ABSTRACT

Title of dissertation: GENERATION AND USES OF DISTRIBUTED ENTANGLEMENT IN QUANTUM INFORMATION

Zachary Eldredge
Doctor of Philosophy, 2019

Dissertation directed by: Professor Alexey Gorshkov
Department of Physics

In this thesis, we focus on the questions of how quantum entanglement can be generated between two or more spatially separated systems and, once generated, how it can be applied in quantum technology. First we will discuss a protocol, which we conjecture to be optimal in some regimes, that quickly creates entangled states across long distances in systems with power-law interactions. We will discuss how this protocol compares with currently known bounds on entangled state generation and how it might be implemented in a three-dimensional lattice of Rydberg atoms.

Next, we will turn our attention to more general questions of how the Lieb-Robinson bound and other limitations on entanglement can be used to inform the design of quantum computers. Quantum computers will be required to create entanglement if they are to realize significant advantages over classical computers, meaning that the generation of entanglement is an important question. First, we will discuss the implications of the Lieb-Robinson bound on graph descriptions of quantum computer architectures, and how the relevant graph parameter (diameter)

compares to likely cost functions for architectures, such as maximum graph degree and total number of necessary connections. We will present a proposed graph architecture, the hierarchical product, which we believe provides excellent balance between these considerations.

We will then introduce new methods of evaluating graphs that allow us to include quantum architectures capable of measurement and feedback operations. After doing so, we will show that the generation of entanglement entropy becomes a limit on computation. We will show that, for several possible physical models of computation, the generation of entanglement can be bounded by simple graph properties. We demonstrate a connection between worst-case scenarios for entanglement generation and a graph quantity called the Cheeger constant or isoperimetric number, which we evaluate for several proposed quantum computing architectures.

Finally, we will look at the scenario of quantum sensing. In particular, we will examine protocols for quantum function estimation, where quantum sensors are available to measure all of the inputs to the function. We will demonstrate that entangled sensors are more capable than non-entangled ones by first deriving a new lower bound on measurement error and then presenting protocols that saturate these bounds. We will first do so for linear functions of the measured quantities and then extend this to general functions using a two-step linearizing protocol.

GENERATION AND USES OF DISTRIBUTED ENTANGLEMENT IN QUANTUM INFORMATION

by

Zachary Eldredge

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Steven Rolston, Chair
Professor Alexey Gorshkov, Advisor
Professor Andrew Childs, Dean's Representative
Professor Mohammad Hafezi
Professor Brian Swingle

Acknowledgments

No one does anything alone. This dissertation and the body of work it represents would never have happened without the support I received throughout my entire life from many people that I wish to thank at this time.

I would first like to thank my family. My parents, grandparents, and everyone else who encouraged me to apply myself, who nurtured my love of learning, and gave me the starting point I needed to succeed.

The most important thing to any student is teachers, and I have been blessed with excellent ones. These include my high school physics teacher, Darrel Negelein, who first encouraged me to major in physics, as well as my undergraduate research advisor Gregory Parker. My entire education from kindergarten to Ph.D. has occurred in public schools. I owe an immense debt to the people of Oklahoma and Maryland for the education systems that gave me a place to learn, thrive, and grow for the past twenty years. No teacher I ever had was paid enough.

In graduate school, I have had many excellent professors and mentors. My advisor, Alexey Gorshkov, has never failed to be a source of energy and enthusiasm in all things. I also have benefited greatly from classes, discussions, and collaborations with many professors, not least my committee: Steve Rolston, Mohammed Hafezi, Andrew Childs, and Brian Swingle. Other great friends, collaborators, and teachers include Jacob Taylor, Ian Spielman, Elizabeth Goldschmidt, and Chandra Turpen.

Postdocs and students both in and out of the Gorshkov group have been my first and most reliable collaborators and mentors, including Michael Foss-Feig, Jim

Garrison, Wenchao Ge, Zhexuan Gong, Aniruddha Bapat, Su-Kuan Chu, Abhinav Deshpande, Chris Flower, Andrew Guo, Kevin Qian, Minh Tran, and Jeremy Young. Finally, my officemates, Alireza Seif and Hwanmun Kim, have both been a joy to work beside throughout my graduate career.

I would like to thank Donna Hammer, Jessica Crosby, Peter Shawhan, and Logan Anbinder for their tireless work in the physics department administration. The physics department has benefited greatly from their work to make it a more welcoming place for all students.

My activity in student life in graduate school was key to my education and helped make my time at UMD meaningful and fulfilling. I would like to thank my colleagues on the Physics Graduate Student Committee, the Mental Health Task-force, the Access Network, and Science for the People, both locally and nationally, for their work to build a better world both within science and without.

Graduate school is only survivable thanks to friends. I want to thank my roommates at Seven Seas for giving me a home when I first arrived in Maryland. Thank you to Courtney Cowper, Chris Flower, Rachel Gartner, Matt Harrington, Kevin Palm, Gina Quan, Steve Ragole, David Somers, and Antony Speranza. Thank you also to many other friends, Rodney Snyder, Molly Carpenter, Sandy Craddock, Clarissa Melkonian, Liz Friedman, Jimmy Juno, Nat Steinstultz, and others I'm sure I've missed.

Finally, I owe more thanks than can fit in these pages to Katie, who has loved and supported me through so much. My life is so much better and brighter with you.

And the rain came down like water from a faucet
I saw him at the crossroads but I never saw him cross it
One minute he was on this side, and the next he was over on the other
So went his rounds, one sad patient after another.

– The Mountain Goats, "The Bad Doctor."
Bitter Melon Farm, 1999

Table of Contents

List of Tables

List of Figures

Citations to Previously Published Work

Much of this dissertation appeared in papers already published. Here we outline those publications, with additional citations to related papers whose text does not appear here.

- Chapter 2: "Fast State Transfer and Entanglement Renormalization Using Long-Range Interactions," Z. Eldredge, Z.-X. Gong, J.T. Young, A.H. Mosavian, M. Foss-Feig, A.V. Gorshkov, Phys. Rev. Lett. 119, 170503. 2017.
  - Further work related to Lieb-Robinson bounds for power-law interacting systems is also presented in "Locality and digital quantum simulation of power-law interactions," M.C. Tran, A.Y. Guo, Y. Su, J.R. Garrison, J.R., Z. Eldredge, M. Foss-Feig, A.M. Childs, A.V. Gorshkov, arXiv: 1808.05225.
  - Further work related to entanglement renormalization is presented in "Scale-Invariant Continuous Entanglement Renormalization of a Chern Insulator," S.-K. Chu, G. Zhu, J.R. Garrison, Z. Eldredge, A.V. Curiel, P. Bienias, I.B. Spielman, A.V. Gorshkov. Accepted for publication in Phys. Rev. Lett.
- Chapter 3: "Unitary Entanglement Construction in Hierarchical Networks," A. Bapat, Z. Eldredge, J.R. Garrison, A. Desphande, F.T. Chong, A.V. Gorshkov, Phys. Rev. A 98, 062328. 2018.

- Chapter 5: "Optimal and Secure Measurement Protocols for Quantum Sensor Networks," Z. Eldredge, M. Foss-Feig, J.A. Gross, S.L. Rolston, A.V. Gorshkov, Phys Rev. A 97, 042337, 2018 and "Heisenberg-Scaling Measurement Protocol for Analytic Functions with Quantum Sensor Networks," K. Qian, Z. Eldredge, A.V. Gorshkov, M. Foss-Feig, arXiv: 1902.09042.

  - Further work applying Fisher information to metrology can be found in "Simultaneous, Full Characterization of a Single-Photon State," T. Thomay, S.V. Polyakov, O. Gazzano, E. Goldschmidt, Z. Eldredge, T. Huber, V. Loo, G.S. Solomon, Phys. Rev. X 7, 041036. 2017.

  - Further work characterizing the application of distributed metrology to linear optical networks can be found in "Distributed Quantum Metrology with Linear Networks and Separable Inputs," W. Ge, K. Jacobs, Z. Eldredge, A.V. Gorshkov, M. Foss-Feig, Phys. Rev. Lett. 121, 043604. 2018.

Chapter 1: Introduction

1.1 Motivation

Since the beginning of the 20th century, physics has been reconstructed on the basis of quantum mechanics. This scientific revolution, in addition to revamping our view of the fundamental properties of nature, has also yielded many technologies based on the new theory. As compensation for letting go of classical physics, we have been gifted with new technologies ranging from magnetic resonance imaging to the electronic transistor.

Now, in the 21st century, a new crop of quantum technologies is bearing fruit. The National Institutes of Standards and Technology has echoed many others by referring to a "second quantum revolution" [1]. This new phase includes quantum computing and quantum communication, which is why "quantum information" is the usual appellation for the field of science underlying these technologies. Whereas as previous technologies may have taken advantage of uniquely quantum properties of matter, the new technologies manipulate and transmit information in ways that are fundamentally different from classical devices.

This latest wave of technologies depends crucially on entanglement. Although subtle, entanglement is one of the most fundamental breaks between quantum and

classical physics. In some ways, quantum entanglement represents the end point of a very long project – reductionism. For centuries, we have used reductionist methodology to break physical systems down into ever-smaller parts, explaining the behavior of substances by appealing to molecules, explaining the behavior of molecules by appealing to atoms, and so forth. This project stretches back to antiquity: Democritus introduced the notion of atoms into Greek natural philosophy to resolve the paradox of changeable everyday experiences arising from unchanging microscopic constituents [2].

Quantum entanglement can be viewed as a rebuke to the reductionist impulse, because entanglement makes the decomposition of systems into smaller parts impossible. To illustrate this, we can imagine taking a "snapshot" of a physical system, yielding a "God's eye view." In classical physics, this set of data would consist of storing the state of each individual particle. However, in quantum physics, we would instead have to store the complex probability amplitude of every possible configuration of all the particles, meaning that a general description is *not* built up from the descriptions of smaller particles. Note that, on the other hand, if we were guaranteed to be dealing with an unentangled state, this exponential scaling would not occur. The many-body state of the system would be like the classical one, describable in terms of independent parts.

Although this causes immense challenges for theorists who would like to calculate the behavior of quantum many-body systems, this is actually a fantastic opportunity. We know that classical physics exists (through the famous correspondence principle) as a subset of quantum physics, an effective description that works

well enough at human scales but only does so by ignoring the rich complexity of the quantum realm. We live in a very tiny corner of the possibilities that quantum mechanics allows for. Technologically, this means that the devices and methods built according to classical physics are using only a small portion of their physical potential. There is no reason why a classical computer, classical measurement, or classical communication channel should be the best option in this context. Quantum technology thus offers the possibility that we can tap into these entangled possibilities and use them to solve human-scale problems.

It is perhaps somewhat surprising to realize that entanglement was not discussed until quantum theory was over thirty years old, if we begin the accounting with Planck's 1900 paper on black body radiation (which does not mention entanglement) [3]. Entanglement does not appear in Einstein's 1905 exploration of the photoelectric effect [4]. Nor does Bohr mention it in his 1913 atomic model [5], or Schrodinger in the 1926 paper introducing his wave equation [6]. It was Einstein, Podolsky, and Rosen's famous 1935 paper, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" [7] which first pointed out and rigorously described the strange correlations of entangled particles. Since then, it has been clear that entanglement is perhaps most interesting when the entangled systems in question are separated across large distances in space. Einstein believed the apparently instantaneous action at a distance that arose in this thought experiment disproved the notion that quantum mechanics could be a final description of reality. However, examination by physicists and philosophers in the decades since have shown that quantum mechanics still yields correct predictions in this setting [8, 9].

Experiments based on EPR and the subsequent work of John Bell [10] have allowed theorists and experimentalists to show that entanglement is a real, intrinsic property of quantum mechanics which is deeply intertwined with the notions of indeterminism within the theory.

All of this leads us to study not just what entangled states are, but how they can be used, and what possibilities they open up technologically. Similarly, we are interested not just in what entanglement can do in general, but what particular possibilities and restrictions emerge when we focus on entanglement's spatial distribution. The inherent difficulty of creating such systems also guides us to study the limitations on entanglement generation and to seek to meet those limits if possible.

1.2 Entanglement

In quantum mechanics, the state of a quantum system is described by a normalized ray in Hilbert space, $|\psi\rangle \in \mathcal{H}$. For any measurement made on the system described by the state $|\psi\rangle$, we can predict the probability of each possible outcome of the measurement through the wavefunction. In general, this involves a positive operator-valued measure (POVM), but in many cases the POVM reduces to projecting $|\psi\rangle$ onto the eigenstates of the observable of interest.

To combine the wavefunction of two systems, a tensor product of two states is taken. Thus, if system 1 is in state $|\psi_1\rangle$ and system 2 is in state $|\psi_2\rangle$, their overall state would be:

$$|\psi_{12}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \,. \tag{1.1}$$

The new overall Hilbert space is $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$. However, it should be noted that not every vector in $\mathcal{H}_{12}$ can be expressed as a tensor product in this manner. Although we can always find a basis in which every basis vector is a product of vectors in $\mathcal{H}_1$ and $\mathcal{H}_2$, Hilbert spaces are vector spaces, and therefore we can add these basis vectors together to yield new wavefunctions and valid kets which are not decomposable. Consider a two-qubit system in which each qubit has basis states $|0\rangle, |1\rangle$. One very simple entangled state might be:

$$|\psi_{12}\rangle = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle . \tag{1.2}$$

This state cannot be written in the form $|\psi_1\rangle \otimes |\psi_2\rangle$. Therefore, we call it *entangled*. Because the two particles are entangled, we cannot describe the two-particle system in terms of two individual particle states. Instead, we have to describe the system as a whole. Unlike the case of a classical mixture, in which the state would be either $|00\rangle$ or $|11\rangle$, but unknown to us, the entangled state persists even if we have the best possible knowledge of the state of the two particles. The individual identities of the two particles, as physical entities with independent, localized properties, have vanished.

1.3 Distributed Entanglement

While entanglement as a mathematical consequence of the algebra of Hilbert space is interesting, it becomes a more consequential physical fact when we suppose that the state $|\psi_{12}\rangle$ describes two qubits which are separated by some macroscopic distance. Suddenly, two objects, which could be arbitrarily far apart, cannot be de-

scribed as two separate entities. This gives rise to what Einstein famously referred to as "spooky action at a distance." Consider the state $|\psi_{12}\rangle$ given in Eq. (1.2). Suppose Alice and Bob are separated by large distances, kilometers or even light-years, and each of them possesses one of the particles described by $|\psi_{12}\rangle$. If Alice measures her particle, she may obtain either outcome $|0\rangle$ or $|1\rangle$ with equal probability, and the same is true of Bob. However, as soon as Alice makes her measurement, Bob's measurement ceases to have an equal probability of either outcome, since his result must be the same as Alice's. In this way, it seems that the act of measurement by Alice has influenced not only her system, but Bob's, no matter how far away he was at the time.

One obvious issue with this scenario is that it would seem to violate the principle of relativity, which forbids signalling faster than light. Indeed, a relativistic perspective means that it is in general impossible to say whether Alice or Bob makes their measurement first if the two are separated by a spacelike interval in spacetime, so if we accept that there is causation in this process, then we could not truly say which measurement was the cause and which was the effect. More philosophically, Einstein, Podolsky, and Rosen objected that, since there was no physical mechanism of causation – no messenger particles or fields, no interaction at all – it made no sense to assert that one particle affected another. As they write, "no reasonable definition of reality could be expected to permit this" [7].

Although there is a correlation between what happens at one qubit and what happens at another, careful analysis shows that signalling is actually impossible. There is no way to use an entangled state to communicate any information in-

stantaneously; in fact, the probability that Bob will receive a particular outcome is independent of anything Alice does. Quantum communication requires classical communication to convey information. Since nothing Alice or Bob does can affect the outcome of the other measurement, the existence of multiple reference frames where one measurement or the other occurs first does not pose an issue.

If no signalling is possible between Alice and Bob, what are we to make of the fact that the statistics of one measurement seem different after we make our first measurement? Quantum mechanics has a long history of interpretations, as many physicists and philosophers have attempted to map the mathematical descriptions of the theory, which are so adept at predicting experimental outcomes, to the intuitive understanding of physical objects we hold in our heads. Different people have come to very different conclusions about what happens in the EPR experiment.

We do not have space here to present every possible interpretation of the EPR thought experiment, but it's worth reflecting on the diversity represented by just a few of them. In Bohr's Copenhagen interpretation, the wavefunction $|\psi_{12}\rangle$ represents all that can be known about the particles, because they had no definitive state before the measurement was made. After the measurement is made, a new wavefunction is appropriate to describe the system, but Bob cannot use this new wavefunction until Alice communicates it to him – so no paradox is really introduced, Bob simply is in a state of ignorance to the new quantum state. More radical is a set of interpretations known as quantum Bayesianism or QBism, which posits that quantum mechanics should be interpreted as a set of suggestions for an individual observer to assign probabilities to events [11]. In this interpretation, it does not make sense to say

that the two particles "have a state"–rather, the state is *assigned* by a particular actor, and then updated when new measurement data reaches that actor. Since the wavefunction is a subjective description, there is no need to pretend that anything "happens" when it changes anywhere except in the mind of the observer.

Other interpretations make more room for an objective reality. For instance, in the "many-worlds" interpretation of Everett, every branch of the wavefunction "really exists." When Alice makes her measurement, one version of her obtains result $|0\rangle$ and another obtains result $|1\rangle$ [12]. The universe "splits" into two separate, non-interfering branches, and the Bob in one universe obtains a different result than the Bob in the other. Meanwhile, hidden-variable interpretations like those of de Broglie-Bohm are willing to discard locality entirely and accept that, despite relativistic objections, there really is a physical mechanism that communicates measurement results from Alice's laboratory to Bob's [13].

Ultimately, these philosophical issues, although important, are unanswerable within the framework of quantum theory itself. All of them are set up in such a way as to reproduce each other in all experiments for the foreseeable future. In this thesis, we are not interested in unpacking the exact meaning of distributed entanglement, metaphysically. Instead, we are interested in how these strange states can be created and, once created, how they can be used. This will mean two aspects: first of all, we will want to examine the uses of entanglement, and how an entangled state might prove a useful resource for technological tasks. Second, we will want to focus on specifically *distributed* entanglement, which raises its own questions. If an entangled state is (in general) useful, what uses persist – or what new uses arise – when the

entangled particles themselves are separated by large distances?

1.4 Limits on the Generation of Entanglement

Because entanglement is such an interesting phenomenon in quantum mechanics, and one which can be used for many different tasks, such as quantum computing [14] and metrology (see Sec. 1.6), an important technological question to address is how quickly entanglement can arise from previously unentangled particles. After all, the existence of an interesting or useful entangled state is of little use to us if the process that creates it takes an extremely long time to complete.

How, precisely, entangled states arise varies in different settings. In this thesis we will examine cases in which particles interact unitarily through a many-body Hamiltonian. For instance, we might consider a system of N spins that interact through the following Hamiltonian:

$$H = \Delta \sum_{i=1}^{N} \sigma_i^z - J \sum_{i=1}^{N-1} \sigma_i^x \sigma_{i+1}^x. \tag{1.3}$$

Here, J is a coupling constant signifying the strength of the spin-spin interaction, Δ is transverse field, and σ_i^k is the Pauli matrix $k \in \{x, y, z\}$ on site i. If the particles interact according to this Hamiltonian, how quickly can entanglement be generated from an initially unentangled state? To quantify this we must choose a particular measure of entanglement. Strictly speaking, the state after any finite amount of time will be "entangled" in most cases, since the spins very rapidly develop *some* amplitude in non-product states. We instead would like to know the time required for a distant set of spins to be significantly entangled. One way to look at this is

to consider how long it takes two operators to stop commuting. For instance, the Pauli operators σ_i^x and σ_{i+1}^y, on sites i and $i+1$, commute at time $t = 0$. If we subject them to evolution in the Heisenberg picture, however, the support of these operators grows, so that for $t > 0$ the commutator of $\sigma_i^x(0)$ and $\sigma_{i+1}^y(t)$ becomes nonzero. The clearest example would be if the evolution in question exchanged the state of spin i and $i+1$ after a fixed amount of time, in which case $\sigma_{i+1}^y(t)$ would actually become $\sigma_i^y(0)$, which is maximally non-commuting with $\sigma_i^x(0)$.

This process has been quantified by Lieb and Robinson in a seminal paper [15] which established that there is a time limit on this non-commutation. In particular, for any Hamiltonian which, like Eq. (1.3), has interactions that are either finite-range or decay exponentially, they showed that:

$$\|[A(t), B(0)]\| \leq \|A\|\|B\|ce^{vt-r}, \tag{1.4}$$

where c and v are constants that do not change with t or r, $\|\cdot\|$ denotes the operator norm, and r is the separation between the initial sites of A and B. We can use this to establish an initial bound on the rate at which entanglement can be generated, by associating v with a velocity at which entanglement can spread. This is known as a "light cone" because it establishes a region within which entanglement is possible. Bravyi has made this treatment based on non-commuting operators applicable to quantum information processing, by showing that the Lieb-Robinson bound can be directly applied to entanglement-related tasks like the generation of non-classical correlations and shifting between different topological states [16].

An interesting question, much studied but still very open, is how this picture

changes if we open up the space of allowed interactions. For instance, Coulomb interactions, rather than being only between site i and site $i+1$, would permit interactions at long range with a strength that decreased proportional to $1/r$; other systems like electric and magnetic dipoles likewise have "power law" interactions of the form $1/r^{\alpha}$ with $\alpha = 3$ for dipolar interactions. It is known that, depending on the spatial dimension of the system, a "light cone" may arise for some values of α, although it may be logarithmic or polynomial, with entanglement possible at $t \sim r^{\beta}$ for some β instead of $t \sim r$ as in the finite-range case [17–19]. Later, we will attempt to investigate the limits of entanglement generation with long-range interactions by presenting and studying an example system.

1.5 Quantum Architectures

The generation of quantum entanglement is of special interest in the field of quantum computing. This is because it is known that, if a quantum process does not produce intermediate states with an entanglement entropy that increases with the number of memory qubits, there are feasible polynomial-time classical algorithms to simulate it [20]. Therefore, any quantum computer that intends to create a sizable advantage over classical computations will need to be able to create large entangled states. Indeed, models of quantum computation exist which use only one large initial entangled state, where the entire process of computation can be performed by classical operations of measurement and feedback [14]. Since quantum computers are also bound to be racing against noise that decoheres useful quantum states into

functionally-classical ones, it is also important that we generate this entanglement as quickly as possible.

Whether or not these states represent "distributed" entanglement depends on precise definitions – different memory registers will necessarily occupy different locations within the computer, just as different bits on a hard disk are stored in different locations. For many situations, however, it is clear that we will want to produce quantum entanglement across long distances as part of our schemes. Most prominently, we may wish to construct networks of computers which are composed of smaller "modules" [21, 22]. This would allow quantum computing to emulate the cluster model of computation deployed with great success in classical supercomputers. These modules may be specialized to perform particular quantum operations, or may simply be employed to allow for a more scalable approach to quantum systems. It is possible that quantum systems of arbitrarily large size will be prevented on physical or technological grounds, leaving the construction of such distributed clusters as the only viable means to creating large systems with the thousands of qubits required to run high-quality quantum computations.

Once we set our sights on the creation of quantum architectures, a rich and interesting set of problems emerges relating to the design of these modular networks. How can we ensure that communication between modules is easy, and that the modules are arranged in ways which maximize the connectivity between different parts of the system? To answer this question, we must begin to grapple with the engineering challenges brought on by building such connectivity – how can we have a well-connected arrangement that does not cost too much or require huge amounts of

communication infrastructure? How can these questions be answered in the context a particular calculation, for instance, if we want to perform Shor's algorithm for factoring numbers on a quantum computer, should we adopt a particular architecture suited for it, and how can we identify that architecture? We will explore all these questions.

1.6 Quantum Metrology

While quantum architecture concerns itself with how entanglement can be generated in universal quantum systems, we are also interested in what novel uses a quantum entangled state might have outside of computing. One which is well-known in the literature [23] is precision measurement and timekeeping.

In general, when a measurement is made N times instead of once, the error on the measurement decreases by a factor of $\sqrt{N}$. In quantum technology, when we are given N systems which are available to measure a quantity – such as a magnetic field or a temperature – we can choose between using them as individual systems or creating entangled states of some subsets of the measurement systems. It can be shown that an entangled strategy can be dramatically more powerful than an unentangled one, with the enhancement factor from using N sensors possibly being as high as N (rather than $\sqrt{N}$ in the classical case).

This improvement can be captured in many different settings, for instance, when there are N photons available to probe a sample in a microscope, or when there are N atoms available to sense a magnetic field. We are particularly interested,

again, in places where we can use the non-local nature of entanglement. That means we want to take advantage of the ability for multiple sensors, in multiple locations in space, to perform novel metrological tasks which quantum technology can enhance.

1.7 Outline of Thesis

The remainder of this thesis is organized as follows:

- Chapter 2, "State Transfer and GHZ Creation With Long-Range Interactions" presents a protocol which creates entanglement by using long-range (power-law) interactions. We then show how this same protocol can be applied to transferring a state as well as executing a circuit created by the MERA (multi-scale entanglement renormalization ansatz). We also present a method for performing the protocol in a real system (Rydberg atoms in three-dimensional optical lattices) and examine the effects of experimental imperfections. This chapter originally appeared as Ref. [24].

- Chapter 3, "Unitary Entanglement Construction in Hierarchical Networks" examines the ways in which entanglement can be used to guide the development of modular architectures for quantum computers. We first begin by introducing a family of graphs dubbed hierarchies, whose properties we then analyze and compare to other potential architectures. This chapter first appeared as Ref. [25].

- Chapter 4, "Entanglement Entropy and Non-Unitary Computations" extends the results of Chapter 3. After first analyzing an entirely unitary computa-

tional setting, we broaden our outlook to include non-unitary settings as well, leading us to develop a new graph quantity dubbed "rainbow time" which allows us to evaluate the worst-case time for the creation of highly entangled states on any architecture.

- Chapter 5, "Sensor Networks" will explore the way entanglement can be used in a sensor setting. We first present a method of estimating a single parameter which is given as a linear function of spatially-separated parameters. We will derive a lower bound on the error allowed by quantum mechanics and then present a protocol which saturates this bound. Next, we will extend this to arbitrary functions by using linearization. This chapter combines work from Refs. [26] and [27].

Chapter 2: State Transfer and GHZ Creation With Long-Range Interactions

2.1 Introduction

Entanglement generation in a quantum system is limited, even in a non-relativistic setting, by the available interactions. In a lattice system with short-range interactions, Lieb and Robinson showed that there exists a linear light cone defined by a speed proportional to both the interaction range and strength [15]. Suppose two operators A and B are supported on single sites separated by a distance r. Then the Lieb-Robinson bound states that, after time t,

$$\|[A(t), B]\| \leq c \|A\| \|B\| e^{vt-r}, \tag{2.1}$$

where c is a constant, v is another constant known as the Lieb-Robinson velocity, and $\|\cdot\|$ represents the operator norm. If a system initially in a product state begins evolving under a short-range Hamiltonian, correlations decrease exponentially outside of the causal cone defined by $r = vt$ [28–30]. However, in physical systems including polar molecules [31–33], Rydberg atoms [34, 35], or trapped ions [36, 37], the interactions fall off with distance r as a power law $1/r^{\alpha}$. For these interactions, generalizations of the Lieb-Robinson bound are known, but they may not be

tight [17–19]. In addition, for sufficiently long-ranged interactions the causal region may even encompass infinite space at finite time, signaling a breakdown of emergent locality [38–41].

These bounds on entanglement have direct implications for quantum information processing. The Lieb-Robinson bound, even if time dependence is allowed [16, 42], limits the speed at which operations can be performed or states created using local Hamiltonians, including states with important applications in quantum metrology and communication [23, 43–46]. In this chapter, we consider the task of using long-range interactions to speed up certain quantum information processes, such as quantum state transfer, GHZ (Greenberger-Horne-Zeilinger) state preparation, and MERA (multiscale entanglement renormalization ansatz) construction.

State transfer is a process by which an unknown quantum state on one site in a lattice is transferred to another site [47–50]. Discussion of possible experimental realizations can be found in Refs. [51–53], and in Ref. [54] a case with long-range interactions is considered. Since state transfer establishes perfect correlation between one site at $t = 0$ and another site after the transfer, it is limited by the Lieb-Robinson bound. In this chapter, we propose a state transfer protocol which makes use of long-range interactions to transfer a state a distance L on a d-dimensional lattice in time proportional to L^0 ($\alpha < d$), $\log L$ ($\alpha = d$), $L^{\alpha-d}$ ($d < \alpha \leq d+1$), or L ($\alpha \geq d$). As an intermediate step of the protocol presented, a GHZ-like state is created, a process also limited by the Lieb-Robinson bound [16]. For polar molecules, Rydberg atoms, or other dipole-dipole interactions in three dimensions, the protocol yields an exponential speed-up in the rate of entanglement generation.

As we will discuss, one powerful application of fast state transfer using long-ranged interactions would be the realization of a circuit described by a MERA [55–57]. MERAs are particularly useful ways to represent entangled states [58–60], such as the ground states of the toric or Haah codes, topological insulators, and quantum Hall states [61–64]. By performing state transfer and then applying a two-qubit gate between nearest neighbors, we can speed up long-range two-qubit gates, which we use to upper bound the minimal time required to create a MERA state. Using dipole-dipole interactions in 3D, our protocol constructs the MERA state exponentially faster than using nearest-neighbor interactions.

2.2 State Transfer

Our state transfer protocol first creates a many-body entangled state including the intended starting and final qubits. We do so by applying a controlled X rotation between pairs of qubits (i, j) using a Hamiltonian

$$H_{ij} = h_{ij} \left(|0\rangle \langle 0|_i \otimes I_j + |1\rangle \langle 1|_i \otimes X_j \right). \tag{2.2}$$

Here h_{ij} is the interaction strength, which may not be identical for all pairs of qubits. In Sec. 2.5, we examine a case where the sign of h_{ij} is variable, but for now we take $h_{ij} > 0$. I_j and X_j are the identity and Pauli X operator acting on qubit j. When the Hamiltonian in Eq. (2.2) is applied for a time $t = \pi/(2h_{ij})$, it realizes a controlled-NOT (CNOT) gate between qubits i and j (up to an unimportant phase). In Eq. (2.2), i is the control qubit for the CNOT while j is the target qubit. When applied to a control qubit in an arbitrary state and a target qubit in the state $|0\rangle$,

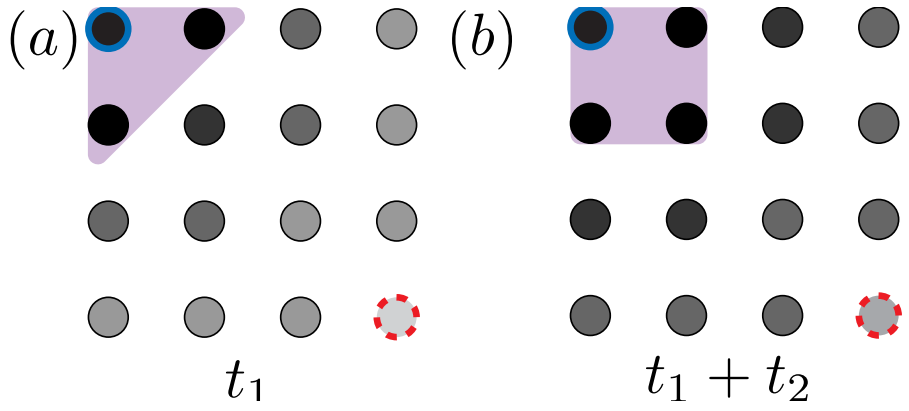

Figure 2.1: Our state transfer protocol using long-range interactions. We want to move a qubit state from the upper-left site (outlined in solid blue) to the lower-right one (outlined in dashed red). After a time t_1 (a), the nearest-neighbor qubits have shifted from target to control (purple region), and continue acting on all other qubits, thereby adding an additional qubit to the set of controls after further time t_2, as shown in (b). After t_2, each qubit has rotated further (shown by darker shading). The growth continues until the original qubit has effectively performed a CNOT on all qubits in the lattice shown.

the CNOT gate results in a two-qubit state encoding the original qubit,

$$\text{CNOT}\left(a\left|0\right\rangle + b\left|1\right\rangle\right)\left|0\right\rangle = a\left|00\right\rangle + b\left|11\right\rangle. \tag{2.3}$$

By continuing this process, we can create a many-body entangled state of N qubits $a\left|0\right\rangle^{\otimes N} + b\left|1\right\rangle^{\otimes N}$ encoding the same state as the initial qubit. The original state can be transferred onto the target qubit by reversing the entangling process and leaving the destination qubit as the final control qubit. If H_{ij} were a nearest-neighbor Hamiltonian, then this procedure would then allow us to transfer a qubit state a distance L by applying L CNOT operations to construct the many-body state and then L other CNOT operations which are properly time-reversed and spatially mirrored, providing a linear scaling which saturates the Lieb-Robinson bound.

By using Hamiltonians with long-range interactions, we can achieve a sublinear state transfer time. We suppose that $h_{ij} = 1/r_{ij}^{\alpha}$, where r_{ij} is the distance between the qubits i and j. [1] Our protocol (Fig. 2.1) starts by acting on all qubits in the lattice with a single control qubit storing the initial state. Once the CNOT operation completes on a qubit, it can be switched from a target to a control and then used to speed up the CNOTs which are still continuing on other qubits. If a single qubit is targeted by many control qubits, then the CNOT operation on that qubit can be completed faster. (Multiple H_{ij} will mutually commute as long as the sets of target qubits and control qubits are disjoint.) If qubit j is targeted by many qubits indexed by i, the time required to complete the CNOT becomes

$$t = \frac{\pi}{2\sum_i h_{ij}} = \frac{\pi}{2\sum_i r_{ij}^{-\alpha}}. \tag{2.4}$$

(By using dimensionless couplings $h_{ij} = 1/r_{ij}^{\alpha}$, we are implicitly giving times in units of the inverse nearest-neighbor coupling strength.) In addition to the progressive inclusion of more control qubits, each subsequent qubit has already been rotated by some angle, reducing the remaining time required to complete the operation. Therefore, additional qubits can be added more quickly to the state as it grows.

As an example, consider beginning with a system of three qubits arranged in

[1] For $\alpha \leq d$, the thermodynamic limit is not well defined unless the Hamiltonian contains a volume-dependent prefactor proportional to $1/L^{d-\alpha}$ for linear system size L (or $\ln L$ if $\alpha = d$) [65, 66]. The inverse of this factor would multiply the required state transfer time. For many physical systems such as polar molecules, this mathematical point will not modify the actually existing interactions over distances of interest, so we do not consider it here.

a line,

$$|\psi(t=0)\rangle = (a\,|0\rangle + b\,|1\rangle)\,|00\rangle\,. \tag{2.5}$$

Simultaneously applying H_{12} and H_{13} for a time $t_1 = \pi/2$, the state becomes

$$|\psi(t_1)\rangle = a\,|000\rangle - ib\,|11\rangle\left(\cos\frac{\pi}{2^{\alpha+1}}\,|0\rangle - i\sin\frac{\pi}{2^{\alpha+1}}\,|1\rangle\right). \tag{2.6}$$

At this point, the second qubit is made a control, so that the acting Hamiltonians are H_{13} and H_{23}. By continuing the evolution under these Hamiltonians for an additional time,

$$t_2 = \frac{\frac{\pi}{2} - \frac{\pi}{2\cdot 2^{\alpha}}}{1 + \frac{1}{2^{\alpha}}} = \frac{\text{rotation remaining}}{\text{sum of interactions}}, \tag{2.7}$$

the system will end in the final state

$$|\psi(t_1 + t_2)\rangle = a\,|000\rangle - b\,|111\rangle\,. \tag{2.8}$$

The entire procedure can be reversed, interchanging the roles of qubits 1 and 3, to transfer the original state,

$$|\psi\,(2\,(t_1 + t_2))\rangle = |00\rangle\,(a\,|0\rangle + b\,|1\rangle)\,. \tag{2.9}$$

We now consider the case of many qubits. Our strategy will be to specify a suboptimal protocol that simplifies the calculation of the time required for state transfer and use that result to bound (from above) the state transfer time of the protocol in Fig. 2.1, which is more difficult to analyze directly. First, we specify that we aim to to construct a GHZ state across a hypercube whose diagonal spans a distance $L\sqrt{d}$. The points on either end of the diagonal are the original and destination sites for state transfer (see Fig. 2.2). Because the state transfer time

using the protocol of Fig. 2.1 is difficult to compute, we use a slightly slower protocol that allows us to easily estimate the transfer time both analytically and numerically. Rather than change a qubit into a control as soon as its evolution completes, we instead halt a qubit's evolution when its rotation finishes. Once we have enough qubits to form a full hypercube of controls, we expand the control set and continue evolution. This scheme is illustrated in Fig. 2.2, and we expect it to perform similarly (in terms of the scaling of transfer time) to the scheme in Fig 2.1. Let $q = 1, 2, \ldots, L$ denote each subsequent expansion of the hypercube, so that after time $t = t_1 + t_2 \cdots + t_q$ we can form a complete control hypercube of edge length q. The times t_q are determined by the condition that each qubit must accumulate a total phase of $\pi/2$,

$$\sum_{p=1}^{q} H(p,q)t_p = \frac{\pi}{2}. \tag{2.10}$$

Here $H(p,q)$ is defined to be the summation of all Hamiltonian strengths h_{ij} for which the control i is in the hypercube with corners $(0,0,0,\ldots)$ and $(p-1,p-1,p-1,\ldots)$ and the target j is at the site $(q,q,q,\ldots)$ at the corner of a larger hypercube containing the first, as illustrated in Fig. 2.2. The qubit j is the slowest-evolving qubit on its layer, so its evolution determines the time required to expand the cube in this scheme.

At this point, we will begin looking for bounds on the times t_q. Our first bound arises by noting that for all p, $t_p > t_{p+1}$. This is because, for each p, the quantity $H(p,p)$ is strictly larger than $H(p-1,p-1)$ – the qubit at $(p,p,\ldots,p)$ has more qubits acting on it than its counterpart in the previous step. We use $t_p > t_{p+1}$ to

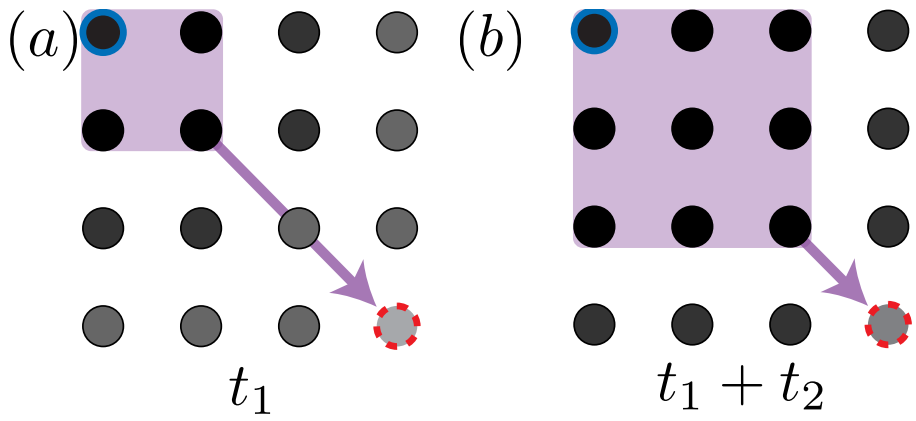

Figure 2.2: (a) The suboptimal protocol used for our bounds, with the same color scheme as Fig. 2.1. After the pth time step, a $(p+1) \times (p+1)$ hypercube of qubits act as controls. The purple arrow represents $H(2,3)$, as it connects a 2×2 square to a qubit at coordinates $(3,3)$. (b) After time $t_1 + t_2$ another set of qubits has been converted from targets to controls. The purple arrow now represents $H(3,3)$.

rewrite the phase condition on times in Eq. (2.10),

$$\frac{\pi}{2} \geq t_q \sum_{p=1}^{q} H(p,q). \tag{2.11}$$

We now construct two complementary bounds for $H(p,q)$. In some cases (small α), $H(p,q)$ will receive appreciable contributions from the entire hypercube of control qubits. In this case, we can obtain a lower bound by pretending that all control qubits are at the same point a distance $q\sqrt{d}$ away, the maximum possible. However, for large α the interaction is dominated by nearby qubits, whose contributions are independent of q. For instance, in $H(q,q)$ there is always one qubit at the nearest vertex of the hypercube whose contribution does not depend on q. These two bounds can be combined to yield:

$$H(p,q) \geq \max\left(\frac{p^d}{\left(q\sqrt{d}\right)^{\alpha}}, \frac{\delta_{pq}}{d^{\alpha/2}} \right). \tag{2.12}$$

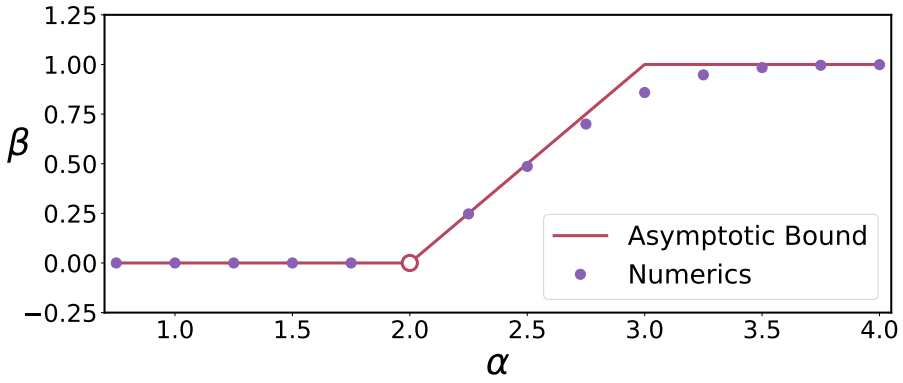

Figure 2.3: Numerical results of solving Eq. (2.10) at different α in $d = 2$. We calculate $\sum_{q \leq L} t_q$ and fit to L^β for L between 900 and 1000; the best-fit exponent is plotted here. The solid line shows the β derived from Eq. (2.13). At $\alpha = d$ (open circle), the numerics are consistent with the expected logarithmic scaling; the fact that the bound is not saturated at $\alpha = 3$ is due to finite L and should vanish in the $L \to \infty$ limit.

After substituting Eq. (2.12) into Eq. (2.11), the sum can be performed. If we discard all constants depending only on d or α, the result is a bound on the scaling of t_q,

$$t_q \leq \min\left(q^{\alpha-(d+1)}, 1\right). \tag{2.13}$$

To obtain the scaling of the entire state transfer process, a sum over t_q is made up to $q = L$. For $\alpha < d$, t_q grows more slowly than q^{-1}, so the sum converges to a constant for asymptotic q. The convergence signals that a state can be transferred any desired distance in a constant time. For $\alpha = d$, $t_q = q^{-1}$, so the sum scales logarithmically in L. For $d < \alpha < d+1$, we obtain a polynomial scaling $L^{\alpha-d}$. Finally, for $\alpha \geq d+1$, the constant lower bound on t_q dominates, and state transfer takes a time proportional to L, just as it does for short-range interacting systems. These scalings are illustrated in Fig. 2.3 along with the exponents of polynomial fits to the numerical solutions of

Eq. (2.10). The time cost of our protocol compares very favorably to the direct use of the long-range interaction, which can create a maximally entangled state in time that scales like L^α. Note that although Hamiltonians turn on and off throughout our protocol, our Hamiltonians always obey the condition that $|h_{ij}| \leq r_{ij}^{-\alpha}$, meaning that the process as a whole obeys the conditions assumed in previous work on speed limits in long-range interacting systems such as Ref. [19].

2.3 Constructing a MERA

We now demonstrate that our state transfer protocol allows for fast construction of a MERA.

In this context, we will interpret a MERA as a quantum circuit for qubits which acts on successively larger length scales, as shown in Fig. 2.4, to produce an entangled state from a product state. More general constructions are possible (e.g. with qudits). Our protocol will also apply to a branching MERA [67] provided that after a constant number of layers the circuit disentangles a constant fraction of the remaining qubits to $|0\rangle$. This condition ensures that there are always sufficient "empty" qubits for our state transfer protocol to scale properly.

A MERA consists of two alternating types of unitary operations and is easiest to understand in reverse (starting at the bottom of the circuit). The first type of unitary, called a disentangler, removes entanglement at the current length scale. The next operation, an isometry, maps a group of ϕ sites into a single site, leaving the other qubits in the state $|0\rangle$. These operations can be repeated, except that now all

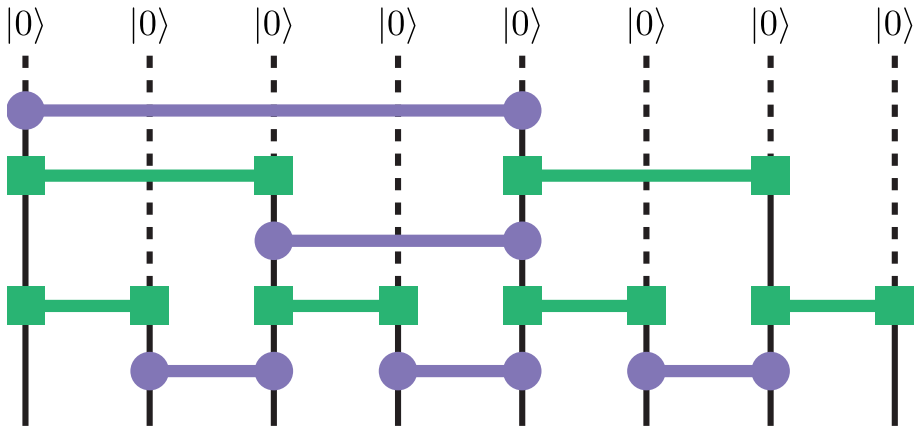

Figure 2.4: Sketch of a MERA circuit, with the disentanglers (purple, circle-capped) and isometries (green, square-capped). All qubits begin in the state $|0\rangle$, indicated by a dashed line. At each length scale, entanglement is created or removed to create a many-body entangled state from a product state after $\log_2 L$ steps. Although we have drawn a simple 1D binary MERA, our protocol can be extended to higher dimensions and more complicated tensor structures.

unitary gates need to be performed over a distance ϕ times larger than previously.

It is clear that MERA produces a circuit with depth $\log_\phi L$, but this apparent logarithmic scaling masks an actual time cost due to the continuously increasing length scale. However, we can replace a long-range two-qubit unitary with state transfer followed by a short-range unitary. This framework allows us to ignore any details of the two-qubit unitary and simply use state transfer as a primitive. The structure of a MERA circuit guarantees that the $|0\rangle$ states required to perform state transfer will be present between any two qubits when we need to perform a unitary on them.

Suppose that t_τ is the maximum time required to perform a two-qubit gate across a distance ℓ_τ at the τth step of the MERA circuit. We can perform all the MERA operations at a given step in parallel, so a single layer of the MERA simply

requires time $2t_\tau$ for the disentanglers and then isometries. The time to perform the entire MERA circuit will then be bounded (up to a constant factor) by

$$t_{\text{MERA}} \lesssim \sum_{\tau=0}^{S-1} t_\tau. \tag{2.14}$$

Here $S = \log_\phi L$. Our state transfer procedure allows for $t_\tau = 2t_{\text{transfer}}$. The time required to perform the final two-qubit gate does not affect the scaling and so is omitted. We can then bound t_{transfer} by considering the length scale at each step, $\ell_\tau = \phi^\tau$. If $\alpha = d$, t_{transfer} scales as $\log_\phi \ell_\tau$ (as in our state transfer bound but with a constant multiple changing the base of the logarithm), and t_{MERA} will be bounded by $\sim \left(\log_\phi L\right)^2$ by considering the largest term in Eq. (2.14) multiplied by the number of terms. For $\alpha \neq d$, t_{transfer} scales polynomially in ℓ_τ with exponent β,

$$t_{\text{MERA}} \lesssim \sum_{\tau=0}^{S-1} \ell_\tau^\beta. \tag{2.15}$$

For $\alpha < d$, $\beta = 0$ and the sum is proportional to $\log_\phi L$. For $\alpha > d$, $\beta = \max(\alpha - d, 1)$. We use $\ell_\tau = \phi^\tau$ and carry out the geometric sum to obtain

$$t_{\text{MERA}} \lesssim \left(\phi^\beta\right)^S = L^\beta. \tag{2.16}$$

Thus we have

$$t_{\text{MERA}} \lesssim \begin{cases} \log_\phi L & \alpha < d \\ \log_\phi^2 L & \alpha = d \\ L^{\alpha-d} & d < \alpha \leq d+1 \\ L & \alpha > d+1. \end{cases} \tag{2.17}$$

2.4 Outlook

We have demonstrated fast state transfer and MERA construction protocols using long-range interactions. Our protocol's exponential speedup for $\alpha = d$ nearly saturates the bound in [17], which gives a logarithmic lightcone for $\alpha > d$. However, we have not shown that our method is the fastest state transfer protocol possible. Such a result would require demonstrating a general Lieb-Robinson-type bound which we would then saturate. Instead, our protocol limits future Lieb-Robinson bounds for long-range systems. The state transfer protocol we have presented establishes that no finite causal region is possible for $\alpha < d$, since a constant amount of time suffices to establish any desired correlation at arbitrary distances. In previous work, causal regions were seen in systems with $d/2 < \alpha \leq d$ as long as the initial state was not entangled [38]. Like our work, Ref. [38] also uses multiple qubits with long-range interactions to reduce state transfer time. We have shown that such causal regions do not persist in general, although it is possible that this violation requires the use of time-dependent Hamiltonians as opposed to the time-independent Hamiltonians in Ref. [38].

For the intermediate value $d < \alpha < d + 1$, our protocol shows that no linear light cone can be drawn, although a polynomial bound may be possible. These results should be compared to Ref. [19], which established a polynomial light cone only for $\alpha > 2D$ that becomes linear only in the limit of $\alpha \to \infty$. Our protocol's linear scaling when $\alpha \geq d+1$ suggests that the tightest possible Lieb-Robinson bound may also possess a critical α with a similar property. Resolving this question could reveal

important facts about the nature of correlations in long-range interacting systems.

It is our hope that this protocol, or a minor variation thereof, could soon be realized experimentally. Such a realization could offer significant technological advantages in, for instance, entanglement-enhanced metrology. In Sec. 2.5, we show how dipole-dipole interactions in three dimensions can be used to implement a variant of our protocol with a focus on Rydberg atoms. Using this protocol, qubits can be entangled exponentially faster than using short-range interactions. In the future, we hope to reduce the local control required to achieve sublinear scaling.

2.5 Application to Dipole-Dipole Interactions

In this section, we show that it is possible to realize a protocol with similar scaling to the one proposed by using Rydberg atoms. Rydberg atoms can be made to interact with a dipole-dipole interaction that has distance dependence $1/r^3$. This suggests that, using our protocol, we could produce a cube of side length L in a GHZ state in time proportional to $\log L$. We will demonstrate that a realistic physical interaction can yield this result. Many details on Rydberg atoms and their applications in quantum information can be found in Refs. [34, 68, 69], and experimental demonstrations can be found in Refs. [35, 70–72]. Our analysis is focused on Rydberg atoms, but much of it should extend to other dipolar systems, such as polar molecules, with appropriate modification of implementation details [31–33, 73].

We select as qubit states the ground state and a highly excited state of a

Rydberg atom under a weak electric field, yielding a purely diagonal atomic interaction [34]. The Hamiltonian of a system of such atoms can be written as:

$$H_{\text{int}} = \sum_{i \neq j} H_{ij} = \sum_{i \neq j} \frac{1 - 3\cos^2 \theta_{ij}}{r_{ij}^3} Z_i Z_j \equiv \sum_{i \neq j} V_{ij} Z_i Z_j. \tag{2.18}$$

Here, r_{ij} is the distance between atoms i and j, while θ_{ij} is the angle between the electric field and the vector separating the two atoms. We have ignored local terms like Z_i and Z_j, which can be removed by applying local rotations. By applying local rotations, this ZZ Hamiltonian can be used to realize CNOT interactions, regardless of whether the overall sign is positive or negative. This is done by applying local rotations to produce a controlled-phase gate and applying Hadamard operations on the target before and after the evolution to yield a controlled-NOT gate [74]. We assume that, while local control fields may be time-dependent, the two-body interaction in Eq. (2.18) is active throughout the entire state transfer process. The individual addressing required to perform these local operations was demonstrated in a 3D optical lattice in Ref. [75]. The roughly 5 μm lattice spacing in that work is also an appropriate spacing for the Rydberg interactions we intend to use in our protocol, as it helps to prevent the dipole-dipole interactions from becoming comparable to the energy level spacing.

To apply the protocol of this chapter, qubits must be separated into controls and targets. Such separation can be performed using an echoing procedure: first, qubits evolve under H_{int} and then under $-H_{\text{int}}$ for an equal amount of time. However, halfway through the second evolution, a π-pulse (X gate) is applied to either all target qubits or all control qubits. This has the effect of swapping Z for

$-Z$. All interactions between controls and controls, or targets and targets, will remain unchanged, but any control-target interactions will be inverted. Thus, during the $-H_{\text{int}}$ time, control-target interactions experience no net evolution, while any control-control or target-target pair evolution due to $+H_{\text{int}}$ is undone. The $-H_{\text{int}}$ evolution time is equal to the initial entangling $+H_{\text{int}}$ time, so the echoing procedure does not change the scaling with L. Even if the negative interaction is not of the same magnitude as the original, we can still accomplish the echoing by adjusting the timescales, and the scaling with L will still not be changed.

To change the sign of the dipole-dipole interaction, realizing $-H_{\text{int}}$, we can encode the computational states into the fine structure of a Rydberg atom. For specificity, we consider the case of Rb^{87} with a weak applied electric field. Ignoring the hyperfine structure, we encode the state $|0\rangle$ in a superposition of $|L = 0, J = 1/2, m_J = 1/2\rangle$ and $|L = 1, J = 3/2, m_J = 3/2\rangle$ created by applying a microwave dressing field, with most of the amplitude being stored in the latter state. The state $|1\rangle$ is then encoded in $|L = 1, J = 1/2, m_J = 1/2\rangle$. All three states have the same principal quantum number. Details can be found in an analogous scheme for polar molecules presented in entry No. 5 of Table II and Fig 3(d) of Ref. [32]. Note that here we are also dropping local Z terms which can be canceled by a local rotation. We have calculated dipole matrix elements for Rb^{87} across a wide range of principal quantum numbers that confirm this scheme remains viable in the Rydberg setting. We also assume that, in addition to changing the overall sign of the interaction, we are able to place qubits in non-interacting electronic ground states to avoid any unwanted interactions or decay from excited states.

If a volume of control qubits exists, this volume will convert a qubit j into a control after time $\pi/2V_j$, where V_j is the sum over all interaction constants V_{ij} for control qubits i. Suppose that enough qubits have been added that the sum of point-to-point interactions is well-approximated by an integral, which is a good approximation in the relevant asymptotic regime. The total interaction on a qubit j in this case can be written as

$$V_j = \int_C V_{ij} \mathrm{d}C. \tag{2.19}$$

Here, C is the volume of control qubits. This quantity has the useful property of scale invariance. If all lengths change by a factor λ, then H_{int} changes by the factor λ^{-3} due to its distance dependence. However, the region of integration expands by λ^3, so the final quantity remains unchanged.

We consider expanding a cube of controls, increasing the side length ℓ by a constant factor λ. After this procedure, we obtain a new cube of side length $\lambda\ell$. Qubits outside of the larger cube have no operations performed on them. Once this expansion has been performed, we expand the cube again. Due to scale invariance, the same operation can be performed in identical time. This means that after n expansion steps, the side length will be $\lambda^n\ell$. Therefore, we can construct a cube of side length L in a time proportional to $\log_\lambda(L/\ell)$ as indicated earlier. The scaling properties of the integral in Eq. (2.19) can be used in cases where $\alpha \neq d$ as well. Equation (2.19) implies that the time required to construct a cube of side length L will be:

$$t_{\mathrm{GHZ}} \sim \sum_{i=1}^{log_\lambda(L/\ell)} \lambda^{n(\alpha-d)}. \tag{2.20}$$

For $\alpha < d$, this saturates to a number independent of L, and for $\alpha > d$, it implies that $t_{\mathrm{GHZ}} \sim L^{\alpha-d}$. Note that for $\alpha > d+1$, a protocol of successive dilations of the cube fails to provide optimal scaling.

All that remains to be shown is that the size of the cube can be increased by a constant factor in finite time. This is not guaranteed because the dipole-dipole interaction changes sign as a function of θ_{ij}, causing V_j to be zero for qubits at some points. If we could only act with the control cube during the expansion time, we would not be able to perform the expansion as outlined above. However, we can use a slightly more complicated scheme in which some intermediate qubits are used. Rather than expand the entire cube at once, we expand the cube outward in the positive x-, y-, and z-directions successively, each time expanding only to qubits which lie on lines perpendicular to the expanding face of the rectangular prism, as illustrated in Fig. 2.5. This works because the interaction can be shown to decrease monotonically (in absolute value) along Cartesian directions, as we prove below. Since at long distances we know that the interaction decays to zero and has the same sign for all target qubits, the monotonicity establishes that there is no zero crossing. As there is no zero crossing, there will be a finite time that suffices to complete the expansion. The logarithmic scaling follows.

2.6 Proof of Interaction Monotonicity

We will now prove that the interaction between a cube of controls and a target qubit decreases monotonically in Cartesian directions. Suppose we begin with a

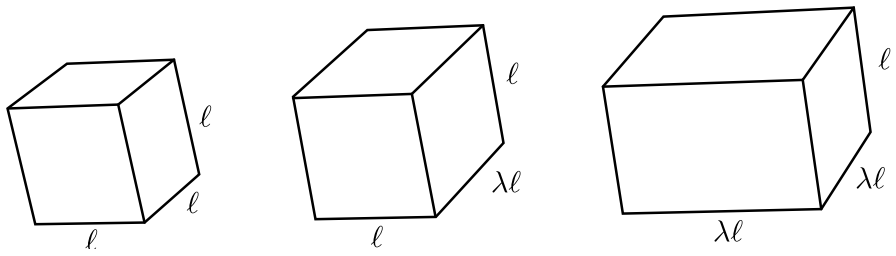

Figure 2.5: Successive transformations of the control cube. A cube of side length ℓ is expanded first in one direction, then the next. After the final step (not shown), the result will be a cube of side length $\lambda\ell$.

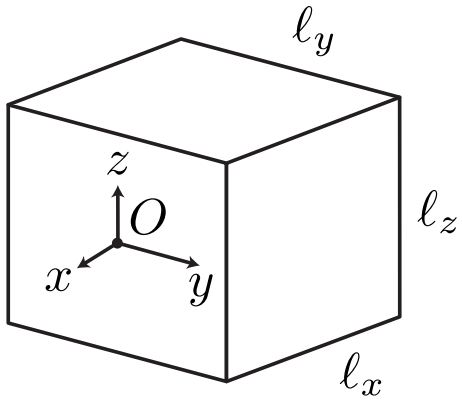

Figure 2.6: Illustration of the coordinate system used in this section.

rectangular prism located in the $y-z$ plane with dimensions $\ell_x \times \ell_y \times \ell_z$ and the origin in the center of one face (see Fig. 2.6 for an illustration). A qubit at point (x, y, z) then has the interaction integral

$$V = \int_{x}^{x+\ell_x} \int_{-\ell_y/2+y}^{\ell_y/2+y} \int_{-\ell_z/2+z}^{\ell_z/2+z} \frac{x'^2 + y'^2 - 2z'^2}{(x'^2 + y'^2 + z'^2)^{5/2}} \mathrm{d}x' \mathrm{d}y' \mathrm{d}z'. \tag{2.21}$$

The integrand in Eq. (2.21) is simply the dipole interaction written in Cartesian coordinates. We choose y and z to fall in $(-\ell_y/2, \ell_y/2)$ and $(-\ell_z/2, \ell_z/2)$ respectively to ensure that their projection to the $y-z$ plane lies on the face of the prism. We consider only positive values of y and z without loss of generality. The derivative of

V with respect to x can be expressed analytically as

$$\partial_x V = D\left(-\frac{\ell_y}{2}+y,-\frac{\ell_z}{2}+z\right)+D\left(\frac{\ell_y}{2}+y,\frac{\ell_z}{2}+z\right) - \left[D\left(-\frac{\ell_y}{2}+y,\frac{\ell_z}{2}+z\right)+D\left(\frac{\ell_y}{2}+y,-\frac{\ell_z}{2}+z\right)\right], \tag{2.22}$$

$$D(a,b) = ab\left(\frac{1}{\left((x+\ell_x)^2+c^2\right)\sqrt{(x+\ell_x)^2+a^2+c^2}}-\frac{1}{(x^2+c^2)\sqrt{(x^2+a^2+c^2)}}\right). \tag{2.23}$$

For $D(a,b)$, the sign is always determined by the prefactor because the factor in parentheses is strictly negative. Using the fact that y and z must be less than $\ell_y/2$ and $\ell_z/2$ respectively, we can assign a negative sign to the first two D to appear in Eq. (2.22) and a positive sign to the second two. Therefore, we find that $\partial_x V$ is always negative in this region, establishing the monotonicity for expansion along one face in the x-direction. This proof also holds for the y-direction immediately from symmetry. For the z-direction, a similar argument holds but with a more complicated parenthetical term in $D(a,b)$.

2.7 Effects of Decoherence

In the next two sections, we will consider the influence of experimental imperfections in qubits and gate operations and examine the implication for our protocols scalability. First, we will consider the influence of decoherence, for instance, due to spontaneous emission out of the Rydberg excited states. The fragile nature of the GHZ state means that a single emission can cause our protocol to fail. We assume that individual qubits fail (spontaneously emit) at a rate γ. This analysis should

extend to any similar failure mechanism that occurs at a constant rate. If each expansion step (dilating the cube by λ) takes time δt, then we can consider whether, in the ith timestep, any of the λ^{3i} qubits currently involved emit. If not, we label the step a success. The protocol succeeds if all of its individual steps succeed. The probability that no spontaneous emissions occur at any of N_t time steps and that the protocol succeeds is

$$P(\text{success}) = \prod_{i=1}^{i=N_t} P(\text{success at step } i) = e^{-\gamma \delta t \sum_i \lambda^{3i}}. \tag{2.24}$$

If we demand that the protocol successfully entangle N qubits with a probability $P > \epsilon$, then Eq. (2.24) becomes

$$\sum_{i=1}^{\log_\lambda N^{1/3}} \lambda^{3i} = \frac{\lambda^3 (N-1)}{\lambda^3 - 1} < \frac{\ln \frac{1}{\epsilon}}{\gamma \delta t}. \tag{2.25}$$

This suggests a limit on the number of qubits which can be entangled with a system of decohering qubits, which we write as

$$N_{\text{lr}} < 1 + \frac{\ln \frac{1}{\epsilon}}{\gamma \delta t} \frac{\lambda^3 - 1}{\lambda^3}. \tag{2.26}$$

Here N_{lr} refers to the number of qubits that can be entangled using our long-range interacting protocol. Note that if ϵ and λ are taken to be of order 1, Eq. (2.26) simply implies that $N_{\text{lr}} \gamma \delta_t \lesssim 1$, which is unsurprising since our largest entangled state decays in a time $1/N_{\text{lr}}\gamma$. We can also consider what this limit looks like in the case of a protocol which uses nearest-neighbor interactions and, at each step, increases the cube's side length by one. In this case, the ith timestep has i^3 qubits entangled, and there are $N^{1/3}$ such steps. A similar argument to the above leads us

to calculate

$$\sum_{i}^{N^{1/3}} i^3 = \frac{1}{4}\left[N^{4/3} + 2N + N^{2/3}\right] < \frac{\ln\frac{1}{\epsilon}}{\gamma\delta t}. \tag{2.27}$$

If we assume we're interested in cases where N is somewhat large *a priori*, then we write the following loose bound by dropping strictly positive terms:

$$N_{\text{nn}} < \left(\frac{4\ln\frac{1}{\epsilon}}{\gamma\delta t}\right)^{3/4}. \tag{2.28}$$

Here the exponent 3/4 arises because we summed over $N^{1/3}$ terms like i^3, yielding $N^{4/3}$ and then inverted that. Suppose we take $\lambda = 2$, in which case the first step of each protocol is the same and we can equate the two δt. Then the ratio of the two threshholds is

$$\frac{N_{\text{lr}}}{N_{\text{nn}}} = \frac{7}{16\sqrt{2}}\left(\frac{\ln\frac{1}{\epsilon}}{\gamma\delta t}\right)^{1/4}. \tag{2.29}$$

To evaluate this figure of merit, we can look at the original proposal for interaction-based Rydberg gates, which suggests a two-qubit gate timescale of less than a nanosecond [68]. Our protocol also requires several one qubit gates in each step, which can also be accomplished on nanosecond timescales using pulsed lasers [76]. Estimating $\delta t \sim 5$ ns, demanding a success probability of 1/2, and taking the Rb^{87} 100s state lifetime of 340 μs at a temperature of 300K [34], we find that $N_{\text{lr}}/N_{\text{nn}} \approx 4.5$, meaning that a long-range protocol can achieve a maximally entangled state containing nearly 4.5 times as many qubits as one constructed by nearest-neighbor interactions. This figure rises to 4.9 if we solve Eq. (2.27) directly rather than using the bound. N_{lr} is about 4×10^4, suggesting a lifetime for the GHZ state of roughly 8 ns. Using δt and $N_t = \log_\lambda N_{\text{lr}}^{1/3}$, we find that constructing such a state would require a total time of about 25 ns.

To estimate the probability of performing state transfer instead of constructing the GHZ state, one must simply replace ϵ with $\sqrt{\epsilon}$ in the above analysis, as a state transfer success is effectively just two successful iterations of the GHZ construction. After state transfer is performed, we can ask whether it survives long enough to be read out or transferred into a non-interacting level. Since the single-atom lifetime of the Rydberg state is 340 μs, this should not be an issue as the time required to complete the transfer is on the order of tens of nanoseconds. Once transfer or GHZ creation is complete, the electric field can be turned off to remove the dipole-dipole interaction in Eq. (2.18).

2.8 Effects of Imperfect Single-qubit Gates

In addition to free evolution under the long-range interaction Hamiltonian [Eq. (2.18)], our protocol requires a number of single-qubit gates to be performed. These can be Hadamard gates which produce the CNOT operation out of our ZZ interaction or the echoing pulses. In any case, a failure of the single-qubit gate can pose a serious problem to the protocol. Suppose we perform N_s single-qubit gates which succeed with a probability P. Then, as in the previous section, we demand that the gate sequence succeed with probability ϵ, obtaining

$$P^{N_s} > \epsilon \implies P > e^{(\ln \epsilon)/N_s}. \tag{2.30}$$

The number of single qubit gates which must be targeted on a qubit in a timestep varies depending on that qubit's role during the step, but let us suppose that on average there are c gates per qubit performed on each of N_t timesteps. We

can count the number of qubits involved in each timestep just as we did in Eq. (2.25) to obtain a criterion for success:

$$P > e^{(\ln \epsilon)/(c\lambda^3(N-1)/(\lambda^3-1))}. \tag{2.31}$$

Theoretical work on composite pulse sequences for atomic qubits suggests achievable fidelities of $1 - 10^{-4}$ [77]. If we assume $c = 4$ as an estimate, $\epsilon = 1/2$, and $\lambda = 2$ as in the last section, Eq. (2.31) suggests that roughly 1500 qubits could be entangled with such gates using our protocol. This is a reduction of several orders of magnitude from the previous section which considered no single-qubit fidelity issues, a limitation which highlights the fact that a version of the protocol requiring less single-qubit control could perhaps entangle dramatically more qubits.

Chapter 3: Unitary Entanglement Construction in Hierarchical Networks

3.1 Introduction

As quantum computers grow from the small, few-qubit machines currently deployed to the large machines required to realize useful, fault-tolerant computations, it will become increasingly difficult for every physical qubit to be part of a single contiguous piece of hardware. Just as modern classical computers do not rely on a single unit of processing and memory, instead using various components such as CPUs, GPUs, and RAM, we expect that a quantum computer will likewise use specialized modules to perform different functions. At a higher level, computers can be organized into clusters, data centers, and cloud services which allow for a distributed approach to computational tasks, another paradigm quantum computers will no doubt emulate. Already, there has been significant interest in how quantum algorithms for elementary operations such as arithmetic perform in distributed-memory situations [78, 79] and how to automate the design of quantum computer architectures [80]. In addition, the construction of a fault-tolerant quantum computer naturally suggests a separation of physical qubits into groups corresponding to

logical qubits, which makes modularity an attractive framework for building fault-tolerant computers [81]. Modular and scalable computing architectures have been explored for both ion trap [82, 83] and superconducting platforms [84–86].

In this chapter, we use tools from graph theory to discuss benefits and drawbacks of different potential architectures for a modular quantum computer. A graph-theoretic approach allows us to flexibly examine a wide range of possible arrangements quantitatively and allows for convenient numerical simulation using existing software packages designed for network analysis [87]. We especially wish to focus on families of graphs that can scale with the desired number of qubits. In general, we assume that connectivity, i.e., being able to quickly perform operations between nodes, is desirable in an architecture, but that building additional graph edges is in some way costly or difficult, and so will try to minimize the number of needed edges to achieve a highly communicative graph.

We will make use of a previously described graph-theoretic binary operation known as the hierarchical product [88, 89]. We will use this iteratively to describe a new family of graphs we dub "hierarchies." We will show that hierarchies perform well by many commonsense graph metrics and argue that they would serve as a plausible and effecient basis for a quantum computing architecture. Furthermore, we will demonstrate that these graphs allow for easily-implemented heuristic procedures to assist in the compilation of quantum algorithms.

We will examine the performance of graphs in generating large entangled states such as the multi-qubit Greenberger-Horne-Zeilinger (GHZ) state (also known as a cat state). The GHZ state has perfect quantum correlations between different qubits;

it thus can be used to perform high-precision metrology [23, 46]. In addition, the creation of a GHZ state can be used as part of a state-transfer protocol, which may be useful as part of large quantum computations [24].

An additional property of GHZ state preparation and state transfer which makes them a useful starting point is that, in nearest-neighbor connected systems, performing these tasks using unitary processes from an initial product state is limited by the Lieb-Robinson bound [16, 90]. It takes a time proportional to the distance between two points to establish maximal quantum correlation between them. By examining these tasks on a range of different graphs, we hope to understand how the graph structure can affect the limitations on quantum processes caused by locality considerations. Prior work has characterized the difficulty of creating graph states [91], but preparation of such states is not limited by Lieb-Robinson considerations.

Our work in this chapter should be contrasted with work on entanglement percolation [92,93]. Entanglement percolation describes the process of using low-quality entanglement between adjacent nodes on a graph to create one unit of long-range, high-quality entanglement (e.g., a Bell pair). The use of entanglement percolation to prepare large cluster states on a lattice was considered in Ref. [94]. The nature of entanglement growth in complex networks was considered in Refs. [95, 96], showing that so-called "scale-free" networks are particularly easy to produce large entangled states in. We are interested in the overall capability of different graph structures to perform large computations and in the use of graph eigenvalue methods to understand the spread of quantum information [97]. GHZ state preparation and state transfer are just two possible benchmark tasks, and it is possible that other tasks

would result in different evaluations of relative performance between graphs.

Our work should also be considered in the context of classical network theory, where much is known about complicated graph structures [98–100]. It remains to be seen to what degree classical network theory can be easily exported to the quantum domain. Quantum effects such as the no-cloning theorem may limit our ability to distribute information, or conversely we can take advantage of teleportation by distributing quantum bandwidth in anticipation of it actually being needed. As further examples of how quantum and classical networks differ, it has been shown that entanglement swapping may be used to permit quantum networks to reshape themselves into interesting and useful topologies [101]. It has also been shown that, in general, the optimal strategy for entanglement generation in quantum networks can be difficult to calculate because many aspects of classical control theory do not apply [102].

The structure of this chapter is as follows. In Sec. 3.2, we will introduce a binary operation on graphs known as the hierarchical product, describe how it can be used to produce families of graphs we call hierarchies, and discuss the properties of these hierarchies. In Sec. 3.3, we will compare hierarchies to other families of graphs, examining how certain graph-theoretic quantities scale with the total number of included qubits. Readers who are not interested in graph theoretic details may wish to skip much of these first two sections. In Sec. 3.4, we will use analytic and numerical methods to examine how long is required to construct GHZ states spanning our graphs or to transfer states across them, using Lieb-Robinson bounds to connect graph-theoretic quantities to bounds on quantum computing performance. Finally,

in Sec. 3.5, we will show how the unique structure of hierarchies allows for simple heuristics to map qubits in an algorithm into physical locations in hardware.

3.2 Hierarchical Products of Graphs

3.2.1 Background and Notation

One of the defining features of modularity in a network is the presence of clusters of nodes that are well-connected. Qualitatively, a modular network can be partitioned into such node clusters, or *modules*, that have a sparse interconnectivity. In quantum networking, it is believed that fully connected architectures will suffer greatly decreasing performance or increasing costs as the number of nodes becomes larger, and this motivates the search for alternative network designs. For instance, Ref. [21] estimates that a single module of trapped-ion qubits will likely contain no more than 10 to 100 ions, noting that the speed at which gates are possible becomes slower as the module is expanded. On the network scale, we might imagine a network of nodes over longer distances connected by quantum repeaters [103]. In such a network, establishing direct links between every possible pair of N nodes would require $\Theta(N^2)$ sets of quantum repeaters, a prohibitive cost as N becomes large.

The state of the art in quantum technologies, such as ion traps and superconducting qubits, is the ability to control a small number ($\approx 10 - 100$) of physical qubits using certain fixed sets of one- and two-qubit operations. Instead of increasing the size of these modules, one could instead build a network out of many small

modules that are connected at a higher level in a sparse way, perhaps by optical communication links [21].

Our first goal will be to describe modular architectures in the language of graph theory. This will then allow us to quantify and compare their connectivity properties against other network designs, notably the nearest-neighbor grid architecture.

Our detour into graph theory in this chapter serves two purposes. First, it will allow to develop a rigorous way to construct families of graphs which we believe are promising quantum computing architectures. Second, we will later (beginning in Sec. 3.4) use these graph properties to connect directly to physical bounds on the generation of states with long-range quantum correlations; phrasing the properties of quantum architectures as graphs allows us to make a direct application of the Lieb-Robinson bound to these cases.

An unweighted graph $G = (V, E)$ is conventionally specified by a set of vertices V, and a set of edges between the vertices E, where an edge between distinct vertices i and j will be denoted by the pair (i, j). In this chapter, we use the terms "vertex" and "node" synonymously. The *order* of a graph is the total number of vertices in the graph, $|V|$. It will be useful for the purposes of this chapter to work with *weighted* graphs, where we specify a weight $w_{ij} \in \mathbb{R}$ for each pair of vertices $(i, j) \in V \times V$. Two vertices i and j are said to be *disconnected* if $w_{ij} = 0$, and connected by an edge with weight $w_{ij} \neq 0$ otherwise. Thus, unweighted graphs may be thought of as graphs with unit weight on every edge.

Finally, the graphs we consider here will be *simple*, meaning:

- The edges have no notion of direction. In other words, $w_{ij} = w_{ji}$ for all $i, j \in V$.
- There are no self-edges, i.e., $w_{ii} = 0$ for all $i \in V$.
- Any two vertices have at most one edge between them.

Henceforth, graphs will be simple and weighted, unless otherwise specified.

The information contained in a graph can be represented as a matrix known as the *adjacency matrix*, whose rows and columns are labeled by the vertices in V and whose entries hold edge weights. Thus, the adjacency matrix is an $n \times n$ matrix where $|V| = n$. The adjacency matrix A_G (or simply A for shorthand) for a graph G is given by

$$A_{ij} = \begin{cases} 0, & \text{if } i = j, \\ w_{ij}, & \text{if } i \neq j. \end{cases} \tag{3.1}$$

An important measure of local connectivity is given by the *valency* v_i of a node i, with $v_i = \sum_{j=1}^{n} w_{ij}$. For unweighted graphs, the valency of any node is simply the number of edges incident at that node, otherwise known as the *degree* of the node. We will also define the graph diameter, $\delta(G)$, as the maximization of the shortest distance between two nodes on the graph over all pairs of nodes.

Graphs may also be described by the *Laplacian*. The algebraic Laplacian L is given by

$$L_{ij} = \begin{cases} v_i, & \text{if } i = j, \\ -w_{ij}, & \text{if } i \neq j. \end{cases} \tag{3.2}$$

The algebraic Laplacian is closely related to the adjacency matrix, since we may write $L = \Delta - A$, where $\Delta = \mathrm{diag}\,(v_1, \ldots, v_n)$ is the diagonal matrix of vertex valencies. The eigenvalues of the algebraic Laplacian give us bounds on various graph properties, as discussed further in Sec. 3.2.2.4.

Finally, we remark that the algebraic Laplacian should not be confused with the normalized Laplacian $\mathcal{L} = \Delta^{-\frac{1}{2}} L \Delta^{-\frac{1}{2}}$, which is frequently seen in the network theory literature. The algebraic properties discussed in the next section (such as associativity of the hierarchical product) apply to the adjacency matrix as well as the algebraic Laplacian, but not to the normalized Laplacian.

3.2.2 Hierarchical Product

Here, we will define the hierarchical product and illustrate it with simple examples. For a fuller exposition, see Ref. [88], where the hierarchical product of graphs was introduced. Note that, in some contexts, the hierarchical product is also known as the rooted product [89].

Given a graph G, let $\mathbb{1}_G$ denote the identity matrix on $n = |V|$ vertices. We will denote by D_G an $n \times n$ diagonal matrix with 1 as the first entry and zero everywhere else. Note that there is no natural notion of order to graph vertices, so the choice of "first" vertex must be specified explicitly. Graphs with such a specified

first vertex are called *rooted graphs* [104]. We write these matrices as

$$\mathbb{1} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & & & & \\ & 0 & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}. \tag{3.3}$$

Definition 3.2.1. Given graphs G and H, the *hierarchical product* $P = G \,\Pi\, H$ is the graph on vertices $V_P = V_G \times V_H$ and edges $E_P \subseteq V_P \times V_P$ specified by the adjacency matrix

$$A_P = A_G \otimes D_H + \mathbb{1}_G \otimes A_H, \tag{3.4}$$

or, equivalently, by the algebraic Laplacian

$$L_P = L_G \otimes D_H + \mathbb{1}_G \otimes L_H. \tag{3.5}$$

We will often use the shorthand $A_P = A_G \,\Pi\, A_H$ and $L_P = L_G \,\Pi\, L_H$.

If G and H are graphs, then $G \Pi H$ may be thought of as one copy of G with $|G|$ copies of H, each attached to a different vertex of G (see Fig. 3.1). Thus, $G \,\Pi\, H$ is a graph which has $|G|$ *modules* of $|H|$ nodes each. The modules' internal connectivity is described by H, and the modules are connected to one another in a manner described by G. The hierarchical product formalism therefore naturally produces modular graphs. Its main advantage comes from the convenience of working with the algebra at the level of adjacency matrices and Laplacians, which in turn makes the computation of important properties of such graphs straightforward.

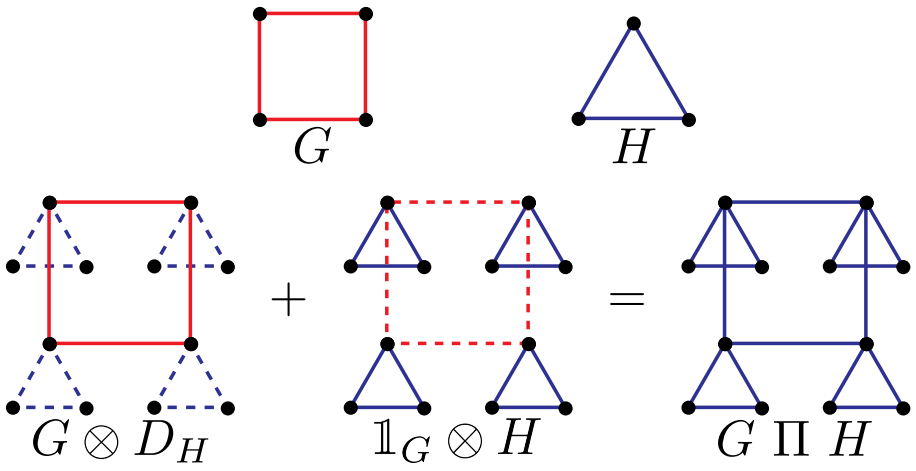

Figure 3.1: A simple example of the hierarchical product $G \,\Pi\, H$ between the cycle graphs $G = C_4$ and $H = C_3$. The first term in Eq. (3.4), $A_G \otimes D_H$, creates one copy of G on the vertex set formed by the first vertices of each H copy, while the second term $\mathbb{1}_G \otimes A_H$ creates the four copies of H.

We now present some properties of the hierarchical product which make it an attractive formalism for practical applications in quantum networking.

3.2.2.1 Structural Properties

At the level of adjacency matrices, the hierarchical product is *associative*. Let A, B, C be three adjacency matrices. Then,

$$(A \,\Pi\, B) \,\Pi\, C = A \,\Pi\, (B \,\Pi\, C)\,. \tag{3.6}$$

For a proof, we refer the reader to Ref. [88].

Associativity implies that a product of multiple graphs does not depend on the order of evaluation. Therefore, we can unambiguously take the hierarchical product over many graphs to produce a graph of the form $G_k \,\Pi\, G_{k-1} \,\Pi \cdots \Pi\, G_1$. We will refer to such graphs as *hierarchies*, and the i-th graph in the product G_i as the

i-th level of the hierarchy, enumerated from the bottom level upwards (symbolically, from right to left). In particular, if all G_i are equal to some graph G, then we write

$$G^{\Pi k} := \underbrace{G \,\Pi \cdots G \,\Pi}_{k-1 \text{ times}} G. \tag{3.7}$$

and refer to $G^{\Pi k}$ as a depth-k (or k-level) hierarchy.

Note that the hierarchical product *does not* satisfy many properties which are commonly assumed for operations on matrices. In particular,

1. Bilinearity: $(A_1 + A_2) \,\Pi\, B = A_1 \otimes D_B + A_2 \otimes D_B + \mathbb{1}_{(A_1+A_2)} \otimes B \neq A_1 \,\Pi\, B + A_2 \,\Pi\, B$. Similarly, $A \,\Pi\, (B_1 + B_2) \neq A \,\Pi\, B_1 + A \,\Pi\, B_2$.

2. Scalar multiplication: For any scalar α, $(\alpha A) \,\Pi\, B = \alpha A \otimes D_B + \mathbb{1}_A \otimes B \neq \alpha\,(A \,\Pi\, B) \neq A \Pi\,(\alpha B)$. Note however that scalar multiplication is distributive in the following way: $\alpha\,(A \,\Pi\, B) = (\alpha A) \,\Pi\, (\alpha B)$.

Hierarchical graphs are also instances of hyperbolic graphs. The Gromov-hyperbolicity [105], which measures curvature and is small for a graph with large negative curvature, is only a constant for hierarchical graphs. Since the hyperbolicity in general is at most half the graph diameter, whereas in this case it is independent of the diameter, it is termed *constantly hyperbolic* in the parlance of Ref. [106]. Hyperbolic graphs are seen in several real-world complex networks [107, 108], most notably the internet [109, 110]. Hyperbolic lattices have also been realized recently in superconducting circuits [111].

Finally, hierarchies have low tree-, clique- and rank-widths, which are each measures of the decomposibility of a graph [112]. These structural properties imply

efficient algorithms for optimization problems expressible in monadic second-order (MSO) logic – a class which, for arbitrary graphs, includes several NP-hard problems. This feature could potentially be used to solve circuit layout and optimization problems on modular architectures without resorting to heuristics. We refer the reader to Ref. [113] for details on these structural results.

3.2.2.2 Scalability

So far we have discussed hierarchies in which the edges in different levels of the hierarchy are equally weighted. However, one useful generalization would be to allow the weight of edges at each layer of the hierarchy to vary. The meaning of this weight could vary depending on the context. In some cases, weights can be used to quantify the costs of an edge (*cost weight*). In others, we may wish to use weighted edges to quantify the power or performance of a network, interpreting edge weights as the strength of terms in a Hamiltonian or, inversely, the time required to communicate between nodes (*time weight*).

In this chapter, we prefer to remain agnostic to the meaning of the weights as much as is possible. When we calculate graph properties in Sec. 3.3, we will do so without reference to the meaning of the weights. In general, we will allow a graph to assign multiple kinds of weights to its edges, and each type of weight might scale differently. For now, we define a generalization of the hierarchical product which will allow us to construct hierarchies that incorporate different weights at different levels of the hierarchy.

Definition 3.2.2. Given graphs G and H, and $\alpha \in \mathbb{R}_+$, the *α-weighted hierarchical product* $P = G \,\Pi_\alpha\, H$ is a graph on vertices $V_P = V_G \times V_H$ and edges $E_P \subseteq V_P \times V_P$ specified by the adjacency matrix

$$A_P = \alpha A_G \otimes D_H + \mathbb{1}_G \otimes A_H, \tag{3.8}$$

or, equivalently, by the algebraic Laplacian

$$L_P = \alpha L_G \otimes D_H + \mathbb{1}_G \otimes L_H. \tag{3.9}$$

We will often use the shorthand $A_P = A_G \,\Pi_\alpha\, A_H$, and $L_P = L_G \,\Pi_\alpha\, L_H$.

As before, we may construct a k-level, *weighted* hierarchy out of k base graphs $G_1, \ldots, G_k$, and k weights $\alpha_i, \ldots, \alpha_k \equiv \vec{\alpha}$, so that the edges of the i-th level graph G_i are weighted by the i-th component of $\vec{\alpha}$, α_i. The adjacency matrix of such a hierarchy may be written as

$$A^{\Pi_{\vec{\alpha}} k} := \sum_{i=1}^{k} \alpha_i \mathbb{1}_{[i+1\,..\,k]} \otimes A_i \otimes D_{[1\,..\,i-1]}, \tag{3.10}$$

where the subscripts $[a\,..\,b]$ on $\mathbb{1}$ and D are shorthand for the Kronecker product of matrices over all descending indices in the integer interval $[a\,..\,b]$. For instance, $D_{[1\,..\,i-1]} := D_{G_{i-1}} \otimes D_{G_{i-2}} \otimes \cdots \otimes D_{G_1}$.

Defined as above, a weighted hierarchy $G^{\Pi_{\vec{\alpha}} k}$ is uniquely and efficiently specified by a real vector of weights $\vec{\alpha} \in {\mathbb{R}_+}^k$ and an ordered tuple of graphs $(G_1, \ldots, G_k)$. It will be the case that our analyses are unaffected by an overall scaling of the weight vector, so that one may identify $\vec{\alpha} \equiv c\vec{\alpha}$ for any real scalar c. As convention, we will always normalize by setting $\alpha_1 = 1$, which corresponds to assigning a unit-weight multiplicative factor to the lowest-level graphs in the hierarchy.

We can construct the adjacency matrix of the graph $G^{\Pi_{\vec{\alpha}} k}$ by repeated application of the two-fold product (Def. 3.2.2) in some well-defined way, analogous to Eq. (3.7). However, unlike before, the weighted product is non-associative, so we must first define an order of operations for manifold weighted products. Unless otherwise specified, we will always evaluate a manifold product from *right to left*, which corresponds to building the hierarchies from the bottom up, and is required in order to ensure that this definition matches Eq. (3.10). For example, in the 3-fold product $A_3 \,\Pi_{\alpha_3}\, A_2 \,\Pi_{\alpha_2}\, (\alpha_1 A_1)$, we will first evaluate the product $A_2 \,\Pi_{\alpha_2}\, (\alpha_1 A_1)$, and then take the product of A_3, weighted by α_3, with the resulting graph. The final result is

$$\alpha_3 A_3 \otimes D_2 \otimes D_1 + \alpha_2 \mathbb{1}_3 \otimes A_2 \otimes D_1 + \alpha_1 \mathbb{1}_3 \otimes \mathbb{1}_2 \otimes A_1. \tag{3.11}$$

In fact, a k-fold product, when evaluated this way, matches the right hand side of Eq. (3.10). Therefore, the k-level weighted hierarchy can also be written unambiguously as

$$A^{\Pi_{\vec{\alpha}} k} = A_k \,\Pi_{\alpha_k}\, A_{k-1} \,\Pi_{\alpha_{k-1}} \cdots \Pi_{\alpha_2}\, (\alpha_1 A_1). \tag{3.12}$$

Henceforth, the weight α_1, which scales the lowest-level adjacency matrix A_1, will be dropped due to our normalization choice of $\alpha_1 = 1$.

An important class of hierarchy graphs is one where the level weights follow a geometric progression of weights, i.e., $\alpha_i = \alpha^{i-1}$. We will denote such hierarchies by $G^{\Pi_\alpha k}$, where the scalar subscript α will be understood to mean the mutual weighting between successive hierarchies. For $\alpha > 1$, this leads to a "fat tree" structure, while for $\alpha < 1$, we instead get a "skinny tree" for which the edge weights decrease between

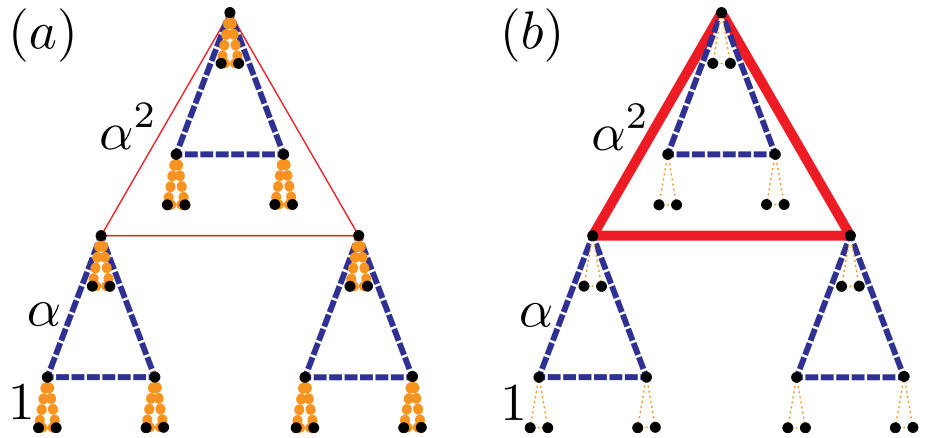

Figure 3.2: An illustration of the use of the hierarchical product to produce (a) "skinny" and (b) "fat" trees. In each case, the hierarchy $K_3^{\Pi_\alpha 3}$ is drawn, with the thickness of edges illustrating the weight of those edges. Depending on whether $\alpha < 1$ or $\alpha > 1$, this can lead to either lower-weighted high-level edges as in (a) or higher-weighted ones as in (b). Note that, for ease of visualization, here we break the usual convention of taking the lowest-level edges as unit weight.

consecutive levels from the leaves to the root. These constructions are illustrated in Fig. 3.2, and mentioned because fat trees are known to be a commonly used architecture in classical networks [114].

Allowing a clear separation of the modular system into hierarchical levels, each of which can be assigned unique edge weight, enables straightforward discussion of computation that occurs both within and between modules in a unified framework. When two nodes interact, we can assign this a cost that depends on the edges between them.

3.2.2.3 Node Addressal

A hierarchy on N nodes gives a natural labeling of the nodes. Suppose the hierarchy H contains k levels and each level is described by a graph G with $|G| = n$

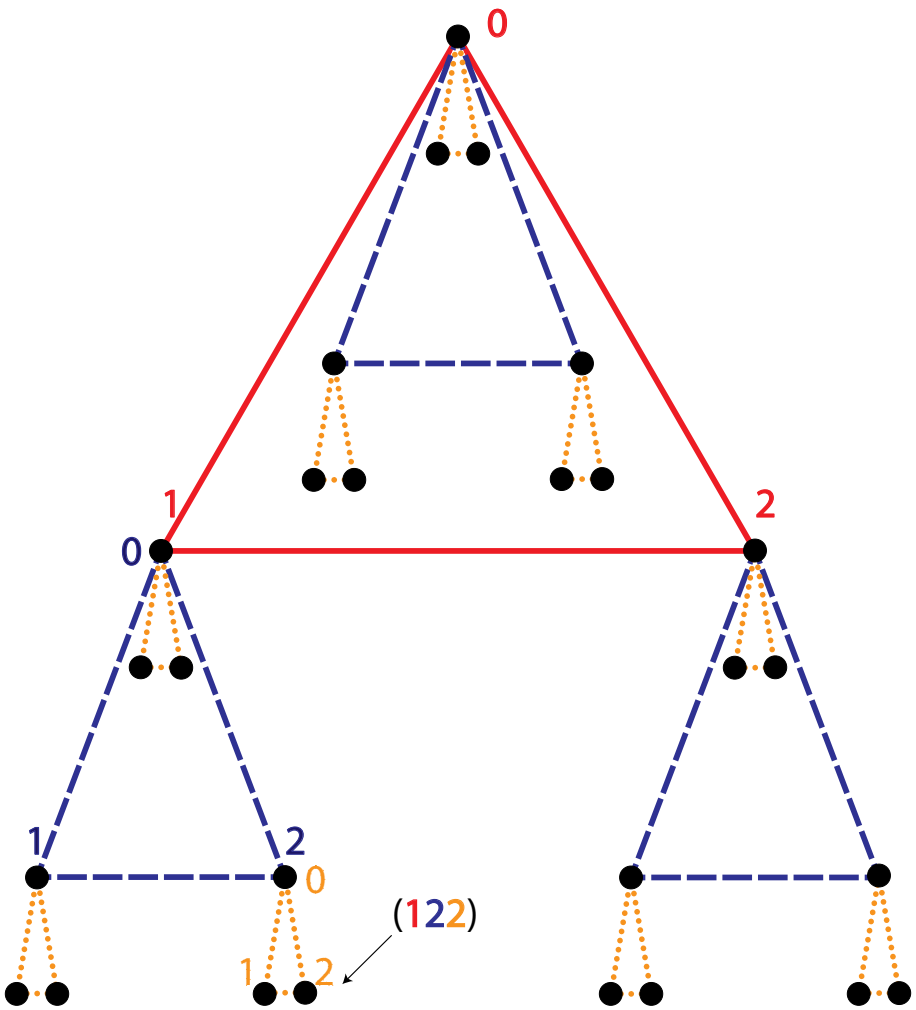

Figure 3.3: Addressing nodes in the hierarchy, layer by layer. Shown is a three-level hierarchy with the triangle graph K_3 as its base. Each vertex is represented as a 3-digit number in base 3. The first digit points to a node at the top level (red solid triangle), the second to a location in the second level (blue dashed triangle), and finally, the last digit (yellow dotted triangle) specifies the node location completely.

nodes, where $n^k = N$. Label the vertices of G by indices $j = 0, 1, \ldots, n-1$. Then, the adjacency matrix $\mathbb{1}_G \otimes G$ (which corresponds to n disjoint copies of G) has vertices which may be labeled as (jk), where $j, k = 0, 1, \ldots n-1$. The first label identifies which copy of G the node occurs in, while the second identifies where in G it appears. The same vertex labeling can then be used for the 2-level hierarchy $G \,\Pi\, G$. In this manner, the k-level hierarchy has n^k vertices with labels of the form $(b_1 b_2 \cdots b_k)$, where $b_i \in \{0, 1, \ldots, n-1\}$ for all i. This is essentially a k-digit, base-n representation of numbers from 0 to $N = n^k - 1$, as illustrated in Fig. 3.3.

This node addressal scheme allows for each node to be uniquely identified in a way that simultaneously describes its connectivity to other nodes and allows for easy counting of how many nodes lie in either the entire graph or in particular subgraphs. This addressal scheme will be important for describing a variant of hierarchies in Sec. 3.2.2.5 and for implementing the graphs in software, e.g. as used to generate the numerical results in Sec. 3.4.3.

3.2.2.4 Spectral Properties

One of the tools frequently used in analyzing large networks is the spectral decomposition of the Laplacian. The behavior of the largest eigenvalue, the first eigenvalue gap, and the distribution of eigenvalues as a function of the network parameters are some of the diagnostics that can provide key information about dynamical processes on the network, and can also be used as points of comparison between competing network topologies [115].

The smallest eigenvalue of a Laplacian is always $\lambda_1 = 0$, which corresponds to the uniform eigenvector $\vec{e}_1 = (1, 1, \ldots, 1)$. In ascending order, the eigenvalues of L may be denoted by $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. We now state some graph properties that can be related to the spectrum of L [115, 116].

The second eigenvalue λ_2 is known as the *algebraic connectivity* of the graph and is closely related to the expansion and connectivity properties of the graph. Broadly, the larger the value of λ_2, the better the connectivity of the network. To illustrate this point, consider the graph diameter, $\delta(H)$, which can be bounded using λ_2 as follows:

$$\frac{4}{N\lambda_2} \leq \delta\left(H\right) \leq 2\left\lceil \frac{\Delta + \lambda_2}{4\lambda_2} \ln\left(N-1\right) \right\rceil, \tag{3.13}$$

where Δ is the maximum degree of H. It can be seen that a larger value for λ_2 will lead to a smaller graph diameter. We also have the following asymptotic bound on the mean distance between nodes, $\bar{\rho}(H)$:

$$\frac{2}{(N-1)\lambda_2(H)} + \frac{1}{2} \lesssim \bar{\rho}(H) \lesssim \left\lceil \frac{\Delta + \lambda_2}{4\lambda_2} \ln\left(N-1\right) \right\rceil. \tag{3.14}$$

Another important diagnostic of a network is given by the *Cheeger constant* $h(H)$ [117], also called the isoperimetric number or the graph conductance. This graph invariant is a measure of how difficult the graph is to disconnect by cutting edges. For a connected graph, this number is always positive. As benchmark values, the complete graph K_N has Cheeger constant $N/2$ while a cycle graph C_N has Cheeger constant $4/N$. The relationship between λ_2 and $h(H)$ can be seen through the following bounds:

$$\frac{\lambda_2}{2} \leq h(H) \leq \sqrt{\lambda_2\left(2\Delta - \lambda_2\right)}. \tag{3.15}$$

Many other graph properties may be derived from the Laplacian spectrum as well (see, e.g., Refs. [115, 116]).

For a large network, finding the eigenvalues can be numerically expensive. However, hierarchies have a special structure which can be exploited for the evaluation of graph spectra. Here, we show (in Theorem 3.2.1) that if the spectra of the base graphs L_i are known, then one can derive the spectrum of the k-level hierarchy efficiently using a recursive procedure. We first present two lemmas. The first lemma generalizes Theorem 3.10 from Ref. [88], which states that the characteristic polynomial $\phi_P(x)$ $(= \det[x\mathbb{1} - P])$ of an unweighted hierarchical product of adjacency matrices A, B is given by

$$\phi_P(x) = \phi_{B'}(x)^{n_A} \phi_A\left(\frac{\phi_B(x)}{\phi_{B'}(x)}\right), \tag{3.16}$$

where A' (resp. B') is the matrix A (resp. B) with the first row and first column removed, and $n_A = |G_A|$ is the order of the graph A. In fact, Eq. (3.16) applies to Laplacians as well as adjacency matrices. The lemma below further generalizes this statement to a weighted product of Laplacians.

Lemma 3.2.1. Let K and L be two graph Laplacians with characteristic polynomials given by $\phi_K(x)$ and $\phi_L(x)$, respectively. Then, the characteristic polynomial $\phi_\Pi(x)$ of the hierarchical product $K\,\Pi_\alpha\,L$ is given by

$$\phi_\Pi(x) = [\alpha\phi_{L'}(x)]^{n_K} \phi_K\left(\frac{1}{\alpha}\frac{\phi_L(x)}{\phi_{L'}(x)}\right), \tag{3.17}$$

where $n_k = \dim\{K\}$, and L' is defined similar to A' and B' above.

Proof. Denote the spectra of K and L by $\{\kappa_j\}$ and $\{\lambda_j\}$, respectively. Recall that

the α-weighted hierarchical product may be written as

$$K \,\Pi_\alpha\, L = \alpha K \otimes D_L + \mathbb{1}_K \otimes L. \tag{3.18}$$

If U_K is a unitary that diagonalizes K, we conjugate the above equation with the unitary $U_K \otimes \mathbb{1}_L$, and look at the resulting block matrix. Each block corresponds to an eigenvalue of K, and thus the j-th block is given by $\alpha\kappa_j D_L + L$. The full spectrum may then be expressed as a disjoint union of the block spectra,

$$\operatorname{spec}\left(K \,\Pi_\alpha\, L\right) = \bigsqcup_{j=1}^{|K|} \operatorname{spec}\left(\alpha\kappa_j D_L + L\right). \tag{3.19}$$

Now, we apply Eq. (3.16) to $K \,\Pi_\alpha\, L \equiv (\alpha K)\,\Pi\, L$ and use the fact that $\phi_{\alpha K}(x) = \det\left[x\mathbb{1} - \alpha K\right] = \alpha^{n_K} \det\left[\frac{x}{\alpha}\mathbb{1} - K\right] \equiv \alpha^{n_K}\phi_K\left(\frac{x}{\alpha}\right)$. This yields Eq. (3.17), as desired.

□

Now we show that if the eigenvalues of K and the polynomials ϕ_L and $\phi_{L'}$ are known, then there is a straightforward procedure to compute the eigenvalues of $K \,\Pi_\alpha\, L$.

Lemma 3.2.2. Let K and L be graph Laplacians, as before. Each eigenvalue of the product characteristic polynomial ϕ_Π can be found as a solution of the equation

$$\alpha\kappa_i = \frac{\phi_L\left(x\right)}{\phi_{L'}\left(x\right)} \tag{3.20}$$

for some K-eigenvalue κ_i.

Proof. Any eigenvalue of the product graph must be a zero of the left-hand side of Eq. (3.17) and, by equality, a zero of the right-hand side. Now, the degree of polynomial ϕ_K is n_K, which implies that the term of degree n_K must be nonzero. Thus,

in the product $\phi_{L'}(x)^{n_K}\,\phi_K\left(\frac{1}{\alpha}\frac{\phi_L(x)}{\phi_{L'}(x)}\right)$, there must be a term which is *indivisible* by the polynomial $\phi_{L'}(x)$. Therefore, the zero of the right-hand side cannot be a root of the polynomial $\phi_{L'}$.

We are seeking values of x such that the polynomial $\phi_K\left(\frac{1}{\alpha}\frac{\phi_L(x)}{\phi_{L'}(x)}\right)$ evaluates to zero. In other words, we are looking for x such that the term $\frac{1}{\alpha}\frac{\phi_L(x)}{\phi_{L'}(x)}$ is a root of ϕ_K. Therefore, we solve Eq. (3.20) for x, for all roots κ_i of K. □

If the forms of ϕ_L and $\phi_{L'}$ are known (and if each have sufficiently low degree), then computing the roots of ϕ_Π becomes tractable, even if K is a large matrix. This suggests a recursive procedure for computing the spectrum of a k-level hierarchy, by writing it as a product of the $(k-1)$-level hierarchy with the k-th base graph. We now frame this as our main result of this section:

Theorem 3.2.1. Suppose we have a k-level hierarchy $L^{\Pi_{\vec{\alpha}}k}$ described by base graph Laplacians $L_1, L_2, \ldots, L_k$ and weights $\vec{\alpha} = (1, \alpha_2, \ldots, \alpha_k)$ as follows,

$$L^{\Pi_{\vec{\alpha}}k} = L_k\,\Pi_{\alpha_k}\,L_{k-1}\,\Pi_{\alpha_{k-1}}\cdots\Pi_{\alpha_3}\,L_2\,\Pi_{\alpha_2}\,L_1. \tag{3.21}$$

Define a new set of weights $\vec{\beta} = (1, \beta_2, \ldots, \beta_k)$ with $\beta_i = \alpha_i/\alpha_{i-1}$, and a new set of Laplacians $M_k, M_{k-1}, \ldots, M_1$ recursively as

$$M_k = L_k,$$

$$M_i = M_{i+1}\,\Pi_{\beta_{i+1}}\,L_i.$$

Then, the following hold:

1. $M_1 = L^{\Pi_{\vec{\alpha}}k}$.

2. Any eigenvalue of M_i (for $i < k$) may be found as a solution to the equation

$$\beta_{i+1}\mu^{(i+1)} = \frac{\phi_{L_i}(x)}{\phi_{L_i'}(x)} \tag{3.22}$$

for some $\mu^{(i+1)} \in \text{spec}\,\{M_{i+1}\}$.

Proof. First, we prove statement 1. It can be seen that

$$\begin{aligned} M_{k-1} &= M_k \,\Pi_{\beta_k}\, L_{k-1} = L_k \,\Pi_{\beta_k}\, L_{k-1} \\ &= \frac{1}{\alpha_{k-1}} \left(\alpha_k L_k \otimes D_{k-1} + \alpha_{k-1} \mathbb{1}_k \otimes L_{k-1}\right), \end{aligned} \tag{3.23}$$

$$\begin{aligned} M_{k-2} &= M_{k-1} \,\Pi_{\beta_{k-1}}\, L_{k-2} \\ &= \frac{1}{\alpha_{k-2}} (\alpha_k L_k \otimes D_{k-1} \otimes D_{k-2} + \\ &\alpha_{k-1} \mathbb{1}_k \otimes L_{k-1} \otimes D_{k-2} + \alpha_{k-2} \mathbb{1}_{k-1} \otimes \mathbb{1}_{k-2} \otimes L_{k-2}), \end{aligned} \tag{3.24}$$

and so on, until we have an $\vec{\alpha}$-weighted sum over all k of the base graphs (with an overall denominator of $\alpha_1 = 1$), which is precisely $L^{\Pi_{\vec{\alpha}} k}$.

The proof of statement 2 follows as a direct consequence of Lemma 3.2.2, with $K = M_{i+1}, L = L_i$, and $\alpha = \beta_{i+1}$. □

Theorem 3.2.1 provides an algorithm to compute the spectrum of $L^{\Pi_{\vec{\alpha}} k}$, namely:

1. Compute the relative weight vector $\vec{\beta}$ from $\vec{\alpha}$.

2. Start with $i = k$, where the spectrum of $M_k = L_k$ is known. Decrease i by one.

3. Compute the spectrum of M_i from the known spectrum of M_{i+1} and Eq. (3.22). Decrease i by one.

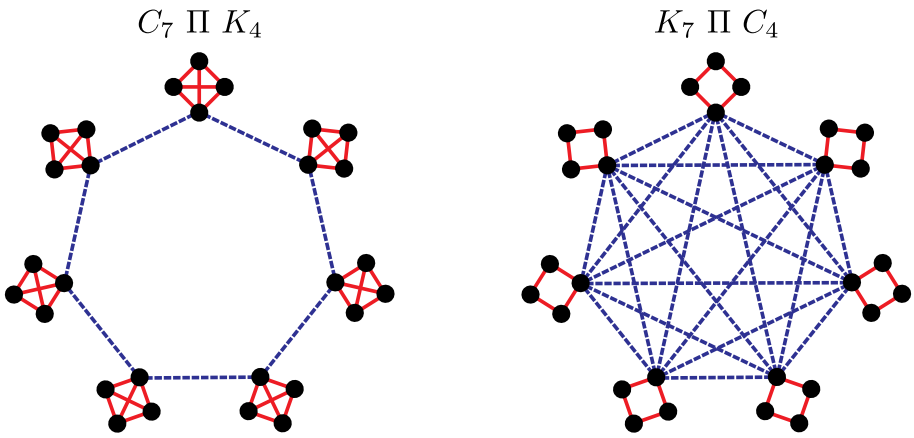

Figure 3.4: Two topologies with the same number of nodes (28) and edges (49). While the diameters for the two graphs are the same, are they equally well-connected? A comparison of the Cheeger constants (see Table 3.1) suggests that the left graph is less interconnected. This is consistent with the spectral gap, which is smaller for the left graph, indicating poorer connectivity.

4. Perform step 3 repeatedly, halting at $i = 0$. Return the spectrum of $M_1 = L^{\Pi_{\vec{\alpha}^k}}$.

Therefore, given a large hierarchy, one can efficiently compute the Laplacian eigenvalues and use them to find bounds on important graph properties. This is a scalable technique for obtaining figures of merit efficiently for hierarchies. Later, in Sec. 3.3, we will present analytic results for some of these figures of merit for simple hierarchies, but the results of the current section can be used even in more complicated cases, such as hierarchies that do not use the same G at every layer or that have heterogeneous scaling parameters.

Due to the structural richness and heterogeneity of graphs, it is not always easy to decide whether one graph is, for instance, more connected than another graph. One aspect of connectivity is how close the nodes are to one another, which is captured by quantities like the diameter and mean distance. In Fig. 3.4, we

compare two graphs, $C_7 \,\Pi\, K_4$ and $K_7 \,\Pi\, C_4$, which have an identical number of nodes (28) and edges (49). The two graphs also have identical diameters (5 each), but the mean distance for the left graph is smaller (see Table 3.1). Under these measures, the left graph appears better connected.

Better connectivity also corresponds to having fewer bottlenecks in the graph, which corresponds to a larger Cheeger constant. In Fig. 3.4, the graph on the right has a larger Cheeger constant, as one would expect given that it has complete connectivity between the seven modules. Note that this metric of connectivity need not agree with the mean distance, as seen in this example.

Similarly, a parameter-by-parameter comparison of the two hierarchy graphs $C_{13} \,\Pi\, K_5$ and $K_{13} \,\Pi\, C_5$ (Table 3.1) reveals that, while both graphs are two-level hierarchies with the same number of nodes and edges, $K_{13} \,\Pi\, C_5$ has the smaller diameter, smaller mean distance, larger cheeger constant, and a larger spectral gap, all of which indicate better connectivity. While structural comparisions for the above examples can be carried out simply by inspection or a quick calculation of graph quantities, general hierarchies may be far too complex to compare this way. In practice, when choosing a modular topology with the best connectivity, one might hope for a single, balanced measure of connectivity that relates to aspects such as node distance and bottleneckedness and is easy to compute. The spectral gap λ_2 meets these requirements. It is asymptotically related to the other invariants discussed here via upper and lower bounds in Eqs. (3.13)–(3.15). Furthermore, λ_2 can be efficiently computed using the recursive procedure described earlier in this section.

Graph Invariant	$C_7 \,\Pi\, K_4$ vs.	$K_7 \,\Pi\, C_4$	$C_{13} \,\Pi\, K_5$ vs.	$K_{13} \,\Pi\, C_5$
Number of edges	49	49	143	143
Number of nodes	28	28	65	65
Diameter	<u>5</u>	<u>5</u>	8	<u>5</u>
Mean distance	<u>2.68</u>	2.71	4.77	<u>3.23</u>
Cheeger constant	0.17	<u>1.0</u>	0.07	<u>1.4</u>
Spectral gap λ_2	0.16	<u>0.46</u>	0.04	<u>0.34</u>

Table 3.1: Comparison of topologies by connectivity measure. In each case, the graphs being compared have an identical number of nodes and edges. The better value for each comparison is underlined.

3.2.2.5 Truncated Hierarchical Product

In some scenarios, there may be physical or technological limitations on the total number of interconnections allowed at a single node of a quantum computer. In our framework, this manifests as a restriction on the maximum degree of a node. We believe that hierarchical structures can still prove useful in this context, but (as we will see in Sec. 3.3) the hierarchy we have described thus far has a maximum degree which grows linearly with the number of levels of the hierarchy.

We now introduce an architecture which maintains the hierarchical properties but also has a bounded maximum node degree (i.e. maximum node degree that does not go to infinity as the number of levels goes to infinity). To model such an architecture, we modify the hierarchical product $G_1 \,\Pi\, G_2$. Whereas previously, $|G_1|$

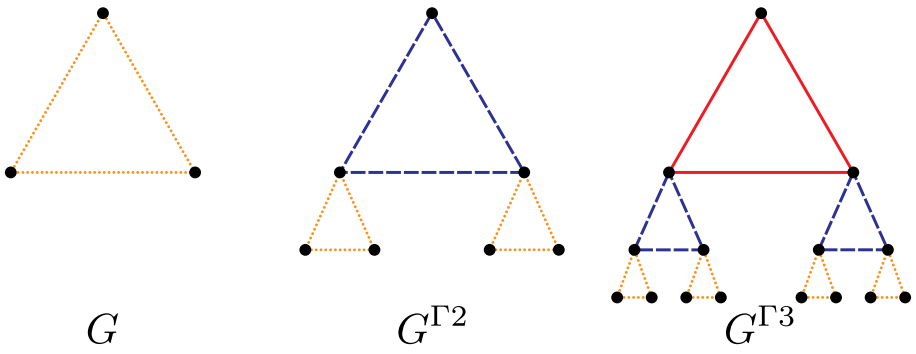

Figure 3.5: A demonstration of how our hierarchical product can be truncated to avoid requiring many interconnections at one node. As the hierarchy grows, the graph is duplicated and then attached to a subset of nodes in a larger version of the base graph, G.

copies of G_2 were connected according to G_1, we now bring together $|G_1| - 1$ copies, which we connect according to G_1, and add the root node of G_1 without an associated subhierarchy (see Fig. 3.5). When extended to a many-level hierarchy, this means that every node will be connected to, at most, two levels, and so its degree will not grow as the hierarchy grows. We will denote this *truncated* hierarchical product by $G_1 \,\Gamma\, G_2$, and its weighted version as $G_1 \,\Gamma_\alpha\, G_2$. It can be written algebraically in terms of adjacency matrices by adopting a more general definition of the hierarchical product.

Definition 3.2.3. Given rooted graphs G and H, the *weighted truncated hierarchical product* $P = G \,\Gamma_\alpha\, H$ is a graph on vertices $V_P = V_G \times V_H$ and edges $E_P \subseteq V_P \times V_P$ specified by the adjacency matrix

$$A_P = \alpha A_G \otimes D_H + P_G \otimes A_H, \tag{3.25}$$

or, equivalently, the algebraic Laplacian

$$L_P = \alpha L_G \otimes D_H + P_G \otimes L_H. \tag{3.26}$$

Here, P_G is a projector onto all nodes in G except the root node. At the level of adjacency matrices, we may also write $A_P = A_G \, \Gamma_\alpha \, A_H$. An unweighted version, $G \, \Gamma \, H$, can be obtained by setting $\alpha = 1$.

An illustration of this architecture can be found in Fig. 3.5. From this definition, we naturally derive both unweighted and weighted truncated hierarchies, $G^{\Gamma k}$ and $G^{\Gamma_{\vec{\alpha}} k}$. We note that a generalization of this definition to allow an arbitrary projector (rather than one that only excludes the root node) is possible, but we do not consider such a case in this chapter.

The addressing scheme outlined in Sec. 3.2.2.3 can also be used for truncated hierarchies. However, since many nodes do not sit atop sub-hierarchies in this case, not all node addresses are valid. We will assume that the node in the i-th level which connects to the level above it has a zero in the i-th digit of its address. In a truncated hierarchy, each node whose address contains a zero (representing the "root" of a hierarchy) must have only zeros in all following positions, as it does not contain any further sub-hierarchies. The base-n addressal scheme can thus be used to specify which nodes are present in a truncated hierarchy.

Note that the truncated hierarchical product adds nodes more slowly than (although with the same scaling as) the hierarchical product structure specified at the beginning of Sec. 3.2.2. When we perform graph comparisons in Sec. 3.3, we will consider all cost functions and optimizations in terms of the total number of nodes so that the two architectures can be compared fairly.

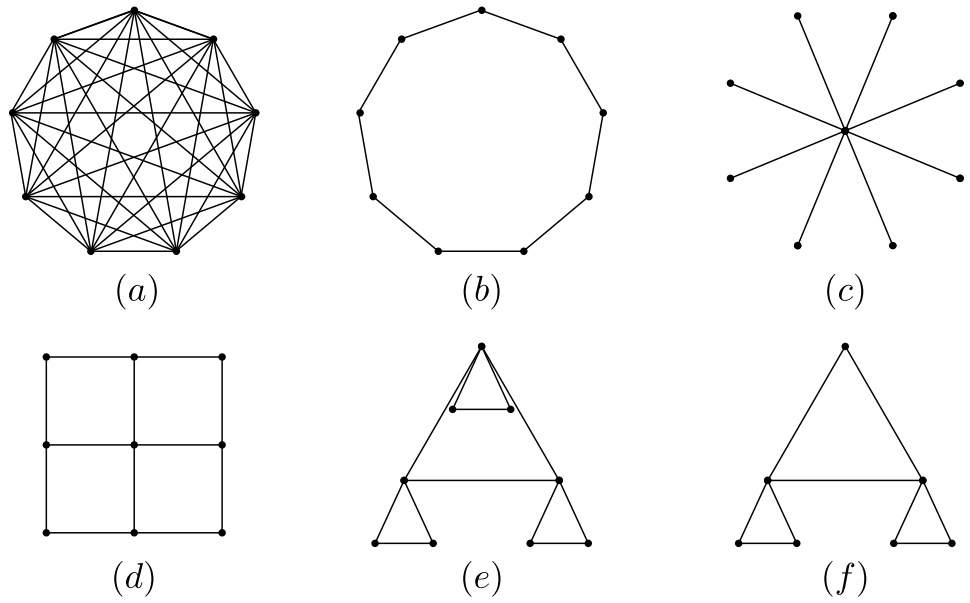

Figure 3.6: Illustration of the graph structures considered in this section, each with nine nodes except (f). (a) The complete graph K_9. (b) The cycle graph C_9. (c) The star graph S_9. (d) The nearest-neighbor grid in two dimensions. (e) The hierarchical product $K_3^{\Pi 2}$. (f) The truncated hierarchical product of Sec. 3.2.2.5, $K_3^{\Gamma 2}$.

3.3 Graph Comparisons

Having developed the machinery to construct hierarchies, we will now evaluate them against other potential architectures. Any evaluation is impossible to do in an absolute sense, since what properties are desirable in a graph and how serious the cost of improving them is will depend on both the application as well as the physical system under consideration. In general, we assume that the most desirable quality of a graph is some measure of connectivity or the ease with which the graph can transport information between nodes. Note that it is always possible to translate between quantum circuit architectures with some overhead. A detailed atlas

summarizing these overheads can be found in Ref. [118].

We will look at the scenario of state transfer, which is an important subroutine that may need to be carried out if an algorithm requires gates to be performed between two qubits that are not directly connected. We consider the worst-case state transfer time on a given graph, which allows us to evaluate graphs without reference to any particular quantum algorithm. If we are interested in the time taken for state transfer in the graph, an appropriate metric can be the diameter of the graph, $\delta(H)$, under the assumption that information transfer takes unit time along any edge in the graph. The diameter then captures the maximum distance, and hence the maximum time required for information to travel between any two nodes in the system.

For graphs produced by the weighted hierarchical product, we will also consider a diameter which takes into account edge weight. This "weighted diameter," $\delta_w(H)$, can be found by considering all pairs of nodes j, k and identifying the two whose least-weighted connecting path has the highest sum weight of edges. If we consider a path between two nodes j and k to be a set of nodes $P = \{j, v_1, v_2 \dots v_n, k\}$ with a weight $W(P)$ given by the sum $w_{j,v_1} + w_{v_1,v_2} + \cdots + w_{v_n,k}$, then the weighted diameter can be written as:

$$\delta_w(H) = \max_{j,k} \min_{P} W(P). \tag{3.27}$$

One way to grasp why the weighted diameter is a useful quantity is to consider the time weights of edges, where the weight signifies the time required to perform a gate between two connected qubits. In this case, the weighted diameter is the maximum

time it will take us to perform a chain of two-qubit gates that connects two different qubits (for instance, using SWAP operations to bring the two qubits to adjacent positions and then performing the final desired operation).

However, optimizing *only* with respect to connectivity yields a trivial result, because a fully connected graph is obviously most capable of communicating information between any two points. Therefore, we will consider a number of different possible "costs" associated with physical implementations of graphs. One potential input to the cost function is the maximum degree of a graph, $\Delta(H)$. As discussed in the previous section, we want to avoid needing to connect too many different communication channels to a single node. Another is total edge weight $w(H)$ – if it costs time, energy, money, coherence, or effort to produce communication between two nodes, we should try to use as few communication channels as possible.

We now walk through the calculations for several important graph quantities for several graphs: an all-to-all connected graph, a cycle graph, a star graph, a square grid, a hierarchy graph with scaling parameter α, and a truncated version of that same hierarchy graph. We calculate how quantities scale with the total number of nodes N. For ease of calculation, we assume that N nodes fit in the architecture of the current graph; for instance, we assume $N = \ell^d$ for some integer ℓ for a d-dimensional square graph. All results of this section are compiled in Table 3.2, and examples of the graphs for small N are illustrated in Fig. 3.6.

Graph H	Diameter δ	Weighted Diameter δ_w	Max Degree Δ	Total Edge Weight $w(H)$
K_N	const.	const.	N	N^2
S_N	const.	const.	N	N
C_N	N	N	const.	N
d-dim grid	$dN^{1/d}$	$dN^{1/d}$	d	dN
$K_n^{\Pi_\alpha k}, \alpha \neq n$	$\log_n N$	$\max\left(\frac{2}{1-\alpha}, N^{\log_n \alpha}\right)$	$n \log_n N$	$nN^{\max(1,\log_n \alpha)}$
$K_n^{\Pi_\alpha k}, \alpha = n$	$\log_n N$	$\max\left(\frac{2}{1-\alpha}, N^{\log_n \alpha}\right)$	$n \log_n N$	$nN \log_n N$
$K_{n+1}^{\Gamma_\alpha k}, \alpha \neq n$	$\log_n N$	$\max\left(\frac{2}{1-\alpha}, N^{\log_n \alpha}\right)$	n	$nN^{\max(1,\log_n \alpha)}$
$K_{n+1}^{\Gamma_\alpha k}, \alpha = n$	$\log_n N$	$\max\left(\frac{2}{1-\alpha}, N^{\log_n \alpha}\right)$	n	$nN \log_n N$

Table 3.2: Summary of scalings of important graph properties with total node number, N. All entries describe only the scaling of the leading coefficient with d, n, and N.

3.3.1 Graph Calculations

3.3.1.1 Complete Graph, K_N

Since all nodes in a complete graph [Fig. 3.6(a)] have edges between them, the diameter is simply 1. This comes at the cost of very high maximum degree, $N-1$, as every node is connected to all $N-1$ other nodes. The total weight of every edge is the same, and there are $N(N-1)/2$ edges because every pair of nodes has a corresponding edge. Therefore, the total edge weight scales as $\Theta(N^2)$.

3.3.1.2 Cycle Graph, C_N

In a cycle graph [Fig. 3.6(b)], the diameter is $\lfloor N/2 \rfloor$, the distance to the opposite side of the circle. The maximum degree is only 2, and the total weight of the edges is likewise only N. This graph is thus able to reduce the cost factors associated with the complete graph, but at the cost of a much higher asymptotic diameter.

3.3.1.3 Star Graph, S_N

The star graph is the graph which has a single central node connected to all others [Fig. 3.6(c)]. Like the complete graph, it also has a constant diameter, although this diameter is two rather than one. The maximum degree of the star graph is $N-1$, the same as the complete graph. However, the star graph improves over the complete graph, as it has a lower total edge weight of $N-1$ rather than $\binom{N}{2}$. Thus, we have improved the cost asymptotically without affecting the overall scaling of the diameter of the graph.

The example of S_N raises a complication which we do not attempt to quantify in this chapter. In a realistic distributed quantum computer, we expect that a significant amount of operations need to be performed at the same time and need to be scheduled on the graph. But in the star graph, all operations between nodes must pass through the single central hub. This is likely to lead to a scheduling bottleneck when performing general quantum algorithms. While we do not attempt to treat scheduling of such algorithms on the network in this chapter, in future work we hope

to consider these complications, which will at times make the star graph unsuitable for real-world use. An experimental comparison of the star graph and the complete graph in existing five-qubit quantum computers can be found in Ref. [119]. In those experiments, the requirement that all information be shuttled through a central node for the S_N connectivity made high-fidelity execution of quantum algorithms more difficult.

3.3.1.4 Square Grid Graph

We consider now a square grid (i.e., a hypercubic lattice) in d dimensions [Fig. 3.6(d)]. Here, the diameter is $d(N^{1/d} - 1)$, since this is the distance from the point in one corner labeled $(1, 1, 1, \dots)$ to the opposite corner at $(N^{1/d}, N^{1/d}, \dots)$ (note that diagonal moves are not allowed). The maximum degree depends on the dimension, as each interior node is connected to $2d$ other nodes. The total edge weight can be found by considering that each node on the interior of the graph corresponds with exactly d edges, and it is these edges that dominate as $N \to \infty$. Therefore, the total edge weight scales as $\Theta(dN)$.

3.3.1.5 Hierarchy Graph, $G^{\Pi_{\vec{\alpha}} k}$

As the hierarchy graph [Fig. 3.6(e)] is built recursively, it is easiest to calculate its properties using recursion relations. We consider a graph that has k levels to it, so that given a base graph G and $n = |G|$, then the overall graph has n^k nodes.

First, we calculate the unweighted diameter of a k-level hierarchy, which we

denote by $\delta\left(G^{\Pi_{\vec{\alpha}}k}\right)$. Since all sub-hierarchies are rooted at their first vertex, we will need to keep track of the *eccentricity* of the root node, which we denote by $\varepsilon(F)$ for any subhierarchy F. The eccentricity of any graph node is defined as the maximum distance from that node to any other node in the graph F. Here, we fix $\varepsilon(F)$ to be the root eccentricity for the graph in question.

Now, we write recursion relations for two quantities, the unweighted diameter $\delta(G^{\Pi_{\vec{\alpha}}i})$ of an i-level hierarchy for some intermediate i, and the eccentricity $\varepsilon(G^{\Pi_{\vec{\alpha}}i})$ of the top-level root node of the current i-level hierarchy.

Consider a diametric path in an i-level hierarchy. This path must ascend and descend the entire hierarchy. That is, using the notation of Sec. 3.2.2.3, two maximally separated qubits have addresses that are different in their first digit. Such a path can always be partitioned into 3 disjoint pieces, the terminal two of which each lie in some $(i-1)$-level subhierarchy, while the middle piece lies in the current top (i.e. i-th) level. These three pieces must be independently maximal, since the path is diametric. The middle piece maximizes to the diameter of the top-level graph, which is simply $\delta(G)$. The two sub-level pieces each maximize to the root eccentricity of the $(i-1)$-th level subhierarchy, which is precisely the quantity $\varepsilon(G^{\Pi_{\vec{\alpha}}(i-1)})$. Therefore, our first recursion reads

$$\delta(G^{\Pi_{\vec{\alpha}}i}) = 2\varepsilon(G^{\Pi_{\vec{\alpha}}(i-1)}) + \delta(G). \tag{3.28}$$

The i-th level root eccentricity may be found by a similar argument. Partition the most eccentric path (starting at the top level root node) into two pieces, one which lies at the top level, and the other which lies exclusively in the lower levels.

Maximizing both pieces, one gets

$$\varepsilon(G^{\Pi_{\vec{\alpha}} i}) = \varepsilon(G^{\Pi_{\vec{\alpha}}(i-1)}) + \varepsilon(G). \tag{3.29}$$

Solving the second relation, we get $\varepsilon(G^{\Pi_{\vec{\alpha}} i}) = i\varepsilon(G)$. By substitution, the first recursion has the solution

$$\delta(G^{\Pi_{\vec{\alpha}} k}) = 2(k-1)\varepsilon(G) + \delta(G). \tag{3.30}$$

Since the total number of levels is given by $k = \log_n N$, and the graph diameter is no greater than twice the eccentricity of any node, we conclude that the diameter scales as $\Theta(\varepsilon(G)\log_n N)$ for a general graph G. If we specifically examine the case when G is a complete graph of order n, $\delta(G) = 1$ and $\varepsilon(G) = 1$, and the exact expression is $\delta\left(G^{\Pi_{\vec{\alpha}} k}\right) = 2\log_n(N) - 1$.

Next we calculate the maximum degree. Again, we proceed by recursion. Iterating the hierarchical product to some level i can be viewed as attaching a copy of the graph $G^{\Pi_{\vec{\alpha}}(i-1)}$ to every point in the graph G. Therefore, the degree of every root node in the $(i-1)$-level subhierarchies increases by the degree of the corresponding node in graph G. The maximal increase achievable thus is the maximum degree $\Delta(G)$ of graph G. Since the root node for an i-level subhierarchy has i distinct copies of G attached to it, its degree is given by $i \cdot \deg(g_1)$, where g_1 is the root node of G. Then, the i-level maximum degree can be expressed as

$$\Delta(G^{\Pi_{\vec{\alpha}} i}) = \max\left\{(i-1)\deg(g_1) + \Delta(G), \Delta(G^{\Pi_{\vec{\alpha}}(i-1)})\right\} \tag{3.31}$$

$$\ldots = \max_{0 \le j \le i-1} \{j \deg(g_1) + \Delta(G)\} \tag{3.32}$$

$$= (i-1)\deg(g_1) + \Delta(G), \tag{3.33}$$

where the second step was obtained by recursion. For a general G, this gives the maximum degree scaling as $\Delta(G^{\Pi_{\vec{\alpha}}k}) = \Theta(\log_n N)$. For $K_n^{\Pi_{\vec{\alpha}}k}$, the root degree and the maximum degree of the base graph K_n are both $n-1$, so $\Delta(K_n^{\Pi_{\vec{\alpha}}k}) = (n-1)\log_n N$.

Now we consider the total edge weight of the hierarchy. We compute this by a recursion relation, first by duplicating the existing edge weight at $i-1$ levels by n (the number of smaller hierarchies we must bring together) and then adding new edges. If the edges at level i have weight α_i, we can write this as:

$$w(G^{\Pi_{\vec{\alpha}}i}) = nw(G^{\Pi_{\vec{\alpha}}(i-1)}) + \alpha_i w(G). \tag{3.34}$$

By counting the number of subhierarchies with different weights, we find the following form for the total edge weight of the weighted hierarchy:

$$w\left(G^{\Pi_{\vec{\alpha}}k}\right) = w(G)\sum_{i=1}^{k}\alpha_i\left|G\right|^{k-i}. \tag{3.35}$$

This can be verified by checking that it satisfies the recursion relation Eq. (3.34). If we now specialize to the case where $G = K_n$ and $\alpha_i = \alpha^{i-1}$, we find

$$w\left(K_n^{\Pi_\alpha k}\right) = \frac{n(n-1)}{2}\sum_{i=1}^{k}\alpha^{i-1}n^{k-i}. \tag{3.36}$$

This behavior can be broken into three regimes. For $\alpha = n$, every term in the sum is identical, and the overall scaling is $\Theta(nN\log_n N)$. Otherwise, we can perform the geometric sum to obtain

$$w\left(K_n^{\Pi_\alpha k}\right) = \frac{n(n-1)}{2}\frac{n^k-\alpha^k}{n-\alpha}. \tag{3.37}$$

Here, the scaling will depend on the relative size of n and α. For $n > \alpha$, the first

term in the numerator dominates, and $w\left(K_n^{\Pi_\alpha k}\right) = \Theta(nN)$. Otherwise, we can write $\alpha^k = N^{\log_n \alpha}$ and find $w\left(K_n^{\Pi_\alpha k}\right) = \Theta(nN^{\log_n \alpha})$.

Finally, we calculate the weighted diameter of a k-level hierarchy $\delta_w(G^{\Pi_{\vec{\alpha}} k})$, just as for the unweighted diameter, by solving recursion relations for the quantities $\delta_w(G^{\Pi_{\vec{\alpha}} i})$ and $\varepsilon_w(G^{\Pi_{\vec{\alpha}} i})$, which are, respectively, the weighted diameter and weighted root eccentricity for an i-level weighted hierarchy. Here, note that the top level (at any intermediate stage i) is weighted by α_i. Therefore, the recursion for the weighted diameter is modified to

$$\delta_w(G^{\Pi_{\vec{\alpha}} i}) = 2\varepsilon_w(G^{\Pi_{\vec{\alpha}}(i-1)}) + \alpha_i \delta_w(G). \tag{3.38}$$

Similarly, the recursion for the weighted eccentricity becomes

$$\varepsilon_w(G^{\Pi_{\vec{\alpha}} i}) = \varepsilon_w(G^{\Pi_{\vec{\alpha}}(i-1)}) + \alpha_i \varepsilon_w(G), \tag{3.39}$$

which has the solution $\varepsilon_w(G^{\Pi_{\vec{\alpha}} i}) = \varepsilon_w(G) \sum_{j=1}^{i} \alpha_j$. Finally, we have

$$\delta_w(G^{\Pi_{\vec{\alpha}} k}) = 2\varepsilon_w(G) \sum_{j=1}^{k-1} \alpha_j + \delta_w(G)\alpha_k. \tag{3.40}$$

For $G = K_n$ and $\alpha_i = \alpha^{i-1}$, this becomes:

$$\delta_w(K_n^{\Pi_\alpha k}) = 2 \sum_{i=1}^{k-1} \alpha^{i-1} + \alpha^{k-1} \tag{3.41}$$

$$= \frac{\alpha^k + \alpha^{k-1} - 2}{\alpha - 1}. \tag{3.42}$$

Therefore, the scaling of the weighted diameter with N has two regimes, depending on α. For $\alpha < 1$ the geometric sum converges as $i \to \infty$ to $\frac{2}{1-\alpha}$. This means that for $\alpha < 1$, a constant time suffices to traverse the entire hierarchy no matter how large

it is. For $\alpha = 1$ the weighted diameter is equal to the (unweighted) diameter, which we have already computed. For $\alpha > 1$, δ_w scales as $\alpha^{k-1} = N^{\log_n \alpha}/\alpha \sim N^{\log_n \alpha}$. Note that the last scaling only applies if α does not scale with n. Since $n > 1$ and $\alpha > 1$, this exponent $\log_n \alpha$ is always positive. Therefore, the total edge weight is asymptotically always either constant (for $\alpha < 1$) or growing (for $\alpha \geq 1$), as expected.

3.3.1.6 Truncated Hierarchy, $G^{\Gamma_{\vec{\alpha}} k}$

Finally, we look at how the results above are modified if we use the truncated hierarchical product discussed in Sec. 3.2.2.5 [Fig. 3.6(f)]. Although many of the calculations in terms of the number of levels k are similar to those for the non-truncated hierarchy, it is no longer the case that $k = \log_n N$ exactly. In order to compare graphs fairly, we will need to recalculate the order of $G^{\Gamma_{\vec{\alpha}} k}$ so that results in this section can be written in terms of the total number of nodes, N.

Under the node addressal scheme of Sec. 3.2.2.3, the nodes of a truncated hierarchy are in one-to-one correspondence with base-n strings of length k that only have trailing zeros. As before, a 0 label points to a root node, but since root nodes do not bear subhierarchies due to truncation, all subsequent labels are forced to be 0. In other words, we only label nodes using strings of the form $(l_1 l_2 \dots l_i 00 \dots 0)$ for some $i \leq k$, and $l_j \neq 0$ for all $j \leq i$. The number of such strings with i nonzero labels followed by $(k - i)$ zero labels is $(n-1)^i$. Therefore, the total number of

nodes is

$$N = \sum_{i=0}^{k} (n-1)^i. \tag{3.43}$$

Since $N = \Theta\left((n-1)^k\right)$, many quantities of a truncated hierarchy with a base graph of order $n+1$ have the same scaling with the number of nodes N as those for a non-truncated hierarchy with a base graph of order n.

In terms of the number of levels k, the maximum diameter will be proportional to k, just as it was in Sec. 3.3.1.5. It follows that the diameter scales with the total number of nodes as $\delta = \Theta\left(\log_{n-1} N\right)$ for a truncated hierarchy.

On the other hand, truncation offers a large improvement in the maximum degree of the hierarchy. As discussed in Sec. 3.2.2.5, the maximum degree of the truncated hierarchy is $\Delta(G^{\Gamma_{\vec{\alpha}} k}) = 2\Delta(G)$, which is constant in N.

The edge weight recursion relation is simply $n-1$ copies of the current graph and then new, additional edges:

$$w(G^{\Gamma_{\vec{\alpha}} i}) = (n-1) w(G^{\Gamma_{\vec{\alpha}} (i-1)}) + \alpha_i w(G). \tag{3.44}$$

This is identical to the recursion relation for the standard hierarchy, Eq. (3.34), except that there are now only $n-1$ copies, and also, for a given number of qubits N, the number of levels k may be different by constant factors and terms. Thus, the only modification to the recursion relation is to replace n with $n-1$, and the solution of the relation is otherwise identical. This means that none of the asymptotic scaling with k is affected, and the scaling with N is only affected by changing the total number of levels required to construct a graph of N nodes.

The recursion relation for weighted diameter is similar to Eq. (3.38). Due

to truncation, one needs to make a careful comparison of paths that do or do not terminate at the root node of the top level, but in any case the weighted diameter's scaling with k is the same as the non-truncated weighted diameter's scaling. The weighted diameter scaling with N can thus be found from Eq. (3.42), using the appropriate value of k for truncated hierarchies with N nodes.

3.3.2 Choosing Among Graphs

3.3.2.1 Graph Embeddings

The long list of comparisons summarized in Table 3.2 can make it difficult to see exactly when different graphs are preferable. To make our calculations more concrete, we would like to compare concrete scenarios for the connection of qubits arranged on a grid in d dimensions. Specifically, in each dimension ($d = 1$, 2, and 3), we examine a hierarchy that is embedded into the grid, comparing its properties to the same grid but with nearest-neighbor connections. We consider building modules where each small module is a complete graph of size n, laid out in cubes on the grid so that the side-length of the cube is $n^{1/d}$. The $d = 1$ and $d = 2$ cases with $n = 2^d$ are illustrated in Fig. 3.7.

As shown, the length of an edge must increase by a factor of $n^{1/d}$ (2 in Fig. 3.7) at every level of the hierarchy in order to make these hierarchies possible. Therefore, to determine the total length of wire used, we can use a cost weight with $\alpha = n^{1/d}$. Keeping factors of N only, Table 3.2 shows that for $d = 1$, we expect a total cost weight $\Theta(N \log_n N)$, while for the higher-dimensional cases we expect a total cost

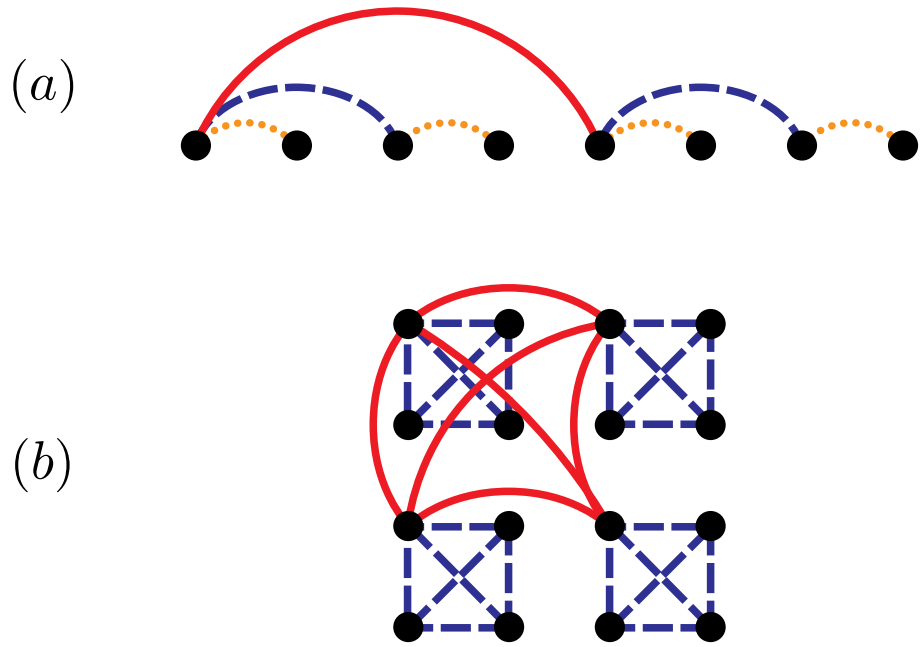

Figure 3.7: An illustration of the embedding of a hierarchy on a (a) one- or (b) two-dimensional lattice of qubits. In both cases, the length of an edge doubles at every level of the hierarchy, but the scaling in total edge length used changes from $\Theta(N \log_2 N)$ to $\Theta(N)$ when going from 1 to 2 dimensions. In $d = 3$, a similar hierarchy with doubling length scales connects modules of eight qubits.

weight $\Theta(N)$ [1]. For the d-dimensional grid, this total cost weight is always $\Theta(N)$.

Now, to consider the performance of the two graphs, we must fix a separate scaling factor for the time weight, β. There are several options which might be reasonable for different physical applications. If $\beta = 1$, i.e., all links act identically in terms of time required to traverse them, then the weighted diameter of the hierarchy is simply $\Theta(\log_n N)$. Another option would be to take $\beta = \alpha$, i.e., to assume that links take as long to move through as they are long. In this case, we find that the

[1] If, for the application at hand, a planar graph is required, cycles such as C_n can yield the same scaling.

hierarchy's weighted diameter scales as $\Theta\left(N^{1/d}\right)$, meaning that the hierarchy and nearest-neighbor graphs match in performance.

We may also want to allow hierarchies to make use of the "fat tree" concept to produce a better-performing graph [114]. Suppose that we allow ourselves to "spend more" on higher-level links, causing their cost weight to increase with a factor α, but improving their performance so that the time weight scales with the factor $\beta = 1/\alpha$. In this case, the question is what range of α allows for the hierarchy to perform better than the nearest-neighbor grid (lower time-weighted diameter) for less cost (lower total edge cost weight)? (Note that this cost weight includes any contribution from "lengthening" wires at higher levels of the hierarchy.)

To answer the first, we compare the two asymptotic diameter scalings, $N^{\max(0,\log_n 1/\alpha)}$ and $N^{1/d}$. This suggests that if $\alpha \geq n^{-1/d}$, the hierarchy will allow for faster traversal than the nearest-neighbor grid. However, we wish to avoid causing the hierarchy to have a total cost weight that scales worse than $\Omega(N)$, which requires $\log_n \alpha < 1$. We find that a winning hierarchy can be constructed if α lies in the range $\alpha \in \left[n^{-1/d}, n\right)$. The optimal α is as large as possible but less than n; at that point an additional logarithmic factor is introduced to the total cost weight scaling.

In these cases, we have not allowed the nearest-neighbor grid to modify the weight (either kind) of its links. This is because any modification in its cost or time weight enters simply as a constant factor; if the individual links have weight c instead of 1, the overall weighted diameter is just $cN^{1/d}$ while the total cost weight is just cN. Of course, one can apply different constants to each figure of merit, or apply c to one and $1/c$ to the other. In order to make the nearest-neighbor grid

match the performance of the hierarchy, the unit-length time weight would have to be $N^{\log_n(\alpha)-1/d}$ while the unit-length cost weight must not scale with N.

3.3.2.2 Pareto Efficiency

Our calculation of various graph parameters suggests that the hierarchy architecture offers significant advantages over others. One way to make this comparison more exact is to appeal to the economics concept of Pareto efficiency, which is used to designate an acceptable set of choices in multiparameter optimization [120]. A choice is Pareto efficient if switching to a different choice will cause at least one parameter to become worse. Suppose we eliminate all constants to focus only on the scaling with N for three parameters: weighted diameter, maximum degree, and total edge weight. By removing these constants, we assume that the small multiplicative factors they provide will not influence decision making. For simplicity, we will assume that both cost and time weights scale with the same factor, α.

For comparison, one could ask: what minimum number of edges is required for a graph on N nodes to have maximum degree Δ and diameter δ? Reference [121] answers this optimization question partially, and constructs what are known as *porcupine graphs* which achieve the optimum, illustrated in Fig. 3.8. We observe here that qualitatively, porcupines are modular, since they may be described by attaching trees to the nodes of a complete graph. In particular, the graph $K_{\sqrt{N}}\Pi S_{\sqrt{N}}$ is a porcupine graph that achieves a diameter $\delta = 3$ and a maximum degree of $\Delta = 2(\sqrt{N} - 1)$ with the minimal number of edges.

Graph		δ_w	Δ	w
K_N		const.	N	N^2
S_N		const.	N	N
C_N		N	const.	N
Square grid		$N^{1/d}$	const.	N
$\star$ $K_{\sqrt{N}} \,\Pi\, S_{\sqrt{N}}$		const.	$\sqrt{N}$	N
$\star$ $K_{n+1}^{\Gamma_\alpha k}$	$\alpha \neq 1$	$N^{\log_n \alpha}$	const.	N
	$\alpha = 1$	$\log_n N$	const.	N

Table 3.3: An illustration of the scaling with N of three key parameters to be used in Pareto optimization. Here δ_w is the weighted diameter, Δ is the maximum degree, and w is the total edge weight of the graph. A star ($\star$) has been placed next to the two graphs we find to be Pareto efficient. We have also included the $\alpha = 1$ (unweighted) hierarchy in the final row, as it has a different scaling for the weighted diameter. Our Pareto efficiency judgment is made assuming $n^{1/d} \geq \alpha \geq 1$.

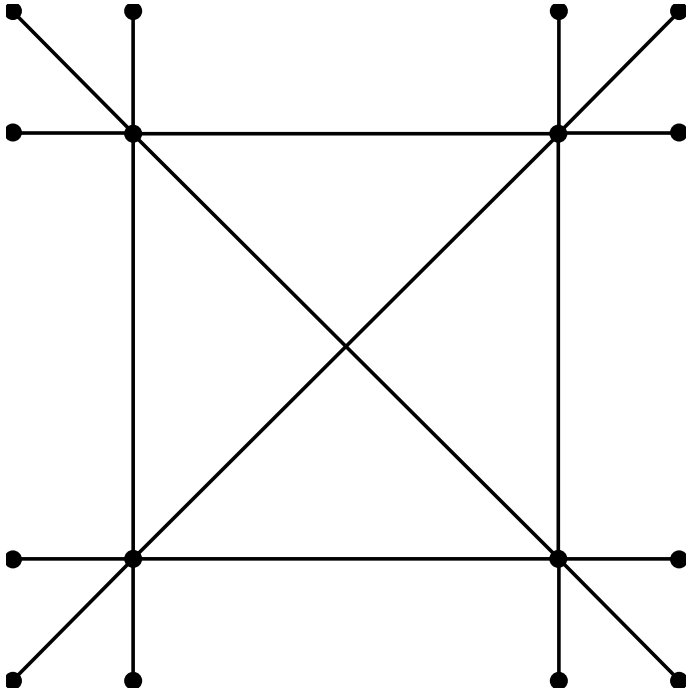

Figure 3.8: An example of a porcupine graph, in this case, $K_4 \amalg S_4$.

We summarize the scalings of these graphs in Table 3.3. Assume that $n^{1/d} \geq \alpha \geq 1$. In this case, we can find the Pareto-efficient solutions by noting which options can be eliminated. We see that K_N is strictly worse than S_N and can be eliminated; S_N is then dominated by the porcupine. C_N is dominated by the square grid, which has identical scaling of total weight and degree but lower diameter. The square grid, in turn, is dominated by the hierarchy due to the assumptions we have made on α. This means that the two Pareto-efficient choices in this case are the truncated hierarchy and the porcupine graph. If we chose any option besides these two, we could improve the scaling with respect to N without any trade-off by switching to one of them. While this framework does not offer a decision rule to choose between the porcupine and $K_n^{\Gamma_\alpha k}$, the latter is clearly preferable if our aim is to create a modular quantum system that does not rely on a few centralized nodes. We stress

that this optimization procedure is only intended to evaluate the quantities and graphs introduced, and the Pareto-efficient choices will change if other figures of merit or other graphs are included in the optimization.

3.3.2.3 Optimality of diameter for hierarchical graphs

The use of $K_n^{\Gamma_\alpha k}$ may be further motivated via the degree-diameter problem [122] (for a survey, see Ref. [123]). Given a graph with a maximum allowed degree Δ on each node and diameter no greater than δ, the degree-diameter problem asks for the maximum number of nodes $N(\Delta, \delta)$ that such a network could hold. This problem is practically well-motivated in the design of networks, and may be answered for special classes of graphs. The Moore bound, which is a bound for general graphs, states that the number of nodes N is at most $\frac{\Delta(\Delta-1)^\delta-2}{\Delta-2}$. This means that for a constant maximum degree $\Delta \geq 3$, the diameter satisfies $\delta = \Omega(\log N)$, meaning that hierarchical graphs have optimal diameter up to a constant factor. Tighter bounds on the number of nodes may be shown, for instance, when the *tree-width* of the graph is bounded. Ref. [124] shows that graphs with small tree-widths t and an odd diameter δ satisfy

$$N\left(\Delta, \delta; t\right) \sim t\left(\Delta - 1\right)^{\frac{\delta-1}{2}}. \tag{3.45}$$

As discussed towards the end of Sec. 3.2.2.1, hierarchies have low tree-widths. In particular, the tree-width of the truncated hierarchy $K_n^{\Gamma_\alpha k}$ is at most $n-1$. Next, the diameter of the truncated hierarchy $K_n^{\Gamma_\alpha k}$ is $\delta(k) = 2k-1$ (which is odd), and the maximum degree is $\Delta(k) = 2(n-1)$. Comparing the number of nodes in this

hierarchy $N(k)$ to the node capacity $N\left(\Delta(k), \delta(k); n-1\right)$ as in Eq. (3.45), we get

$$\frac{N(k)}{N\left(\Delta(k), \delta(k); n-1\right)} \gtrsim \frac{n^k}{(n-1)\left(2n-3\right)^{k-1}}. \tag{3.46}$$

Keeping the total number of nodes N fixed, consider two limits: one, a shallow hierarchy in which the number of levels k is $O(1)$, and two, a deep hierarchy, in which the size n of the base graph is $O(1)$ [i.e., $k = O(\log N)$]. We see that when the hierarchy is shallow, the right side of Eq. (3.46) is $\Theta(1)$, which indicates optimality. For a deep hierarchy, the above ratio scales as $2^{-\log_n N} = N^{\frac{-1}{\log(n)}}$, which is polynomially suboptimal. However, when $n = 3$, the ratio in Eq. (3.46) is again $\Theta(1)$, and the truncated hierarchy $K_3^{\Gamma_\alpha k}$ is degree-diameter optimal in this case.

3.4 Entangled State Construction

3.4.1 Setup

Although some of the graph properties calculated in the previous section give a heuristic sense for the capabilities of the hierarchical graph versus the nearest-neighbor or all-to-all graphs, we would like to examine their performance directly in terms of a quantum information processing task. The task we have chosen as a benchmark is the creation of a many-qubit GHZ state. Since this entangled state is difficult to create across long distances when using nearest-neighbor interactions, we hope that it can serve as a useful yet basic benchmark for processing quantum information with unitary evolution [24]. As shown in Chapter 2, preparation of a GHZ state also provides a means of transferring a state across the graph. Thus,

the results of this section also bound state transfer time. However, in this chapter, unlike Chapter 2, we focus on the use of discrete unitary operations (gates) rather than Hamiltonian interactions. This means that we cannot take advantage of the many-body interference which provided a speed-up in Chapter 2.

Using GHZ state creation as a benchmark for potential quantum architectures allows us to use physical limitations (represented by the Lieb-Robinson bound) to place computational limits on information processing. The GHZ state is directly useful on its own [23, 24, 46], but even in systems which do not directly produce the GHZ state, it is likely that quantum operations will require the creation of long-range correlations between distant sites. For example, the same physical bounds which govern the creation of the GHZ state also restrict the speed at which topological order can be produced [16]. We focus on the GHZ state as an easy-to-analyze example for the problem of creating these nonlocal correlations, but we stress that our results generalize to any state which possesses non-local correlations of the kind whose creation is limited by the Lieb-Robinson bound.

We adopt a framework in which every vertex of the graph represents one logical qubit, while an edge of the graph represents the ability to perform a two-qubit gate between nodes. For the purposes of this chapter, we assume that we can ignore single-qubit operations, instead focusing on the cost imposed by the required two-qubit gates between nodes.

3.4.2 Analytical Results for Deterministic Entanglement Generation

In order to create the GHZ state, we assume that we begin with all qubits in the state $|0\rangle$ except for one qubit that we place in the initial state $|+\rangle$. By performing controlled-NOT operations between this qubit and its neighbors, a GHZ state of those qubits is created. The state can be expanded by continuing to use further CNOT operations to expand the "bubble" of nodes contained in the GHZ state until it eventually spans the entire graph. For state transfer, we instead assume the initial state $|\psi\rangle$ to be transferred sits on one qubit, which is then transferred through the graph using SWAP operations until it reaches its destination.

We first consider a graph which has been assigned time weights, so that a gate between two linked edges can be performed deterministically in a time given by the weight of the edge between them. We assume that one node can act as the control qubit for several CNOT operations at once. Therefore, according to our protocol above, the time t_{GHZ} required to construct the GHZ state is found by identifying the qubit that will take the longest to reach from the initial qubit by hopping on the graph. A similar argument holds for the state transfer time.

This implies that a GHZ state can be created, or a state transferred, in time that scales like the (time-)weighted eccentricity of the node we choose as the initial $|0\rangle + |1\rangle$ state. However, if we take the further step in our analysis of maximizing over weighted eccentricities (identifying the worst-case starting node), then the time will simply be the weighted diameter of the graph as calculated in the previous section. Note that the difference between the best-case weighted eccentricity (the

weighted graph radius) and the worst-case weighted eccentricity (the weighted graph diameter) over all nodes is at most a factor of two – if we look at the midpoint of the path that realizes the graph diameter, its distance to the endpoints of the path is bounded by the radius – so from the perspective of how this time scales asymptotically with N, the two are interchangeable.

3.4.3 Numerical Results for Probabilistic Entanglement Generation

As shown in the previous subsection, in a deterministic setting of entanglement generation where a gate between two nodes of our graph H can be performed in fixed time, the time required to create a GHZ state is equal to the weighted diameter $\delta_w(H)$. However, in many situations in long-distance quantum information processing, probabilistic or heralded methods might be used instead. We might suppose that, in a small time step, the network succeeds in performing a desired two-qubit gate with probability p (and that we know whether the gate succeeded or not). Upon failure, one can try performing the gate again in the next time step without having to rebuild the state from the beginning. In this setting, we expect that the scaling will likely be similar to the deterministic case but more difficult to calculate exactly. Fortunately, it is easy to re-interpret the meanings of the edge weights to account for this.

The main complication arising from the inclusion of unitaries that do not get completed in a fixed amount of time is that multiple paths between two nodes can all contribute to the total probability that entanglement has been produced, making

it a harder problem to solve exactly. However, we can turn to numerical simulation to get an idea of the behavior. In the following, we define a new edge weight called the probability weight, p_{ij}, which is the probability of success of edge (i, j) in one time step.

The algorithm for simulating the creation of a GHZ state is as follows:

- At each time step t, identify the subgraph F of nodes that have already joined the GHZ state.
- For each edge between a GHZ node $i \in F$ and a non-GHZ node $j \notin F$, identify the probability edge weight p_{ij}. With probability p_{ij}, allow node j to join the GHZ state in the current time step, t.
- Once all edges have been tested, repeat the procedure for the next time step on the new, possibly larger, set of GHZ nodes.

A single number p_0 is chosen as the base probability, so that the probability weights on the lowest level are p_0, and edges on the i-th level of the hierarchy succeed with probability $p_0\alpha^{i-1}$. Note that we must fix $\alpha < 1$. As a first step toward evaluating the performance of a graph, we estimate its time weights as $w_{ij} = 1/p_{ij}$, the time required to perform a two-qubit unitary on average. The overall estimate of the expected time taken is then δ_T/p_0, where δ_T is the time taken for the deterministic case with time weights scaling by a factor $\beta = 1/\alpha$ at each level. We find that this predicts very well the rate at which the GHZ state can be constructed over a wide range of α values (Fig. 3.9). The expected time remains $\Theta\left(N^{\log_n(1/\alpha)}\right)$.

For graphs with multiple potential paths between two nodes, such as a two-dimensional grid, the expected time is not simply the deterministic time scaled by the extra time factor the probabilistic setup requires in each step. We can however still bound the expected time to build the GHZ state $\mathbb{E}[t_{\mathrm{GHZ}}]$ above and below for a graph H. We will bound it above by considering a modified graph in which the only path between the initial qubit and the qubit farthest from the starting point has distance $\delta_w(H)$. Such a path completes in time $\delta_w(H)/p_0$ on average. Since H has strictly more paths than this, the expected time will be lower. However, the shortest path between the initial and final qubits has total distance $\delta_w(H)$, which would take time $\delta_w(H)$ to complete even if $p_0 = 1$ and all gates were deterministic. Therefore, no path can finish faster than this, and the expected outcome over all possible paths cannot improve over $\delta_w(H)$. We can therefore write the following restriction on the expected time:

$$\delta_w(H) \leq \mathbb{E}\left[t_{\mathrm{GHZ}}\right] \leq \frac{\delta_w(H)}{p_0}, \tag{3.47}$$

where $\mathbb{E}[\cdot]$ denotes the expected value. This implies $\mathbb{E}\left[t_{\mathrm{GHZ}}\right] = \Theta\left(\delta_w(H)\right)$. Therefore, although the prefactor is difficult to calculate, we can tell that the time required to complete the creation of a GHZ state on the nearest-neighbor graph with $d = 2$ is $\Theta(\sqrt{N})$. This scaling implies that the condition for the hierarchy to outperform the nearest-neighbor grid in 2D is $\alpha \geq n^{-1/2}$, which is reflected in Fig. 3.9.

Using the GHZ-creation time and state transfer as examples, we can see many of the advantages of hierarchical graphs as network topologies. Such architectures are able to rapidly incorporate a very large number of qubits (exponential in the

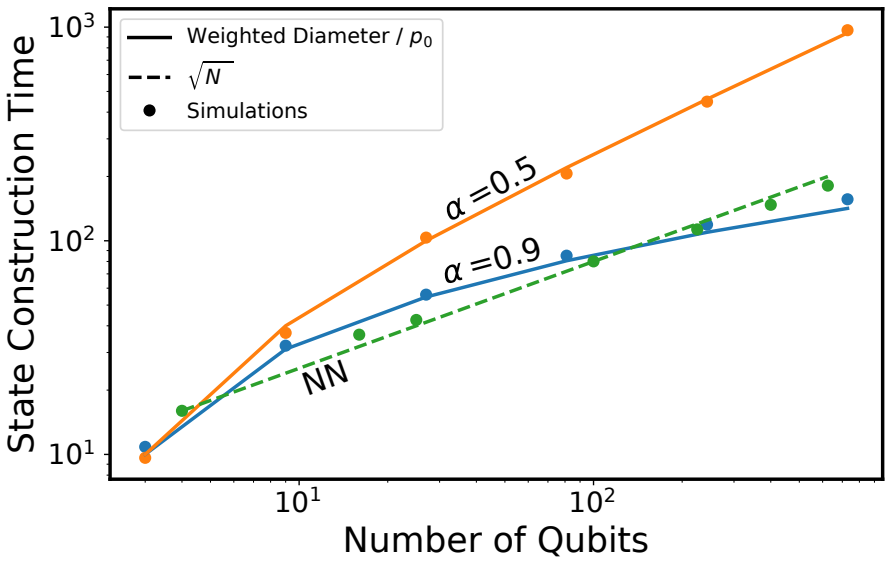

Figure 3.9: Graph-theoretic predictions and simulation of t_{GHZ} for the hierarchy $K_3^{\Pi_\alpha k}$ at various α, and a two-dimensional nearest-neighbor (NN) grid; $p_0 = 0.1$. The $\sqrt{N}$ fit shows the scaling of t_{GHZ} for the nearest-neighbor case, with a prefactor in the range suggested by the text's argument. Note that since $n = 3$, the crossover for the hierarchy to asymptotically outperform the nearest-neighbor grid is at $\alpha \geq 1/\sqrt{3} \approx 0.58$, which is seen in the numerical results. Code for generating this figure can be found at [125].

number of hierarchy levels), while the time-weighted diameter (and thus communication time) grows linearly with the number of levels. Since the weighted diameter is not substantially changed even if we use the truncated hierarchical product of Sec. 3.2.2.5, these benefits can also be realized in that setup.

3.5 Circuit Placement on Hierarchies

A final reason we believe hierarchies could be a useful way to organize modular quantum systems is that they may be able to take advantage of straightforward methods for circuit placement. Circuit placement is a problem that arises when

a quantum circuit or algorithm must be translated onto a physical system [126]. Suppose we are given a specification for a quantum algorithm in the form of a circuit diagram, and we wish to run that algorithm on a given quantum computer (which presumably has enough quantum memory to perform that algorithm). In order to translate the circuit into instructions for our machine, we must identify each algorithm qubit with a machine qubit and then determine how the individual quantum gates can be realized in our machine [2].

Circuit placement is an important part of the quantum software stack, just as the compilation to machine code is in classical computers. By placing qubits which must operate on each other often close together in the real-world machine, we can minimize the amount of time spent performing long-range quantum gates. However, this problem is generally quite difficult for arbitrary instances and in fact has been shown to be NP-complete [126].

However, since we are interested in the sub-problem of circuit placement on hierarchies, it is possible that the hardness results of Ref. [126] do not apply and the exact solution can be found in polynomial time, just as the problem can be solved tractably in linear qubit chains [127]. Whether or not an exact algorithm exists, we can appeal to heuristics to efficiently place circuits as well as possible. Such an approach is promising because hierarchies are extremely structured with clear

[2] We studiously avoid referring to the machine qubits as "physical" in this chapter, as we do not want to confuse this conceptual distinction with the physical/logical qubit divide in error correction. All of the qubits referenced in this section are logical qubits in the error-correcting sense.

prioritization of clustering between small groups of qubits, which can be recognized in the algorithm and matched to the physical architecture.

To explain further, we consider the following model. We suppose that we begin with a weighted circuit graph C with a vertex set V_C and an edge set E_C, in which an edge exists between two vertices if there is at least one two-qubit gate between them in the circuit, with the weight of the edge corresponding to the number of gates. We then specify a machine graph, M, with vertex set V_M and edge set E_M, in which each edge (u, v) indicates that the machine can perform two-qubit gates between u and v.

We now seek a mapping $f : V_C \to V_M$ that assigns algorithm qubits to machine qubits. A mapping f has a total cost found by considering, for every edge in E_C between vertices c_i and c_j, the shortest-path distance between $f(c_i)$ and $f(c_j)$ in M, multiplying that distance by the weight of the edge in C and summing over all edges. Thus, it captures the total distance that must be traversed by all gates in order to execute the circuit when the current mapping is used. Reducing this is expected to reduce the amount of time spent performing SWAP gates in order to connect two distant qubits. Performing this mapping is an important subroutine in any quantum programming framework, and at least one existing quantum compiler has a "mapper" phase that takes into account the actual graph that a program must be compiled onto [128, 129].

Our cost function is a choice made from convenience, and others are possible. Using this cost function ignores several important aspects of quantum circuits. First, our cost function does not account for the fact that a different mapping might allow

for more parallelism, since it evaluates the cost of each gate individually. In addition, we take the circuit graph C as a given, when in fact many different circuits exist for any given quantum operation. In fact, it is likely that optimization of C could be performed, possibly by using the structure of M itself. A more realistic model for circuit placement may require a back-and-forth in which a circuit is first placed, then optimized, then re-placed, and so forth. A more advanced placement algorithm may even permit the swapping of qubits throughout the circuit, thus optimizing the placement of the quantum algorithm without constructing a circuit connectivity graph as an intermediate step.

For this chapter, we will ignore these concerns and proceed with a heuristic approach to circuit placement for hierarchies. We describe our algorithm as "partition and rotate," as it requires these two basic subroutines. First, qubits are partitioned into sub-hierarchies by examining whether they are connected by many gates in C. This process continues recursively, with each partition being subdivided and so on until every qubit is identified with its point in the hierarchy. This top-down process is then followed by a bottom-up process in which each small cluster is rotated so that its most-communicative qubit is at the root of the sub-hierarchy, and then the partitions themselves are rotated, and then clusters of clusters, etc. Ideally, this results in a mapping in which every qubit is (a) placed close to qubits it needs to communicate with and (b) placed in easy access to other modules if that qubit requires such access. We will now explore in detail these subroutines and the circuit speed-ups that result. We will place algorithms on a machine graph M which we take to be defined by $K_n^{\Pi k}$ for some integer k. Note that we examine unweighted

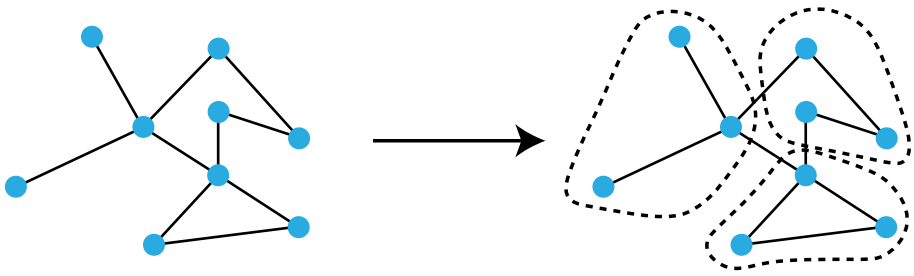

Figure 3.10: Illustration of how we might divide a hypothetical graph into smaller clusters. This process is repeated many times, recursively.

hierarchies, but these methods can be applied to weighted hierachies as well.

3.5.1 Partitioning

For the first step of our algorithm, we wish to divide the computational graph C into n subgraphs which are as disconnected as possible. In addition, since we wish to assign each node in C to physically separate and limited qubit registers, it is important that each of the subsets has precisely $|C|/n$ nodes. This problem is known as *balanced graph partitioning*, and the problem of finding the optimal solution is NP-complete for $n \geq 3$ [130]. However, heuristic methods exist which approximate the solution, and are widely used in the field of parallel computing and circuit design [131]. We have illustrated this process in Fig. 3.10.

Our method for performing circuit placement on hierarchies relies on a subroutine that performs balanced graph partitioning. There are many algorithms and software packages from which to choose. Here, we have used a software package called Metis, which implements an algorithm called recursive bipartitioning [131].

We begin by supposing that we have the circuit graph C and we wish to

identify groups of $|C|/n$ nodes which have high connection to each other but low connection outside of the group. This is accomplished by finding a balanced graph partition in which the weight of the edges connecting each group is minimized. If we call the initial set of all nodes S, then we wish to identify subsets $S_0, S_1, \ldots, S_n$. In terms of the addressal scheme of Sec. 3.2.2.3, all the nodes in set S_i will have have digit i in their base-n representation. In the next section, we will discuss the choice of which digit to assign to each set.

Once the subsets S_i are found, partitioning can be run again on that relevant subgraph, creating n new subsets of this subset. Eventually, every node in the graph will be identified with a lowest-level module of size n, a next-level module of size n^2, and so forth.

Here we have used a generalized, pre-existing algorithm for graph partitioning. It is possible that the specifics of this problem, and the possibility of co-designing the precise quantum circuit implementing the algorithm (and thus C) with the architecture, enable more specific, better-performing approaches.

3.5.2 Rotation

Drawing partitions between qubits is not enough to fully specify their placement into a hierarchy. If we consider using the i-digit representation, we can imagine that partitioning essentially describes the process of deciding, from a set of qubits, which ones will share a digit in the next level. However, these digits are more than arbitrary markers, because there is one node in any sub-hierarchy which connects

Figure 3.11: An illustration of how and why the process of rotation works in our circuit placement algorithm. In this diagram, red links represent gates to be performed (edges in C) and black ones are available communicative links (edges in M). In the graph C, the qubits 1, 2, and 3 are all connected, and 3 is connected with 4. These qubits have been correctly placed into clusters $(1, 2, 3)$ and (4). However, if they are not rotated correctly (see left), the link between 3 and 4 can become quite long, necessitating a long-range quantum gate. By properly rotating (right), the gate between links 3 and 4 becomes much shorter, improving the placement.

to the hierarchy above. This node (which we say has digit 0) has privileged access to communication with other sub-hierarchies. Therefore, in order for our circuit placement to succeed, we should ensure that the qubit on top of each sub-hierarchy is the one which requires the most access.

In order to do this, we implement a second subroutine, the "rotate" part of the algorithm. This is called rotation because, once we know which qubits will be together in a module, we must choose how to orient them relative to the larger modular structure. Whereas partitioning is top-down (the full graph is broken into small subgraphs which are then themselves partitioned), rotation is bottom-up. Suppose the modular structure is $K_n^{\Pi k}$. We begin with sets of n qubits and must choose which will be the top of each smallest instance of K_n. We then take each partition of n instances of K_n and decide which instance of K_n will connect to the

next level up, and so on. This process is illustrated in Fig. 3.11.

Note that the general structure of our algorithm is to first go down the hierarchy, partitioning nodes, and then to go up, re-arranging sub-hierarchies in the proper order. We perform this procedure only once to obtain our circuit mapping.

3.5.3 Results

Now that the placement algorithm is specified, we turn toward examining its performance on quantum circuits. We consider two separate questions. First, we investigate whether the algorithm is effective – does it actually reduce, relative to a random assignment, the amount of distance that must be traversed in a circuit to execute all the requested gates? Second, we will examine whether the algorithm executes efficiently on a classical computer. This second point is important because in general the problem can be solved by brute-force search, but such a search requires a time $\mathcal{O}(N!)$ to perform (although, as we stated earlier, it is possible that an exact algorithm exists with a lower time cost for the special case of hierarchies).

To investigate the above concerns, we examine the algorithm's performance on random circuits. For each trial, we first generate a random circuit of N_g two-qubit gates on N total qubits. The precise type of two-qubit gate is irrelevant in this framework. Likewise, single-qubit gates require no communication overhead, so we do not consider them. The random circuit then implies a computational graph C, where, as described above, the vertices represent the algorithm qubits and the edge weights represent the number of gates that must be applied between each pair

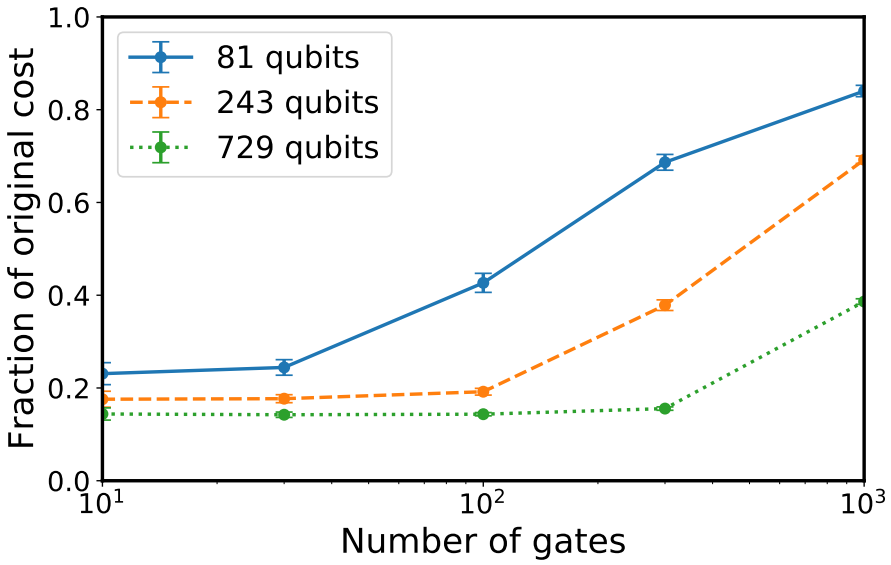

Figure 3.12: Plot of the average ratio (total gate distance after partition-and-rotate)/(total gate distance before) given 100 trials each for different numbers of random gates and random qubits. Error bars represent one standard deviation. As the number of gates begins to saturate the number of qubits, the possible improvement from optimization begins to decrease.

of qubits. Once this computational graph has been generated, we first attempt to map it blindly to the hierarchy graph, using the addressing scheme of Sec. 3.2.2.3 and an arbitrary order of the graph C. Then, we apply partition-and-rotate and calculate the new cost function. By comparing the cost function between these two, we develop an idea of how much long-range quantum information processing is eliminated by partition-and-rotate. We perform this several times to build up statistics on average time costs and average improvement. Code which performs circuit placement and generates the profiling figures included in this section can be found at [125].

In our simulations, we test hierarchies $K_3^{\Pi k}$ up to 729 qubits ($k = 6$). We find that as gates are added, the improvement over the initial cost is decreased.

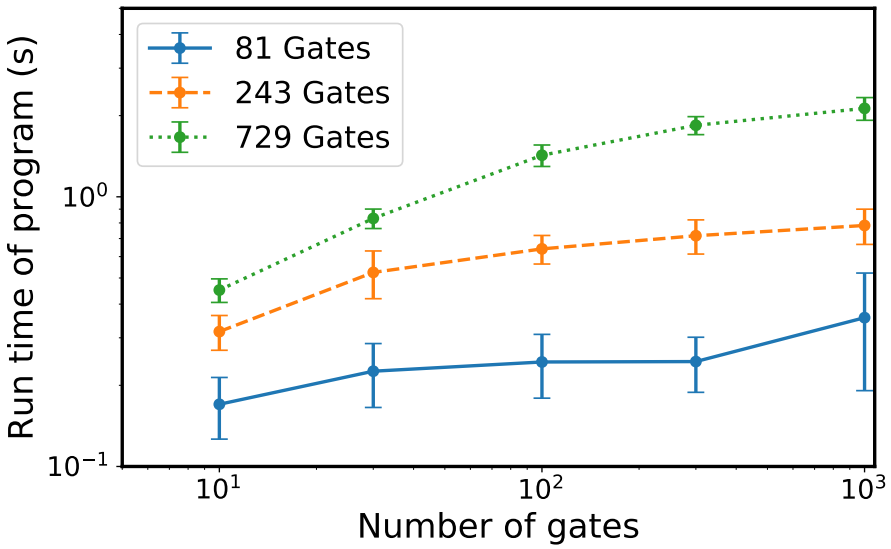

Figure 3.13: Average run times over 100 trials for partition-and-rotate on a 2015 MacBook Pro with a 2.6 GHz processor. Each line represents an increasing number of gates for a constant circuit size as measured by the number of qubits. Error bars represent one standard deviation.

This is sensible, because as more randomly placed gates are present, different node mappings become more similar. Such an effect will likely not be present for quantum algorithms which do not have their gates placed randomly. For cases in which the number of gates is significantly fewer than the number of qubits, partition-and-rotate is able to significantly reduce the cost function. We find that 100 gates can be placed on a 729 qubit hierarchy with a total cost less than 20 % of the original on average. When 1000 gates are placed on a 729 qubit hierarchy, the final cost is still only 40 % of the initial one. Results for $K_3^{\Pi 4}$, $K_3^{\Pi 5}$, and $K_3^{\Pi 6}$ can be seen in Fig. 3.12.

Next, we examine the time required to place such a circuit. Our code, most of which is written in Python3 but which uses a C implementation of Metis for graph partitioning, can place 1000 gates on a 729-qubit hierarchy in roughly two seconds

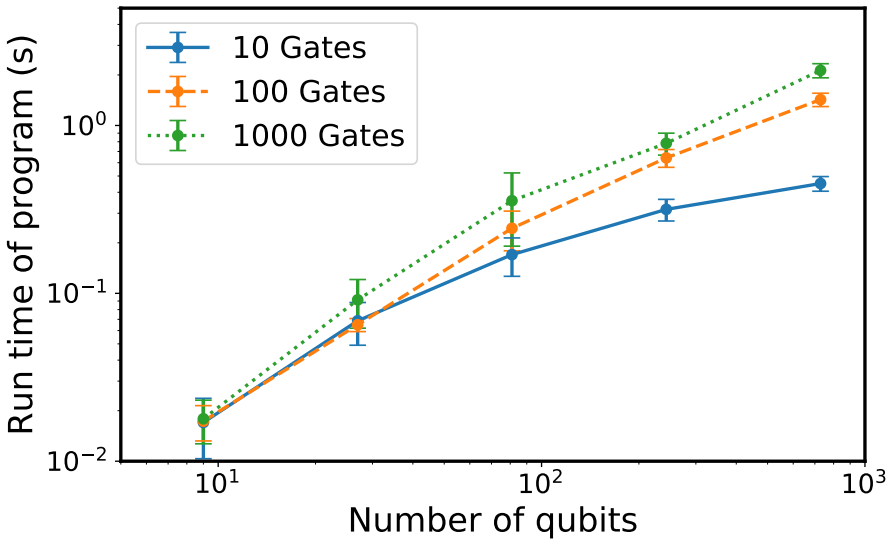

Figure 3.14: Average run times over 100 trials for partition-and-rotate on a 2015 MacBook Pro with a 2.6 GHz processor. Each line represents an increasing number of qubits for a constant number of gates. Error bars represent one standard deviation.

when running on a 2015 MacBook Pro. Although the algorithm seems naturally suited to parallelization, our implementation uses only a single core. Our current implementation appears to scale with the number of qubits as $\mathcal{O}(N)$ and not to depend on the number of gates included at all once there are a sizable number of gates. We illustrate these two relationships in Figs. 3.13 and 3.14. These times compare favorably to the times reported in Ref. [126], with much optimization still possible in our implementation.

Note that using random graphs as described above means that our results may not be valid for more general quantum algorithms. It is possible that practical quantum algorithms have structure that makes them either particularly amenable or particularly difficult for partition-and-rotate algorithms to place, depending on the actual algorithm being examined.

3.6 Conclusions and Outlook

In this chapter, we have developed the theory of hierarchies using the existing binary operation of the graph hierarchical product. We have shown that hierarchies may be a promising architecture for large quantum information processing systems. To demonstrate this, we analyzed both properties of the underlying graph (such as diameter, maximum degree, total edge weight) as well as the time it would require to perform a representative quantum information process (constructing the GHZ state/state transfer) in both deterministic and probabilistic settings. We have also computed and tabulated these properties for many other graphs which appear as potential architectures, for comparison. We have shown that, for much of parameter space, hierarchies have favorable scalings in cost and performance with the total number of qubits N compared to these competitors. Also, since hierarchical graphs are hyperbolic, they share many of the advantages of hyperbolic graphs such as efficient routing schemes [132], network security [133], and node addressal [134].

We have also presented a conceptually simple circuit placement algorithm which allows for simple optimization using existing graph-partitioning software packages. Our partition-and-rotate algorithm scales well with the number of qubits and gates in the circuit and reliably reduces the total distance that needs to be traversed by random quantum circuits, which we verified by simulation.

One significant limitation of our analysis in this chapter has been that we remained confined to unitary operations. Non-unitary operations (for instance, measurements which are then fed forward to choose future unitary operations) are

capable of establishing long-range correlations like those present in the GHZ state much more quickly than unitary ones if measurements and classical communication are fast. In Chapter 4, we extend our results into non-unitary domains.

In addition, we have made the assumption that the primary way in which quantum architectures will differ is the speed with which two qubits can communicate (as represented by our time weights on edges). Another important case might be one in which the primary way edges are enhanced is by improving bandwidth or duplicating nodes to provide parallel routes rather than affecting gate speed directly. For some schemes, our abstract notion equating the time of a two-qubit gate with the edge weight may still be a useful tool of analysis, but in other cases bandwidth and speed may not be interchangeable. In Chapter 4 bandwidth, rather than latency, will be our primary concern.

In this chapter, we limited ourselves to consideration of a few quantum processes (generation of a large entangled state or transfer of a state across the graph), which might not be representative of other, more general distributed quantum information tasks. Some algorithms, such as Shor's algorithm, are known to be able to run with little additional overhead even on one-dimensional, nearest-neighbor graphs [135]. Therefore, when selecting an architecture for a practical quantum computer, care will need to taken to select the proper benchmarking task.

In future work, we hope to look at a wider variety of quantum circuits and use those to better benchmark different modular architectures. In addition, we hope to gain a better understanding of the treatment of probabilistic links for general graphs. For instance, as we discussed briefly when assessing the star graph S_N, one

real concern in a networked setting is whether some parts of the network will form bottlenecks. To analyze the impact of this in a general way will require a better understanding of realistic quantum algorithms and the demands they place on a network. Analyzing more complex quantum algorithms could also shed light on the performance of partition-and-rotate placement algorithms in realistic settings when sequencing and scheduling also enter into consideration.

Finally, in addition to asking ourselves how current circuits and algorithms can be executed on highly modular systems, we also hope to explore the possibility that highly modular architectures open up new possibilities for parallelized quantum algorithms. For instance, Ref. [136] shows that quantum fan-out gates can be used to parallelize gate sequences, decreasing the time to perform an algorithm at the cost of requiring additional memory qubits. Hierarchies could implement such schemes by using high-level connections to perform the initial fan-out gates and then performing the various parallelized operations in each individual module.

Chapter 4: Entanglement Entropy and Non-Unitary Computations

4.1 Introduction

As the development of quantum computers progresses from the construction of qubits to the construction of intermediate-scale devices, quantum information science has increasingly begun to explore schemes for scalable quantum computing [22, 80, 83]. Unlike the situation presented in introductory textbooks, where quantum circuit diagrams are drawn without concern for connectivity, in reality a quantum architecture will limit the connections available in realistic quantum devices. Researchers have quantified the cost imposed by moving from one architecture to another [118], as well as optimizing the placement of qubits on a fixed architecture [127]. Experimentalists have also begun to test different architectures in the real world, for instance, Ref. [119] presents a comparison of the same algorithms run on small quantum devices that differ in connectivity and Ref. [137] discusses the task of compiling a circuit on ion trap quantum computers.

In this chapter, we are interested in identifying desirable candidate architectures directly, rather than adapting algorithms to architectures which are fixed beforehand. The construction of quantum networks is particularly well-studied in relation to building graph states, which are a valuable resource for quantum commu-

nication [91,138]. However, here we are most interested in constructing architectures for computation rather than communication. Researchers in quantum computing and information need to develop tools to evaluate the appropriateness of different architectures and their strengths and weaknesses. Metrics should allow us to understand both sides of a cost-benefit analysis, quantifying both some aspect of how hard a given network is to construct and how capable it is of performing information processing tasks.

One option for evaluating architectures would be to attempt to compile quantum algorithms to low-level instructions limited by the connectivity of the architecture, as can be accomplished by already-existing quantum software packages [128,129]. Afterwards, the circuit depth after compilation of benchmark algorithms could serve as a useful metric. Although this approach is usefully connected to direct application (the main reason we care about architecture design is to run algorithms, after all) it suffers from several technical issues. First of all, it is not clear what algorithms to use as benchmarks, and indeed it may be highly application-dependent. In a machine implementing Grover's algorithm for searching an unsorted dataset, testing the architecture in terms of a compiled example would require specifying a single oracle which may or may not be representative of the full range of oracles the machine may encounter once deployed. Second, our evaluation of a quantum architecture will be dependent on a particular compilation of the algorithm rather than strictly a property of the architecture itself. As the science of quantum compilation is still in its infancy, a compiled example may not be a reliable test of the quantum computer. It is desirable to instead produce a metric founded in physical results

and limitations if possible, which will avoid these compiler-based issues.

In Chapter 3, we used the diameter of the graphs as one diagnostic of performance, reasoning that graphs with low diameter would take less time to establish long-range quantum correlations. This limitation was suggested by the Lieb-Robinson bound, which can be applied to graphs to limit the generation of entanglement between distant sites [15, 16, 90]. However, the Lieb-Robinson bound as originally formulated applies only to unitary evolution. Although it can be extended to cases where the system evolution is governed by a master equation including dissipation [139], it does not apply to systems undergoing active measurement and feedback. Quantum computers must be able to perform non-unitary operations to meet the DiVincenzo criteria for initializing qubits in known states and measuring qubits at the end of the computation [140]. In particular, the ability to perform measurement and feedback is necessary for most schemes that perform error correction in scalable quantum computers. (However, autonomous error correction suggests this is not an absolute necessity, see Refs. [141, 142].) Although it might not be required in every case, it is clear we ought to develop a framework to evaluate architectures for the general class of quantum devices that are capable of measurement and feedback.

In this chapter, we develop a more sophisticated approach, providing a new physically-motivated benchmark for quantum architecture. We will use results from entanglement generation [143] and quantum channels literature [144] to establish a minimum amount of time required for a particular state to be generated on a quantum computer specified by a connectivity graph G, and in turn we will connect

this to a graph-theoretic quantity called the Cheeger constant or isoperimetric number [145]. We will propose a metric based on entanglement measures and show that when several physical models are represented by a graph G with edges corresponding to two-qubit interactions, a useful metric is given by what we dub the "rainbow time,"

$$\tau_{\mathrm{rb}}(G) = \max_{F \subset G, |F| \leq \frac{1}{2}|G|} \frac{|F|}{|\partial F|}, \tag{4.1}$$

We will then apply this analysis to the hierarchies described in Chapter 3 and discuss the possibility that $\tau_{\mathrm{rb}}(G)$, a bound on the speed with which maximally-entangled states can be created, is saturable.

4.2 Physical Setup

In this chapter, we describe a quantum architecture by a graph G with vertex sets V and E, which we write $G = (V, E)$. Each edge of the graph also possesses a weight function, w, which we will sometimes write $G = (V, E, w)$. We assume every graph introduced is weighted, connected, and has at least two vertices. Every vertex in the graph represents a single data qubit, and an edge exists between two vertices if two-qubit gates can be performed between them. We write the weight of an edge between i and j as w_{ij}. We interpret the edge weight as representing bandwidth, so that higher-weighted edges are capable of performing more gates in a single unit of time.

There are several possible physical interpretations of the graph model. One could interpret the edges as representing the distribution of entangled pairs [146],

as the number of two-qubit unitaries that can be performed in a gate model, or as a Hamiltonian system, with a graph edge describing the maximum operator norm of the interaction term between two qubits. We will focus on the first example as it is the simplest to analyze, but we will also present results for the other two.

In a Bell pair exchange model, graph vertices contain one data qubit, with other qubits within the module serving as auxiliary communication qubits, generating entanglement which is then used as a resource. In every interval of time t, each of these auxiliary qubits is able to form a Bell pair with a qubit in another module as dictated by the graph structure. These Bell pairs can then be consumed to perform two-qubit gates using teleportation [147,148]. Our graph G can be found by simplifying the "microscopic" model which includes all of the physical qubits into a more abstract representation that focuses on connectivity between data qubits, as illustrated in Fig. 4.1.

We assume that arbitrary unitaries and measurements within a module can occur arbitrarily quickly and that these operations can depend on the results of measurements from other modules. This is justified if the timescale t required to generate entanglement is much larger than the time required for intra-module unitaries, one-qubit measurements, and classical communication.

In this framework, several graph properties have clear physical interpretations. The degree (or valency) of a vertex now corresponds to the number of ancilla auxiliary qubits required in that module for communication overhead. Therefore, penalizing a graph for its maximum degree limits the size of an individual module. Similarly, adding up the total weight of all edges also gives a measure of the required

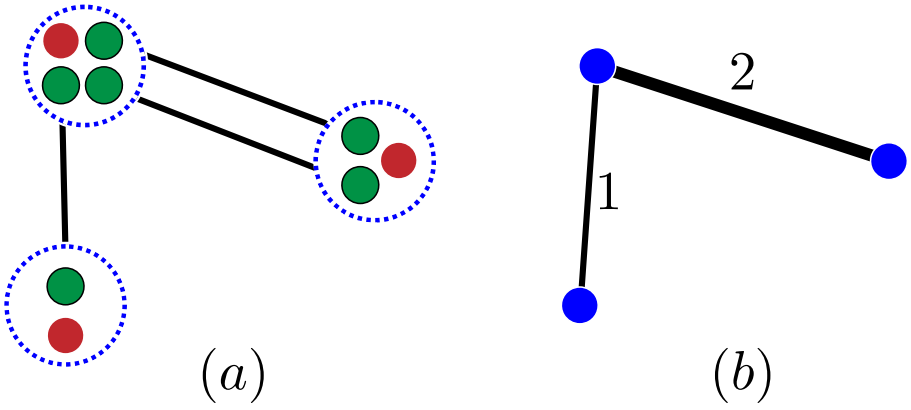

Figure 4.1: Illustration of how a model with communicator qubits can be abstracted into one in which only data qubits and edge weights are tracked. In (a) communicator qubits (green with black outline) form Bell pairs with other modules (blue dashed circles), in (b) the module as a whole is represented by blue circles, while the communicator ancilla qubits are now represented by edge weights. The remaining state that is tracked is the state on the data qubits, the red circles in (a).

communication overhead in terms of ancilla qubits. This framework provides a clear physical motivation for several cost functions used in Chapter 3.

In Chapter 3, we used the Lieb-Robinson bound to limit the time required to generate a Greenberger-Horne-Zeilinger state, which allowed us to connect an architecture's performance to its graph diameter. In this chapter, we will instead use a bound on entanglement generation to assess the performance of a graph in creating highly entangled states. This new bound is more robust becaue it does not assume the use of only unitary operations.

4.3 Entanglement Capacity

In Chapter 3, we connected the graph diameter to the time required to build a multi-qubit GHZ state. We were able to make this connection due to our assumption that the graph described a set of allowable unitary operations. We then noted that the GHZ state has maximal connected correlations in spin for every qubit, $\left\langle \sigma_i^z \sigma_j^z \right\rangle - \left\langle \sigma_i^z \right\rangle \left\langle \sigma_j^z \right\rangle = 1$ for all i and j. The Lieb-Robinson bound tells us that if two qubits are separated by a distance L, it takes a time proportional to L to create this perfect correlation [16]. This allowed us to establish a benchmark task (the creation of the GHZ state) whose performance could be limited on physical grounds.

However, if we allow for fast measurement, classical communication, and feedback, it is possible to construct an algorithm which can create the GHZ state of N qubits in time that does not scale with N. This occurs because we use measurement and feedback to establish correlations. The algorithm is:

1. Determine an ordering of the qubits, i, from 1 to N.
2. Put every qubit in the initial state $|+\rangle$.
3. Measure the product $Z_i Z_{i+1}$ for $i = 1$ to $N - 1$.
4. From the measurements, determine whether every qubit i is aligned or anti-aligned with the first qubit.
5. Perform single-qubit X gates on every qubit anti-aligned with the first one.

If only single-qubit measurements are allowed, some qubits can be set aside as

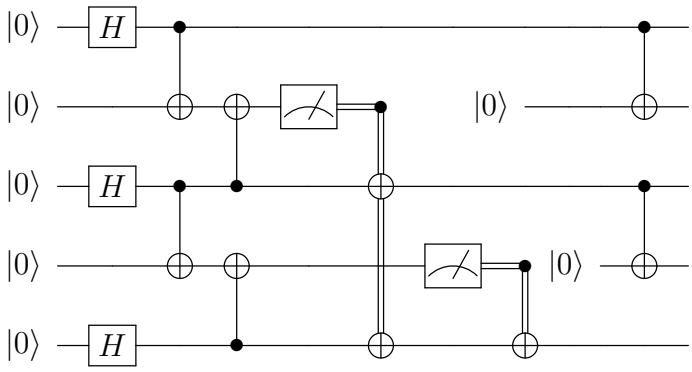

Figure 4.2: A quantum circuit on five qubits that creates the GHZ state. The second and fourth qubits are used to hold two-qubit ZZ measurements and then reset after being measured. The measurement results are used to bring the rest of the chain in alignment with the first qubit. Notice that only two layers of two-qubit quantum operations occur: the first to make all of the two-qubit measurements and the last to absorb the measurement qubits into the final GHZ state. This scheme can be extended for any number of qubits without requiring additional quantum circuit depth, only increasing the classically-controlled single qubit operations.

ancillas to assist in making the measurements, and then added to the GHZ state afterwards. We illustrate a realization of this protocol on five qubits in Fig. 4.2. This type of measurement-and-feedback approach to the creation and manipulation of large entangled states has also been applied to error-corrected color codes [149]. We conclude that the GHZ state – when we allow for non-unitary operations – is a poor benchmark task, and the graph diameter itself, although probably still a useful quantity to minimize, cannot be directly linked to minimum state creation time.

This suggests that rather than the non-local correlations which we targeted by proposing the GHZ state, we must find a different measure for network capability. In the following subsections we will derive the entanglement capacity for several

different physical models that can correspond to a graph. Consider a graph, $G = (V, E)$, and select a subset of the graph, F. We then want to show that the maximum amount of entanglement that can be created between F and $\bar{F}$ in unit time is proportional to the size of the boundary, $|\partial F|$,

$$|\partial F| = \sum_{i \in F, j \in \bar{F}} w_{ij}. \tag{4.2}$$

We will allow arbitrary constant factors, and discuss how this bound arises in three different physical situations. We quantify this capacity by considering any entanglement measure S on two regions so long as S obeys the following axioms:

- Additively distributive over the tensor product, so $S(\rho \otimes \tau) = S(\rho) + S(\tau)$ if ρ and τ are supported on both sides of the bipartation.
- Zero for states which are a product of states on each region, $S(\rho_F \otimes \rho_{\bar{F}}) = 0$.
- Non-increasing after any operation which is local to each region, even if we permit classical communication.

Entanglement measures which obey these conditions include the entanglement cost, the distillable entanglement, and the entanglement of formation [144]. All of these measures are identical to the entanglement entropy for pure states. First we show how to apply these axioms to the analysis of a case in which computation is performed by the production and consumption of Bell pairs. We will also look at a gate model of computation and a case in which the graph describes the limits on a time-dependent interaction Hamiltonian.

4.3.1 Bell Pairs

For this case, the edge w_{ij} between two nodes denotes that w_{ij} Bell pairs can be distributed between node i and node j in one unit of time. These can then be consumed to perform unitaries or simply distributed elsewhere in the graph through entanglement swapping. This is the case outlined in Fig. 4.1.

We can assume that the state after t time steps can be written as ρ_t. Since the only non-local operation is distributing Bell pairs, the entanglement of state at time $t+1$ can be bounded by the entanglement of ρ_t plus the new entanglement from introduced Bell pairs:

$$S(\rho_{t+1}) \leq S\left(\rho_t \otimes \rho_{\text{Bell}}^{\otimes|\partial F|}\right) \tag{4.3}$$

$$\leq S(\rho_t) + |\partial F|\, S(\rho_{\text{Bell}}) \tag{4.4}$$

$$\implies \Delta S \leq |\partial F|\, S(\rho_{\text{Bell}}) \tag{4.5}$$

This completes the proof, with the entanglement of the Bell pair under the measure in question providing the proportionality constant.

4.3.2 Unitaries

In this model, each graph edge of weight w_{ij} represents the capability to perform w_{ij} unitaries between qubits i and j in a time step. These unitaries are freely chosen by the experimenter.

We note that every two-qubit unitary can be performed using two Bell pairs as a shared resource and applying local operations. This can be easily seen in the

following process:

1. Alice and Bob start with a data qubit each and two Bell pairs shared between them. They wish to implement an arbitrary two-qubit unitary using only local operations and classical control.

2. Alice uses one Bell pair and classical communication to teleport her qubit to Bob.

3. Bob uses his local operations to perform the desired two-qubit gate.

4. Bob teleports Alice's qubit back to her.

Therefore, the state ρ' can be obtained from the state ρ by using LOCC and consuming up to $2\,|\partial F|$ Bell pairs in the process. Since LOCC cannot increase S, it follows that:

$$S(\rho') \leq S\left(\rho \otimes \rho_{\text{Bell}}^{\otimes 2|\partial F|}\right) \tag{4.6}$$

$$\implies \Delta S \leq 2\,|\partial F|\, S(\rho_{\text{Bell}}). \tag{4.7}$$

This suggests that the ability to perform arbitrary unitaries is potentially twice as powerful as the ability to distribute arbitrary Bell pairs, which makes sense, as an arbitrary two-qubit gate cannot necessarily be performed with one Bell pair (for instance, SWAP requires two) [150]. In any case, this still yields an entanglement capacity $\Delta S = \mathcal{O}(|\partial F|)$ bound as desired.

4.3.3 Hamiltonians

We will now consider a case in which the graph describes a two-body Hamiltonian, possibly time-dependent. The graph will restrict the strength of these Hamiltonians. If we assume that $G = (V, E)$, then the Hamiltonian can be written as a sum over the two-qubit operations:

$$H(t) = \sum_{(i,j)\in E} h_{ij}(t). \tag{4.8}$$

We then impose the condition:

$$\forall t : \|h_{ij}(t)\| \leq w_{ij}, \tag{4.9}$$

where w_{ij} is the i-j edge weight.

We can then apply the "small incremental entanglement" (SIE) theorem [143], in particular, we can apply the special case used in Ref. [151] to bound the total amount of entanglement generated by this Hamiltonian. If H is a sum of pairwise Hamiltonians h_{ij} acting on qubits, then the time-rate of entanglement generation on a set F of sites is:

$$\left|\frac{\mathrm{d}S_F}{\mathrm{d}t}\right| \leq 36 \log 2 \sum_{i\in F, j\in \bar{F}} \|h_{ij}\|. \tag{4.10}$$

Here, S_F is the entropy of the reduced density matrix on the region F, and *not* a general measure of entanglement as in the other two cases. This can be derived from Eq. 3 of Ref. [151] and specifying two-body terms a local Hilbert space dimension of two, but the result could be extended to qudits or to general k-body interactions. The sum over Hamiltonian norms, in the graph context, corresponds to a sum over graph edges, meaning that $\sum \|h_{ij}\| \leq \sum w_{ij} = |\partial F|$. Therefore, we can specifically

say that for this case, $\Delta S_F = \mathcal{O}(|\partial F|)$. Many other entanglement measures, such as entanglement of formation or entanglement cost, can be related to the entanglement entropy [152]. For example, many entanglement measures on mixed states can be defined as a weighted sum over pure state components; since none of the pure states can increase dramatically in entanglement under this process, the entanglement measure on the mixed state is similarly limited.

Having proven a limitation on our entanglement capacity, we will now define a final state ρ_f with a large entropy of entanglement across a specified bipartition. As this entanglement measure obeys the conditions we outlined above, we can calculate the time it will take to generate this state on a system with an arbitrary connectivity graph. In the next section, we specify the state of interest. Although we will discuss the time constraints in terms of the final pure state's entanglement entropy, the analysis of this subsection shows that we can use more general entanglement measures which coincide with entanglement entropy for pure states, meaning that our bound will apply even to protocols that produce mixed states as an intermediate or final case.

4.4 Rainbow States

In order to take advantage of the entanglement capacity results, we wish to establish a particular task in quantum information processing which requires the creation of a highly entangled state. Entanglement makes a useful benchmark for any quantum computer because it can be shown that computations without much en-

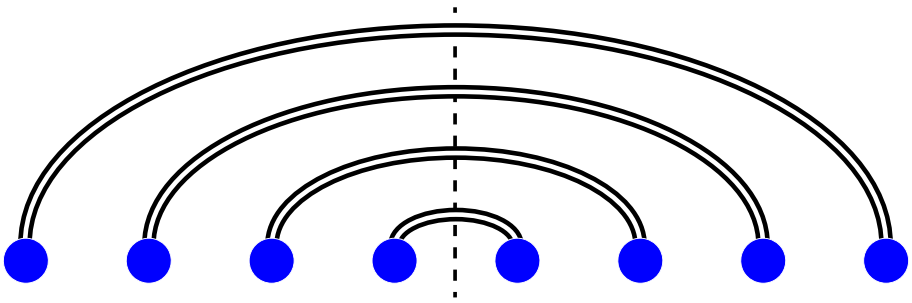

Figure 4.3: An illustration of the rainbow state in 1D, where doubled lines between qubits represent maximally entangled pairs. The dashed line shows the partition that the entanglement entropy is evaluated with respect to.

tanglement can be simulated efficiently classically, meaning any significant quantum speedup requires the generation of entanglement [20]. Note that from the standpoint of entanglement entropy, the GHZ state is *not* very entangled, even though all the qubits are maximally correlated. If any subregion A is traced out, the reduced density matrix is always a two-component classical mixture with $S_A(\rho_{\text{GHZ}}) = \ln 2$. Even an arbitrarily large GHZ state of N qubits does not possess much entanglement entropy.

Instead, consider the so-called "rainbow state," which is defined on N qubits in 1D contexts for even N [153] as

$$|\psi_{\text{rb}}\rangle = \bigotimes_{i=1}^{N/2} \frac{1}{\sqrt{2}} \left(|0\rangle_i |0\rangle_{N-i} + |1\rangle_i |1\rangle_{N-i} \right). \tag{4.11}$$

This state is easiest to understand graphically, as in Fig. 4.3. Each qubit is half of a Bell pair with a qubit on the opposite side of a partition that evenly divides the qubits.

Rainbow states have entanglement entropy that scales with the total volume of the state, $S_A(|\psi_{\text{rb}}\rangle \langle\psi_{\text{rb}}|) = N \ln 2$, where A is taken to be qubits 1 through $N/2$

in 1D. This "volume law" scaling on entanglement entropy is the maximum possible, as can be seen from the fact that the reduced density matrix of one region, ρ_A, is proportional to identity (maximally mixed).

Having calculated the entanglement entropy for rainbow states, the minimum time required to create them can be limited once we specify a connectivity. For instance, in an all-to-all connected system, we can create the rainbow state in one unit of time by simply directly entangling the pairs desired. Alternately, we might restrict ourself to a nearest-neighbor connected graph, where a vertex i is only capable of sharing Bell pairs (or equivalently performing CNOT) with the qubits $i\pm1$. Since we must create $\mathcal{O}(N)$ units of entanglement entropy, it follows that it will take $\mathcal{O}(N)$ operations across the boundary to create it. If we can only distribute one Bell pair for consumption per time step across the boundary, it will take a time $\mathcal{O}(N)$ to create the rainbow state, no matter what classical communication or measurement is allowed.

For the purposes of identifying a difficult task to test the 1D architecture, any state of N qubits that has some possible bipartition with $\mathcal{O}(N)$ entanglement entropy would share this same lower-bound derived from the entanglement capacity. However, the rainbow state has the advantage that it is easy to describe and specify. Algorithms which create the state are also intuitively easy to understand, since they involve only the creation and arrangement of Bell pairs. The rainbow state, specifically, makes a good choice for the benchmarking of quantum architectures because this simplicity provides a hope that the difficulty of creating the rainbow state can be reduced entirely to the bipartite entanglement which we can limit by

the entanglement capacity. We will explore the question of when the rainbow state can be saturated in Sec. 4.7

By contrast, as we saw previously, the GHZ state can be created in constant time by using measurement and feedback. This agrees with the fact that the GHZ state itself has only a constant amount of entanglement entropy and so $\mathcal{O}(1)$ is the bound suggested by the entanglement capacity.

4.5 Rainbow Times and Isoperimetric Number

Above, we argued that the entanglement capacity offers an easy way to understand the limitations on creating the rainbow state on N qubits in a line. We now wish to extend this to arbitrary graphs. Suppose we define an arbitrary architecture of qubits $G = (V, E, w)$ and any set of $F \subset G$, with the requirement that $|F| \leq \frac{1}{2}|G|$. Denote by F_i the ith vertex of F by an arbitrary ordering and similarly use $\bar{F}_i$ to index vertices in the complement $\bar{F}$. We can then define a "rainbow" state:

$$|\psi_{\text{rb}}(F)\rangle = \bigotimes_{i=1}^{|F|} \frac{1}{\sqrt{2}} \left(|0\rangle_{F_i} |0\rangle_{\bar{F}_i} + |1\rangle_{F_i} |1\rangle_{\bar{F}_i}\right) \bigotimes_{i=|F|+1}^{|\bar{F}|} |0\rangle_{\bar{F}_i} . \tag{4.12}$$

Now we have specified a state which has maximal entanglement entropy for the bipartition of the entire system (G) into a set F and its complement $\bar{F}$. Note that if $|F| \geq \frac{1}{2}|G|$, we cannot define a rainbow state as in Eq. (4.12) because there will not be enough qubits in $\bar{F}$ to form Bell pairs with all qubits in F. To choose a $|\psi_{\text{rb}}(F)\rangle$ with maximal entanglement entropy, we would increase F until $|F| = \frac{1}{2}|G|$. However, our goal is instead to identify a difficult task for quantum computers, so we instead maximize the required state creation time, rather than the entropy itself.

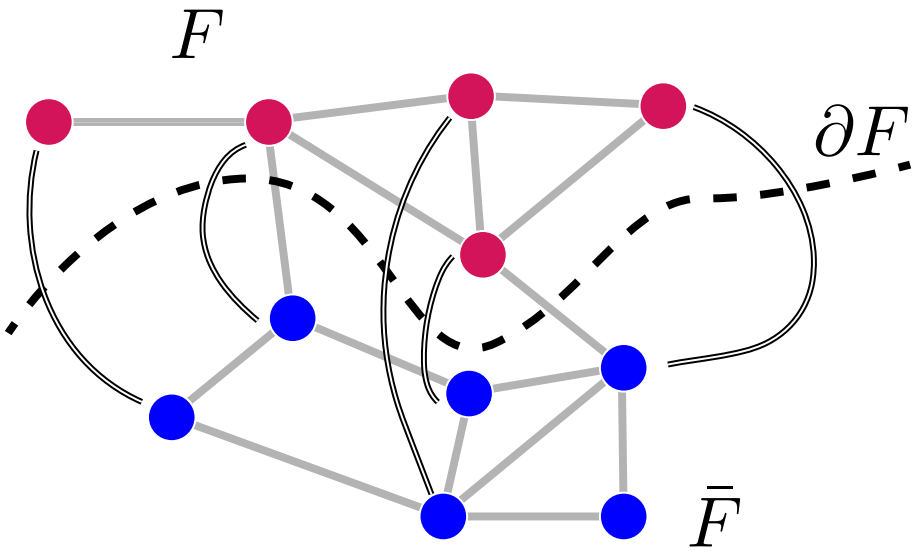

Figure 4.4: An illustration of how a rainbow state is constructed on an arbitrary subgraph F. Here, grey lines represent the connectivity graph of allowed two-qubit interactions, while doubled black lines represent maximally entangled qubit pairs. Qubits without a doubled line are assumed to be in state $|0\rangle$.

If we treat G as a graph with its edges being the allowed CNOT operations, we can calculate the lower bound on the time required to create the rainbow state $|\psi_{\text{rb}}(F)\rangle$, according to the entanglement capacity. For any set F we define the time, $t(F)$,

$$t(F) = \frac{|F|}{|\partial F|} = \frac{\text{set size}}{\#\text{ of connections}}. \tag{4.13}$$

Here, the numerator is simply the size of the set F, which is proportional to the entanglement entropy that must be created, while the denominator, $|\partial F|$, is the total size of the boundary of F. The boundary defines the total number of Bell pairs that can be distributed to the region F from its complement $\bar{F}$ in a single time step.

Note that Eq. (4.12) does not completely describe a physical state, which

would require an ordering of vertices in F and $\bar{F}$. However, that ordering does not affect the time required to create the state according to the entanglement capacity, $t(F)$. While different rainbow states which share a common F may differ in how easy they are to create, $t(F)$ serves as the common lower bound on creation time for all of them, and we will focus on that metric here.

Rather than the properties of a particular rainbow state or a particular subgraph, we wish to gain insight on the behavior of the graph G itself. To do this, we seek to find the maximum $t(F)$ given G. This will tell us: of all the maximally-entangled states we can build from sets of G, which family is hardest to build according to the entanglement capacity? We call this the *rainbow time* of the graph G and denote it $\tau_{\mathrm{rb}}(G)$. It is defined as:

$$\tau_{\mathrm{rb}}(G) = \sup_{F \subset G, |F| \leq \frac{1}{2}|G|} t(F). \tag{4.14}$$

The rainbow time has a simple and attractive interpretation, can be directly connected to quantum computing tasks, and is robust to various models of computation by combining the theorems of Refs. [143, 144]. In addition, it can be directly connected to a quantity known as the Cheeger constant or isoperimetric number, $h(G)$ [145] which is well-studied in graph theory and computer science [154–156]. As we have defined it, the rainbow time is simply $1/h(G)$ (Note that as we have defined it so far $\tau_{\mathrm{rb}}(G)$ can take on any nonnegative real value. In reality, the creation of a quantum state will always take an integer number of steps greater than or equal to one in the Bell pair exchange model. Therefore, $\lceil \tau_{\mathrm{rb}} \rceil$ can be used as a measure of the "number of rounds" required in cases where this is important.) Thus, aiming

to minimize the rainbow time (so that large entangled states can be easily constructed) in a quantum architecture is equivalent to maximizing the isoperimetric number. An "isoperimetric set" is a set F that achieves $t(F) = \tau_{\text{rb}}(G)$. Often, the Cheeger constant appears in the context of expander graphs, which are constructed to possess large Cheeger constants [157] and can be used for proofs in complexity theory [158–160].

4.6 Application to Hierarchical Product and Hierarchies

We now calculate the rainbow times for the hierarchical products and hierarchies of Chapter 3. To summarize the construction presented in that chapter, a hierarchical product is a graph product denoted $G \,\Pi\, H$ in which $|G|$ copies of H are connected at their root (first) vertices by the graph G. By iterating this process, we can create a hierarchy, in which a high-level graphs connect lower-level identical sub-hierarchies. We also extend our concept to that of a weighted hierarchy, in which the edges on level i have weight α_i. We write a k-level hierarchy with a vector of weights $\vec{\alpha}$ as $G^{\Pi_{\vec{\alpha}} k}$, where G is the base graph. Finally, if $\alpha_i = \alpha^{i-1}$, so that edge weight scales geometrically with the level of the hierarchy, we simply write $G^{\Pi_\alpha k}$. Some examples are shown in Fig. 4.5.

In order to calculate the rainbow time for a hierarchical product, we make use of the result from Ref. [145] there must exist an isoperimetric set [an F such that $t(F) = \tau_{\text{rb}}(F)$] that is connected and whose complement $\bar{F}$ connected. Therefore, we will look at all possible subgraphs of $H_1 \,\Pi\, H_2$ where both F and $\bar{F}$ are connected.

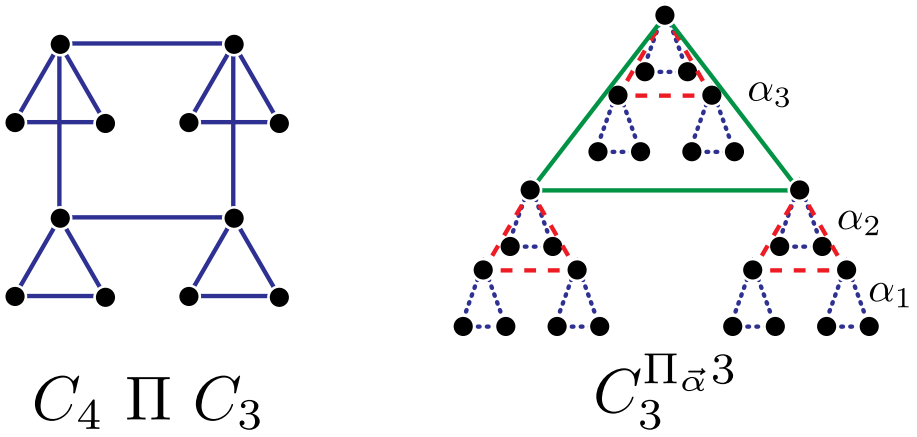

Figure 4.5: Examples of a hierarchical product (left) and a weighted hierarchy (right).

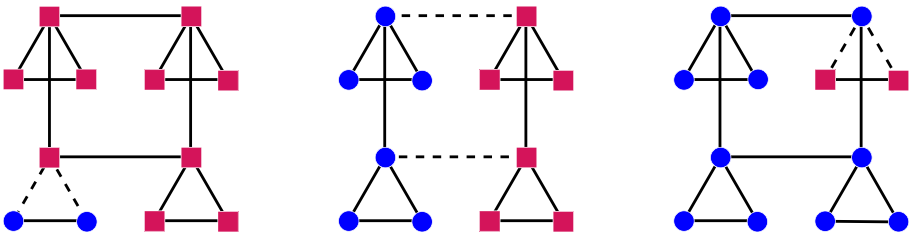

Figure 4.6: Three classes of subgraph used in our proof. Circles represent vertices in F, squares are vertices in $\bar{F}$, and dashed lines are edges in ∂F. (Left) A setting in which part of one copy of H_2 is covered. (Center) A division that lies entirely in H_1. (Right) Here, F covers all but one of the copies of H_2.

From these, we will search for the one with the largest $t(F)$. Since some isoperimetric set is guaranteed to exist in this set of subgraphs, this maximization over $t(F)$ in this set will also give us $\tau_{\text{rb}}(F)$. We will begin by specifying three cases, illustrated in Fig. 4.6. These cases cover all possible sets with the right connectedness properties and therefore allow us to find the maximizing set for the graph and $\tau_{\text{rb}}(H_1 \,\Pi\, H_2)$.

One such set would cover part of one copy of H_1. However, note that if the root vertex of H_2 were included, we would have to include the entire hierarchy, since otherwise $\bar{F}$ would not be connected. We leave situations in which an entire sub-

hierarchy is included for the next case. In the meantime, we must maximize over all possible sets of H_2 to find the best $t(F)$. This may seem like it would yield $\tau_{\mathrm{rb}}(H_2)$; however, in this instance we can pick subsets of H_2 which make up a majority of H_2, which is not allowed for τ_{rb}. We define the *unrestricted-size rainbow time* as

$$u_{\mathrm{rb}}(G) = \sup_{F \subset G \setminus G_1} t(F), \tag{4.15}$$

which is just Eq. (4.14) with no bound on the size of the subset (but with the restriction that the root vertex is excluded). Therefore, any set from this class will offer a candidate rainbow time of at most $t(F) = u_{\mathrm{rb}}(H_2)$.

The second class of candidate sets would cover multiple copies of H_2. Since F must be connected, the path between these copies must be included in F, which means the root vertices of each H_2 that connect to each other via H_1 must also be in F. Then, as argued above, each copy of H_2 must be included entirely. As a result, this case is equivalent to choosing copies of H_2 and either entirely including them in F or entirely excluding them. This problem reduces to dividing up H_2, and then calculating as if each vertex had an effective volume of $|H_1|$. Therefore, we can find the maximum $t(F)$ of these sets by simply finding $\tau_{\mathrm{rb}}(H_2)$ and scaling it by $|H_1|$.

The final class of F which meets the connectedness criteria would be an F which includes all of H_2 and then all but one copy of H_1 completely, with perhaps some of the remaining H_1 also included. However, this F would necessarily be larger than half of the total graph $H_1 \amalg H_2$, and therefore we can discard it as a possible F for determining the rainbow time.

We combine the first two options and conclude that:

$$\tau_{\mathrm{rb}}(H_1 \,\Pi\, H_2) = \max\left(u_{\mathrm{rb}}(H_1), |H_1|\,\tau_{\mathrm{rb}}(H_2)\right). \tag{4.16}$$

We now seek to apply this to hierarchies $G^{\Pi_{\vec{\alpha}}k}$. Just as before, if a vertex is included in F, we must include all of the qubits "below" it in the hierarchy, otherwise the complement of F will not be connected. Therefore, all bipartitions can be reduced to choosing a particular level of the hierarchy to cut – on that level, either a vertex will be included or not included, and this must apply to all of its descendants as well. Every bipartition can then be mapped to a bipartition of G, but one where every vertex is scaled by $|G|^{i-1}$ due to the size of each sub-hierarchy. (Note that the large number of vertices not in F do not contribute to $t(F)$.)

There is one important difference between the top (kth) level and all others, which arises from the constraint that $|F| \leq \frac{1}{2}\left|G^{\Pi_{\vec{\alpha}}k}\right|$. A cut on the top level must not include more than half of the highest-level copy of G, while all lower levels can use any cut at all as long as it does not include the root vertex. Whatever level we cut, our cut depends only on base graph G, with each node standing for $|G|^{i-1}$ total nodes below it. In addition, $t(F)$ must also be modified by the edge weight, which we define to be α_i on level i. Therefore, we can write the overall τ_{rb} as a maximization over these options:

$$\tau_{\mathrm{rb}}(G^{\Pi_{\vec{\alpha}}k}) = \max\left(\frac{|G|^{k-1}}{\alpha_k}\tau_{\mathrm{rb}}(G), \sup_{i<k}\frac{|G|^{i-1}}{\alpha_i}u_{\mathrm{rb}}(G)\right). \tag{4.17}$$

For specificity, we will evaluate the case where $G = K_n$ and $\alpha_i = \alpha^{i-1}$. Here, the maximization over lower levels [the second term in Eq. (4.17)] can be reduced to either to the first level or the $(k-1)$th level, since we simply have to pick the

largest element in a geometric sequence defined by n/α. We can write the resulting maximization as a choice between three options,

$$\tau_{\mathrm{rb}}(K_n^{\Pi_\alpha k}) = \max\left(\left(\frac{n}{\alpha}\right)^{i-1}\frac{2}{n}, \left(\frac{n}{\alpha}\right)^{i-2}, 1\right). \tag{4.18}$$

Whereas one might have expected two options to arise (cut at the top or at the bottom), we actually have three. For $\alpha > n$, the edges grow in capacity too quickly for the increased volume to make a higher-level cut worthwhile, so the optimal cut is at the bottom, yielding a constant scaling. Two other options appear at $n > \alpha$, where cutting higher up the hierarchy allows for greater volume of qubits in F without too much penalty caused by changing edge weights. The reason there are two strategies is that it may be possible to cut a larger portion of a lower hierarchy and exploit the split between τ_{rb} and u_{rb}. (For K_n in particular, the cut that includes all but the root vertex satisfies $u_{\mathrm{rb}}(K_n)$.)

Graph Name	τ_{rb}	w	Δ
K_N	N^{-1}	N^2	N
S_N	1	N	N
d-dimensional Grid	$N^{1/d}$	N	$2d$
$K_n^{\Pi_\alpha k}$	$N^{\max(0,\log_n \alpha)}$	$N^{\max(1,\log_n \alpha)}$	$\log_n N$

Table 4.1: Important statistics for graphs. Here, only the asymptotic scaling with N is written. In addition to the rainbow time τ_{rb} for each graph, we also include the total weight of all edges, w and the maximum graph degree Δ. Rainbow times for graphs other than hierarchies can be found in terms of isoperimetric number in Refs. [145, 155].

To place these results in context, we compare the rainbow time of $K_n^{\Pi_\alpha k}$ to

the total rainbow time of other graphs. To do this, we write the rainbow time in terms of the total number of qubits in a graph, N, and concern ourself with the overall scaling. In this language, $\tau_{\mathrm{rb}}(K_n^{\Pi_\alpha k}) = \Omega\left(N^{\max(0,1-\log_n \alpha)}\right)$. We compare this to the rainbow time of some other graphs in Table 4.1. References [145, 155] give the isoperimetric number for K_N, S_N, and grids (which are Cartesian products of paths). Satisfying sets for those graphs are: for K_N and S_N, an arbitrary half of the nodes; for grids, a hypercube placed in one corner that takes up precisely half the total volume.

One goal would be to identify a set of parameters where a hierarchy outperforms a d-dimensional grid architecture. We are most concerned with comparing to the d-dimensional grid because the other candidates we present, K_N and S_N, both have very large degree, making them impractical for scalable architectures. We find that the rainbow time of the hierarchy with base graph K_n and scaling constant α will be better (smaller) than that of the grid if $\alpha > n^{(d-1)/d}$. If it also holds that if $n > \alpha$, then the hierarchy will accomplish this while also having lower total edge weight, i.e., less overall connection overhead. We conclude that a hierarchy $K_n^{\Pi_\alpha k}$ with $\alpha \in \left[n^{(d-1)/d}, n\right)$ has both lower rainbow time and less total edge weight than a d-dimensional grid of qubits.

4.7 Saturability of Rainbow Time

We now know that the time required to create a highly entangled state across a bipartition F, $\bar{F}$ of G is lower-bounded by $t(F)$. However, we have not proven that

it is possible to create a rainbow state on an isoperimetric set in the time $\tau_{\mathrm{rb}}(G)$. It is possible that the limit is an unattainable lower bound on time required. We would like to provide an algorithm which can, in time $\mathcal{O}(\tau_{\mathrm{rb}}(G))$, produce a state with maximal entanglement entropy. This would then show that rainbow time is not just a limit on the generation of entanglement, but a reliable diagnostic that can be used to evaluate architectures.

Unfortunately we will not be able to present a general proof of this fact, however, we will outline early progress towards the following conjecture: that for any quantum computer whose architecture can be represented as a graph $G = (V, E, w)$, where an edge e can distribute $w(e)$ Bell pairs per time step, for any isoperimetric subgraph F, it is possible to construct a state with entanglement entropy $\mathcal{O}(|F|)$ in time $\mathcal{O}(\tau_{\mathrm{rb}}(G))$ using entanglement swapping.

4.7.1 Flow Results

We first begin by proving a pair of results about flows on a graph. A *flow* between two nodes s (a source) and t (a sink) is a function on the edges of a graph G such that, at every edge e, the flow function satisfies the following constraints:

- The flow through a graph edge never exceeds the weight of that edge, which we interpret as a capacity: $f(e) \leq w(e)$.

- The flow into and out of every node except s and t is balanced, so that $\sum f(e) = 0$ if the sum runs over all e including a given node.

The value of the flow, denoted $|f|$, is the total amount of flow that emerges in s and is deposited in t. From the flow, we can create a directed graph which tracks the flow as it moves from s to t.

4.7.1.1 Maximal Flow of Entanglement

We are most interested in examining a quantum computer in which nodes communicate by exchanging Bell pairs. Bell pairs are indivisible, meaning that the weights $w(e)$ of our edges are necessarily integers. In this example, we ignore this restriction temporarily, considering an example in which entanglement behaves more like a fluid being distributed. We will show that for any isoperimetric set F, it is possible to define a flow that delivers $1/\tau_{\mathrm{rb}}(G)$ units of entanglement to each node in F per time step. Therefore, in $\tau_{\mathrm{rb}}(G)$ steps each node in F would become maximally entangled with a node in $\bar{F}$. We illustrate a sample flow in Fig. 4.7.

We begin by creating a set K which is an arbitrary subset of $\bar{F}$ such that $|K| = |F|$. We now attach to our graph G two fictitious nodes: a source s, which connects to every node in F, and a sink t, which connects to every node in K. We assign all of these new edges a weight (capacity) $1/\tau_{\mathrm{rb}}(G)$. We now seek to prove that there is a flow f of value $|F|/\tau_{\mathrm{rb}}(G)$. Such a flow saturates all of the edges we have added, ensuring that every node in F is being reached by the flow.

To do this we will use the MaxFlow-MinCut theorem. A *cut* is a division of the graph into two sets, S and T, such that $s \in S$ and $t \in T$. The value of the cut is the weight of all edges which connect S and T. By MaxFlow-MinCut, the

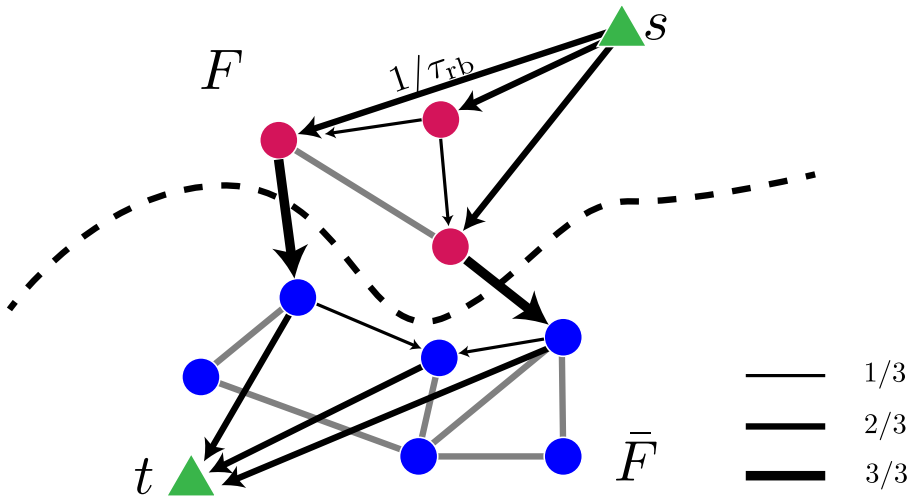

Figure 4.7: An illustration of the fictitious nodes added to the isoperimetric set, F and a set of equal size K. The new nodes, s and t, appear as green triangles; the original nodes and edges are pink (in F) and blue (in $\bar{F}$) circles. The new edges connected to s, t are of weight $1/\tau_{\rm rb} = 2/3$, while all other edges have weight one. The flow is illustrated by arrows of differing thickness and transfers $|F|/\tau_{\rm rb}$ units of entanglement across the graph. Grey edges are not used by the flow.

maximum possible value of the flow f is the smallest possible value of a cut that separates s and t. Intuitively, this is simply a statement that the worst bottleneck in the graph bounds the flow. Note that MaxFlow-MinCut also guarantees that a flow of this value exists [161].

Consider an arbitrary cut, $\mathrm{Cut}(S,T)$. Let Q denote $S \setminus \{s\}$ and R denote $T \setminus \{t\}$. Furthermore, define the following:

$$a \equiv |R \cap F|\,, b \equiv |Q \cap K|\,. \tag{4.19}$$

Observe that

$$|Q| \geq |Q \cap F| + |Q \cap K| = |F| - a + b, \tag{4.20}$$

$$|R| \geq |R \cap F| + |R \cap K| = a + |K| - b = a + |F| - b. \tag{4.21}$$

Now, for any sets A and B in the graph, we will use the following notation for the weight of the edges between them,

$$|A \to B| = \sum_{i \in A, j \in B} w(i,j). \tag{4.22}$$

We can now define the value of the cut:

$$|\mathrm{Cut}(S,T)| = |(\{s\} \cup Q) \to (R \cup \{t\})| = |s \to R| + |Q \to R| + |Q \to t|\,. \tag{4.23}$$

By definition, s connects to every node in R which is also in F, and t connects to every node in Q that is also in K. All of these edges have weight $1/\tau_{\mathrm{rb}}(G)$, allowing us to write:

$$|\mathrm{Cut}(S,T)| = \frac{a+b}{\tau_{\mathrm{rb}}(G)} + |Q \to R|\,. \tag{4.24}$$

We now want to use the fact that the isoperimetric number $(1/\tau_{\mathrm{rb}}(G))$ limits the possible size of $|Q \to R| = |\partial Q| = |\partial R|$. Since $\tau_{\mathrm{rb}}(G)$ is defined as a maximization, any set which is less than half of the graph has ratio of volume to perimeter that is upper-bounded by $\tau_{\mathrm{rb}}(G)$. Suppose that $|R| \leq |G|/2$. Then applying the isoperimetric condition yields:

$$|Q \to R| = |\partial R| \geq \frac{|R|}{\tau_{\mathrm{rb}}(G)} \geq \frac{a + |F| - b}{\tau_{\mathrm{rb}}(G)}. \tag{4.25}$$

Alternately, if $|R| > |G|/2$, it follows that $|Q| \leq |G|/2$, since these two sets partition the entire graph. In this case, it follows that:

$$|Q \to R| = |\partial Q| \geq \frac{|Q|}{\tau_{\mathrm{rb}}(G)} \geq \frac{|F| - a + b}{\tau_{\mathrm{rb}}(G)}. \tag{4.26}$$

No matter which result is applicable, substitution into Eq. (4.24) yields $|\mathrm{Cut}(S,T)| \geq |F|/\tau_{\mathrm{rb}}(G)$. Since every cut is larger than $|F|/\tau_{\mathrm{rb}}(G)$ it follows that the maximum flow must have at least $|F|/\tau_{\mathrm{rb}}(G)$ units of flow.

This then defines a "protocol" of sorts, although not one that fits into the model of computation we have outlined. If entanglement can be transmitted on the edges of G, this flow allows us to selects a subset of the same size of F and, every time step, deliver $1/\tau_{\text{rb}}(G)$ units of entanglement between every node in F and some partner in K. In $\tau_{\text{rb}}(G)$ time steps a total of $|F|$ entanglement has been created. The individual paths that combine to produce this flow can be found using the Ford-Fulkerson algorithm [162].

It is easy to create an integer flow that corresponds to our Bell pair exchange physical interpretation. If we scale all of the edges in the graph by $\lceil \tau_{\text{rb}}(G) \rceil$, and make the fictitious edges that connect to s and t be of weight 1 instead of $1/\tau_{\text{rb}}(G)$, an identical argument to the above allows us to conclude that a flow exists of value $|F|$. We can also invoke the Integral Flow Theorem to assert that, since in this case all of the edge weights are integers, there exists an optimal flow such that the flow on each edge is also integer-valued [163]. Since we intend to perform a total protocol in $\lceil \tau_{\text{rb}}(G) \rceil$ time steps, this calculation shows that, given all of our capacity "at once" we can define a flow that delivers one unit of entanglement (one Bell pair) to each node. However, it is not clear whether or not this can then be decomposed into $\lceil \tau_{\text{rb}}(G) \rceil$ steps which can all be simultaneously executed.

4.7.2 Algorithm for Flow Decomposition

We will now present a proposed algorithm for performing the task of creating $\lceil \tau_{\text{rb}}(G) \rceil$ separate steps, each of which entangles $|F| / \lceil \tau_{\text{rb}}(G) \rceil$ nodes. The algorithm

we will present works by dividing the problem into two sub-tasks:

1. We create a list of paths between nodes. These paths specify which nodes in F are paired with which nodes in K as well as the paths that are taken when performing entanglement swapping to create the Bell pair between those two nodes.

2. We partition the paths into $\lceil \tau_{\text{rb}}(G) \rceil$ subsets. Each subset can be simultaneously performed without requiring any edge in G to be used beyond capacity. Therefore, the rainbow time can be created in time $\lceil \tau_{\text{rb}}(G) \rceil$.

We will outline specific methods to perform each step, but this general structure can accommodate many variants of the two subroutines. At this time, we do not have a proof that our methods succeed in all cases.

4.7.2.1 Building a Path Matrix

We accomplish the first step by performing a depth-first search on the directed graph defined by the flow f. We simply begin at a node $u \in F$ and perform the search until we reach some $v \in K$ which has not yet been paired off. We then remove one unit of flow from every edge on the path found between these nodes and continue. The result is two matrices. One matrix, M, is simply a $|F| \times |K|$ matrix which identifies which node in F pairs to which node in K. $M_{ij} = 1$ if i and j are paired, and $M_{ij} = 0$ otherwise. The second matrix, A, is called a path matrix. If $G = (V, E)$, then A is an $|E| \times |F|$ matrix. Each row of A corresponds to one edge in G, while each column represents one of the paths we identified. An element A_{ij}

is 1 if the jth path passes through the ith edge, and 0 otherwise. The ordering of the paths and edges has no significance. A formal specification of this procedure is found below in Algorithm 1.

4.7.2.2 Partitioning Paths Into Steps

Now that we have the path matrix A which defines the set of paths, we can cast the problem of dividing the paths into simultaneously-executable steps as a linear programming problem. Suppose that $\mathbf{x}$ is a binary vector of length $|F|$ which we use to identify a set of paths we wish to simultaneously execute. Let $x_i = 1$ if the ith path is selected to execute, and $x_i = 0$ otherwise. Then the product $A\mathbf{x}$ will yield the total number of times each edge will be used in this time step. In order to identify a feasible set of paths, we require $A\mathbf{x} \leq \mathbf{w}$, where $\mathbf{w}$ is a vector of the edge weights (the inequality is element-wise).

On the other hand, we also wish to ensure that the edges that are *not* used will remain feasible. If there are n steps remaining after this one, then if any of the unused edges needs to be used more than $nw(e)$ times, the procedure will fail. Therefore, by using $\mathbf{1} - \mathbf{x}$ as the set of paths not used, we will also require that $A(\mathbf{1} - \mathbf{x}) \leq n\mathbf{w}$, again using an elementwise inequality.

This procedure can simply be iterated by deleting columns of A which correspond to already-used paths. By optimizing $\mathbf{x}$ to capture as many paths as possible while respecting both edge capacity in each time step and future demand for all other edges, we can complete the partitioning process. A rigorous explanation of

this procedure can be found in Algorithm 2.

Algorithm 1 The function which creates a path matrix describing paths between two sets which are meant to be entangled. The function inputs are the vertex and edge sets V and E, a flow f, and the sets F and K that are to be entangled.

function PATHMATRIX(V, E, f, F, K)
    Initialize a $|E| \times |F|$-matrix A and a $|F| \times |F|$ matrix M to all zeros.
    Find any directed cycles in $G_f = (V, E, f)$, and reduce all weights in the cycles to get $G_{f'} = (V, E, f')$ which has no cycles.
    $K_{res} \leftarrow K$
    for $u \in F$ **do**
        Perform depth-first-search starting from u on graph $G_{f'} = (V, E, f')$, and find a path $P_{u,v}$ from u to some $v \in K_{res}$. Set $M(u, v) \leftarrow 1$.
        for $e \in P_{u,v}$ **do**
            $f'(e) \leftarrow f'(e) - 1$
            $A(e, u) \leftarrow 1$
        end for
        $K_{res} \leftarrow K \setminus v$
    end for
    return (A, M) ▷ A is what we call the path matrix
end function

Algorithm 2 Creating the rainbow state from an integral max flow. Inputs are a graph $G = (V, E)$, a weight function w, an isoperimetric set F, a set of size $|F|$ $K \subset \bar{F}$, and an optimal integral flow f^*.

1: Let $(A, M) = \textsc{PathMatrix}(V, E, f^*, F, K)$.

2: **for** $t = T, T-1, \ldots, 2, 1$ **do**

3: Solve the following linear program:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\{ -\mathbf{1}^T A\mathbf{x} : A\mathbf{x} \leq \mathbf{w}, A(\mathbf{1}-\mathbf{x}) \leq (t-1)\mathbf{w} \text{ and } \mathbf{x} \in \{0,1\}^{|F|} \right\} \tag{4.27}$$

where $\mathbf{w} \in \mathbb{Z}^{|E|}$ is the vector of edge weights in G, and $\mathbf{1}$ is the all-one vector (of appropriate dimension). Inequalities are element-wise.

4: Let $\mathbf{g} = A\mathbf{x}^*$. We generate $g(e) \leq w(e)$ Bell pairs across every edge e.

5: We can now simultaneously entangle qubits in F corresponding to the non-zero entries in $\mathbf{x}^*$ to their counterparts in K. This is done by performing entanglement swapping along the paths indicated by the column u of A for all u where $x^*(u) = 1$, reaching $v = M(u) \in K$.

6: $A \leftarrow A \cdot \text{diag}(\mathbf{1} - \mathbf{x}^*)$. ▷ Set columns of A corresponding to depleted paths to zero.

7: **end for**

4.7.3 Algorithm Optimality

Unfortunately, Algorithm 2 is not guaranteed to succeed. This is because the linear program Eq. (4.27) cannot necessarily be guaranteed to have integer solutions. Non-integer solutions have no meaning in our context, because we have to either execute an entire path or not. In this section, we will present some partial results and identify a condition that can be placed on A which, if satisfied, guarantees integer solutions.

4.7.3.1 Feasibility of the Linear Program

We first wish to show that the linear program defined in Eq. (4.27) always does have some feasible solution, although this does not establish that any of those solutions are integers. In order to show that there exists a solution, we will appeal to Farkas' lemma; this lemma states that the following implication is true for linear equations [164]:

$$\exists \mathbf{x} \geq 0, A\mathbf{x} = b \iff \forall \mathbf{p} : A^{\dagger}\mathbf{p} \geq 0 \implies \mathbf{b}^{\dagger}\mathbf{p} \geq 0. \tag{4.28}$$

In order to apply Farkas' lemma, we must modify the linear program we originally presented. Our first task is to combine the two inequalities into a single larger inequality:

$$\tilde{A} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}, \tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathbf{1} - \mathbf{x} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} \mathbf{w} \\ (t-1)\mathbf{w} \end{pmatrix}. \tag{4.29}$$

Now we can write the linear program as optimizing over $\tilde{A}\tilde{\mathbf{x}} \leq \mathbf{b}$. However, we need to convert this to an equality. We can do this by adding "slack variables," $\mathbf{s_1}$ and $\mathbf{s_2}$,

which are constrained to be positive, representing the difference that a particular solution may have from its upper bound. Adding these in, the system of equations is defined by:

$$\tilde{A} = \begin{pmatrix} A & 0 & I & 0 \\ 0 & A & 0 & I \\ I & I & 0 & 0 \end{pmatrix}, \quad \tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{w} \\ (t-1)\mathbf{w} \\ \mathbf{1} \end{pmatrix}, \tilde{\mathbf{x}} \geq 0. \tag{4.30}$$

We are now ready to apply Farkas' lemma to the equality $\tilde{A}\tilde{\mathbf{x}} = \mathbf{b}$. Suppose there is a vector $\mathbf{p}$ such that $\tilde{A}^\dagger \mathbf{p} \geq 0$. We will show that this necessarily implies $\mathbf{b}^\dagger p \geq 0$, which by Farkas' lemma implies a solution exists. Let $\mathbf{p} \equiv (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$. Expanding out $\tilde{A}$ in terms of the original matrix then yields,

$$A^\dagger \mathbf{p}_1 + \mathbf{p}_3 \geq 0, \; A^\dagger \mathbf{p}_2 + \mathbf{p}_3 \geq 0, \; \mathbf{p}_1 \geq 0, \; \mathbf{p}_2 \geq 0. \tag{4.31}$$

Since $\mathbf{p}_3$ is greater than both $-A^\dagger \mathbf{p}_1$ and $-A^\dagger \mathbf{p}_2$, it follows that

$$\forall \alpha \in [0,1] : \mathbf{p}_3 \geq -A^\dagger \left[\alpha \mathbf{p}_1 + (1-\alpha)\mathbf{p}_2\right]. \tag{4.32}$$

We are now ready to begin computing $\mathbf{b}^\dagger \mathbf{p}$. We will simply compute each of its three parts and then apply the inequality Eq. (4.32).

$$\mathbf{b}^\dagger \mathbf{p} = \mathbf{w}^\dagger \mathbf{p}_1 + (t-1)\mathbf{w}^\dagger \mathbf{p}_2 + \mathbf{1}^\dagger \mathbf{p_3} \tag{4.33}$$

$$\geq \mathbf{w}^\dagger \mathbf{p}_1 + (t-1)\mathbf{w}^\dagger \mathbf{p}_2 - \mathbf{1}^\dagger A^\dagger \mathbf{p}_2 [\alpha \mathbf{p}_1 + (1-\alpha)\mathbf{p}_2] \tag{4.34}$$

Now note that $A\mathbf{1} \leq t\mathbf{w}$, since $(A\mathbf{1})_i$ is just the total use of edge i by all paths over all time steps. Therefore,

$$\mathbf{b}^\dagger \mathbf{p} \geq \mathbf{w}^\dagger \mathbf{p}_1 + (t-1)\mathbf{w}^\dagger \mathbf{p}_2 - t\mathbf{w}^\dagger [\alpha \mathbf{p}_1 + (1-\alpha)\mathbf{p}_2] \tag{4.35}$$

Let us then choose $\alpha = 1/t$, and $1 - \alpha = (t-1)/t$, which yields

$$\mathbf{b}^\dagger \mathbf{p} \geq 0, \tag{4.36}$$

completing the argument and showing that the linear program in Eq. (4.27) is feasible.

4.7.3.2 Total Unimodularity and Planar Graphs

A known sufficient condition for a linear program to have integer solutions is known as total unimodularity. A totally unimodular (TU) matrix is a matrix in which every square submatrix has determinant -1, 0, or +1. If the matrix A were totally unimodular, then the linear program it defined would be guaranteed to have an integer optimal solution [164]. By combining this knowledge with the guarantee from Farkas' lemma that the program has some feasible solution, we would be guaranteed that our algorithm would be capable of producing rainbow states in $\lceil \tau_{\text{rb}}(G) \rceil$ time steps.

Although our algorithm produces totally unimodular matrices on randomly-generated test graphs, we have yet to prove that this is definitively the case. However, for an important class of input graphs, this condition can be proved. Those graphs are planar graphs, which are any graph that can be drawn on a piece of paper without any of the edges crossing each other [165].

In order to see that total unimodularity can be guaranteed for planar graphs, we consider the maximal flow f^* on a graph $G = (V, E)$. Suppose we add the fictitious nodes s and t and the attendant edges, and that after this addition the

graph G is planar. Then embed this graph in the plane, oriented so that the flow moves from left to right. We can then identify the topmost path from s to t in the graph that has positive flow and remove that path from the flow. Repeating this yields $|F|$ paths which are ordered from top to bottom. Although these paths may share edges, they never cross each other. As a result, when we construct the path matrix A, there is an ordering of the paths (columns) such that any paths that share an edge are consecutive in the ordering. Therefore, any 1's that appear in a row will appear consecutively. This is a known condition that guarantees unimodularity [164], so that when using Algorithm 2, the solutions to Eq. (4.27) will be integers, and the resulting output can be used to schedule the creation of entangled pairs to produce $|F|$ Bell pairs between F and $\bar{F}$ in time $\lceil \tau_{\mathrm{rb}}(G) \rceil$, as desired.

4.8 Conclusion

In this chapter, we have presented a new metric for evaluating proposed architectures for quantum computers. By analyzing the architecture in terms of a graph whose edges represent bandwidth and whose vertices represent qubits, we were able to put a rigorous physical lower bound on time required to create families of states inspired by the "rainbow state" from 1D systems. Notably, this physical bound is robust to the inclusion of non-unitary operations such as classical control, measurement, and feedback. Having found this bound, we then connected the creation of those states to a known property of graphs, called the Cheeger constant or isoperi-

metric number. By making use of earlier results in graph theory, we were able to evaluate this isoperimetric number for graphs formed from the hierarchical product and for hierarchies formed from many iterations of that product.

In the future, we would like to explore the nature of the rainbow time as a bound. In addition to the saturability of the rainbow time considered above, another interesting question is whether there are significant differences between different permutations of qubits within F that might allow us to prove larger lower bounds on some particular physical states. In general, we expect that there are likely other computational or physical ways to lower-bound the required time for quantum algorithms, even after entanglement is taken into account. We hope that by clarifying the role of entanglement entropy these other constraints can be better understood. We note that for the examples tabulated in Table 4.1, it is easy to outline algorithms which create a rainbow state in $\lceil \tau_{\mathrm{rb}} \rceil$ rounds. These algorithms simply consist of producing Bell pairs along the boundary and immediately distributing them away from the boundary.

Finally, we would like to learn more about how the entanglement entropy, which was used here in terms of the rainbow time, can be applied to the analysis of quantum algorithms, for instance, by showing that a certain amount of entanglement is required to perform other tasks in quantum computing. References [166, 167] explore this question in Shor's algorithm and in adiabatic quantum computing. These complement other results which show that low-entanglement systems can be simulated efficiently on a classical computer [20, 168]. Our metric can evaluate the amount of time different quantum architectures would require in order to perform

entanglement-generating processes. Rainbow time could also be a fruitful way to test algorithms for compilation and gate decomposition, by comparing their realized circuit depth to this theoretical minimum required time.

Chapter 5: Sensor Networks

5.1 Introduction

Entanglement is a valuable resource in precision measurement, as measurements using entangled probe systems have fundamentally higher optimal sensitivity than those using unentangled states [23]. A generic measurement using N unentangled probes will have a standard deviation from the true value asymptotically proportional to $1/\sqrt{N}$. By using N maximally entangled probes, a single parameter coupled independently to each probe system can be measured with an uncertainty proportional to $1/N$. This is the best possible scaling consistent with the Heisenberg uncertainty principle and is known as the Heisenberg limit [23, 169]. The procedure can also be reversed–enhanced sensitivity to disturbances can provide experimental evidence of entanglement [170–172].

Measurements making use of entanglement usually couple one parameter to N different systems [23, 45, 173]. However, the emerging potential of long-range quantum information opens new avenues for metrology [44, 174] and entanglement distribution [175]. The ability to distribute entanglement across spatially separated regions has already been used for recent loophole-free tests of Bell's inequality [8, 9, 176]. In this chapter, we are interested in coupling N parameters to N different

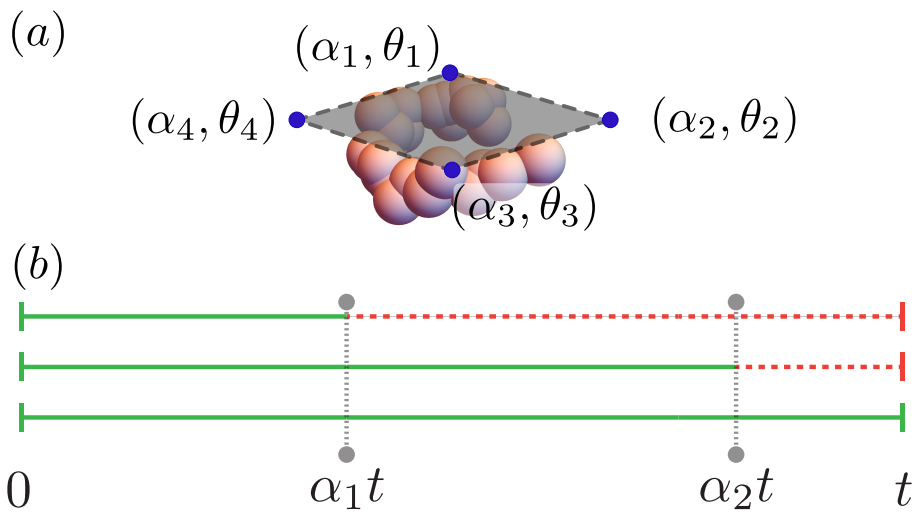

Figure 5.1: (a) An illustration of the network setup in a nanoscale NMR setting. Nodes, located at different points relative to a large molecule, share an entangled state; at each node there is both an unknown parameter θ_i and a known relative weight α_i. We are concerned with estimating $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$. (b) Illustration of the partial time evolution protocol for three qubits. Solid green segments of the timeline represent periods when a qubit is evolving due to coupling to the local parameter θ_i, while dashed red segments represent periods after the qubit stops evolving. The switches occur at times corresponding to the qubits' weights in the final linear combination. The weight of the last qubit is $\alpha_3 = 1$.

systems, which may be spatially separated, and measuring a function of all of them (see Fig. 5.1a) such as a single mode of a spatially varying field. For now, we focus on linear functions, but in Section 5.6 we will use our linear-function protocol as a subroutine in a more general procedure. Such measurements may be of interest in geodesy, geophysics, or medical imaging [177–181], but in this chapter we focus on potential application to nanoscale nuclear magnetic resonance (NMR) imaging. Later in this chapter we will discuss precisely how our method might apply in this setting.

The function q we wish to measure is a weighted sum of the deterministic individual parameters θ_i, where i indexes the individual systems and each weight is denoted by a known real number α_i,

$$q = \sum_{i=1}^{N} \alpha_i \theta_i = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}. \tag{5.1}$$

In this chapter, we characterize the advantage entanglement provides in this setting and construct an optimal strategy equivalent to turning some qubits' evolution "on" and "off" for time proportional to the weight with which their parameter contributes to the function q (see Fig. 5.1b). With this scheme of "partial time evolution," we can measure a linear function with the minimum variance permitted by quantum mechanics, which can be viewed as an extension of the Heisenberg limit to linear combinations. We will also show that our method can protect the secrecy of the result, allowing the network as a whole to perform a measurement without eavesdroppers learning any details of $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$.

5.2 Setup

We consider a system in which there are N sensor nodes. Each sensor node i possesses a single qubit coupled to an unknown parameter θ_i unique to each node. We suppose that the state evolves unitarily under the Hamiltonian

$$\hat{H} = \hat{H}_c(t) + \sum_{i=1}^{N} \frac{1}{2} \theta_i \hat{\sigma}_i^z. \tag{5.2}$$

Here, $\hat{H}_c(t)$ is a time-dependent control Hamiltonian chosen by us, which may include coupling to additional ancilla qubits and $\hat{\sigma}_i^{x,y,z}$ are the Pauli operators acting

on qubit i. We wish to measure the quantity q defined in Eq. (5.1). We assume that $\forall i : |\alpha_i| \leq 1$ and additionally that there is at least one α_i such that $\alpha_i = 1$. These conditions simply set a scale for the function, and for an arbitrary $\boldsymbol{\alpha}$ all that is needed is division by the largest α_i to meet this requirement. As an example, a network with two nodes interested in measuring the contrast between those nodes would set $\boldsymbol{\alpha} = (1, -1)$ to measure $\theta_1 - \theta_2$. We would like to establish how well an arbitrary measurement of $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ can be made and what the best measurement protocol is for doing so. By "protocol" we mean three different choices: (1) which input state we begin with, (2) what auxiliary control Hamiltonian $\hat{H}_c(t)$ we implement, and (3) how the final measurement is made.

We define the quality of measurement in terms of an estimator, Q, constructed from experimental data. (Throughout this chapter, we denote operators with hats, vectors by boldface, quantities to be estimated by lowercase, and corresponding estimators by uppercase.) We assume that the estimator is unbiased, so that its expectation value is the true value $\mathbf{E}[Q] = q$. Then our metric for the quality of the measurement is the average squared error, or variance, of the estimator,

$$\operatorname{Var} Q = \mathbf{E}\left[(Q - q)^2\right]. \tag{5.3}$$

If measurements of θ_i can be made locally with accuracy $\operatorname{Var} \Theta_i$ for an estimator Θ_i, then we could compute the linear combination by local measurements and classical computation. In this case, the variance is given by classical statistical theory as $\operatorname{Var} Q = \|\boldsymbol{\alpha}\|^2 \operatorname{Var} \Theta_0$ assuming that $\operatorname{Var} \Theta_i$ is identical at each site and equal to $\operatorname{Var} \Theta_0$. A measurement of an individual θ_i in Eq. (5.2) can be made in time t with

a variance of $1/t^2$ [169]. Therefore, our entanglement-free figure of merit is

$$\operatorname{Var} Q \geq \frac{\|\boldsymbol{\alpha}\|^2}{t^2}. \tag{5.4}$$

We consider this the standard quantum limit for networks. To compare to the typical case where N independent qubits measure a single parameter, consider the average $\bar{\theta}$, which is equivalent to setting all $\alpha_i = 1$ and then using $\bar{\Theta} = Q/N$ to obtain $\operatorname{Var} \bar{\Theta} = 1/Nt^2$. It is our goal in this chapter to present a means to improve on the limit in Eq. (5.4).

5.3 Heisenberg Limit for Sensor Networks

5.3.1 Using Fisher Information Matrix

Our task is to perform parameter estimation on a quantum system evolving under some set of parameters $\{\theta_i\}$ linearly coupled to sensor qubits as in Eq. (5.2) [182–185]. Although we are only interested in measuring a single number, we still need to treat a system that has many parameters in the evolution, necessitating the use of a multi-parameter theory as in Refs. [186–194]. It is known from classical estimation theory that, given a probability distribution $p(z)$ over a set of outcomes z that depends on a number of parameters, all estimators of the parameters obey the Cramér-Rao inequality [195, 196],

$$\Sigma \geq \frac{F^{-1}}{M}. \tag{5.5}$$

Here, M is the number of experiments performed, F is the Fisher information matrix (see below), and Σ is the covariance matrix, where $\Sigma_{ij} = \mathbf{E}\left[(\Theta_i - \theta_i)\,(\Theta_j - \theta_j)\right]$. The

inequality is a matrix inequality, meaning that $M\Sigma - F^{-1}$ is positive semidefinite. We will concern ourselves with the single-shot Fisher information, and set $M = 1$ from now on. The Fisher information matrix captures how each parameter changes the probability distribution of outcomes,

$$F_{ij} = \int p(z) \left(\frac{\partial \ln p(z)}{\partial \theta_i} \right) \left(\frac{\partial \ln p(z)}{\partial \theta_j} \right) \mathrm{d}z. \tag{5.6}$$

This bound is a purely classical statement about probability distributions, and is saturated asymptotically using a maximum-likelihood estimator [197]. Note that although we have presented the formulas for the Fisher information matrix, in the case of a single parameter the Fisher information will be a scalar which can be obtained by setting $i = j$ in Eq. (5.6).

Quantum theory bounds the probability distributions that can result from a state evolved under a parameter-dependent unitary operation [182]. We thus define the quantum Fisher information F_Q for a process with a given initial state as the maximization of the Fisher information over all possible measurement schemes. This gives rise to the quantum Cramér-Rao bound (QCRB), which simply replaces F with F_Q in Eq. (5.5). A matrix element of F_Q for a pure state evolving under a Hamiltonian $\hat{H}$ is given by

$$(F_Q)_{ij} = 4t^2 \left[\langle \hat{g}_i \hat{g}_j \rangle - \langle \hat{g}_i \rangle \langle \hat{g}_j \rangle \right], \tag{5.7}$$

where $\hat{g}_i = \left(\partial \hat{H} / \partial \theta_i \right)$ is the generator corresponding to parameter i. For instance, in Eq. (5.2) the generator $\hat{g}_i$ is the operator $\frac{1}{2}\hat{\sigma}_i^z$. Unlike the Cramér-Rao bound, the QCRB cannot always be satisfied, even asymptotically. However, in the setting

of this chapter, where all generators commute, it can be [186]. Equation (5.5) then takes the form:

$$\Sigma \geq F^{-1} \geq F_Q^{-1}. \tag{5.8}$$

To formulate the appropriate Cramér-Rao bound in the case where the quantity we wish to estimate is a linear combination of the θ_i, we simply use the fact that the variance of a linear combination $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ can be written as $\boldsymbol{\alpha}^T \Sigma \boldsymbol{\alpha}$. It follows immediately from Eq. (5.8) that

$$\operatorname{Var} Q \geq \boldsymbol{\alpha}^T F_Q^{-1} \boldsymbol{\alpha}. \tag{5.9}$$

Note that although we began by considering the full covariance matrix, we now focus on just a *single* scalar $\boldsymbol{\alpha}^T F_Q^{-1} \boldsymbol{\alpha}$ because our quantity of interest is a *single* linear transformation of the original parameters.

In order to properly define the Cramér-Rao bound, it is necessary to consider the fact that F and F_Q are only positive semi-definite and not necessarily invertible. For instance, if a parameter has no effect on probabilities at all, then it cannot be estimated from experimental results and the bound on the variance of its estimator is undefined. To sidestep this issue, we can instead look at $\tilde{F}_Q$, the quantum Fisher information projected onto its own image [194], assuming that $\boldsymbol{\alpha}$ has no overlap with the kernel of F_Q. This matrix (and its inverse) are now both positive definite, meaning they can always be inverted. Equation (5.9) is therefore always well-defined if $\tilde{F}_Q$ is used.

Since $\tilde{F}_Q$ is Hermitian and positive definite, $\sqrt{\tilde{F}_Q}$ is Hermitian. We can then

write the following for an arbitrary real $\boldsymbol{b}$ by invoking the Cauchy-Schwarz inequality:

$$\boldsymbol{\alpha}^T \tilde{F}_Q^{-1} \boldsymbol{\alpha} = \frac{\|\sqrt{\tilde{F}_Q^{-1}}\boldsymbol{\alpha}\|^2 \|\sqrt{\tilde{F}_Q}\boldsymbol{b}\|^2}{\boldsymbol{b}^T \tilde{F}_Q \boldsymbol{b}} \tag{5.10}$$

$$\geq \frac{\|\boldsymbol{\alpha}^T \sqrt{\tilde{F}_Q^{-1}}\sqrt{\tilde{F}_Q}\boldsymbol{b}\|^2}{\boldsymbol{b}^T \tilde{F}_Q \boldsymbol{b}} \tag{5.11}$$

$$\geq \frac{\|\boldsymbol{\alpha}^T \boldsymbol{b}\|^2}{\boldsymbol{b}^T \tilde{F}_Q \boldsymbol{b}}. \tag{5.12}$$

Taking $\boldsymbol{b}$ to be the bth element of the standard basis gives

$$\operatorname{Var} Q \geq \boldsymbol{\alpha}^T \tilde{F}_Q^{-1} \boldsymbol{\alpha} \geq \frac{\alpha_b^2}{\left(\tilde{F}_Q\right)_{bb}}. \tag{5.13}$$

Here, $\left(\tilde{F}_Q\right)_{bb}$ is the quantum Fisher information for a single parameter, as defined by Eq. (5.7). In Ref. [184], it was shown that for any time-dependent control Hamiltonian $\hat{H}_c(t)$, including those with ancilla qubits,

$$\left(\tilde{F}_Q\right)_{bb} \leq t^2 \|\hat{g}_b\|_s^2. \tag{5.14}$$

Here $\|\hat{g}_b\|_s$ is the operator seminorm (difference between the largest and smallest eigenvalues) of the generator corresponding to parameter θ_b. Our final bound comes from applying this condition and recognizing that the formula must hold for all b:

$$\operatorname{Var} Q \geq \max_b \frac{\alpha_b^2}{t^2 \|\hat{g}_b\|_s^2}. \tag{5.15}$$

We emphasize that Eq. (5.15) remains true no matter what time-dependent control $\hat{H}_c(t)$ is applied.

In Eq. (5.2), all $\hat{g}_b = \frac{1}{2}\hat{\sigma}_b^z$, $\|\hat{g}_b\|_s = 1$, and we find a bound,

$$\operatorname{Var} Q \geq \max_i \frac{\alpha_i^2}{t^2} = \frac{1}{t^2}. \tag{5.16}$$

Here we have used the fact that the largest $\alpha_i = 1$. If we want to estimate the average of the θ_i, then all qubits are equally weighted and the desired quantity is $\bar{\theta} = q/N$, so $\operatorname{Var}\bar{\Theta} \geq 1/N^2t^2$ and we reproduce the desired Heisenberg scaling which is more precise than the $1/N$ in Eq. (5.4). However, note that if we wanted to estimate only a single θ_i, then we would not benefit from the entanglement. In general, we can, for some situations, greatly improve the precision of parameter esitmation with nonlocal techniques if the parameter itself is also non-local. Our bound allows us to explore the full range of possible $\boldsymbol{\alpha}$ between these two extremes. Compared to the bound on unentangled states [Eq. (5.4)], Eq. (5.15) simply picks out the largest contribution due to uncertainty from a single site. Equation (5.15) can be viewed as an extension of the usual Heisenberg bound to linear combinations of parameters.

We can illustrate the above argument by optimizing over the space of all control Hamiltonians $\hat{H}_c(t)$. As this is computationally expensive, we limit ourselves to a two-qubit sensor network with no ancillas. The Hamiltonians we optimize over include enough operators to provide universal control on two qubits, meaning we can effectively modify the input state as well as the final measurement basis in order to optimize the Fisher information. In order to test the form of our bound, Eq. (5.15), which depends both on relative weights of each parameter and the underlying generator, we couple θ_1 to a generator $\hat{\sigma}_1^z$ which has $\|\hat{\sigma}_1^z\|_s = 2$. We leave the second qubit coupled to a generator $\frac{1}{2}\hat{\sigma}_2^z$ as in Eq. (5.2). The bound corresponding to the first qubit from Eq. (5.15) is $\alpha_1^2/4t^2$ and that of the second qubit is α_2^2/t^2. In our numerics, we set $\alpha_1 = t = 1$, meaning the two bounds are $1/4$ and α_2^2. Our analytic

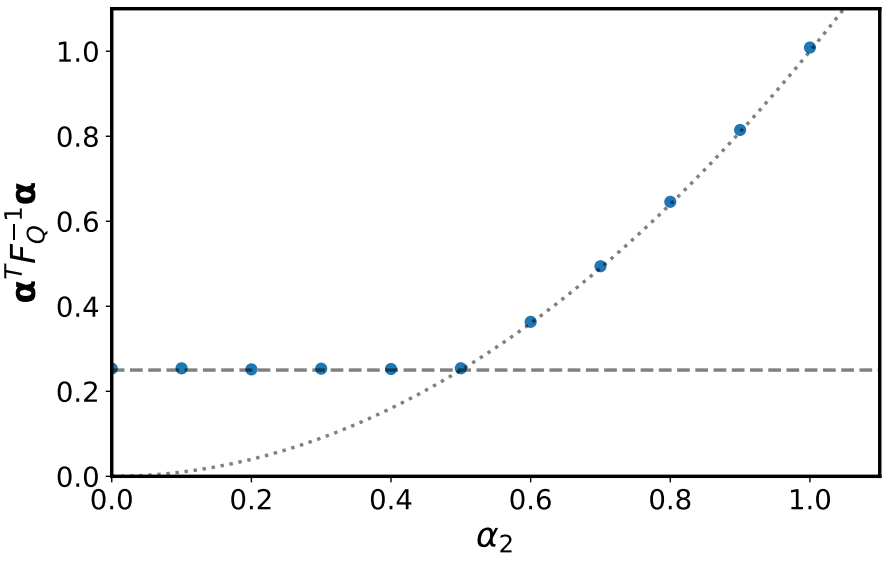

Figure 5.2: Numerical optimization of $\boldsymbol{\alpha}^T F_Q^{-1} \boldsymbol{\alpha}$ for two qubits with $\alpha_1 = 1$ compared to the bound predicted by our analytic result. Each point is generated by running a gradient descent algorithm until convergence; the control parameters begin at small random values. The dashed (dotted) line is the analytic bound derived from the first (second) qubit. As α_2 increases, the second qubit becomes the source of the relevant bound.

result leads us to believe therefore that if $\alpha_2^2 > 1/4$, the minimum possible variance should be α_2^2. However, if $\alpha_2^2 < 1/4$, then the lower bound should be $1/4$. That behavior is precisely what we find through the numerical optimization shown in Fig. 5.2, confirming Eq. (5.15).

5.3.2 Using Single-Parameter Bounds

It is tempting to dismiss the above argument as unnecessarily complicated, as the ultimate quantity of interest is only a single parameter. Why not simply apply the Cramér-Rao bound directly to $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ instead of using the matrix approach? We will now show that the single-parameter bound that arises from naive application

of the Cramér-Rao bound is looser than Eq. (5.15). This gap occurs because the single-parameter bound can only be applied if there is only one unknown parameter controlling the evolution of the input state, which implicitly places a constraint on the other components of the field. Later, we will discuss how the single-parameter approach can be amended to take this into account and agree with Eq. (5.15).

To apply the single-parameter Cramér-Rao bound to our evolution Hamiltonian Eq. (5.2), we consider the Hamiltonian as $\frac{1}{2}\boldsymbol{\theta}\cdot\hat{\boldsymbol{\sigma}}$ where $\hat{\boldsymbol{\sigma}}$ is simply a vector of operators whose ith element is $\hat{\sigma}_i^z$. We then rewrite $\boldsymbol{\theta}$ in a new basis,

$$\boldsymbol{\theta} = \sum_{i=0}^{N-1} (\boldsymbol{\alpha}_i \cdot \boldsymbol{\theta})\, \boldsymbol{\beta}_i. \tag{5.17}$$

We assume that $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}$ and that the other $\boldsymbol{\alpha}_{i>0}$ make up a basis. The set of vectors $\boldsymbol{\beta}_i$ is then a dual basis such that $\boldsymbol{\alpha}_i \cdot \boldsymbol{\beta}_j = \delta_{ij}$. (For this basis as well, we will drop the subscript 0 to indicate that this particular vector corresponds to the parameter of interest.) The advantage of rewriting $\boldsymbol{\theta}$ this way is that we can now identify the term in the Hamiltonian which is proportional only to $\boldsymbol{\alpha}\cdot\boldsymbol{\theta}$. The generator corresponding to the quantity $\boldsymbol{\alpha}\cdot\boldsymbol{\theta}$ is:

$$\hat{g} = \frac{\partial \hat{H}}{\partial\,(\boldsymbol{\alpha}\cdot\boldsymbol{\theta})} = \frac{\boldsymbol{\beta}\cdot\hat{\boldsymbol{\sigma}}}{2}. \tag{5.18}$$

To obtain the quantum Fisher information corresponding to this generator, we consider the variance of the operator $\hat{g}$. The maximum variance of this generator is given by the operator seminorm [184]. Using this fact, we can write:

$$F_Q \leq t^2 \|\hat{g}\|_s^2 = t^2 \left(\sum_i |\beta_i|\right)^2. \tag{5.19}$$

In general the bound on $\operatorname{Var} Q$ derived from Eq. (5.19) is a looser lower bound

than Eq. (5.16). For example, with $\boldsymbol{\alpha} = (1, \frac{1}{2})$ and $\boldsymbol{\alpha}_1 = (\frac{1}{2}, -1)$, this implies that $\boldsymbol{\beta} = (\frac{4}{5}, \frac{2}{5})$. Equation (5.19) would suggest that

$$\operatorname{Var} Q \geq \frac{25}{36t^2}, \tag{5.20}$$

which is looser than the $1/t^2$ given by Eq. (5.15). This discrepancy can be addressed by thinking more closely about the process of choosing a new basis. We will use the seminorm condition again to bound the maximum possible Fisher information. To start calculating the seminorm, we express it in terms of the elements of $\boldsymbol{\beta}$:

$$\|\hat{g}\|_s = \|\sum_j \beta_j \frac{1}{2}\hat{\sigma}_j^z\|_s = \sum_j |\beta_j| \,. \tag{5.21}$$

We will now show that it is possible to choose a basis such that the seminorm in Eq. (5.21) goes to infinity. This shows that the approach which led us to Eq. (5.19) should not be applied blindly, and we will then discuss how to control for this issue. First, an illustration of the bound diverging. Suppose that in a two-parameter problem, the basis vectors we choose are $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$. It can then be shown by direct computation of the matrix inverse that yields the dual basis that the implied maximum Fisher information from Eq. (5.21) is:

$$F \leq \|\hat{g}\|_s = \frac{\alpha_2' + \alpha_1'}{|\alpha_1\alpha_2' - \alpha_1'\alpha_2|}. \tag{5.22}$$

If we then choose $\boldsymbol{\alpha}' = (\alpha_1/\alpha_2 + \varepsilon, 1)$, it follows that:

$$F \leq \frac{1 + \varepsilon + \frac{\alpha_1}{\alpha_2}}{\varepsilon\alpha_2}. \tag{5.23}$$

As $\varepsilon \to 0$, this becomes arbitrarily large. From this we conclude that our previous approach was ill-advised as it can yield arbitrarily small lower bounds on the es-

timator variance – using this basis, we would conclude that the right-hand side of Eq. (5.19) could be ∞.

In order to produce a useful bound from Eq. (5.21), we recognize that any possible choice of basis must yield a valid bound. Therefore, rather than look at one particular basis (as we did in deriving Eq. (5.19)) we instead need to optimize for the highest lower bound over all possible choices of basis. Finding the tightest bound on Fisher information will then produce the highest lower bound on parameter uncertainty. To do this, we first write the following chain of inequalities using the relationship of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$1 = \sum_j \alpha_j \beta_j \leq \sum_j |\alpha_j \beta_j| \leq \sum_j |\beta_j| \,, \tag{5.24}$$

where the last line follows due to the fact that $|\alpha_j| \leq 1$. Note that we can achieve equality,

$$\sum_j |\beta_j| = 1, \tag{5.25}$$

by taking the other $N-1$ basis vector $\boldsymbol{\alpha}_j$ to be unit vectors $\boldsymbol{e}_j$ in the standard basis, making sure that the j that does not appear has $\alpha_j \neq 0$ to ensure the entire space is spanned. Now we look to the minimum possible value of $\|\hat{g}\|_s$. The minimum possible value is interesting to us because the minimum $\|\hat{g}\|_s$ will be the choice of basis for which the bound on Fisher information is tightest.

It follows from Eq. (5.21) and Eq. (5.24) that the minimum seminorm $\|\hat{g}\|_s$ is equal to 1, implying that the maximum value for $\operatorname{Var} \hat{g}$ is 1/4 [184]. Using this to optimize the bound in Eq. (5.19) over all possible choices of re-parameterization implies that $\operatorname{Var} \boldsymbol{\alpha} \cdot \boldsymbol{\theta} \geq 1/t^2$, just as we found in Eq. (5.16).

The single-parameter bound *is* applicable in our situation, but it requires careful accounting of the influence of other parameters in the problem. The reason that our previous results such as Eq. (5.15) do not hold in this case is that Cramér-Rao bound does not apply if we can take advantage of constraints on the signal field $\boldsymbol{\theta}$ to improve our estimation strategy.

5.3.2.1 Single-Parameter Bounds With Prior Information

These naive single-parameter bounds can be applied and saturated if the field structure is known before the measurement takes place. To demonstrate, suppose that for a set of fields $\boldsymbol{\theta}$ where we wish to learn $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$, we know that the fields are proportional to α_i. Then we can write the total field as:

$$\boldsymbol{\theta} = q\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|^2}, \tag{5.26}$$

and our goal is to estimate $q = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$. This is now a truly one-parameter problem, enabling a new strategy which saturates Eq. (5.19). By defining $\boldsymbol{w}$ as a new vector such that $w_i = \operatorname{sgn}(\alpha_i)$, we can measure the quantity

$$\boldsymbol{w} \cdot \boldsymbol{\theta} = q\frac{\sum |\alpha_i|}{\|\boldsymbol{\alpha}\|^2}. \tag{5.27}$$

Since $\boldsymbol{w}$ is a linear combination which satisfies the condition $|w_i| \leq 1$, we can estimate $q' = \boldsymbol{w} \cdot \boldsymbol{\theta}$ with accuracy bounded by $1/t^2$ as shown in Sec. 5.4. Then:

$$\operatorname{Var}(Q') = \operatorname{Var}\left(Q\frac{\sum |\alpha_i|}{\|\boldsymbol{\alpha}\|^2}\right) \geq \frac{1}{t^2}, \tag{5.28}$$

$$\implies \operatorname{Var} Q \geq \frac{\|\boldsymbol{\alpha}\|^4}{t^2\left(\sum |\alpha_i|\right)^2}. \tag{5.29}$$

This saturates the bound in Eq. (5.19). The reason we are able to outperform Eq. (5.15) is that we have assumed something about the structure of the field which reduces it to a lower-dimensional problem. This is only possible by using knowledge about components of $\boldsymbol{\theta}$ not parallel to $\boldsymbol{\alpha}$. Otherwise, there is no guarantee that $\boldsymbol{\theta}$ will be proportional to $\boldsymbol{\alpha}$. In general cases $\boldsymbol{w}\cdot\boldsymbol{\theta}$ will contain noise from "undesired" components.

In many situations where the field structure is known, new strategies can be introduced which may outperform our previous results, even asymptotically. Consider as an example a case with a field:

$$\boldsymbol{\theta} = \theta\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|^2} + \theta_\gamma\boldsymbol{\gamma}, \tag{5.30}$$

where $\boldsymbol{\alpha}\cdot\boldsymbol{\gamma} = 0$ and θ_γ is a nuisance parameter describing the field magnitude orthogonal to $\boldsymbol{\alpha}$. Any field can be written in this way to separate out the $\boldsymbol{\alpha}$ component. Suppose we measure $\boldsymbol{w}\cdot\boldsymbol{\theta}$. We know that:

$$\operatorname{Var} Q' \geq \frac{1}{t^2} \tag{5.31}$$

is an achievable bound. By writing $\boldsymbol{w} = c_\alpha\boldsymbol{\alpha} + c_\gamma\boldsymbol{\gamma}$, decomposing $\boldsymbol{w}$ into its only relevant components, we can obtain the following bound on $\operatorname{Var} Q$:

$$\operatorname{Var} Q \geq \frac{1}{t^2 c_\alpha} + c_\gamma^2\|\boldsymbol{\gamma}\|^2. \tag{5.32}$$

Therefore, the optimal strategy is to pick a $\boldsymbol{w}$ which maximizes $\boldsymbol{w}\cdot\boldsymbol{\alpha}$ while minimizing (preferably to zero) $\boldsymbol{w}\cdot\boldsymbol{\gamma}$. However, in general, learning the structure of the field perpendicular to $\boldsymbol{\alpha}$ is just as difficult as learning the component parallel to $\boldsymbol{\alpha}$,

so beginning from a state of ignorance, it is still optimal to measure $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ rather than a different linear combination.

The bound in Eq. (5.19) can actually be found by other statistical methods which fully treat the initial multi-parameter structure, for instance, the constrained Cramér-Rao bound of Ref. [198]. It can also be derived from the Van Trees inequality [199] by assuming that we have pre-existing knowledge that the components of $\boldsymbol{\theta}$ perpendicular to $\boldsymbol{\alpha}$ have a normal distribution of width ε and then taking the limit $\varepsilon \to 0$.

If rather than a constraint we simply have some initial information in the form of a prior distribution, the Van Trees inequality (which takes into account that prior information) will reduce to the Cramér-Rao bound in the limit of many measurements. This is because the information gained from measurements scales linearly with the number of measurements while the prior information is static.

5.4 Protocols

We now present two protocols that saturate the bound of Eq. (5.16) and are therefore optimal. The first begins from the conceptually simple Greenberger-Horne-Zeilinger (GHZ) state or a spin-squeezed state and uses time-dependent control during phase accumulation to produce an output state sensitive to the desired $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$, while the second method uses a more complicated initial state but requires no control during the phase accumulation.

5.4.1 Protocols Involving Time-Dependent Control

5.4.1.1 Using GHZ Input State

We start by considering an N-qubit GHZ state:

$$\frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}\right). \tag{5.33}$$

Under $\hat{\sigma}^z$ evolution, each $|1\rangle$ accumulates a phase relative to $|0\rangle$. By allowing qubits to accumulate phase proportional to the desired weight α_i, we obtain a final state in which $|1\rangle^{\otimes N}$ has accumulated a total phase of $\boldsymbol{\alpha}\cdot\boldsymbol{\theta}t$ relative to $|0\rangle^{\otimes N}$. We refer to our protocol as "partial time evolution" because it relies on a qubit undergoing evolution for a fraction of the total measurement time (see Fig. 5.1). We can realize this by applying $\hat{\sigma}_i^x$ to a qubit at time $t_i = t\,(1+\alpha_i)\,/2$ so that the qubit evolution will be identical to evolving it for a time $\alpha_i t$. Note that if there is a fixed experimental time t, this scheme can realize values of $\alpha_i \in [-1, 1]$, which motivates our restrictions on the values of individual α_i. Specifying this sequence of gates identifies the $\hat{H}_c(t)$ which defines the protocol. The result of this protocol is an effective evolution according to the unitary operator

$$\hat{U}(t) = e^{-i\frac{t}{2}\sum_{i=1}^{N}\alpha_i\theta_i\hat{\sigma}_i^z}. \tag{5.34}$$

Under this evolution, the final state is:

$$\frac{1}{\sqrt{2}}\left(e^{-i\frac{t}{2}q}\,|0\rangle^{\otimes N} + e^{i\frac{t}{2}q}\,|1\rangle^{\otimes N}\right). \tag{5.35}$$

Now we make a measurement of the overall parity of the state, $\hat{P} = \bigotimes_{i=1}^{N}\hat{\sigma}_i^x$. The details of this measurement and calculation of $\langle\hat{P}\rangle$ are given in Ref. [23]; notably,

the measurement can be performed locally at each site. Measurement of the time-dependent expectation value $\langle \hat{P} \rangle(t)$ allows for the estimation of Q with accuracy [200]

$$\operatorname{Var} Q = \frac{\operatorname{Var} \hat{P}(t)}{\left(\partial \left\langle \hat{P} \right\rangle / \partial q\right)^2} = \frac{\sin^2 qt}{t^2 \sin^2 qt} = \frac{1}{t^2}, \tag{5.36}$$

saturating the bound in Eq. (5.16) and Fig. 5.2.

We can also directly evaluate F_Q and F for this protocol. F_Q can be found by noting that this protocol is identical to evolution under the Hamiltonian $\frac{1}{2}\sum \alpha_i \theta_i \hat{\sigma}_i^z$. Therefore the quantum Fisher information matrix F_Q is simply

$$(F_Q)_{ij} = t^2 \left[\left\langle \alpha_i \hat{\sigma}_i^z \alpha_j \hat{\sigma}_j^z \right\rangle - \left\langle \alpha_i \hat{\sigma}_i^z \right\rangle \left\langle \alpha_j \hat{\sigma}_j^z \right\rangle\right] = \alpha_i \alpha_j t^2. \tag{5.37}$$

Furthermore, we can show that this F_Q satisfies the second inequality in Eq. (5.13). The inverse of $\tilde{F}_Q$ can be easily written, as F_Q simply projects onto $\boldsymbol{\alpha}$. In order to get $\tilde{F}_Q^{-1}\tilde{F}_Q = \boldsymbol{\alpha}^T\boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|^2$ (identity on the image of F_Q), we must have

$$\tilde{F}_Q^{-1} = \frac{\alpha_i \alpha_j}{t^2 \|\boldsymbol{\alpha}\|^4}. \tag{5.38}$$

$\boldsymbol{\alpha}^T \tilde{F}_Q^{-1} \boldsymbol{\alpha}$ is then equal to $1/t^2$, saturating the second inequality in Eq. (5.13) for the basis vector $\boldsymbol{b}$ corresponding to the largest $\boldsymbol{\alpha}$ component, $\alpha_b = 1$.

To evaluate the classical Fisher information in this case, we note that the final measurement [200] projects onto one of two outcomes with probability $\sin^2(\boldsymbol{\alpha}\cdot\boldsymbol{\theta}t/2)$ and $\cos^2(\boldsymbol{\alpha}\cdot\boldsymbol{\theta}t/2)$. Therefore the classical Fisher information is simply:

$$F = \frac{\left(\frac{\partial\left(\sin^2 \frac{\boldsymbol{\alpha}\cdot\boldsymbol{\theta}t}{2}\right)}{\partial \boldsymbol{\alpha}\cdot\boldsymbol{\theta}}\right)^2}{\sin^2 \boldsymbol{\alpha}\cdot\boldsymbol{\theta}t/2} + \frac{\left(\frac{\partial\left(\cos^2 \frac{\boldsymbol{\alpha}\cdot\boldsymbol{\theta}t}{2}\right)}{\partial \boldsymbol{\alpha}\cdot\boldsymbol{\theta}}\right)^2}{\cos^2 \boldsymbol{\alpha}\cdot\boldsymbol{\theta}t/2} \tag{5.39}$$

$$= t^2. \tag{5.40}$$

This Fisher information also implies the variance bound in Eq. (5.16).

It may seem surprising that an optimal measurement can be one in which most qubits spend some of the measurement time idle. Since more time yields more signal, intuition suggests that the most effective strategy would make better measurements on the less-weighted qubits rather than keep them off for much of the measurement time. For example, by disentangling a qubit from the larger state halfway through the protocol, a separate measurement could be made on $\theta_1 + \frac{1}{2}\theta_2$ and $\frac{1}{2}\theta_2$, which appears to yield more information than just measuring the quantity of interest $\theta_1 + \frac{1}{2}\theta_2$. This reasoning fails because there is no way to use information about θ_2 to improve an estimate of $\theta_1 + \frac{1}{2}\theta_2$ without also knowing about θ_1. Because we do not know about the individual parameters, only a measurement of the entire function is usable and our scheme is optimal in this case. However, once we account for pre-existing knowledge about the parameter values (drawn from physically-motived estimates or less-precise previous measurements) our bound will instead apply in the regime of asymptotically many measurements ($M \gg 1$) and in that setting our scheme will also saturate it [201]. This is because the value of prior knowledge becomes increasingly low as we accumulate measurement data.

One advantage of this protocol is that an eavesdropper cannot learn the result of the network measurement by capturing a subset of the nodes' $\hat{\sigma}^x$ measurement results. This privacy can be shown by tracing out the first qubit in Eq. (5.35), which leaves no phase information in the resulting mixed state. The central node can receive the measurement outcomes from all other nodes but keep its own secret, and no eavesdropper is able to extract information from the broadcasted results.

This is true even if the central node's qubit is unweighted (i.e., $\alpha_i = 0$), which follows simply from the properties of the GHZ state.

5.4.1.2 Using Spin Squeezed States

The perfect security of the GHZ state arises because obtaining the measurement result requires every qubit, but this also implies an extreme sensitivity to noise. This noise can be a serious problem for metrological applications [202,203]. Because the GHZ state decoheres faster than an individual qubit, the advantage provided by entanglement is nullified if the interrogation time of the qubit is limited by its coherence time [204]. However, in many settings the time spent on a single measurement will be much shorter than the decoherence time, for instance, to gather data on short timescales. In these cases, GHZ states still provide a metrological advantage. Note that dynamical decoupling [205] or quantum error correction [206, 207] could be used to lengthen the effective decoherence time in some cases.

In other situations, however, it may be that decoherence is the dominant concern. In these situations, the best strategy uses a highly-symmetric entangled state which is more robust to noise than the GHZ state [204]. Under dephasing, these states can still offer a constant factor improvement over unentangled metrology. In this section, we show that spin-squeezed states can also function as inputs to the partial time evolution protocol, and so may be good candidates for a sensor network operating in a situation where decoherence limits the interrogation time. Squeezed states are collective spin states which, due to entanglement, have reduced variance

along one axis of the collective Bloch sphere at the cost of increased variance along an orthogonal axis [169, 208]. Recently, it has been shown that these states may allow Heisenberg-scaling measurements even without single-particle detection, which makes them very attractive for experimental implementations [209].

We consider a state whose overall spin vector is aligned along $+x$, such that $\langle \hat{\sigma}_i^x \rangle \approx 1$. We assume that the other spin components have zero expectation value, but that the variance of the collective spin projection $\hat{J}_y = \frac{1}{2}\sum_i \hat{\sigma}_i^y$ is decreased while the variance of $\hat{J}_z$ is increased. We quantify this effect through the spin-squeezing parameter ξ [169],

$$\xi = \sqrt{\frac{\mathrm{Var}\, \hat{J}_y}{N/4}}. \tag{5.41}$$

Suppose that we perform Ramsey interferometry on such a state [169, 200]. The protocol includes both partial time evolution $\hat{U}(t)$ and a final rotation pulse $\hat{R}_x\left(\frac{\pi}{2}\right) = \exp\left(-i\frac{\pi}{4}\sum_i \hat{\sigma}_i^x\right)$. A final measurement is made of the total spin projection $\hat{J}_z$ after applying these operations:

$$\left\langle \hat{J}_z(t) \right\rangle = \left\langle \hat{U}^\dagger(t) \hat{R}_x^\dagger\left(\frac{\pi}{2}\right) \hat{J}_z(0) \hat{R}_x\left(\frac{\pi}{2}\right) \hat{U}(t) \right\rangle, \tag{5.42}$$

$$= \frac{1}{2}\left\langle \sum_{i=1}^{N} \hat{\sigma}_i^x \sin\alpha_i\theta_i t + \hat{\sigma}_i^y \cos\alpha_i\theta_i t \right\rangle. \tag{5.43}$$

If we specify that this expectation is to be taken over a squeezed state with $\langle \hat{\sigma}_i^x \rangle \approx 1$ and $\langle \hat{\sigma}_i^y \rangle = 0$, then our signal will be sensitive only to $\boldsymbol{\alpha}\cdot\boldsymbol{\theta}$ if each individual phase is small:

$$\left\langle \hat{J}_z(t) \right\rangle_{\text{squeezed}} \approx \frac{1}{2}\sum_{i=1}^{N} \sin\alpha_i\theta_i t \approx \frac{t}{2}\sum_{i=1}^{N} \alpha_i\theta_i. \tag{5.44}$$

This shows that a squeezed state can be used for measurements of linear functions.

The sensitivity can then be calculated just as in Eq. (5.36),

$$\operatorname{Var} Q = \left.\frac{\operatorname{Var} \hat{J}_z(t)}{\left(\partial \left\langle \hat{J}_z(t) \right\rangle / \partial q\right)^2}\right|_{q=0} = \frac{\operatorname{Var} \hat{J}_y}{t^2/4} = \frac{N\xi^2}{t^2}. \tag{5.45}$$

We evaluate the sensitivity at $q = 0$ because we are interested in small signals. Partial time evolution with spin-squeezed input beats the standard quantum limit if $\xi \leq \|\boldsymbol{\alpha}\|/\sqrt{N}$. Note that there are N components of $\boldsymbol{\alpha}$ and therefore $\|\boldsymbol{\alpha}\|$ will generally be of order $\sqrt{N}$ assuming that the moments of the field being measured are well distributed. Squeezed states can achieve squeezing proportional to $N^{-1/2}$ [169, 208], which approaches the bound in Eq. (5.16) up to numerical prefactors not scaling with N.

Other highly-entangled states such as Dicke states also have metrological value in the presence of noise and could also serve as input states to partial time evolution with similarly favorable scaling [193, 210–213].

5.4.2 Time-Independent Protocols

In this section, we present two other possible measurement schemes for linear combinations of parameters. Both of these differ from the protocols of Sec. 5.4.1 because they prepare a particular state and then allow for free evolution during phase accumulation, rather than using pulses to evolve for an effective time of $\alpha_i t$ on qubit i. We will present time-independent schemes that begin with both a GHZ-like state and the spin-squeezed state. Note that these protocols rely on assumptions about the size of signals θ_i or the evolution time t.

5.4.2.1 Using GHZ-like Input State

We begin by defining a single-qubit state $|\boldsymbol{\tau}\rangle$, where $\boldsymbol{\tau}$ is a vector whose elements are $\tau_j = -1, 0, 1$:

$$|\boldsymbol{\tau}\rangle = \bigotimes_{j=1}^{N} \begin{cases} |0\rangle & \tau_j \neq -1 \\ \\ |1\rangle & \tau_j = -1 \end{cases} . \tag{5.46}$$

We then define the entangled state $|\psi(\boldsymbol{\tau})\rangle$ as

$$|\psi(\boldsymbol{\tau})\rangle = \frac{1}{\sqrt{2}} \left(|\boldsymbol{\tau}\rangle + |-\boldsymbol{\tau}\rangle \right) . \tag{5.47}$$

This state can be understood as a general class that includes the GHZ state as the case $\tau_j = 1$ for all j. For every $\tau_j = -1$, spin j is flipped relative to the GHZ state, while for every $\tau_j = 0$, spin j is entirely disentangled.

In order to measure $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$, we will evolve $|\psi(\boldsymbol{\tau})\rangle$ under the Hamiltonian in Eq. (5.2) and then measure the following observable $\hat{\Pi}(\boldsymbol{\tau})$:

$$\hat{\Pi}(\boldsymbol{\tau}) = \bigotimes_{j} \left(\hat{\sigma}_x^j \right)^{\tau_j} . \tag{5.48}$$

That is, we multiply the outcomes of the individual projective σ^x measurements for each qubit which was originally entangled with the others ($\tau_j \neq 0$). It can be shown that probability distribution of this observable is

$$P\left(\hat{\Pi} = \pm 1 \middle| \boldsymbol{\tau}, \boldsymbol{\theta} \right) = \begin{cases} \cos^2 \left(\boldsymbol{\theta} \cdot \boldsymbol{\tau} t/2 \right) & 1, \\ \\ \sin^2 \left(\boldsymbol{\theta} \cdot \boldsymbol{\tau} t/2 \right) & -1 \end{cases} . \tag{5.49}$$

To create a final protocol, we will now randomize the choice of $\boldsymbol{\tau}$, which in turn means we will randomly select both the initial state and the final measurement.

An overall sensitivity to $\boldsymbol{\alpha}\cdot\boldsymbol{\theta}$ can be realized if the probability distribution for every individual spin to be τ_j is given by:

$$P(\tau_j) = \begin{cases} \frac{\alpha_j(\alpha_j \pm 1)}{2} & \tau_j = \pm 1 \\ 1 - \alpha_j^2 & \tau_j = 0 \end{cases} . \tag{5.50}$$

By then summing over $P(\boldsymbol{\tau})$, we find that $\left\langle \hat{\Pi}(\boldsymbol{\theta}, t) \right\rangle = 1 - t^2 \left(\boldsymbol{\alpha}\cdot\boldsymbol{\theta}\right)^2$ to lowest order in t. Since $\hat{\Pi}^2 = 1$, we can use the same approach as Eq. (5.36) to find that the sensitivity for this measurement is $\operatorname{Var} Q = 1/t^2$, leading to the same sensitivity as the time-independent protocol.

5.4.2.2 Using Spin-Squeezed States

To implement a time-independent protocol that makes use of a spin-squeezed input state, we will actually use a two-part measurement protocol. First we will derive a general expression that applies to both parts, and then show how they can be combined.

Much as in Sec. 5.4.1.2, we will use the Heisenberg evolution of the total angular momentum along one axis to evaluate the final observable. We can begin with the result of Eq. (5.43), but with two alterations. First, rather than $\hat{U}$ representing a partial time evolution on each qubit, instead it will be the full time evolution operator $\hat{U} = \exp\left(-it\sum \theta_i \hat{\sigma}_i^z\right)$. Second, we will add an additional operator at the beginning of the protocol, which we write as $\hat{Q}(\boldsymbol{\eta})$:

$$\hat{Q}\left(\boldsymbol{\eta}\right) = \bigotimes_{i=1}^{N} \hat{r}_z^i\left(\eta_i\right). \tag{5.51}$$

Here, $\hat{r}_z^i$ is the single-qubit rotation about the z axis. That is, we apply a qubit-dependent rotation about the z axis before we begin the evolution. The final operator $\hat{J}_z(t)$ will be:

$$\hat{J}_z(t) = \hat{Q}^\dagger(\boldsymbol{\eta})\hat{U}^\dagger(t)\hat{R}_x^\dagger\left(\frac{\pi}{2}\right)\hat{J}_z(0)\hat{R}_x\left(\frac{\pi}{2}\right)\hat{U}(t)\hat{Q}(\boldsymbol{\eta}). \tag{5.52}$$

The effect of $\hat{Q}$ is to add an additional phase to the evolution, meaning the final value for $\left\langle \hat{J}_z(t) \right\rangle$ can be found by substituting the angles $\theta_i t + \eta_i$ for $\alpha_i \theta_i t$ in Eq. (5.43). As a result, we find that the final expectation value is:

$$\left\langle \hat{J}_z(t) \right\rangle = \frac{1}{2}\left\langle \sum_{i=1}^{N} \hat{\sigma}_i^x \sin\left(\theta_i t + \eta_i\right) + \hat{\sigma}_i^y \cos\left(\theta_i t + \eta_i\right) \right\rangle. \tag{5.53}$$

By using the conditions that $\langle \hat{\sigma}_i^x \rangle \approx 1$ and $\langle \hat{\sigma}_i^y \rangle \approx 0$, we find that:

$$\left\langle \hat{J}_z(t) \right\rangle \approx \frac{1}{2}\sin\left(\theta_i t + \eta_i\right). \tag{5.54}$$

Now we introduce a two-step protocol. In the first step, we perform this sequence (prepare a spin-squeezed state, add qubit-dependent rotations, evolve, measure $\hat{J}_z$) with $\eta_i = \phi_i$, where $\cos\phi_i = \alpha_i$. We will call the quantity measured $\hat{J}_z^+$. Then, we repeat the process with $\eta_i = -\phi_i$, and call the resulting quantity $\hat{J}_z^-$. The expectation value of the sum of these quantities is:

$$\left\langle \hat{J}_z^+ + \hat{J}_z^- \right\rangle \approx \frac{1}{2}\sum_{i=1}^{N} \sin\left(\theta_i t + \phi_i\right) + \sin\left(\theta_i t - \phi_i\right) \tag{5.55}$$

$$= \sum_{i=1}^{N} \cos\phi_i \sin\theta_i t \approx \sum_{i=1}^{N} \alpha_i \theta_i t. \tag{5.56}$$

Here, as in Sec. 5.4.1.2, we have assumed that the phases to be detected, $\theta_i t$, are small enough to make the small-angle approximation.

In order to evaluate the sensitivity of this measurement, we look at the point of zero signal as in Eq. (5.45). At zero signal, $J_z^+ + J_z^-$ gives $\sum \alpha_i \sigma_i^y$. It can be shown that $\mathrm{Var} \sum \alpha_i \sigma_i^y \leq 4 \, \mathrm{Var} \, J_y$, and so, by the same calculations used in Eq. (5.45), the variance is no more than $4N\xi^2/t^2$. Note, however, that this assumes that both $\hat{J}_+$ and $\hat{J}_-$ are measured for time t. For a fairer comparison, we can replace t with $t/2$ so the time required for the two-step protocol is the same as for one time-dependent round. In this case, the sensitivity is no worse than $16N\xi^2/t^2$.

Interestingly, this two-step protocol requires only single-qubit operations once the initial squeezed state is created. This may make it a more tractable scheme for experimental realizations of quantum enhancements in measurements of linear combinations of parameters.

5.5 Entanglement-enhanced molecular NMR

Many applications of entangled sensor networks may emerge as distributed entanglement becomes easier to achieve. In this section we focus on an application which may be viable in the near future: nanoscale nuclear magnetic resonance (NMR) as a form of molecular microscopy. NMR has long been used to investigate the chemical composition of molecular structures and perform medical imaging [214]. The spatial resolution of NMR had been limited to a few micrometers until the recent advent of nitrogen-vacancy (NV) center magnetometers [215–217]. These magnetometers are sensitive to nanotesla magnetic fields with spatial resolution on the nanometer scale and can be used to image molecules or single proteins deposited on

a diamond layer with embedded NV centers [218–220].

Nanoscale NMR applications are a promising setting for entanglement-enhanced sensor networks. The electronic spin associated with an NV center in diamond can be operated as a two-level system whose free evolution results in the accumulation of phase dependent on the local magnetic field [217]. Because NV centers are useful platforms for quantum information processing, entangling protocols already exist and have been demonstrated experimentally [221–224]. Our protocol is particularly useful for studies of chemical or magnetic dynamics, such as Ref. [225], because the measurement timescale may be much shorter than the decoherence time of the GHZ state, making our noise-free treatment applicable.

Linear combinations of spatially separated field values are interesting measurement quantities in nanoscale NMR. Reference [226] describes an imaging protocol which combines many different Fourier spatial modes, and Ref. [219] similarly combines many signals to perform molecular microscopy. These measurements could be performed more accurately using entangled NV sensors. In addition, our entanglement scheme can perform simple subtraction of the signal between two qubits. This allows common mode noise subtraction between a sensor qubit and another qubit exposed only to environmental noise. In general, even if a full GHZ state of all sensors is not feasible, smaller clusters of entangled sensors can still enhance sensitivity.

Entanglement-enhanced imaging of objects larger than single molecules may also be a fruitful area of research. An experiment detecting the firing of a single animal neuron with accuracy near the standard quantum limit has already been

performed [227], making exploration of techniques surpassing the limit a natural next step. Similar experiments could demonstrate an enhancement due to distributed entanglement in the near future.

5.6 Extension to Arbitrary Functions

In this section, we generalize our previous work, which demonstrated a lower bound on the variance of an estimator of a linear combination of d parameters coupled to d qubits. We will generalize this approach to measuring an arbitrary real-valued, analytic function of d parameters and show that entanglement can reduce the variance of such an estimate by a factor of $\mathcal{O}(d)$. Finally, we present a protocol which achieves optimal variance asymptotically in the limit of long measurement time. In addition, when the parameters are coupled to d interferometers or to a combination of interferometers and qubits, we propose an analogous Heisenberg-scaling protocol to improve measurement noise. However, in this case, we lack a proof of optimality. We also can use the protocol presented in Ref. [228] to couple the parameters to continuous variables detected by homodyne measurements.

We will also examine the application of such a protocol to field interpolation. Suppose sensors are placed at d spatially separated locations, but we wish to know the field at a point with no sensor. We may pick a reasonable ansatz for the field with no more than d parameters, use our d measurements to fix the degrees of freedom of that ansatz, and compute the field at our desired point. Because the field of interest is a function of the field at d other locations, our protocol offers reduced noise over

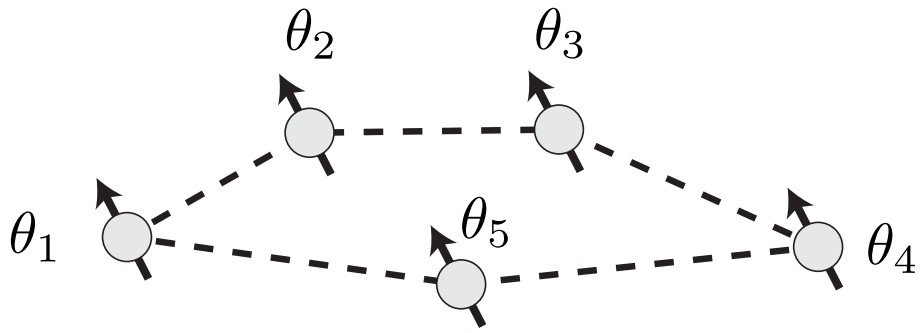

Figure 5.3: An illustration of a quantum sensor network of spatially separated nodes. At each node, there is an unknown parameter θ_i coupled to a qubit, which accumulates phase proportional to θ_i.

performing the same procedure without using entanglement.

5.6.1 Setup

In this section, the notation $\mathbb{E}_Y[X]$ means the expected value of X over all possible Y. If we merely write $\mathbb{E}[X]$, then we average over all parameters required to define X (e.g. if Y depended on Z, then $\mathbb{E}_Z[\mathbb{E}_Y[X]]$). We define the variance, $\mathrm{Var}_Y[X]$, similarly.

We consider a system with d sensor nodes, where node i consists of a single qubit coupled to a real parameter θ_i (see Fig. 5.3), and suppose that the state evolves under the Hamiltonian

$$\hat{H} = \hat{H}_c(t) + \frac{1}{2}\sum_{i=1}^{N} \theta_i \hat{\sigma}_i^z, \tag{5.57}$$

where $\hat{\sigma}_i^{x,y,z}$ are the Pauli operators acting on qubit i and $\hat{H}_c(t)$ is a time-dependent control Hamiltonian that we choose, which may include coupling to ancilla qubits. We want to measure an arbitrary real-valued, analytic function $f(\boldsymbol{\theta})$ of d unknown parameters $\boldsymbol{\theta} = \langle \theta_1, \ldots, \theta_d \rangle$ for time t_{total}. We would like to determine how well

the quantity $f(\boldsymbol{\theta})$ can be estimated, and find a protocol for doing so. To specify a protocol, we choose an input state, a control Hamiltonian $\hat{H}_c(t)$, and a final measurement.

For a general estimator, we use the mean squared error (MSE) M of our estimate $\tilde{f}$ from the true value $f(\boldsymbol{\theta})$ as a figure of merit. [1] Explicitly,

$$M = \mathbb{E}[(\tilde{f} - f(\boldsymbol{\theta}))^2] = \text{Var}\, \tilde{f} + (\mathbb{E}[\tilde{f}] - f(\boldsymbol{\theta}))^2. \tag{5.58}$$

Thus the MSE accounts for both the variance and the bias of the estimator $\tilde{f}$.

5.6.2 Lower Bound on Error

We now identify the minimum possible error of an estimator of $f(\boldsymbol{\theta})$ which measures for time t_{total}. For any estimator $\tilde{f}$, biased or otherwise, which uses samples from a probabilistic process (such as physical experiments) to estimate the value $f(\boldsymbol{\theta})$, the MSE is bounded by [197]

$$\mathbb{E}[(\tilde{f} - f(\boldsymbol{\theta}))^2] \geq \frac{1}{F} \geq \frac{1}{F_Q}, \tag{5.59}$$

where F is the Fisher information for the parameter f and F_Q is the quantum Fisher information evaluated over our input state, where $F_Q \geq F$ always [186]. The Fisher information matrix measures the sensitivity of the sampled probability distribution to changes in the parameters $\vec{\theta}$. While F tells us something about a particular experimental setup, F_Q is maximized over all possible experiments that could be performed on a state.

[1] Since the capital letter F is used for Fisher information, we break our usual notational conventions and use $\tilde{f}$ for the function estimator.

In order to evaluate the Fisher information for our function of interest f, we evaluate the generator $\hat{g} = \partial\hat{H}/\partial f$ as defined in Ref. [184]. By first writing the chain rule, we find that

$$\hat{g} = \frac{\partial \hat{H}}{\partial f} = \sum_{i=1}^{N} \frac{\partial \hat{H}}{\partial \theta_i}\frac{\partial \theta_i}{\partial f} = \frac{1}{2}\sum_{i=1}^{N} \hat{\sigma}_i^z \frac{\partial \theta_i}{\partial f}. \tag{5.60}$$

Note that F_Q can be upper-bounded by the seminorm of this generator, $F_Q \leq t^2\|\hat{g}\|_s^2$ [184]. However, to evaluate the seminorm, we will need to evaluate the partial derivative in Eq. (5.60). To do so we must specify a full basis of functions so that the partial derivative can be defined in terms of what variables are held constant during differentiation. We suppose that a set of functions $f_1, f_2, f_3 \ldots f_N$ are created, with the f of interest equal to f_1, defining an invertible coordinate transformation on $\mathbb{R}^N$. The seminorm is then:

$$\|\hat{g}\|_s = \sum_{i=1}^{N} \left|\frac{\partial \theta_i}{\partial f}\right| = \sum_{i=1}^{N} \left|J_{i1}^{-1}\right|. \tag{5.61}$$

Here, J_{ij}^{-1} is an element of the Jacobian matrix of the inverse transformation to that defined by the f functions. Depending on which functions are chosen, the value of $\|\hat{g}\|_s$ can vary, as can be seen in Sec. 5.3.2 for linear functions. We therefore wish to find the smallest possible $\|\hat{g}\|_s$, which will provide the tightest possible bound on F_Q. To do so, we note that J^{-1} and J must obey an inverse relationship, meaning that the following chain of inequalities holds,

$$1 = J_{1i}J_{i1}^{-1} \leq |J_{1i}|\left|J_{i1}^{-1}\right| \leq \max_j |J_{1j}| \sum_{i=1}^{N} \left|J_{i1}^{-1}\right|. \tag{5.62}$$

By using the definition of the Jacobian, we can rewrite this as a lower bound on the

value of $\|g\|_s$ in terms of partial derivatives of f:

$$\|\hat{g}\|_s = \sum_{i=1}^{N} \left| J_{i1}^{-1} \right| \geq \left(\max_j \left| \frac{\partial f}{\partial \theta_j} \right| \right)^{-1}. \tag{5.63}$$

All that remains is to note that if we label the θ_i that yields the maximum first derivative as θ_1, and then choose $f_i = \theta_i$ for $i > 1$, the lower bound in Eq. (5.63) is met, since $\partial\theta_i/\partial f_1$ must be evaluated holding the other f_j constant. Invoking the resulting bound on the quantum Fisher information, we find that Eq. (5.59) becomes

$$M = \mathbb{E}[(\tilde{f} - f(\boldsymbol{\theta}))^2] \geq \frac{1}{F_Q} \geq \max_j \frac{\left| \frac{\partial f}{\partial \theta_j} \right|}{t^2}. \tag{5.64}$$

Although the quantum Cramér-Rao Bound derived in Eq. (5.64) cannot always be saturated, it can when the generators $\partial\hat{H}/\partial\theta_i$ commute, as in Eq. (5.57) [186]. We will show later that the inequality in Eq. (5.64) can be saturated at asymptotic time t_{total}.

From this point forward, to simplify later calculation, we define $f_i(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_i}$. This definition also generalizes to multiple partial derivatives (i.e. $f_{ij} = \frac{\partial}{\partial \theta_j}\frac{\partial f}{\partial \theta_i}$).

Before moving on to the optimal protocol, we will consider a protocol which does not use entanglement and does not saturate Eq. (5.64) as a useful contrast to an entangled strategy. Suppose we estimate each parameter individually, without bias. Then the MSE $\mathbb{E}[(f(\tilde{\boldsymbol{\theta}}) - f(\boldsymbol{\theta}))^2]$ can be written as

$$M_{\text{unentangled}} = f_i(\boldsymbol{\theta})^2 \operatorname{Var} \tilde{\theta}_i. \tag{5.65}$$

A measurement of a single parameter can be made in time t with variance $\operatorname{Var} \tilde{\theta}_i = \frac{1}{t^2}$

[169]. Therefore, our entanglement-free figure of merit is

$$M_{\text{unentangled}} = \frac{\|\nabla f(\boldsymbol{\theta})\|^2}{t_{\text{total}}^2}, \tag{5.66}$$

where the $\|\cdot\|$ in Eq. (5.66) denotes the Euclidean norm. More generally, we use $\|\vec{v}\|_p$ to denote the p-norm of vector $\vec{v}$. Since Eq. (5.66) only saturates Eq. (5.64) in trivial cases where $\nabla f(\boldsymbol{\theta})$ is zero in all but one component, the unentangled protocol described is not optimal.

5.6.3 Two-step Protocol

We now present a protocol which asymptotically saturates Eq. (5.64). Our protocol consists of two steps. First, we make an unbiased estimate $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}$ for time t_1. Second, given our estimates $\boldsymbol{\Theta}$, we make an unbiased measurement Q of the quantity $q = \nabla f(\boldsymbol{\Theta}) \cdot (\boldsymbol{\theta} - \boldsymbol{\Theta})$ using the linear combination protocol in Ref. [26], which takes time t_2. Our final estimate is $\tilde{f} = f(\boldsymbol{\Theta}) + Q$.

It can be shown that our protocol is optimal (in terms of scaling with the total time $t_1 + t_2$) provided that the individual estimations of the parameters satisfy $\mathbb{E}[(\Theta_i - \theta_i)^4] = \mathcal{O}(t_1^{-4})$ and that t_1 and t_2 are chosen properly. To simplify our computations, we will make the more concrete assumption that our initial estimates $\boldsymbol{\Theta}$ are each normally distributed as $\mathcal{N}(\theta_i, \operatorname{Var}\Theta_i)$.

For the sake of concision, let $\boldsymbol{\Delta} = \boldsymbol{\Theta} - \boldsymbol{\theta}$ which satisfies $\mathbb{E}[\boldsymbol{\Delta}] = \mathbf{0}$. Furthermore, let T_k be $k!$ times the k-th term of the Taylor expansion of f (so $T_1 = f_i(\boldsymbol{\theta})\Delta_i$, $T_2 = f_{ij}(\boldsymbol{\theta})\Delta_i\Delta_j$, $T_3 = f_{ijk}(\boldsymbol{\theta})\Delta_i\Delta_j\Delta_k$, etc.). Thus, the Taylor expansion of $f(\boldsymbol{\Theta})$

would be

$$f(\boldsymbol{\Theta}) = f(\boldsymbol{\theta}) + T_1 + \frac{T_2}{2} + \frac{T_3}{6} + \dots. \tag{5.67}$$

We compute our figure of merit:

$$M = \mathbb{E}[(f(\boldsymbol{\Theta}) + Q - f(\boldsymbol{\theta}))^2] \tag{5.68}$$

$$= \underbrace{\mathbb{E}[(f(\boldsymbol{\Theta}) - f(\boldsymbol{\theta}))^2]}_{\text{term 1}} + \underbrace{\mathbb{E}[Q^2]}_{\text{term 2}} + 2\underbrace{\mathbb{E}[f(\boldsymbol{\Theta})Q]}_{\text{term 3}} - 2f(\boldsymbol{\theta})\mathbb{E}[Q] \tag{5.69}$$

$$= \left(\underbrace{\mathbb{E}[T_1^2] + \mathbb{E}[T_1T_2] + \frac{1}{3}\mathbb{E}[T_1T_3] + \frac{1}{4}\mathbb{E}[T_2^2] + \mathcal{O}(\boldsymbol{\Delta}^5)}_{\text{term 1}}\right) + \left(\underbrace{\mathbb{E}[\mathrm{Var}_Q\, Q] + \mathbb{E}[q^2]}_{\text{term 2}}\right)$$
$$+ 2\left(\underbrace{f(\boldsymbol{\theta})\mathbb{E}[q] + \mathbb{E}[T_1q] + \frac{1}{2}\mathbb{E}[T_2q] + \frac{1}{6}\mathbb{E}[T_3q] + \mathcal{O}(\boldsymbol{\Delta}^5)}_{\text{term 3}}\right) - 2f(\boldsymbol{\theta})\mathbb{E}[q] \tag{5.70}$$

$$= \mathbb{E}[\mathrm{Var}_Q\, Q] + \mathbb{E}[(q + T_1)^2] + \mathbb{E}[(q + T_1)T_2] + \frac{1}{3}\mathbb{E}[(q + T_1)T_3] + \frac{1}{4}\mathbb{E}[T_2^2] + \mathcal{O}(\boldsymbol{\Delta}^5). \tag{5.71}$$

The actual computation of the labeled terms is rather involved and space consuming, so it is presented later, in Sec. 5.6.3.1. Notice that we may simplify

$$q + T_1 = \sum_{i,j} \Delta_i (f_i(\boldsymbol{\theta}) - f_i(\boldsymbol{\Theta})) \tag{5.72}$$

$$= -\sum_{i,j} \Delta_i (f_{ij}(\boldsymbol{\theta})\Delta_j + \mathcal{O}(\boldsymbol{\Delta}^2)) \tag{5.73}$$

$$= -T_2 + \mathcal{O}(\boldsymbol{\Delta}^3), \tag{5.74}$$

so Eq. (5.71) evaluates to

$$M = \mathbb{E}[\mathrm{Var}_Q\, Q] + \mathbb{E}[T_2^2] - \mathbb{E}[T_2^2] - \frac{1}{3}\mathbb{E}[T_2T_3] + \frac{1}{4}\mathbb{E}[T_2^2] + \mathcal{O}(\boldsymbol{\Delta}^5) \tag{5.75}$$

$$= \mathbb{E}[\mathrm{Var}_Q\, Q] + \frac{1}{4}\mathbb{E}[T_2^2] + \mathcal{O}(\boldsymbol{\Delta}^5) \tag{5.76}$$

since $\mathbb{E}[T_2 T_3]$ is $\mathcal{O}(\boldsymbol{\Delta}^5)$. Now, this simplifies further as

$$M = \mathbb{E}[\mathrm{Var}_Q\, Q] + \frac{1}{4}\mathbb{E}[T_2^2] \tag{5.77}$$

$$= \mathbb{E}[\mathrm{Var}_Q\, Q] + \frac{1}{4}\sum_{i,j}\mathbb{E}[(f_{ij}(\boldsymbol{\theta})\Delta_i\Delta_j)^2] \tag{5.78}$$

$$= \mathbb{E}[\mathrm{Var}_Q\, Q] + \frac{1}{4}\mathbb{E}\left[4\sum_{i<j} f_{ij}(\boldsymbol{\theta})^2\Delta_i^2\Delta_j^2 + 2\sum_{i<j} f_{ii}(\boldsymbol{\theta})f_{jj}(\boldsymbol{\theta})\Delta_i^2\Delta_j^2 + \sum_i f_{ii}(\boldsymbol{\theta})^2\Delta_i^4\right] \tag{5.79}$$

since all terms with some Δ_i to a single power will factor out as $\mathbb{E}[\Delta_i] = 0$. We will assume that $\Delta_i \sim \mathcal{N}(0, \frac{1}{t_1^2})$ is normally distributed. This is not strictly necessary as long as the distribution of errors satisfies $\mathbb{E}[\Delta_i^4] \leq \mathcal{O}(t_1^{-4})$, a condition that is satisfied by phase estimation procedures like those in Ref. [229]. However, assuming normality allows the calculation to proceed easily, as we will be able to simplify $\mathbb{E}[\Delta_i^4] = 3\,\mathrm{Var}\,\Theta_i^2$. Thus, we arrive at

$$M = \mathbb{E}[\mathrm{Var}_Q\, Q] \tag{5.80}$$

$$+ \frac{1}{4}\left(4\sum_{i<j} f_{ij}(\boldsymbol{\theta})^2\,\mathrm{Var}\,\Theta_i\,\mathrm{Var}\,\Theta_j + 2\sum_{i<j} f_{ii}(\boldsymbol{\theta})f_{jj}(\boldsymbol{\theta})\,\mathrm{Var}\,\Theta_i\,\mathrm{Var}\,\Theta_j + \sum_i 3f_{ii}(\boldsymbol{\theta})^2\,\mathrm{Var}\,\Theta_i^2\right) \tag{5.81}$$

Then, the figure of merit for this protocol is

$$M = \mathbb{E}[(f(\boldsymbol{\Theta}) + Q - f(\boldsymbol{\theta}))^2] \tag{5.82}$$

$$= \mathbb{E}[\mathrm{Var}_Q\, Q] + \sum_{i.j}\frac{2f_{ij}(\boldsymbol{\theta}) + f_{ii}(\boldsymbol{\theta})f_{jj}(\boldsymbol{\theta})}{4}\,\mathrm{Var}\,\Theta_i\,\mathrm{Var}\,\Theta_j. \tag{5.83}$$

In Eq. (5.83), the first term is the error resulting from the second phase of the protocol, estimating the linear combination. The second term is a residual error remaining from the first phase of the protocol after it is corrected by the linear combination measurement.

5.6.3.1 Simplification of Labeled Terms

In this subsection, we present the simplification of the labeled terms from Eqs. (5.69-5.71) in full detail.

Term 2 is simplified by using the definition of $\mathrm{Var}_Q\, Q$. One needs to be careful as there are two layers of expected values - one for the values of $\boldsymbol{\Theta}$ and one for the estimator Q:

$$\underbrace{\mathbb{E}[Q^2]}_{\text{term 2}} = \mathbb{E}_{\boldsymbol{\Theta}}[\mathbb{E}_Q[Q^2]] \tag{5.84}$$

$$= \mathbb{E}_{\boldsymbol{\Theta}}[\mathrm{Var}_Q\, Q + \mathbb{E}_Q[Q]^2] \tag{5.85}$$

$$= \mathbb{E}_{\boldsymbol{\Theta}}[\mathrm{Var}_Q\, Q + q^2] \tag{5.86}$$

$$= \mathbb{E}[\mathrm{Var}_Q\, Q] + \mathbb{E}[q^2]. \tag{5.87}$$

Terms 1 and 3 are simplified by expanding the Taylor series for $f(\boldsymbol{\Theta})$ up to $\boldsymbol{\Delta}^4$ terms; note that $q = \mathcal{O}(\boldsymbol{\Delta})$, so we only need to expand the Taylor series up to $\mathcal{O}(\boldsymbol{\Delta}^3)$ terms:

$$\underbrace{\mathbb{E}\left[(f(\boldsymbol{\Theta}) - f(\boldsymbol{\theta}))^2\right]}_{\text{term 1}} = \mathbb{E}[f(\boldsymbol{\Theta})^2] - 2f(\boldsymbol{\theta})\mathbb{E}[f(\boldsymbol{\Theta})] + f(\boldsymbol{\theta})^2 \tag{5.88}$$

$$\begin{aligned} &= f(\boldsymbol{\theta})^2 + \mathbb{E}[T_1^2] + f(\boldsymbol{\theta})\mathbb{E}[T_2] + \mathbb{E}[T_1T_2] + \frac{1}{3}f(\boldsymbol{\theta})\mathbb{E}[T_3] \\ &\quad + \frac{1}{12}f(\boldsymbol{\theta})\mathbb{E}[T_4] + \frac{1}{3}\mathbb{E}[T_1T_3] + \frac{1}{4}\mathbb{E}[T_2^2] + \mathcal{O}(\boldsymbol{\Delta}^5) \\ &\quad - 2f(\boldsymbol{\theta})\left(f(\boldsymbol{\theta}) + \frac{1}{2}\mathbb{E}[T_2] + \frac{1}{6}\mathbb{E}[T_3] + \frac{1}{24}\mathbb{E}[T_4] + \mathcal{O}(\boldsymbol{\Delta}^5)\right) + f(\boldsymbol{\theta})^2 \end{aligned} \tag{5.89}$$

$$= \mathbb{E}[T_1^2] + \mathbb{E}[T_1T_2] + \frac{1}{3}\mathbb{E}[T_1T_3] + \frac{1}{4}\mathbb{E}[T_2^2] + \mathcal{O}(\boldsymbol{\Delta}^5). \tag{5.90}$$

$$\underbrace{\mathbb{E}[f(\boldsymbol{\Theta})Q]}_{\text{term 3}} = \mathbb{E}_{\boldsymbol{\Theta}}[\mathbb{E}_Q[f(\boldsymbol{\Theta})Q]] \tag{5.91}$$

$$= \mathbb{E}_{\boldsymbol{\Theta}}[f(\boldsymbol{\Theta})q] \tag{5.92}$$

$$= \mathbb{E}\left[\left(f(\boldsymbol{\theta}) + T_1 + \frac{T_2}{2} + \frac{T_3}{6} + \mathcal{O}(\boldsymbol{\Delta}^4)\right) q\right] \tag{5.93}$$

$$= f(\boldsymbol{\theta})\mathbb{E}[q] + \mathbb{E}[T_1 q] + \frac{\mathbb{E}[T_2 q]}{2} + \frac{\mathbb{E}[T_3 q]}{6} + \mathcal{O}(\boldsymbol{\Delta}^5). \tag{5.94}$$

For our particular Hamiltonian $\hat{H} = \frac{1}{2}\theta_i \hat{\sigma}_i^z$, as per Ref. [26], we know that the minimum variance of an unbiased estimator of some linear combination $\boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ given time t is

$$\operatorname{Var} \boldsymbol{\alpha} \cdot \boldsymbol{\theta} \geq \frac{\max_i \alpha_i^2}{t^2}, \tag{5.95}$$

which can be achieved with the entangled GHZ state $|\psi_{\text{spin}}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes d} + |1\rangle^{\otimes d})$. We can apply this linear combination protocol to the second phase of our protocol by setting $\boldsymbol{\alpha} = \nabla f(\boldsymbol{\Theta})$. For the individual estimators of the first phase, we use the fact that an individual estimation can be made in time t with variance $1/t^2$ [23]. Using these results, we simplify Eq. (5.83):

$$M = \mathbb{E}\left[\frac{\max_i f_i(\boldsymbol{\Theta})^2}{t_2^2}\right] + \sum_{i,j} \frac{\frac{2f_{ij}(\boldsymbol{\theta}) + f_{ii}(\boldsymbol{\theta}) f_{jj}(\boldsymbol{\theta})}{4}}{t_1^4} \tag{5.96}$$

$$= \frac{\mathbb{E}[\max_i f_i(\boldsymbol{\Theta})^2]}{t_2^2} + \frac{g_1(\boldsymbol{\theta})}{t_1^4}, \tag{5.97}$$

where we have absorbed the second derivatives into $g_1(\boldsymbol{\theta})$, which does not depend on time. Without loss of generality, we designate $f_1(\boldsymbol{\Theta})$ as the largest $f_i(\boldsymbol{\Theta})$. We then expand $\mathbb{E}[f_1(\boldsymbol{\Theta})^2]$ as

$$\sum_i f_1(\boldsymbol{\theta})^2 + \frac{f_1(\boldsymbol{\theta}) f_{1ii}(\boldsymbol{\theta})}{t_1^2} + \frac{f_{1i}(\boldsymbol{\theta})^2}{t_1^2} + \mathcal{O}((\boldsymbol{\Theta} - \boldsymbol{\theta})^3). \tag{5.98}$$

We may substitute Eq. (5.98) into Eq. (5.97) to obtain

$$M = \frac{g_2(\boldsymbol{\theta})}{t_2^2} + \frac{g_3(\boldsymbol{\theta})}{t_1^2 t_2^2} + \frac{g_1(\boldsymbol{\theta})}{t_1^4} + \mathcal{O}((\boldsymbol{\Theta} - \boldsymbol{\theta})^3), \tag{5.99}$$

where $g_2(\boldsymbol{\theta}) = f_1(\boldsymbol{\theta})^2$ and $g_3(\boldsymbol{\theta})$ have been introduced to absorb more time-independent factors.

Finally, we must specify how the total time t_{total} is to be allocated between t_1 and t_2. We want to choose the t_1, t_2, under the constraint that $t_1 + t_2 = t_{\text{total}}$, which minimize the MSE

$$M = \frac{g_2(\boldsymbol{\theta})}{t_2^2} + \frac{g_3(\boldsymbol{\theta})}{t_1^2 t_2^2} + \frac{g_1(\boldsymbol{\theta})}{t_1^4}. \tag{5.100}$$

Notice that the g_1, g_2, g_3 functions are only dependent on $\boldsymbol{\theta}$ and not t_1, so we may set the derivative of M with respect to t_1 equal to 0 and obtain

$$\frac{2g_2(\boldsymbol{\theta})}{t_2^3} + \frac{2g_3(\boldsymbol{\theta})}{t_2^3 t_1^2} = \frac{2g_3(\boldsymbol{\theta})}{t_2^2 t_1^3} + \frac{4g_1(\boldsymbol{\theta})}{t_1^5}. \tag{5.101}$$

Let $r = t_1/t_2$. Then we may rearrange to obtain

$$g_2(\boldsymbol{\theta}) t_1^2 = \frac{g_3(\boldsymbol{\theta})}{r} + \frac{2g_1(\boldsymbol{\theta})}{r^3} - g_3(\boldsymbol{\theta}). \tag{5.102}$$

Since $t_1 \gg 1$, then $r \ll 1$, so the r^{-3} term dominates the RHS. Thus, $g_2(\boldsymbol{\theta}) t_1^2 \approx \frac{2g_1(\boldsymbol{\theta})}{r^3}$, which implies

$$t_1 \approx \left(\frac{2g_1(\boldsymbol{\theta})}{g_2(\boldsymbol{\theta})}\right)^{1/5} t_2^{3/5} \approx \left(\frac{2g_1(\boldsymbol{\theta})}{g_2(\boldsymbol{\theta})}\right)^{1/5} t_{\text{total}}^{3/5}. \tag{5.103}$$

Therefore, the best possible allocation satisfies

$$t_1 = g(\boldsymbol{\theta}) t_{\text{total}}^{3/5}, \tag{5.104}$$

where g is a function which depends only on f and $\boldsymbol{\theta}$. In particular, $t_1 = \mathcal{O}(t_{\text{total}}^{3/5})$, so the fraction of time spent on t_1 vanishes as $t_{\text{total}} \to \infty$. Almost all of the time is

spent on t_2, the linear combination step of the two-step protocol. It can readily be shown that Eq. (5.99) is asymptotically dominated by the first term, which (since $t_2 \to t_{\text{total}}$) is equal to the right-hand-side of the bound in Eq. (5.64). In other words, this distribution of time asymptotically achieves the optimal MSE.

The two-step protocol exhibits Heisenberg scaling as defined for distributed sensing [26, 230, 231]. Comparing Eq. (5.66) to Eq. (5.64) shows an improvement of $\mathcal{O}(d)$, maximized when all components of $\nabla f(\boldsymbol{\theta})$ are approximately equal. Intuitively, the advantage is maximal when all parameters contribute, but minimal (i.e. no advantage) when only one parameter affects the function value. Similar behavior was noted in the linear combination case [26].

Note that when actually implementing the protocol, the optimal t_1 is unknown since the function g that determines it depends on the true parameters $\boldsymbol{\theta}$. However, we do not need to use the optimal t_1 to saturate the bound in Eq. (5.64). If t_1 is a function ct_{total}^p of the total time where $\frac{1}{2} < p < 1$ and some constant c, then the protocol will saturate Eq. (5.83). Although these different times do result in a higher MSE, the additional error is $\mathcal{O}\left(t_{\text{total}}^{-4}\right)$, which is insignificant asymptotically. The two-step protocol will therefore be asymptotically optimal for a wide range of time allocations. Recall that the MSE derived was

$$M = \frac{g_2(\boldsymbol{\theta})}{t_2^2} + \frac{g_3(\boldsymbol{\theta})}{t_1^2 t_2^2} + \frac{g_1(\boldsymbol{\theta})}{t_1^4}, \tag{5.105}$$

Suppose that $t_1 = ct_{\text{total}}^p$ for some $\frac{1}{2} < p < 1$ and some constant c. Since $p < 1$, we see that $\lim_{t\to\infty} \frac{t_2}{t_{\text{total}}} = 1$. Therefore, we may substitute our t_1 into the MSE

formula in Eq. (5.105) and simplify:

$$\lim_{t_{\text{total}}\to\infty} M = \lim_{t_{\text{total}}\to\infty} \frac{g_2(\boldsymbol{\theta})}{t_{\text{total}}^2} + \frac{g_3(\boldsymbol{\theta})}{c^2 t_{\text{total}}^{2+2p}} + \frac{g_1(\boldsymbol{\theta})}{c^4 t_{\text{total}}^{4p}}. \tag{5.106}$$

Since $p > \frac{1}{2}$, the t_{total}^2 term is dominant. Thus, as we defined $g_2 := f_1(\boldsymbol{\theta})^2 = \max_i f_i(\boldsymbol{\theta})^2$ under the assumption that $f_1(\boldsymbol{\theta})^2$ was maximal, our asymptotic error is

$$M = \frac{\max_i f_i(\boldsymbol{\theta})^2}{t_{\text{total}}^2}, \tag{5.107}$$

which saturates the bound of Eq. (5.83).

Other possible power-law scalings for t_1 fail. If $p \leq \frac{1}{2}$, the last term in Eq. (5.106) becomes significant asymptotically and prevents the protocol from saturating Eq. (5.83). If $p = 1$, then of course $c \leq 1$ or $t_1 > t_{\text{total}}$. In this case, we can no longer claim that t_2 approaches t_1. Even though the $\frac{g_2(\boldsymbol{\theta})}{t_2^2}$ would remain dominant, it would be scaled larger by $\frac{1}{(1-c)^2}$, which is always larger than 1. Hence, the protocol would no longer saturate the bound.

5.6.4 Function Measurement in Other Physical Settings

We now consider a different physical setting for function estimation. Rather than d qubits which accumulate phase for some time t, we instead pass n photons through d Mach-Zehnder interferometers and accumulate some fixed phase θ_i encoded into each interferometer (see Fig. 5.4). For single parameters, the use of entangled states to reduce noise in this setting has been explored in Refs. [173, 211, 232–234] with multiparameter cases explored in Refs. [230, 231]. In this setting, the relevant limitation is the total number of photons used in the measurement, rather

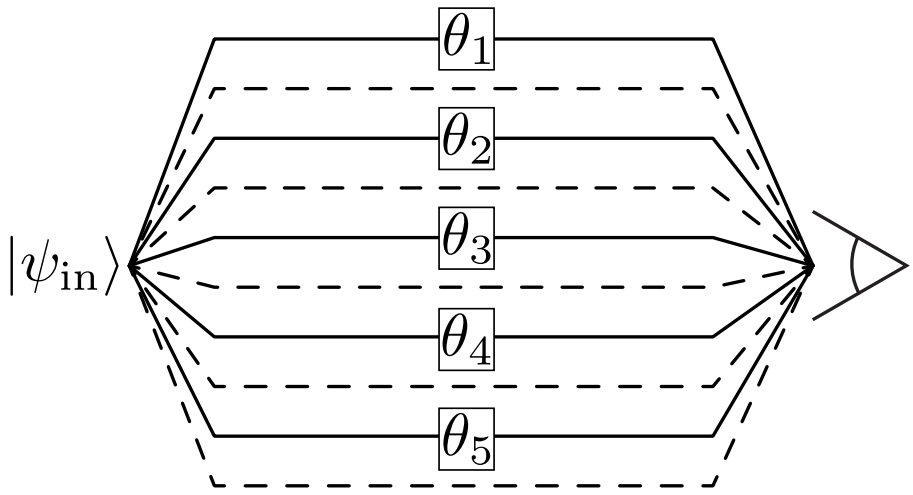

Figure 5.4: An example illustration of a quantum sensor network composed of separate interferometers. In each, one arm accumulates an unknown phase θ_i and the other arm is a reference port with no phase.

than time. This constraint is particularly relevant when analyzing a biological or chemical sample which is sensitive to light, making it desirable to reduce noise with as few photons as possible. Similar biologically motivated situations are presented in Refs. [227, 235, 236].

For photons, a two-step protocol with similar structure to the protocol for qubits yields reduced noise compared to any estimate of f derived entirely from local measurements. Suppose we allot N_1 photons for the first step (individual measurement) and N_2 photons for the second step (linear combination), for a total of $N_{\text{total}} = N_1 + N_2$ photons. We again begin from the general result of Eq. (5.83). However, the use of photons which can be apportioned between modes introduces new structure to the problem. We need to partition the N_1 photons into $N_1 = n_1 + \cdots + n_d$, putting n_i photons into the i-th interferometer, as some parameters may affect our final result more than others. Thus, in the second term of Eq. (5.83),

we replace $\operatorname{Var}\Theta_i$ with $\frac{1}{n_i^2}$ instead of $\frac{1}{t_1^2}$ [211].

The optimal variance when measuring the linear combination $\boldsymbol{\alpha}\cdot\boldsymbol{\theta}$ using N total photons is unknown. However, Ref. [231] conjectures the optimal variance to be

$$\operatorname{Var}\boldsymbol{\alpha}\cdot\boldsymbol{\theta} \geq \frac{\|\boldsymbol{\alpha}\|_1^2}{N^2}. \tag{5.108}$$

Furthermore, Ref. [231] provides a protocol achieving the bound in Eq. (5.108) using a proportionally weighted GHZ state: $|\psi_{\text{photon}}\rangle = \frac{1}{\sqrt{2}}(|n_1, 0, n_2, 0\ldots\rangle + |0, n_1, 0, n_2, \ldots\rangle)$, where $n_i = N_{\text{total}}\frac{\alpha_i}{\sum\alpha_j}$ and where, in reference to Fig. 5.4, the modes are listed from top to bottom. Note that this will only work for $\boldsymbol{\alpha}$ proportional to some rational vector as photons are discrete. Since Eq. (5.108) is saturable, we may simplify the first term of Eq. (5.83) to obtain

$$M = \frac{\mathbb{E}\left[\|\nabla f(\boldsymbol{\Theta})\|_1^2\right]}{N_2^2} + \sum_{i,j}\frac{2f_{ij}(\boldsymbol{\theta})^2 + \frac{f_{ii}(\boldsymbol{\theta})f_{jj}(\boldsymbol{\theta})}{4}}{n_i^2 n_j^2}. \tag{5.109}$$

For fixed f and θ, the $\frac{1}{n_i n_j}$ terms in Eq. (5.109) are minimized for the same ratio of $n_1 : n_2 : \cdots : n_d$ regardless of the value of the total number of photons used, N_1. Each term is proportional to N_1^{-4} multiplied by some function of $f, \boldsymbol{\theta}$, and d. Therefore, the structure of Eq. (5.109) becomes identical to the structure of Eq. (5.99), with N_1 and N_2 replacing t_1 and t_2. As a result, the optimal allocation of photons between N_1 and N_2 will yield $N_1 = \mathcal{O}(N_{\text{total}}^{3/5})$ and $N_2 = \mathcal{O}(N_{\text{total}})$, meaning that the N_2^{-2} term in Eq. (5.109) is dominant asymptotically. Therefore, for photons, we may asymptotically achieve

$$M = \frac{\|\nabla f(\boldsymbol{\Theta})\|_1^2}{N_{\text{total}}^2} + \mathcal{O}\left(\frac{1}{N_{\text{total}}^{12/5}}\right). \tag{5.110}$$

This strategy is optimal if the linear combination estimation strategy presented in Ref. [231] is optimal, as conjectured in that work.

Eq. (5.110) also exhibits Heisenberg scaling. Suppose we were to measure each parameter individually and then calculate the function. When measuring the parameters individually, we obtain the same error formula as Eq. (5.65), except now we set $\operatorname{Var}\Theta_i = \frac{1}{n_i^2}$ to get

$$M_{\text{unentangled}} = \sum_i \frac{f_i(\boldsymbol{\theta})^2}{n_i^2}. \tag{5.111}$$

The optimal distribution requires an n_i proportional to the weight $f_i(\boldsymbol{\theta})^{2/3}$, yielding an entanglement-free error of

$$M_{\text{unentangled}} = \frac{\|\nabla f(\boldsymbol{\theta})\|_{2/3}^2}{N_{\text{total}}^2}. \tag{5.112}$$

As with qubits, by comparing Eq. (5.110) with Eq. (5.112) in the case where all of the $f_i(\boldsymbol{\theta})$ are approximately equal, we find that the photonic two-step protocol yields a $\mathcal{O}(d)$ improvement in error over measuring each parameter individually. This improvement when all quantities are equally important can also be seen in Ref. [230] for the special case of f being a linear combination. As in the qubit case, the improvement in error is lessened when $\nabla f(\boldsymbol{\theta})$ is not approximately equal in all components.

In fact, this method can be extended still more generally. Rather than cases where the signal is imprinted on photons by a phase shift, we can consider the protocol developed in Ref. [228], which is capable of entanglement-enhanced distributed sensing of continuous variables by using homodyne measurements. Besides measuring parameters in different physical settings, we may also measure functions

of variables coupled to spins, phase-shifts of photons, continuous variables, and any combination of these.

In such a hybrid scenario, we can still make use of the two-step protocol. The first step, obtaining initial estimates for the individual parameters, proceeds equivalently, since the measurements of the spins and of the photons can be viewed as occurring in parallel. For the linear combination case, we can assume that the optimal spin and photon input states can be entangled as follows:

$$|\psi_{\text{spin}-\text{photon}}\rangle = \frac{1}{\sqrt{2}}\left(|n_1, 0, n_2, 0\ldots\rangle \otimes |1, 1, 1, \ldots\rangle + |0, n_1, 0, n_2, \ldots\rangle \otimes |0, 0, 0, \ldots\rangle\right). \tag{5.113}$$

Here, $n_i = N_{\text{total}}\frac{\alpha_i}{\sum \alpha_j}$, where the sum runs over only the j corresponding to photonic modes, denotes the number of photons which pass through the arms of the i-th interferometer. The state in Eq. (5.113) is designed in such a way that the two branches of the overall wavefunction accumulate relative to each other a phase equal to the total linear combination we are interested in. In order to extract this final phase, the state can be unitarily mapped onto a qubit, which contains all of the accumulated phase and is then measured.

One caveat is that the linear combination protocol will accumulate phase proportional to time for the qubits and phase proportional to the number of photons for interferometers. For instance, if θ_1 is coupled to a qubit (and therefore has units of frequency) and θ_2 is coupled to an interferometer (and is therefore unitless), then the two branches of our state accumulate a relative phase $\theta_1 t + \theta_2 n$. Therefore, one may have to adjust t or n in order to get the desired linear combination

5.6.5 Applications

As our protocol can measure any analytic function of $\boldsymbol{\theta}$, it is widely applicable. In fact, there is no requirement that different θ_i have the same physical origin. For instance, a θ_1 representing an electric field and θ_2 measuring a magnetic field could be used to measure the Poynting vector.

One potential application of function measurements is the interpolation of non-linear functions. Suppose that an ansatz with d tunable parameters is made for the strength of the field in a region. With readings from $\geq d$ different points, one could determine the parameters of the ansatz and therefore determine the value of the field at other points. Estimations of these ansatz parameters, which are functions of the measured fields, may potentially be improved using entangled states depending on the figure of merit [186, 237]. Note that this procedure can be carried out even if it is difficult to invert the ansatz in terms of the d measurements. Suppose that $\boldsymbol{\theta} = f(\boldsymbol{c}, \boldsymbol{x})$ and that $\boldsymbol{c} = f^{-1}(\boldsymbol{\theta}, \boldsymbol{x})$ exists, but has no closed-form solution which can be easily evaluated. First, we make measurements $\hat{\boldsymbol{\theta}}$. To create an initial estimate of the values $\boldsymbol{c}$, we use a numerical root-finder to find estimates $\boldsymbol{C}$. We can now implement the second step of our protocol by finding the first derivatives $\partial c_i/\partial\theta_j$ using the matrix identity $\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{c}} \cdot \frac{\partial\boldsymbol{c}}{\partial\boldsymbol{\theta}} = I$. Since f is known, $\partial\boldsymbol{\theta}/\partial\boldsymbol{c}$ can be inverted to yield the $\partial\boldsymbol{c}/\partial\boldsymbol{\theta}$ needed to estimate $\boldsymbol{Q} = \partial\boldsymbol{c}/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\Theta}} \cdot (\boldsymbol{\theta} - \boldsymbol{\Theta})$. Our final estimate is $\boldsymbol{C} + \boldsymbol{Q}$, which was obtained without having to compute f^{-1} in general.

Interpolation in this manner can proceed by two different schemes. We can either attempt to measure the ansatz parameters themselves, which allows compu-

tation of the field at all other points, or we can skip the final computation step by writing the field at a point of interest as a function of all the points that can be measured. This final function can then be directly measured using an entangled protocol, which will be more accurate. However, the first approach has the advantage that knowing the ansatz parameters allows estimation of all points in the space in question.

One particular interpolation of interest arises in ion trap quantum computing. In trapped ion chains, qubits are manipulated using Gaussian laser beams, and two primary sources of error are intensity and beam pointing fluctuations [74, 238, 239]. Our protocol offers better ways to characterize this noise. In order to detect the field error at a qubit's position without disturbing the qubit, we can perform interpolation by measuring the field's effect on other ions, possibly of a different atomic species, positioned nearby. Given the ansatz of the Gaussian beam profile, we are able to calculate the field at the qubit of interest and perhaps correct the error.

5.7 Outlook

We have presented measurement protocols for quantum networks which are useful for measuring arbitrary analytic functions of many parameters and developed a Heisenberg limit for this problem linear combinations. We have presented a Heisenberg-scaling measurement protocol using quantum sensor networks for measuring any multivariate, real-valued, analytic function, and this protocol is consistent with the Heisenberg limit when measuring functions with comparably-sized gradi-

ents in each component. Future work may include proving the optimality of the two-step protocol when constrained by the number of photons, which would require extending the results of Ref. [231].

Our protocol can be considered a generalization of entanglement-enhanced Ramsey spectroscopy, as in Ref. [23], to the measurement of spatially varying quantities. In the future, we hope to search for further protocols and to remove the requirements of small signal or evolution time where we have imposed them. We identified magnetometry in general and nanoscale NMR in particular as candidate applications of our protocol, but we wish to stress our protocol's significantly broader scope. In particular, we expect that our protocol will be useful for measuring spatially varying quantities in contexts such as gravimetry [240,241], spectroscopy [173], and rotation sensing [242–244]. Note there is also no requirement that the parameters measured in a linear combination be of the same physical source. For instance, a sensor network could measure a linear combination of both electric and magnetic fields. In general, our protocol can be applied in any setting where Ramsey spectroscopy can be applied if the quantity of interest is nonlocal. In addition, recent work [245] indicates that spatial correlations in measurements may be a useful tool for noise-filtering and error correction in quantum sensors.

We specifically identified field interpolation as a promising application of our arbitrary-function protocol, but we stress that our protocol can assist in the measurement of any analytic function. More work remains to determine when it is optimal to measure the coefficients of interpolation and when it is optimal to directly measure the final function. We are also interested in fleshing out possible

intersections between quantum function estimation and machine learning. Supervised machine learning is a type of interpolation: estimating functional outputs for unknown inputs by extracting information from known input-output pairs [246]. It is possible our protocol could be used to improve the accuracy of training a machine learning model if the necessary quantity for training was a function of physical measurements. Additionally, the final output of many machine learning algorithms, such as neural networks, is a non-linear but infinitely differentiable function of the inputs [247]. Our work could aid in computing this complicated function for new input when making predictions.

Bibliography

[1] Robin Materese. The Second Quantum Revolution. https://www.nist.gov/topics/physics/introduction-new-quantum-revolution/second-quantum-revolution, March 2018.

[2] Sylvia Berryman. Democritus. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2016.

[3] D. ter Haar. *The Old Quantum Theory*. The Commonwealth and International Library. Selected Readings in Physics. Pergamon Press, Oxford, New York, 1st ed. edition, 1967.

[4] A. Einstein. über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Ann. Phys.*, 322(6):132–148, 1905.

[5] Nielsls Bohr. I. On the constitution of atoms and molecules. *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, 26(151):1–25, July 1913.

[6] E. Schrödinger. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.*, 28(6):1049–1070, December 1926.

[7] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.*, 47(10):777–780, May 1935.

[8] Lynden K. Shalm, Evan Meyer-Scott, Bradley G. Christensen, et al. Strong Loophole-Free Test of Local Realism. *Phys. Rev. Lett.*, 115(25):250402–250402, December 2015.

[9] Marissa Giustina, Marijn A. M. Versteegh, Sören Wengerowsky, et al. Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. *Phys. Rev. Lett.*, 115(25):250401–250401, December 2015.

[10] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195–200, November 1964.

[11] Christopher A. Fuchs. QBism, the Perimeter of Quantum Bayesianism. *ArXiv10035209 Quant-Ph*, March 2010.

[12] Hugh Everett. "Relative State" Formulation of Quantum Mechanics. *Rev. Mod. Phys.*, 29(3):454–462, July 1957.

[13] David Bohm. A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I. *Phys. Rev.*, 85(2):166–179, January 1952.

[14] Richard Jozsa. An introduction to measurement based quantum computation. arXiv:0508124, 2005.

[15] Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Commun. Math. Phys.*, 28(3):251–257, September 1972.

[16] S. Bravyi, M. B. Hastings, and F. Verstraete. Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order. *Phys. Rev. Lett.*, 97(5):050401–050401, July 2006.

[17] Matthew B. Hastings and Tohru Koma. Spectral Gap and Exponential Decay of Correlations. *Commun. Math. Phys.*, 265(3):781–804, April 2006.

[18] Zhe-Xuan Gong, Michael Foss-Feig, Spyridon Michalakis, and Alexey V. Gorshkov. Persistence of Locality in Systems with Power-Law Interactions. *Phys. Rev. Lett.*, 113(3):030602–030602, July 2014.

[19] Michael Foss-Feig, Zhe-Xuan Gong, Charles W. Clark, and Alexey V. Gorshkov. Nearly Linear Light Cones in Long-Range Interacting Quantum Systems. *Phys. Rev. Lett.*, 114(15):157201–157201, April 2015.

[20] Guifré Vidal. Efficient Classical Simulation of Slightly Entangled Quantum Computations. *Phys. Rev. Lett.*, 91(14):147902, October 2003.

[21] C. Monroe, R. Raussendorf, A. Ruthven, et al. Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects. *Phys. Rev. - At. Mol. Opt. Phys.*, 89(2):1–16, 2014.

[22] A. Pirker, J. Wallnöfer, and W. Dür. Modular architectures for quantum networks. *New J. Phys.*, 20(5):053054, May 2018.

[23] J. J . Bollinger, Wayne M. Itano, D. J. Wineland, and D. J. Heinzen. Optimal frequency measurements with maximally correlated states. *Phys. Rev. A*, 54(6):R4649–R4652, December 1996.

[24] Zachary Eldredge, Zhe-Xuan Gong, Jeremy T. Young, et al. Fast Quantum State Transfer and Entanglement Renormalization Using Long-Range Interactions. *Phys. Rev. Lett.*, 119(17):170503–170503, October 2017.

[25] Aniruddha Bapat, Zachary Eldredge, James R. Garrison, et al. Unitary entanglement construction in hierarchical networks. *Phys. Rev. A*, 98(6):062328, December 2018.

[26] Z. Eldredge, M. Foss-Feig, J.A. Gross, S.L. Rolston, and A.V. Gorshkov. Optimal and secure measurement protocols for quantum sensor networks. *Phys. Rev. A*, 97(4), 2018.

[27] Kevin Qian, Zachary Eldredge, Wenchao Ge, et al. Heisenberg-Scaling Measurement Protocol for Analytic Functions with Quantum Sensor Networks. arXiv:1901.09042, January 2019.

[28] Bruno Nachtergaele and Robert Sims. Lieb-Robinson Bounds in Quantum Many-Body Physics. *Contemp. Math.*, 529:141–176, April 2010.

[29] David Poulin. Lieb-Robinson Bound and Locality for General Markovian Quantum Dynamics. *Phys. Rev. Lett.*, 104(19):190401–190401, May 2010.

[30] Bruno Nachtergaele, Anna Vershynina, and Valentin Zagrebnov. Lieb-Robinson bounds and existence of the thermodynamic limit for a class of irreversible quantum dynamics. *Contemp. Math.*, 552:161–175, 2011.

[31] Bo Yan, Steven A. Moses, Bryce Gadway, et al. Observation of dipolar spin-exchange interactions with lattice-confined polar molecules. *Nature*, 501(7468):521–525, September 2013.

[32] Alexey V. Gorshkov, Salvatore R. Manmana, Gang Chen, et al. Quantum magnetism with polar alkali-metal dimers. *Phys. Rev. A*, 84(3):033619, September 2011.

[33] M. A. Baranov, M. Dalmonte, G. Pupillo, and P. Zoller. Condensed Matter Theory of Dipolar Quantum Gases. *Chem. Rev.*, 112(9):5012–5061, September 2012.

[34] M. Saffman, T. G. Walker, and K. Mølmer. Quantum information with Rydberg atoms. *Rev. Mod. Phys.*, 82(3):2313–2363, August 2010.

[35] Antoine Browaeys, Daniel Barredo, and Thierry Lahaye. Experimental investigations of dipole–dipole interactions between a few Rydberg atoms. *J. Phys. B At. Mol. Opt. Phys.*, 49(15):152001–152001, August 2016.

[36] R. Islam, C. Senko, W. C. Campbell, et al. Emergence and Frustration of Magnetism with Variable-Range Interactions in a Quantum Simulator. *Science*, 340(6132):583–587, May 2013.

[37] D. Porras and J. I. Cirac. Effective Quantum Spin Systems with Trapped Ions. *Phys. Rev. Lett.*, 92(20):207901–207901, May 2004.

[38] Jens Eisert, Mauritz van den Worm, Salvatore R. Manmana, and Michael Kastner. Breakdown of Quasilocality in Long-Range Quantum Lattice Models. *Phys. Rev. Lett.*, 111(26):260401–260401, December 2013.

[39] P. Hauke and L. Tagliacozzo. Spread of Correlations in Long-Range Interacting Quantum Systems. *Phys. Rev. Lett.*, 111(20):207202–207202, November 2013.

[40] Philip Richerme, Zhe-Xuan Gong, Aaron Lee, et al. Non-local propagation of correlations in quantum systems with long-range interactions. *Nature*, 511(7508):198–201, July 2014.

[41] David Métivier, Romain Bachelard, and Michael Kastner. Spreading of Perturbations in Long-Range Interacting Classical Lattice Models. *Phys. Rev. Lett.*, 112(21):210601–210601, May 2014.

[42] Sven Bachmann, Spyridon Michalakis, Bruno Nachtergaele, and Robert Sims. Automorphic Equivalence within Gapped Phases of Quantum Lattice Systems. *Commun. Math. Phys.*, 309(3):835–871, February 2012.

[43] Daniel Gottesman, Thomas Jennewein, and Sarah Croke. Longer-Baseline Telescopes Using Quantum Repeaters. *Phys. Rev. Lett.*, 109(7):070503–070503, August 2012.

[44] P. Kómár, E. M. Kessler, M. Bishof, et al. A quantum network of clocks. *Nat. Phys.*, 10(8):582–587, June 2014.

[45] E M Kessler, P Kómár, M Bishof, et al. Heisenberg-limited atom clocks based on entangled qubits. *Phys. Rev. Lett.*, 112(19):190403–190403, May 2014.

[46] Z. Eldredge, P. Solano, D. Chang, and A.V. Gorshkov. Self-organization of atoms coupled to a chiral reservoir. *Phys. Rev. - At. Mol. Opt. Phys.*, 94(5), 2016.

[47] J. I. Cirac, P. Zoller, H. J. Kimble, and H. Mabuchi. Quantum State Transfer and Entanglement Distribution among Distant Nodes in a Quantum Network. *Phys. Rev. Lett.*, 78(16):3221–3224, April 1997.

[48] Matthias Christandl, Nilanjana Datta, Artur Ekert, and Andrew J. Landahl. Perfect State Transfer in Quantum Spin Networks. *Phys. Rev. Lett.*, 92(18):187902–187902, May 2004.

[49] Sougato Bose. Quantum Communication through an Unmodulated Spin Chain. *Phys. Rev. Lett.*, 91(20):207901–207901, November 2003.

[50] Sougato Bose. Quantum communication through spin chain dynamics: An introductory overview. *Contemp. Phys.*, 48(1):13–30, January 2007.

[51] G. D. de Moraes Neto, F. M. Andrade, V. Montenegro, and S. Bose. Quantum state transfer in optomechanical arrays. *Phys. Rev. A*, 93(6):062339–062339, June 2016.

[52] Umer Farooq, Abolfazl Bayat, Stefano Mancini, and Sougato Bose. Adiabatic many-body state preparation and information transfer in quantum dot arrays. *Phys. Rev. B*, 91(13):134303–134303, April 2015.

[53] Daniel Burgarth and Sougato Bose. Conclusive and arbitrarily perfect quantum-state transfer using parallel spin-chain channels. *Phys. Rev. A*, 71(5):052315–052315, May 2005.

[54] Giulia Gualdi, Vojtech Kostak, Irene Marzoli, and Paolo Tombesi. Perfect state transfer in long-range interacting spin chains. *Phys. Rev. A*, 78(2):022325–022325, August 2008.

[55] G Vidal. Entanglement renormalization. *Phys. Rev. Lett.*, 99(22):220405–220405, November 2007.

[56] G. Vidal. Class of Quantum Many-Body States That Can Be Efficiently Simulated. *Phys. Rev. Lett.*, 101(11):110501–110501, 2008.

[57] V. Giovannetti, S. Montangero, and Rosario Fazio. Quantum Multiscale Entanglement Renormalization Ansatz Channels. *Phys. Rev. Lett.*, 101(18):180503–180503, October 2008.

[58] G. Evenbly and G. Vidal. Algorithms for entanglement renormalization. *Phys. Rev. B*, 79(14):144108–144108, April 2009.

[59] Robert N. C. Pfeifer, Glen Evenbly, and Guifré Vidal. Entanglement renormalization, scale invariance, and quantum criticality. *Phys. Rev. A*, 79(4):040301–040301, April 2009.

[60] Brian Swingle. Interplay between short- and long-range entanglement in symmetry-protected phases. *Phys. Rev. B*, 90(3):035451–035451, July 2014.

[61] Miguel Aguado and Guifré Vidal. Entanglement renormalization and topological order. *Phys. Rev. Lett.*, 100(7):070404–070404, February 2008.

[62] Jeongwan Haah. Bifurcation in entanglement renormalization group flow of a gapped spin model. *Phys. Rev. B*, 89(7):075119–075119, February 2014.

[63] Xueda Wen, Gil Young Cho, Pedro L. S. Lopes, et al. Holographic Entanglement Renormalization of Topological Insulators. May 2016.

[64] Brian Swingle and John McGreevy. Renormalization group constructions of topological quantum liquids and beyond. *Phys. Rev. B*, 93(4):045127–045127, January 2016.

[65] Alessandro Campa, Thierry Dauxois, and Stefano Ruffo. Statistical mechanics and dynamics of solvable models with long-range interactions. *Phys. Rep.*, 480(3):57–159, 2009.

[66] Sergio A. Cannas and Francisco A. Tamarit. Long-range interactions and nonextensivity in ferromagnetic spin models. *Phys. Rev. B*, 54(18):R12661–R12664, November 1996.

[67] G. Evenbly and G. Vidal. Class of Highly Entangled Many-Body States that can be Efficiently Simulated. *Phys. Rev. Lett.*, 112(24):240502–240502, June 2014.

[68] D. Jaksch, J. I. Cirac, P. Zoller, et al. Fast Quantum Gates for Neutral Atoms. *Phys. Rev. Lett.*, 85(10):2208–2211, September 2000.

[69] M Saffman. Quantum computing with atomic qubits and Rydberg interactions: Progress and challenges. *J. Phys. B At. Mol. Opt. Phys.*, 49(20):202001–202001, October 2016.

[70] K. M. Maller, M. T. Lichtman, T. Xia, et al. Rydberg-blockade controlled-not gate and entanglement in a two-dimensional array of neutral-atom qubits. *Phys. Rev. A*, 92(2):022336–022336, August 2015.

[71] T. Wilk, A. Gaëtan, C. Evellin, et al. Entanglement of Two Individual Neutral Atoms Using Rydberg Blockade. *Phys. Rev. Lett.*, 104(1):010502–010502, January 2010.

[72] L. Isenhower, E. Urban, X. L. Zhang, et al. Demonstration of a Neutral Atom Controlled-NOT Quantum Gate. *Phys. Rev. Lett.*, 104(1):010503–010503, January 2010.

[73] D. DeMille. Quantum Computation with Trapped Polar Molecules. *Phys. Rev. Lett.*, 88(6):067901, January 2002.

[74] J. I. Cirac and P. Zoller. Quantum Computations with Cold Trapped Ions. *Phys. Rev. Lett.*, 74(20):4091–4094, May 1995.

[75] Yang Wang, Xianli Zhang, Theodore A. Corcovilos, Aishwarya Kumar, and David S. Weiss. Coherent Addressing of Individual Neutral Atoms in a 3D Optical Lattice. *Phys. Rev. Lett.*, 115(4):043003, July 2015.

[76] B. Huber, T. Baluktsian, M. Schlagmüller, et al. GHz Rabi Flopping to Rydberg States in Hot Atomic Vapor Cells. *Phys. Rev. Lett.*, 107(24):243001–243001, December 2011.

[77] J. True Merrill, S. Charles Doret, Grahame Vittorini, J. P. Addison, and Kenneth R. Brown. Transformed composite sequences for improved qubit addressing. *Phys. Rev. A*, 90(4):040301–040301, October 2014.

[78] Rodney Van Meter and Kohei M. Itoh. Fast quantum modular exponentiation. *Phys. Rev. A*, 71(5):052320, May 2005.

[79] Rodney Van Meter, W. J. Munro, Kae Nemoto, and Kohei M. Itoh. Arithmetic on a Distributed-memory Quantum Multicomputer. *J Emerg Technol Comput Syst*, 3(4):2:1–2:23, January 2008.

[80] Muhammad Ahsan and Jungsang Kim. Optimization of Quantum Computer Architecture using a Resource-Performance Simulator. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015*, pages 1108–1113. IEEE Conference Publications, 2015.

[81] T.S. Metodi, D.D. Thaker, and A.W. Cross. A Quantum Logic Array Microarchitecture: Scalable Quantum Data Movement and Computation. In *38th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'05)*, pages 305–318. IEEE, 2005.

[82] L.-M. Duan and C. Monroe. Colloquium: Quantum networks with trapped ions. *Rev. Mod. Phys.*, 82(2):1209–1224, April 2010.

[83] C Monroe and J Kim. Scaling the ion trap quantum processor. *Science*, 339(6124):1164–9, March 2013.

[84] M H Devoret and R J Schoelkopf. Superconducting circuits for quantum information: An outlook. *Science*, 339(6124):1169–74, March 2013.

[85] Teresa Brecht, Wolfgang Pfaff, Chen Wang, et al. Multilayer microwave integrated quantum circuits for scalable quantum computing. *Npj Quantum Inf.*, 2(1):16002–16002, November 2016.

[86] P. Kurpiers, P. Magnard, T. Walter, et al. Deterministic quantum state transfer and remote entanglement using microwave photons. *Nature*, 558(7709):264, June 2018.

[87] Aric A Hagberg, Daniel A Schult, and Pieter J Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, page 5, 2008.

[88] L. Barrière, F. Comellas, C. Dalfó, and M.A. Fiol. The hierarchical product of graphs. *Discrete Appl. Math.*, 157(1):36–48, January 2009.

[89] C. D. Godsil and B. D. McKay. A new graph product and its spectrum. *Bull. Aust. Math. Soc.*, 18(1):21–28, February 1978.

[90] Gregory Bentsen, Yingfei Gu, and Andrew Lucas. Fast scrambling on sparse graphs. arXiv:1805.08215, May 2018.

[91] Martí Cuquet and John Calsamiglia. Growth of graph states in quantum networks. *Phys. Rev. A*, 86(4):042304–042304, October 2012.

[92] Antonio Acín, J. Ignacio Cirac, and Maciej Lewenstein. Entanglement percolation in quantum networks. *Nat. Phys.*, 3(4):256–259, April 2007.

[93] S. Perseguers, G J Lapeyre, D. Cavalcanti, M. Lewenstein, and A. Acín. Distribution of entanglement in large-scale quantum networks. *Rep. Prog. Phys.*, 76(9):096001–096001, 2013.

[94] K. Kieling, T. Rudolph, and J. Eisert. Percolation, Renormalization, and Quantum Computing with Nondeterministic Gates. *Phys. Rev. Lett.*, 99(13):130501–130501, September 2007.

[95] Martí Cuquet and John Calsamiglia. Entanglement Percolation in Quantum Complex Networks. *Phys. Rev. Lett.*, 103(24):240503–240503, December 2009.

[96] Martí Cuquet and John Calsamiglia. Limited-path-length entanglement percolation in quantum complex networks. *Phys. Rev. A*, 83(3):032319–032319, March 2011.

[97] Yang Wang, D. Chakrabarti, Chenxi Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34. IEEE Comput. Soc.

[98] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small world' networks. *Nature*, 393(6684):440–442, June 1998.

[99] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439), 1999.

[100] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, January 2002.

[101] S. Perseguers, M. Lewenstein, A. Acín, and J. I. Cirac. Quantum random networks. *Nat. Phys.*, 6(7):539–543, July 2010.

[102] C. Di Franco and D. Ballester. Optimal path for a quantum teleportation protocol in entangled networks. *Phys. Rev. A*, 85(1):010303–010303, January 2012.

[103] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller. Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication. *Phys. Rev. Lett.*, 81(26):5932–5935, December 1998.

[104] Frank Harary. The number of linear, directed, rooted, and connected graphs. *Trans Amer Math Soc*, 78:445–463, 1955.

[105] Mikhail Gromov. *Metric Structures for Riemannian and Non-Riemannian Spaces*. Modern Birkhäuser Classics. Birkhäuser Basel, 2007.

[106] Wei Chen, Wenjie Fang, Guangda Hu, and Michael W. Mahoney. On the Hyperbolicity of Small-World and Treelike Random Graphs. *Internet Math*, 9(4):434–491, 2013.

[107] Poonsuk Lohsoonthorn. *Hyperbolic Geometry of Networks*. PhD Thesis, University of Southern California, Los Angeles, CA, USA, 2003.

[108] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic Geometry of Complex Networks. *Phys Rev E*, 82(3):036106, 2010.

[109] Fabien de Montgolfier, Mauricio Soto, and Laurent Viennot. Treewidth and Hyperbolicity of the Internet. In *Proceedings of the 2011 IEEE 10th International Symposium on Network Computing and Applications*, NCA '11, pages 25–32, Washington, DC, USA, 2011. IEEE Computer Society.

[110] Marián Boguñá, Fragkiskos Papadopoulos, and Dmitri Krioukov. Sustaining the Internet with Hyperbolic Mapping. *Nat Commun*, 1:62, 2010.

[111] Alicia J. Kollár, Mattias Fitzpatrick, and Andrew A. Houck. Hyperbolic Lattices in Circuit Quantum Electrodynamics. arXiv: 1802.09549, 2018.

[112] Sang-il Oum. Rank-Width and Vertex-Minors. *J Comb Theory Ser B*, 95(1):79–100, 2005.

[113] B. Courcelle, J. A. Makowsky, and U. Rotics. Linear Time Solvable Optimization Problems on Graphs of Bounded Clique Width. pages 1–16. Springer, Berlin, Heidelberg, 1998.

[114] Charles E. Leiserson. Fat-trees: Universal networks for hardware-efficient supercomputing. *IEEE Trans. Comput.*, C-34(10):892–901, October 1985.

[115] Michael William Newman. *The Laplacian Spectrum of Graphs*. Master's Thesis, University of Manitoba, Winnipeg, Canada, 2000.

[116] Bojan Mohar. The Laplacian Spectrum of Graphs. In *Graph Theory, Combinatorics, and Applications*, pages 871–898. Wiley, 1991.

[117] Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[118] Donny Cheung, Dmitri Maslov, and Simone Severini. Translation Techniques Between Quantum Circuit Architectures. In *Proceedings of the Workshop on Quantum Information*, 2007.

[119] Norbert M. Linke, Dmitri Maslov, Martin Roetteler, et al. Experimental comparison of two quantum computing architectures. *PNAS*, 114(13):3305–3310, March 2017.

[120] Robert S. Pindyck and Daniel L. Rubinfeld. *Microeconomics*. Pearson, 2013.

[121] Zoltán Füredi. Graphs of Diameter 3 with the Minimum Number of Edges. *Graphs Comb*, 6(4):333–337, 1990.

[122] A. J. Hoffman and R. R. Singleton. On Moore Graphs with Diameters 2 and 3. *IBM J Res Dev*, 4(5):497–504, 1960.

[123] Mirka Miller and Jozef Sirán. Moore Graphs and beyond: A Survey of the Degree/Diameter Problem. *Electron J Comb*, 1000:DS14–Dec, 2005.

[124] Guillermo Pineda-Villavicencio and David R. Wood. The Degree-Diameter Problem for Sparse Graph Classes. *Electron J Comb*, 22(2):2.46, 2015.

[125] Zachary Eldredge et al. Unitary-modular. GitHub repository.

[126] Dmitri Maslov, Sean M. Falconer, and Michele Mosca. Quantum Circuit Placement. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, 27(4):752–763, April 2008.

[127] Massoud Pedram and Alireza Shafaei. Layout Optimization for Quantum Circuits with Linear Nearest Neighbor Architectures. *IEEE Circuits Syst. Mag.*, 16(2):62–74, 2016.

[128] Thomas Häner, Damian S. Steiger, Krysta Svore, and Matthias Troyer. A software methodology for compiling quantum programs. *Quantum Sci. Technol.*, 3(2):020501, 2018.

[129] Damian S. Steiger, Thomas Häner, and Matthias Troyer. ProjectQ: An Open Source Software Framework for Quantum Computing. *Quantum*, 2:49, January 2018.

[130] Konstantin Andreev and Harald Racke. Balanced Graph Partitioning. *Theory Comput Syst*, 39(6):929–939, 2006.

[131] G. Karypis and V. Kumar. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs. *SIAM J Sci Comput*, 20(1):359–392, 1998.

[132] R. Kleinberg. Geographic Routing Using Hyperbolic Space. In *IEEE INFOCOM 2007*, pages 1902–1909. IEEE, 2007.

[133] E. Jonckheere and P. Lohsoonthorn. Geometry of Network Security. In *Proceedings of the 2004 American Control Conference*, volume 2, pages 976–981 vol.2, 2004.

[134] Victor Chepoi, Feodor F. Dragan, Bertrand Estellon, et al. Additive Spanners and Distance and Routing Labeling Schemes for Hyperbolic Graphs. *Algorithmica*, 62(3-4):713–732, 2012.

[135] A. G. Fowler, S. J. Devitt, and L. C. L. Hollenberg. Implementation of Shor's Algorithm on a Linear Nearest Neighbour Qubit Array. *Quantum Inf Comput*, 4(4):237–251, 2004.

[136] Peter Høyer and Robert Špalek. Quantum Fan-out is Powerful. *Theory Comput.*, 1(1):81–103, August 2005.

[137] Dmitri Maslov. Basic circuit compilation techniques for an ion-trap quantum machine. *New J. Phys.*, 19(2):023035–023035, February 2017.

[138] Michael Epping, Hermann Kampermann, and Brus Dagmar. Large-scale quantum networks based on graphs. *New J. Phys.*, 18(5):053036, May 2016.

[139] Benoıt̂ Descamps. Asymptotically decreasing Lieb-Robinson velocity for a class of dissipative quantum dynamics. *J. Math. Phys.*, 54(9):092202–092202, May 2013.

[140] David P. DiVincenzo and IBM. The Physical Implementation of Quantum Computation. *Fortschritte Phys.*, 48(9-11):771–783, September 2000.

[141] Joseph Kerckhoff, Hendra I. Nurdin, Dmitri S. Pavlichin, and Hideo Mabuchi. Designing Quantum Memories with Embedded Control: Photonic Circuits for Autonomous Quantum Error Correction. *Phys. Rev. Lett.*, 105(4):040502, July 2010.

[142] Eliot Kapit. Hardware-Efficient and Fully Autonomous Quantum Error Correction in Superconducting Circuits. *Phys. Rev. Lett.*, 116(15):150501, April 2016.

[143] Karel Van Acoleyen, Michaël Mariën, and Frank Verstraete. Entanglement Rates and Area Laws. *Phys. Rev. Lett.*, 111(17):170501–170501, October 2013.

[144] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin. On the capacities of bipartite Hamiltonians and unitary gates. *IEEE Trans. Inf. Theory*, 49(8):1895–1911, August 2003.

[145] Bojan Mohar. Isoperimetric Numbers of Graphs. *J. Comb. Theory*, 47:274–291, 1989.

[146] K. S. Chou, J. Z. Blumoff, C. S. Wang, et al. Deterministic teleportation of a quantum gate between two logical qubits. January 2018.

[147] Daniel Gottesman and Isaac L. Chuang. Quantum Teleportation is a Universal Computational Primitive. *Nature*, 402(6760):390–393, November 1999.

[148] Liang Jiang, Jacob M. Taylor, Anders S. Sørensen, and Mikhail D. Lukin. Distributed quantum computation based on small quantum registers. *Phys. Rev. A*, 76(6), December 2007.

[149] Ali Javadi-Abhari, Pranav Gokhale, Adam Holmes, et al. Optimized Surface Code Communication in Superconducting Quantum Computers. In *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO-50 '17, pages 692–705, New York, NY, USA, 2017. ACM.

[150] Michael J. Bremner, Christopher M. Dawson, Jennifer L. Dodd, et al. Practical Scheme for Quantum Computation with Any Two-Qubit Entangling Gate. *Phys. Rev. Lett.*, 89(24):247902–247902, November 2002.

[151] Zhe-Xuan Gong, Michael Foss-Feig, Fernando G. S. L. Brandão, and Alexey V. Gorshkov. Entanglement Area Laws for Long-Range Interacting Systems. *Phys. Rev. Lett.*, 119(5):050501, July 2017.

[152] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81(2):865–942, June 2009.

[153] Giovanni Ramírez, Javier Rodríguez-Laguna, and Germán Sierra. Entanglement over the rainbow. *J. Stat. Mech.*, 2015(6):P06002, 2015.

[154] Bojan Mohar. Isoperimetric inequalities, growth, and the spectrum of graphs. *Linear Algebra and its Applications*, 103:119–131, May 1988.

[155] F. R. K. Chung and Prasad Tetali. Isoperimetric Inequalities for Cartesian Products of Graphs. *Comb. Probab. Comput.*, 7(2):141–148, June 1998.

[156] Fan Chung. Laplacians and the Cheeger Inequality for Directed Graphs. *Ann. Comb.*, 9(1):1–19, April 2005.

[157] Oded Goldreich. Basic Facts about Expander Graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, Lecture Notes in Computer Science, pages 451–464. Springer, Berlin, Heidelberg, 2011.

[158] M. Ajtai, J. Komlós, and E. Szemerédi. An 0(N Log N) Sorting Network. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC '83, pages 1–9, New York, NY, USA, 1983. ACM.

[159] Omer Reingold. Undirected Connectivity in Log-space. *J ACM*, 55(4):17:1–17:24, September 2008.

[160] Irit Dinur. The PCP Theorem by Gap Amplification. *J ACM*, 54(3), June 2007.

[161] P. Elias, A. Feinstein, and C. Shannon. A note on the maximum flow through a network. *IRE Trans. Inf. Theory*, 2(4):117–119, December 1956.

[162] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Can. J. Math.*, 8(0):399–404, January 1956.

[163] D. R. Fulkerson and Rand Corporation. *Notes on Linear Programming. Part XVL, A Network-Flow Feasibility Theorem and Combinatorial Applications.* ASTIA Document ; No. AD 156011. Rand Corp., Santa Monica, Calif., 1958.

[164] Schrijver, Alexander. *Theory of Linear and Integer Programming.* Wiley-Interscience Series in Discrete Mathematics. Wiley-Interscience, Essex, 1986.

[165] Richard J. Trudeau. *Introduction to Graph Theory.* Dover Publications, New York, 2nd edition edition, February 1994.

[166] Román Orús and José I. Latorre. Universality of entanglement and quantum-computation complexity. *Phys. Rev. A*, 69(5):052308, May 2004.

[167] Vivien M. Kendon and William J. Munro. Entanglement and its Role in Shor's Algorithm. *Quantum Info Comput.*, 6(7):630–640, 2006.

[168] N. Schuch, M. M. Wolf, K. G. H. Vollbrecht, and J. I. Cirac. On entropy growth and the hardness of simulating time evolution. *New J. Phys.*, 10(3):033032, 2008.

[169] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen. Spin squeezing and reduced quantum noise in spectroscopy. *Phys. Rev. A*, 46(11):R6797–R6800, December 1992.

[170] Luca Pezzé and Augusto Smerzi. Entanglement, nonlinear dynamics, and the heisenberg limit. *Phys. Rev. Lett.*, 102(10):100401–100401, March 2009.

[171] Helmut Strobel, Wolfgang Muessel, Daniel Linnemann, et al. Quantum metrology. Fisher information and entanglement of non-Gaussian spin states. *Science*, 345(6195):424–7, July 2014.

[172] Géza Tóth. Multipartite entanglement and high-precision metrology. *Phys. Rev. A*, 85(2):022322–022322, February 2012.

[173] Hossein T. Dinani, Manish K. Gupta, Jonathan P. Dowling, and Dominic W. Berry. Quantum-enhanced spectroscopy with entangled multiphoton states. *Phys. Rev. A*, 93(6):063804–063804, June 2016.

[174] P. Kómár, T. Topcu, E. M. Kessler, et al. Quantum Network of Atom Clocks: A Possible Implementation with Neutral Atoms. *Phys. Rev. Lett.*, 117(6):060506, August 2016.

[175] Raju Valivarthi, Marcel.li Grimau li Grimau Puigibert, Qiang Zhou, et al. Quantum teleportation across a metropolitan fibre network. *Nat. Photonics*, 10(10):676–680, May 2016.

[176] B. Hensen, H. Bernien, A. E. Dréau, et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526(7575):682–686, October 2015.

[177] M. N. Nabighian, V. J. S. Grauch, R. O. Hansen, et al. 75th Anniversary The historical development of the magnetic method in exploration. *Geophysics*, 70(6):33ND–61ND, November 2005.

[178] H. Ding and B.F. Chao. The Slichter mode of the Earth: Revisit with optimal stacking and autoregressive methods on full superconducting gravimeter data set. *J Geophys Res Solid Earth*, 120:1–16, 2013.

[179] Tim J Wright, Barry E Parsons, and Zhong Lu. Toward mapping surface deformation in three dimensions using InSAR. *Geophys. Res. Lett.*, 31(October 2003):L01607–L01607, 2004.

[180] Frank D. Stacey. The seismomagnetic effect. *Pure Appl. Geophys.*, 58(1):5–22, 1964.

[181] Matti Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Olli V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.*, 65(2):413–497, April 1993.

[182] Samuel L. Braunstein, Carlton M. Caves, and G.J. Milburn. Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance. *Ann. Phys.*, 247(1):135–173, April 1996.

[183] Matteo G. A. Paris. Quantum estimation for quantum technology. *Int. J. Quantum Inf. Suppl.*, 7:125–137, April 2008.

[184] Sergio Boixo, Steven T Flammia, Carlton M Caves, and J M Geremia. Generalized limits for single-parameter quantum estimation. *Phys. Rev. Lett.*, 98(9):090401–090401, March 2007.

[185] Shengshi Pang and Todd A. Brun. Quantum metrology for a general Hamiltonian parameter. *Phys. Rev. A*, 90(2):022117–022117, August 2014.

[186] Tillmann Baumgratz and Animesh Datta. Quantum enhanced estimation of a multi-dimensional field. *Phys. Rev. Lett.*, 116(3):9–9, July 2015.

[187] Philip J. D. Crowley, Animesh Datta, Marco Barbieri, and I. A. Walmsley. Tradeoff in simultaneous quantum-limited phase and loss estimation in interferometry. *Phys. Rev. A*, 89(2):023845–023845, February 2014.

[188] Yang Gao and Hwang Lee. Bounds on quantum multiple-parameter estimation with Gaussian state. *Eur. Phys. J. D*, 68(11):347–347, November 2014.

[189] M. G. Genoni, M. G. A. Paris, G. Adesso, et al. Optimal estimation of joint parameters in phase space. *Phys. Rev. A*, 87(1):012107–012107, January 2013.

[190] Peter C Humphreys, Marco Barbieri, Animesh Datta, and Ian A Walmsley. Quantum enhanced multiple phase estimation. *Phys. Rev. Lett.*, 111(7):070403–070403, August 2013.

[191] Pieter Kok, Jacob Dunningham, and Jason F. Ralph. Role of entanglement in calibrating optical quantum gyroscopes. *Phys. Rev. A*, 95(1):012326, January 2017.

[192] Jie-Dong Yue, Yu-Ran Zhang, and Heng Fan. Quantum-enhanced metrology for multiple phase estimation with noise. *Sci. Rep.*, 4:5933–5933, January 2014.

[193] Yu-Ran Zhang and Heng Fan. Quantum metrological bounds for vector parameters. *Phys. Rev. A*, 90(4):043818–043818, October 2014.

[194] T. J. Proctor, P. A. Knott, and J. A. Dunningham. Networked quantum sensing. February 2017.

[195] C. Helstrom. The minimum variance of estimates in quantum signal detection. *IEEE Trans. Inf. Theory*, 14(2):234–242, March 1968.

[196] Carl W. Helstrom. Quantum detection and estimation theory. *J. Stat. Phys.*, 1(2):231–252, 1969.

[197] SL Braunstein and CM Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72(22):3439–3443, May 1994.

[198] Petre Stoica, Boon Chong Ng, and Student Member. On the Cramer – Rao Bound Under Parametric Constraints. *IEEE Signal Process. Lett.*, 5(7):177–179, 1998.

[199] Harry L Van Trees and Kristine L Bell. *Detection Estimation and Modulation Theory, Part I.* Wiley,, Hoboken :, 2nd ed. nv edition, 2013.

[200] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. J. Heinzen. Squeezed atomic states and projection noise in spectroscopy. *Phys. Rev. A*, 50(1):67–88, July 1994.

[201] Richard D. Gill and Boris Y. Levit. Applications of the van Trees inequality: A Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2):59–79, 1995.

[202] B. M. Escher, R. L. de Matos Filho, and L. Davidovich. General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology. *Nat. Phys.*, 7(5):406–411, March 2011.

[203] Mihai D Vidrighin, Gaia Donati, Marco G Genoni, et al. Joint estimation of phase and phase diffusion for quantum metrology. *Nat. Commun.*, 5:3532–3532, January 2014.

[204] S. F. Huelga, C. Macchiavello, T. Pellizzari, et al. Improvement of Frequency Standards with Quantum Entanglement. *Phys. Rev. Lett.*, 79(20):3865–3868, November 1997.

[205] Lorenza Viola, Emanuel Knill, and Seth Lloyd. Dynamical Decoupling of Open Quantum Systems. *Phys. Rev. Lett.*, 82(12):2417–2421, March 1999.

[206] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin. Quantum Error Correction for Metrology. *Phys. Rev. Lett.*, 112(15):150802–150802, April 2014.

[207] Xiao-Ming Lu, Sixia Yu, and C H Oh. Robust quantum metrological schemes based on protection of quantum Fisher information. *Nat. Commun.*, 6:7282–7282, January 2015.

[208] Masahiro Kitagawa and Masahito Ueda. Squeezed spin states. *Phys. Rev. A*, 47(6):5138–5143, June 1993.

[209] Emily Davis, Gregory Bentsen, and Monika Schleier-Smith. Approaching the Heisenberg Limit without Single-Particle Detection. *Phys. Rev. Lett.*, 116(5):053601–053601, February 2016.

[210] R. H. Dicke. Coherence in Spontaneous Radiation Processes. *Phys. Rev.*, 93(1):99–110, January 1954.

[211] MJ Holland and K Burnett. Interferometric detection of optical phase shifts at the Heisenberg limit. *Phys. Rev. Lett.*, 71(9):1355–1358, August 1993.

[212] S Campbell, M S Tame, and M Paternostro. Characterizing multipartite symmetric Dicke states under the effects of noise. *New J. Phys.*, 11(7):073039–073039, July 2009.

[213] Bernd Lücke, Jan Peise, Giuseppe Vitagliano, et al. Detecting Multiparticle Entanglement of Dicke States. *Phys. Rev. Lett.*, 112(15):155304–155304, April 2014.

[214] Peter Mansfield. Snapshot Magnetic Resonance Imaging (Nobel Lecture). *Angew. Chem. Int. Ed.*, 43(41):5456–5464, October 2004.

[215] H. J. Mamin, M. Kim, M. H. Sherwood, et al. Nanoscale Nuclear Magnetic Resonance with a Nitrogen-Vacancy Spin Sensor. *Science*, 339(6119), 2013.

[216] L Rondin, J-P Tetienne, T Hingant, et al. Magnetometry with Nitrogen-Vacancy Defects in Diamond. *Rep. Prog. Phys.*, 77(5):056503–056503, 2014.

[217] Stephen J. DeVience, Linh M. Pham, Igor Lovchinsky, et al. Nanoscale NMR spectroscopy and imaging of multiple nuclear species. *Nat. Nanotechnol.*, 10(2):129–134, January 2015.

[218] I. Lovchinsky, A. O. Sushkov, E. Urbach, et al. Nuclear magnetic resonance detection and spectroscopy of single proteins using quantum logic. *Science*, 351(6275):836–841, February 2016.

[219] Andrii Lazariev and Gopalakrishnan Balasubramanian. A nitrogen-vacancy spin based molecular structure microscope using multiplexed projection reconstruction. *Sci. Rep.*, 5:14130–14130, September 2015.

[220] V. S. Perunicic, C. D. Hill, L. T. Hall, and L.C.L. Hollenberg. A quantum spin-probe molecular microscope. *Nat. Commun.*, 7:12667–12667, October 2016.

[221] F. Dolde, I. Jakobi, B. Naydenov, et al. Room-temperature entanglement between single defect spins in diamond. *Nat. Phys.*, 9(3):139–139, 2013.

[222] Florian Dolde, Ville Bergholm, Ya Wang, et al. High-fidelity spin entanglement using optimal control. *Nat. Commun.*, 5:345–380, February 2014.

[223] C. Zu, W.-B. Wang, L. He, et al. Experimental realization of universal geometric quantum gates with solid-state spins. *Nature*, 514(7520):72–75, October 2014.

[224] H. Bernien, B. Hensen, W. Pfaff, et al. Heralded entanglement between solid-state qubits separated by three metres. *Nature*, 497(7447):86–90, April 2013.

[225] C. S. Wolfe, S. A. Manuilov, C. M. Purser, et al. Spatially resolved detection of complex ferromagnetic dynamics using optically detected nitrogen-vacancy spins. *Appl. Phys. Lett.*, 108(23):232409–232409, June 2016.

[226] K. Arai, C. Belthangady, H. Zhang, et al. Fourier magnetic imaging with nanoscale resolution and compressed sensing speed-up using electronic spins in diamond. *Nat. Nanotechnol.*, 10(10):859–864, August 2015.

[227] Kasper Jensen, Rima Budvytyte, Rodrigo A. Thomas, et al. Non-invasive detection of animal nerve impulses with an atomic magnetometer operating near quantum limited sensitivity. *Sci. Rep.*, 6:29638–29638, January 2016.

[228] Quntao Zhuang, Zheshen Zhang, and Jeffrey H. Shapiro. Distributed quantum sensing using continuous-variable multipartite entanglement. *Phys Rev A*, 97(3):032329, March 2018.

[229] Shelby Kimmel, Guang Hao Low, and Theodore J. Yoder. Robust Calibration of a Universal Single-Qubit Gate-Set via Robust Phase Estimation. February 2015.

[230] Wenchao Ge, Kurt Jacobs, Zachary Eldredge, Alexey V. Gorshkov, and Michael Foss-Feig. Distributed Quantum Metrology with Linear Networks and Separable Inputs. *Phys. Rev. Lett.*, 121(4):043604, July 2018.

[231] Timothy J. Proctor, Paul A. Knott, and Jacob A. Dunningham. Multiparameter Estimation in Networked Quantum Sensors. *Phys Rev Lett*, 120(8):080501, 2018.

[232] Taesoo Kim, Olivier Pfister, Murray J. Holland, Jaewoo Noh, and John L. Hall. Influence of decorrelation on Heisenberg-limited interferometry with quantum correlated photons. *Phys. Rev. A*, 57(5):4004–4013, May 1998.

[233] A. R. Usha Devi and A. K. Rajagopal. Quantum target detection using entangled photons. *Phys. Rev. A*, 79(6):062320–062320, June 2009.

[234] Rafal Demkowicz-Dobrzański, Marcin Jarzyna, and Jan Kołodyński. Chapter Four - Quantum Limits in Optical Interferometry. volume 60 of *Progress in Optics*, pages 345–435. Elsevier, 2015.

[235] Tak W. Kee and Marcus T. Cicerone. Simple approach to one-laser, broadband coherent anti-Stokes Raman scattering microscopy. *Opt. Lett.*, 29(23):2701–2701, December 2004.

[236] Orang Alem, Tilmann H Sander, Rahul Mhaskar, et al. Fetal Magnetocardiography Measurements with an Array of Microfabricated Optically Pumped Magnetometers. *Phys. Med. Biol.*, 60(12):4797–4811, 2015.

[237] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi. Optimal estimation of group transformations using entanglement. *Phys. Rev. A*, 72(4):042338, October 2005.

[238] H. Häffner, C. F. Roos, and R. Blatt. Quantum computing with trapped ions. *Physics Reports*, 469(4):155–203, December 2008.

[239] K. R. Brown, A. C. Wilson, Y. Colombe, et al. Single-qubit-gate error below 10^{-4} in a trapped ion. *Phys Rev A*, 84(3):030303, September 2011.

[240] M de Angelis, A Bertoldi, L Cacciapuoti, et al. Precision gravimetry with atomic sensors. *Meas. Sci. Technol.*, 20(2):022001–022001, February 2008.

[241] J. E. Debs, P. A. Altin, T. H. Barter, et al. Cold-atom gravimetry with a Bose-Einstein condensate. *Phys. Rev. A*, 84(3):033610–033610, September 2011.

[242] Stephen Ragole and Jacob M. Taylor. Interacting Atomic Interferometry for Rotation Sensing Approaching the Heisenberg Limit. *Phys. Rev. Lett.*, 117(20):203002, November 2016.

[243] T. L. Gustavson, P. Bouyer, and M. A. Kasevich. Precision Rotation Measurements with an Atom Interferometer Gyroscope. *Phys. Rev. Lett.*, 78(11):2046–2049, March 1997.

[244] S. Kolkowitz, I. Pikovski, N. Langellier, et al. Gravitational wave detection with optical lattice atomic clocks. *Phys. Rev. D*, 94(12):124043, December 2016.

[245] David Layden and Paola Cappellaro. Spatial noise filtering through error correction for quantum sensing. *Npj Quantum Inf.*, 4(1):30, July 2018.

[246] Stuart Russell and Peter Norvig. In *Artificial Intelligence, A Modern Approach*. Prentice Hall, 2010.

[247] Jürgen Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural Netw.*, 61:85–117, 2015.