is allowed so that each qubit in the computer can be causally influenced by every other qubit. For concreteness we will focus on square and cube geometries with a regular spacing of d between adjacent qubits. However, the results are insensitive to the precise shape of the geometry and also to the regular nature of the array (i.e., the qubits can be taken as randomly distributed within the considered region). We denote each individual qubit with the index j and consider a continuum of bosonic modes with annihilation and creation operators $\hat{a}_{\kappa\epsilon}$ and $\hat{a}^\dagger_{\kappa\epsilon}$ respectively. These operators act on the mode of the field with wave-vector κ and polarization ϵ. The total Hamiltonian for the system when only the energy conserving terms are retained (under the rotating wave approximation) is:

$$\begin{aligned}\hat{H}_{total} &= \sum_j \frac{1}{2}\hbar\omega_a\hat{\sigma}^z_j + \sum_{\kappa\epsilon}\hbar\nu_{\kappa\epsilon}\left(\hat{a}^\dagger_{\kappa\epsilon}\hat{a}_{\kappa\epsilon} + \frac{1}{2}\right) \\ &- \sum_j\sum_{\kappa\epsilon}\hbar g_{\kappa\epsilon}\left[\hat{a}_{k\epsilon}\exp\left(i\vec{\kappa}\cdot\vec{r}_j\right)\hat{\sigma}^+_j + \hat{a}^\dagger_{\kappa\epsilon}\exp\left(-i\vec{\kappa}\cdot\vec{r}_j\right)\hat{\sigma}^-_j\right] \quad , \quad \text{(E.1)}\end{aligned}$$

where

$$\begin{aligned}\hat{\sigma}^z_j &= |1\rangle_{j\ j}\langle 1| - |0\rangle_{j\ j}\langle 0| \quad , \\ \hat{\sigma}^+_j &= |1\rangle_{j\ j}\langle 0| \quad , \\ \hat{\sigma}^-_j &= |0\rangle_{j\ j}\langle 1| \quad . \qquad \text{(E.2)}\end{aligned}$$

In Eq. (E.1), the first two terms describe the qubit array and the bosonic modes in the absence of any interaction whereas the third term describes the coupling between the two systems. $\vec{r}_j$ is the position of the j'th atom and the energies of the qubit states $|0\rangle$ and $|1\rangle$ are taken to be $-\frac{1}{2}\hbar\omega_a$ and $\frac{1}{2}\hbar\omega_a$, respectively. The Dicke limit of the above equations is obtained when the total size of the sample is assumed to be small compared to the κ-vector of the relevant modes, i.e., $\vec{\kappa}\cdot\vec{r}_j \to 0$. It is now well-understood that the key physical effect that describes many different aspects of correlated decay and superradiance is the exchange interaction. Starting

with the Hamiltonian of Eq. (E.1), this interaction has been derived using a variety of approaches by a number of authors [52, 53, 71, 72]. This derivation follows steps that closely mimic the Wigner-Weiskoppf theory of spontaneous decay [73]. We take the the initial state of the qubit system to be an arbitrary (in general entangled) superposition state and assume that we start with zero photons in each field mode $\kappa\epsilon$. The initial state of the combined atom-radiation field system can be written as:

$$|\psi(t=0)\rangle = \sum_{q=0}^{2^N-1} c_{q,0}|q\rangle \otimes |0\rangle \quad . \tag{E.3}$$

Here, the index q runs through all possible 2^N combinations for the qubits and c_q are the expansion coefficients. We define the following parameter for each atomic state $|q\rangle$:

$$2M_q \equiv \#\text{ of atoms in state } |1\rangle - \#\text{ of atoms in state } |0\rangle \quad . \tag{E.4}$$

With this definition, the energy of the atomic state $|q\rangle$ is $M_q\hbar\omega_a$. Working in the interaction picture, we expand the wavefunction as:

$$\begin{aligned} |\psi(t)\rangle = &\sum_{q=0}^{2^N-1} c_{q,0}(t) \exp\left[-i(M_q\omega_a)t\right] |q\rangle \otimes |0\rangle \\ &+ \sum_{\kappa\epsilon} \sum_{q'=0}^{2^N-1} c_{q',1_{\kappa\epsilon}}(t) \exp\left[-i(M_{q'}\omega_a + \nu_{\kappa\epsilon})t\right] |q'\rangle \otimes 1_{\kappa\epsilon}\rangle \quad . \end{aligned} \tag{E.5}$$

Here, $|1_{\kappa\epsilon}\rangle$ represents the state of the radiation field in which the field mode $\kappa\epsilon$ has one photon while all the other modes are in vacuum state and the quantity $\nu_{\kappa\epsilon}$ is the frequency of this mode. With these definitions, the Schrödinger equation

$$i\hbar\frac{d|\psi(t)\rangle}{dt} = \hat{H}_{total}|\psi(t)\rangle \quad , \tag{E.6}$$

yields the following continuum of coupled equations:

$$\begin{aligned}\frac{dc_{q,0}}{dt} &= i\sum_{\kappa\epsilon} g_{\kappa\epsilon}\sum_j \exp\left(i\vec{\kappa}\cdot\vec{r}_j\right)c_{q\ominus j,1_{\kappa\epsilon}}(t)\exp\left[-i(\nu_{\kappa\epsilon}-\omega_a)t\right] \quad , \\ \frac{dc_{q\ominus j,1_{k\epsilon}}}{dt} &= ig_{\kappa\epsilon}\sum_k \exp\left(-i\vec{\kappa}\cdot\vec{r}_k\right)c_{q\ominus j\oplus k,0}(t)\exp\left[-i(\omega_a-\nu_{\kappa\epsilon})t\right] \quad . \end{aligned} \tag{E.7}$$

Here we use the notation $|q\ominus j\rangle$ to mean the configuration that equals $|q\rangle$ at all qubits except the qubit j which has undergone $|1\rangle^j \to |0\rangle^j$ transition. The quantity $\vec{r}_j$ is the position vector of the qubit j. Similarly, the configuration $|q\ominus j\oplus k\rangle$ differs from $|q\ominus j\rangle$ by raising qubit k (at the position $\vec{r}_k$) from $|0\rangle^k \to |1\rangle^k$. The coupled equations of above are intuitive. Each specific configuration $|q\rangle$ is coupled to configurations where a qubit is lowered (configurations $|q\ominus j\rangle$) by emitting a photon into the bath. Similarly, each configuration $|q\ominus j\rangle$ is coupled to configurations where a qubit is raised (configurations $|q\ominus j\oplus k\rangle$) by absorbing a photon from the bath. We then formally integrate the equations that contain bath excitations $c_{q\ominus j,1_{\kappa\epsilon}}$

$$c_{q\ominus j,1_{\kappa\epsilon}}(t) = ig_{\kappa\epsilon}\sum_k \exp\left(-i\vec{\kappa}\cdot\vec{r}_k\right)\int_0^t \exp\left[-i(\omega_a-\nu_{\kappa\epsilon})t'\right]c_{q\ominus j\oplus k,0}(t')dt' \tag{E.8}$$

Here, it is assumed that the coupling to the radiation modes is turned on at $t=0$. It is also assumed that the initial conditions for the continuum mode amplitudes are $c_{q\ominus j,1_{\kappa\epsilon}}(t=0)=0$. Using Eq. (E.8), the differential equation for $c_{q,0}$ in Eq. (E.7) now reads

$$\frac{dc_{q,0}}{dt} = -\sum_{\kappa\epsilon}|g_{\kappa\epsilon}|^2\sum_j\sum_k \exp\left(i\vec{\kappa}\cdot\vec{r}_{jk}\right)\int_0^t \exp\left[-i(\nu_{\kappa\epsilon}-\omega_a)(t-t')\right]c_{q\ominus j\oplus k,0}(t')dt' \tag{E.9}$$

where we have defined $\vec{r}_{jk}\equiv\vec{r}_j-\vec{r}_k$. This way, the problem is reduced to a set of coupled integro-differential equations for the initial probability amplitudes $c_{q,0}$. Each state $|q\rangle$ is coupled to all states that differ by lowering one qubit and raising another qubit in the configuration (i.e., to states $|q\ominus j\oplus k\rangle$). We next rewrite the

summation as an integral since we are considering a continuum of radiation modes:

$$\sum_{\kappa\epsilon} |g_{\kappa\epsilon}|^2 \to \frac{V}{(2\pi)^3} \int_{\kappa\epsilon} |g_{\kappa\epsilon}|^2 d^3\kappa \quad . \tag{E.10}$$

Here, $\kappa = |\vec{\kappa}|$ and V is the quantization volume which is assumed to be much larger than the qubit ensemble. We proceed in spherical coordinates and replace $d^3\kappa$ integral with $d^3\kappa = \kappa^2 \sin\theta d\kappa d\theta d\phi$. The coupling constants $g_{\kappa\epsilon}$ depend on the matrix element between the two levels and also the angle between the polarization of a particular electromagnetic mode and the orientation of the atomic dipole:

$$|g_{\kappa\epsilon}|^2 = \frac{\nu_{\kappa\epsilon}\mu^2}{2\epsilon_0\hbar V}(\vec{\epsilon}\cdot\vec{\epsilon}_a)^2 \quad . \tag{E.11}$$

Here, μ is the coupling matrix element between the two levels (electric-dipole or magnetic-dipole), $\vec{\epsilon}$ is the polarization vector of the mode with frequency $\nu_{k\epsilon}$, and $\vec{\epsilon}_a$ is the orientation vector of the atomic dipole. With these definitions, Eq. (E.9) reads:

$$\frac{dc_{q,0}}{dt} = -\frac{\mu^2}{2(2\pi)^3\epsilon_0\hbar c^3} \sum_j \sum_k \int_0^\infty \nu_\kappa^3 \left[\int_0^\pi \int_0^{2\pi} \sin\theta(\vec{\epsilon}\cdot\vec{\epsilon}_a)^2 \exp\left(i\vec{\kappa}\cdot\vec{r}_{jk}\right) d\theta d\phi \right] \times \left[\int_0^t \exp\left[-i(\nu_\kappa - \omega_a)(t-t')\right] c_{q\ominus j\oplus k,0}(t') dt' \right] d\nu_\kappa \tag{E.12}$$

We have used the identity $\kappa = \nu_\kappa/c$ to convert the outermost integral from $d\kappa$ to $d\nu_{\kappa\epsilon}$. The middle angular integral can be evaluated analytically:

$$\int_0^\pi \int_0^{2\pi} \sin\theta(\vec{\epsilon}\cdot\vec{\epsilon}_a)^2 \exp\left(i\vec{\kappa}\cdot\vec{r}_{jk}\right) d\theta d\phi = 4\pi(1-\cos^2\theta_{jk})\frac{\sin(\kappa r_{jk})}{\kappa r_{jk}} + 4\pi(1-3\cos^2\theta_{jk}) \left[\frac{\cos(\kappa r_{jk})}{(\kappa r_{jk})^2} - \frac{\sin(\kappa r_{jk})}{(\kappa r_{jk})^3} \right] \quad . \tag{E.13}$$

Here $r_{jk} = |\vec{r}_{jk}|$ and the angle θ_{jk} is the angle between the atomic dipole moment

vector $\vec{\epsilon}_a$ and the separation vector $\vec{r}_{jk}$:

$$\cos\theta_{jk} = \frac{(\vec{\epsilon}_a \cdot \vec{r}_{jk})^2}{r_{jk}^2} \quad . \tag{E.14}$$

To evaluate the time and frequency integrals in Eq. (E.12), we employ the usual Born-Markov approximation. For this purpose we replace $c_{q\ominus j\oplus k,0}(t')$ with $c_{q\ominus j\oplus k,0}(t)$ and take this quantity outside the integral. We also consider the $t \to \infty$ limit of the inner time integral, which results in $\int_0^\infty \exp\left[i(\nu_\kappa - \omega_a)t'\right] dt' = \pi\delta(\nu_\kappa - \omega_a) + iP\{\frac{1}{\nu_\kappa - \omega_a}\}$ (δ is the Dirac delta function and $P\{\}$ stands for the principal value). Using all of these results and simplifications, Eq. (E.12) reduces to:

$$\frac{dc_{q,0}}{dt} = -\sum_j \sum_k F_{jk} c_{q\ominus j\oplus k,0} \quad , \tag{E.15}$$

where F_{jk} are the exchange coupling coefficients,

$$\begin{aligned} F_{jk} = F_{kj} = -(i\frac{\Gamma}{2} + \delta\omega)(\frac{3}{8\pi})\Big[4\pi(1-\cos^2\theta_{jk})\frac{\sin\kappa_a r_{jk}}{\kappa_a r_{jk}} \\ + 4\pi(1-3\cos^2\theta_{jk})(\frac{\cos\kappa_a r_{jk}}{(\kappa_a r_{jk})^2} - \frac{\sin\kappa_a r_{jk}}{(\kappa_a r_{jk})^3})\Big] \end{aligned} \tag{E.16}$$

In the large N and constant density limit where almost all pairs of subsystems are very far from one another, this expression approaches the simpler $F_{jk} \sim \sin(kr_{jk})/kr_{jk}$ which is used in the main text.

E1 Beyond the Born-Markov Approximation

The Born-Markov approximation that allows the derivation of the effective exchange Hamiltonian assumes the evolution time-scales for the state coefficients to be slow compared to the correlation time scales of the bath. It is important to understand our system beyond this approximation, since as $N \to \infty$ the system dynamics become ever faster and Born-Markov approximation is violated. Without the Born-Markov

approximation, the state coefficients evolve according to an integro-differential equation of the form:

$$\frac{dc_q}{dt} = -\left(\frac{3}{8\pi}\right)\frac{\Gamma}{2}\sum_j\sum_k(1-\cos^2\theta_{jk})\int_0^t \exp\left[i\omega_a(t-\tau)\right]G(t-\tau)c_{q\ominus j\oplus k}(\tau)d\tau \quad \text{(E.17)}$$

where

$$G(t-\tau) \equiv \frac{2\pi}{(r_{jk}/c)}\text{Box}\left[\frac{t-\tau}{(r_{jk}/c)}\right] - i\frac{2}{(r_{jk}/c)}\ln\left[\frac{(r_{jk}/c)+(t-\tau)}{|(r_{jk}/c)-(t-\tau)|}\right] \quad . \quad \text{(E.18)}$$

Here, the Box function equals one when its argument is between 0 and 1, and equals zero otherwise. The integration Kernel $G(t-\tau)$ captures the correlation (memory) time scales of the bath, which has a width of order r_{jk}/c. We have verified that the above model is consistent with many other formulations of large-sample superradiance, as discussed in detail, for example, in the review article Ref. [41], Section 7. The above model clarifies the effect of the Born-Markov approximation in large samples. If the evolution time scales for the state coefficients are slow compared to the width of the function $G(t-\tau)$, then the Kernel effectively acts like a delta-function. Under these conditions, $c_{q\ominus j\oplus k}(\tau)$ coefficients can be taken outside the integral and be replaced by $c_{q\ominus j\oplus k}(\tau) \approx c_{q\ominus j\oplus k}(t)$. If the coefficients are not slowly varying, then this approximation cannot be applied. When Born-Markov approximation is not satisfied, the rate of change in each coefficient c_q is not only influenced by the values of the coefficients at that particular time, $c_{q\ominus j\oplus k}(t)$, but instead depend the values of the coefficients over the time window $[t, t-r_{jk}/c]$. We note that all the scaling results that we discuss in the manuscript result from N dependent sums over the prefactors in fronts of these coefficients, which have almost random phases (due to large spacings) and $1/r_{jk}$ scaling factors. It is very hard to see how any of these scaling results could be altered when one considers not just the values of the coefficients at that particular time, but also takes into account their evolution history, since neither phases nor the average strengths of the prefactors are altered.

A good physical model for correlated decay in large samples can be summarized as follows: consider a specific qubit j in the ensemble. The radiated fields from all the other qubits in the computer interfere at the position of the qubit j producing a randomly fluctuating field. This fluctuating field results in random phase rotations and population transfer in qubit j, causing bit-flip and decoherence errors. When Born-Markov approximation is satisfied, we evaluate the field values at a specific point in time and consider their interference. When Born-Markov approximation is violated, we need to take into account not only each emitted field at a specific point in time, but also the history of the field over the window $[t, t - r_{jk}/c]$. But for both cases, neither the number of fields that are interfering, nor the average strength of each field is altered. As a result, the scaling results that we discuss should remain valid even when the Born-Markov approximation is violated.

References

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press).

[2] S. Lloyd, Science **261**, 1569, 17798117.

[3] D. P. DiVincenzo, Science **270**, 255.

[4] C. H. Bennett and D. P. DiVincenzo, Nature **404**, 247.

[5] A. Wallraff *et al.*, Nature **431**, 162.

[6] J. Clarke and F. K. Wilhelm, Nature **453**, 1031.

[7] E. Knill, R. Laflamme, and G. J. Milburn, Nature **409**, 46.

[8] T. B. Pittman, B. C. Jacobs, and J. D. Franson, Physical Review Letters **88**, 257902.

[9] P. W. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pp. 124–134.

[10] S. Lloyd, Science **273**, 1073, 8688088.

[11] M. Sipser, The History and Status of the P Versus NP Question, in *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '92, pp. 603–618, ACM.

[12] P. W. Shor, SIAM Journal on Computing **26**, 1484, quant-ph/9508027.

[13] A. Ekert and R. Jozsa, Reviews of Modern Physics **68**, 733.

[14] R. W. Hamming, The Bell System Technical Journal **29**, 147.

[15] B. M. Terhal and G. Burkard, Phys. Rev. A **71**, 1094 (2005).

[16] P. Aliferis, D. Gottesman, and J. Preskill, Quantum Information and Computation **6** (2006).

[17] D. Aharonov, A. Kitaev, and J. Preskill, Physical Review Letters **96**.

[18] H. K. Ng and J. Preskill, Physical Review A **79**.

[19] J. Preskill, Quantum computing and the entanglement frontier, 2012, arXiv:1203.5813.

[20] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, Physical Review Letters **75**, 4714.

[21] F. Schmidt-Kaler *et al.*, Nature **422**, 408.

[22] R. Blatt and D. J. Wineland, Nature **453**.

[23] D. Schrader *et al.*, Physical Review Letters **93**.

[24] T. Wilk *et al.*, Physical Review Letters **104**.

[25] D. Loss and D. P. DiVincenzo, Physical Review A **57**, 120.

[26] J. Berezovsky *et al.*, Science **314**, 1916.

[27] J. Preskill, Quantum Information and Computation **13**, 0181 (2013).

[28] G. Kalai, How quantum computers fail: Quantum codes, correlations in physical systems, and noise accumulation, 2011, arXiv:1106.0485.

[29] G. Kalai, The quantum computer puzzle (expanded version), 2016, arXiv:1605.00992.

[30] S. Aaronson, Proceedings of the thirty-sixth annual ACM symposium on Theory of computing - STOC '04 .

[31] R. Bousso, Black hole entropy and the bekenstein bound, 2018, arXiv:1810.01880.

[32] P. C. W. Davies, Fluctuation and Noise Letters **7**, 04 (2007).

[33] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Physical Review A **86**.

[34] T. F. Gallagher and P. Pillet, DipoleâDipole Interactions of Rydberg Atoms, in *Advances In Atomic, Molecular, and Optical Physics*, volume 56 of *Advances in Atomic, Molecular, and Optical Physics*, pp. 161–218, Academic Press.

[35] R. H. Dicke, Phys. Rev. **93**, 99 (1954).

[36] K. Buschow and F. de Boer, *Physics of Magnetism and Magnetic Materials* (Springer).

[37] Y. Sekino and L. Susskind, JHEP **0810**, 065.

[38] P. Hayden and J. Preskill, Journal of High Energy Physics **2007**, 120, 0708.4025.

[39] E. R. Livine and D. R. Terno, Nuclear Physics B **741**, 131, gr-qc/0508085.

[40] P. Strack and S. Sachdev, Physical Review Letters **107**.

[41] M. Gross and S. Haroche, Physics Reports **93**, 301.

[42] F. Alet and N. Laflorencie, Comptes Rendus Physique **19**, 498, 1711.03145.

[43] M. Rangamani and T. Takayanagi, *Holographic Entanglement Entropy* (Springer).

[44] B. Lemberger and D. D. Yavuz, Physical Review A **96**, 062337 (2017).

[45] A. Steane, Proc. Roy. Soc. Lond. **A452**, 2551 (1996).

[46] M. Schlosshauer, Reviews of Modern Physics **76**, 1267, quant-ph/0312059.

[47] S. Eibenberger, S. Gerlich, M. Arndt, M. Mayor, and J. Tuxen, Physical Chemistry Chemical Physics **15**, 14696.

[48] S. Boughn, Making sense of the many worlds interpretation, 2018, arXiv:1801.08587.

[49] C. Fuchs, Mind and Matter **15** (2017).

[50] L. Marchildon, Foundations of Physics **45**, 754.

[51] W. H. Zurek, Progress of Theoretical Physics **89**, 281.

[52] Line Hjortshoj Pedersen and Klaus Molmer, Physical Review A **79**.

[53] I. E. Mazets and G. Kurizki, J. Phys. B: At. Mol. Opt. Phys. **40**, F105.

[54] E. Novais, E. R. Mucciolo, and H. U. Baranger, Physical review letters **98**, 040501.

[55] R. Gambini, L. P. García-Pintos, and J. Pullin, Phys. Rev. A **100**, 012113 (2019).

[56] K. V. Kuchar, Int.J.Mod.Phys. **D20**, 3.

[57] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, Journal of High Energy Physics **04**, 22.

[58] C. Cao, S. M. Carroll, and S. Michalakis, Physical Review D **95**, 024031, 1606.08444.

[59] M. Raasakka, Symmetry, Integrability and Geometry: Methods and Applications , 1605.03942.

[60] Athanasios Papoulis and S. Unnikrishna Pillai, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill).

[61] T. Wang and S. F. Yelin, Phys. Rev. A **72**, 043804 (2005).

[62] E. Wolfe and S. F. Yelin, Phys. Rev. Lett. **112**, 140402 (2014).

[63] R. T. Sutherland and F. Robicheux, Phys. Rev. A **93**, 023407 (2016).

[64] R. T. Sutherland and F. Robicheaux, Phys. Rev. A **95**, 033839 (2017).

[65] N. Ghazanfari and O. E. Mustecaplioglu, Phys. Rev. A **89**, 043619 (2014).

[66] O. E. Mustecaplioglu and L. You, Phys. Rev. A **62**, 063615 (2000).

[67] R. K. M. O. Araujo, I. Kresic and W. Guerin, Phys. Rev. A **117**, 073002 (2016).

[68] M. H. S. J. Roof, K. J. Kemp and I. Sokolov, Phys. Rev. A **117**, 073003 (2016).

[69] B. C. S. T. Hill and H. Deng, Phys. Rev. A , 033832 (2017).

[70] K. P. M. J. J. Choquette and B. C. Sanders, Phys. Rev. A **82**, 023827 (2010).

[71] C. P. M. S. Utsunomiya and Y. Yamamoto, J. Opt. Soc. Am. B **24**, 198 (2007).

[72] D. V. K. I. M. Sokolov, M. D. Kupriyanova and M. D. Havey, Phys. Rev. A **79**, 053405 (2009).

[73] Y. Yamamoto and A. Imamoglu, *Mesoscopic Quantum Optics* (John Wiley & Sons).