ABSTRACT

Title of Dissertation: QUANTUM ADVANTAGE IN SENSING AND SIMULATION

Adam Ehrenberg
Doctor of Philosophy, 2024

Dissertation Directed by: Professor Alexey V. Gorshkov
Department of Physics

Since the discovery of Shor's factoring algorithm, there has been a sustained interest in finding more such examples of *quantum advantage*, that is, tasks where a quantum device can outperform its classical counterpart. While the universal, programmable quantum computers that can run Shor's algorithm represent one direction in which to search for quantum advantage, they are certainly not the only one. In this dissertation, we study the theory of quantum advantage along two alternative avenues: sensing and simulation.

Sensing refers to the task of measuring some unknown quantity with the smallest possible error. In many cases, when the sensing apparatus is a quantum device, this ultimate achievable precision, as well as specific protocols producing estimators with this precision, are unknown. In this dissertation, we help close this gap for both qubit-based and photonic quantum sensors for the specific task of measuring a linear function of unknown parameters. We use quantum Fisher information and the quantum Cramér-Rao bound to derive limits on their ultimate precision. We further develop an algebraic framework that allows us to derive protocols saturating these

bounds and better understand the quantum resources, such as entanglement, that are necessary to implement these protocols. In doing so, we help clarify how quantum resources like entanglement lead to more precise sensing.

We also study a specific form of simulation called Gaussian Boson Sampling, which is a member of the broad framework of random sampling tasks that have become a popular method for demonstrating quantum advantage. While many of the theoretical underpinnings of these random sampling tasks, including Gaussian Boson Sampling, are well understood, many questions remain. Anticoncentration, which is strongly related to the moments of the output distribution, is a particularly relevant property when it comes to formally proving the existence of quantum advantage. We develop a graph-theoretic framework to calculate these moments, and we show that there is a transition in the strength of anticoncentration as a function of how many of the photonic modes are initially squeezed. We therefore demonstrate a transition in the evidence for the so-called approximate average-case hardness of Gaussian Boson Sampling, hence clarifying in what regimes we have the strongest evidence for quantum advantage.

Finally, we also discuss the simulation complexity of Many-Body Localized systems. Many-Body Localization is a widely studied phase of matter that is often characterized by the appearance of a large number of quasilocal integrals of motion (operators that commute with the Hamiltonian) that interact via exponentially decaying interactions. In this dissertation, we study a phenomenological form of Many-Body Localization and show three main results. First, we demonstrate that, for polynomially long evolution times under a Hamiltonian in the Many-Body Localized phase, there is a quasipolynomial-time classical algorithm that can perform strong simulation of the output state. On the flip side, our second result is that, when the evolution time is exponentially long, weak simulation of the output state becomes formally classically hard.

Finally, as a consequence of our classical results, we show the approximate quantum circuit complexity of these Hamiltonians grows sublinearly in the evolution time (in contrast with the proposed linear growth for chaotic Hamiltonians). Thus, this work helps clarify whether and how we might find quantum advantage via simulating certain types of condensed matter systems.

QUANTUM ADVANTAGE IN SENSING AND SIMULATION

by

Adam Ehrenberg

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
2024

Advisory Committee:
Professor Steven L. Rolston, Chair
Professor Alexey V. Gorshkov, Co-Chair and Research Advisor
Professor Andrew M. Childs, Dean's Representative
Professor Nathan Schine
Professor Michael J. Gullans

Acknowledgments

First, it is a pleasure to thank my advisor, Professor Alexey Gorshkov, for everything that he has done to make this journey possible. Alexey has been a consistently supportive and instructive guide. He is an exceptionally gifted physicist who has taught me a great deal about how to conduct research and, more generally, how to solve problems. I must also thank him for his unwavering patience and kindness in the face of the many challenges that I faced during this process. When I told him I needed to take time away from the program after my second year in order to reset and figure out what I wanted, he did not hesitate to give me that space and help insulate me from some of the fallout that decision could have caused. Without his willingness to welcome me back to his group even after this extended absence, I would certainly not have completed this dissertation. I cannot imagine a better holistic advisor than the one I was lucky enough to have.

It is also a great pleasure to thank the many collaborators I have had the chance to work with over these past years. In particular, I would like to thank all those who worked on the projects that comprise the majority of this dissertation: Jacob Bringewatt, Taurshii Goel, Joseph Iosue, Abhinav Deshpande, Dominik Hangleiter, Christopher Baldwin, Dmitry Abanin, and, of course, Alexey. There are many other collaborators that I have had the pleasure of working with on projects that are not contained in this dissertation, and I thank everyone with whom I had the privilege of working for spending their time and academic energy with me. I also want to thank all the members of the Gorhskov group, even if we did not have a chance to work on a project together, for creating a fun, supportive, and collaborative atmosphere in which to pursue new knowledge.

A few collaborators are worthy of special mention. Abhinav Deshpande, Dominik Hangleiter, and Minh Tran have been the three best mentors I could have possibly asked for. So much of what I have learned and what I have accomplished can be directly attributed to your teaching. Your patience, generosity, and overall kindness are admirable and have facilitated so much of my academic journey. Jacob Bringewatt and Joe Iosue were absolute pleasures to work with on multiple projects, and you both played a significant role in improving the science that I was able to accomplish over these years. I am very appreciative of the time, energy, thoughtfulness, and excellence that you brought to our collaborations. You both have a bright future as physicists and mathematicians (should that be what you want).

I would like to thank the members of the Mental Health Task Force, Professor Chandra Turpen, Elizabeth Bennewitz, Kellen O'Brien, Patrick Banner, and Zach Steffen, for working to try to make the department a better place, and for also creating a simultaneously serious and silly space to vent as I have gotten deeper and deeper into the Ph.D.

Beyond those who have supported my work as a student, I would also like to thank those who have made life outside of work fulfilling, meaningful, and joyful. In Maryland/D.C.: Alex Wikner, Cara DeCusatis, Dhruvit Patel, Elizabeth Bennewitz, Grace Brown, Jacob Bringewatt, Martin Ritter, Nicholas Mennona, Peter Mancini, Selu Villaescusa-Nadal, Stefano Antonini, and Zach Steffen, you have been a wonderful support network that has made it possible to relax and enjoy day-to-day life outside of research. Meeting all of you has made me glad that I chose to attend the University of Maryland. It is also important for me to thank those from Atlanta and Cambridge with whom I have maintained such close relationships even from afar: Alex Goldstein, Andrew Abbot, Anna Figueroa, Ari Nichol, Ben Banavige, Cam Futral, Ezra Zigmond, Jessie Anderson, John Rahill, Milan Ravenell, Nick Padula, Naomi Carolan, Noah Dezen, Priyanka Farrell, Sam Orr, Sam Tatum, and Zach Schnell. Being able to pick up like no time has passed every time I see you or talk on the phone has made breaks from work more restful and restorative experiences. Truly special thanks go to Ezra Zigmond, Sam Orr, and Zach Schnell who have been such consistent shoulders to lean on for more than a decade. I also owe a great debt of gratitude to the Orr, Schnell, and Scanlon/Sinks families for supporting me and providing homes away from home.

Most importantly, I would like thank my own family. Abbe, Betsy, and Beth, you all have provided support and encouragement from afar, both to me individually and to the whole family. Philip, you have consistently provided a sounding board for me to work through whatever happens to be bothering me. You also helped me find a meaningful space to contribute to Democratic causes in Georgia when I needed time away from school. Without your help making my break from physics worthwhile, I probably would not have finished this degree. Mom and Dad, you made this entire endeavor possible. I would never have begun a Ph.D. in the first place without you two instilling in me a love of science and a commitment to excellence, and I never would have finished it without you pushing me to achieve my best while never wavering in your support. I love all of you (and Tim and Sasha) dearly, even if I do not often say it.

Finally, I must thank Simone for providing much needed comfort and calm for the past three years. She helped me readjust to the life of a graduate student after I returned, and she has been a truly wonderful (if occasionally too loud) companion.

Summary of Research Contributions

Chapters 2 to 6 and their associated Appendices A to E (respectively) are based on papers that are either published in peer-review journals or listed on the arXiv while they are under review for publication. These Chapters and the associated Appendices have been copied nearly verbatim from their published/pre-printed counterparts. The only changes consist of minor grammatical or stylistic changes to ensure consistency with the format of this dissertation or to correct typos. Where necessary, footnotes have been added with the phrase "Note added:" to help clarify any details, including adding information that either corrects minor mistakes or adds some new insight. For this reason, this dissertation uses, almost exclusively, the plural "we," even in Chapters 1 and 7, where I am the sole author. Additionally, because the papers upon which Chapters 2 and 3 are based have me listed as a co-first author, I detail below the specific contributions that I made to each of these Chapters.

Chapter 2 and the associated Appendix A were originally written together with Jacob Bringewatt and Alexey V. Gorshkov and published as Ref. [1]. Jacob and I are listed as co-first author on this paper. I was responsible for the original concept of the paper, that is, the idea of minimizing entanglement in optimal protocols for quantum sensing. Correspondingly, I was responsible for the development of these new time-dependent protocols and proving that they satisfy the necessary information-theoretic conditions for optimality. I also specifically showed how to minimize average entanglement. However, the proof of the main theorem, specifically, how instantaneous entanglement is minimized, was developed fully jointly. Additionally, the discussion of the CNOT costs of the protocols was also addressed collaboratively; I focused more on the analytical analysis, while Jacob focused more on the numerics. Jacob took the lead on the remaining sections. Specifically, he focused on the introduction and conclusion as well as the derivation of the Fisher Information conditions for optimal protocols. He also wrote on the appropriate accounting for resources for saturating the quantum Cramér-Rao bound, which led to the Appendix on robust phase estimation and the section on the insufficiency of time-independent protocols as currently formulated in the literature.

Chapter 3 and the associated Appendix B were originally written together with Jacob Bringewatt, Tarushii Goel, and Alexey V. Gorshkov and published as Ref. [2]. Jacob, Tarushii, and I are listed as co-first authors on this paper. This paper was originally a summer project for Tarushii designed by Jacob, Alexey, and me to extend the work of the previous Chapter to photonic degrees of freedom. As such, I (along with Jacob and Alexey) was an advisor on much of the original work and exploration that Tarushii did to begin and advance the project. I provided the original conception of the proof for the bounds on optimal protocols for the case of number operator couplings (which is the central case of interest in the work), and Tarushii extended this outline to the first formal version of the proof. I also extended some of the key lemmas from the previous Chapter/Appendix into the photonic context and wrote the first version of the section regarding minimum entanglement based on ideas that Tarushii developed. Jacob formalized the proof of the lower bounds and provided the first sketch of the proofs for the quadrature operator case in our algebraic language. He also wrote the first draft of most of the main manuscript. The families of optimal protocols and the corresponding proofs of optimality were developed fully collaboratively with Jacob.

Citations to Previously Published Work

As stated in the Summary of Research Contributions, the majority of this dissertation appears in papers that have either already been published or are under consideration for being published (and currently exist on the arXiv preprint repository). These Chapters and the associated Appendices have been copied nearly verbatim from their published/pre-printed counterparts. The only changes consist of minor grammatical or stylistic changes to ensure consistency with the format of this dissertation or to correct typos. We now provide citations to these works, separating by the two major themes of this dissertation. We also provide additional citations to related work that the author of this dissertation has contributed to, but that is not explicitly covered in this dissertation.

Quantum Sensing

- Chapter 2 and Appendix A: A. Ehrenberg*, J. Bringewatt*, and A.V. Gorshkov. Minimum-Entanglement Protocols for Function Estimation. *Phys. Rev. Res.*, 5:033228, 2023. [1]
- Chapter 3 and Appendix B: J. Bringewatt*, A. Ehrenberg*, T. Goel*, and A.V. Gorshkov. Optimal Function Estimation with Photonic Quantum Sensor Networks. *Phys. Rev. Res.*, 6:013246, 2024. [2]

Other works related to the study of Fisher information and quantum metrology include:

[1] A. Ehrenberg and J. Bringewatt are listed as having contributed equally to this publication. See the Summary of Research Contributions for details on the specific contributions from the author of this dissertation.

[2] J. Bringewatt, A. Ehrenberg, and T. Goel are listed as having contributed equally to this publication. See the Summary of Research Contributions for details on the specific contributions from the author of this dissertation.

- P. Niroula, J. Dolde, X. Zheng, J. Bringewatt, A. Ehrenberg, K. C. Cox, J. Thompson, M. J. Gullans, S. Kolkowitz, and A. V. Gorshkov. Quantum Sensing with Erasure Qubits. *arXiv:2310.01512*, 2023.
    - Studies how one can achieve metrological gains by using qubits whose dominant error source is erasure (which is easily detectable and removes the state from the computational subspace) rather than more commonly considered sources, such as depolarization and dephasing.
- L. P. García-Pintos, K. Bharti, J. Bringewatt, H. Dehghani, A. Ehrenberg, N. Yunger Halpern, and A. V. Gorshkov. Estimation of Hamiltonian Parameters from Thermal States. *arXiv:2401.10343*, 2023.
    - Uses Fisher information and the quantum Cramér-Rao bound to place upper and lower bounds on how well one can estimate unknown parameters in a Hamiltonian using Gibbs states that are thermalized to a known temperature with respect to this Hamiltonian.

Quantum Simulation

- Chapter 4 and Appendix C: A. Ehrenberg, J. T. Iosue, A. Deshpande, D. Hangleiter, and A. V. Gorshkov. Transition of Anticoncentration in Gaussian Boson Sampling. *arXiv:2312.08433*, 2023.
- Chapter 5 and Appendix D: A. Ehrenberg, J. T. Iosue, A. Deshpande, D. Hangleiter, and A. V. Gorshkov. The Second Moment of Hafnians in Gaussian Boson Sampling. *arXiv:2403.13878*, 2024.

- Chapter 6 and Appendix E: A. Ehrenberg, A. Deshpande, C. L. Baldwin, D. A. Abanin, and A. V. Gorshkov. Simulation Complexity of Many-Body Localized Systems. *arXiv:2205.12967*, 2022.

Other works related to quantum simulation (especially considered through the lens of using Lieb-Robinson bounds in long-range interacting systems) and the study of Gaussian bosonic systems include:

- N. Maskara, A. Deshpande, A. Ehrenberg, M. C. Tran, B. Fefferman, and A. V. Gorshkov. Complexity Phase Diagram for Interacting and Long-Range Bosonic Hamiltonians. *Phys. Rev. Lett.* 129:150604, 2022.
  - Studies the computational complexity of simulating systems of interacting bosonic particles with long-range hoppings. Some of the results are used to improve the scope of the classical hardness results in Chapter 6.
- J. T. Iosue, A. Ehrenberg, D. Hangleiter, A. Deshpande, and A. V. Gorshkov. Page Curves and Typical Entanglement in Linear Optics. *Quantum*, 7:1017, 2023.
  - Studies the Page curve, or average entanglement, and the typicality of entanglement in a system of single-mode squeezed vacuum states (this system is nearly identical to that described in Chapters 4 and 5, but the analysis requires that all modes be squeezed with the same squeezing parameter). Entanglement here is measured by the Rényi-2 entropy of the reduced state between a partition of the bosonic modes.
- J. Youm, J. T. Iosue, A. Ehrenberg, Y.-X. Wang, and A. V. Gorshkov. Average Rényi Entanglement Entropy in Gaussian Boson Sampling. *arXiv.2403.18890*, 2024.

  - Generalizes the previous paper to the case of arbitrary positive integer Rényi entropies, including the von Neumann entropy.

- J. T. Iosue*, T. C. Mooney*, A. Ehrenberg, and A. V. Gorshkov. Projective Toric Designs, Difference Sets, and Quantum State Designs. *arXiv:2311.13479*, 2023.

  - Provides various constructions of t-designs on projective tori, which can be combined with simplex designs to form quantum state designs. These quantum state designs play a role in many topics in quantum information theory and the geometry of quantum states, and they often relate to various aspects of the theory of quantum simulation.

- C. L. Baldwin, A. Ehrenberg, A. Y. Guo, and A. V. Gorshkov. Disordered Lieb-Robinson Bounds in One Dimension. *PRX Quantum*, 4:020349, 2023.

  - Shows that an explicit consideration of the disorder (a lack of translation-invariance) in one-dimensional quantum spin chains can lead to tighter Lieb-Robinson bounds for these systems based on the prevalence of so-called "weak links." These bounds are saturable by time-dependent "SWAP-based" Hamiltonians.

- M. C. Tran, A. Y. Guo, C. L. Baldwin, A. Ehrenberg, A. V. Gorshkov, and A. Lucas. Lieb-Robinson Light Cone for Power-Law Interactions. *Phys. Rev. Lett.*, 127:160401, 2021.

  - Finds the optimal Lieb-Robinson light cone in a previously undetermined regime of long-range interacting systems.

- M. C. Tran, C.-F. Chen, A. Ehrenberg, A. Y. Guo, A. Deshpande, Y. Hong, Z.-X. Gong,

A. V. Gorshkov, and A. Lucas. Hierarchy of Linear Light Cones with Long-Range Interactions. *Phys. Rev. X*, 10:031009, 2020.

  - Studies quantum information transfer in long-range interacting systems as defined by the Lieb-Robinson and Frobenious light cones. Also contains a robust discussion of applications of these bounds, including to digital quantum simulation.

- M. C. Tran, A. Ehrenberg, A. Y. Guo, P. Titum, D. A. Abanin, and A. V. Gorshkov. Locality and Heating in Periodically Driven, Power-Law-Interacting Systems. *Phys. Rev. A*, 100:052103, 2019.

  - Uses Lieb-Robinson bounds to show that, when interactions decay quickly enough, the heating time for periodically driven, long-range interacting systems scales exponentially in the drive frequency. We show this using both linear response theory and a Magnus-like expansion.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1D	One Dimension (One-Dimensional)
2D	Two Dimensions (Two-Dimensional)
AL	Anderson-Localization (Anderson-Localized)
BPP	Bounded-Error Probabilistic Polynomial Time
BQP	Bounded-Error Quantum Polynomial Time
BS	Boson Sampling
COE	Circular Orthogonal Ensemble
CRB	Cramér-Rao Bound
EOM	Electro-Optical Modulator
FI	Fisher Information
FIM	Fisher Information Matrix
GBS	Gaussian Boson Sampling
GNFS	General Number Field Sieve
IQP	Instantaneous Quantum Polynomial
LHS	Left-Hand Side
LIOM	(Quasi-)Local Integral of Motion
LXEB	Linear Cross-Entropy Benchmarking
MBL	Many-Body Localization (Many-Body Localized)
MBQC	Measurement-Based Quantum Computing
NIST	National Institute of Standards and Technology
P	Polynomial-Time
QCRB	Quantum Cramér-Rao Bound
QFI	Quantum Fisher Information
QFIM	Quantum Fisher Information Matrix
RHS	Right-Hand Side
RSA	Rivest–Shamir–Adleman
SBS	Scattershot Boson Sampling
SM	Supplemental Material
SMSV	Single-Mode Squeezed Vacuum
TMSV	Two-Mode Squeezed Vacuum
TVD	Total Variation Difference
XEB	Cross-Entropy Benchmarking

Chapter 1: Introduction

In this dissertation, we discuss some of the theoretical concepts behind quantum advantage in various contexts. Quantum advantage, broadly defined, refers to some sort of information-theoretic task in which a device that operates under the rules of quantum mechanics can outperform the best possible classical device.[1] Since the development of Shor's algorithm in 1994 [3], which shows that there exists a BQP[2] algorithm for factoring integers, there has been an explosion of interest in quantum advantage through the specific lens of quantum computing. This is because, despite much effort, the best-known (in terms of asymptotic scaling in the size of the integer) classical factoring algorithm, the General Number Field Sieve (GNFS), runs in superpolynomial time [5]. Therefore, a formal proof that there exists *no* classical algorithm that runs in polynomial time (i.e., the decision problem version of factoring is not in either of the complexity classes P or BPP[3], which many conjecture to be equivalent) would demonstrate a

[1] It is a bit early for a footnote, but we would be remiss not to add: Of course, to the best of our knowledge, *all* devices operate by the rules of quantum mechanics; here, we are distinguishing between devices where the use of quantum mechanics is required to accurately describe the operation of this device versus those that admit an accurate effective classical theory.

[2] BQP, or Bounded-Error Quantum Polynomial Time, is the complexity class of languages that can be decided with error at most $1/3$ in quantum polynomial time [4]. More intuitively, it is the class of decision problems that are efficiently solvable (with high probability) by a quantum computer. It is, in some sense, the quantum version of BPP, described below.

[3] P, or Polynomial Time, refers to the complexity class of languages decidable by a Turing machine in polynomial time. BPP, or Bounded-Error Probabilistic Polynomial Time, is the class of languages decidable with error at most $1/3$ using a polynomial-time probabilistic Turing machine [4]. Again, for some intuition, BPP and P are the classes of decision problems solvable in classical polynomial time with or without the use of randomness, respectively.

superpolynomial separation between the best quantum and classical algorithms.[4] This would disprove the (classical) complexity-theoretic Church-Turing thesis (also referred to as the extended Church-Turing thesis), which states that any reasonable model of computing can be simulated up to polynomial overhead by a Turing machine (see, e.g., Ref. [7]). Such a proof of quantum advantage would have implications beyond just complexity theory, as factoring is at the heart of the security of the popular Rivest–Shamir–Adleman (RSA) cryptosystem [8]. The existence of an efficient quantum algorithm for factoring, therefore, places the security of RSA at risk, prompting the National Institute of Standards and Technology (NIST) to update their cryptography standards to account for threats posed by Shor's algorithm [9].

However, quantum advantage, conceptually, need not be restricted to solving some sort of decision problem like factoring. Indeed, an insular focus on Shor's algorithm and the factoring problem would crucially overlook many of the other potential benefits that quantum devices can provide. For one thing, actually implementing Shor's algorithm requires using quantum error correction, which is a method by which quantum information can be preserved despite its natural tendency to decohere over time. While the threshold theorem [10–12], shows that quantum error correction is possible (that is, it is possible to correct errors at a rate faster than they reappear through the correction process), these schemes can be quite complicated, and they often require huge overheads in the number of qubits or gates. While fundamental improvements in quantum error correction, gate design, or qubit architecture might eventually improve fidelity enough to

[4] We note one more small subtlety here. The problem of finding a factor of or listing all factors of a given integer is not actually a decision problem, but instead a function problem. Therefore, if phrased in this way, factoring is technically not in BQP. However, one can easily turn this function version of factoring into a decision problem; does there exist a factor of the given integer that is between 1 and some other given value? Being able to solve this decision problem also allows one to solve the functional version by a sort of binary search procedure. Therefore, Shor's algorithm, which does actually produce a factor of the integer, can easily be turned into a BQP algorithm, which is why we refer to the decision problem version of factoring. See Ref. [6] for a discussion of this matter.

lower these numbers to a more reasonable regime, the time required to achieve these gains is undetermined.

Therefore, while there is strong interest in developing more examples of quantum advantage in the quantum computation realm, there is also an interest in exploring the advances that quantum advantage can provide in other domains, such as cryptography, communication, sensing, and non-universal simulation. In this dissertation, we focus on the latter two topics of sensing and simulation. Accordingly, this dissertation contains two main parts, the first of which consists of two Chapters and two associated Appendices, and the latter of which contains three of each. Each individual Chapter contains its own introduction that provides even more details and references regarding the topics at hand. Therefore, here we find it appropriate to discuss the broad strokes of what is contained in those Chapters with a focus on the main contributions and new results that are proven therein.

The first part of this dissertation, which consists of Chapters 2 and 3 and their respective Appendices A and B, discusses quantum advantage in the domain of quantum sensing, specifically for the problem of measuring a linear function of unknown parameters. We assume that there are d unknown parameters $\{\theta_i\}_{i=1}^{d}$ that are coupled to d quantum sensors, and the goal is to construct an estimator $\tilde{q}$ of the function $q = \sum_{i=1}^{d} \alpha_i \theta_i$, where $\alpha_i \in \mathbb{R}$ are arbitrary real parameters that define the linear function. Simple examples include the case where $\alpha_i = \delta_{ij}$, meaning we are actually only interested in a single parameter, and $\alpha_i = 1/d$, where we are actually interested in an average of all the parameters. Additionally, it is desired that this estimator $\tilde{q}$ be unbiased (i.e., for $\mathbb{E}[\tilde{q}] = q$, where the expectation value is taken with respect to the probabilistic outcomes of the positive operator-valued measure that is used to construct the estimator), and for the mean-squared error of the estimator to be as small as possible. This is a problem that has been studied before [13–22], and it is known that one can use quantum entangled states to improve on the achieveable mean-squared error of the estimator compared to an unentangled strategy.[5]

We study this problem for both the case of qubit-based (Chapter 2) and photonic (Chapter 3) sensors. Furthermore, we study each type of sensor from two complementary perspectives: bounds and protocols. Bounds refer to lower limits on how precise $\tilde{q}$ can be as measured by its mean-squared error $\mathcal{M}$. The bounds proven in this dissertation are based on the quantum Cramér-Rao bound, which itself follows from the calculation of the quantum Fisher information with respect to the function q, $\mathcal{F}(q)$: $\mathcal{M} \geq 1/\mathcal{F}(q)$.[6] In Chapters 2 and 3, we calculate the quantum Fisher information with respect to q, and we use the quantum Cramér-Rao bound to find the ultimate precision limits of quantum sensors, which we can then compare to unentangled strategies, thus demonstrating an advantage via entanglement.

Protocols, on the other hand, refer to specific sequences of operations that actually construct an estimator for the function of interest. Ideally, these protocols lead to an estimator that achieves the proven bounds (that is, an estimator whose mean-squared error is as small as possible). For quantum sensors, there are, typically, four main steps to these protocols: state preparation, unitary control, measurement, and classical postprocessing of the measurement results. While each of these steps is important, our main focus in this dissertation is on the second stage, unitary control.

[5] Here, we are technically not comparing quantum vs. classical strategies, but instead entangled vs. unentangled strategies. Strictly speaking, even without entanglement the device itself is still quantum, as the sensor still consists of qubits or photonic modes (in Fock or squeezed states, not semiclassical coherent states), and we still treat the problem in a fully quantum mechanical way. However, we still consider this separation an example of quantum advantage because it is known that entanglement, a purely quantum resource, is necessary for quantum sensors to outperform those that are connected by only classical correlations [23, 24]. Because unentangled strategies achieve a scaling given by the classical summing of errors in quadrature, we treat the unentangled strategy as a proxy for an equivalent classical sensor.

[6] Technically, this is the single-shot quantum Cramér-Rao bound; the full bound contains a factor μ in the denominator, where μ is the number of samples. The quantum Cramér-Rao bound is only saturable in the limit of asymptotically many shots, but one can get around this subtlety at the expense of a multiplicative constant. This is discussed in depth in both Appendices A and B.

The reasoning for this is as follows: we can absorb the initial state preparation into this unitary control (assuming some fiducial initial state, such as $|0\rangle^d$ in the case of qubit sensors). The measurement and classical postprocessing follow from earlier results in the literature. Therefore, our main contributions are in better understanding what sequences of unitary control lead to protocols that saturate the bounds, and then characterizing the resources needed to implement this control.

We can now summarize our main results. In Chapter 2 and Appendix A, we focus on the case of qubit sensors that are coupled to the parameters $\{\theta_i\}_{i=1}^d$ (often listed as a vector $\boldsymbol{\theta}$) via the Hamiltonian $\hat{H} = \frac{1}{2}\sum_{i=1}^d \theta_i \hat{\sigma}_i^z + \hat{H}_c$. Note that the factor of $\frac{1}{2}$ is merely a convention (it ensures that the seminorm of the generator on each qubit is normalized to 1). Note also that $\hat{H}_c$ is an arbitrary control Hamiltonian that can, generically, vary with time, but which has no $\boldsymbol{\theta}$-dependence; $\hat{H}_c$ is the source of the unitary control that is so crucial to the construction of our protocols. We use an algebraic approach to the Fisher information to rederive known bounds (see, e.g., Refs. [14, 19]) on the performance of these sensors. This rederivation of known bounds is not useless, however, as the algebraic framework that we utilize allows us to make powerful statements on the protocol side. Specifically, we use it derive an infinite family of time-dependent protocols that saturate these bounds; we therefore refer to these protocols as "optimal." We then characterize how much entanglement these optimal protocols use, specifically finding saturable upper and lower bounds on the amount of instantaneous and average entanglement needed over the course of these protocols. We prove the surprising result that fully entangled states are not necessary to achieve optimality (even though, naively, one might expect this to be the case given that all previous protocols utilized such states). We also characterize how many entangling gates are needed to run these protocols, though our results there are a bit more heuristic. In this way, this Chapter offers significant clarification in the role that entanglement plays in achieving the metrological gains that are possible via quantum sensors.

In Chapter 3 and Appendix B, we extend the previous results to the case of photonic sensors. There, we consider coupling the parameters to the photonic modes via both the number operator $\hat{n}$ and displacement operator $\hat{p}$ (we choose the momentum operator without loss of generality). In these cases, the actual optimal bounds were not known in all cases (for generators of the form $\hat{n}$, it was conjectured in Ref. [13] for functions with non-negative coefficients, but it was not proven; for generators of the form $\hat{p}$ Refs. [25, 26] provide some results, but they do not allow the function $\boldsymbol{\alpha}$ to have negative coefficients, nor do they consider arbitrary probe states). Therefore, our derivations of bounds using our algebraic approach to the quantum Fisher information do not just serve an ancillary purpose for finding optimal protocols, but are themselves useful results. As in Chapter 2, we also use the quantum Fisher information to find optimal time-dependent protocols and characterize their entanglement, again clarifying the role of quantum resources in achieving metrological gains.

The second part of this dissertation, which consists of Chapters 4 to 6 and their respective Appendices C to E, switches focus to simulation. Chapters 4 and 5 (Appendices C and D) and Chapter 6 (Appendix E) are about two different systems, which we introduce separately now.

Chapters 4 and 5 and Appendices C and D discuss quantum advantage in random sampling experiments, specifically in Gaussian Boson Sampling, which is a generalization of the famous framework of Boson Sampling developed by Aaronson and Arkhipov in Ref. [27]. Random sampling experiments are a promising avenue for near-term demonstrations of quantum advantage using non-universal quantum simulators (see Ref. [28] and references therein). Indeed, many random sampling experiments, such as those in Refs. [29–33], already have been performed. The general concept of a random sampling task is as follows: (1) some fiducial initial state, say $|0\rangle^m$ for a system of m qubits or the state $|1\dots10\dots0\rangle$ for a system of photonic modes, is prepared; (2) a unitary operator randomly drawn from some distribution, typically the Haar (that is, the uniform) distribution, is applied to this state; (3) the post-unitary state is measured in some simple basis, such as the computational basis for a system of qubits, or the Fock basis for a system of photonic modes. This measurement results in a given basis state with probability according to the Born rule. The task of random circuit sampling, speaking in highly general terms, is to mimic this sampling procedure and produce a sample from the output distribution described in this way.

In this work, we drill deeply into a specific feature of the typical arguments for quantum advantage through these random sampling schemes. Specifically, we study the property of anticoncentration in Gaussian Boson Sampling. Anticoncentration plays a crucial role in the current arguments that random sampling schemes are not just hard in the exact or worst case, but also hard in the approximate and average case.[7] That is, anticoncentration helps to show that it is hard to sample from a distribution that is close (in some suitable sense) to the exact measurement distribution induced at the end of the above-described protocol [approximate-case], and that this is true not only for a single possible random unitary, but for sufficiently many of them [average-case]. Roughly speaking, anticoncentration ensures that sufficiently many instances of a random sampling experiment have many output probabilities that nontrivially contribute to the distribution (which, intuitively, makes it harder to mimic the sampling procedure on a classical device). In its simplest formulation, anticoncentration can be thought of as a property of the *moments* of the output probability distribution. How precisely anticoncentration arises is a result

[7] Note here that "hard" means that the task admits no polynomial-time algorithm. A task can be hard quantumly (meaning it is not in BQP) or classically (meaning it is not in P or BPP). By contrast, a task being "easy" means it does fall into the respective quantum or classical polynomial-time class.

of a long, complicated argument connecting the hardness of sampling with the hardness of computing/approximating output probabilities through an algorithm of Stockemyer [34] and a host of other computational complexity related results. While a full review of this reasoning is beyond the scope of this dissertation, Ref. [28] contains an excellent overview.

In Chapter 4 and Appendix C, we introduce a graph-theoretic framework to analyze the moments of the output probabilities in Gaussian Boson Sampling, hence probing anticoncentration in this scheme. Specifically, we convert an algebraic problem about computing combinatorial functions of random matrices into a problem about the number of connected components of suitably defined graphs. We are able to use this graph-theoretic mindset to derive a closed form for the first moment of the output probabilities, and we are also able to derive certain properties of the second moment; we show that it admits a polynomial expansion in k, the number of modes that are squeezed in the initial state, and we also compute the leading-order term in this expansion. These two results are sufficient to prove that there is a transition in anticoncentration (as defined by the moments of the output distribution). That is, we prove that there is a transition in the strength of evidence for quantum advantage in Gaussian Boson Sampling.

In Chapter 5 and Appendix D, we significantly expand upon this graph-theoretic framework to better understand the second moment, about which the previous Chapter only provides small amounts of information. In particular, through a much more thorough analysis of the graphs originally defined in the Chapter 4, we set up a recursion relation to calculate the connected components of these graphs and efficiently evaluate numerically exactly the second moment. Using the results of this recursion, we are able to numerically investigate various properties of the second moment and how it compares to the first moment. This allows us to more precisely pin down where the transition in anticoncentration, and, hence, the transition in evidence for quantum advantage, actually occurs.

Chapter 6 and Appendix E change focus to a different simulation task, namely, trying to mimic the evolution of product states in a phase of matter called Many-Body Localization (MBL). In the MBL phase, the Hamiltonian is characterized by the emergence of an extensive number of so-called quasilocal integrals of motion (LIOMs) that commute with the Hamiltonian. In contrast to, say, Anderson-Localized (AL) systems [35], these LIOMs have exponentially decaying interactions that can slowly spread information and entanglement through the system. While MBL is typically described in the context of local disorder (creating the possibility for interference patterns that induce these LIOMs), we take a much more phenomenological approach, and use the work in Ref. [36] to simply define MBL Hamiltonians to be those that can be diagonalized into a quasilocal, commuting Hamiltonian by a quasilocal unitary (these terms are all defined more precisely in Chapter 6).

With this definition in hand, we can prove three main results about such Hamiltonians (given some technical assumptions requiring the localization length to be sufficiently small): (1) When the evolution time t under an MBL Hamiltonian is logarithmic in the size of the system N (and the Hamiltonian is short-range in its original basis before diagonalization by the quasilocal unitary), there is a classically efficient algorithm for strong simulation of the evolved state. When the evolution time t is polynomial (or even quasipolynomial) in N, we prove that there is a quasipolynomial-time classical algorithm that can perform strong simulation of the evolved state. Here, strong simulation refers to the task of estimating (with small error) the marginal and conditional probabilities of the distribution induced by measuring the output state in the computational basis; (2) When the evolution time t is exponentially large in N, we use a construction in the literature from Ref. [37] to show that weak simulation of, or sampling from, the evolved state becomes formally hard in the worst case (i.e., for a particular family of MBL Hamiltonians). This shows that there is a transition in the worst-case complexity of simulating MBL Hamiltonians; (3) Finally, we switch focus from classical simulation to quantum simulation, and we show that MBL Hamiltonians have approximate quantum circuit complexity that scales sublinearly in the evolution time, placing them in contrast to general chaotic Hamiltonians, where there is good evidence that the approximate circuit complexity grows linearly in t (until saturation)—see, e.g., Ref. [38] and more citations within Chapter 6.

These results, then, help us understand where we might find quantum advantage when it comes to the task of simulating condensed matter systems. In particular, it seems that MBL Hamiltonians might, at least for short evolution times, be classically simulable. Therefore, while understanding the properties of these systems is quite interesting, it might not be the best place to look for quantum advantage. This serves as an interesting counterpart to the previous two Chapters.

Therefore, this dissertation should be viewed as a targeted, mathematically rigorous investigation into some of the details behind quantum advantage in different schemes. While there is still much work to be done, the results in this dissertation move us closer to a full understanding of the power of quantum devices. A discussion of how to further build upon the work in this dissertation is included at the end of each Chapter and in Chapter 7.

Chapter 2: Minimum-Entanglement Protocols for Function Estimation

Abstract: We derive a family of optimal protocols, in the sense of saturating the quantum Cramér-Rao bound, for measuring a linear combination of d field amplitudes with quantum sensor networks, a key subprotocol of general quantum sensor network applications. We demonstrate how to select different protocols from this family under various constraints. Focusing primarily on entanglement-based constraints, we prove the surprising result that highly entangled states are not necessary to achieve optimality in many cases. Specifically, we prove necessary and sufficient conditions for the existence of optimal protocols using at most k-partite entanglement. We prove that the protocols which satisfy these conditions use the minimum amount of entanglement possible, even when given access to arbitrary controls and ancilla. Our protocols require some amount of time-dependent control, and we show that a related class of time-independent protocols fail to achieve optimal scaling for generic functions.

2.1 Introduction

Entanglement is a hallmark of quantum theory and plays an essential role in many quantum technologies. Consider single-parameter metrology, where one seeks to determine an unknown phase θ that is independently and identically coupled to d sensors via a linear Hamiltonian $\hat{H}$. Given a probe state $\hat{\rho}$, evolution under $\hat{H}$ encodes θ into $\hat{\rho}$ where it can then be measured. If the sensors are classically correlated the ultimate attainable uncertainty is the so-called standard quantum limit $\Delta\theta \sim 1/\sqrt{d}$ [39], which can be surpassed only if the states are prepared in an entangled state [23, 24]; if $O(d)$-partite entanglement is used, the Heisenberg limit $\Delta\theta \sim 1/d$ can be achieved [40–42]. The necessity of entanglement for optimal measurement has also been explored in numerous other contexts [43, 44]; for instance, in sequential measurement schemes (where one may apply the encoding unitary multiple times) [45, 46], in the presence of decoherence [47–50], when the coupling Hamiltonian is non-linear [51–53], or in reference to resource theories for metrology [54–57].

In this Chapter, we consider the amount of entanglement required to saturate the quantum Cramér-Rao bound, which lower bounds the variance of measuring an unknown quantity [58–61], in the prototypical multiparameter setting of a quantum sensor network, where d independent, unknown parameters $\boldsymbol{\theta}$ (boldface denotes vectors) are each coupled to a unique quantum sensor. Specifically, we revisit the problem of optimally measuring a single linear function $q(\boldsymbol{\theta})$ [13–22], which is a crucial element of optimal protocols for more general quantum sensor network problems (the case of measuring one or multiple analytic functions [62, 63] and the case where the parameters $\boldsymbol{\theta}$ are not independent [64] reduce asymptotically to the linear problem considered here). Therefore, we focus on measuring a single linear function of independent parameters for ease of presentation while emphasizing that our results generalize.

Given the similarity of measuring a single linear function to the single-parameter case and the fact that such functions of local parameters are global properties of the system, one might expect (provided all the local parameters non-trivially appear in q) that d-partite entanglement is necessary. This intuition is reinforced by the fact that all existing optimal protocols for this problem do, in fact, make use of d-partite entanglement [13, 14, 19].

We show that such intuition is faulty and only holds in the case where q is approximately an average of the unknown parameters. In particular, we derive a family of protocols that saturate necessary and sufficient algebraic conditions to achieve optimal performance in this setting, and we prove necessary and sufficient conditions on q for the existence of optimal protocols using at most $(k < d)$-partite entanglement. The more uniformly distributed q is amongst the unknown parameters, the more entanglement is required. We also consider other resources of interest, such as the average entanglement used over the course of the protocol, as well as the number of entangling gates needed to perform these protocols, and discuss optimizing them within our scheme.

Finally, we address the impracticality of certain assumptions that have typically been made in the more theoretically-focused literature on function estimation protocols. Specifically, we show that so-called probabilistic protocols fail to achieve the Heisenberg limit except for a narrow class of functions.

2.2 Problem Setup

We first briefly review the problem of measuring a linear function of unknown parameters in a quantum sensor network [13, 14, 16–19]. Consider a network of d qubit quantum sensors coupled to d independent, unknown parameters $\boldsymbol{\theta} \in \mathbb{R}^d$ via a Hamiltonian of the form

$$\hat{H}(s) = \sum_{i=1}^{d} \frac{1}{2}\theta_i \hat{\sigma}_i^z + \hat{H}_c(s), \tag{2.1}$$

where $\hat{\sigma}_i^{x,y,z}$ are the Pauli operators acting on qubit i and $\hat{H}_c(s)$ for $s \in [0,t]$ is any choice of time-dependent, $\boldsymbol{\theta}$-independent control Hamiltonian, potentially including coupling to an arbitrary number of ancilla. That is, $\hat{H}_c(s)$ accounts for any possible parameter-independent contributions to the Hamiltonian, including those acting on any extended Hilbert space with a (finite) dimension larger than that of the network of d qubit sensors directly coupled to the unknown parameters.[1] We encode the parameters $\boldsymbol{\theta}$ into a quantum state $\hat{\rho}$ via the unitary evolution generated by a Hamiltonian of this form for a time t. Given some choices of initial probe state, control $\hat{H}_c(s)$, final measurement, and classical post-processing, we seek to construct an estimator for a linear combination $q(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ of the unknown parameters, where $\boldsymbol{\alpha} \in \mathbb{R}^d$ is a set of known coefficients. Throughout this Chapter, we assume without loss of generality that $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Ref. [14] established that the fundamental limit for the mean square error $\mathcal{M}$ of an estimator for q is

$$\mathcal{M} \geq \frac{\|\boldsymbol{\alpha}\|_\infty^2}{t^2}, \tag{2.2}$$

[1] Thus, the Hilbert space under consideration is a $(d + n_a)$-qubit Hilbert space of dimension 2^{d+n_a}, where n_a is the number of ancilla.

where t is the total evolution time.

Equation (2.2) is derived via the single-parameter quantum Cramér-Rao bound [51,58–61]. This is somewhat surprising: while we seek to measure only a single quantity $q(\boldsymbol{\theta})$, d parameters control the evolution under Eq. (2.1), so we do not *a priori* satisfy the conditions for the use of the single-parameter quantum Cramér-Rao bound. However, we can justify its validity for our system: consider an infinite set of imaginary scenarios, each corresponding to a choice of artificially fixing $d-1$ degrees of freedom and leaving only $q(\boldsymbol{\theta})$ free to vary. Under any such choice, our final quantum state depends on a single parameter q, and we can apply the single-parameter quantum Cramér-Rao bound. While this requires giving ourselves information that we do not have, additional information can only reduce $\mathcal{M}$, and, therefore, any such choice provides a lower bound on $\mathcal{M}$ when we do not have such information. To obtain the tightest possible bound there must be some choice(s) of artificially fixing $d-1$ degrees of freedom that gives us no (useful) information about $q(\boldsymbol{\theta})$. We will derive algebraic conditions that characterize such choices.

Thus, we may apply the single-parameter quantum Cramér-Rao bound

$$\mathcal{M} \geq \frac{1}{\mathcal{F}(q)} \geq \frac{1}{t^2 \|\hat{g}_q\|_s^2}, \tag{2.3}$$

where $\mathcal{F}$ is the quantum Fisher information, $\hat{g}_q = \partial \hat{H}/\partial q$ (the partial derivative fixes the other $d-1$ degrees of freedom), and the seminorm $\|\hat{g}_q\|_s$ is the difference of the largest and smallest eigenvalues of $\hat{g}_q$ [51]. For our problem, the best choice of fixing extra degrees of freedom—in the sense of yielding the tightest bound via Eq. (2.3)—gives $\|\hat{g}_q\|_s^2 = 1/\|\boldsymbol{\alpha}\|_\infty^2$, yielding Eq. (2.2) [14]. The proof of this fact is provided in Appendix A.6 for completeness.

2.3 Conditions for Saturable Bounds

While the argument above justifies applying the single-parameter bound in our multiparameter scenario, it offers no roadmap for constructing optimal protocols. The quantum Fisher information matrix $\mathcal{F}(\boldsymbol{\theta})$ provides an information-theoretic solution to this issue. When calculating $\mathcal{F}(\boldsymbol{\theta})$ we restrict to pure probe states, as the convexity of the quantum Fisher information matrix implies mixed states fail to produce optimal protocols [65, 66]. For pure probe states and unitary evolution for time t under the Hamiltonian in Eq. (2.1), it has matrix elements [66]

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4\left[\frac{1}{2}\langle\{\hat{\mathcal{H}}_i(t), \hat{\mathcal{H}}_j(t)\}\rangle - \langle\hat{\mathcal{H}}_i(t)\rangle\langle\hat{\mathcal{H}}_j(t)\rangle\right], \tag{2.4}$$

where $\{\cdot,\cdot\}$ denotes the anti-commutator and

$$\hat{\mathcal{H}}_i(t) = -\int_0^t ds \hat{U}^\dagger(s)\hat{g}_i\hat{U}(s), \tag{2.5}$$

with $\hat{g}_i = \partial\hat{H}/\partial\theta_i = \hat{\sigma}_j^z/2$ and $\hat{U}$ the time-ordered exponential of $\hat{H}$. The expectation values in Eq. (2.4) are taken with respect to the initial probe state.

Choosing $d-1$ degrees of freedom to fix in hopes of using the single-parameter bound then corresponds to a basis transformation $\boldsymbol{\theta} \to \boldsymbol{q}$, where we take $q_1 = q$ to be our quantity of interest, and the other arbitrary $q_{j>1}$ are the extra degrees of freedom. This basis transformation has a corresponding Jacobian J such that $\mathcal{F}(\boldsymbol{q}) = J^\intercal\mathcal{F}(\boldsymbol{\theta})J$. To obtain the bound in Eq. (2.2) and have no information about $q(\boldsymbol{\theta})$ from the extra degrees of freedom $q_{j>1}$, $\mathcal{F}(\boldsymbol{q})$ must have the following properties:

$$\mathcal{F}(\boldsymbol{q})_{11} = \frac{t^2}{\alpha_1^2}, \tag{2.6}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = 0 \quad (\forall\, i \neq 1) \tag{2.7}$$

(recall $|\alpha_1| = \|\boldsymbol{\alpha}\|_\infty$ without loss of generality). Via the inverse basis transformation $\boldsymbol{q} \rightarrow \boldsymbol{\theta}$, we find Eqs. (2.6)-(2.7) are satisfied if and only if

$$\mathcal{F}(\boldsymbol{\theta})_{1j} = \mathcal{F}(\boldsymbol{\theta})_{j1} = \frac{\alpha_j}{\alpha_1} t^2, \tag{2.8}$$

where we assume here and for the rest of the Chapter that $|\alpha_1| > |\alpha_j|\ \forall j > 1$ for ease of presentation. Our main result (see Theorem 2.1) is unchanged by this assumption, although its proof and that of several other results becomes more tedious. The explicit derivation of Eq. (2.8), along with the generalization of our results beyond this assumption, is provided in Appendix A.6.

Finally, we remark that the problem of function estimation is mathematically equivalent to the concept of nuisance parameters in the literature on classical (c.f. [67]) and quantum estimation theory [68–70]. One finds similarly derived bounds in these contexts.[2] However, the protocols we now describe, and especially their entanglement features, are new to this Chapter.

2.4 A Family of Optimal Protocols

We now derive a family of protocols that achieve Eq. (2.8). A particular protocol consists of preparing a pure initial state $\hat{\rho}_0 = |\psi(0)\rangle\langle\psi(0)|$, evolving $\hat{\rho}_0$ under the unitary generated by $\hat{H}(s)$

[2] For instance, the conditions in Eqs. (2.6)-(2.7) are equivalent to the so-called global parameter orthogonality condition discussed in Section 5.5 of Ref. [70].

for time t, performing some positive operator-valued measurement, and computing an estimator for q from the measurement outcomes. Given $\hat{\rho}_0$ and $\hat{H}(s)$, $\mathcal{F}(\boldsymbol{\theta})$ can be computed via Eq. (2.4).

The protocols we propose will use $\hat{H}_c(s)$ to coherently switch between probe states with different sensitivities to the unknown parameters $\boldsymbol{\theta}$, thereby accumulating an overall sensitivity to the unknown function of interest q. In particular, we consider the following set $\mathcal{T}$ of $N = 3^{d-1}$ one-parameter families of cat-like states:

$$|\psi(\boldsymbol{\tau};\varphi)\rangle = \frac{1}{\sqrt{2}}\left(|\boldsymbol{\tau}\rangle + e^{i\varphi}\,|-\boldsymbol{\tau}\rangle\right), \tag{2.9}$$

where each family of states is labeled by a vector $\boldsymbol{\tau} \in \{0, \pm 1\}^d$ such that

$$|\boldsymbol{\tau}\rangle = \bigotimes_{j=1}^{d} \begin{cases} |0\rangle\,, & \tau_j \neq -1 \\ \\ |1\rangle\,, & \tau_j = -1 \end{cases}, \tag{2.10}$$

and $\varphi \in \mathbb{R}$ parameterizes individual states in the family. We require that $\tau_1 = 1$, as any optimal protocol must always be sensitive to this most important parameter; see Lemma A.1 in Appendix A.1. Each of the probe states described in Eqs. (2.9) and (2.10) is a superposition of exactly two states in the $\hat{\sigma}^z$ basis (which we call "branches"). Note that these states use no ancilla.

Our protocols proceed in three main stages: a state initialization stage, a parameter encoding stage, and, finally, a measurement stage. In the state initialization stage, we prepare the probe state $|\psi(\boldsymbol{\tau};0)\rangle$ that is then coupled to the parameters in the parameter encoding stage via a Hamiltonian of the form of Eq. (2.1). During this parameter encoding stage, we use the control Hamiltonian to coherently switch between families of probe states at particular (optimized) times, such that the relative phase between the branches is preserved during the switches (that is, $\hat{H}_c(s)$ changes $\boldsymbol{\tau}$, but not φ). This can be done using finitely many controlled-NOT (CNOT) and $\hat{\sigma}^x$ gates. We stay in the family of states $|\psi(\boldsymbol{\tau}^{(n)};\varphi)\rangle$ for time $p_n t$, where $p_n \in [0,1]$ such that $\sum_n p_n = 1$. Here n indexes some enumeration of the families of states in $\mathcal{T}$. There are three possibilities for the relative phase that qubit j induces between the two branches due to the time spent in family n. If $\tau_j^{(n)} = 0$, then no relative phase is accrued because qubit j is disentangled. If $\tau_j^{(n)} = 1$, the relative phase imprinted by $\hat{\sigma}_j^z/2$ is $p_n\theta_j t$, while if $\tau_j^{(n)} = -1$, the relative phase is $-p_n\theta_j t$. Thus, the j-th qubit always induces a relative phase of $p_n\tau_j^{(n)}\theta_j t$. Accounting for all qubits, being in family n for time $p_n t$ induces a relative phase

$$\phi_n = \sum_j p_n t\tau_j^{(n)}\theta_j. \tag{2.11}$$

Given some time-dependent probe $|\psi(t)\rangle$ which is in each family $|\psi(\boldsymbol{\tau}^{(n)};\varphi)\rangle$ for time $p_n t$, the total phase ϕ accumulated between the branches over the course of the entire parameter encoding stage of the protocol is

$$\phi = \sum_n \phi_n = \sum_n \sum_j p_n t\tau_j^{(n)}\theta_j = \sum_j (T\boldsymbol{p})_j\theta_j t, \tag{2.12}$$

where we implicitly defined $\boldsymbol{p} = (p_1,\ldots,p_N)^\intercal$ and the $d \times N$ matrix T with matrix elements $T_{mn} = \tau_m^{(n)}$. If $\boldsymbol{p}$ is chosen such that $T\boldsymbol{p} \propto \boldsymbol{\alpha}$ this total phase is $\propto qt$. More formally, choosing $\boldsymbol{p}$ such that

$$T\boldsymbol{p} = \frac{\boldsymbol{\alpha}}{\alpha_1} \tag{2.13}$$

achieves the saturability condition in Eq. (2.8). Algebraic details of this calculation are provided in Appendix A.2.

Any nonnegative solution (in the sense that $p_n \geq 0 \; \forall \, n$) to Eq. (2.13) specifies a valid set of states and evolution times satisfying Eq. (2.8). Because the system in Eq. (2.13) is highly underconstrained, such protocols do not necessarily use all 3^{d-1} families of states in $\mathcal{T}$. As an illustrative example, consider the solutions to Eq. (2.13) for two qubits. The available families of states are described by

$$T = \begin{pmatrix} \boldsymbol{\tau}^{(1)} & \boldsymbol{\tau}^{(2)} & \boldsymbol{\tau}^{(3)} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}. \tag{2.14}$$

By Eq. (2.13), the fraction of time spent in each family of states must satisfy

$$p_1 + p_2 + p_3 = 1, \tag{2.15}$$

$$p_1 - p_2 = \frac{\alpha_2}{\alpha_1}. \tag{2.16}$$

Solving in terms of p_1 leads to the 1-parameter family of solutions $p_2 = p_1 - \frac{\alpha_2}{\alpha_1}$ and $p_3 = 1 + \frac{\alpha_2}{\alpha_1} - 2p_1$, where $p_n \in [0,1]$ for all n. Without loss of generality, assume $\alpha_1 = 1$. Then non-negativity is achieved by

$$p_1 \in \begin{cases} \left[\alpha_2, \frac{1+\alpha_2}{2}\right] & \alpha_2 \geq 0 \\ \left[0, \frac{1+\alpha_2}{2}\right] & \alpha_2 < 0 \end{cases}. \tag{2.17}$$

There are many solutions satisfying these constraints. Of particular note, there is a two-family protocol that does not require using exclusively maximally entangled states: for $\alpha_2 > 0$, let $p_1 = \alpha_2$ so that $p_2 = 0$ and $p_3 = 1 - \alpha_2$; for $\alpha_2 < 0$, let $p_1 = 0$ so that $p_2 = -\alpha_2$ and $p_3 = 1 + \alpha_2$.

We refer to protocols achieving Eq. (2.13) (or, equivalently, Eq. (2.8)) as optimal. Note, however, that achieving these conditions is a property of the probe state(s) used and does not *a priori* guarantee the existence of measurements to extract q. Therefore, we now move on to describing the third main stage of our protocols, which is the explicit measurement scheme: apply a sequence of $\hat{\sigma}_i^x$ and CNOT gates to the final state of a protocol to transform it into $1/\sqrt{2}(|0\rangle + e^{iqt/\alpha_1}|1\rangle)(|0\ldots0\rangle)$. Then perform single qubit phase estimation to measure q.[3]

Such phase estimation is not as simple as it might appear, however. Because we are interested in how our error scales in the $t \to \infty$ limit, a naive approach loses track of which 2π interval the phase is in [74–76]. We could assume that this information is known *a priori* [14], but this is unjustified in practice as the required knowledge is of precision $\sim |\alpha_1|/t$, i.e. it is already within the Heisenberg limit. More realistically, starting with any t-independent prior knowledge of the unknown phase, we use the so-called robust phase estimation protocols from Refs. [71–73] to saturate Eq. (2.2) up to a modest constant factor. Such protocols work by optimally dividing the total time t into K stages with stage k using a time $2\nu_k t_k$ such that $2\sum_{k=1}^{K}\nu_k t_k = t$. In each stage, one encodes the parameters into the state for a time t_k and then makes a ($\hat{\sigma}^x$ or $\hat{\sigma}^y$) measurement. This is repeated $2\nu_k$ times in order to obtain an estimate of q, which in each stage becomes a more and more precise estimate. Provided the time of the final stage scales linearly with the total time, i.e., $t_K \sim t$, Heisenberg scaling in time is still achieved and we can estimate q with a mean square error achieving the bound in Eq. (2.2) up to a constant factor. For completeness, we review this measurement scheme in more detail in Appendix A.3.

[3] It is worth pointing out that it is not strictly necessary to reduce the problem to single qubit phase estimation. The reason we consider disentangling all qubits is to reduce fully to the single qubit phase estimation problem of the robust phase estimation papers in Refs. [71–73], described below. However, one could apply essentially equivalent protocols by forgoing the disentangling of the qubits and simply performing parity measurements on the final cat-like state. Such parity measurements can be carried out by simply measuring all qubits individually.

To summarize, a full optimal protocol is as follows:

1. Using any relevant experimental desiderata and optimization algorithm, find a nonnegative solution $\boldsymbol{p}$ to Eq. (2.13).

2. Restrict $\boldsymbol{p}$ to its $\overline{N}$ nonzero elements, and restrict T to the corresponding columns. If desired, reorder the elements of $\boldsymbol{p}$ and the columns of T. The $\overline{N}$ $\boldsymbol{\tau}$ corresponding to the columns of T will be the families of states used in the protocol.

3. Initialize a quantum state on the d qubits to $|0\rangle^{\otimes d}$.

4. Using CNOT and $\hat{\sigma}^x$ gates, prepare $|\psi(\boldsymbol{\tau}^{(1)}; 0)\rangle$, the first state of the protocol. Couple the state to the Hamiltonian $\hat{H}$ and remain in this family for time $p_1 t_k$, leading to state $|\psi(\boldsymbol{\tau}^{(1)}; \phi_1)\rangle$, where $\phi_1 = \sum_j p_1 t_k \tau_j^{(1)} \theta_j$. Here, t_k is the time required by the current step of the robust phase estimation protocol.

5. Using CNOT and $\hat{\sigma}^x$ gates, coherently switch to $|\psi(\boldsymbol{\tau}^{(2)}; \phi_1)\rangle$ from $|\psi(\boldsymbol{\tau}^{(1)}; \phi_1)\rangle$. Remain in this family for time $p_2 t_k$, leading to state $|\psi(\boldsymbol{\tau}^{(2)}; \phi_1 + \phi_2)\rangle$, with $\phi_2 = \sum_j p_2 t_k \tau_j^{(2)} \theta_j$.

6. Repeat this process for all states in the restricted T, staying in the family parameterized by $\boldsymbol{\tau}^{(n)}$ for time $p_n t_k$, leading to a final state $\left|\psi(\boldsymbol{\tau}^{(\overline{N})}; q t_k)\right\rangle$.[4]

7. Using CNOT and $\hat{\sigma}^x$ gates, convert this final state to $1/\sqrt{2}(|0\rangle + e^{iqt_k}|1\rangle)|0\rangle^{\otimes d-1}$.

8. Make a measurement on the first qubit of the final state (see Appendix A.3 for more details) and repeat starting from step 3. After $2\nu_k$ repetitions, move to the next stage of the robust phase estimation protocol, and use an updated evolution time t_k. After a number of stages

[4] Note added: As written, this value of φ (here and written directly as the relative phase in item 7 below) technically assumes that $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1| = 1$. This can be fixed by simply replacing q with $q/|\alpha_1|$.

K as prescribed by the robust phase estimation protocol, extract a final estimate of q with a mean square error achieving the bound in Eq. (2.2) up to a constant factor.

Having described the full details of the protocol, including the subtleties involved in subdividing the total time t into different stages in order to implement robust phase estimation, in the rest of the Chapter, for simplicity of presentation, we will simply consider the total encoding time t and act as if the parameters can be encoded into the state in one step, using evolution for this full time. This should be viewed as a notational shorthand such that t can be replaced with the relevant t_k at any given stage when implementing the full protocol.

2.5 Minimum-Entanglement Solutions

We now focus on solutions from our family of protocols that require the minimum amount of entanglement. Specifically, we prove necessary and sufficient conditions on $\boldsymbol{\alpha}$ for the existence of a protocol that uses at most k-partite entanglement. This is the primary technical result of this Chapter. We emphasize that, while the protocols in the previous Section use a particular choice of controls that does not include ancilla qubits, Theorem 2.1 applies to any protocol making use of a Hamiltonian described via Eq. (2.1).

Theorem 2.1 (Main result)**.** *Let $q(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$. Without loss of generality, let $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Let $k \in \mathbb{Z}^+$ so that*

$$k - 1 < \frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty} \leq k. \tag{2.18}$$

An optimal protocol to estimate $q(\boldsymbol{\theta})$, where the parameters $\boldsymbol{\theta}$ are encoded into the probe state via unitary evolution under the Hamiltonian in Eq. (2.1) requires at least, but no more than, k-partite entanglement.

Theorem 2.1 justifies our claim that d-partite entanglement is not necessary unless $\|\boldsymbol{\alpha}\|_1$ is large enough, i.e. in the case of measuring an average ($\alpha_i = \frac{1}{d} \; \forall \, i$). We now sketch the proof, providing full details in Appendix A.4. The proof comes in two parts. First, using k-partite entangled states from the set of cat-like states considered above, we show the existence of an optimal protocol, subject to the upper bound of Eq. (2.18). Second, we show that, subject to the conditions in the theorem statement, there exists no optimal protocol using at most $(k-1)$-partite entanglement, proving the lower bound of Eq. (2.18).

Part 1. Define $T^{(k)}$ to be the submatrix of T with all columns n such that $\sum_m |T_{mn}| > k$ are eliminated, which enforces that any protocol derived from $T^{(k)}$ uses only states that are at most k-partite entangled. Define System $A(k)$ as

$$T^{(k)}\boldsymbol{p}^{(k)} = \boldsymbol{\alpha}/\alpha_1, \tag{2.19}$$

$$\boldsymbol{p}^{(k)} \geq 0. \tag{2.20}$$

Let $\boldsymbol{\alpha}' = \boldsymbol{\alpha}/\alpha_1$ and define System $B(k)$ as

$$(T^{(k)})^\top \boldsymbol{y} \geq 0, \tag{2.21}$$

$$\langle \boldsymbol{\alpha}', \boldsymbol{y} \rangle < 0. \tag{2.22}$$

By the Farkas-Minkowski lemma [77,78], System $A(k)$ has a solution if and only if System $B(k)$ does not, so it is sufficient to show that System $B(k)$ does not have a solution if $\sum_{j>1} |\alpha_j'| \leq k-1$, where we used that $\alpha_1' = 1$. This can be shown by contradiction.

Part 2. The probe state must always be maximally sensitive to the first sensor qubit (see

Lemma A.1 in Appendix A.1), so $\mathcal{F}(\boldsymbol{\theta})_{1j}$ only accumulates in magnitude when qubit j is entangled with the first qubit (intuitively, Eq. (2.4) is similar to a connected correlator). Using this, we show that satisfying the condition in Eq. (2.8) requires $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty > k - 1$. □

Theorem 2.1 provides conditions for the existence of solutions to Eq. (2.13) with limited entanglement, but it is not constructive. To obtain an explicit protocol, simply solve the system of linear equations $T^{(k)}\boldsymbol{p} = \boldsymbol{\alpha}$.

Of course, instantaneous entanglement is not the only resource that one might want to minimize. For instance, one might also be interested in minimizing average entanglement over the entire protocol. This possibility is considered in Section 2.6. Other, more general, resource restrictions can be handled by setting up a constrained optimization problem that picks out certain solutions to the system of linear equations $T^{(k)}\boldsymbol{p} = \boldsymbol{\alpha}$ subject to a cost function $\mathcal{E}(\boldsymbol{p})$. For example, if certain pairs of sensors are easier to entangle than others, due to, for instance, their relative spatial location in the network, that could be encoded into $\mathcal{E}(\boldsymbol{p})$. More complicated optimizations could also take into consideration the ordering of the states used in the protocols. For example, because our protocols require coherently applying CNOT gates to move between different families of entangled states, and these gates may be costly or error-prone resources, one might wish to find protocols that minimize the usage of these gates. We discuss this possibility and the potential tradeoff between minimizing entanglement and CNOT gates in Section 2.7.

2.6 Average Entanglement

As mentioned above, one might also wish to minimize not just the size of the most-entangled family of states, but also the average entanglement used (given by weighting the size of each entangled family by the proportion of time that the family is used in the protocol). In this Section (with some details deferred to Appendix A.5), we show that there exists a class of optimal protocols, ones that we name "non-echoed," that minimize this average entanglement. The formal definition is as follows:

Definition 2.1 (Non-Echoed Protocols)**.** *Consider some $\boldsymbol{\alpha} \in \mathbb{R}^d$ encoding a linear function of interest. Let T be the matrix which describes our families of cat-like probe states, and let $\boldsymbol{p}$ specify a valid protocol such that $\boldsymbol{p} > 0$ and $T\boldsymbol{p} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_\infty$. We say that the protocol defined by $\boldsymbol{p}$ is "non-echoed" if $\forall i$ such that p_i is strictly greater than 0,* $\mathrm{sgn}(T_{ij}) \in \{0, \mathrm{sgn}(\alpha_j)\}$.

At any stage of a non-echoed protocol, letting the portion of the relative phase accumulated between the two branches of the probe state associated to the parameter θ_i be given by $c_i\theta_i$, two conditions must hold: (1) $|c_i| < |\alpha_i|$; (2) $\mathrm{sgn}(c_i) = \mathrm{sgn}(\alpha_i)$. More intuitively, sensitivity to each parameter is accumulated "in the correct direction" at all times, meaning one does not use any sort of spin echo to produce a sensitivity to the function of interest, hence the name "non-echoed."

We now prove two useful statements about non-echoed protocols.

Lemma 2.1 *Non-echoed protocols use minimium average entanglement.*

Proof. We start with $T\boldsymbol{p} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_\infty$. Then

$$\begin{aligned} \|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty &= \mathrm{sgn}(\boldsymbol{\alpha})^\top(T\boldsymbol{p}) \\ &= (\mathrm{sgn}(\boldsymbol{\alpha})^\top T)\boldsymbol{p} = \boldsymbol{w}^\top\boldsymbol{p}, \end{aligned} \tag{2.23}$$

where we have defined $w_j = \sum_i |T_{ij}|$ to be the sum of the absolute value of the elements of the jth column of T. That is, w_j represents how entangled the corresponding cat-like family of states is. But, then, clearly $\boldsymbol{w}^\top \boldsymbol{p}$ is the average entanglement of the entire protocol. Furthermore, the second half of the proof of Theorem 2.1, given in Appendix A.4 shows that the minimum average entanglement of any optimal protocol is given by $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty$ (see the discussion after the completion of the proof). □

The intuition behind this lemma is that if one always accumulates phase in the "correct direction," then the total amount of entanglement used over the course of the protocol must be minimized, as any extra entanglement would lead to becoming overly sensitive to some parameter, which would require some sort of echo to correct.

We further have the following theorem, which can be viewed as an extension of Theorem 2.1.

Theorem 2.2. *For any $\boldsymbol{\alpha} \in \mathbb{R}^d$, there exists an optimal non-echoed protocol with minimum instantaneous entanglement for measuring $q = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$.*

The proof of this theorem is given in Appendix A.5, and it proceeds in a very similar way to the proof of Theorem 2.1. The main difference is that one also restricts the allowed state families to be those with the correct sign so as to be non-echoed. And, analogously to how one can find a protocol with minimum entanglement, one can also obtain a solution that minimizes average entanglement by restricting T to only include columns such that $\operatorname{sgn}(T_{ij}) = \operatorname{sgn}(\alpha_i)$ for all i, j and then solving the corresponding system of linear equations.

2.7 CNOT Costs of Minimum Entanglement Protocols

We now address another resource of potential interest: how many entangling (CNOT) gates are required to perform our protocols with a focus on the minimum entanglement protocols.

We will again assume, for simplicity, that $\|\boldsymbol{\alpha}\|_\infty = \alpha_1 = 1 > |\alpha_2| \geq |\alpha_3| \geq \cdots \geq |\alpha_d|$. Furthermore, without loss of generality, we will adopt the convention that an optimal protocol specified by a $\boldsymbol{p} \geq 0$ such that $T\boldsymbol{p} = \boldsymbol{\alpha}$ begins by preparing a state in the family described by the first column of T and evolving for time $p_1 t$, and then proceeds to the appropriate state (i.e., the one with phase $p_1 t$) in the family described by the second column, then evolving for time $p_2 t$, and so on, until eventually moving to the measurement state. If $p_i = 0$, the corresponding state family is skipped and not prepared. By construction, the number of CNOT gates needed to perform this protocol is the number of gates required to generate the first state, plus the number needed to convert from the first state to the second state, and so on. Finally, one should add the number of gates needed to prepare the measurement state, which disentangles all qubits, from the final probe state.[5] The number of gates required to move from state i to state $i+1$ corresponds to the number of elements of $\boldsymbol{\tau}_i$ that are ± 1 but 0 in $\boldsymbol{\tau}_{i+1}$ and vice versa. In what follows, we will often consider only the gates that are used to convert between probe states (i.e., we will not consider the initial state preparation or final measurement preparation). This is physically motivated by the fact that these intermediate gates may be more difficult to perform or may be more susceptible to noise. Furthermore, assuming one is interested in the value of q at some particular moment (and not, say, continuously), one might be free to prepare and purify the initial probe state in advance of the actual sensing task, which also justifies ignoring the initial CNOT cost.

Assume that $\overline{N}$ states used in the protocol, i.e. $\boldsymbol{p}$ is such that it contains at most $\overline{N}$ nonzero elements. It is clear that at most $\mathcal{O}(\overline{N}^2)$ CNOT gates are needed. However, this is not necessarily optimal. In fact, Ref. [14] provides a protocol that uses d states and only $(d-1) = \mathcal{O}(d)$ intermediate CNOT gates. This "disentangling protocol" consists of using a maximally entangled

[5] These gates are not strictly necessary. See footnote 3.

Greenberger-Horne-Zeilinger state (up to $\hat{\sigma}^x$ rotations) for a time $|\alpha_d|t$, then disentangling the last qubit and using the $(d-1)$-entangled state for time $(|\alpha_{d-1}|-|\alpha_d|)t$ before disentangling the next-to-last qubit and so on until reaching the final state corresponding to $\boldsymbol{\tau} = (1,0,\dots,0)^\intercal$. This final state is used for time $(|\alpha_1|-|\alpha_2|)t = (1-|\alpha_2|)t$. The disentangling protocol does not minimize the instantaneous entanglement, but it does minimize average entanglement (as it is a non-echoed protocol—see Section 2.6).

Even more interestingly, Ref. [14] also provides a protocol, which we refer to as the "echoing" protocol, that uses *zero* intermediate CNOT gates. It proceeds by using d exclusively maximally entangled states (thereby minimizing neither average nor, in most cases, instantaneous entanglement), but judiciously echoing away the extra sensitivity that this extra entanglement induces.

To illustrate these protocols in the language of the current Chapter, we provide T and $\boldsymbol{p}$ (where, for simplicity of notation, we restrict T and $\boldsymbol{p}$ to the states that are used for a non-zero fraction of time) for the case $d = 8$ and $\alpha_i > 0$:

$$T^{\text{disentangling}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \boldsymbol{p}^{\text{disentangling}} = \begin{pmatrix} \alpha_8 \\ \alpha_7 - \alpha_8 \\ \alpha_6 - \alpha_7 \\ \alpha_5 - \alpha_6 \\ \alpha_4 - \alpha_5 \\ \alpha_3 - \alpha_4 \\ \alpha_2 - \alpha_3 \\ \alpha_1 - \alpha_2 \end{pmatrix} \tag{2.24}$$

and

$$T^{\text{echoing}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}, \qquad \boldsymbol{p}^{\text{echoing}} = \begin{pmatrix} \frac{1+\alpha_8}{2} \\ \frac{\alpha_7-\alpha_8}{2} \\ \frac{\alpha_6-\alpha_7}{2} \\ \frac{\alpha_5-\alpha_6}{2} \\ \frac{\alpha_4-\alpha_5}{2} \\ \frac{\alpha_3-\alpha_4}{2} \\ \frac{\alpha_2-\alpha_3}{2} \\ \frac{\alpha_1-\alpha_2}{2} \end{pmatrix}. \tag{2.25}$$

In the case of the disentangling protocol, the number of CNOTs needed is heavily dependent on the ordering of the states. For example, consider, instead, ordering the states in the following way:

$$T^{\text{disentangling}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.26}$$

Here, the number of CNOTs required is now $(d-1)+(d-2)+\cdots+1=\Theta(d^2)$. Thus, it is not only the choice of states that affects the CNOT cost of a protocol, but also their ordering. Naively, finding an optimal set of states and their optimal ordering is a difficult problem, as if one finds a protocol using $\overline{N}$ states, there are $\overline{N}!$ orders to check.

While we were unable to find a general solution to this optimization problem, numerics allow us to provide a pragmatic analysis of the cost. To begin, we considered the naive approach of finding a random (non-echoed) minimum entanglement solution using d states for random problem instances and, then, using this solution set, we brute-force searched over all column orderings of T restricted to families of states specified by this solution to find an optimal ordering in terms of CNOT cost. This was done for $d \in [3, 10]$ sensors with twenty random instances each. Without loss of generality, the random problem instances were taken to have all positive coefficients. We observe a CNOT cost scaling $\sim d^2$, indicating that a random minimum entanglement solution, even with optimal ordering, does not have the optimal linear in d scaling. See Figure 2.1.

Consequently, more nuanced algorithms for finding a minimum entanglement solution with better CNOT costs are desirable. To this end, we considered a greedy algorithm that yields a $\Theta(d)$ CNOT cost whenever it does not fail. The algorithm works by building up the full sensitivity to one parameter before switching coherently to a new state family (in this way, it is non-echoed—see Section 2.6). Consequently, each time we switch to a new state, one sensor qubit can be disentangled and never re-entangled. In particular, we seek to build up sensitivity to the parameters according to their weight in q, i.e. we build up sensitivity to parameters going from the smallest corresponding $|\alpha_j|$ to the largest. The full algorithm is completed in at most d steps.[6]

However, this greedy algorithm can fail to produce a valid protocol, as it does not enforce the condition that $\|\boldsymbol{p}\|_1 = 1$. This condition will be violated for some functions—typically those with many coefficients with approximately equal magnitude. Still, when it works, this algorithm

[6] Code is available upon request.

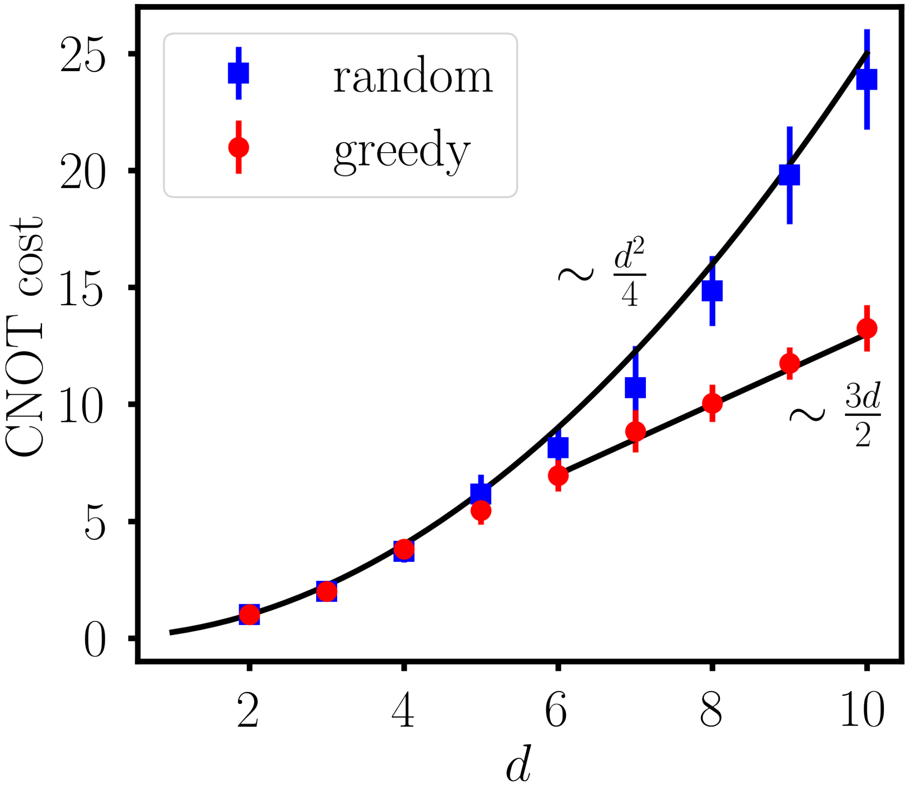

Figure 2.1: CNOT costs versus number of sensors d for minimum entanglement protocols using d optimally ordered states chosen either randomly or via the greedy algorithm described above. Twenty randomly chosen instances (that do not fail) to yield a valid protocol via the greedy algorithm. When it returns a valid protocol, the greedy algorithm recovers optimal linear scaling with d for the CNOT cost, whereas randomly chosen states have quadratic scaling, even with optimal state ordering.

succeeds in producing CNOT-efficient minimum entanglement protocols, as shown in Figure 2.1. Finding more general algorithms that always succeed for this task remains an interesting open problem.

Independent of the algorithm used to minimize the CNOT count of an optimal protocol, the takeaway message is the same: there is an apparent tradeoff between entanglement- and gate-based resources. The disentangling protocol minimizes average entanglement, but not necessarily instantaneous entanglement, and requires only $\mathcal{O}(d)$ intermediate entangling gates; the echoing protocol uses maximal entanglement, but requires only single-particle intermediate gates. Protocols that minimize instantaneous entanglement do so at the cost of more intermediate entangling gates. Depending on the primary sources of error or the physical constraints on any given quantum sensor network implementation, one of these resources might be more important to minimize than the other. In general, determining the optimal CNOT scaling for protocols that minimize instantaneous and/or average entanglement is a crucial open question for future work.

2.8 Time-Independent Protocols

Another approach to constructing protocols is to use so-called probabilistic protocols. These protocols eschew control and instead exploit the convexity of the quantum Fisher information by staying in one family throughout any given run of the protocol, but by letting this family vary over different runs. Intuitively, each family is sensitive to a different function q_n such that $q = \sum_{n=1}^{\overline{N}} p_n q_n$, where $\overline{N}$ is the number of families from $\mathcal{T}$ used in the protocol, and p_n is the frequency that family n is used. In this way, one can create an estimator for q using those for q_n. In order to generate a Fisher information matrix satisfying Eq. (2.8) [14, 19], the p_n should come from a solution to Eq. (2.13). These protocols have the advantage of requiring no control, but, unfortunately, suffer worse scaling with d than ours for generic functions when the available resources are comparable.

In particular, to fairly account for resources, we must fix a total time t to perform *all* stages of our protocol. Therefore, when considering a probabilistic protocol that uses multiple families from $\mathcal{T}$, but does not switch coherently between them, we must assign a time t_n to family n such that

$$\sum_{j=1}^{\overline{N}} t_n = t. \tag{2.27}$$

Note, we have used the fact that no stages of a probabilistic protocol with the families in $\mathcal{T}$ can be performed simultaneously. One could imagine protocols that parallelize the measurement of some q_j that involve disjoint sets of sensors. However, such protocols are necessarily non-optimal given Lemma A.1 in Appendix A.1, which says that any optimal protocol requires entanglement with the first qubit at all times.

We can bound the maximum of the Fisher information matrix element $\mathcal{F}(\boldsymbol{\theta})_{11}$ obtainable via such a probabilistic protocol as

$$\begin{aligned} \max \mathcal{F}(\boldsymbol{\theta})_{11} \leq & \max_{p_n, t_n} \sum_{n=1}^{\overline{N}} p_n t_n^2, \\ & \text{subject to: } \sum_{n=1}^{\overline{N}} t_n = t, \\ & \qquad\qquad\quad \sum_{n=1}^{\overline{N}} p_n = 1. \end{aligned} \tag{2.28}$$

where we used that $\tau_1^{(n)} = 1$ for all n. The inequality arises due to the fact that the maximization problem on the right hand side of the inequality does not enforce that $T\boldsymbol{p} = \boldsymbol{\alpha}/\alpha_1$. We could add this as an additional constraint, but it will not be necessary.

To perform the necessary optimization, consider the Lagrangian:

$$\mathcal{L} = \sum_{n=1}^{\overline{N}} p_n t_n^2 + \gamma_1 \left(t - \sum_{n=1}^{\overline{N}} t_n \right) + \gamma_2 \left(1 - \sum_{n=1}^{\overline{N}} p_n \right), \tag{2.29}$$

where γ_1, γ_2 are Lagrange multipliers. Therefore, we obtain the system of equations

$$\begin{aligned} 2p_n t_n - \gamma_1 &= 0, \quad (\forall\, n), \\ t_n^2 - \gamma_2 &= 0, \quad (\forall\, n), \\ \sum_{n=1}^{\overline{N}} t_n &= t, \\ \sum_{n=1}^{\overline{N}} p_n &= 1, \end{aligned} \tag{2.30}$$

which can be solved to yield the solution

$$\max_{p_n, t_n} \sum_{n=1}^{\overline{N}} p_n t_n^2 = \frac{t^2}{\overline{N}^2}, \tag{2.31}$$

for $p_n = 1/\overline{N}$ and $t_n = t/\overline{N}$ for all n. Therefore,

$$\mathcal{F}(\boldsymbol{\theta})_{1j} \leq \frac{t^2}{\overline{N}^2}, \quad (\forall j), \tag{2.32}$$

which clearly fails to achieve the saturability condition for $j = 1$, unless $\overline{N} = 1$, which is only possible for a very small set of functions (generic functions require $\overline{N}$ that scale nontrivially with d). Therefore, provided one considers cases where each q_n must be learned sequentially (which is a requirement for any possibly optimal protocol via Lemma A.1), we fail to achieve saturability even up to a d-independent constant for generic functions via time-independent protocols.

Note that we have, for simplicity, again restricted ourselves to the case where $\boldsymbol{\alpha}$ has a single maximal magnitude element. The more general proof follows almost identically, with some notational overhead, when generalizing beyond this condition.

2.9 Conclusion and Outlook

We have proven that maximally entangled states are not necessary for the optimal measurement of a linear function with a quantum sensor network unless the function is sufficiently uniformly supported on the unknown parameters. While the uniformly distributed case has been considered extensively in the literature, as it provides the largest possible separation in performance between entangled and separable protocols, there is no *a priori* reason why one should be interested in only these sorts of quantities. Our results demonstrate that while the precision gains to be had are less away from the uniformly distributed regime, the required resources are also less. This result is of particular relevance to the development of near-term quantum sensor networks, where creating large-scale entangled states may not be practical. Furthermore, while algebraic approaches like the one we consider here have been used before to generate bounds for the function estimation problem [14, 64], leveraging this approach to derive protocols that achieve these bounds subject to various experimental constraints is a new and widely applicable technique. We emphasize again that these results are also useful in more general settings, such as the measurement of analytic functions, as these measurements reduce to the case studied here [62–64].

To the best of the authors' knowledge, all information-theoretically optimal protocols for the estimation of a single linear function that are currently in the literature are subsumed by the framework that we develop in this Chapter. What protocol one chooses to use will depend heavily on the experimental context; if decoherence is more problematic than the number of entangling gates that one must perform, then minimum entanglement protocols will be preferred to the conventional protocols. However, if decoherence is mild, but two-qubit gates introduce significant errors, then a protocol such as the echoing protocol presented in Ref. [14] will be preferred. Consequently, the extent to which minimum entanglement protocols are more or less valuable than their more highly entangled counterparts depends on the details of the physical implementation of a quantum sensor network. Either way, the development of a framework to address these questions is, in and of itself, an important contribution of this Chapter.

We also briefly point out one more resource-related constraint of protocols that rely on time-dependent control (whether in the form of $\hat{\sigma}^x$ gates, CNOT gates, or others): these protocols require precise timing of the gate applications. Uncertainty in the timing leads directly to a systematic error in the function being measured. Importantly, however, this timing issue is a limitation of all known optimal protocols for the linear function estimation task (see e.g. Ref. [14]). We therefore view these limitations as more pertinent to experimental implementation than the theory of resource tradeoffs that we are considering here.

So far, we have not discussed the situation where we are constrained to k-partite entanglement, but k is not sufficient to achieve optimality (for any protocol) via Theorem 2.1. We propose the following protocol for such a scenario: Let R be a partition of the sensors into independent sets where we do not allow entanglement between sets and allow, at most, k-partite entanglement within each $r \in R$. Let $\boldsymbol{\alpha}^{(r)}$ denote $\boldsymbol{\alpha}$ restricted to r. Pick the optimal R such that the condition of Theorem 2.1 is satisfied for all r; that is, we ensure that *within* each independent set we obtain the optimal variance for the linear function restricted to that set. The result is a variance

$$\mathcal{M} = \frac{1}{t^2} \sum_{r \in R} \left\| \boldsymbol{\alpha}^{(r)} \right\|_\infty^2 . \tag{2.33}$$

The optimal R is a partition of the sensors into contiguous sets (assuming for simplicity that

$|\alpha_i| \geq |\alpha_j|$ for $i < j$) such that for all $r \in R$, $\sum_{i \in r} |\alpha_i| / \max_{i \in r} |\alpha_i| \leq k$, satisfying Theorem 2.1. We conjecture that this protocol is optimal, and it is clearly so if partitioning the problem into independent sets is optimal. However, one could imagine protocols that use different partitions for some fraction of the time. Intuitively, this should not improve the performance, but we leave analyzing this as an open question.

Finally, no optimal time-independent protocols for arbitrary linear functions exist in the literature. Finding such protocols (or proving their non-existence) remains an open problem of interest.

Acknowledgments

We thank Michael Foss-Feig, Zachary Eldredge, Tarushii Goel, Pradeep Niroula, Luis Pedro García-Pintos, and Michael Gullans for helpful discussions. We thank the anonymous referee who pointed out that the CNOT cost of our minimum entanglement protocols deserved a more detailed analysis. This work supported in part by AFOSR MURI, AFOSR, DARPA SAVaNT ADVENT, NSF PFCQC program, ARO MURI, DoE ASCR Accelerated Research in Quantum Computing program (award No. DE-SC0020312), NSF QLCI (award No. OMA-2120757), U.S. Department of Energy Award No. DE-SC0019449, and the DoE ASCR Quantum Testbed Pathfinder program (award No. DE-SC00119040). Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. This research was supported in part by the Heising-Simons Foundation, the Simons Foundation, and National Science Foundation Grant No. NSF PHY-1748958. J.B. acknowledges support by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship (award No. DE-SC0019323).

Chapter 3: Optimal Function Estimation with Photonic Quantum Sensor Networks

Abstract: The problem of optimally measuring an analytic function of unknown local parameters each linearly coupled to a qubit sensor is well understood, with applications ranging from field interpolation to noise characterization. Here, we resolve a number of open questions that arise when extending this framework to Mach-Zehnder interferometers and quadrature displacement sensing. In particular, we derive lower bounds on the achievable mean square error in estimating a linear function of either local phase shifts or quadrature displacements. In the case of local phase shifts, these results prove, and somewhat generalize, a conjecture by Proctor *et al.* [arXiv:1702.04271 (2017)]. For quadrature displacements, we extend proofs of lower bounds to the case of arbitrary linear functions. We provide optimal protocols achieving these bounds up to small (multiplicative) constants and describe an algebraic approach to deriving new optimal protocols, possibly subject to additional constraints. Using this approach, we prove necessary conditions for the amount of entanglement needed for any optimal protocol for both local phase and displacement sensing.

3.1 Introduction

In quantum metrology, entangled states of quantum sensors are used to try to obtain a performance advantage in estimating an unknown parameter or parameters (e.g., field amplitudes) coupled to the sensors. In addition to this practical advantage of quantum sensing, the theory of the ultimate performance limits for parameter estimation tasks is deeply related to a number of topics of theoretical interest in quantum information science, such as resource theories [79], the geometry of quantum state space [59], quantum speed limits [80–82], and quantum control theory [81].

Initial experimental and theoretical work on quantum sensing focused on optimizing the estimation of a single unknown parameter (see, e.g., Ref. [83] for a review). More recently, the problem of distributed quantum sensing has become an area of particular interest [25]. Here, one considers a network of quantum sensors, each coupled to a local unknown parameter. The prototypical task in this setting is to measure some function or functions of these parameters. In this context, the task of optimally measuring a single linear function $q(\boldsymbol{\theta})$ of d independent local parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$ is particularly well studied both theoretically [1, 13–16, 18–22, 26, 84–86] and experimentally [87–90].[1] In addition to its independent utility (i.e., for measuring an average of local fields in some region), linear function estimation serves as a key subtask of more general metrological tasks, such as measuring an analytic function of the unknown parameters [62], measuring an analytic function of dependent parameters [64, 91], or measuring multiple functions [63, 92].

For qubit sensors, the asymptotic limits on performance for these function estimation tasks

[1] Note added: Reference [1] refers to the published version of Chapter 2.

are rigorously understood, and techniques for generating optimal protocols subject to various constraints, such as limited entanglement between sensors, are known [1]. However, despite extensive theoretical and experimental research on distributed quantum sensing for photonic quantum sensors (see, e.g., [25, 93] for reviews), the asymptotic performance limits for function estimation are not yet rigorously established. Here, we close this gap, proving an ultimate bound on asymptotic performance, as measured by the mean square error of the estimator, for measuring a linear function of unknown parameters each coupled to a different photonic mode via either (1) the number operator $\hat{n}$ or (2) a field-quadrature operator, chosen without loss of generality to be the momentum quadrature $\hat{p} \coloneqq i(\hat{a}^\dagger - \hat{a})/2$. That is, we are interested in determining a function of either unknown local phase shifts or unknown quadrature displacements. For case (1), our primary focus, we derive this bound subject to a strict constraint on photon number, proving a long-standing conjecture appearing in Ref. [13]. In case (2), we derive our bound subject to a constraint on the average photon number, which is more natural in this setting as quadrature displacements are not photon-number conserving. Here, our results are consistent with existing bounds in the literature [26], but, for completeness, we include derivations in this setting using an equivalent mathematical framework to the number operator case and the qubit sensor case [1]. This allows for a natural comparison of the various performance limits and resource requirements of function estimation in quantum sensor networks and opens the door to designing new, information-theoretically optimal protocols in the asymptotic limit of sufficient data.

The rest of this Chapter proceeds as follows. In Section 3.2, we formally set up the problem of interest and provide useful notation. In Section 3.3 we prove lower bounds on the mean-squared error of an estimator for arbitrary linear functions for both number operator and displacement operator generators. We then study protocols that saturate these bounds in Section 3.4.

Finally, we discuss other entanglement-restricted optimal protocols in Section 3.5.

3.2 Problem Setup

Consider a sensor network of d optical modes each coupled to an unknown parameter θ_j for $j \in \{1, \dots, d\}$ via

$$\hat{H}(s) = \sum_{j=1}^{d} \theta_j \hat{g}_j + \hat{H}_c(s) =: \boldsymbol{\theta} \cdot \hat{\boldsymbol{g}} + \hat{H}_c(s), \tag{3.1}$$

where $\hat{g}_j$ is the local coupling Hamiltonian and boldface denotes vectors. Here, we consider the following two cases:

$$\hat{g}_j := \hat{n}_j = \hat{a}_j^\dagger \hat{a}_j, \tag{3.2a}$$

$$\hat{g}_j := \hat{p}_j = \frac{i}{2}(\hat{a}_j^\dagger - \hat{a}_j), \tag{3.2b}$$

where $\hat{a}_j^\dagger, \hat{a}_j$ are the bosonic creation and annihilation operators acting on mode j, $\hat{n}_j$ is the number operator acting on mode j, and $\hat{p}_j$ is the momentum- ($\hat{p}$-) quadrature on mode j. The choice of $\hat{p}$ quadrature is, of course, arbitrary. All results apply equally well for coupling to any quadrature. The $\boldsymbol{\theta}$-independent, time-dependent Hamiltonian $\hat{H}_c(s)$ is a control Hamiltonian, possibly including coupling to an arbitrary number of ancilla modes. Here, $s \in [0, t]$, where t is the total sensing time.

In either case, our task is to measure a linear function $q(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ of the local field amplitudes $\boldsymbol{\theta}$ where $\boldsymbol{\alpha} \in \mathbb{Q}^d$ is a vector of rational coefficients. (The restriction to rational coefficients is due to the discreteness of the resources—the number of photons—available in this problem; in the case we are interested in—large photon numbers—this is only a technical point.) To accomplish this task, we consider probe states with either fixed photon number N or fixed average photon number $\overline{N}$. Given such probe states, we consider encoding the unknown parameters into the state via the unitary evolution generated by the Hamiltonian in Eq. (3.1).

We will consider both an unrestricted control Hamiltonian and a control Hamiltonian fixed to have the form

$$\hat{H}_c(s) = \hat{h}_c(s)\delta(s - j\Delta t), \tag{3.3}$$

where $\hat{h}_c(s)$ is a (unitless) Hermitian operator, $\delta(s)$ is the Dirac delta function, $\Delta t := t/M$ is the time for a single application of the encoding unitary $\exp(-iH\Delta t)$. The index $j \in \{1, \ldots, M\}$ indexes these applications, where M is the total number of applications. This construction is motivated by the fact that typical physical implementations of a number operator coupling, e.g., in a Mach-Zehnder interferometer, and displacement operator coupling, e.g., via an electro-optical modulator (EOM), often do not allow for intermediate controls at arbitrary times. Therefore, when we fix our control Hamiltonians to be described by Eq. (3.3), we have limited any controls to be applied between each pass through these optical elements; for simplicity, we have assumed that these control operations can be implemented on a timescale much shorter than the timescale of phase accumulation. Without loss of generality, we will let $\Delta t = 1$ for the rest of this Chapter, implying that (in this setting) $t = M$. Therefore, the parameter encoding procedure for the photon number coupling[2] is done via the unitary

$$U = U^{(M)}VU^{(M-1)}V \ldots U^{(1)}V = \prod_{m=1}^{M}(U^{(m)}V), \tag{3.4}$$

[2] Note added: This is a typo in the original paper, as the expression also holds for the displacement coupling given that V listed below uses the general generator $\hat{\boldsymbol{g}}$.

where $V := \exp(-i\hat{\boldsymbol{g}}\cdot\boldsymbol{\theta})$ and $U^{(m)}$ for $m \in \{1,\ldots,M\}$ denote the unitaries applied between passes. Here, by pass, we mean a single application of the unitary V. We use the convention that the product operation left multiplies.

In both settings, it is worth emphasizing that, while our information-theoretic results lower bounding the asymptotically achievable mean square error of an estimate $\tilde{q}$ of q will apply to any protocol within the framework(s) described above, the explicit protocols we will develop will use finite ancillary modes and finite controls.

3.3 Lower Bounds

Following the approach of Refs. [1, 14], we compute lower bounds on the mean square error $\mathcal{M}$ of an estimator $\tilde{q}$ of q by rewriting the Hamiltonian in Eq. (3.1) as

$$H(s) = \sum_{j=1}^{d}(\boldsymbol{\alpha}^{(j)}\cdot\boldsymbol{\theta})(\boldsymbol{\beta}^{(j)}\cdot\hat{\boldsymbol{g}}) + \hat{H}_c(s), \tag{3.5}$$

for some (time-independent) choice of basis vectors $\{\boldsymbol{\alpha}^{(j)}\}_{j=1}^{d}$, where $\boldsymbol{\alpha}^{(1)} := \boldsymbol{\alpha}$ and $\{\boldsymbol{\beta}^{(j)}\}_{j=1}^{d}$ is a dual basis such that $\boldsymbol{\alpha}^{(i)}\cdot\boldsymbol{\beta}^{(j)} = \delta_{ij}$. The vectors $\{\boldsymbol{\alpha}^{(j)}\}_{j=1}^{d}$ are associated with a change of basis $\boldsymbol{\theta}\to\boldsymbol{q}$ where $q_j := \boldsymbol{\alpha}^{(j)}\cdot\boldsymbol{\theta}$ such that $q_1 = q$; that is, $\boldsymbol{\alpha}^{(1)} =: \boldsymbol{\alpha}$ with corresponding dual vector $\boldsymbol{\beta}^{(1)} =: \boldsymbol{\beta}$. Then we can define a $\boldsymbol{\beta}$-parameterized generator of translations with respect to the quantity q as

$$\hat{g}_{q,\boldsymbol{\beta}} := \min_{q^{(2)},\ldots,q^{(d)}} \left.\frac{\partial \hat{H}}{\partial q}\right|_{q^{(2)},\ldots,q^{(d)}} = \boldsymbol{\beta}\cdot\hat{\boldsymbol{g}}. \tag{3.6}$$

Armed with Eq. (3.6), we can write a bound on $\mathcal{M}$ in terms of a single-parameter quantum

Cramér-Rao bound [23, 51, 93]

$$\mathcal{M} \geq \frac{1}{\mu \mathcal{F}(q|\boldsymbol{\beta})}, \tag{3.7}$$

where $\mathcal{F}(q|\boldsymbol{\beta})$ is the quantum Fisher information with respect to q, given some choice of fixing the extra $d-1$ degrees of freedom in our problem, as specified by the vector $\boldsymbol{\beta} \in \mathbb{R}^d$ such that $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1$. Any such single-parameter bound is a valid lower bound as fixing extra degrees of freedom can only give us more information about the parameter q (see below for mathematical details). μ is the number of experimental repetitions. This bound holds for an unbiased estimator $\tilde{q}$. When deriving our bounds, we will restrict ourselves to single-shot Fisher information and set $\mu = 1$.[3] Quantum Fisher information is maximized for pure states, so restricting ourselves to pure states and unitary encoding of the unknown parameters into the state we can write

$$\mathcal{F}(q|\boldsymbol{\beta}) \leq 4t^2 \max_{\rho}[(\Delta \hat{g}_{q,\boldsymbol{\beta}})_\rho]^2, \tag{3.8}$$

where $\hat{g}_{q,\boldsymbol{\beta}}$ is the $\boldsymbol{\beta}$-parameterized generator of translations with respect to the unknown function q. The variance $[\Delta(\hat{g}_{q,\boldsymbol{\beta}})_\rho]^2$ is taken with respect to a pure probe state $\rho = |\psi\rangle\langle\psi|$.

Ultimately, we seek a choice of new basis that yields the tightest possible bound on the

[3] Clearly, with $\mu = 1$, we are not guaranteed the existence of an unbiased estimator, so there is some subtlety in this restriction. The choice is sufficient for determining bounds and optimal probe states, but, when considering measurements to extract the quantity of interest, realistic protocols must use more than one shot. For instance, robust phase estimation allows for $\mu = \mathcal{O}(1)$, while still allowing us to obtain an unbiased estimator that achieves the quantum Cramér-Rao bound up to a multiplicative constant [71–73]. In Appendix B.7, for completeness, we briefly summarize this approach. See also, Refs. [68, 70] and Ref. [1] for further discussion of these issues.

quantum Fisher information $\mathcal{F}(q)$. This choice is determined by the solution to[4]

$$\min_{\boldsymbol{\beta}} \max_{\rho} [\Delta(\boldsymbol{\beta}\cdot\hat{\boldsymbol{g}})_{\rho}]^2, \quad \text{subject to } \boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 1. \tag{3.9}$$

Let $(\boldsymbol{\beta}^*, \rho^*)$ be a solution for this optimization problem. Then we can rewrite the single-shot version of Eq. (3.7) as

$$\mathcal{M} \geq \frac{1}{4t^2[\Delta(\boldsymbol{\beta}^*\cdot\hat{\boldsymbol{g}})_{\rho^*}]^2}. \tag{3.10}$$

This bound can be understood as corresponding to the optimal choice of an imaginary single parameter scenario, where we have fixed $d-1$ of the d parameters controlling the evolution of the state, leaving only the parameter of interest q free to vary. While this requires giving ourselves information that we do not have, additional information can only reduce $\mathcal{M}$, and, therefore, any such choice provides a lower bound on $\mathcal{M}$ (via single-parameter bounds) when we do not have such information. While not guaranteed by this method of derivation, we shall see that such bounds are saturable, up to small multiplicative constants.

Constraints can be placed on the probe state ρ depending on the physical generators coupled to the parameters of interest: as previously discussed, in this work we consider the constraints of fixed photon number N for the generator $\hat{n}_j$ and fixed average photon number $\overline{N}$ for the generator $\hat{p}_j$. The rationale behind these constraints is as follows. $\hat{p}$ does not conserve photon number, hence it does not make sense to restrict to a fixed photon number sector when coupling to quadrature operators, and, thus, average photon number is the natural constraint. For $\hat{n}$, on the other hand, we must work in the fixed photon sector, as using fixed average photon number

[4] Note the use of a minimax as opposed to a maximin in Eq. (3.9). This follows from the fact that the minimax of some objective function is always greater than or equal to the maximin and we seek to maximize the quantum Fisher information.

allows for the construction of pathological probe states enabling arbitrarily precise sensing. In particular, consider the state

$$|\psi_a\rangle = \sqrt{\frac{a-1}{a}}\,|0\rangle + \sqrt{\frac{1}{a}}\,\big|a\overline{N}\big\rangle\,. \tag{3.11}$$

It is easy to see that $|\psi_a\rangle$ has mean photon number $\overline{N}$ and variance $(a-1)\overline{N}^2$. Hence, even for fixed $\overline{N}$, letting a get arbitrarily large allows for an arbitrarily large variance, and hence arbitrarily precise sensing.

Leaving the details of the calculation to Appendix B.1, solving the above optimization problem for $\hat{g}_j = \hat{n}_j$ restricted to probe states with exactly N photons yields

$$\mathcal{M} \geq \frac{\max\left\{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}^2, \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}^2\right\}}{N^2 t^2}, \tag{3.12}$$

where $\mathcal{P} := \{j \,|\, \alpha_j \geq 0\}$ and $\mathcal{N} := \{j \,|\, \alpha_j < 0\}$. In the second line, we use the notation

$$\|\boldsymbol{\alpha}\|_{1,\mathcal{S}} := \sum_{i\in\mathcal{S}} |\alpha_i|, \tag{3.13}$$

where $\mathcal{S} \in \{\mathcal{P}, \mathcal{N}\}$. For the rest of this Chapter, we assume without loss of generality that we are in the case that $\|\boldsymbol{\alpha}\|_{1,\mathcal{P}} \geq \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}$ to simplify our expressions. In the special case where $\boldsymbol{\alpha}$ possesses only positive coefficients (i.e., $\mathcal{N} = \varnothing$),

$$\mathcal{M} \geq \frac{\|\boldsymbol{\alpha}\|_1^2}{N^2 t^2}, \tag{3.14}$$

proving a long-standing conjecture from Ref. [13] that this is the minimum attainable variance for $\boldsymbol{\alpha} \in \mathbb{Q}^d$ with $\boldsymbol{\alpha} \geq 0$ and $N\boldsymbol{\alpha} \in \mathbb{N}^d$. This is our primary result.

Similarly, for the case of local quadrature displacements restricted to probe states with average photon number $\overline{N}$, we obtain the following bound:

$$\mathcal{M} \geq \frac{\|\boldsymbol{\alpha}\|_2^2}{4\overline{N}t^2} - \mathcal{O}\left(\frac{d\|\boldsymbol{\alpha}\|_2^2}{\overline{N}^2 t^2}\right). \tag{3.15}$$

Equation (3.15) is a minor generalization of the results in Refs. [25, 26], extended to allow for negative coefficients and for arbitrary non-Gaussian probe states. Therefore, for completeness, we include a reminder of the arguments from Refs. [25,26] along with our more general derivation in Appendix B.2.

We can compare the bounds in Eqs. (3.12) and (3.15) to the corresponding bounds on the mean square error obtainable by separable protocols—that is, those using separable probe states such that each parameter θ_i is measured individually using an optimized partition of the available photons, and then these estimates are used to compute q. In particular, for number operator coupling and fixed photon number states, using $\eta_j = \frac{|\alpha'_j|}{\|\boldsymbol{\alpha}'\|_1}N$ photons ($\alpha'_j := \alpha_j^{2/3}$) in mode j, it holds that [13]

$$\mathcal{M}_{\text{sep}} \geq \frac{\|\boldsymbol{\alpha}'\|_{2/3}^2}{N^2 t^2}, \tag{3.16}$$

where $\|\cdot\|_{2/3}$ denotes the Schatten p-function

$$\|\boldsymbol{v}\|_p = \left(\sum_i v_i^p\right)^{1/p} \tag{3.17}$$

with $p = 2/3$. When $p \in [1, \infty]$, this function is a norm, but for $p \in (0, 1)$ it is not, as it does not satisfy the property of absolute homogeneity, but it still provides a convenient notational shorthand.

Performing a similar optimization for the case of displacement coupling and fixed average photon number, one obtains

$$\mathcal{M}_{\mathrm{sep}} \geq \frac{\|\boldsymbol{\alpha}\|_1^2}{4\overline{N}t^2} + \mathcal{O}\left(\frac{1}{\overline{N}^2 t^2}\right), \tag{3.18}$$

where the optimum division of photons is given by using $\eta_j = \frac{|\alpha_j|}{\|\boldsymbol{\alpha}\|_1} N$ photons in mode j. A non-closed-form version of this bound can be found in Ref. [84] in the case where $\overline{N}$ is finite. One recovers our result in the asymptotic in $\overline{N}$ limit.

Consequently, in both the phase and displacement sensing settings, the achievable advantage due to entanglement between modes is fully characterized by the difference between the vector p norm of $\boldsymbol{\alpha}$ with $p = \frac{2}{3}, 1$ or $p = 1, 2$, respectively. By generalized Hölder's inequality, $\|\boldsymbol{\alpha}\|_{2/3}^2 \leq d\|\boldsymbol{\alpha}\|_1^2$ and $\|\boldsymbol{\alpha}\|_1^2 \leq d\|\boldsymbol{\alpha}\|_2^2$. Both inequalities are saturated for any "average-like" function with $|\boldsymbol{\alpha}| \propto (1, 1, \ldots 1)^T$. In both cases, we obtain a $\mathcal{O}(1/d)$ improvement in precision due to entanglement, consistent with the so-called Heisenberg scaling in the number of sensors d. This is consistent with results for qubits in Ref. [14], where the best improvement between the separable and entangled bounds occurs when measuring an average-like function. For the case of phase sensing, the optimal performance, including constants, is obtained when $\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}^2 = \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}^2 = \|\boldsymbol{\alpha}\|_1/2$ (which occurs when the vector $\boldsymbol{\alpha}$ is half positive ones and half negative ones).

3.4 Protocols

3.4.1 Existing Protocols

The bounds established in the previous section are all saturable, up to small multiplicative constants, using protocols that exist in the literature, or slight variations thereof. In particular, Refs. [13,16] present a protocol for estimating a linear function of local phase shifts with positive coefficients (i.e., $\boldsymbol{\alpha} \geq 0$) which achieves the bound in Eq. (3.12) up to a small multiplicative constant. This protocol makes use of a so-called proportionally weighted N00N state over $d+1$ modes,

$$|\psi\rangle \propto \left| N\frac{\alpha_1}{\|\boldsymbol{\alpha}\|_1}, \ldots, N\frac{\alpha_d}{\|\boldsymbol{\alpha}\|_1}, 0 \right\rangle + \left| 0, \ldots, 0, N \right\rangle, \tag{3.19}$$

where we have expressed the state in an occupation number basis over $d+1$ modes and have dropped the normalization for concision. The last mode serves as a reference mode. Observe that, for this state to be well defined, it is essential that $\boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_1 \in \mathbb{Q}^d$ and that N is sufficiently large that the resulting occupation numbers are integers. Details of how protocols using this probe state work and how they generalize to the case of negative coefficients are provided in Appendix B.4. A description of how to achieve the separable bound in Eq. (3.16) is provided in Appendix B.2.[5]

Similarly, in the case of measuring a linear function of displacements using states with fixed average photon number, Ref. [84] provides a protocol that, up to small multiplicative constants,

[5] Note added: This statement is a typo. We do discuss how to derive the separable bound for the case of quadrature displacements, as it is relevant to our more general proof. However, we do not similarly review the derivation of the separable bound for the case of local phase shifts. The interested reader can consult Ref. [13] for details on this matter.

saturates the bound in Eq. (3.15), and a separable protocol that, again up to small constants, achieves the bound in Eq. (3.18). Interestingly, these protocols require only Gaussian probe states, indicating that these states are optimal. In particular, these protocols make use of an initial single-mode squeezed state, followed by a properly constructed beam-splitter array to prepare a multimode entangled probe state with the appropriate sensitivity to quadrature displacements in each mode. Homodyne measurements on each mode can then be used to extract the function of interest. Consistent with this fact, our separable lower bound matches the Gaussian state-restricted bound obtained in Ref. [84] and the bound for arbitrary states derived in Ref. [26] for the particular case of measuring an average.

3.4.2 Algebraic Conditions for New Protocols

Other protocols are possible and can be derived via a simple set of algebraic conditions. In particular, for a probe state to exist saturating the bound in Eq. (3.10), or its specific versions in Eqs. (3.12) and (3.15), we require the existence of an optimal choice of basis transformation $\boldsymbol{\theta} \to \boldsymbol{q}$ such that knowing q_j for $j > 1$ yields no information about $q = q_1$. Mathematically, this means that the quantum Fisher information matrix [66] with respect to the parameters $\boldsymbol{q}$ must have the following properties:

$$\mathcal{F}(\boldsymbol{q})_{11} = 4t^2[\Delta(\boldsymbol{\beta}^* \cdot \hat{\boldsymbol{g}})_{\rho^*}]^2, \tag{3.20a}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = 0 \quad (\forall\, i \neq 1), \tag{3.20b}$$

Recall that $(\boldsymbol{\beta}^*, \rho^*)$ are the solution to the minimax problem in Eq. (3.9). We can reexpress these conditions in terms of the quantum Fisher information matrix with respect to $\boldsymbol{\theta}$ as

$$(\boldsymbol{\beta}^*)^T \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta}^* = 4t^2 [\Delta(\boldsymbol{\beta}^* \cdot \hat{\boldsymbol{g}})_{\rho^*}]^2, \tag{3.21a}$$

$$(\boldsymbol{\beta}^*)^T \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta}^{(i)} = (\boldsymbol{\beta}^{(i)})^T \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta}^* = 0 \quad (\forall\, i \neq 1). \tag{3.21b}$$

Then, using $\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\beta}^{(j)} = \delta_{ij}$, we obtain the condition

$$\mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta}^* = 4t^2 [\Delta(\boldsymbol{\beta}^* \cdot \hat{\boldsymbol{g}})_{\rho^*}]^2 \boldsymbol{\alpha}. \tag{3.22}$$

Matrix elements of $\mathcal{F}(\boldsymbol{\theta})$ for pure probe states and unitary evolution are given via

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4\left[\frac{1}{2}\langle\{\mathcal{H}_i, \mathcal{H}_j\}\rangle - \langle\mathcal{H}_i\rangle\langle\mathcal{H}_j\rangle\right], \tag{3.23}$$

where $\mathcal{H}_i = -iU^\dagger \partial_i U$ with $\partial_i := \partial/\partial\theta_i$, U is the unitary generated by Eq. (3.1) and the expectation values are taken with respect to the initial probe state [66].

We refer to protocols that make use of probe states and controls so that Eq. (3.22) is satisfied as optimal. However, we caution that the existence of an optimal probe state does not imply the existence of measurements on this state that allow one to extract an estimate of the parameter q saturating the lower bounds we have derived. This issue of the optimal measurements to extract parameters is also discussed extensively in, e.g. Ref., [75], with some convenient, nearly optimal, protocols presented in Refs. [71–73]. Such methods are the origin of the "small multiplicative constants" that arise in the explicit protocols above. In fact, lower bounds derived via the quantum

Cramér-Rao bound can be obtained only up to a constant $\geq \pi^2$ [76]. See Appendix B.7 for a brief explanation of these ideas.

For the particular cases considered in this Chapter, $\boldsymbol{\beta}^*$ has been explicitly calculated (see Appendices B.1 and B.2), so Eq. (3.22) can be expressed in a more meaningful form. For number operator coupling, we obtain the condition

$$\sum_{i\in\mathcal{P}} \mathcal{F}(\boldsymbol{\theta})_{ij} = \frac{N^2t^2}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\alpha_j, \tag{3.24}$$

for all j. Similarly, for the quadrature coupling, an optimal protocol requires

$$\mathcal{F}(\boldsymbol{\theta})\boldsymbol{\alpha} \sim 4\overline{N}t^2\boldsymbol{\alpha}, \tag{3.25}$$

where $\sim$ denotes asymptotically in $\overline{N}$. Equations (3.24) and (3.25) provide a generic route to finding new protocols: consider a set of parameterized families of probe states $\mathcal{T}$ that one can coherently switch between using available controls $\hat{H}_c(t)$ (here, a "family" of states refers to a particular superposition of Fock states with an arbitrary relative phase). One can then calculate $\mathcal{F}(\boldsymbol{\theta})$ via Eq. (3.23) and allocate the time spent in a particular family of states such that the associated quantum Fisher information condition is achieved. As a limiting case, one could consider $|\mathcal{T}| = 1$, removing the necessity of coherent control; the protocols considered in the previous section are of this sort (and, in Appendix B.4, we show that these protocols do, indeed, achieve the saturability conditions).

The possible choices for families of states $\mathcal{T}$ that allow for such a solution are actually quite limited, even given access to arbitrary control Hamiltonians and ancilla modes. In particular, we prove the following in the case where $\hat{g}_j := \hat{n}_j$:

Lemma 3.1 [6] *Any optimal protocol using N photons and M passes through interferometers with a coupling as in Eq. (3.1) with $\hat{g}_j = \hat{n}_j$ requires that, for every pass m, the probe state $|\psi_m\rangle$ be of the form*

$$|\psi_m\rangle \propto |\boldsymbol{N}(m)\rangle_{\mathcal{P}} |\mathbf{0}\rangle_{\mathcal{NR}} + e^{i\varphi_m} |\mathbf{0}\rangle_{\mathcal{P}} |\boldsymbol{N}'(m)\rangle_{\mathcal{NR}}, \tag{3.26}$$

where $\mathcal{P}$, $\mathcal{N}$, and $\mathcal{R}$ represent the modes with $\alpha_j \geq 0$, $\alpha_j < 0$, and the (arbitrary number of) reference modes, respectively, $\boldsymbol{N}(m)$ and $\boldsymbol{N}'(m)$ are strings of occupation numbers such that $|\boldsymbol{N}(m)| = |\boldsymbol{N}'(m)| = N$ for all passes m. φ_m is an arbitrary phase.

The proof follows straightforwardly from an explicit calculation of the Fisher information matrix for $\hat{g}_j = \hat{n}_j$, but is somewhat algebraically tedious so we relegate it to Appendix B.5.

Lemma 3.1 suggests a particular choice of $\mathcal{T}$ from which we can pick an optimal protocol for function estimation in the $\hat{g}_j = \hat{n}_j$ case. In particular, define a set of vectors

$$\mathcal{W} := \left\{\boldsymbol{\omega} \in \mathbb{Z}^d \,\middle|\, \|\boldsymbol{\omega}\|_{1,\mathcal{P}} = N,\ \|\boldsymbol{\omega}\|_{1,\mathcal{N}} \leq N,\ \omega_j\alpha_j \geq 0 \,\forall\, j\right\}. \tag{3.27}$$

Further, consider the restriction $\boldsymbol{\omega}|_{\mathcal{P}} \in \mathbb{Z}^d$ with components

$$(\boldsymbol{\omega}|_{\mathcal{P}})_j = \begin{cases} \omega_j, & j \in \mathcal{P} \\ 0, & \text{otherwise,} \end{cases} \tag{3.28}$$

and the restriction $\boldsymbol{\omega}|_{\mathcal{N}}$, defined similarly. Armed with these vectors, we can define a particular

[6] Note added: Note the similarity to Lemma A.1 in Appendix A, which plays a crucial role in the proof of Theorem 2.1 in Chapter 2.

choice $\mathcal{T}$ of one-parameter families of probe states in an occupation number basis where each $|\psi(\boldsymbol{\omega};\varphi)\rangle \in \mathcal{T}$ is labeled by a particular choice of $\boldsymbol{\omega}$ such that

$$|\psi(\boldsymbol{\omega};\varphi)\rangle \propto |\boldsymbol{\omega}|_{\mathcal{P}}\rangle\,|0\rangle + e^{i\varphi}\,|{-\boldsymbol{\omega}}|_{\mathcal{N}}\rangle\,|N - \|\boldsymbol{\omega}|_{\mathcal{N}}\|_1\rangle\,, \tag{3.29}$$

where $\varphi \in \mathbb{R}$ is an arbitrary parameter and the last mode is a reference mode. It should be clear that these families of states are of the form specified by Lemma 3.1. Furthermore, note that the proportionally weighted N00N state in Eq. (3.19) is also of this form.

Our protocols proceed as follows: starting in a state $|\psi(\boldsymbol{\omega};0)\rangle$, after any given pass through the interferometers we use control unitaries to coherently switch between families of probe states such that the relative phase between the branches is preserved (that is, we change $\boldsymbol{\omega}$, but not φ). The fact that an optimal protocol must coherently map between such states is proven in Lemma B.4 in Appendix B.5. We stay in the family of states $|\psi(\boldsymbol{\omega}_n;\varphi)\rangle$ for a fraction p_n of the passes where $p_n = \frac{r_n}{M}$ for $r_n \in \{0, 1, \ldots, M\}$ such that $\sum_n p_n = 1$. Here n indexes some enumeration of the families of states in $\mathcal{T}$.

The value of the component ω_j in a given probe state determines the contribution of the parameter θ_j coupled to sensor j to the relative phase between the two branches of the probe state during a single pass. In particular, in a single pass with a probe state in the family $|\psi(\boldsymbol{\omega};\varphi)\rangle$, the relative phase between the two branches of the probe state becomes $\boldsymbol{\omega}\cdot\boldsymbol{\theta}+\varphi$. Assuming an initial probe state with $\varphi = 0$ and summing over all passes we obtain a total relative phase

$$\begin{aligned}\varphi_{\mathrm{tot}} &= M \sum_n p_n(\boldsymbol{\omega}_n \cdot \boldsymbol{\theta}) && (3.30)\\ &=: (W\boldsymbol{r}) \cdot \boldsymbol{\theta}. && (3.31)\end{aligned}$$

In the second line, we implicitly defined W as a matrix whose columns are the vectors $\boldsymbol{\omega}_n \in \mathcal{W}$ and $\boldsymbol{r} := M\boldsymbol{p} \in \mathbb{Z}^{|\mathcal{T}|}$. Explicitly computing the Fisher information matrix for these states demonstrates that the optimality condition in Eq. (3.24) is satisfied if

$$W\boldsymbol{r} = NM\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}; \tag{3.32}$$

see Appendix B.4 for details. Consequently, any integer solution $\boldsymbol{r}$ to Eq. (3.32) such that

$$\begin{aligned}\|\boldsymbol{r}\|_1 &= M,\\ \boldsymbol{r} &\geq 0,\end{aligned} \tag{3.33}$$

yields an optimal protocol. The protocols of Ref. [13], described above and generalized in Appendix B.4, are a particularly simple case within this class with $M = 1$ and $\boldsymbol{\omega} = \frac{N\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}$, i.e. we select out only a single column of $\mathcal{W}$.

Solutions to Eqs. (3.32) and (3.33) are not guaranteed to exist for all N, M. In particular, we require that

$$NM\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}} \in \mathbb{Z}^d. \tag{3.34}$$

For $\boldsymbol{\alpha} \in \mathbb{Q}$ and sufficiently large N or M this hold true. Setting up the system of equations in

Eqs. (3.32) and (3.33) that must be solved to pick out explicit protocols requires identifying the set of vectors $\mathcal{W}$ defined in Eq. (3.27). While computationally straightforward, if expensive, to construct and enumerate this set, the number of states is extremely large, yielding a correspondingly large set of linear Diophantine equations in Eq. (3.32). Consequently, it is reasonable to place further, experimentally motivated constraints to limit this set of states and pick out advantageous protocols. For instance, one such constraint is to limit the amount of entanglement between modes on any given pass. We consider this case in the following section.

It is also important to note that integer linear programming is NP-hard [4], so finding a particular solution once we add additional constraints is not a computationally easy task. Regardless, in applications one can apply standard (possibly heuristic) algorithms for integer linear programming to seek solutions. If a solution is found, it is known to be optimal. Consequently, proving the existence or lack thereof of a solution with certain additional constraints may be intractable for large problem instances.

Similar arguments to those that go into proving Lemma 3.1 allow us to show that, for quadrature sensing, the condition in Eq. (3.25) can be reduced to the condition that

$$\mathcal{F}(\boldsymbol{\theta})_{ij} \sim \frac{4\overline{N}t^2}{\|\boldsymbol{\alpha}\|_2^2}\alpha_i\alpha_j, \tag{3.35}$$

which is proven in Appendix B.6. However, there is not a clearly interesting family of states that can be leveraged to achieve this quantum Fisher information, as in the case of number operator coupling or qubit sensors [1]. However, the existing optimal protocols described above do obey this condition asymptotically in average photon number $\overline{N}$.

3.5 Entanglement Requirements

The remaining flexibility in the choice of optimal probe states enabled by some control also allows us to impose further experimentally relevant constraints. One reasonable constraint is the amount of intermode entanglement required during the sensing process. This was considered in Ref. [1] for the case of qubit sensors.

The answer to the entanglement question in the current context depends crucially on the sorts of control operations we allow. In the number operator case, with arbitrary time-dependent control, only two-mode entanglement is needed at any given time, as one can simply prepare a N00N state between the reference and one of the sensing modes and coherently switch which sensing mode is entangled with the reference mode such that the time spent entangled with mode j is given by $t_j = |\alpha_j| t / \|\boldsymbol{\alpha}\|_1$. For similar reasons, no entanglement is needed for displacement sensing; here, no reference mode is needed and one can simply sequentially apply displacement operators for a time $t_j = |\alpha_j| t / \|\boldsymbol{\alpha}\|_1$ on a single-mode squeezed state, followed by a homodyne measurement. When control operations to change the probe state are allowed only at M discrete time intervals, as described by Eq. (3.3), the problem becomes more interesting. For number operator coupling, subject to a fixed photon number constraint, any optimal protocol requires at least ($\lceil \|\boldsymbol{\alpha}\|_0 / M \rceil$+1)-mode entanglement. This bound is fairly trivial: it merely states that one must be entangled with each nontrivial mode for at least one pass. For displacement operator coupling, subject to a fixed average photon number constraint, an essentially identical argument allows us to prove that any optimal protocol requires at least $\lceil \|\boldsymbol{\alpha}\|_0 / M \rceil$-mode entanglement. The difference of one is because, unlike displacement sensing, phase sensing generally requires entanglement with a reference mode. In the $M \to \infty$ limit, we recover the continuous control

	Qubit phase sensing	Phase sensing	Displacement sensing
Parameter coupling	$\frac{1}{2}\hat{\sigma}_i^z\theta_i$	$\hat{n}_i\theta_i$	$\frac{i}{2}(\hat{a}_i^\dagger - \hat{a}_i)\theta_i$
Resources	Qubit number, d sensing time, t	Photon number, N sensing time, t	Avg. photon number, $\overline{N}$ sensing time, t
MSE (separable)	$\geq \frac{\lVert\boldsymbol{\alpha}\rVert_2^2}{t^2}$ [14]	$\geq \frac{\lVert\boldsymbol{\alpha}\rVert_{2/3}^2}{N^2t^2}$ [13]	$\geq \frac{\lVert\boldsymbol{\alpha}\rVert_1^2}{4\overline{N}t^2}$
MSE (entangled)	$\geq \frac{\lVert\boldsymbol{\alpha}\rVert_\infty^2}{t^2}$ [14]	$\geq \frac{\lVert\boldsymbol{\alpha}\rVert_{1,\mathcal{P}}^2}{N^2t^2}$	$\geq \frac{\lVert\boldsymbol{\alpha}\rVert_2^2}{4\overline{N}t^2}$ [26]
Entanglement needed (discrete controls)	$k \geq \max\left\{\left\lceil \frac{\lVert\boldsymbol{\alpha}\rVert_1}{\lVert\boldsymbol{\alpha}\rVert_\infty}\right\rceil, \left\lceil \frac{\lVert\boldsymbol{\alpha}\rVert_0}{M}\right\rceil\right\}$	$k > \left\lceil \frac{\lVert\boldsymbol{\alpha}\rVert_0}{M}\right\rceil$	$k \geq \left\lceil \frac{\lVert\boldsymbol{\alpha}\rVert_0}{M}\right\rceil$
Entanglement needed (arbitrary controls)	$\frac{\lVert\boldsymbol{\alpha}\rVert_1}{\lVert\boldsymbol{\alpha}\rVert_\infty} \in (k-1, k]$ [1]	$k = 2$	No entanglement
k-partite entanglement protocol always exists?	Yes [1]	No	Yes

Table 3.1: Comparison of the lower bounds on the mean square error and entanglement requirements for an (asymptotically) optimal protocol obeying the corresponding conditions on the quantum Fisher information for the task of estimating a linear function $q = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ with qubit, phase sensing, and displacement sensing quantum sensor networks.

case, so these trivial bounds can be tight. This triviality is in contrast to the qubit case, where results analogous to Lemma 3.1 lead to significantly tighter constraints on the minimum amount of necessary entanglement for optimal protocols [1].[7] This discrepancy arises due to the fact that, unlike with photonic resources which must be distributed in a zero-sum way between modes, for qubit sensors one can be maximally sensitive to all coupled parameters simultaneously.

3.6 Conclusion and Outlook

We have determined the fundamental achievable performance limits for phase sensing and have extended proofs of lower bounds for displacement sensing beyond just an average to arbitrary functions. In the process, we proved a long-standing conjecture regarding function es-

[7] Note added: See Theorem 2.1 in Chapter 2.

timation with number operator coupling [13] and showed that some of the protocols that exist in the literature [13, 16, 84], are, in fact, optimal in the asymptotic limit. By considering different implementations of a quantum sensor network within a single framework, we reveal the role of entanglement and controls as they relate to the type of coupling and whether the relevant resource is "parallel" (as in qubit sensor networks, where all parameters can simultaneously be measured to maximal precision) or "sequential" (as in photonic sensor networks, where the photons must be optimally distributed between modes). Our approach to proving our bounds also enables an algebraic framework for developing further optimal protocols, subject to various constraints. Here, we considered the particular case of entanglement-based constraints, enabling comparison to similar work in the case of qubit sensors [1]. These results, and how they fit into the landscape of known results for quantum sensor networks, are summarized in Table 3.1. How other constraints impact the existence of and control requirements for optimal protocols remains an interesting open question deserving of further study.

Acknowledgments

We thank Luis Pedro García-Pintos for helpful discussions. This work was supported in part by DARPA SAVaNT ADVENT, AFOSR MURI, AFOSR, NSF STAQ program, DoE ASCR Accelerated Research in Quantum Computing program (award No. DE-SC0020312), NSF QLCI (award No. OMA-2120757), and the DoE ASCR Quantum Testbed Pathfinder program (awards No. DE-SC0019040 and No. DE-SC0024220). Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator.

Chapter 4: Transition of Anticoncentration in Gaussian Boson Sampling

Abstract: Gaussian Boson Sampling is a promising method for experimental demonstrations of quantum advantage because it is easier to implement than other comparable schemes. While most of the properties of Gaussian Boson Sampling are understood to the same degree as for these other schemes, we understand relatively little about the statistical properties of its output distribution. The most relevant statistical property, from the perspective of demonstrating quantum advantage, is the *anticoncentration* of the output distribution as measured by its second moment. The degree of anticoncentration features in arguments for the complexity-theoretic hardness of Gaussian Boson Sampling, and it is also important to know when using cross-entropy benchmarking to verify experimental performance. In this Chapter, we develop a graph-theoretic framework for analyzing the moments of the Gaussian Boson Sampling distribution. Using this framework, we show that Gaussian Boson Sampling undergoes a transition in anticoncentration as a function of the number of modes that are initially squeezed compared to the number of photons measured at the end of the circuit. When the number of initially squeezed modes scales sufficiently slowly with the number of photons, there is a lack of anticoncentration. However, if the number of initially squeezed modes scales quickly enough, the output probabilities anticoncentrate weakly.

4.1 Introduction

There is a hope that quantum computation will be able to outperform classical computation on certain tasks. In particular, there has been a recent explosion of interest in so-called sampling problems given the strong theoretical evidence for an exponential speedup of quantum algorithms over the best possible classical algorithms; see Ref. [28] for an overview. Aaronson and Arkhipov introduced one of the most deeply studied sampling frameworks in their seminal work on Boson Sampling [27]. The Boson Sampling task is to approximately sample from the outcome distribution of measuring n single photons in m optical modes transformed by a Haar-random linear-optical unitary, which can be implemented as a random network of beamsplitters and phase shifters [94]. Reference [27] focused on single-photon input states, but these can be challenging to produce experimentally [95] because existing single-photon sources are not sufficiently reliable to avoid an exponential amount of post-selection [96]. Therefore, there has been an interest in generalizing the original Boson Sampling setup to other input states.

The currently most feasible generalization is Gaussian Boson Sampling (GBS) [97–100], which uses Gaussian input states. These states are significantly easier to prepare reliably than single-photon states. At the same time, similar statements can be made about the hardness of sampling from the corresponding output distribution [100–103], and several large-scale GBS experiments have been performed recently [30–33].

Broadly speaking, the hardness of Boson Sampling is based on the connection between output probabilities and the permanent, which is, classically, $\#\mathsf{P}$-hard to compute exactly [104]. Similarly, the hardness of GBS arises from the fact that output probabilities are controlled by a generalization of the permanent called the hafnian; while the permanent counts the number of perfect matchings in a weighted bipartite graph, the hafnian counts the number of perfect matchings in an arbitrary weighted graph [105]. Because the hafnian generalizes the permanent, it is also difficult to compute classically.

However, the complexity of classically computing an individual output probability defined in terms of the permanent or the hafnian is not itself sufficient to prove hardness of sampling from the overall probability distributions. The standard hardness argument based on Stockmeyer's algorithm [28, 34] requires that outcome probabilities of *random* Boson Sampling instances be hard to *approximate*. Jointly with provable hardness of nearly exactly computing output probabilities [101], anticoncentration of the outcome probabilities serves as evidence for this. Intuitively, if most outcome probabilities are comparable to the uniform probability, then a good classical sampling algorithm needs a very precise idea of each probability's relative magnitude because all of them are important. Anticoncentration quantifies this idea as, most concisely, the outcome-collision probability (i.e. the probability of getting the same outcome from two independent samples) of the GBS distribution averaged over the choice of linear-optical unitary and normalized by the size of the sample space [28, Section D]. While a weak form of anticoncentration holds in Boson Sampling [27], under what conditions anticoncentration holds in GBS is an open question.

In this Chapter, we analyze the moments of GBS in the photon-collision-free limit. In this limit, the output distribution is dominated by outcomes with at most a single photon in each mode, and the moments of GBS approximately reduce to moments of squared hafnians of Gaussian random matrices. We show that evaluating those moments reduces to counting the connected components of certain graphs. Using this perspective, we find a closed-form expression for the first moment and derive analytical properties of the second moment. We then identify a transi-

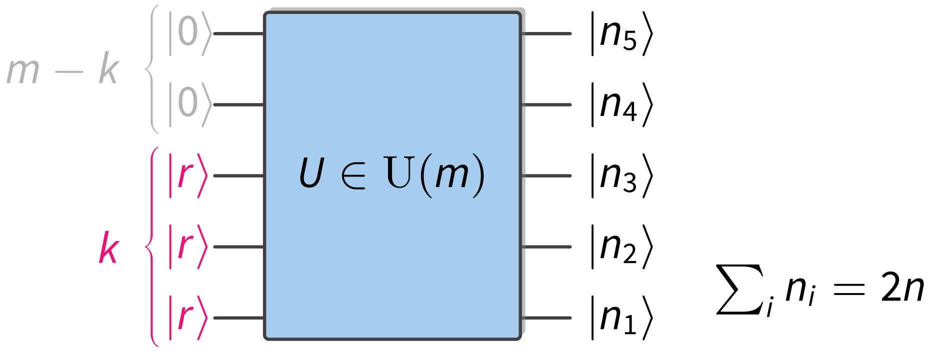

Figure 4.1: In Gaussian Boson Sampling (GBS), k out of m modes are prepared in single-mode squeezed states with parameter r, while the remaining modes are prepared in the vacuum state $|0\rangle$. The modes are then transformed by a Haar random linear-optical unitary U and measured in the Fock basis with outcome counts n_i summing to $2n$.

tion in anticoncentration in GBS: when the number of initially squeezed modes is large enough compared to the measured number of photons n, a weak version of anticoncentration holds where the normalized average outcome-collision probability scales as $\sqrt{n}$. However, when sufficiently few modes are initially squeezed, there is a lack of anticoncentration, as the normalized second moment scales exponentially in n.

The rest of this Chapter proceeds as follows. We first provide background information and set up the system and problem of interest. We then derive the graph-theoretic formalism for computing the first moment of the output probabilities. We proceed to discuss how to apply the formalism to calculate certain properties of the second moment. These results let us finally prove the transition in anticoncentration.

4.2 Setup

We consider a photonic system with m modes that is transformed by a Haar-random linear optical unitary $U \in \mathrm{U}(m)$ acting on the modes of the system, see Fig. 4.1. In the paradigmatic version of GBS [99, 100], the first k of the m modes are prepared in single-mode squeezed states with equal squeezing parameter r, while the remaining $m - k$ modes are prepared in the vacuum state. After applying U, all m modes are measured in the Fock basis.

Reference [99] proves that, given a unitary U, the probability of obtaining an outcome count vector $\boldsymbol{n} = (n_1, n_2, \ldots, n_m) \in \mathbb{N}_0^m$ with total photon count $2n = \sum_{i=1}^m n_i$ is given by

$$P_U(\boldsymbol{n}) = \frac{\tanh^{2n} r}{\cosh^k r} \left|\mathrm{Haf}(U_{1_k,\boldsymbol{n}}^\top U_{1_k,\boldsymbol{n}})\right|^2, \tag{4.1}$$

where $U_{1_k,\boldsymbol{n}}$ denotes the $k \times 2n$ submatrix of U given by its first k rows and the columns selected according to the nonzero entries of $\boldsymbol{n}$ each copied n_i times.[1] Moreover,

$$\mathrm{Haf}(A) = \frac{1}{2^n n!} \sum_{\sigma \in S_{2n}} \prod_{j=1}^{n} A_{\sigma(2j-1),\sigma(2j)} \tag{4.2}$$

denotes the hafnian of a $2n \times 2n$ symmetric matrix A.[2]

We will work in the regime in which the output states are, with high probability.(photon-)collision-free, meaning that $n_i \in \{0, 1\}$ for all i. A sufficient condition for this to be the case is that the expected number of photons $\mathbb{E}[2n] = k \sinh^2(r) = o(\sqrt{m})$. In this regime, Ref. [101] provides evidence that, for any fixed $n = o(\sqrt{m})$, the distribution over submatrices is well-captured by a generalization of the circular orthogonal ensemble (COE).[3]

Conjecture 4.1 (Hiding [101])**.** *For any k such that $1 \leq k \leq m$ and $2n = o(\sqrt{m})$, the distribution of the symmetric product $U_{1_k,\boldsymbol{n}}^\top U_{1_k,\boldsymbol{n}}$ of submatrices of a Haar-random $U \in \mathrm{U}(m)$ closely approx-*

[1] Note that squeezed states are supported only on even Fock states, so the total photon count $2n$ must always be even.

[2] Other equivalent definitions exist, but Eq. (4.2) is the most convenient one for our purposes.

[3] Note that, strictly speaking, the conjecture is only formulated for the regime $n \leq k$ in Ref. [101]. However, the evidence provided there for the case $k = n$—Ref. [27] proves that $n \times n$ submatrices of Haar-random unitaries are approximately Gaussian—clearly also holds for the case $k \leq n$.

imates in total variation distance the distribution of the symmetric product $X^\intercal X$ of a complex Gaussian matrix $X \sim \mathcal{N}(0,1/m)_c^{k\times 2n}$ with mean 0 and variance $1/m$.

We work under the assumption that Conjecture 4.1 is true. Fixing the measured number of photons $2n$, the normalized average (outcome-)collision probability, which quantifies anticoncentration, can be written as $|\Omega_{2n}|\mathbb{E}_{U\in U(m)}[\sum_{\boldsymbol{n}\in\Omega_{2n}} P_U(\boldsymbol{n})^2]$, where $|\Omega_{2n}|$ is the size of the photon-non-collisional sample space of $2n$ photons in m modes, which is the dominant space of outputs when $n = o(\sqrt{m})$. Conjecture 4.1 implies that, with respect to random choices of U, all outcomes are equally distributed over the unitaries, the so-called hiding property. This implies that the inverse size of the sample space is given by the first moment $\mathbb{E}_U[P_U(\boldsymbol{n})]$. See Appendix C.4 for more details. Under Conjecture 4.1, the anticoncentration property thus reduces to computing the moments

$$M_t(k,n) \coloneqq \mathbb{E}_{X\sim\mathcal{G}^{k\times 2n}}[|\mathrm{Haf}(X^\intercal X)|^{2t}] \tag{4.3}$$

of the squared hafnian as a function of the parameters k and n for $t = 1,2$, where we have abbreviated $\mathcal{N}(0,1)_c^{k\times 2n}$ as $\mathcal{G}^{k\times 2n}$. We consider unit variance because rescaling X by $1/\sqrt{m}$ just leads to an overall prefactor that, like the prefactor in Eq. (4.1), is irrelevant to the normalized average outcome-collision probability. We will phrase our discussion in terms of the inverse of the average collision probability

$$m_2(k,n) \coloneqq M_1(k,n)^2/M_2(k,n). \tag{4.4}$$

4.3 First Moment and Graph-Theoretic Formalism

We begin by analyzing the (rescaled) first moment M_1 of the output probabilities. In order to derive our graph-theoretic formalism, we use Eq. (4.2) to expand the hafnian in Eq. (4.3) as a sum over permutations of a product of matrix elements. From there, the key is to use that the matrix elements are independent complex Gaussian, meaning that $\mathbb{E}_{X\sim\mathcal{G}^k}[X_i X_j^*] = \delta_{ij}$ and $\mathbb{E}_{X\sim\mathcal{G}^k}[X_i X_j] = \mathbb{E}_{X\sim\mathcal{G}^k}[X_i^* X_j^*] = 0$. This yields

$$M_1(k,n) = \frac{(2n)!}{(2^n n!)^2} \sum_{\tau\in S_{2n}} \sum_{\{o_i\}_{i=1}^{n}}^{k} \prod_{j=1}^{n} \delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}}. \tag{4.5}$$

Let us briefly discuss Eq. (4.5), see Appendix C.1 for details. The sum over $\tau \in S_{2n}$ and the product over index j come from Eq. (4.2); the sum over the indices $o_i \in [k] \coloneqq \{1, 2, \dots, k\}$ is due to an expansion of $X^\top X$ as a matrix product. Note that, when $\tau(2j-1)$ and $\tau(2j)$ form a tuple $(2\ell-1, 2\ell)$, then the Kronecker δ always equals 1 for index o_ℓ, such that summing over o_ℓ yields a factor of k. When $\tau(2j-1)$ and $\tau(2j)$ do not form such a tuple, we get a nontrivial relationship between indices that decreases the number of independent degrees of freedom, thus decreasing the number of factors of k in the final answer. Therefore, to evaluate this expression, one must determine the number of "free indices" over all the permutations in S_{2n}. We accomplish this with our graph-theoretic approach.

Specifically, define a graph G_τ as follows, see Fig. 4.2(a). Let G_τ have $2n$ vertices labeled O_1 through O_{2n}. These upper-case vertices are not directly equivalent to the lower-case indices in Eq. (4.5)—instead, each index o_j splits into two vertices O_ℓ and $O_{\ell'}$ such that $\lceil \tau(\ell)/2 \rceil = j = \lceil \tau(\ell')/2 \rceil$ (in other words, $o_{\lceil \tau(\ell)/2 \rceil}$ maps to a vertex O_ℓ). Let G_τ have a black edge between O_{2j-1}

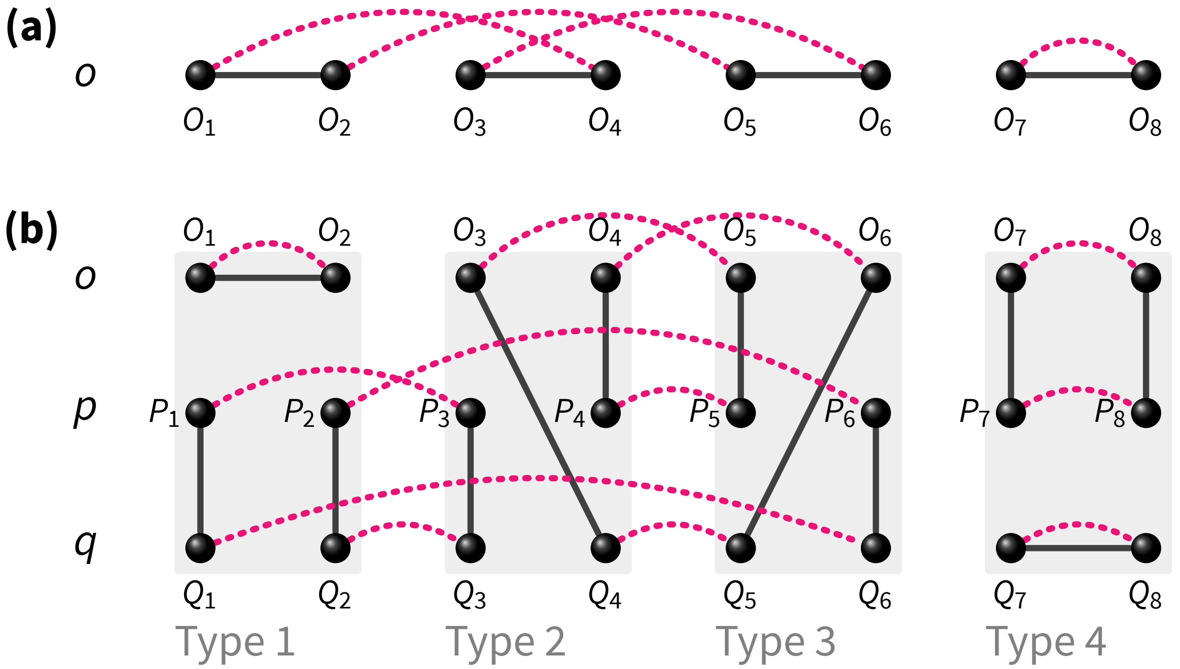

Figure 4.2: Examples of graphs used to calculate the (a) first and (b) second moments of GBS outcome probabilities. (a) $G \in \mathbb{G}_4^1$. There are eight vertices O_1 to O_8 representing the index o (labeled in the left column). The black (solid) edges connect only adjacent pairs, whereas the red (dashed) edges form an arbitrary perfect matching. This graph has two connected components, meaning it contributes k^2 to the first moment. (b) $G \in \mathbb{G}_4^2$. The black (solid) edges are, from left to right, type-1, type-2, type-3, and type-4, as denoted by the gray background. $z = 1 + 0 \times 4^3 + 1 \times 4^2 + 2 \times 4^1 + 3 \times 4^0 = 28$ (this is calculated by converting 0123 from base-4 into base-10 and then adding 1 such that the final result is in $[4^4]$). Note that black (solid) edges stay within two adjacent columns. Red (dashed) edges form an arbitrary perfect matching in each row. This graph contributes k^5 to the second moment, as there are five connected components.

and O_{2j} for all $j \in [n]$, and a red edge between O_ℓ and $O_{\ell'}$ if $\lceil \tau(\ell)/2 \rceil = \lceil \tau(\ell')/2 \rceil$. These two kinds of edges mimic two types of ways that dependencies in Eq. (4.5) can be induced through an index j. Red edges identify the ℓ and ℓ' that map to the same value via τ and the ceiling function, hence red edges identify which vertices came from the same o index. Black edges identify that Eq. (4.5) has a Kronecker δ between $o_{\lceil \tau(2j-1)/2 \rceil}$ and $o_{\lceil \tau(2j)/2 \rceil}$.

We see, then, that the number of connected components of G_τ, $C(G_\tau)$, is equivalent to the number of free indices in the sum in Eq. (4.5). Therefore,

$$M_1(k,n) = \frac{(2n)!}{(2^n n!)^2} \sum_{\tau \in S_{2n}} k^{C(G_\tau)}. \tag{4.6}$$

Now, there is a degeneracy where many permutations induce the same final graph. Each graph has the same fixed set of black edges and then one of $(2n-1)!!$ possible sets of red edges (this is the number of ways of pairing $2n$ elements when order does not matter). For each graph G corresponding to some assignment of the red edges, there are $2^n n!$ permutations τ such that $G_\tau = G$. Therefore, instead of studying G_τ as instantiated by permutations τ, we study the underlying graphs G. Define $\mathbb{G}_n^1$ to be the set of graphs on $2n$ vertices with one perfect matching defined by the fixed set of black edges and one perfect matching defined by the arbitrary red edges. We can thus rewrite $M_1 = (2n-1)!! \sum_{G \in \mathbb{G}_n^1} k^{C(G)}$ and state our first result.

Theorem 4.1. *The sum over graphs in $\mathbb{G}_n^1$ satisfies*

$$\sum_{G \in \mathbb{G}_n^1} k^{C(G)} = k(k+2)\dots(k+2n-2), \tag{4.7}$$

and hence $M_1(k,n) = (2n-1)!!(k+2n-2)!!/(k-2)!!$.

The proof proceeds by induction over n, where the inductive step reduces a graph in $\mathbb{G}_n^1$ to one in $\mathbb{G}_{n-1}^1$ through an analysis of the red edge connected to O_1. There are two options for this red edge: either it connects to O_2, the vertex with which O_1 shares a black edge, or it attaches to some $O_{x>2}$. The former creates a connected component of size two; the latter reduces to a graph in $\mathbb{G}_{n-1}^1$ by merging vertices O_1, O_2, and O_x (which does not change the number of connected components). Full details can be found in Appendix C.3.

4.4 Second Moment

We now sketch the application of our graph-theoretic formalism to the second moment M_2, deferring the details to Appendix C.2. We expand $|\mathrm{Haf}(X^\intercal X)|^4$ using Eq. (4.2) which becomes a sum of products of matrix elements that are indexed by four permutations in S_{2n}. The independence of matrix elements again ensures that we must have an equal number of copies of X_{ij} and X_{ij}^* for the expectation value not to vanish on a given product. However, because there are more copies of X, there are more ways of matching the indices. Accounting for these possibilities leads to an expression analogous to Eq. (4.5). The key differences are the following: (1) instead of summing over a single permutation, we now sum over three permutations, labeled τ, α, β (as in the first moment, one of the original four permutations eventually becomes redundant); (2) instead of summing over n indices $\{o_i\}_{i=1}^n$, we now sum over $3n$ indices $\{o_i, q_i, p_i\}_{i=1}^n$; (3) each factor is a sum of four possible Kronecker δ terms corresponding to the different types of index matching.

As before, we will define a useful set of graphs; see Fig. 4.2(b) for an example. We expand each index in $\{o_i, q_i, p_i\}_{i=1}^n$ to two graph vertices $\{O_i, Q_i, P_i\}_{i=1}^{2n}$, and we organize them into $2n$

columns and three rows assigned to O, P, and Q vertices, respectively. We then use the Kronecker δs to define black and red edges between these vertices. Fixing permutations τ, α, β, there is a red edge between O_ℓ and $O_{\ell'}$ if $\lceil \tau(\ell)/2 \rceil = \lceil \tau(\ell')/2 \rceil$, and similarly for the P and Q vertices using permutations α and β, respectively. This means that the red edges are constrained to lie within rows in the graph. Furthermore, these red edges again identify the vertices originating from the same index. Because each factor has four Kronecker δ terms, each factor contributes one of four patterns of black edges, which we refer to as type-1, type-2, type-3, and type-4. Due to the nature of these Kronecker δ terms, the black edges are constrained to lie within pairs of adjacent columns. Each graph then has one of 4^n possible sets of black edges indexed by an integer z. We therefore call these graphs $G_{\tau,\alpha,\beta}(z)$.

As in the first moment, the number of connected components $C(G_{\tau,\alpha,\beta}(z))$ of the graph $G_{\tau,\alpha,\beta}(z)$ gives the number of free indices of its corresponding term in the expansion of the hafnian, meaning that graph contributes $k^{C(G_{\tau,\alpha,\beta}(z))}$ to the sum. The second moment then becomes

$$M_2(k,n) = \frac{(2n)!}{(2^n n!)^4} \sum_{\tau,\alpha,\beta \in S_{2n}} \sum_{z \in [4^n]} k^{C(G_{\tau,\alpha,\beta}(z))}. \tag{4.8}$$

We also again use the fact that many permutations induce the same final graph. We thus define $\mathbb{G}_n^2(z)$ to be the set of graphs for the zth set of black edges and $\mathbb{G}_n^2 \coloneqq \bigcup_{z \in [4^n]} \mathbb{G}_n^2(z)$. Because there are now three permutations associated to each graph, we obtain a degeneracy factor of $(2^n n!)^3$ and find

$$M_2(k,n) = (2n-1)!! \sum_{G \in \mathbb{G}_n^2} k^{C(G)}. \tag{4.9}$$

We can now state our second result.

Theorem 4.2. *The second moment $M_2(k,n)$ is a degree-$2n$ polynomial in k and can be written as $M_2(k,n) = (2n-1)!!\sum_{i=1}^{2n} c_i k^i$, where c_i is the number of graphs $G \in \mathbb{G}_n^2$ that have i connected components.*

Theorem 4.2 follows from Eq. (4.9) and verifying that the limits of summation are correct, which is done in Appendix C.3.

4.5 Transition in Anticoncentration

We now use Theorems 4.1 and 4.2 in order to show that anticoncentration in GBS undergoes a transition as a function of k. Roughly speaking, $m_2(k,n)$ upper-bounds the fractional support of the outcome distribution on outcomes with probabilities larger than uniform, i.e. those most relevant to the sampling task, as we explain in detail in Appendix C.5.1. We speak of strong anticoncentration if $m_2(k,n) \geq \text{const.}$. We speak of weak anticoncentration if $m_2(k,n) \geq 1/\text{poly}(n)$. If $m_2(k,n) = O(1/n^a)$ for any constant $a > 0$, however, we say that there is a lack of anticoncentration; in this regime, only a negligible fraction of the probabilities are nontrivial. While our definition of anticoncentration in terms of m_2 is stronger than the standard definition, it captures the essence of anticoncentration, see Appendix C.5.1. We show a transition between a lack of anticoncentration for $k = 1$ and weak anticoncentration for $k \to \infty$ (which, of course, requires $m \to \infty$ as well).

In order to do so, we analyze the polynomial coefficients c_i, observing that for $k = 1$, $M_2(k,n) = (2n-1)!!\sum_{i=1}^{2n} c_i$, and for $k \to \infty$[4], $M_2(k,n) \approx (2n-1)!!\, c_{2n} k^{2n}$.[5] The following lemma states our results for these regimes.

[4] For any n, there exists some sufficiently large k for which the leading order term dominates. The exact required scaling of k with n is investigated more thoroughly in the companion piece Ref. [106].

[5] Note added: The companion piece Ref. [106] in the previous footnote refers to Chapter 5.

Lemma 4.1 *We have that*

i. $M_2(1,n) = ((2n-1)!!)^4 4^n$

ii. $c_{2n} = (2n)!!$

Part (i) of the lemma follows from a simple, direct computation using the expansion of the second moment in terms of Kronecker δs; part (ii) follows by reducing the graph counting problem to a special instance of the first moment with $k = 2$. This reduction happens because the types of edges that are allowed in order to get $2n$ connected components are quite restrictive, see Appendix C.3 for details.

Theorem 4.2 and Lemma 4.1 imply that, when $k = n^0 = 1$, the inverse normalized second moment is negligible:

$$m_2(1,n) = \frac{((2n-1)!!^2)^2}{(2n-1)!!^4 4^n} = 4^{-n}. \tag{4.10}$$

Now take $k = n^a \leq m$, for some large a. In this limit, $M_2(k,n)$ is dominated by the behavior of its leading order in k, which is $(2n-1)!!(2n)!!k^{2n}$. Additionally, $M_1(k,n) = (2n-1)!!(k+2n-2)!!/(k-2)!! \sim (2n-1)!!k^n$ and, hence, the k-dependence of $m_2(k,n)$ vanishes. Using Stirling's approximation on the remaining n-dependence yields

$$m_2(k,n) \sim \frac{(2n-1)!!}{(2n)!} = \frac{(2n)!}{4^n (n!)^2} \sim \frac{1}{\sqrt{\pi n}}. \tag{4.11}$$

This proves the central claim of this Chapter. In Appendix C.5.2, we also show how anticoncentration of the approximate GBS distribution relates to anticoncentration of the true distribution.

4.6 Discussion and Conclusion

In this Chapter, we have shown a transition in anticoncentration in the output probabilities of GBS as a function of the number of initially squeezed modes. The presence of anticoncentration is additional evidence for the hardness of GBS, and our results thus yield clear advice for experiments in the collision-free regime: given a desired average photon number, distribute the required squeezing for this number across all modes.

Our results give rise to an interesting state of affairs when considered in conjunction with the hiding property: in both GBS and standard Boson Sampling, the hiding property is known to fail outside of the highly dilute collision-free regime which is characterized by $m = O(n^2)$ [107, 108], while it is conjectured to hold for any $m = \omega(n^2)$ [27, 101]. Standard Boson Sampling anticoncentrates weakly with inverse normalized second moment $1/n$ in the same regime [27, Lemma 8.8]. The only relevant scale is thus the relative size of the number of modes to the number of photons. In GBS, we now find an additional relevant scale, the number of squeezed modes in the input state. This scale does not seem to be relevant to the hiding property in GBS which holds for $m = \omega(n^2)$ and *any* k under Conjecture 4.1. We find, however, that it is very relevant to the anticoncentration property of GBS.

For a potential explanation of the relevance of this scale, we refer to Scattershot Boson Sampling, which is "intermediate" between standard Boson Sampling and GBS. In Scattershot Boson Sampling, n single photons are distributed randomly across the input modes using post-selection on two-mode squeezed states. In order to satisfy collision-freeness in the input state with high probability, the total squeezing in the input needs to be distributed across $\omega(n^2)$ initial squeezed states [97], see Appendix C.6 for details. It is not clear to what extent this explanation generalizes to GBS, however, because the distribution of photons in the input state of GBS is only supported on (collision-full) integer multiples of photon pairs in every mode. Therefore, this connection warrants future study.

Our results also connect to the classical simulability of GBS. The hafnian of A can be computed in time exponential in the rank of A [109]. The absence of anticoncentration for small $k \ll n$ overlaps with this regime of efficient classical simulability, as the rank of $X^\intercal X$ is upper-bounded by k.

But does it also extend beyond this regime? While we have been able to prove the existence of this transition, our above work is not sufficient to pin down its precise location. However, we conjecture that, weak anticoncentration holds for $k = \omega(n^2)$, but there is a lack of anticoncentration for $k = O(n^2)$. In a companion work [106],[6] we give evidence for this conjecture by fully analyzing the coefficients c_i of $M_2(k, n)$.

Acknowledgments

We are grateful to Changhun Oh for sharing his independent calculation of the first moment. We also thank Changhun Oh, Bill Fefferman, Marcel Hinsche, Max Alekseyev, and Benjamin Banavige for helpful discussions, and Marcel Hinsche for comments on an earlier version of the manuscript. A.E., J.T.I., and A.V.G. were supported in part by the DoE ASCR Accelerated Research in Quantum Computing program (award No. DE-SC0020312), DARPA SAVaNT ADVENT, AFOSR MURI, DoE ASCR Quantum Testbed Pathfinder program (awards No. DE-SC0019040 and No. DE-SC0024220), NSF QLCI (award No. OMA-2120757), NSF STAQ program, and AFOSR. Support is also acknowledged from the U.S. Department of Energy, Office of

[6] Note added: Again, Ref. [106] refers to Chapter 5 and the associated Appendix D.

Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. J.T.I thanks the Joint Quantum Institute at the University of Maryland for support through a JQI fellowship. D.H. acknowledges funding from the US Department of Defense through a QuICS Hartree fellowship.

Chapter 5: The Second Moment of Hafnians in Gaussian Boson Sampling

Abstract: Gaussian Boson Sampling is a popular method for experimental demonstrations of quantum advantage, but many subtleties remain in fully understanding its theoretical underpinnings. An important component in the theoretical arguments for approximate average-case hardness of sampling is anticoncentration, which is a second-moment property of the output probabilities. In Gaussian Boson Sampling these are given by hafnians of generalized circular orthogonal ensemble matrices. In a companion work [arXiv:2312.08433][1], we develop a graph-theoretic method to study these moments and use it to identify a transition in anticoncentration. In this work, we find a recursive expression for the second moment using these graph-theoretic techniques. While we have not been able to solve this recursion by hand, we are able to solve it numerically exactly, which we do up to Fock sector $2n = 80$. We further derive new analytical results about the second moment. These results allow us to pinpoint the transition in anticoncentration and furthermore yield the expected linear cross-entropy benchmarking score for an ideal (error-free) device.

[1] Note added: Any use in this Chapter of 2312.08433, "companion work," or Ref. [110] refers to Chapter 4 and the associated Appendix C.

5.1 Introduction

One of the major goals of quantum computer science is to find examples of certain tasks on which quantum devices can outperform classical computers. While the ultimate goal is to develop quantum computers that can run, say, Shor's algorithm [3], the qubit numbers, gate fidelities, and error correction needed to accomplish such a task fault-tolerantly are well beyond the current state of the art. Therefore, there is interest in finding near-term examples of quantum advantage.

One area of focus that has strong theoretical evidence for an exponential speedup over the best possible classical algorithms comprises the so-called sampling problems. Aaronson and Arkhipov introduced one such promising framework called Boson Sampling [27]. The Boson Sampling task is to produce a sample (that is, a valid output Fock state) according to the outcome distribution generated by measuring indistinguishable photons that have been subjected to a random linear optical network of beam-splitters and phase shifters. In Boson Sampling, the input states consist of single photons on many input modes. However, because single-photon sources have imperfect efficiency, these states are difficult to produce experimentally, requiring an exponential amount of post-selection [28]. Therefore, generalizing this framework to other inputs that are more reliably produced has been an important topic of study.

Gaussian Boson Sampling represents one such popular generalization. There, the input states are quadratic, meaning they are generated from the vacuum by some combination of displacement and squeezing (assuming pure input states that have no thermal contribution) [111]. Typically, the displacements are ignored because they do not contribute to entanglement between the modes. Hence, the input states are simply squeezed vacuum states, which are much easier to prepare in a lab than many parallel single-photon states [28]. Much theoretical work has been done to generalize the original statements from Ref. [27] about the computational complexity of sampling in the Fock basis to this Gaussian setting [97–103]. In due course, many labs have performed experiments claiming to show quantum advantage using Gaussian Boson Sampling [30, 31].

Broadly speaking, the hardness of sampling schemes in general, and therefore of both Fock state and Gaussian Boson Sampling, is based on certain statistical properties of the output probability distributions. Fock state Boson Sampling and Gaussian Boson Sampling have output probabilities defined by permanents and hafnians, respectively, which are combinatorial functions mapping matrices over a field to an element of that field. If one treats the input matrix as a weighted adjacency matrix, then the permanent and the hafnian count the number of perfect matchings in the bipartite and generalized weighted graph, respectively, defined by this adjacency matrix [105]. These functions are, in general, difficult to compute. The permanent is #P-hard to compute exactly [104], and this hardness extends to the hafnian because one can encode the permanent of a matrix as the hafnian of a matrix that is twice as big. Even further, Ref. [27] extended this exact hardness to a proof that it is GapP-hard to approximate the modulus squared of the permanent up to inverse polynomial multiplicative error (which similarly extends to the hafnian). However, showing that it is hard to compute or approximate specific output probabilities is not, in and of itself, enough to demonstrate hardness of actually producing a sample from the Fock or Gaussian Boson Sampling distributions; many theoretical tools are needed to show that a difficulty in computing probabilities further implies a difficulty in sampling.

One such crucial tool is called anticoncentration. Anticoncentration is a property of the output distribution that says, roughly, that the outputs are not too clustered on individual probabilities, hence making it more difficult to adequately mimic this distribution in a sampling procedure, and it is commonly used as evidence for approximate average-case hardness of sampling [28]. Anticoncentration is usually proven by analyzing the moments of the outcome probability distribution. In Chapter 4 and Appendix C, we study anticoncentration in the non-collisional limit (where the outcome states are very likely to have at most a single photon in each mode). We develop a graph-theoretic technique to find a closed form for the first moment and a few simple analytical results about the second moment; most saliently, we show that the second moment admits a polynomial expansion in the number of initially squeezed modes, and we derive the leading order in this expansion. These simple results are sufficient to show that there is actually a transition in whether or not anticoncentration holds based on how many of the initial modes are squeezed; when few are squeezed, there is a lack of anticoncentration, but, in the opposite limit, a weak version of anticoncentration holds.

However, the second moment itself deserves a more thorough treatment beyond the few analytic results needed to prove this transition in anticoncentration. For example, linear cross-entropy benchmarking (LXEB) is a tool that has been used to characterize the performance of sampling experiments, most notably in the random circuit sampling experiment of Ref. [29]. It can be shown that the LXEB score that an error-free sampler would achieve when averaged over all possible random networks is precisely given by the second moment of the output probabilities normalized by the square of the first moment. Therefore, a better understanding of the second moment is crucial to achieving a better understanding this popular benchmarking scheme.

To that end, we develop a classically efficient recursion relation that allows us to exactly calculate the second moment up to any desired Fock sector n, which is the main technical contribution of this Chapter. The recursion relation follows from the graph-theoretic approach we introduce in Chapter 4, which we generalize and expand upon here. This approach reduces the algebraic evaluation of the hafnian to simply counting the number of connected components of a certain class of graphs. We then carefully study how higher-order graphs reduce to lower-order ones under certain operations, and the effect that this has on the number of connected components, in order to recursively solve for the second moment. Not only does this allow us to make statements about the average LXEB score for an error-free sampler, but it also allows us to pin down more precisely *where* the aforementioned transition in anticoncentration occurs. If k is the number of initially squeezed modes, we provide strong evidence that this transition occurs at $k = \Theta(n^2)$.

The rest of the Chapter proceeds as follows. In Section 5.2, we provide some background information, set up the system and problem of interest, and briefly summarize our main results. In Section 5.3, we review our results from Chapter 4 and Appendix C; specifically, in Section 5.3.1, we review results about the first moment, and in Section 5.3.2, we discuss how to calculate the second moment. This latter Section sets up the discussion of the recursion in Section 5.4 (though most of the technical details are addressed in Appendices D.1 and D.2). Section 5.5.1 discusses the actual exact numerical evaluation of the recursion. Complementing this, Section 5.5.2 discusses some preliminary analytical results and scaling properties of the second moment. Finally, in Section 5.6, we apply these results to give evidence for the exact location of the transition in anticoncentration we derive in Chapter 4 and Appendix C.

5.2 The Output Distribution of Gaussian Boson Sampling

In this Section, we provide some necessary background information on Gaussian Boson Sampling and set up our system of interest. We also motivate the study of the moments of the output probabilities. Finally, we provide a brief summary of our main results.

5.2.1 Gaussian Boson Sampling

We consider a paradigmatic Gaussian Boson Sampling system on m modes [99,100]. These modes pass through a random sequence of beamsplitters and phase shifters that effect a linear optical (i.e. photon-number-conserving Gaussian) unitary $U \in \mathrm{U}(m)$ and are then measured in the Fock basis (this non-Gaussian operation is necessary for classical hardness of sampling [103]). We consider the typical case where the initial state on the first k modes consists of single-mode squeezed states of equal squeezing parameter r, and the remaining $m - k$ modes are initialized to the vacuum state.

Reference [99] calculates the outcome probability of the Fock measurement of such a system. Given a unitary U, the probability of obtaining an outcome $\boldsymbol{n} = (n_1, n_2, \ldots, n_m) \in \mathbb{N}_0^m$ with total photon count $2n = \sum_{i=1}^m n_i$ is given by

$$P_U(\boldsymbol{n}) = \frac{\tanh^{2n} r}{\cosh^k r} \left|\mathrm{Haf}(U_{1_k,\boldsymbol{n}}^\intercal U_{1_k,\boldsymbol{n}})\right|^2. \tag{5.1}$$

$U_{1_k,\boldsymbol{n}}$ is the $k \times 2n$ submatrix of U corresponding to its first k rows and its columns determined by the nonzero elements of $\boldsymbol{n}$ (appropriately repeated n_i times). Haf refers to the hafnian, which, for a $2n \times 2n$ symmetric matrix A, is

$$\mathrm{Haf}(A) = \frac{1}{n! 2^n} \sum_{\sigma \in S_{2n}} \prod_{j=1}^{n} A_{\sigma(2j-1),\sigma(2j)}, \tag{5.2}$$

with S_{2n} the permutation group on $2n$ elements. We specify that the dimensions of A are even because the hafnian of an odd matrix vanishes; it also vanishes if the input matrix is not symmetric. In our setting, this aligns with the physical fact that single-mode squeezed vacuum states are supported only on even Fock states. The hafnian generalizes the permanent (whose computational complexity controls the hardness of Fock state Boson Sampling) because one can prove that [99]

$$\mathrm{Per}(A) = \mathrm{Haf}\left[\begin{pmatrix} 0 & A \\ A^\intercal & 0 \end{pmatrix}\right]. \tag{5.3}$$

Hence, computing the hafnian is at least as hard as computing the permanent.

We work in the regime where the measured output states are, with high probability, photon-collision-free, which means that the output vector $\boldsymbol{n}$ has $n_i \in \{0,1\}$. That is, $U_{1_k,\boldsymbol{n}}$ has no repeated columns. It suffices for $\mathbb{E}[2n] = k \sinh^2 r = o(\sqrt{m})$ for photon-collision-freeness to hold with high probability. When $n = o(\sqrt{m})$, Ref. [101] provides strong numerical and theoretical evidence that the distribution of submatrices $U_{1_k,\boldsymbol{n}}$ is well-captured by a generalization of the circular orthogonal ensemble (COE):

Conjecture 5.1 (Hiding [101])**.** [2] *For any k such that $1 \leq k \leq m$ and $2n = o(\sqrt{m})$, the distribution of the symmetric product $U_{1_k,\boldsymbol{n}}^\intercal U_{1_k,\boldsymbol{n}}$ of submatrices of a Haar-random $U \in \mathrm{U}(m)$ closely approximates in total variation distance the distribution of the symmetric product $X^\intercal X$ of a complex Gaussian matrix $X \sim \mathcal{N}(0, 1/m)_c^{k\times 2n}$ with mean 0 and variance $1/m$.*

We note that, in Ref. [101], this conjecture is only formulated for the case $n \leq k \leq m$. However, here we allow k to reach 1. The reasoning is that the evidence for Conjecture 5.1 in the regime $k = n$ is based on a proof from Ref. [27] showing that $n \times n$ submatrices of Haar-random unitaries are approximately Gaussian. Clearly the proof must still hold in the case $k < n$ (if $n \times n$

[2] Note added: This is the same as Conjecture 4.1.

submatrices are approximately Gaussian, then so too are smaller submatrices), meaning we can safely extend the conjecture to all $k \leq m$.

Roughly speaking, the intuition behind the conjecture and the original proof of the $k = n$ regime in Ref. [27] is that, if one looks at a small enough submatrix of a unitary, this submatrix no longer "notices" the unitary constraints. Multiplying this small submatrix by its transpose washes out the remaining correlations between elements of the unitary. Hence, the product of the submatrices is approximately the same as a product of i.i.d. Gaussian matrices. Observe also that working in the non-collisional regime, $n \in o(\sqrt{m})$, is crucial for this argument to hold; an output state with more than one photon in a given mode leads to a repeated column/row in the respective submatrix, which, of course, destroys the independence of these elements. In what follows, we work under the assumption that Conjecture 5.1 holds. We are therefore interested in the statistical properties of $X^\top X$ when the elements of X are i.i.d. Gaussian.

5.2.2 Moments of the Gaussian Boson Sampling Distribution and Their Significance

In order to understand the statistical properties of the outcome probabilities of Gaussian Boson Sampling, we must study not just the distribution over individual matrix elements of $X^\top X$, but how they interact with one another through the hafnian. Under Conjecture 5.1 and Eq. (5.1), the outcome probabilities of Gaussian Boson Sampling are given by (up to a prefactor that is mostly irrelevant for our purposes)

$$M_t(k,n) \coloneqq \mathbb{E}_{X \sim \mathcal{G}^{k \times 2n}}\left[|\mathrm{Haf}(X^\top X)|^{2t}\right], \tag{5.4}$$

where we use $\mathcal{G}^{k\times 2n}$ as shorthand for $\mathcal{N}(0,1)_c^{k\times 2n}$ (we consider unit variance for computational simplicity; rescaling X by $1/\sqrt{m}$ leads to another overall prefactor that can be dealt with independently). Specifically, we are most interested in the first and second moments, $t = 1$ and $t = 2$, respectively. We motivate this interest in two ways: the study of anticoncentration and linear cross entropy benchmarking in Gaussian Boson Sampling.

We first recall the framework for anticoncentration established in Chapter 4 and Appendix C.5.1. There, the key definition is p_2, the inverse average collision probability in the output, which, under the hiding conjecture (Conjecture 5.1), is approximately given by the ratio of the square of the first moment to the second moment:

$$p_2(\mathrm{U}(m)) = \frac{\mathbb{E}_{U\in\mathrm{U}(m)}[P_U(\boldsymbol{n})]^2}{\mathbb{E}_{U\in\mathrm{U}(m)}[P_U(\boldsymbol{n})^2]} \approx \frac{M_1(k,n)^2}{M_2(k,n)} =: m_2(k,n). \tag{5.5}$$

We refer to $m_2(k,n)$ as the inverse normalized second moment. Chapter 4 and Appendix C.5.1 use p_2 to define three different classes of anticoncentration:

(A) We say that $P_U, U \in \mathrm{U}(m)$, *anticoncentrates* if $p_2 = \Omega(1)$;

(WA) We say that P_U *anticoncentrates weakly* if $p_2 = \Omega(1/n^a)$ for some $a = O(1)$;

(NA) We say that P_U *does not anticoncentrate* if $p_2 = O(1/n^a)$ for any constant $a > 0$.

Appendix C.5.1 contextualizes these definitions in relation to the approximate average-case hardness necessary for formal hardness of Gaussian Boson Sampling.

We note also that, of course, it is important how precise this approximation in Eq. (5.5) really is. That is, exactly how close in total variation distance the exact and approximate distributions are is important to formalizing the complexity theoretic implications of our work. In particular, if the distribution $U_{1_k,\boldsymbol{n}}^{\top} U_{1_k,\boldsymbol{n}}$ is not close enough in total variation distance to the distribution $X^{\top}X$, then it is not possible to transfer statements about, say, anticoncentration between the two distributions. We address this subtlety in Appendix C.5.2, but, in short, we can formalize and sharpen Conjecture 5.1 such that statements made about anticoncentration of the approximate distribution via m_2 imply anticoncentration of the exact distribution via p_2 as well.

Beyond understanding anticoncentration, calculations of $M_1(k,n)$ and $M_2(k,n)$ also allow one to study linear cross-entropy benchmarking in Gaussian Boson Sampling. Recall that linear cross-entropy benchmarking is a method by which one can compare the outputs of a potentially noisy Gaussian Boson Sampling experiment with the output of a perfect, error-free experiment. Cross-entropy benchmarking was introduced in the context of random circuit sampling in Refs. [112, 113] and later linearized in Ref. [29]. We review this linearized form now, translating from the random circuit sampling language to that of bosonic sampling.

Let $\{\boldsymbol{n}\}$ be the possible output photon strings sampled in some Gaussian Boson Sampling experiment that are produced with respective experimental probabilities $\tilde{P}_U(\boldsymbol{n})$. Let $P_U(\boldsymbol{n})$ be the ideal probabilities for these outputs; that is, these are the probabilities for an output $\boldsymbol{n}$ given by Eq. (5.1). The linear cross-entropy score F_{XEB} for such an experiment is

$$F_{\text{XEB}} = |\Omega_{2n}| \sum_{\boldsymbol{n}\in\Omega_{2n}} P_U(\boldsymbol{n})\tilde{P}_U(\boldsymbol{n}) - 1, \tag{5.6}$$

where Ω_{2n} is the non-collisional sample space with $2n$ output photons in m modes. If the noisy outputs are correct, i.e. the experiment is error-free, then $\tilde{P}(\boldsymbol{n}_i) = P(\boldsymbol{n})$. The ideal cross-entropy score, then, is

$$F_{\text{XEB}}^{\text{ideal}} = |\Omega_{2n}| \sum_{\boldsymbol{n}\in\Omega_{2n}} P_U(\boldsymbol{n})^2 - 1. \tag{5.7}$$

The expected value of the ideal cross-entropy over all possible unitaries is, therefore,

$$\mathbb{E}_{U\in U(m)}[F_{\text{XEB}}^{\text{ideal}}] = |\Omega_{2n}| \sum_{\boldsymbol{n}\in\Omega_{2n}} \mathbb{E}_{U\in U(m)}[P_U(\boldsymbol{n})^2] - 1. \tag{5.8}$$

Assuming that one operates in the hiding regime, then two facts are true: first, $|\Omega_{2n}| \sim M_1(k,n)$; second, $\mathbb{E}_{U\in U(m)}[P_U(\boldsymbol{n})^2]$ is independent of $\boldsymbol{n}$ (see Appendix C.4 for more details). Therefore,

$$\mathbb{E}_{U\in U(m)}[F_{\text{XEB}}^{\text{ideal}}] = \frac{M_2(k,n)}{M_1^2(k,n)} - 1 = m_2(k,n)^{-1} - 1. \tag{5.9}$$

Thus, anticoncentration and the expected ideal linear cross-entropy benchmarking score both depend on this inverse average collision probability. Therefore, a precise calculation of the second moment beyond asymptotics is valuable to a more fine-grained understanding of both anticoncentration and cross-entropy benchmarking.

5.2.3 Summary of Results

We now come to a brief summary of our main results.

In Chapter 4 and Appendix C, we develop a graph-theoretic formalism that allows us to derive various analytic properties of the first and second moments, $M_1(k,n)$ and $M_2(k,n)$. We use this formalism to find a closed form expression for $M_1(k,n)$ and to show that $M_2(k,n)$ admits a polynomial expansion in k; we also calculate the leading order of this expansion. This allows us to show the transition in anticoncentration. We review these results in more depth in Section 5.3.

In this Chapter, we significantly expand upon this graph-theoretic formalism and derive an efficiently evaluable recursion relation that allows us to numerically exactly calculate all coefficients of the polynomial expansion of the second moment. We then apply this algorithm and calculate these expansions up to photon sector $2n = 80$. In the photon-non-collisional regime, where $n \in o(\sqrt{m})$, this corresponds to approximately 6400 modes, which is well beyond the current state-of-the-art experiments. Therefore, the technique that we develop in this Chapter yields results that can help characterize the output distribution of any near-term Gaussian Boson Sampling experiment. The recursion is developed in Section 5.4, with details about its efficiency and construction deferred to Appendices D.1 and D.2, respectively.

We then discuss some simple analytic results about the scaling of the second moment in Section 5.5.1. We follow this with substantial numerical investigation of the results of the recursion up to $2n = 80$ in Section 5.6. In particular, we are able to give strong evidence that the transition in anticoncentration occurs at $k = \Theta(n^2)$. We accomplish this with numerical plots of $m_2(k,n)$, the quantity that controls anticoncentration, when k scales polynomially with n. We also provide a brief analytic argument that this transition occurs somewhere between $k = \Omega(n)$ and $k = O(n^2)$.

This result, along with the fact that we operate in the conjectured hiding regime where $2n = o(\sqrt{m})$ and $k \leq m$, implies concrete advice for experimental demonstrations of quantum advantage via Gaussian Boson Sampling. Namely, one should squeeze all m modes with squeezing parameter $\sinh^2 r = o(m^{-1/2})$.

5.3 Graph-Theoretical Analysis of Gaussian Boson Sampling Moments

In this Section, we lay out the graph-theoretic framework for analyzing the moments of Gaussian Boson Sampling output probabilities. This is a review of the same framework we develop in Chapter 4 and Appendix C. We first briefly recall the derivation of the closed form of the first moment $M_1(k,n)$, and we follow this with a discussion of how an extension of this framework also allows us to analyze the second moment $M_2(k,n)$.

5.3.1 First Moment

In this Section, we discuss the first moment of the output probabilities, which is, up to some multiplicative factors, $\underset{X\sim\mathcal{G}^{k\times 2n}}{\mathbb{E}}\left[\left|\mathrm{Haf}(X^\intercal X)\right|^2\right]$. We calculate and analyze this moment in Chapter 4 and Appendix C.1, but we review the key elements of that discussion because they are a useful point of reference for the calculation of the second moment.

Using the definition of the hafnian and properties of the expectation value of complex Gaussians, we reduce the first moment to a sum over Kronecker δs:

$$M_1(k,n) = \frac{(2n)!}{(2^n n!)^2} \sum_{\tau\in S_{2n}} \sum_{\{o_i\}_{i=1}^{n}}^{k} \prod_{j=1}^{n} \delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}}. \tag{5.10}$$

We ascribe a graph-theoretic interpretation to this equation; see Fig. 5.1[3] for an example. Each permutation τ instantiates a graph G_τ on $2n$ vertices labeled O_1 to O_{2n} with edges defined by two perfect matchings: one fixed black set of edges, and one set of red edges determined by τ. More specifically, each index o_j in the sum splits into two vertices O_ℓ and $O_{\ell'}$ such that

[3] Note added: This is a recreation of Fig. 4.2(a).

$\lceil \tau(\ell)/2 \rceil = j = \lceil \tau(\ell')/2 \rceil$ (that is, $o_{\lceil \tau(\ell)/2 \rceil}$ maps to a vertex O_ℓ). One perfect matching consists of black edges between O_{2j-1} and O_{2j} for all $j \in [n] := \{1, 2, \ldots, n\}$; these edges enforce that $o_{\lceil \tau(2j-1)/2 \rceil}$ and $o_{\lceil \tau(2j)/2 \rceil}$ are linked by a Kronecker δ. The other perfect matching has red edges between O_ℓ and $O_{\ell'}$ if $\lceil \tau(\ell)/2 \rceil = \lceil \tau(\ell')/2 \rceil$; these edges ensure that there is an edge between the ℓ, ℓ' mapped to the same value under τ and the ceiling function, meaning the vertices arose from the same lower-case-o index.

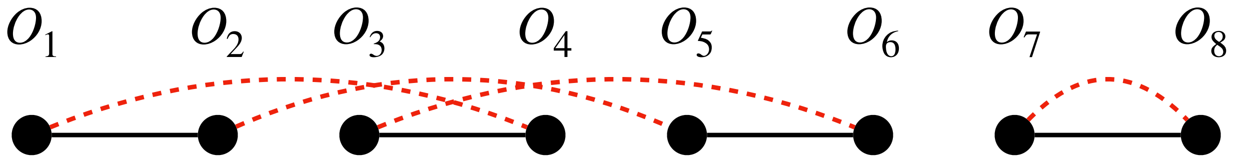

Figure 5.1: Graph $G \in \mathbb{G}^1_n$. One of $2^n n!$ permutations that induces this graph is $\tau = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 3 & 5 & 2 & 4 & 6 & 8 & 7 \end{smallmatrix}\right)$. This graph has two connected components, therefore contributing k^2 to the first moment.

This definition of G_τ ensures that the number of connected components of G_τ, $C(G_\tau)$, is equivalent to the number of unconstrained indices in the interior sum in Eq. (5.10), and, hence, the number of factors of k that τ contributes overall. Therefore,

$$M_1(k, n) = \frac{(2n)!}{(2^n n!)^2} \sum_{\tau \in S_{2n}} k^{C(G_\tau)}. \tag{5.11}$$

We simplify this expression using a degeneracy whereby $2^n n!$ different τ all induce the same final graph; the factor of $n!$ corresponds to choosing which tuple $(2j-1, 2j)$ corresponds to which index $\lceil \tau(\ell)/2 \rceil = \lceil \tau(\ell')/2 \rceil$, and the factor of 2^n comes from ordering within each tuple. Therefore, we study only these final sets of graphs, which we label $\mathbb{G}^1_n$ (1 refers to the first moment, and n indexes the order). We study the connected components of graphs in $\mathbb{G}^1_n$ by writing down a recursion relation in n and k that, when solved, yields Theorem 4.1.

Theorem 5.1. [4] *The sum over graphs in $\mathbb{G}_n^1$ satisfies*

$$\sum_{G \in \mathbb{G}_n^1} k^{C(G)} = k(k+2)\dots(k+2n-2), \tag{5.12}$$

and hence $M_1(k,n) = (2n-1)!!(k+2n-2)!!/(k-2)!!$.

To summarize: Eq. (5.10) gives an expression for the first moment of the outcomes of Gaussian Boson Sampling probabilities in terms of sums of products of Kronecker δs. We then reinterpret this as counting the number of connected components of a certain type of graph with two perfect matchings. We solve this counting problem by developing and evaluating a recursion relation. We use the same overall technique to calculate the second moment, as we explain in the next Section.

5.3.2 Second Moment

We now move on to analyzing the second moment of the output probabilities. Using similar techniques as described for the first moment, in Chapter 4 and Appendix C.2 we derive an

[4] Note added: This is a restatement of Theorem 4.1.

expression for the second moment that is equivalent to Eq. (5.10):[5]

$$M_2(k,n) \coloneqq \mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\left|\mathrm{Haf}(X^\top X)\right|^4\right] = \left(\frac{1}{2^n n!}\right)^4 (2n)! \sum_{\tau,\alpha,\beta\in S_{2n}} \sum_{\{\ell_i,o_i,p_i\}_{i=1}^{n}=1}^{k} \left[\prod_{j=1}^{n}\right.$$
$$\left(\delta_{o_{\lceil\frac{\tau(2j-1)}{2}\rceil} o_{\lceil\frac{\tau(2j)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j-1)}{2}\rceil} q_{\lceil\frac{\beta(2j-1)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j)}{2}\rceil} q_{\lceil\frac{\beta(2j)}{2}\rceil}} + \delta_{o_{\lceil\frac{\tau(2j-1)}{2}\rceil} q_{\lceil\frac{\beta(2j)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j-1)}{2}\rceil} q_{\lceil\frac{\beta(2j-1)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j)}{2}\rceil} o_{\lceil\frac{\tau(2j)}{2}\rceil}} +\right.$$
$$\left.\left.\delta_{q_{\lceil\frac{\beta(2j-1)}{2}\rceil} o_{\lceil\frac{\tau(2j)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j-1)}{2}\rceil} o_{\lceil\frac{\tau(2j-1)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j)}{2}\rceil} q_{\lceil\frac{\beta(2j)}{2}\rceil}} + \delta_{q_{\lceil\frac{\beta(2j-1)}{2}\rceil} q_{\lceil\frac{\beta(2j)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j-1)}{2}\rceil} o_{\lceil\frac{\tau(2j-1)}{2}\rceil}} \delta_{p_{\lceil\frac{\alpha(2j)}{2}\rceil} o_{\lceil\frac{\tau(2j)}{2}\rceil}}\right)\right]. \tag{5.13}$$

. The main differences between Eq. (5.13) and Eq. (5.10) are threefold:

1. We sum over three permutations (instead of a single one) labeled τ, α, β;

2. There are now $3n$ indices to sum over, $\{o_i, q_i, p_i\}_{i=1}^{n}$, instead of just the n given by $\{o_i\}_{i=1}^{n}$;

3. Each factor is a sum of four possible terms instead of just one.

However, this expression still possesses a natural graph-theoretic interpretation, as we now review. See Fig. 5.2 [6] for an example graph as a guide to the following discussion.

Each index in $\{o_i, q_i, p_i\}_{i=1}^{n}$ is again split into two graph vertices $\{O_i, Q_i, P_i\}_{i=1}^{2n}$ that are placed into $2n$ columns and three rows labeled o, p, and q, respectively. As for the first moment, we define two perfect matchings on these vertices given by black and red edges. The black edges are between vertices whose labels are linked under the Kronecker δs, and the red edges connect graph vertices that came from the same original summation index.

More specifically, consider fixing a set of three permutations τ, α, β. There is a red edge between O_ℓ and $O_{\ell'}$ if $\lceil\tau(\ell)/2\rceil = \lceil\tau(\ell')/2\rceil$. An analogous statement holds for P and Q vertices,

[5] Note added: This is Eq. (C.24).

[6] Note added: This is a recreation of Fig. 4.2(b).

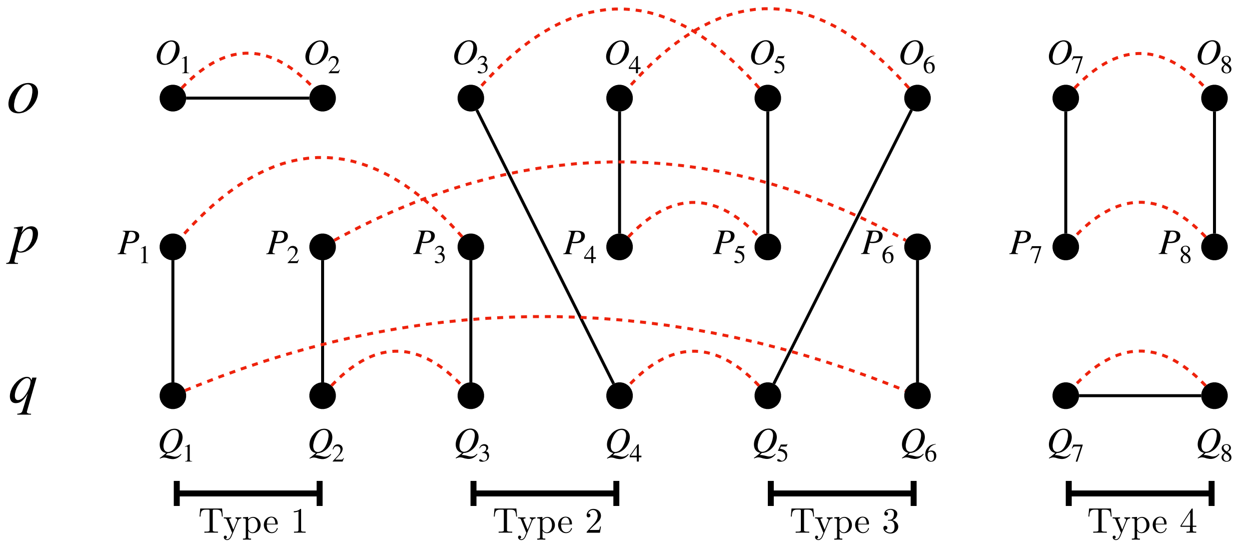

Figure 5.2: Example graph on $n = 4$ used in the calculation of the second moment. Each of the four possible sets of black edges are shown. An example of three permutations that would induce this graph is: $\tau = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 2 & 4 & 5 & 3 & 6 & 8 & 7 \end{smallmatrix}\right)$, $\alpha = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 8 & 6 & 7 & 3 & 4 & 5 & 1 & 2 \end{smallmatrix}\right)$, and $\beta = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 8 & 5 & 6 & 2 & 1 & 7 & 3 & 4 \end{smallmatrix}\right)$. This graph has 5 connected components, so it contributes k^5 to the second moment.

though one uses permutations α and β, respectively, instead of τ. Note that this implies that red edges are always contained within a single row. Now, the black edges are slightly more complicated. There is only a single Kronecker δ term in each factor in the product Eq. (5.10), meaning there is only a single set of black edges for the graphs in $\mathbb{G}_n^1$. However, because the second moment as expressed in Eq. (5.13) contains factors with four Kronecker δ terms, each value of $j \in [n]$ can lead to one of four different patterns of black edges on columns $2j - 1$ and $2j$. We refer to these patterns of black edges on a single pair of adjacent columns as type-1, type-2, type-3, and type-4; see Fig. 5.2 for an example graph that has one of each type. The Kronecker δ

terms and their corresponding black edges, listed in order from type-1 to type-4, are given by

$$\delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \to \{(O_{2j-1}, O_{2j}), (P_{2j-1}, Q_{2j-1}), (P_{2j}, Q_{2j})\}, \tag{5.14}$$

$$\delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \to \{(O_{2j-1}, Q_{2j}), (P_{2j-1}, Q_{2j-1}), (O_{2j}, P_{2j})\}, \tag{5.15}$$

$$\delta_{q_{\lceil \frac{\beta(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \to \{(O_{2j}, Q_{2j-1}), (P_{2j-1}, O_{2j-1}), (P_{2j}, Q_{2j})\}, \tag{5.16}$$

$$\delta_{q_{\lceil \frac{\beta(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \to \{(O_{2j-1}, P_{2j-1}), (O_{2j}, P_{2j}), (Q_{2j-1}, Q_{2j})\}. \tag{5.17}$$

Because there are four patterns of black edges per pair of adjacent columns, and n such pairs, there are 4^n possible arrangements of black edges on the entire graph. We label these arrangements by an integer $z \in [4^n]$, and we label a graph as $G_{\tau,\alpha,\beta}(z)$.

Analogously to the first moment, we can rewrite the sum over products of Kronecker δs in Eq. (5.13) as a sum over these graphs, where $G_{\tau,\alpha,\beta}(z)$ contributes a factor of k raised to its number of connected components. Therefore, Eq. (5.13) becomes

$$M_2(k,n) = \frac{(2n)!}{(2^n n!)^4} \sum_{\substack{\tau,\alpha,\beta \in S_{2n} \\ z \in [4^n]}} k^{C(G_{\tau,\alpha,\beta}(z))}. \tag{5.18}$$

There is again a degeneracy where many permutations all lead to the same set of red edges in a given row, and, hence, the same graph. Specifically, this degeneracy is again $2^n n!$, but for each copy of S_{2n}. We can therefore again ignore the permutations and look only at the underlying graphs. For any given z, we define $\mathbb{G}_n^2(z)$ to be the graphs on $6n$ vertices with two perfect matchings: the zth set of black edges and red edges that pair vertices in the same row. We then define $\mathbb{G}_n^2 = \cup_{z\in[4^n]}\mathbb{G}_n^2(z)$. Thus, accounting for the described degeneracy and these definitions, we get

$$M_2(k,n) = (2n-1)!! \sum_{G\in\mathbb{G}_n^2} k^{C(G)}. \tag{5.19}$$

This implies the following theorem.

Theorem 5.2. [7] *The second moment $M_2(k,n)$ is a degree-$2n$ polynomial in k and can be written as $M_2(k,n) = (2n-1)!!\sum_{i=1}^{2n} c_i k^i$, where c_i is the number of graphs $G \in \mathbb{G}_n^2$ that have i connected components.*

Our goal, then, is to determine these coefficients c_i. It is possible to directly compute c_{2n} and c_{2n-1}, that is, the number of graphs $G \in \mathbb{G}_n^2$ with $2n$ or $2n-1$ connected components, respectively (see Appendix D.3). However, these calculations do not easily generalize to the other c_i. Therefore, we take a different approach, which is to derive a recursion relation that is similar in spirit to the one we use to compute the first moment.

5.4 Recursion for the Second Moment

We now move on to the recursion relation that builds the c_i for larger n from those of smaller n. It is useful to refer to Fig. 5.3 for the following discussion. We are interested in the connected components of the graphs in $\mathbb{G}_n^2$, and the number of connected components does not change if one takes a graph and then "collapses" vertices that are connected via an edge into a

[7] Note added: This is a restatement of Theorem 4.2.

single larger vertex. The graphs that we have defined for the second moment are composed of $2n$ columns of 3 vertices each. Therefore, if one performs this collapsing operation on all of the vertices in, say, the first two columns, this converts a graph with $2n$ columns into one with $2n - 2$ columns. Let us refer to these first two columns as $\mathbb{C}_{1,2}$; that is, $\mathbb{C}_{1,2} = \{O_1, O_2, P_1, P_2, Q_1, Q_2\}$. Two facts follow from the approach we have just described: (1) there are only a finite number of ways that the two columns can connect into the rest of the graph, (2) if one "integrates out" $\mathbb{C}_{1,2}$ by collapsing all of the vertices, one can write the number of connected components of the original graph as the sum of the remaining connected components plus the number of connected components contained entirely within $\mathbb{C}_{1,2}$. This is a generalization of the approach used to prove Theorem 5.1.

However, this recursion is substantially more complicated than the one we use to calculate the first moment, as illustrated in Fig. 5.3. In particular, we must generalize the types of graphs that we consider in order to build a recursion that "closes" on itself, that is, to build a recursion that consistently produces valid graphs. Consider the graphs that we have described so far in the context of this "integration" procedure whereby sets of vertices are collapsed onto one another. As stated, this procedure can induce a graph with red edges that cross between rows, which is not allowed in our current formulation. In Fig. 5.3, the first figure shows a graph in $\mathbb{G}_4^2$ where $\mathbb{C}_{1,2}$ is integrated out, as denoted by the hashing, and the second figure depicts the consequence of this integration. Consider the path P_3—P_1—Q_1—Q_6 that passes through the first column. Collapsing the vertices P_1 and Q_1 into P_3 and Q_3, respectively, does not change the number of connected components, but it induces an edge P_3—Q_6 that is heretofore unallowed because it crosses between rows 2 and 3. Therefore, the newly induced graph is not an element of $\mathbb{G}_3^2$, hence why we must generalize what kinds of graphs we consider.

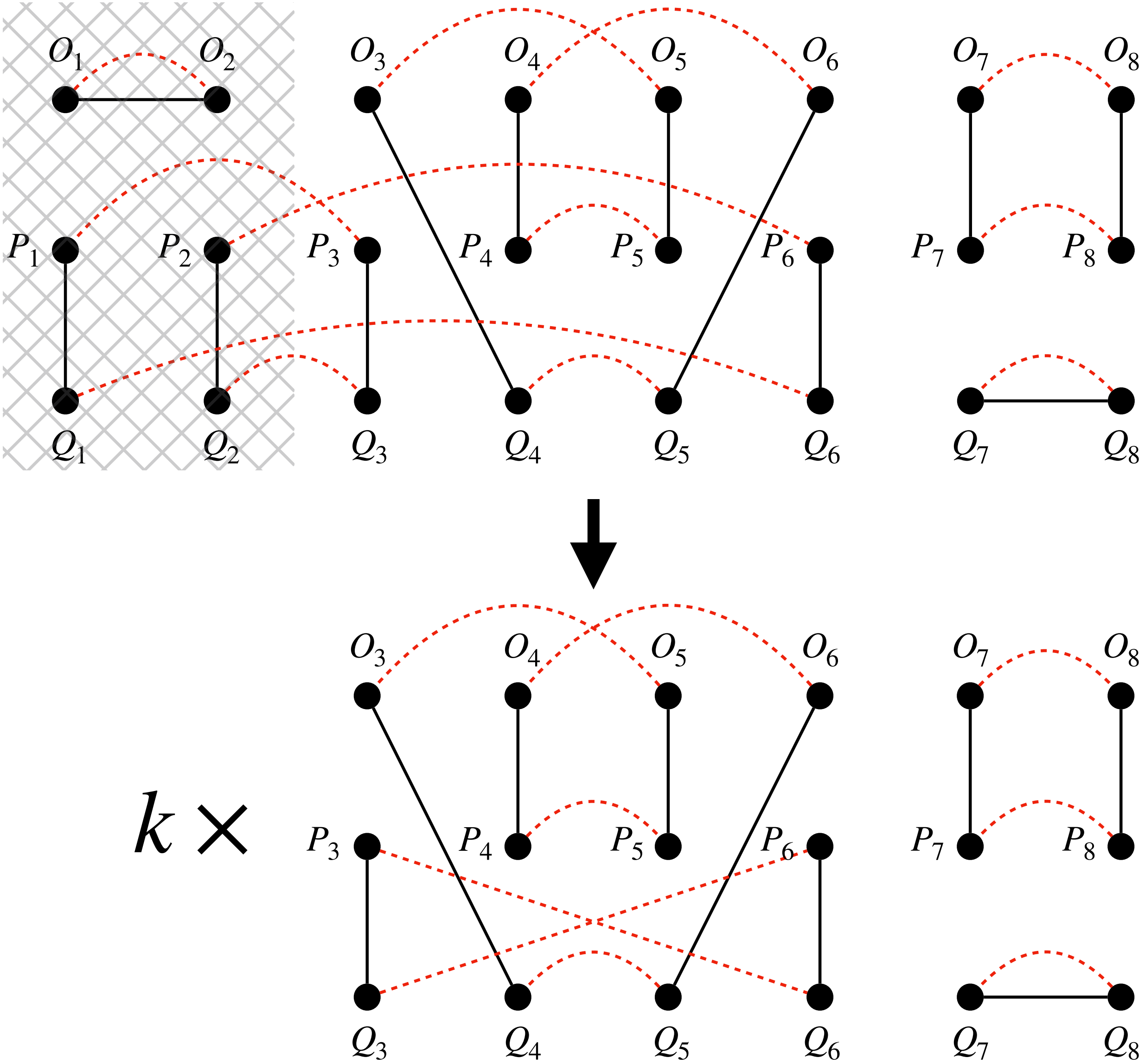

Figure 5.3: Example showing why the simple graphs with red edges that do not cross between rows are not sufficient to develop a recursion. Trying to "integrate out" or collapse the edges that connect the six vertices in the leftmost two columns (represented here with a crosshatch pattern over those vertices) induces a multiplicative factor of k due to the connected component O_1—O_2 as well as red edges P_3—Q_6 and P_6—Q_3. Such red edges are not allowed for graphs in $\mathbb{G}_n^2$, so we must expand the set of graphs we consider in the recursion.

To that end, we define a simple generalization of our graphs, where we allow *all* possible perfect matchings of red edges across the $6n$ vertices. That is, we no longer restrict red edges to connect only vertices of the same letter (i.e., in the same row); we now allow the red edges to cross between two different rows. However, we still demand that each vertex still possess exactly one red edge.

Let a_{12}, a_{13}, a_{23} be the number of edges that span between the first and second, first and third, and second and third rows, respectively. We can then define a set of graphs $\mathbb{G}_n^2(a_{12}, a_{13}, a_{23}, z)$ on $6n$ vertices, where the z again indexes the 4^n possible sets of black edges. We can again write $\mathbb{G}_n^2(a_{12}, a_{13}, a_{23}) = \bigcup_{z \in [4^n]} \mathbb{G}_n^2(a_{12}, a_{13}, a_{23}, z)$. Finally, then, we have

$$g(n, a_{12}, a_{13}, a_{23}) := \sum_{\lambda \in \mathbb{G}_n^2(a_{12}, a_{13}, a_{23})} k^{C(\lambda)} \tag{5.20}$$

The second moment we desire is then, of course, proportional to $g(n, 0, 0, 0)$.

A few constraints on a_{12}, a_{13}, a_{23} are apparent immediately:

- $a_{12} + a_{13}$, $a_{12} + a_{23}$, and $a_{13} + a_{23}$ (that is, the number of edges coming out of the first, second, and third row, respectively) must be even;

- $a_{12} + a_{13}, a_{12} + a_{23}, a_{13} + a_{23}$ must all be less than or equal to $2n$ (there cannot be more than $2n$ edges coming out of a row with only $2n$ vertices given that there is exactly one red edge incident on every vertex).

We also observe that, while we do not explicitly keep track of these edges, we can also define a_{11}, a_{22}, a_{33} as the number of "proper" edges that map between vertices in the first, second, and third rows, respectively. These edges have a simple relationship to the ones we do keep track of that can be derived by simply counting how many vertices in a given row are left after subtracting those that are used in edges that cross between rows:

$$a_{11} = \frac{2n - a_{12} - a_{13}}{2}, \tag{5.21}$$

$$a_{22} = \frac{2n - a_{12} - a_{23}}{2}, \tag{5.22}$$

$$a_{33} = \frac{2n - a_{13} - a_{23}}{2}. \tag{5.23}$$

Because we have the constraints that $a_{12} + a_{13}, a_{12} + a_{23}, a_{13} + a_{23}$ must all be even, a_{11}, a_{22}, a_{33} are all integral. Also, the fact that $a_{12} + a_{13}, a_{12} + a_{23}, a_{13} + a_{23}$ must all be less than or equal to $2n$ ensures that a_{11}, a_{22}, a_{33} are all non-negative as well.

It is also useful to write down the total number of graphs of each type. There are $6n$ total vertices, and $2n$ in each row. Given a vector $\boldsymbol{a} = (a_{12}, a_{13}, a_{23})$, we need to choose a_{12} vertices in row 1 and row 2 to link to one another, a_{13} in rows 1 and 3 (with no overlap between the vertices chosen in the first row corresponding to a_{12} vs. a_{13}), and a_{23} in rows 2 and 3 (again, no overlap with previously chosen vertices is allowed). Once these vertices are chosen, it also remains to choose how to connect them. Finally, one must pair off the remaining vertices in each row, then multiply by 4^n to account for the black edges. The result is

$$\begin{aligned} |\mathbb{G}_n^2(a_{12}, a_{13}, a_{23})| = &\binom{2n}{a_{12}}\binom{2n - a_{12}}{a_{13}}\binom{2n}{a_{12}}\binom{2n - a_{12}}{a_{23}}\binom{2n}{a_{13}}\binom{2n - a_{13}}{a_{23}} a_{12}! a_{13}! a_{23}! \\ &\times (2n - a_{12} - a_{13} - 1)!!(2n - a_{12} - a_{23} - 1)!!(2n - a_{13} - a_{23} - 1)!! 4^n. \end{aligned} \tag{5.24}$$

This result is useful because, if one sets $k = 1$ in Eq. (5.20), then every graph is put on equal footing; that is, any number of connected components contributes equally to the sum. Therefore,

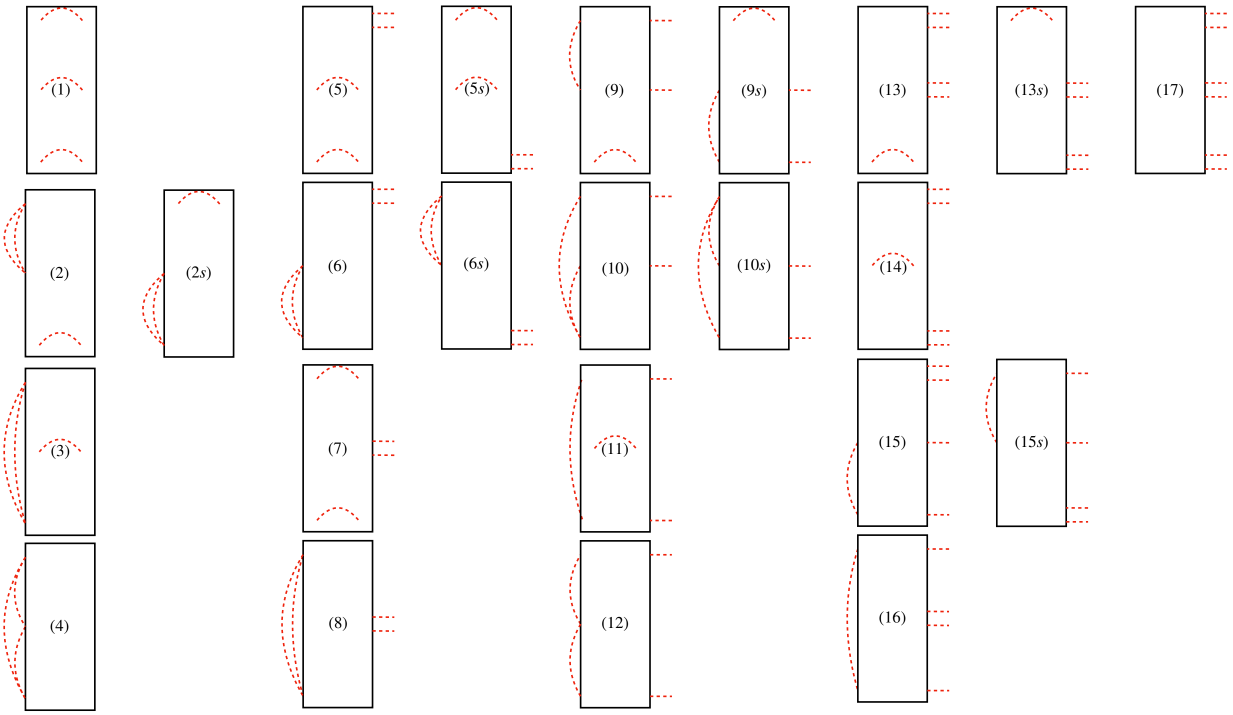

Figure 5.4: List of 17 cases (up to symmetry) for how the first two columns in a graph of order n can connect into the rest of the graph.

given a polynomial expansion in k for any $g(n, a_{12}, a_{13}, a_{23})$ (note that Theorem 5.2 still holds for the generalized graphs, except the highest order term need not be $2n$ anymore—generically it can reach $3n$), Eq. (5.24) gives the sum of the coefficients on the monomials.

We now describe the recursion using the following equation

$$g(n, a_{12}, a_{13}, a_{23}) = \sum_{b_{12}, b_{13}, b_{23}} c(a_{12}, a_{13}, a_{23}, b_{12}, b_{13}, b_{23}) g(n-1, b_{12}, b_{13}, b_{23}). \tag{5.25}$$

The goal is to determine the coefficients c for all valid sets a_{12}, a_{13}, a_{23} allowed by n. In order to do this, one must effectively determine the various ways integrating out two columns changes the possible red edge configurations.

Specifically, there are 17 ways (24 if one disambiguates symmetric cases) in which $\mathbb{C}_{1,2}$ can attach into a graph of order $n-1$ (that is, one with $2n-2$ columns). We classify these by the number of red edges that "protrude" from $\mathbb{C}_{1,2}$. We illustrate these cases in Fig. 5.4 and now describe how to interpret these images.

Red edges that attach within a row inside the block are fixed, as there is only one possible edge that can connect two vertices in the same row. We depict the red edges that connect $\mathbb{C}_{1,2}$ to the rest of the graph as protruding from the same row on the right side of the block, as shown in Fig. 5.4. We depict red edges that go between different rows in $\mathbb{C}_{1,2}$ on the left of the box, again shown in Fig. 5.4. We do not draw the four possible sets of black edges within the block, but understanding their effect is crucial to the actual mechanics of the recursion.

We must determine how each of these cases leads to a relationship between $\boldsymbol{a}$ and $\boldsymbol{b}$, as well as the coefficient c in Eq. (5.25), which is related to the number of possible graphs of order n that, when integrated out, lead to *the same* graph at order $n-1$. The coefficient out front is also affected by how many internal loops the given case has, as that of course leads to extra connected components that yield factors of k. There are overall three different contributions to $c(\boldsymbol{a},\boldsymbol{b}) \coloneqq c(a_{12},a_{13},a_{23},b_{12},b_{13},b_{23})$:

- Loop: This corresponds to the number of connected components in $\mathbb{C}_{1,2}$. This is the easiest contribution to determine;

- Vectorial: This corresponds to the relationship between $\boldsymbol{a}$ and $\boldsymbol{b}$ in Eq. (5.25), and, while somewhat simple in spirit, it often requires significant casework. In short, when integrating out $\mathbb{C}_{1,2}$, one loses contributions from internal edges that are lost by collapsing the vertices, but one gains edges of the types that are induced between the remaining vertices;

- Combinatorial: This corresponds to the combinatorial factors that are associated with how many ways a given case leads to the same graph at lower order. This depends both on the number of protruding edges and how the red and black edges interact via the vertices in $\mathbb{C}_{1,2}$.

With these ideas set forth, the evaluation of the recursion proceeds as follows. We first evaluate the base cases when $n = 1$. We then determine the loop, vectorial, and combinatorial contributions to each of the 17 cases depicted in Fig. 5.4, thus determining how that case contributes to the overall recursion. Finally, we evaluate the recursion numerically exactly, which is classically efficient (see Section 5.5.1 and Appendix D.1 for details). Note that, while it is, in principle, possible to write down analytically the contribution of each of the 17 cases depicted in Fig. 5.4, the terms are sufficiently numerous and complicated that we could not actually solve the recursion analytically; for more details, see Appendix D.2, where the loop, vectorial, and combinatorial contributions are worked out for the cases.

5.5 Analysis of the Second Moment

In this Section, we analyze the results derived from the numerically exact evaluation of the recursion described in the previous Section. Specifically, we first discuss the code behind the recursion and provide some checks to gain confidence that code is accurate. We then derive some analytic results upper and lower bounding the second moment, which we then compare to the numerically exact data to understand how well they capture the scaling of the second moment.

5.5.1 Numerical Evaluation of the Recursion

Once the theoretical principles behind the recursion in Eq. (5.25) are developed, we simply account for the contributions from each case and evaluate the recursion numerically exactly. We accomplish this using the Julia programming language [114] and find $g(n,0,0,0)$ from $n = 1$ to $n = 40$ (which, recall, means up to photon sector 80).

We now briefly describe our implementation of the exact numerical recursion; the code is available on GitHub [115]. As a consequence of Eq. (5.24), the polynomial coefficients in $g(n,a_{12},a_{13},a_{23})$ grow at most factorially, so the number of bits needed to store the integers grows polynomially. Therefore, to ensure exact accuracy of all of the integer calculations, we use Julia's `BigInt` type, which allows us to achieve arbitrary-precision arithmetic [114]. Next, in order to avoid performing slow symbolic arithmetic operations, we represent polynomials in k as `BigInt` arrays, where the i^{th} element of the array corresponds to the coefficient in front of the k^i term in the polynomial. Multiplication and addition of polynomials in k is then done at the array level. We begin with $n = 1$ and store the base case values of $g(1,a_{12},a_{13},a_{23})$ given in Appendix D.2.1. To compute the value of $g(n,a_{12},a_{13},a_{23})$, we iterate through the 17 cases described in Appendix D.2 and compute the various combinatorial factors and values of b_{12},b_{13},b_{23} that show up in the sum in Eq. (5.25). We then recursively compute the values of $g(n-1,b_{12},b_{13},b_{23})$. The algorithm utilizes memoization every time any value of $g(n,a_{12},a_{13},a_{23})$ is computed so that the recursion rarely needs to go particularly deep. In the end, in order to compute up to $g(40,0,0,0)$, we compute $g(n,a_{12},a_{13},a_{23})$ for around 50 000 combinations of arguments, resulting in almost 200 megabytes of (uncompressed) data.

As mentioned, the evaluation of the recursion is classically efficient. In short, the number of allowed $\boldsymbol{a}$ (i.e. those that satisfy the necessary bounds and parity constraints) is polynomially bounded, the size of the coefficients cannot be more than factorially large (meaning they can be stored with polynomial space), and the array-based multiplication and addition is classically tractable. More details are presented in Appendix D.1.

We can check the computed values of $g(n, a_{12}, a_{13}, a_{23})$ derived via the recursion in a few ways. First, we note again that for any value of $g(n, a_{12}, a_{13}, a_{23})$, setting $k = 1$ (i.e., summing the coefficients in front of each monomial) yields the total number of graphs of this type, which is given in Eq. (5.24). Furthermore, Lemma 5.1(ii)[8] (to be introduced below) gives the coefficient in front of the leading order term in $g(n, 0, 0, 0)$. Our numerically exact computation of these numbers using the recursion matches these predicted values.

Second, for various n and k, we numerically sample 10^5 random $X \in \mathcal{G}^{k\times 2n}$, compute $|\mathrm{Haf}[X^\top X]|^4$ using the code provided by Ref. [116], and average the results. This gives a numerical approximation to $(2n-1)!!g(n, 0, 0, 0)$. We perform this calculation for $n, k \in \{1, 2, \ldots, 9\}$. The result is shown in Fig. 5.5, and we see good agreement between the approximate numerical calculations (data points and error bars) and the theoretical values predicted by the recursion (solid lines).

5.5.2 Scaling of the Second Moment

While we have not found a closed form for the solution to the recursion, we are able to derive a few simple analytic results about the values of the coefficients of the polynomial expansion as well as the overall scaling of the second moment. The former are covered in Appendix C.3, as they are crucial to demonstrating the transition in anticoncentration that is the central result of Chapter 4. The latter are new to this Chapter.

We recall Lemma 4.1.

Lemma 5.1 *We have that*

[8] Note added: This is the same as Lemma 4.1.

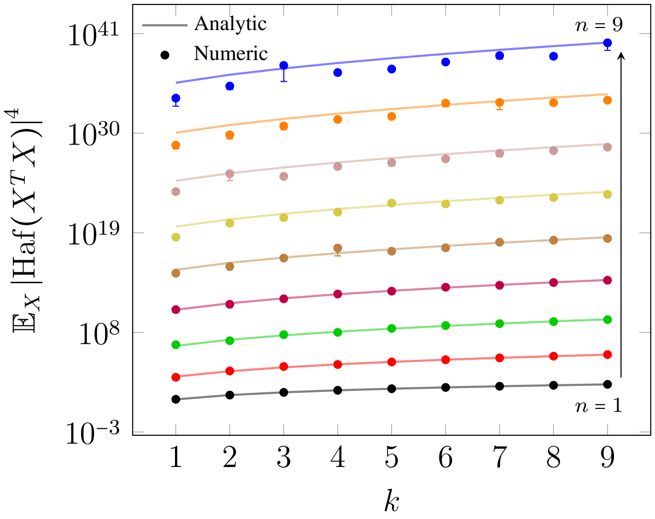

Figure 5.5: Numerical test of recursion. The x-axis represents k, and the y axis represents $\underset{X\sim\mathcal{G}^{k\times 2n}}{\mathbb{E}}[|\mathrm{Haf}(X^\intercal X)|^4]$. Solid lines, from $n = 1$ through $n = 9$ are the theoretical predictions derived from the recursion relation (see [115] for the code). Dots and bars represent the expected value and standard error, respectively, estimated by sampling 10^5 random Gaussian matrices and computing the second moment using the code provided by Ref. [116]. Note that, for many points, the size of the error bar is smaller than its associated dot. Further, there is an asymmetry in the error bars due to the log nature of the plot. We see excellent alignment between theory and numerics for $n = 1$ through $n = 5$. For larger n, the agreement is still good, but we seem to undersample the true value in many cases. We suspect that this is because the distribution of the second moment has a long tail, meaning we do not suspect that the given error bars are indicative of the true difference between the sampled and numerically exact data. We believe that were we able to either take sufficiently more samples we would see stronger agreement between the sampled and true means, but this option is too computationally demanding given the size of the matrices involved and the exponential complexity of classically computing the hafnian [109].

i. $M_2(1,n) = ((2n-1)!!)^4 4^n$*;*

ii. $c_{2n} = (2n)!!$*.*

The proof of part (i) consists of a direct calculation using Eq. (5.13); it also follows from the graph-theoretic framework by simply counting the number of possible graphs of type $\boldsymbol{a} = (0,0,0)$ (see Eq. (5.24)). The proof of part (ii) follows from a reduction of the problem of counting connected components to a special case of the first moment using $k = 2$. We also reprove this result in a slightly different way in Appendix D.3.

As a corollary of Lemma 5.1, we can derive upper and lower bounds on the second moment:

Lemma 5.2 *Lemma 5.1 implies*

$$M_2(k,n) \leq (2n-1)!!^4 4^n k^{2n}, \tag{5.26}$$

$$M_2(k,n) \geq (2n)! k^{2n}, \tag{5.27}$$

$$M_2(k,n) \geq (2n-1)!!^4 4^n. \tag{5.28}$$

Proof. We first prove the upper bound. The leading term in $g(n,0,0,0)$ is of the form k^{2n}, and the total number of graphs with no red edges crossing between rows is $(2n-1)!!^3 4^n$. Thus, the upper bound comes from saying that all graphs have $2n$ connected components.

We next prove the lower bounds. The first lower bound comes from considering only the leading order term in the polynomial expansion, which is given in Lemma 5.1(ii). Because each term in the expansion is non-negative, this is a valid lower bound. The second lower bound comes from observing that $g(n,0,0,0)$ is monotonically increasing with k, as there are no negative coefficients in the polynomial expansion. Therefore, we can also take a lower bound which is simply the value at $k = 1$, which we know counts the total number of possible graphs and follows from Lemma 5.1(i). □

Stirling's approximation tells us when each lower bound is most useful:

$$(2n)!k^{2n} \sim (nk)^{2n}\left(\frac{4}{e^2}\right)^n, \tag{5.29}$$

$$(2n-1)!!^4 4^n \sim n^{4n}\left(\frac{64}{e^4}\right)^n. \tag{5.30}$$

For $k \in o(n)$, Eq. (5.30) is larger, and when $k \in \omega(n)$, Eq. (5.29) is instead larger. When $k \in \Theta(n)$, then both lower bounds have a leading dependence of n^{4n}, so which is better depends on the constant of proportionality.

Armed with our analytical results and the exact numerical data from the recursion, we can now investigate how the second moment scales with k and n. In Fig. 5.6(a), we plot the logarithm of the upper and lower bounds, as well as the numerically exactly computed values for $(2n-1)!!g(n,0,0,0)$, for our largest available n, which is $n = 40$. We set $k = n^a$ with $a \in [0,4]$. We see that, except for when $k = n^0$ and the upper bound is exactly correct (as is the lower bound based on the number of graphs), the lower bound is a much better approximation. In fact, as expected, the lower bound based on the leading order appears to become a very good approximation as k gets larger.

We should also point out that the logarithmic scaling of the y-axis of Fig. 5.6(a) means that small differences between the exact values and the corresponding lower bound actually represent large multiplicative differences between the true values. For this reason, in Fig. 5.6(b),

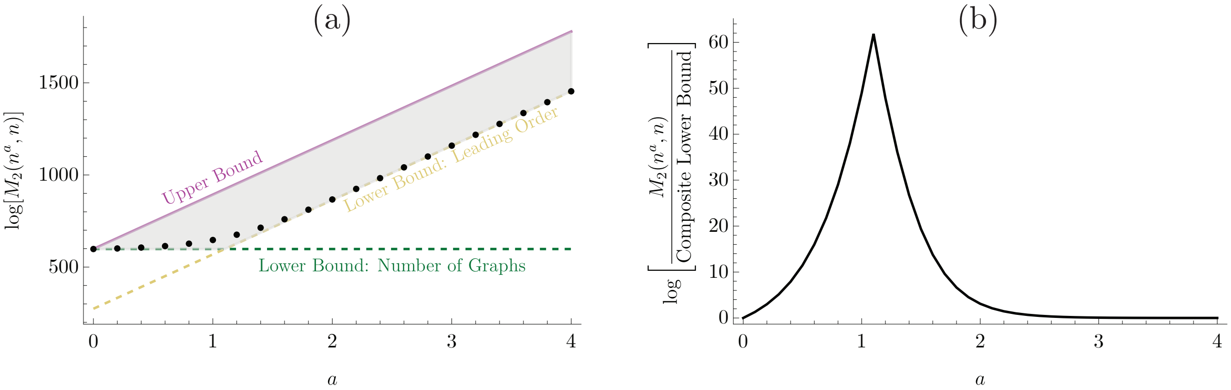

Figure 5.6: Plots showing scaling of the second moment compared to upper and lower bounds. For both plots, physically, k should be an integer, but we here simply use the polynomial expansion of the second moment as a function of arbitrary real k. (a) Scaling of logarithm of the second moment and its upper and lower bounds for $n = 40$ and $k = n^a$ with $a \in [0, 4]$. The green horizontal dashed line and the yellow slanted dashed line represent the lower bounds based on the number of graphs [Eq. (5.28)] and the leading order term [Eq. (5.27)], respectively. The maroon solid line represents the upper bound Eq. (5.26). The bound region is, therefore, highlighted in gray. Numerically exact data is given for $n = 40$ by the black dots [115]. Notice that the black dots representing the exact data stay within the gray region and, for most values of a, closely track the lower bound. (b) Difference between the logarithms of the exact data and the combined lower bound. The peak of the curve shows where the lower switches from the number of graphs to the leading order term. We see that, around $a = 2$, the lower bound based on the leading order becomes a good approximation.

we also plot the difference between the logarithms of the exact data and the composite lower bound defined by $\max\{$Eq. (5.27), Eq. (5.28)$\}$. This helps show how the exact data trends toward Eq. (5.27) as k grows.

Relatedly, we can actually show analytically that Eq. (5.27) cannot fully capture the scaling of the second moment when $k = O(n^2)$. In Appendix D.3, we discuss how to compute individual coefficients in the polynomial expansion of the second moment. There, we give a new proof that $c_{2n} = (2n)!!$, and we also prove for the first time that $c_{2n-1} = (2n)!!(3n-2)n$. Together, these two results mean

$$\frac{c_{2n}k^{2n}}{c_{2n-1}k^{2n-1}} = \frac{k}{(3n-2)n} \sim \frac{k}{n^2}. \tag{5.31}$$

Therefore, in order for the leading term $c_{2n}k^{2n}$ to asymptotically dominate $c_{2n-1}k^{2n-1}$, we require $k = \omega(n^2)$. *A fortiori*, for the leading term to dominate all other terms, and, therefore, for the leading-order lower bound to be a good approximation for the second moment, k must be $\omega(n^2)$.

In summary, then, the lower bounds in Eqs. (5.27) and (5.28) typically track the true value of the second moment much better than the upper bound in Eq. (5.26). When $k = \omega(n^2)$, the first lower bound, Eq. (5.27), which is based on the leading order term, appears to be a very good approximation to second moment.

5.6 Locating the Transition in Anticoncentration

We now move on to some of the concrete consequences of this Chapter The main result of Chapter 4 is identifying a transition in anticoncentration in Gaussian Boson Sampling as a function of k, the number of initially squeezed modes. This result follows entirely from analytic results. Specifically, in Chapter 4, we show through direct computation that, when $k = 1$, the probabilities do not anticoncentrate, and we use the leading order term to show that these probabilities weakly anticoncentrate in the limit that $k \to \infty$. Hence, we show the existence of a transition, but we do not isolate its exact location. We do conjecture that it occurs at $a = 2$, where k scales with n as $k = \Theta(n^a)$, based on an allusion to Scattershot Boson Sampling [97], which is another generalization of Fock state Boson Sampling; there the initial state is composed of two-mode squeezed states where one half of each state is measured and postselected on measurements with at most one photon. In short, one can roughly draw a connection between the presence of hiding in Scattershot Boson Sampling and the number of initially squeezed modes (this is detailed more thoroughly in Appendix C.6).

The main contribution of this Chapter is to show convincingly that the location of the transition is indeed at $k = \Theta(n^2)$. We accomplish this through numerical arguments based on the exact data generated through the recursion for the second moment and a few more analytic results. We formalize this with the following conjecture:

Conjecture 5.2 (Anticoncentration in Gaussian Boson Sampling)**.** *Let $2n = o(\sqrt{m})$ such that one operates in the (conjectured) hiding regime. Then Gaussian Boson Sampling does not anticoncentrate for $k = O(n^2)$, but it weakly anticoncentrates with inverse normalized second moment, $m_2(k,n) := M_1^2(k,n)/M_2(k,n)$, scaling as $1/\sqrt{\pi n}$ for $k = \omega(n^2)$.*

Our evidence for Conjecture 5.2 is twofold and based on results regarding the anticoncentration of the approximate distribution (see Appendix C.5.2 for details on how to convert these statements to those about anticoncentration of the exact distribution):

1. We provide a sequence of numerical plots of $\log[(m_2(k,n)\sqrt{\pi n})^{-1}]$ and its symmetric difference with respect to n for various polynomial scalings of k with n. The numerical

plots of the function itself show an exponential scaling when $k = O(n^2)$, but that the function becomes approximately constant when $k = \omega(n^2)$. Similarly, the plots of the symmetric difference are positive in the $k = O(n^2)$ regime, but approximately vanish when $k = \omega(n^2)$.

2. We show that, assuming the lower bound for $M_2(k,n)$ is a good approximation, weak anticoncentration holds for $k = \omega(n^2)$. We also show that there is a lack of anticoncentration when $k = o(n)$.

We begin with the numerical evidence. In Fig. 5.7, we set $k = n^a$ and plot $\log[(m_2(k,n)\sqrt{\pi n})^{-1}]$ for various values of a. We choose this quantity because, in the asymptotic limit of large k, $(m_2(k,n)\sqrt{\pi n})^{-1} \sim 1$, but when $k = 1$, it is exponentially big.[9] Therefore, we hope to use Fig. 5.7 to understand how this quantity interpolates between the exponential and polynomial behavior of $m_2(k,n)^{-1}$. In Fig. 5.7(a), we plot $\log[(m_2(k,n)\sqrt{\pi n})^{-1}]$ for $a = 0.5$ to $a = 4.0$ with spacing 0.5. We see that for $a \le 2$, this quantity seems to linearly increase with n, meaning that $m_2(k,n)^{-1}$ is exponentially large in n. However, for $a > 2$, it trends to a small constant. Because $m_2(k,n) \sim 1/\sqrt{\pi n}$ is derived in the limit of asymptotically large k using the leading order lower bound for the second moment in Eq. (5.27), this suggests that the use of this lower bound is a good approximation to the second moment when $a > 2$; this aligns well with Fig. 5.6. Thus, we see that, when $a > 2$, the normalized second moment trends to its asymptotic-in-k value of $\sqrt{\pi n}$. In Fig. 5.7(b), we zoom in on the suspected transition point and plot the same quantity when $a \in \{1.95, 1.99, 2.00, 2.01, 2.05, 2.10, 2.15, 2.20\}$. We see similar behavior in this plot; namely, at approximately $a = 2$, the curves transition from growing in n to decreasing toward 0. To clarify

[9] Note added: See Eq. (4.10).

this point even further, we also plot the symmetric difference of the above quantity as a function of n (excluding the minimum and maximum values of n). Here, the symmetric difference of a function $f(n)$, which we refer to as $\Delta_n f(n)$, is defined as $(f(n+1) - f(n-1))/2$. Fig. 5.7(c) and Fig. 5.7(d) use the same values of a as Fig. 5.7(a) and Fig. 5.7(b), respectively. We see that, up to some finite size effects, when $a > 2$ this symmetric difference trends to 0, but it remains positive for $a \leq 2$. We next plot in Fig. 5.8 the symmetric difference $\Delta_n \log[(m_2(k,n)\sqrt{\pi n})^{-1}]$

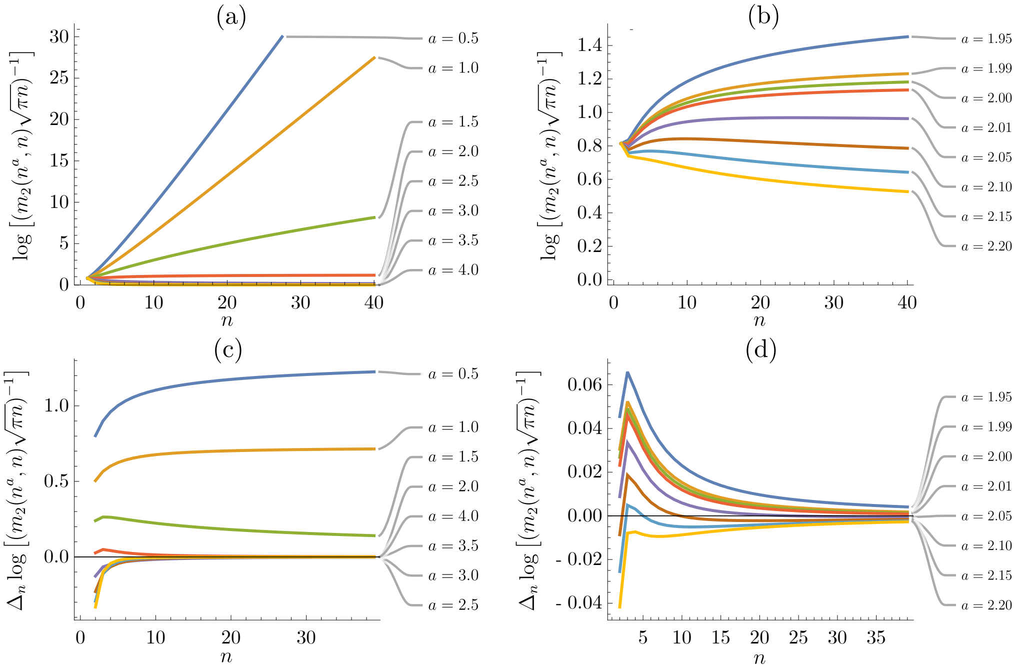

Figure 5.7: Plots of $\log[(m_2(k,n)\sqrt{\pi n})^{-1}]$ and its symmetric difference, notated as Δ_n, as a function of n for $k = n^a$. Recall that $m_2(k,n) := M_1(k,n)^2/M_2(k,n)$ and, for asymptotically large k, $m_2(k,n) \sim 1/\sqrt{\pi n}$.[10](a) $a \in [0.5, 4.0]$, equally spaced by 0.5. (b) $a \in \{1.95, 1.99, 2.00, 2.01, 2.05, 2.10, 2.15, 2.20\}$ to show the regime around $a = 2$ more clearly. (c) The symmetric difference of $\log[(m_2(k,n)\sqrt{\pi n})^{-1}]$ with respect to n, again with $a \in [0.5, 4.0]$. (d) Zooming in on the symmetric difference when a is around 2, with the same values as plot (b). Note that each of the curves in plots (a) and (b) are composed of numerically exact data at 40 points ($n \in \{1, \ldots, 40\}$) that are smoothed for visualization. The same holds for plots (c) and (d), except there are only 38 points ($n = 1$ and $n = 40$ are excluded because we compute the symmetric difference). Finally, while k physically must be an integer, we do not enforce that for these plots; we instead just using the polynomial expansion of the moments to extend k to arbitrary real numbers.

[10] Note added: See Eq. (4.11).

with respect to n at $n = 39$ (the largest n for which we can compute the symmetric difference) as a function of a. We see the symmetric difference vanish near $a = 2$, as would be expected if the transition occurs at $k = \Theta(n^2)$. The inset of Fig. 5.8 clarifies this by plotting the logarithm of this symmetric difference such that its vanishing instead becomes a divergence.

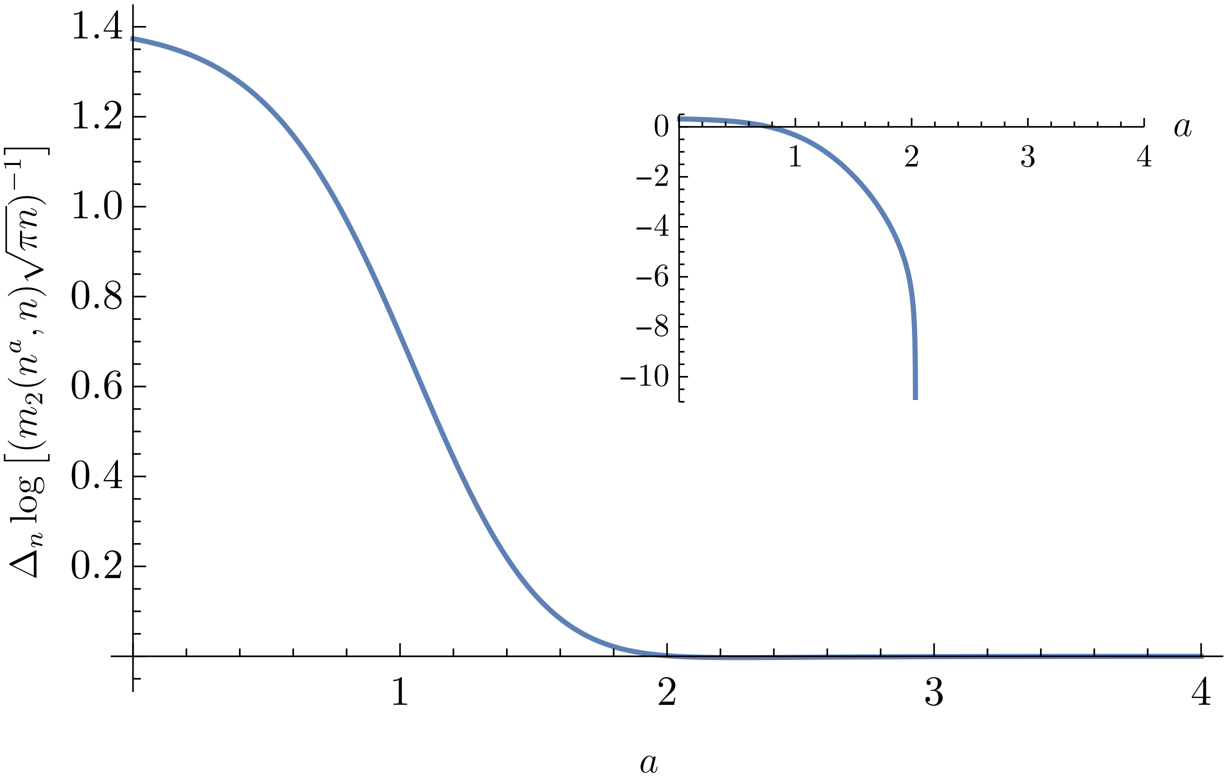

Figure 5.8: Symmetric difference $\Delta_n \log[(m_2(k,n)\sqrt{\pi n})^{-1}]$ evaluated at $n = 39$. Here, $k = n^a$, and a represents the x-axis. Again, physically, k must be an integer, but for this plot we are simply using the polynomial expansions of the moments where k can be an arbitrary real number. This symmetric difference vanishes very close to $a = 2$, suggesting that, when $k = \Omega(n^2)$, the quantity $m_2(k,n)\sqrt{\pi n}$ is a constant, meaning the normalized second moment appears to scale as $\sqrt{\pi n}$. The inset simply plots the $\log$ of the y-axis in the main plot (still with a along the x-axis) in order to visualize more clearly the transition. The divergence occurs somewhere around $a = 2.03$, but we suspect this difference is due solely to finite-size effects. Beyond this divergence, the symmetric difference is negative, meaning the logarithm is complex and, hence, unplotted.

For our second, more analytic argument, we show that if the lower bound is a good approximation to the second moment, then weak anticoncentration holds for $k = \omega(n^2)$ and there is a lack of anticoncentration when $k = o(n)$.

First, consider the case $a < 1$. Note that $k = n^a$ is negligible to n (asymptotically in n). Therefore, up to subleading order,

$$\frac{(k+2n-2)!!}{(k-2)!!} \sim (2n)!!. \tag{5.32}$$

Using Eq. (5.28), which is a valid lower bound, we get

$$\begin{aligned}
\frac{M_2(k,n)}{M_1(k,n)^2} &\gtrsim \frac{(2n-1)!!^4 4^n}{((2n-1)!!(2n)!!)^2} && (5.33)\\
&= 4^n \frac{(2n-1)!!^2}{(2n)!!^2} && (5.34)\\
&\sim \frac{4^n}{\pi n}, && (5.35)
\end{aligned}$$

which is exponentially big, demonstrating a lack of anticoncentration (accounting for the subleading contribution of k does not change the conclusion). Here, we have used Stirling's approximation and

$$\frac{(2n)!!}{(2n-1)!!} \sim \frac{\sqrt{2\pi n}(2n/e)^n}{\sqrt{2}(2n/e)^n} = \sqrt{\pi n}. \tag{5.36}$$

We now examine the case where $k = n^a$ with $a > 2$. We use that, according to Fig. 5.6, the lower bound $M_2(k,n) \geq (2n)!k^{2n}$ is actually an extremely good approximation to the second moment. Here, k now dominates n, so

$$\frac{(k+2n-2)!!}{(k-2)!!} \sim \sqrt{k^{2n}} = n^{an}. \tag{5.37}$$

Correspondingly, the normalized second moment scales as

$$\frac{M_2(k,n)}{M_1(k,n)^2} \sim \frac{(2n-1)!!(2n)!!k^{2n}}{(2n-1)!!^2 k^{2n}} \tag{5.38}$$

$$= \frac{(2n)!!}{(2n-1)!!} \tag{5.39}$$

$$\sim \sqrt{\pi n}. \tag{5.40}$$

Therefore, when $k = \omega(n^2)$, weak anticoncentration holds (again, the inclusion of any subleading terms does not change the conclusion). Note that this argument is similar to the argument used to demonstrate the existence of the transition in the first place, but it uses the fact that the second moment is already well approximated by the leading order lower bound at $k = \omega(n^2)$ instead of just in the asymptotic limit of large k. Unfortunately, our current results are insufficient to more formally handle the regime $a \in [1,2]$ regime.

To recap, we have shown the following results. First, we have provided numerics in Figs. 5.7 and 5.8 that suggest that $\sqrt{\pi n}$ is a good approximation to the normalized second moment when $k = \omega(n^2)$. This is the value of the normalized second moment that is calculated when one uses the lower bound in Eq. (5.27) that is based on the leading order term. Similarly, these plots numerically indicate that when $k = O(n^2)$, the normalized second moment grows exponentially in n, meaning there is a lack of anticoncentration. Next, we have shown that, if the leading order is a good approximation to the second moment, which, according to Fig. 5.6 occurs when $k = \omega(n^2)$, then the normalized second moment scales as $\sqrt{\pi n}$, meaning weak anticoncentration holds in that regime. We have also shown that for $k = O(n)$, there is a lack of anticoncentration. All together, the totality of the evidence presented here strongly suggests the veracity of Conjecture 5.2 and that the transition between lack of anticoncentration and weak anticoncentration in the approximate output distribution occurs at $k = \Theta(n^2)$.

5.7 Conclusion

In this Chapter, we have studied the output distribution of the prototypical setup for Gaussian Boson Sampling in the hiding regime. Our main theoretical contribution is the development of a recursion relation that allows one to compute numerically exactly in polynomial time the second moment of these output probabilities for any photon Fock sector. We additionally detail separate ways to calculate individual coefficients of the polynomial expansion of the second moment. Together, these results provide strong evidence for our conjecture that the transition in anticoncentration, whose existence is proven in Chapter 4 and Appendix C, occurs at $k = \Theta(n^2)$.

Ideally we would have been able to derive a closed-form expression for the polynomial description of the second moment akin to Theorem 5.1, as this might have allowed us to formally prove this conjecture, but we leave this important question to future work. It would also be nice to develop a better, more intuitive understanding for why this transition occurs. It appears to be related to the transition between collisional and collision-free outputs in Scattershot Boson Sampling, but the connection is not perfect, and further investigation seems worthwhile.

Related to all of these points, the precise nature of the crossover at $k = \Theta(n^2)$ is an interesting realm of future study. Specifically, we conjecture that weak anticoncentration holds for $k = \omega(n^2)$ and there is a lack of anticoncentration when $k = O(n^2)$, which of course places the transition at $k = \Theta(n^2)$. But precisely how the normalized moment behaves as we tune a through $a = 2$ deserves special attention.

Our results open the door for answering other questions of interest. In particular, our results may make it possible to evaluate how well certain classical algorithms may sample from the output distribution or evaluate spoofing cross-entropy benchmarking in Gaussian Boson Sampling. Further exploration here is worthwhile. We also note that we have studied Gaussian Boson Sampling with no noise and number-resolving detectors. It would be interesting to see whether our techniques can be expanded to imperfect settings, such as when photons are partially distinguishable [117], or when the measurement detectors only distinguish between the presence or absence of photons [118].

Finally, the graph-theoretic approach that we have developed in this Chapter is surprisingly flexible, and it deserves continued treatment. In Appendix D.4, we present another way to use the graphs in $\mathbb{G}_2^n$ in order to develop a recursion that can solve for the second moment. In short, this other approach observes that there are really only five types of black edges in our graphs: ones that stay in row 1, ones that stay in row 3, and ones that go between rows 1 and 2, rows 1 and 3, and rows 2 and 3. Because we are interested only in the number of connected components, and because we sum over *all* perfect matchings defined by red edges in each row, we are free to drag the black edges around and order them in new, convenient ways. Therefore, looking at these graphs from the perspective of the total number of each type of black edge allows us to conceive of a different kind of recursion for the second moment. While we only sketch the idea behind this alternative recursion, we believe that it may be a promising new way of looking at the problem. In particular, this new approach allows us to find an, admittedly, somewhat complicated, expression for c_1 (that reproduces our expression for c_1 found via the original recursion up to $n = 40$). However, this new approach should not be viewed as a strict alternative to what we have derived in this manuscript, but a complementary approach that might yield new insights.

We leave exploring it to future work.

Acknowledgments

We thank Changhun Oh, Bill Fefferman, Marcel Hinsche, Max Alekseyev, and Benjamin Banavige for helpful discussions. We thank Jacob Bringewatt for providing feedback on Appendix A. A.E., J.T.I., and A.V.G. were supported in part by the DoE ASCR Accelerated Research in Quantum Computing program (award No. DE-SC0020312), DARPA SAVaNT ADVENT, AFOSR MURI, DoE ASCR Quantum Testbed Pathfinder program (awards No. DE-SC0019040 and No. DE-SC0024220), NSF QLCI (award No. OMA-2120757), NSF STAQ program, and AFOSR. Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator. JTI thanks the Joint Quantum Institute at the University of Maryland for support through a JQI fellowship. D.H. acknowledges funding from the US Department of Defense through a QuICS Hartree fellowship. Specific product citations are for the purpose of clarification only, and are not an endorsement by the authors or NIST.

Chapter 6: Simulation Complexity of Many-Body Localized Systems

Abstract: We use complexity theory to rigorously investigate the difficulty of classically simulating evolution under many-body localized (MBL) Hamiltonians. Using the defining feature that MBL systems have a complete set of quasilocal integrals of motion (LIOMs), we demonstrate a transition in the classical complexity of simulating such systems as a function of evolution time. On one side, we construct a quasipolynomial-time tensor-network-inspired algorithm for strong simulation of 1D MBL systems (i.e., calculating the expectation value of arbitrary products of local observables) evolved for any time polynomial in the system size. On the other side, we prove that even weak simulation, i.e. sampling, becomes formally hard after an exponentially long evolution time, assuming widely believed conjectures in complexity theory. Finally, using the consequences of our classical simulation results, we also show that the quantum circuit complexity for MBL systems is sublinear in evolution time. This result is a counterpart to a recent proof that the complexity of random quantum circuits grows linearly in time.[1]

[1] Note added: This further answers in the affirmative an open question in the literature on whether MBL Hamiltonians may be fast-forwarded.

6.1 Introduction

As quantum computers become larger-depth, less error-prone, and eventually fully fault-tolerant, it will become increasingly important to understand which computational problems admit quantum speedups over the best possible classical algorithms. This question broadly falls under the domain of computational complexity theory, which studies how easy or hard it is to solve certain problems under various computational assumptions. More specifically, *sampling complexity*, the study of how difficult it is to draw samples from classes of probability distributions, is a useful framework for studying the classical hardness of simulating quantum systems, and can help to narrow the parameter space where quantum advantage may be obtained. At their core, many quantum experiments reduce to repeatedly preparing a certain quantum state, measuring it (thus generating a probability distribution of outcomes), and classically post-processing on the measurement results. This high-level viewpoint motivates the systematic study of quantum systems via the lens of sampling complexity. Indeed, the past ten years have seen significant interest in sampling after the proof (up to widely believed mathematical conjectures) that one could obtain a quantum advantage in the famous Boson Sampling problem [27], leading to the recent demonstration of quantum sampling experiments believed to be beyond the accessibility of classical simulations [29–31].

With the same motivation in mind, Ref. [119] considered a system of indistinguishable non-interacting bosons distributed on a lattice and evolved under a local Hamiltonian (also see Refs. [120, 121] for variants of this problem). Intuitively, one expects that classical simulation is initially easy while the particles are separated, but grows more difficult as the system evolves. Reference [119] formalized this idea by showing that sampling remains easy until the particles have evolved for long enough to travel the distance initially separating them, whereafter their fundamental indistinguishability leads to quantum interference that is hard to classically simulate. A key corollary of this result is that classical sampling is easy in single-particle-localized systems, where the particle wavepackets do not spread out [122–125]. Thus, while single-particle localized systems are fascinating from a condensed matter perspective, we do not necessarily expect them to encode hard computational problems, and we will likely have to look to other types of systems to find useful quantum speedups.

This Chapter is concerned with the more subtle situation of *many-body* localization (MBL) [36, 126, 127] in spin systems, which we take to mean any spin Hamiltonian having a complete set of local integrals of motion (precisely defined below) [128–132]. These systems differ from the single-particle-localized situation described above in a crucial way: the quasilocal commuting operators that fully describe the dynamics of these systems interact with one another through non-trivial exponentially decaying interactions. These interactions can spread entanglement through the system and destroy separability of an initial state over exponentially long time-scales.

Suppose we time-evolve an initial product state under an MBL Hamiltonian acting on N spins and then measure the result in a product basis, generating a probability distribution. We will explore the algorithmic time complexity of both *strong simulation* and *weak simulation* of this physical system. Weak simulation is the ability to sample from the distribution of outcomes, whereas strong simulation is the ability to calculate all marginal and conditional probabilities of the outcomes. The ability to strongly simulate a system implies the ability to sample from it [133], but not vice versa—one can, in principle, sample from a distribution without ever knowing the values of the probabilities.

Observe that in describing the problem of interest, we have introduced two types of time:

Evolution Time t	Complexity	Task
$\mathcal{O}(\log N)$	Easy [134]	Strong Simulation
$\mathcal{O}(\text{poly} N)$	Quasi-easy	Strong Simulation
$\mathcal{O}(\text{quasipoly} N)$	Quasi-easy	Strong Simulation
$\Omega(\exp N)$	Hard	Weak Simulation

Table 6.1: Summary of our results for classical simulation. We define "quasi-easy" to be those problems admitting a quasipolynomial-time algorithm but which may yet possess a polynomial-time algorithm.

evolution and computational. For clarity in the remainder of this Chapter, we will use a lower-case t to refer to the physical evolution time, or the time for which the MBL Hamiltonian acts on the initial state. We denote the time complexity of a classical algorithm for a given simulation task with an upper-case T.

We now present our main results. Using techniques inspired by tensor networks, we present an algorithm that can strongly simulate (and thus sample from) any one-dimensional MBL system in quasipolynomial computer time (i.e., times of the form $T = \exp[\mathcal{O}(\log^c N)]$ [2] for some $c > 1$), for any evolution time t polynomial in the system size N. It is interesting that even this algorithm does not run in strictly polynomial time, and we are not aware of any algorithm which (provably) can. Conversely, by using ideas inspired by the hardness of the Instantaneous Quantum Polynomial (IQP) sampling problem in Ref. [37], we also show that the MBL sampling problem becomes hard in the worst case after evolution time $t = \Omega(\exp[N^\delta])$ for arbitrarily small $\delta > 0$ (by "worst case," we mean that we demonstrate that a specific family of MBL Hamiltonians becomes hard to simulate, but this family does not contain all possible MBL Hamiltonians). These results are summarized in Table 6.1.

Interestingly, as a consequence of our proof techniques, we can also derive results on the

[2] We say that $f = \mathcal{O}(g)$ if $f/g \nrightarrow \infty$ as $n \to \infty$, and $f = \Omega(g) \iff g = \mathcal{O}(f)$. Similarly, $f = o(g)$ means $f/g \to 0$ as $n \to \infty$, and $f = \omega(g) \iff g = o(f)$. Finally, if $f = \mathcal{O}(g)$ and $g = \mathcal{O}(f)$, we say $f = \Theta(g)$ (and $g = \Theta(f)$). The precise asymptotic dependence on n can be arbitrary. Additionally, a tilde over the asymptotic symbol, such as $\tilde{\mathcal{O}}(g)$, means that we are ignoring logarithmic factors in g.

quantum circuit complexity of implementing time evolution due to an MBL Hamiltonian. The quantum circuit complexity of a unitary U is the minimum number of gates (from a predefined universal gate set) required to approximate U. In many-body physics, it is of great significance to understand how the quantum circuit complexity of a time-evolution operator e^{-iHt} grows with respect to the time t for various Hamiltonians H. In the context of high-energy physics, gravitational physics, and the AdS/CFT correspondence, it was conjectured [135, 136] that the circuit complexity of a conformal field theory is dual to the action of a gravitational theory describing the bulk. More specifically, it has been conjectured that the circuit complexity of fast-scrambling dynamics grows linearly in time until a timescale exponential in system size. This conjecture has gathered support due to recent work [38, 137].[3] In stark contrast with these fast scramblers, we show in this Chapter that the circuit complexity for sufficiently localized MBL Hamiltonians grows only sublinearly with evolution time.[4] Therefore, our work suggests that, in addition to classical complexity, studying the quantum complexity of simulating time evolution can also serve as a basis for classifying the ergodicity of quantum dynamics.

Others have investigated the simulation of MBL systems. For a few examples, see Refs. [143–147], which introduce efficient methods for classically simulating both spin and weakly-interacting fermionc MBL systems. However, while these works demonstrate empirically good numerical alternatives to computationally demanding exact diagonalization schemes, they stop short of formal proofs that these algorithms can maintain accuracy for all MBL systems as the system size grows (though Ref. [145] does contain some formal proofs in the case of exactly local integrals of motion, as opposed to the more general quasilocal integrals of motion we consider here). Overall,

[3] Note added: A more complete (but not exhaustive) list of works studying this topic include Refs. [38, 137–141].

[4] Note added: We note that this answers in the affirmative an open question raised by Ref. [142] about whether MBL Hamiltonians may be "fast-forwarded," that is, whether it is possible to quantumly simulate the evolution for time t of an MBL Hamiltonian in such a way that the simulation time is parametrically smaller than t.

our work is the first to systematically investigate the simulation of generic MBL systems from a rigorous complexity-theoretic perspective.

The rest of this Chapter is organized as follows. In Section 6.2, we formally define the simulation problem. We then prove in Section 6.3 crucial mathematical results that we use in Section 6.4 to demonstrate the quasipolynomial runtime of our tensor-network algorithm for strong simulation. Correspondingly, in Section 6.5 we demonstrate that generic MBL Hamiltonians are hard to sample from after exponentially long evolution time t. In Section 6.6 we also show that that the quantum circuit complexity of the time-evolution operator of a sufficiently localized MBL Hamiltonians is sublinear in time. Finally, in Section 6.7 we synthesize these results and consider directions for future work.

6.2 Setup

Consider a 1D lattice of N spin-1/2 particles (with spin operators σ_i^α, $\alpha = x, y, z$) that evolve under some Hamiltonian H. We say that H is MBL if there exists a quasilocal unitary U (defined below) that brings H to the form

$$H = \sum_i J_i \tau_i^z + \sum_{i<j} J_{ij} \tau_i^z \tau_j^z + \sum_{i<j<k} J_{ijk} \tau_i^z \tau_j^z \tau_k^z + \dots, \qquad (6.1)$$

with $[\tau_i^z, \tau_j^z] = 0$ and $\left|J_{i_1 \dots i_p}\right| \leq \exp\left(-(i_p - i_1)/\xi\right)$. We call the σ_i^z the *physical bits* (p-bits) because they represent the experimentally accessible basis of observables, and we call the τ_i^z the *local integrals of motion* (LIOMs) or *localized bits* (l-bits) because they commute with the Hamiltonian and thus represent a set of N conserved quantities that constrain the dynamics.

We define a quasilocal unitary, which we schematically depict in Fig. 6.1, as follows:

Definition 6.1 (Quasilocal unitary [36])**.** *A unitary U is quasilocal if it can be decomposed on a finite 1D lattice with N sites as*

$$U = \prod_{n=1}^{N} \prod_{j=1}^{n} \prod_{i=0}^{\lfloor (N-n)/n \rfloor} U_{in+j}^{(n)}, \tag{6.2}$$

where $U_k^{(n)}$ acts on sites $k, k+1, \ldots, k+n-1$ such that

$$\left\| \mathbb{1} - U_k^{(n)} \right\|^2 < q e^{-\frac{(n-1)}{\xi}}, \tag{6.3}$$

where $\|\cdot\|$ is the operator norm (i.e., the largest singular value of the operand) and q is some $\mathcal{O}(1)$ constant.[5] *When $k+n-1 > N$, $U_k^{(n)}$ should be interpreted as a tensor product of two unitaries, one acting on sites k through N, and the other on 1 to $k+n-1-N$.*

This means that we can decompose U into a sequence of n layers of n-site unitaries, where the more sites a constituent unitary acts on, the closer it is to the identity. We call U "quasilocal" because, though any two distant sites may be entangled, the amount of entanglement generated decays rapidly with distance.

Having defined the properties of our Hamiltonian H, consider now an experiment whereby the system is initially prepared in the physical state $|0\ldots0\rangle$ (i.e., $\forall i\ \sigma_i^z |0\ldots0\rangle = |0\ldots0\rangle$), then time-evolved into $e^{-iHt}|0\ldots0\rangle$, and finally measured in the physical basis. The probability of observing an outcome $|\sigma\rangle$ after a time t is $\mathcal{D}(\sigma) := |\langle\sigma|e^{-iHt}|0\ldots0\rangle|^2$. As previously discussed, we want to assess the difficulty of both drawing a sample from (weak simulation) and calculating marginals of (strong simulation) the distribution $\mathcal{D} := \{\mathcal{D}(\sigma)\}_\sigma$. However, even a quantum

[5] To be precise, our definition is for a family of quasilocal unitaries U with respect to the size N of the system, otherwise all unitaries U would be quasilocal given a sufficiently large, but still constant, q.

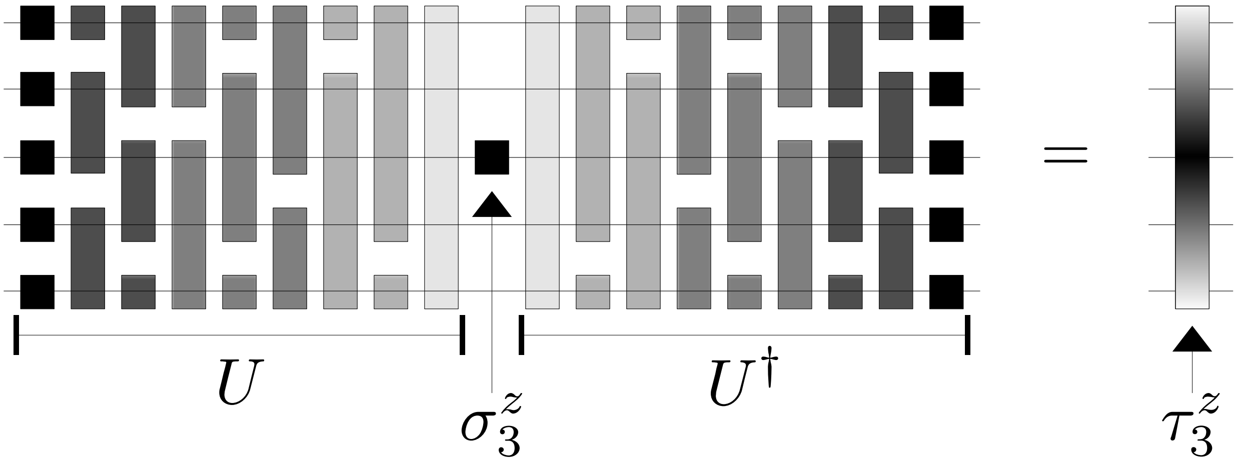

Figure 6.1: Schematic depiction of a quasilocal unitary U on $N = 5$ sites converting between the physical and localized bases, $U\sigma_3^z U^\dagger = \tau_3^z$. As described in Definition 6.1, U decomposes into constituents, and the opacity of each constituent block represents its proximity to the identity with respect to the norm $\|\cdot\|$; the lighter the block, the closer it is to the identity.

computer directly performing such an experiment will be subject to at least small errors, and will thus be unable to draw a sample from this distribution perfectly. Therefore, we will only assess the difficulty of *approximate* sampling from a distribution $\mathcal{D}_\epsilon$ that is ϵ-close to $\mathcal{D}$ in total variation distance (TVD):

$$\|\mathcal{D}_\epsilon - \mathcal{D}\|_{\mathrm{TVD}} = \frac{1}{2}\sum_\sigma |\mathcal{D}_\epsilon(\sigma) - \mathcal{D}(\sigma)| < \epsilon. \tag{6.4}$$

We state our sampling problem formally.

Problem 6.1. *Let H be an MBL Hamiltonian (according to the above definition) on an N-site chain and U its corresponding quasilocal unitary. Consider the distribution $\mathcal{D} = \{|\langle\sigma|e^{-iHt}|0\ldots0\rangle|^2\}_\sigma$. Given a description of H in terms of physical operators, an efficient algorithm to compute any element of any constituent $U_k^{(n)}$ of U, and an efficient algorithm to compute any coupling $J_{i_1\ldots i_p}$, output a sample from a distribution $\mathcal{D}_\epsilon$ that is ϵ-close to $\mathcal{D}$ in total variation distance for any $\epsilon > 0$.*

A few comments on Problem 6.1 are worthwhile. We need these efficient algorithms to calculate any desired constituent $U_k^{(n)}$ and any desired coupling $J_{i_1 \dots i_p}$ because knowledge of these quantities will be crucial for our algorithm, and it is too computationally expensive to calculate and naively list out all exponentially many of them. Formally, we assume that we have an *oracle* for these properties of the system.

Ideally we would be able to extract $J_{i_1 \dots i_p}$ and U efficiently from the description of H in the physical basis. However, MBL is typically considered in the context of disordered spin chains where it may not always be possible to efficiently compute these quantities (though there is some evidence that this may be possible—see Refs. [145–149]). Therefore, we do not restrict ourselves to this particular mechanism for producing LIOMs, and our results will apply to any Hamiltonian that can be diagonalized by quasilocal unitary U into the form Eq. (6.1). Finally, neither the specific initial state nor the measurement basis are critical to our formulation of Problem 6.1 as long as they are a product state and a product basis. This is because we allow U to contain a layer of $\mathcal{O}(1)$ 1-site terms so that we do not pick out any particular basis as special. Our main results concern the classical time complexity T of solving Problem 6.1 as a function of evolution time t and system size N.

6.3 Truncating the Canonical Hamiltonian

We proceed to characterize the classical complexity of solving Problem 6.1 in two ways depending on the evolution time t. If $t = \mathcal{O}(\log N)$ and H is finite-range in the physical basis, Ref. [134] proves there exists an efficient matrix-product operator representation of the propagator e^{-iHt}. This representation may be used to approximately sample from the outcome distribution of evolution under H. See Appendix E for more details.

For longer times $t = \omega(\log N)$, we construct a Hamiltonian $\tilde{H}$ for which the time-evolved probability distribution is $\tilde{\mathcal{D}} := \{|\langle\sigma|e^{-i\tilde{H}t}|0\ldots0\rangle|^2\}$, such that (a) $\|\mathcal{D} - \tilde{\mathcal{D}}\|_{\mathrm{TVD}} \leq \epsilon$ and (b) the distribution associated with evolution under $\tilde{H}$ can be sampled from in computer time scaling quasipolynomially with the number of spins N. The total variation distance between the probability distributions associated with two pure states $|\psi\rangle$ and $|\phi\rangle$ can be upper bounded by the 2-norm distance [150], which in turn can be bounded [121] as $\||\psi(t)\rangle - |\phi(t)\rangle\|_2 \leq ||H - \tilde{H}||t := \|\Delta H\|t$, where $\|\cdot\|$ is the standard operator norm. Therefore, if we want the two distributions to be ϵ-close in total variation distance up to a time t, it is sufficient to ensure $\|\Delta H\| \leq \epsilon/t$.

We construct this approximate Hamiltonian $\tilde{H}$ by *truncating* the exact Hamiltonian in two ways: via the coupling constants and the LIOMs. In particular, we set the coupling constants equal to zero if they connect sites beyond a certain radius r_J, and we set equal to the identity those constituents of U supported on more than r_U sites. Mathematically:

$$\tilde{J}_{i_1\ldots i_p} = \begin{cases} J_{i_1\ldots i_p} & \text{if } i_p - i_1 < r_J \\ \\ 0 & \text{if } i_p - i_1 \geq r_J \end{cases}, \tag{6.5}$$

$$\tilde{U} = \prod_{n=1}^{r_U}\prod_{j=1}^{n}\prod_{i=0}^{\lfloor (N-n)/n \rfloor} U^{(n)}_{in+j}, \tag{6.6}$$

$$\tilde{\tau}_i^z = \tilde{U}\sigma_i^z\tilde{U}^\dagger. \tag{6.7}$$

We can now bound the norm of

$$\Delta H := H - \tilde{H} = \sum_I J_I \tau_I^z - \tilde{J}_I \tilde{\tau}_I^z \tag{6.8}$$

by applying the triangle inequality:

$$\|\Delta H\| \leq \sum_I \Big(|J_I - \tilde{J}_I| + |\tilde{J}_I| \|\tau_I^z - \tilde{\tau}_I^z\|\Big), \tag{6.9}$$

where we have introduced I as a general multi-index for brevity. Before continuing, it is useful to define $S_{p,n_0} := \sum_{n=n_0}^{\infty} \binom{n}{p} e^{-\frac{n}{\xi}}$. Intuitively, this sum appears because we will often be interested in summing over couplings of a range exceeding some n_0, and each coupling comes with an associated exponential decay. Assuming that the localization length $\xi < 1/\log 2$, we have:

$$S_{p,n_0} \leq C \begin{cases} e^{-\frac{n_0}{\xi}} & p = 0 \\ p e^{-ap} & n_0 < n_*, p > 0 \\ \frac{n_0^{p+1}\sqrt{p}}{p!} e^{-\frac{n_0}{\xi}} & n_0 \geq n_*, p > 0 \end{cases}, \tag{6.10}$$

where $a := \log(e^{1/\xi} - 1)$, $n_* := p(1 - e^{-1/\xi})^{-1}$, and C is some $\mathcal{O}(1)$ constant. See Lemma E.2 in Appendix E for a detailed proof.

We now separately bound the two contributions to Eq. (6.9). The details, which are in Appendix E, make heavy use of Eq. (6.10), and the result is

$$\|\Delta H\| \leq C_J N r_J e^{-k r_J} + C_U N^2 e^{-\frac{r_U}{2\xi}}, \tag{6.11}$$

where C_U, C_J, and k are constants independent of N. Intuitively, the factors of N come from summing over sites, and the exponential decay factors come from the decay properties of H and

U. To ensure that $\|\Delta H\| \le \epsilon/t$ for some polynomially long time $t = \mathcal{O}(N^b)$, it suffices to choose

$$r_U = \Omega(\xi b \log N), r_J = \Omega(bk^{-1} \log N). \tag{6.12}$$

Therefore, truncating the coupling coefficients and the diagonalizing quasilocal unitary to a scale logarithmic in the system size is sufficient to produce a distribution that is close in total variation distance to the true distribution.

6.4 Quasipolynomial-Time Sampling

Having defined an appropriate approximation $\tilde{H}$ we now describe how to sample from the distribution generated by $\tilde{H}$. More precisely, we provide an algorithm for strong simulation, meaning it can calculate all probabilities and marginal probabilities of the distribution generated by measuring the simulated system in any local basis. Equivalently, it can estimate the expectation value of arbitrary products of local observables. Strong simulation implies the ability to solve the easier problem of weak simulation, i.e. sampling, which itself implies the ability to calculate the expectation values of local observables [133]. Specifically, our algorithm will calculate

$$\langle \tilde{O} \rangle_t = \langle \psi(0)| e^{i\tilde{H}t} O e^{-i\tilde{H}t} |\psi(0)\rangle , \tag{6.13}$$

where O is a product of single-site observables in the p-bit basis of the form $O = \sigma_i^z \prod_{j<i} P_j$, with P_j a projector of qubit j onto the 0 or 1 outcome when measuring in the appropriate local basis, and the tilde indicates that we evolve with the approximate Hamiltonian $\tilde{H}$. Intuitively, O is selected such that Eq. (6.13) calculates the conditional probability $P(z_i|z_{i-1} \dots z_1)$, and drawing

a sample given these conditional probabilities is equivalent to flipping N biased coins, where the bias of each coin is conditioned on the previous outcomes. For $t = \mathcal{O}(\log N)$, we use the algorithm implied by results in Ref. [134] and elucidated in Appendix E.

In short, when H is short-range in the p-bit basis, the propagator for the true Hamiltonian e^{-iHt} can be efficiently approximated by a matrix product operator M. Because a product of local observables also admits a matrix product operator form, $\langle\psi(0)| M^\dagger O M |\psi(0)\rangle|$ may be calculated in computational time $T = \mathcal{O}(\text{poly} N)$.

For the more complicated problem of $t = \omega(\log N)$, we provide a different algorithm where each unitary in the circuit is interpreted as a tensor, making the quantum circuit for time evolution a tensor network. Specifically, we now insert copies of the identity to rewrite Eq. (6.13) as

$$\langle\tilde{O}\rangle_t = \langle\psi(0)| \tilde{U}^\dagger e^{i\tilde{H}_\sigma t} \tilde{U} O \tilde{U}^\dagger e^{-i\tilde{H}_\sigma t} \tilde{U} |\psi(0)\rangle\,, \tag{6.14}$$

where $\tilde{H}_\sigma := \tilde{U}\tilde{H}\tilde{U}^\dagger$ (in words, $\tilde{H}_\sigma$ takes the form of Eq. (6.1) but with σ_j in place of $\tilde{\tau}_j$ and $\tilde{J}_I$ in place of J_I). We calculate these expectation values using a quantum circuit of the form in Fig. 6.2. We order the qubits going from bottom to top and evolution time from left to right. Following the structure of Eq. (6.14), the first section of the circuit applies $\tilde{U}$ to convert to the truncated LIOM basis. The second section evolves under the truncated Hamiltonian. After converting back to the original basis by using $\tilde{U}^\dagger$, the operator O is applied. Then the previous steps are repeated in reverse. Because the terms of $\tilde{H}_\sigma$ pairwise commute, we are allowed to choose the order in which each term appears. Our choice is the following. Place all evolution under terms supported on site 1 first, refer to these terms as $\tilde{H}_1$, and define $\tilde{V}_1 := e^{-i\tilde{H}_1 t}$. Then, place all evolution under terms supported on site 2, but not site 1, and refer to this as $\tilde{H}_2$. Similarly,

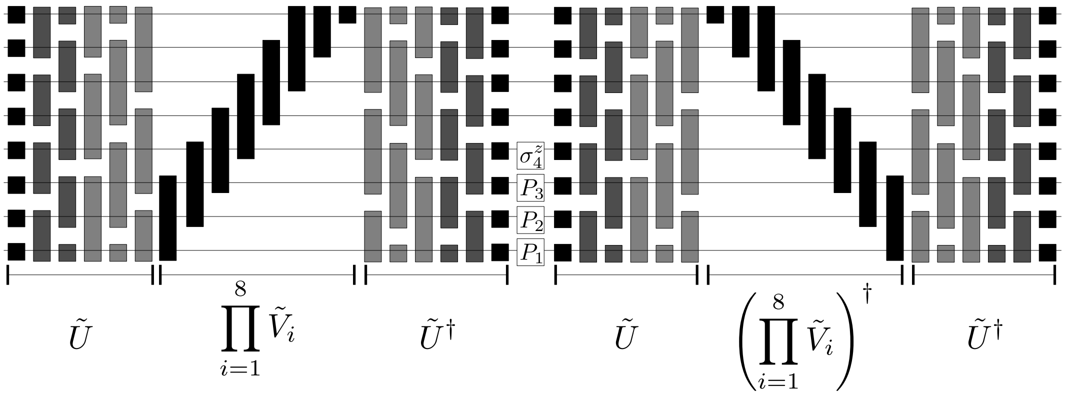

Figure 6.2: Example of the quantum circuit that calculates a relevant product of local observables O on a lattice of $N = 8$ sites. Here $O = \sigma_4^z P_3 P_2 P_1$.

define $\tilde{V}_2 := e^{-i\tilde{H}_2 t}$. Continue in this way until all Hamiltonian evolution is accounted for. See Fig. 6.2 for a depiction of the circuit for $O = \sigma_4^z P_3 P_2 P_1$ and $N = 8$. Note that generating $\tilde{V}_i$ is an efficient process; there are at most $\binom{r_J}{k}$ k-site terms that involve site i (but no site before i) and have physical range at most r_J. Thus, there are at most $2^{r_J} \sim \text{poly}N$ Hamiltonian evolution unitaries that must be multiplied together to generate each of the N unitaries $\tilde{V}_i$. We treat each unitary in the evolution as a tensor, and we contract these tensors "qubit-wise" as opposed to "time-wise." That is, instead of contracting tensors in the order that they appear in Eq. (6.14), we first contract together every tensor that intersects qubit 1. We then contract this much larger tensor with every other tensor that intersects qubit 2, and so forth. Contracting the tensors "time-wise" would quickly lead us to an extensively sized tensor spanning some $\Theta(N)$ portion of the system, and evaluating a contraction involving this extensive tensor would take an exponentially long amount of time; contracting the tensors "qubit-wise" avoids this issue. Ensuring that our algorithm only ever produces tensors with $\mathcal{O}(\log N)$ legs would be sufficient to demonstrate a polynomial time algorithm. This is because $\tilde{U}$ and $\tilde{U}^\dagger$ each contain $\mathcal{O}(N \log N)$ constituents, $e^{-i\tilde{H}t}$ contains only $\mathcal{O}(N)$ terms (as we have decomposed it into $\{\tilde{V}_i\}$), and there are at most N

tensors coming from O. Thus, the total number of tensors, and, correspondingly, the total number of legs that could be contracted, is only $\tilde{\mathcal{O}}(\mathrm{poly}N)$ (where the tilde indicates that we are ignoring logarithmic factors of N).[6] Therefore, the maximum amount of time this algorithm could take would be $\tilde{\mathcal{O}}(\mathrm{poly}N) \cdot 2^{\mathcal{O}(\log N)} = \tilde{\mathcal{O}}(\mathrm{poly}N)$.[7]

Unfortunately we can only guarantee that our algorithm produces tensors with $\mathcal{O}(\mathrm{polylog}N)$-many legs. Intuitively, we cannot guarantee against an adversarial placement of constituents in $\tilde{U}, \tilde{U}^\dagger$ whereby there is a jagged "skyline" of tensors leading to $\mathrm{polylog}N$ leftover legs after a qubit is contracted. Repeating the above analysis means the algorithm can take as long as $\mathcal{O}(\mathrm{poly}N) \cdot 2^{\mathcal{O}(\mathrm{polylog}N)}$. This is not a polynomial-time algorithm; it is quasipolynomial, which means it is faster than any exponential-time algorithm, but slower than any polynomial-time algorithm. Lemma 6.1 formalizes this rough argument.

Lemma 6.1 *Given the truncation of an MBL Hamiltonian and the quasilocal unitary that diagonalizes it, as in Eqs.* (6.5) *and* (6.6)*, following the qubit-wise contraction scheme never creates a tensor with more than* $\mathcal{O}([\log N]^3)$ *leftover legs.*

Proof. We will crudely upper-bound the total number of legs at any stage of the algorithm. It is simple to see that the largest possible tensor occurs at the end of contracting all tensors intersecting a qubit k. At this point consider a bound on the worst-case scenario where each of the n-site constituents in $\tilde{U}$ extends $n-1$ sites above qubit k, and $\tilde{V}_k$ extends $r_J - 1$ sites above qubit k. By naively ignoring that the internal legs should be contracted, it is straightforward to verify that this tensor possesses fewer than $4[2(2-1)+3(3-1)+\cdots+r_U(r_U-1))+2(r_J-1)+2] = \mathcal{O}([\log N]^3)$

[6] Note added: The inclusion of the tilde here does not meaningfully change the conclusions, so we can just as easily bound this as $\mathcal{O}(\mathrm{poly}N)$.

[7] Note added: Again, the tildes do not meaningfully change the conclusion, so we could simply say the algorithm time would be $\mathcal{O}(\mathrm{poly}N) \cdot 2^{\mathcal{O}(\log N)} = \mathcal{O}(\mathrm{poly}N)$

legs. Because this is the worst-case scenario, the bound is thus proven. □

Lemma 6.1 bounds the size of any one tensor contracted in the algorithm, thus placing a quasipolynomial-time bound on any individual contraction. The total number of contraction operations is itself bounded by a polynomial in N. Finally, we proved earlier that the distributions generated by H and $\tilde{H}$ are ϵ-close for polynomial evolution time. Thus, the following theorem holds:

Theorem 6.1. *For evolution time $t = \mathcal{O}(\mathrm{poly} N)$, the contraction algorithm takes time quasipolynomial in N, which means Problem 6.1 can be solved in quasipolynomial time.*

Additionally, observe that Theorem 6.1 can be extended to quasipolynomial evolution time with little effort. Tracking the rest of the proof, we see that truncating the quasilocal unitary and the MBL couplings to length scales polylogarithmic in N will make $\|\Delta H\|$ small enough to counteract the larger evolution time t. A polylogarithmic truncation distance, however, does not change the quasipolynomial conclusion of Lemma 6.1. Finally, we note Theorem 6.1 holds in the worst case, meaning for any possible choice of coupling strengths and quasilocal unitary that obey our definition of MBL.

6.5 Hardness After Exponential Time

In contrast to the quasi-easiness result for strong simulation in Section 6.4, it is also possible to show, via a comparison to Instantaneous Quantum Polynomial (IQP) circuits [151], that weak simulation of, or sampling from, MBL systems becomes formally hard on a classical computer after a time exponential in the system size.

Theorem 6.2. *Problem 6.1 is classically hard when the evolution time $t \geq \Omega(e^{N^\delta/\xi})$ for any $\delta > 0$.*

Proof. For simplicity, we start with $\delta = 1/2$ and give a family of hard instances of the problem, described by the couplings $J_{i_1 \dots i_p}$ in the τ basis and the quasilocal unitaries U that satisfy our definition of MBL. We rely on the hardness construction of Ref. [37], which shows that evolution under a nearest-neighbor, commuting 2D Hamiltonian for constant time can be hard to classically simulate. We implement the nearest-neighbor 2D dynamics using selective long-range interactions in 1D to generate an effective square grid of size $\sqrt{N} \times \sqrt{N}$, as depicted in Fig. 6.3. The 1D Hamiltonian H_1 is an MBL Hamiltonian of the form in Eq. (6.1) with coupling coefficients given by

$$J_{i_1} = h_{i_1} = \mathcal{O}(1), \tag{6.15}$$

$$J_{i_1 i_2} = \begin{cases} -e^{-\frac{\sqrt{N}}{\xi}} & i_2 - i_1 = 1,\ i_1 \neq 0 \bmod \sqrt{N} \\ -e^{-\frac{\sqrt{N}}{\xi}} & i_2 - i_1 = \sqrt{N} \end{cases}, \tag{6.16}$$

$$J_{i_1 \dots i_p} = 0 \text{ if } p \geq 3, \tag{6.17}$$

(where we have assumed, for simplicity, $\sqrt{N}$ is an integer) and l-bits given by

$$\tau_i^z = \sigma_i^x, \tag{6.18}$$

$$\tau_i^x = \sigma_i^z. \tag{6.19}$$

The Hamiltonian H_1 clearly satisfies our definition of a canonical MBL Hamiltonian; the coupling coefficients decay sufficiently quickly, and it is easy to verify that the Hadamard gate

$U_i^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \mathrm{H}$ is unitary, satisfies Eq. (6.3) with $q = 4$, and effects Eqs. (6.18) and (6.19). It can be seen that (up to a local basis change $\sigma_i^z \leftrightarrow \sigma_i^x$) time-evolving $|0\rangle^N$ under H_1 for time $t = \pi e^{\frac{\sqrt{N}}{\xi}}/4$ is equivalent to time-evolving $|+\rangle^N$ (with $|+\rangle$ the +1 eigenstate of σ^x) under the 2D Hamiltonian $H = -\sum_{\langle i,j \rangle} \frac{\pi}{4} \sigma_i^z \sigma_j^z + \sum_i \frac{\pi}{4} e^{\frac{\sqrt{N}}{\xi}} h_i \sigma_i^z$ for time 1, where $\langle i,j \rangle$ denotes neighboring sites. If the local fields h_i are chosen randomly such that $e^{\frac{\sqrt{N}}{\xi}} h_i \in \{1, 3/2\} \bmod 4$ with equal probability, evolution under H_1 on the initial state $|0\rangle^N$ implements Architecture I of Ref. [37]. This Architecture is a Measurement-Based Quantum Computing (MBQC) scheme that is based on the hardness of IQP sampling.[8] Essentially, a disordered product state is prepared on a 2D grid after which controlled σ^z gates are applied across each edge and a measurement in the σ^x is performed. Sampling from the output distribution of this scheme is hard assuming two plausible complexity-theoretic conjectures (namely: the Polynomial Hierarchy is infinite and approximating partition functions of Ising models is average-case hard—the original paper contained a third conjecture related to anticoncentration of certain classes of random circuits, but this conjecture was proven in a later work [152]). Therefore, for times $t = \Omega(e^{\sqrt{N}/\xi})$, Problem 6.1 is hard, assuming certain plausible conjectures in computational complexity [27, 37, 153, 154].

Recent work in Ref. [121] allows us to extend $\delta = 1/2$ to any $0 < \delta < 1$. Because Architecture I of Ref. [37] may be implemented on any rectangular grid with non-constant dimensions, we may sculpt an effective 2D grid of size $N^\delta \times N^{1-\delta}$, where the long-range coefficients in Eq. (6.15) now couple sites at a distance of only N^δ. The rest of our arguments go forward unchanged, except the time it takes to implement the architecture is now exponential in N^δ/ξ. □

[8] We could choose h_i such that this relationship holds exactly, i.e. without working modulo 4, but that would require on-site terms that are exponentially small in the system size.

Theorem 6.2 thus proves that there is a family of MBL Hamiltonians that are hard to classically simulate after an exponentially long evolution time. Note that while it is hard to simulate this particular family of Hamiltonians in the average case, per the results in Ref. [37], we observe that this family of Hamiltonians is itself somewhat fine-tuned. We therefore say that classically simulating MBL Hamiltonians for exponentially long evolution time is hard in the worst case. However, Theorem 6.1 provided a quasipolynomial time algorithm to simulate MBL Hamiltonians for polynomially long evolution time, even in the worst case (as the results were indepdendent of the couplings and quasilocal unitary definining the Hamiltonian). Together, Theorems 6.1 and 6.2 point toward a possible transition in the classical worst-case hardness of Problem 6.1 between polynomial and exponential evolution times (and prove such a transition between logarithmic and exponential times for Hamiltonians that are short-range in the p-bit basis). Furthermore, Theorem 6.2 stands in stark contrast to the easiness result from Ref. [119] that single-particle localized systems of bosons admit an efficient sampling algorithm for all evolution times. However, it matches the intuition behind the hardness result in Ref. [119], where sampling free boson systems becomes difficult when the system is no longer approximately separable. Similarly, Problem 6.1 becomes provably hard when the system has evolved sufficiently for entanglement to spread across a distance scaling polynomially with N, where this long-range entanglement means the state of the system is no longer approximately separable [155].

6.6 Quantum Complexity of Simulating MBL Systems

In this section, we focus on the quantum circuit complexity of approximately implementing the time-evolution operation e^{-iHt} for an MBL Hamiltonian H.

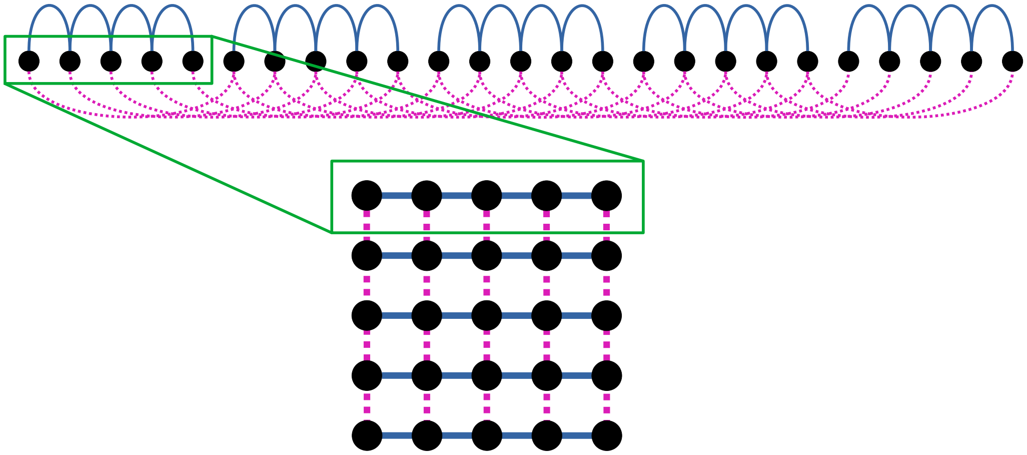

Figure 6.3: Example illustrating the 1D-to-2D mapping of a Hamiltonian H with coefficients given in Eq. (6.15) acting on $N = 25$ qubits. The solid blue (dotted pink) lines depict the interactions with $|i_1 - i_2| = 1$ ($|i_1 - i_2| = \sqrt{N}$) in the true 1D lattice. While the interactions differ in their locality, they have the same magnitude for simplicity in implementing the proposed architecture. The single-site terms are not depicted.

Definition 6.2 (Approximate circuit complexity)**.** *The ϵ-approximate circuit complexity C_ϵ of a unitary U is the minimum circuit size k of a circuit $G = G_k \dots G_2 G_1$ composed of the standard gate set containing* CNOT*, Hadamard, and $\pi/8$-phase gates ($\{\mathrm{CNOT}, \mathrm{H}, \mathrm{T}\}$) that approximates U up to error ϵ. More formally, let*

$$S_\epsilon(U) = \{G = G_k \dots G_2 G_1 \textit{ such that} \tag{6.20}$$

$$\|G - U\| \leq \epsilon \textit{ and } G_i \in \{\mathrm{CNOT}, \mathrm{H}, \mathrm{T}\}\}$$

be the set of all gate decompositions of U over the standard gate set achieving error $\leq \epsilon$. For a gate decomposition G, let $|G| := k$ denote its size. Then

$$C_\epsilon(U) := \min_{G \in S_\epsilon(U)} |G|. \tag{6.21}$$

We show that for evolution under MBL Hamiltonians, the complexity growth with respect to evolution time is slower than linear, which we denote through the symbol $o(t)$[9] in the theorem below (while the gate complexity ultimately depends on the chosen gateset, the Solovay-Kitaev theorem ensures that this dependence is weak enough to not change this sublinear scaling).

Theorem 6.3 (Sublinear growth of MBL circuit complexity)**.** *For a Hamiltonian H satisfying the criterion of MBL as defined in Eq.* (6.1) *and Definition 6.1 with $\xi < 1/(4\log 2)$, the approximate circuit complexity C_ϵ for constant ϵ obeys the bound*[10]

$$C_\epsilon(e^{-iHt}) \leq \mathrm{poly}(N)\mathrm{polylog}(N^2 t) \times o(t). \tag{6.22}$$

Proof. We leverage results from Section 6.3. Our strategy to approximate the time-evolution unitary e^{-iHt} is to apply instead the truncated evolution $e^{-i\tilde{H}t}$. We have already argued that $\|e^{-iHt} - e^{-i\tilde{H}t}\| \leq \|\Delta H\| t$, so, therefore, it suffices to choose $\tilde{H}$ so that $\|\Delta H\| \leq \epsilon/t$. In order to ensure that the unitary e^{-iHt} can be applied with small circuit complexity, we make use of the fact that the (truncated) quasilocal unitary (approximately) diagonalizes the Hamiltonian:

$$e^{-i\tilde{H}t} = \tilde{U}^\dagger e^{-i\tilde{H}_\sigma t}\tilde{U}. \tag{6.23}$$

The cost of implementing the evolution under the MBL Hamiltonian comes from two parts: the first part stems from the cost of diagonalizing the Hamiltonian by implementing the quasilocal unitary $\tilde{U}$, and the second part comes from the complexity of applying time evolution under the truncated Hamiltonian in the physical basis, namely implementing $e^{-i\tilde{H}_\sigma t}$. This is the cost of

[9] Note added: See earlier footnote for a reminder of the definition of $o(t)$.

[10] Note added: We note that we can simply the middle term to $\mathrm{polylog}(N)$ because we can move the time dependence into $o(t)$ and the square on N to a constant out front.

implementing the last three sections (after the column of single-site observables) of the circuit depicted in Fig. 6.2.

The cost of applying $\tilde{U}$ can be upper bounded from the fact that it consists of gates that act on no more than $r_U = \Theta(\xi b \log N)$ many qubits at a time. In the decomposition of $\tilde{U}$ as a quasilocal unitary, there are N single-qubit unitaries, $2\lceil N/2 \rceil = \mathcal{O}(N)$ two-qubit unitaries, and so on until the last layer of $\mathcal{O}(N)$ unitaries acting on r_U qubits at a time. Every unitary acting on k qubits can be decomposed exactly into an $\mathcal{O}(k^2 2^{2k})$-long sequence of single-qubit and CNOT unitaries [156]. Using approximate synthesis algorithms over the Clifford+T gate set [157], each of the single-qubit unitaries can be further decomposed into single-qubit gates from the standard gate set at only polylogarithmic overhead in the achieved error. More precisely, the circuit complexity is upper bounded by

$$\begin{aligned} N \log\left(\delta^{-1}\right) + 4N \log\left(\delta^{-1}\right) \cdot 2^{2\cdot 2} + 9N \log\left(\delta^{-1}\right) \cdot 2^{2\cdot 3} + \dots \\ + N r_U^2 \log\left(\delta^{-1}\right) \cdot 2^{2\cdot r_U}, \end{aligned} \tag{6.24}$$

where δ is the error made in approximating each local unitary. The terms in Eq. (6.24) correspond sequentially to the complexity of simulating the single-site, two-site, ..., r_U-site terms. The first term does not contain the factor 2^{2k} because it corresponds to single-qubit unitaries. The total error made in approximating $\tilde{U}$ then sums to

$$\delta \times (N + 4N \cdot 2^{2\cdot 2} + 9N \cdot 2^{2\cdot 3} + \ldots r_U^2 N \cdot 2^{2\cdot r_U}) \tag{6.25}$$

$$\leq \delta N \times (1^2 \cdot 4^1 + 2^2 \cdot 4^2 + 3^2 \cdot 4^3 + \ldots {r_U}^2 \cdot 4^{r_U}) \tag{6.26}$$

$$= N\delta \times \frac{4}{27}\left((9r_U^2 - 6r_U + 5)4^{r_U} - 5\right) \tag{6.27}$$

$$\leq 2N\delta r_U^2 4^{r_U}, \tag{6.28}$$

which we set to be $\epsilon/6$ by choosing $\delta = \epsilon/(12Nr_U^2 4^{r_U})$. Hence

$$C_{\epsilon/6}(\tilde{U}) \leq N\log\left(\delta^{-1}\right) \times (4 + 4\cdot 4^2 + 9\cdot 4^3 + \ldots r_U^2 \cdot 4^{r_U}) \tag{6.29}$$

$$= \mathcal{O}(N\log\left(\delta^{-1}\right) r_U^2 4^{r_U}) \tag{6.30}$$

$$= \mathcal{O}\left(N4^{r_U} r_U^2 \left(r_U \log(4) + \log\left(\frac{12Nr_U^2}{\epsilon}\right)\right)\right). \tag{6.31}$$

The cost of implementing $e^{-i\tilde{H}_\sigma t}$ can also similarly be upper bounded. Here, for simplicity, we use the decomposition of $e^{-i\tilde{H}_\sigma t}$ from Section 6.4, where we combined unitaries acting on site i (but not before i) into $\tilde{V}_i$. This decomposition has N unitaries of size at most r_J, meaning the gate complexity for $e^{-i\tilde{H}_\sigma t}$ is upper bounded by $\mathcal{O}(N\log(\delta^{-1})r_J^2 4^{r_J})$, and the total error made in approximating these gates is thus $\mathcal{O}(N\delta r_J^2 4^{r_J})$. We again set this error equal to $\epsilon/6$ with a choice now of $\delta = \epsilon/(12Nr_J^2 4^{r_J})$, similarly yielding a gate complexity of

$$C_{\epsilon/6}(e^{-i\tilde{H}_\sigma t}) = \mathcal{O}\left(N4^{r_J} r_J^2 \left(r_J \log(4) + \log\left(\frac{12Nr_J^2}{\epsilon}\right)\right)\right). \tag{6.32}$$

Combining everything, the total error for implementing the decomposition in Eq. (6.23) is

$\epsilon/6 \times 3 = \epsilon/2$. The total error in implementing e^{-iHt} is thus upper bounded by the sum of the error in approximating e^{-iHt} by $e^{-i\tilde{H}t}$ plus the error in decomposing $e^{-i\tilde{H}t}$ into a sequence of single and two-qubit gates:

$$\epsilon/2 + \|\Delta H\| t \leq \epsilon/2 + tC_J N r_J e^{-kr_J} + tC_U N^2 e^{-\frac{r_U}{2\xi}}, \tag{6.33}$$

where we used Eq. (6.11) to bound the second term. We make the choices $r_J = (1.01)\log(Nt)/k$ and $r_U = 2.02\xi\log(N^2 t)$ so that the total error is at most

$$\begin{aligned} &\epsilon/2 + C_J (Nt)^{-0.01}\log(Nt)/k + C_U (N^2 t)^{-0.01} \\ &< \epsilon. \end{aligned} \tag{6.34}$$

With these choices, the total gate cost of simulating the entire circuit becomes $2C_{\epsilon/6}(\tilde{U}) + C_{\epsilon/6}(e^{-i\tilde{H}_\sigma t})$:

$$\begin{aligned} C_\epsilon(e^{-iHt}) \leq \mathcal{O}\big(&N(N^2 t)^{2.02\xi\log 4}\text{polylog}(N^2 t) \\ &+ N(Nt)^{1.01\log 4/k}\,\text{polylog}(Nt)\big). \end{aligned} \tag{6.35}$$

As long as $\xi < 1/(2.02\log 4) = 1/(4.04\log 2)$, the exponent of t in the first term is smaller than 1. The same choice also ensures that the exponent of t in the second term is smaller than 1 because $1.01\log 4/k = 1.01\log 4/(1/\xi - \log 2) < 2.02/3.04 < 1$. □

Thus, for sufficiently localized MBL Hamiltonians, the quantum circuit complexity is sublinear in time. Such sublinear scaling contrasts MBL systems with chaotic Hamiltonians, which are conjectured to have quantum circuit complexity growing linearly with time, as supported by recent work in [38, 137].[11] This provides a complexity-theoretic understanding of why MBL systems are unlikely to generate such chaotic dynamics. This conclusion is intuitively consistent with the slow logarithmic spread of entanglement that is characteristic of MBL systems.[12]

6.7 Conclusion and Outlook

In this Chapter, we have developed the best known formal results on the complexity of simulating MBL systems. We have applied results in the literature to show that MBL systems evolved for time logarithmic in the system size admit an efficient classical strong simulaion, and, hence, sampling, algorithm. Further, we have demonstrated a quasipolynomial-time algorithm that can strongly simulate sufficiently localized MBL systems that have evolved for any (quasi)polynomially long time. While we have not quite provided a polynomial-time algorithm, the quasipolynomial-time algorithm is suggestive that possible improvements may lead to a formal proof of easiness. In particular, either the algorithm may be improved, potentially by leveraging the work on spectral tensor networks in Refs. [145–147] to make formal complexity statements in the case of quasilocal integrals of motion, or it may be possible to develop an algorithm that samples directly instead of going through the harder task of strong simulation. We leave these possible improvements (or the proof that they are impossible) as important open questions for future work. Furthermore, our proof holds only for Hamiltonians with LIOMs that

[11] Note added: Again, a more complete (but not exhaustive) list of works studying this topic include Refs. [38, 137–141].

[12] Note added: Furthermore, this result also has implications in the complexity literature on "fast-forwarding." Specifically, fast-forwarding means quantumly simulating the evolution of a system for time t in such a way that the simulation time is parametrically smaller than t. Ref. [142] found that both commuting local Hamiltonians and Anderson-Localized Hamiltonians may both be fast-forwarded, but left for future work whether this also holds true for MBL Hamiltonians. Our results immediately provides an answer in the affirmative, thus settling this open question.

are highly localized to a distance of about $\xi < 1/\log 2$, in units of the lattice spacing. We do not consider this restriction to be too problematic, as previous work, e.g., Ref. [158], has demonstrated that LIOMs may need to be highly localized for MBL systems to remain stable. It would be interesting, however, to understand more fully if this restriction is an artifact of our techniques, or if it is explained by some physical transition in MBL systems. Additionally, all of our results are based on bounding the worst-case scenario without explicitly accounting for disorder in our couplings, and studying the effect of disorder is an interesting open question. Finally, it is also crucial to explore the easiness of simulating MBL systems when one only has access to H in the p-bit basis.

Apart from our easiness results, we have shown by a comparison to the problem of sampling from IQP circuits that a family of random MBL systems becomes hard to simulate after a time exponentially long in the system size. This family, while entirely consistent with our definition of MBL, is rather fine-tuned and likely has little overlap with the family of MBL Hamiltonians induced by disorder in the physical basis. Therefore, it would be quite valuable to determine in future work whether average-case hardness at exponential evolution times also holds for a more natural family of disorder-induced MBL Hamiltonians.

Additionally, we have also detailed the gate complexity of quantum simulation of MBL systems, and we have shown that for systems with localization length $\xi < 1/(4\log 2)$, this gate complexity is sublinear. As for our results on classical simulation, it would be interesting to determine whether this localization length restriction is an artifact of our proof techniques or is physical. It would also be enlightening to investigate the connection between these results and the literature on fast-forwarding Hamiltonian evolution [142].[13]

[13] Note added: As per some of the other added notes, it is now more clear that this result resolves this open

Finally, so far we have specified entirely to MBL systems defined in 1D. Indeed, there is significant debate over whether disorder-induced MBL can even exist in higher dimensions [36] (for example, the proof of MBL and LIOM structure in Ref. [132] relies crucially on the 1D nature of the system). However, the natural generalization of our definition of MBL to higher dimensions would allow for MBL Hamiltonians that implement Architecture I of Ref. [37] directly (i.e., without sculpting an effective 2D grid using exponentially decaying interactions) in constant time. Thus, sampling from higher-dimensional MBL systems becomes hard very quickly, after evolution time $t = \mathcal{O}(1)$. However, other less natural extensions might exclude fast implementations of Architecture I, so the hardness of simulating higher-dimensional MBL systems still deserves further examination.

Acknowledgments

We thank Eli Chertkov, Elizabeth Crosson, Bill Fefferman, James Garrison, Vedika Khemani, Nishad Maskara, Paraj Titum, Minh C. Tran, and Brayden Ware for helpful discussions. A. E., C. L. B., and A. V. G. acknowledge funding from the DoD, DoE ASCR Accelerated Research in Quantum Computing program (award No. DE-SC0020312), NSF PFCQC program, AFOSR, DoE QSA, NSF QLCI (award No. OMA-2120757), DoE ASCR Quantum Testbed Pathfinder program (award No. DE-SC0019040), AFOSR MURI, U.S. Department of Energy Award No. DE-SC0019449, ARO MURI, and DARPA SAVaNT ADVENT. A. D. acknowledges support from the National Science Foundation RAISE-TAQS 1839204 and Amazon Web Services, AWS Quantum Program. This research was performed in part while C. L. B. held an NRC

question, so we would amend this to say "It would also be enlightening to investigate more deeply the various implications of the fact that MBL Hamiltonians may be fast-forwarded."

Research Associateship award at the National Institute of Standards and Technology. The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center PHY-1733907. D. A. acknowledges support from the Swiss National Science Foundation and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 864597).

Chapter 7: Discussion

In this dissertation, we have studied the power and limits of quantum technology with a specific emphasis on determining the capabilities of quantum sensors and quantum simulators as compared to unentangled sensors and classical simulators, respectively. For quantum sensors measuring a linear function of unknown parameters, we have considered bounds on their ultimate performance (derived either ourselves or already in the literature) and determined optimal protocols that saturate them. In doing so, we have also determined more precisely what resources are needed to run these protocols, clarifying the role of quantum entanglement in achieving metrological performance gains. When it comes to Gaussian Boson Sampling, a form of simulation of some photonic quantum systems, we have made strong progress on showing when anticoncentration, an important part of state-of-the-art hardness arguments, may or may not hold in certain regimes. We have also worked to outline when Many-Body Localized Hamiltonians may or may not be classical simulable. Taken together, these results help better define when and how we can expect quantum devices to outperform classical ones.

Substantial questions, of course, remain to be answered. Indeed, each Chapter ends with a discussion of the results therein and also poses exciting open questions that remain, which we now partially review and expand upon.

To begin with a more general comment, the Chapters in this dissertation have focused almost entirely on highly idealized scenarios in the asymptotic regimes of sensing and simulation where there is no noise. While it is important to understand these ideal scenarios (which, in many cases, are already sufficiently mathematically complicated), they can not perfectly reflect how real, physical quantum devices operate, which means that these works, taken alone, cannot fully describe achievable quantum advantage.

We shall now be a bit more precise, starting with our discussion of quantum sensing. In Chapters 2 and 3, we work in the idealized limit of no noise and where our sensors are coupled to the parameters of interest by simple commuting operators. Extending beyond this regime to deal with noise or more complicated (perhaps even non-commuting) generators is an important aspect of future work. We have taken some steps toward understanding this in other works in the Gorshkov group—for example, in Ref. [159], we show how erasure errors, which are errors that are easily detectable and take a quantum state outside of a computational subspace, allow for more precise sensing than other errors, such as depolarization or dephasing, that keep one inside the computational subspace (assuming that the noise levels are equivalent). However, there still remain questions about how noise affects the protocols and bounds we have developed in this dissertation. Specifically, it is not clear that our protocols, or even the overarching structure of our protocols, are still optimal once noise is introduced.[1] However, much work has been done connecting the error-correction schemes that are so crucial to fault-tolerant quantum computation with various quantum sensing schemes in order to combat the effects of noise—see, e.g., [160–165]. To the best of our knowledge, these works have focused on either the single-parameter scenario or the fully multi-parameter scenario. Because the task of single function

[1] In the case of depolarizing noise, which, intuitively, affects all states in an equal way, there is good evidence to suggest that our protocols still are optimal (up to a loss of precision due to the noise); see Ref. [159] for a discussion of this. However, other noise models require more attention. We thank members of the Committee for pointing this out.

estimation actually sits somewhere in between these two scenarios, understanding how to connect these results for noisy sensing with the algebraic framework, optimal protocols, and resource estimations that we have derived in the quantum sensing portion of this dissertation is an important open question. It also is becoming increasingly timely as these error-correction schemes become more and more prevalent—consider the recent demonstration in Ref. [166].[2]

Furthermore, our discussion of the power of quantum sensors in Chapters 2 and 3 is based on the quantum Cramér-Rao bound, but this bound is only saturable in the limit of an asymptotic number of samples. While the robust phase estimation protocols of Refs. [71–73] are able to partially address this issue, using a finite number of samples comes at the expense of a multiplicative constant. Recent work in the literature has looked at moving away from this framework based on the quantum Cramér-Rao bound by using more sophisticated statistics and schemes. In particular, the estimation process in Refs. [71–73] that we use for the measurement and classical post-processing stages of our protocols is non-adaptive, meaning it does not change based on the samples that are received. There is evidence that a more Bayesian approach, where future measurements are conditioned in some way on the results of previous ones, might help improve these schemes at the cost of a more complex protocol, especially in the regime of a non-asymptotic number of samples. For example, a Bayesian approach can be used in both the single parameter [167] and fully multiparameter [92] scenarios in regimes where other bounds, such as the quantum Cramér-Rao and Holevo Cramér-Rao bounds fail to be valid. Investigating how our work fits into these Bayesian schemes is a useful direction for future work.

[2] Note that combining error correction and sensing is a non-trivial endeavor. In particular, not all error-correcting codes that work for computation can immediately be lifted to the sensing scenario, as one must ensure that one corrects only the noise, but not the fragile signal. See, e.g., Ref. [161] for a discussion of this. Therefore, we are not claiming that the particular code utilized in Ref. [166] is directly useful for quantum sensing; we merely point out that quantum error correction schemes in general are getting closer to reality.

Tying these threads together, we are working on understanding how the protocols and framework that we develop in Chapters 2 and 3 can be converted into truly useful, end-to-end applications. Specifically, we are considering how we might use the protocols that we have developed to measure quickly and accurately the gravitational distribution of an active volcano [168]. Doing this requires addressing all of the concerns we have mentioned above.

In Chapters 4 and 5, we work in the idealized setting of Gaussian Boson Sampling with no noise and fully number-resolving detectors. The outcome distribution changes if one assumes either some sort of experimental imperfection (such as photon loss or photons no longer being fully indistinguishable particles) or if the measurement devices only detect the presence or absence of photons, but not how many are in any particular mode. In particular, the combinatorial function that defines the output probabilities (which is the hafnian in idealized Gaussian Boson Sampling and the permanent in idealized Fock Boson Sampling) can change in these instances. Thus, our work would need to be generalized in order to more accurately describe real experiments that operate in this noisy or otherwise less ideal regime.[3] Additionally, while the complexity-theoretic distinction between hardness and easiness that plays a crucial role in Chapters 4 and 5 is only truly defined in the asymptotic limit of resources (as these definitions are based on how computation time *scales* with these resources), any individual experiment that might seek to demonstrate quantum advantage using the Gaussian Boson Sampling scheme discussed therein operates at some finite level of resources. Therefore, understanding how our results about the scaling of

[3] We do work in the regime where collisions in the output are very rare by construction. Therefore, it is not entirely clear that a lack of number-resolving detectors alone would significantly change the results of these Chapters. Formally proving this and understanding precisely how the calculations using non-number-resolving detectors reduce to our calculations would be interesting. Furthermore, our choice to work in the non-collisional limit was partially due a significant (conjectured) simplification of the form of the output probabilities—see Conjecture 4.1. It is not inconceivable that, for non-number-resolving detectors, one can calculate the moments of the output probabilities without this simplification. Our results, then, would serve as a useful limiting case in this scenario.

anticoncentration fit into these finite-size regimes is a crucial open question.

One of the more concrete options for future investigation is in understanding more precisely how the results in Chapter 5 about the second moment and the calculation of the expected linear cross-entropy benchmarking score for an ideal sampler relate to both experimental implementations and classical spoofing of these scores. Specifically, it would be interesting to look at actual experiments that have operated in the non-collisional regime and calculated the linear cross-entropy benchmarking score and see how their results compare to the expected score for an ideal, error-free device. We could potentially use this comparison as a way of understanding the kinds of errors present in the experiment. On the other hand (and in keeping with our goal of delineating quantum vs. classical power), it is also useful to understand whether our exact calculation of the expected error-free score on linear cross-entropy benchmarking offers any insight into the classical algorithms that try to spoof this measure, such as in Ref. [169]. Can our results either help classical spoofers or show that there is some regime in which they fail? This is an extremely important and interesting direction of research.

We are also curious as to how the results in Chapters 4 and 5 relate to other research in the Gorshkov group that the author of this dissertation has been involved in, but that is not included in this dissertation. Specifically, in Refs. [170, 171], we study the so-called Page curve for entanglement in systems of Gaussian Bosonic states. The Page curve is, essentially, a measure of the average entanglement between two partitions of a quantum system as a function of the size of the partitions, where the average is taken over a suitable ensemble on the relevant Hilbert space (typically the full state is assumed to be Haar-random). It was originally defined and its functional form conjectured by Page in Ref. [172], and this conjecture was then proven in Refs. [173–175]. One can define and study Page curves in a wide variety of systems—see Ref. [176] and references therein. In the case of Refs. [170, 171], we build upon and generalize Refs. [177–179] to study Page curves and the typicality of entanglement (roughly, how likely it is that any individual random unitary creates close to the average amount of entanglement calculated in the Page curve) in a setup that is nearly identical to that of Chapters 4 and 5. Specifically, we apply a Haar-random linear optical unitary to an initial state of equally squeezed single-mode squeezed vacuum states on all modes of a photonic system and consider the average entanglement between two partitions of modes as a function of the size of this partition. Entanglement is measured using either the Rényi-2 entropy [170] or arbitrary Rényi-α entropy with $\alpha \in \mathbb{Z}_{\geq 1}$ [171]. However, there is a crucial difference between the setup discussed in this dissertation and that in Refs. [170, 171]; in the latter, *all* initial modes are equally squeezed, whereas the setup in this dissertation allows the number of squeezed modes to vary. This raises the question of whether the results of Refs. [170, 171] can be generalized to the case where only $k < m$ of m total modes are squeezed with some equal strength s, but the rest remain in vacuum. This generalization would allow us to see whether there is any interesting transition in entanglement akin to the transition in anticoncentration, thus hopefully shedding light on any connections between entanglement and complexity. Surprisingly, even this modest change to our setup seems to make the calculation of the Page curve essentially untenable. It would be interesting to see whether a change of theoretical approach might solve this problem. Alternatively, one could simply study the entanglement numerically. We are currently pursuing both of these avenues.

Finally, in the study of Many-Body Localized Hamiltonians in Chapter 6, we have made use of a purely phenomenological definition in terms of a quasilocal diagonalizing unitary that takes the Hamiltonian from its original basis to one composed of quasilocal integrals of motion. However, we have not specified physically how these integrals of motion arise. Typically, Many-

Body Localization is considered to be disorder-induced. While this disorder is not the same concept of “noise” that we have mentioned for the other topics in this dissertation, spiritually, the work in Chapter 6 is similarly disconnected from real-world examples of Many-Body Localization because of this disparity. Therefore, an important topic of future work is to try to recreate our results for a more realistic disorder-based model. Indeed, this might actually *improve* our results, as many of our proofs are based on a worst-case analysis that would be highly unlikely for disorder-induced systems. We believe that this is the central topic of future work, though a few others are also mentioned at the end of Chapter 6.

Therefore, while this dissertation has improved the understanding of quantum advantage in sensing and simulation schemes, there is still much work to be done in characterizing all of the possible benefits that quantum devices might offer. We look forward to seeing some or all of these questions addressed.

Appendix A: Appendices Associated with Chapter 2

A.1   A Useful Lemma Regarding Optimal Probe States

In this Appendix, we prove a useful lemma restricting the structure of the probe state for an optimal protocol.

Lemma A.1 *Any optimal protocol, independent of the choice of control, requires that* $\langle\hat{\mathcal{H}}_1(t)\rangle = 0$*, where* $\mathcal{H}_1(t)$ *is the time-evolved generator of the first parameter and the expectation value is taken with respect to the initial probe state. Furthermore, the probe state must be of the form*

$$|\psi\rangle = \frac{|0\rangle|\chi_0\rangle + e^{i\phi}|1\rangle|\chi_1\rangle}{\sqrt{2}}, \tag{A.1}$$

for all times $s \in [0, t]$*, where* $|\chi_0\rangle, |\chi_1\rangle$ *are arbitrary states on the* $d-1$ *remaining sensor qubits plus, potentially, the arbitrary number of ancillas, and* ϕ *is an arbitrary phase in* $\mathbb{R}$*—they can be* s*-dependent.*

Proof. Consider the expression for the matrix elements of the quantum Fisher information matrix at time t [Eq. 2.4]:

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4[\frac{1}{2}\langle\{\hat{\mathcal{H}}_i(t), \hat{\mathcal{H}}_j(t)\}\rangle - \langle\hat{\mathcal{H}}_i(t)\rangle\langle\hat{\mathcal{H}}_j(t)\rangle], \tag{A.2}$$

where the expectation values are taken with respect to the initial probe state $|\psi(0)\rangle$. Using the integral form of $\hat{\mathcal{H}}_j(t)$ (Eq. 2.5), we can write

$$\mathcal{F}(\boldsymbol{\theta})_{11} = 4\mathrm{Var}\left[\hat{\mathcal{H}}_1(t)\right] \tag{A.3}$$

$$= 4\left[\int_0^t ds \int_0^t ds' \langle\psi(0)| \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)\hat{U}^\dagger(s')\hat{g}_1\hat{U}(s') |\psi(0)\rangle\right] \tag{A.4}$$

$$- 4\left[\int_0^t ds \langle\psi(0)| \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s) |\psi(0)\rangle\right]^2$$

$$= 4\int_0^t ds \int_0^t ds' \mathrm{Cov}_{|\psi(0)\rangle}[\hat{g}_1(s), \hat{g}_1(s')], \tag{A.5}$$

where we recall

$$\hat{g}_1(s) := \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s), \tag{A.6}$$

and $\hat{g}_1 = \partial\hat{H}/\partial\theta_1$ is the initial generator with respect to the first parameter. Once again, the covariance is with respect to the initial probe state $|\psi(0)\rangle$. We can then upper bound this as

$$\mathcal{F}(\boldsymbol{\theta})_{11}(t) \leq 4\int_0^t ds \int_0^t ds' \sqrt{\mathrm{Var}_{|\psi(0)\rangle}[\hat{g}_1(s)]\mathrm{Var}_{|\psi(0)\rangle}[\hat{g}_1(s')]} \tag{A.7}$$

$$= 4\left[\int_0^t ds \sqrt{\mathrm{Var}_{|\psi(0)\rangle}[\hat{g}_1(s)]}\right]^2 \tag{A.8}$$

$$\leq \left[\int_0^t ds \|\hat{g}_1\|_s\right]^2 \tag{A.9}$$

$$= t^2\|\hat{g}_1\|_s^2 \tag{A.10}$$

$$= t^2, \tag{A.11}$$

where the first inequality bounds the covariance as the square root of the product of the variances, the second inequality bounds the standard deviation of an operator by half the seminorm [51], and the final equality uses the fact that $\hat{g}_1 = \hat{\sigma}_1^z/2$ has seminorm 1.[1] [2]

Via Eq. (2.8) (rigorously derived in Appendix A.6) we know that an optimal protocol must have $\mathcal{F}_{11}(\boldsymbol{\theta})(t) = t^2$. Therefore, an optimal protocol must saturate the inequalities in Eq. (A.7) and Eq. (A.9). Equation (A.9) is saturated when $\mathrm{Var}[\hat{g}_1(s)] = \|\hat{g}_1(s)\|_s = \|\hat{g}_1\|_s$ for all s. This holds if and only if $|\psi(0)\rangle = \frac{1}{\sqrt{2}}(|\lambda_{\mathrm{min}}\rangle + e^{i\phi}|\lambda_{\mathrm{max}}\rangle)$, where $|\lambda_{\mathrm{min}}\rangle$ and $|\lambda_{\mathrm{max}}\rangle$ are the eigenstates corresponding to the minimum and maximum eigenvalues of $\hat{g}_1(s)$ for all $s \in [0,t]$ and ϕ is an arbitrary phase. Given this condition, $\hat{g}_1(s)$ and $\hat{g}_1(s')$ act identically on the state $|\psi(0)\rangle$ and consequently are fully correlated when one considers the covariance of these operators with respect to the state. The Cauchy-Schwarz inequality in Eq. (A.7) is immediately saturated as well.

Importantly, under this condition on the probe state, any operator in the one-parameter family $\hat{g}_1(s) = \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)$ acts identically on $|\psi(0)\rangle$ (the unitary does not change the eigenvalues, and the eigenstates are shared by all $\hat{g}_1(s)$, as argued above). Thus, one can freely substitute any operator in the one-parameter family $\hat{g}_1(s) = \hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)$ for another. Therefore, for such an optimal probe state,

$$\langle\mathcal{H}_1(t)\rangle = -\int_0^t ds\,\langle\psi(0)|\,\hat{g}_1(s)\,|\psi(0)\rangle = t\langle\hat{g}_1\rangle = 0 \tag{A.12}$$

because $\hat{g}_1 \propto \hat{\sigma}_1^z$ and, consequently, by the argument that we can replace $\hat{g}_1$ by $\hat{g}_1(s)$ when acting on the probe state,

$$\langle\psi(s)|\,\hat{g}_1\,|\psi(s)\rangle = 0 \quad (\forall s). \tag{A.13}$$

[1] Note that the above block of equations relies on the fact that we are using the fixed Hilbert space of qubit sensors. Were one to extend this derivation to photonic sensors with indefinite particle number, the results would not immediately follow.

[2] Note added: We address the generalization to photonic sensors in Chapter 3 and Appendix B.

The statement of the lemma immediately follows. □

Note that Lemma A.1 holds for any optimal protocol, not just those using our catlike states. However, it also justifies our choice of probe states and why we specifically set $\tau_1 = 1$ for all $\boldsymbol{\tau}$ (i.e., to maintain an equal superposition between $|0\rangle$ and $|1\rangle$ on the first qubit).

A.2 Proof of the Optimality of Cat-State Protocols

In this Appendix, we will rigorously prove the optimality of the time-dependent protocols considered in Chapter 2. In particular, we show that the Fisher information matrix condition for saturability in Eq. (2.8) is satisfied by solutions to Eq. (2.13) when we consider protocols that use $\hat{\sigma}^x$ and CNOT controls to switch between families of catlike states in $\mathcal{T}$. That is, we show the following mapping between saturability conditions:

$$T\boldsymbol{p} = \frac{\boldsymbol{\alpha}}{\alpha_1} \quad \Longrightarrow \quad \mathcal{F}(\boldsymbol{\theta})_{1j} = \frac{\boldsymbol{\alpha}}{\alpha_1} t^2, \tag{A.14}$$

where we recall that we have assumed that $|\alpha_1| = \|\boldsymbol{\alpha}\|_\infty > |\alpha_j|$ for all $j > 1$ (in Appendix A.6, we will generalize beyond the assumption of a single maximum magnitude α_j at the cost of some notational inconvenience).

Using Lemma A.1, we can show that for *any* optimal protocol (i.e., not just those using our cat-like states)

$$
\begin{aligned}
\mathcal{F}(\boldsymbol{\theta})_{1j} &= 2\langle\{\hat{\mathcal{H}}_1, \hat{\mathcal{H}}_j\}\rangle && \text{(A.15)}\\
&= 2\int_0^t ds \int_0^t ds' \langle\psi(0)| \{\hat{g}_1(s), \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\} |\psi(0)\rangle && \text{(A.16)}\\
&= 2\int_0^t ds \int_0^t ds' \langle\psi(0)| \{\hat{g}_1, \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\} |\psi(0)\rangle && \text{(A.17)}\\
&= 2t\int_0^t ds' \langle\psi(0)| \{\hat{g}_1, \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\} |\psi(0)\rangle && \text{(A.18)}\\
&= 2t\int_0^t ds' \langle\psi(0)| \{\hat{g}_1(s'), \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\} |\psi(0)\rangle && \text{(A.19)}\\
&= 4t\int_0^t ds' \langle\psi(s')| \hat{g}_1\hat{g}_j |\psi(s')\rangle && \text{(A.20)}\\
&= t\int_0^t ds' \langle\psi(s')| \hat{\sigma}_1^z\hat{\sigma}_j^z |\psi(s')\rangle . && \text{(A.21)}
\end{aligned}
$$

The third and fifth equalities come from the argument in the proof of Lemma A.1 that we may replace $\hat{g}_1(s)$ with $\hat{g}_1$ (and vice versa) when acting on optimal probe states. The penultimate equality is just a consequence of the commutativity of the initial generators.

We now apply these general results to our specific protocols. Saturating the initial Fisher information conditions in Eq. (A.14) implies that we must show

$$\int_0^t ds' \langle\psi(s')| \hat{\sigma}_1^z\hat{\sigma}_j^z |\psi(s')\rangle = \frac{\alpha_j}{\alpha_1}t. \tag{A.22}$$

Let the gates in our protocols be labeled as $\hat{G}_i$ where $\hat{G}_i$ is either a CNOT or $\hat{\sigma}^x$ gate. The gate $\hat{G}_i$ is applied at a time $s = t_i^*$. Then, for $s \in (t_k^*, t_{k+1}^*)$, we can write the time-dependent state as

$$|\psi(s)\rangle = \left|\psi(\boldsymbol{\tau}^{(k)};\varphi)\right\rangle \equiv \prod_{i=0}^{k}\hat{G}_i\left|\psi(\boldsymbol{\tau}^{(0)};\varphi)\right\rangle, \tag{A.23}$$

where $|\psi(\boldsymbol{\tau}^{(0)};0)\rangle$ is the initial state of the protocol, φ is the relative phase between the two branches of the state that has accumulated up to time s, and, therefore, $|\psi(\boldsymbol{\tau}^{(k)};\varphi)\rangle$ is the state produced after applying the first k gates. Because our protocols explicitly use only $\hat{\sigma}^x$ and CNOT gates to move between families in $\mathcal{T}$, we have that $|\psi(\boldsymbol{\tau}^{(k)};\varphi)\rangle = (|0\rangle|\chi_0^{(k)}\rangle + e^{i\varphi}|1\rangle|\chi_1^{(k)}\rangle)/\sqrt{2}$, and

$$\int_0^t ds' \langle\psi(s')|\hat{\sigma}_1^z\hat{\sigma}_j^z|\psi(s')\rangle = \sum_{i=0}^{n}(t_{i+1}^* - t_i^*)\tau_j^{(i)}, \tag{A.24}$$

where we implicitly define $t_0^* = 0$ and $t_{n+1}^* = t$ as the initial and final times of the protocol and $|\chi_0^{(k)}\rangle$ and $|\chi_1^{(k)}\rangle$ are some states defined on the Hilbert space which excludes the first qubit sensor. The time $t_{i+1}^* - t_i^*$ corresponds to the time we are in the probe family $|\psi(\boldsymbol{\tau}^{(i)};\varphi)\rangle$, which in our protocols is $p_i t$. Thus, to satisfy the Fisher information conditions, we need

$$\sum_i p_i\tau_j^{(i)} = \frac{\alpha_j}{\alpha_1} \implies (T\boldsymbol{p})_j = \frac{\alpha_j}{\alpha_1}. \tag{A.25}$$

This formally proves optimality of our time-dependent protocols that satisfy $T\boldsymbol{p} = \boldsymbol{\alpha}/\alpha_1$.

A.3 Review of Robust Phase Estimation

In this Appendix, we review, for completeness, the phase estimation protocols of Refs. [71–73] described in Chapter 2 as a method to extract the quantity of interest, q, from the state

$$1/\sqrt{2}(|0\rangle + e^{iqt/\alpha_1}|1\rangle)(|0\ldots0\rangle), \tag{A.26}$$

which is the final state obtained from our family of optimal protocols.

Again, when we refer to our protocols as optimal, we mean this in the sense that our protocols achieve the conditions on the quantum Fisher information matrix that allow the maximum possible quantum Fisher information with respect to the parameter q to be obtained. However, to completely specify the procedure by which one obtains the quantity q, an explicit phase estimation protocol is needed. As explained in Chapter 2, such a task is complicated by the fact that for large times and/or small $\alpha_1 = \|\boldsymbol{\alpha}\|_\infty$, it is unclear what 2π interval the relative phase between the branches of Eq. (A.26) is in [74, 75]. The phase estimation protocols of Refs. [71–73] demonstrate how to optimize resources to deal with this issue, while still saturating the single-shot bound in Eq. (2.2) up to a small d- and t-independent constant. In particular, such protocols allow us to reach a mean square error of

$$\mathcal{M} = \frac{c^2 \|\boldsymbol{\alpha}\|_\infty^2}{t^2}, \tag{A.27}$$

for some small (explicitly known) constant c. Reference [76] proves that this constant factor c^2 in Eq. (2) can be reduced to, at best, π^2.

While reviewing such phase estimation protocols, we follow the presentation of Ref. [73], which corrects a few minor errors in Ref. [71], as noted in the corresponding erratum [72]. We refer the reader to Ref. [73] for further details. Conveniently, by putting the final state into the form of Eq. (A.26), we have reduced this problem completely to the single qubit, multipass version of the problem described in that reference. Consequently, everything follows practically identically to their presentation.

Consider dividing the total time t, which is the relevant resource in our problem, into K stages where we evolve for a time $M_j \delta t$ in the jth stage (δt is some small basic unit of time and $M_j \in \mathbb{N}$). We assume that we have (d, t)-independent, prior knowledge of q such that we can set

δt to satisfy

$$\frac{\delta t q}{\|\boldsymbol{\alpha}\|_\infty} \in [0, 2\pi). \tag{A.28}$$

In the jth stage, using one of our protocols for a time $M_j \delta t$, we prepare $2\nu_j$ independent copies of the state

$$|\psi_j\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + e^{iqM_j\delta t/\|\boldsymbol{\alpha}\|_\infty} |1\rangle \right) |0 \ldots 0\rangle . \tag{A.29}$$

From now on we will drop the $d - 1$ qubit sensors in the state $|0 \ldots 0\rangle$, as they are irrelevant; however, it is worth noting that it is not necessary to put the state in this form before performing measurements. We do so to make the comparison to Ref. [73] particularly transparent. We then perform a single-qubit measurement on the first qubit sensor of each of these state copies, yielding $2\nu_j$ measurement outcomes, which we can use to estimate q. The total time of this K stage protocol is consequently given by

$$t = 2 \sum_{j=1}^{K} \nu_j M_j \delta t. \tag{A.30}$$

Given this setup, we choose single-qubit measurements and optimize the choice of ν_j, M_j per stage so that we can learn q bit by bit, stage by stage, in such a way that optimal scaling in d, t is still obtained [Eq. (A.27)]. In particular, consider making two measurements, each ν_j times per stage (thus explaining the factor of 2 we introduced earlier): (i) a $\hat{\sigma}^x$ measurement and (ii) a $\hat{\sigma}^y$ measurement. These measurements each give us outcomes that are Bernoulli variables (i.e., with values $\in \{0,1\}$) with outcome probabilities

$$\begin{aligned}
p^{(x)}(0) &= \frac{1+\cos\left(M_j q\delta t/\|\boldsymbol{\alpha}\|_\infty\right)}{2}, \\
p^{(x)}(1) &= 1 - p^{(x)}(0), \\
p^{(y)}(0) &= \frac{1+\sin\left(M_j q\delta t/\|\boldsymbol{\alpha}\|_\infty\right)}{2}, \\
p^{(y)}(1) &= 1 - p^{(y)}(0),
\end{aligned} \tag{A.31}$$

where the first two probabilities are for the $\hat{\sigma}^x$ measurement and the latter two are for the $\hat{\sigma}^y$ measurement. Using both of these measurements allows us to resolve the twofold degeneracy in the phase $qM_j\delta t/\|\boldsymbol{\alpha}\|_\infty$ within a given $[0,2\pi)$ interval that would arise from, e.g., a $\hat{\sigma}^x$ measurement alone. The observed probabilities of obtaining 0 for the $\hat{\sigma}^x$ and $\hat{\sigma}^y$ are independent random variables that converge in probability to their associated expectation values for $\nu_j \to \infty$. Let these observed probabilities be labeled $f_0^{(x)}$ and $f_0^{(y)}$, respectively. These measurements are nonadapative, which makes this particular phase estimation protocol especially appealing.

At each stage, we extract an estimator $\tilde{\phi}$ of $\phi := M_j q\delta t/\|\boldsymbol{\alpha}\|_\infty$ as

$$\tilde{\phi} := \mathrm{atan2}(2f_0^{(y)} - 1, 2f_0^{(x)} - 1) \in [0, 2\pi), \tag{A.32}$$

where $\mathrm{atan2}$ is the two-argument arctangent with range $[0,2\pi)$. In the limit $\nu_j \to \infty$, this estimator indeed converges to ϕ, but the "magic" of this phase estimation scheme lies in the correct reprocessing of data stage-by-stage so that ν_j can be kept (d,t) independent. Reference [73] demonstrates rigorously that picking $M_j = 2^{j-1}$ for $j \in \{1,\ldots,K\}$ and optimizing over ν_j one can, at each stage, estimate $q/\|\boldsymbol{\alpha}\|_\infty$ with a confidence interval of size $2\pi/(3\times 2^{j-1})$ so that in each stage we learn another bit of this quantity. The results of this optimization are ν_j that decrease linearly with the step j so that as the time spent in a stage grows, the statistics we employ shrink. Importantly, it so happens that we can scale $K \to \infty$ (i.e., take an asymptotic-in-t limit) while maintaining ν_K constant. The net result is a mean-square error given by Eq. (A.27) with $c = 24.26\pi$, which is a factor of 24.26 greater than the theoretical optimal value [76], but with the convenient feature that the protocol uses nonadaptive measurements. We refer the interested reader to Ref. [73] for detailed derivation of the results sketched here.

It is also worth noting that other protocols are possible. For instance, in Ref. [70], a similar two-step method is described for the estimation of global parameters (i.e., where the parameter is not restricted to a local neighborhood of parameter space). This protocol provides an explicit method to use some (ultimately negligible) fraction of the sensing time available to narrow down the location of the parameter q in parameter space, followed by an optimal local estimation. We emphasize that the explicit estimation scheme we propose (i.e., the one in Refs. [71–73]) does not require adaptive measurements, which is one of its key advantages.

A.4 Full Proof of the Main Theorem

In this Appendix, we expand on the proof sketch of Theorem 2.1 to fully prove the result. For reference, this theorem is restated here.

Theorem A.1. *Let $q(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$. Without loss of generality, let $\|\boldsymbol{\alpha}\|_\infty = |\alpha_1|$. Let $k \in \mathbb{Z}^+$ so that*

$$k - 1 < \frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty} \leq k. \tag{A.33}$$

An optimal protocol to estimate $q(\boldsymbol{\theta})$, where the parameters $\boldsymbol{\theta}$ are encoded into the probe state via unitary evolution under the Hamiltonian in Eq. (2.1), requires at least, but no more than, k-partite entanglement.

Proof. We divide our proof into two parts. First, using k-partite-entangled states from the set of catlike states considered in Chapter 2, we show the existence of an optimal protocol, subject to the upper bound of Eq. (A.33). Second, we show that there exists no optimal protocol using at most $(k-1)$-partite entanglement, proving the lower bound of Eq. (A.33).

Part 1. Define $T^{(k)}$ to be the submatrix of T with all columns n such that $\sum_m |T_{mn}| > k$ are eliminated, which enforces that any protocol derived from $T^{(k)}$ uses only states that are at most k-partite entangled. Define system $A(k)$ as

$$T^{(k)}\boldsymbol{p}^{(k)} = \boldsymbol{\alpha}/\alpha_1, \tag{A.34}$$

$$\boldsymbol{p}^{(k)} \geq 0. \tag{A.35}$$

Let $\boldsymbol{\alpha}' = \boldsymbol{\alpha}/\alpha_1$ and define system $B(k)$ as

$$(T^{(k)})^\top \boldsymbol{y} \geq 0, \tag{A.36}$$

$$\langle \boldsymbol{\alpha}', \boldsymbol{y} \rangle < 0. \tag{A.37}$$

By the Farkas-Minkowski lemma [77,78], system $A(k)$ has a solution if and only if system $B(k)$ does not. In particular, this lemma, which, geometrically, is an application of the hyperplane separation theorem [180], is as follows:

Lemma A.2 (Farkas-Minkowski) *Consider the system*

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{A.38}$$

$$\boldsymbol{x} \geq 0, \tag{A.39}$$

with $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{x} \in \mathbb{R}^{n}$, and $\boldsymbol{b} \in \mathbb{R}^{m}$. The above system has a solution if and only if there is no solution $\boldsymbol{y}$ to

$$A^{\top}\boldsymbol{y} \geq 0, \tag{A.40}$$

$$\langle \boldsymbol{b}, \boldsymbol{y} \rangle < 0. \tag{A.41}$$

Therefore, to prove the result it is sufficient to show that system $B(k)$ does not have a solution if $\sum_{j>1} |\alpha'_j| \leq k-1$, where we used that $\alpha'_1 = 1$. We assume that a solution $\boldsymbol{y}$ exists and will arrive at a contradiction. Without loss of generality, we assume that $|y_j| \geq |y_{j+1}|$ for all $1 < j < d$. Equation (A.37) implies $\sum_{j>1} \alpha'_j y_j < -y_1$. $(T^{(k)})^{\top}$ has a row n^* given by $\boldsymbol{\tau}^{(n^*)} = (1, 0, \dots, 0)$, so by Eq. (A.36) any solution $\boldsymbol{y}$ to system B has $y_1 \geq 0$. Therefore, $\left|\sum_{j>1} \alpha'_j y_j\right| > y_1$, which, by the triangle inequality, implies

$$\sum_{j>1} |\alpha'_j||y_j| > y_1. \tag{A.42}$$

Because $|\alpha'_j| \leq 1$ for all j, because $\sum_{j>1} |\alpha'_j| \leq k-1$, and because $|y_j|$ for $j > 1$ are ordered in descending order, the largest the left-hand side of Eq. (A.42) can be is $\sum_{j=2}^{k} |y_j|$, leading to

$$\sum_{j=2}^{k} |y_j| > y_1. \tag{A.43}$$

This directly contradicts Eq. (A.36) for the column of $T^{(k)}$ given by $\boldsymbol{\tau} = (1, -\mathrm{sgn}(y_2), \ldots, -\mathrm{sgn}(y_k), 0, 0, \ldots)$.

Part 2. Using Eq. (A.21), we have that, for any optimal protocol,

$$\mathcal{F}(\boldsymbol{\theta})_{1j} = t \int_0^t ds' \langle\psi(s')| \hat{\sigma}_1^z \hat{\sigma}_j^z |\psi(s')\rangle \,, \tag{A.44}$$

where we recall that $|\psi(s)\rangle = U(s)|\psi(0)\rangle$. Because $\langle\psi(s')| \hat{\sigma}_1^z |\psi(s')\rangle = 0$ for all s' [see Eq. (A.13)], the integrand is nonzero if and only if $|\psi(s')\rangle$ is such that the first qubit is entangled with the jth. Define the indicator variable

$$E_j(s') = \begin{cases} 1 & |\psi(s)\rangle \text{ entangles qubit } j \text{ and } 1 \\ 0 & \text{else,} \end{cases} \tag{A.45}$$

for all j, including any possible ancilla qubits. Here, we define $E_1 = 1$ even though the first qubit is not "entangled" with itself. Further define

$$E(s') = \sum_j E_j(s') \leq (k-1), \tag{A.46}$$

where $E(s')$ is the total number of sensor qubits entangled with the first qubit at time s' and the upper bound comes from our assumption on the partiteness of our probe states. We then have that

$$\mathcal{F}(\boldsymbol{\theta})_{1j} \leq t \int_0^t ds' E_j(s'). \tag{A.47}$$

Furthermore, for any optimal protocol using at most $(k-1)$-partite entanglement, we require that

$$\sum_j \left|\frac{\alpha_j}{\alpha_1} t^2\right| = \sum_j |\mathcal{F}(\boldsymbol{\theta})_{j1}| \leq t \sum_j \int_0^t ds' E_j(s') = t \int_0^t ds' \sum_j E_j(s) \leq t \int_0^t ds'(k-1) = (k-1)t^2. \tag{A.48}$$

We now have a contradiction, however, as the theorem statement assumed that

$$\sum_j \left|\frac{\alpha_j}{\alpha_1} t^2\right| = \frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty} t^2 > (k-1)t^2. \tag{A.49}$$

This concludes the proof that $(k-1)$-partite entanglement in any form (i.e., not just from catlike probe states) is insufficient to generate an optimal protocol. □

We also observe that the lower bound on the size of the least-entangled state used in an optimal protocol is really, at its core, a lower bound on the *average* entanglement required to saturate the conditions on the quantum Fisher information matrix. Here, average entanglement refers to weighting the size of the entangled state by the proportion of time it is used in the protocol. This lower bound is simply $\|\boldsymbol{\alpha}\|_1/\boldsymbol{\alpha}_\infty$. The lower bound on the size of the most-entangled state, or the bound on *instantaneous* entanglement, comes from ensuring that this lower bound on average entanglement is achievable (that is, if the instantaneous entanglement is too small at each stage, then the average entanglement required cannot be reached).

A.5 Minimum-Entanglement Non-Echoed Protocols

In this Appendix, we prove that there exist protocols that minimize both instantaneous and average entanglement. We recall from Section 2.6 the definition of the non-echoed protocols that minimize average entanglement.

Definition A.1 (Non-echoed protocols)**.** *Consider some $\boldsymbol{\alpha} \in \mathbb{R}^d$ encoding a linear function of interest. Let T be the matrix which describes our families of catlike probe states, and let $\boldsymbol{p}$ specify a valid protocol such that $\boldsymbol{p} \geq 0$ and $T\boldsymbol{p} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_\infty$. We say that the protocol defined by $\boldsymbol{p}$ is "non-echoed" if, $\forall i$ such that p_i is strictly greater than 0, $\operatorname{sgn}(T_{ij}) \in \{0, \operatorname{sgn}(\alpha_j)\}$.*

We now prove Theorem 2.2, which we again repeat for simplicity.

Theorem A.2. *For any function encoding $\boldsymbol{\alpha}$, there exists a non-echoed optimal protocol with minimum instantaneous entanglement.*

Proof. We proceed with a relatively simple tweak of the proof of the main theorem. As in that theorem, we assume without loss of generality that $\alpha_1 = \|\boldsymbol{\alpha}\|_\infty = 1$. Also assume, for computational simplicity, that $\alpha_{i>1} < 1$ (i.e., there is only a single maximal-magnitude element of $\boldsymbol{\alpha}$) and that $\alpha_i > 0\,\forall i$. These latter assumptions can easily be lifted, as we describe at the end of the proof.

We will again use the Farkas-Minkowski lemma [77, 78] to show that no vector $\boldsymbol{y}$ exists such that

$$(T_+^{(k)})^\top \boldsymbol{y} \geq 0, \tag{A.50}$$

$$\langle \boldsymbol{\alpha}, \boldsymbol{y} \rangle < 0, \tag{A.51}$$

proving the existence of a non-echoed protocol. Here, $T_+^{(k)}$ is T restricted to non-echoed vectors [i.e., $(T_+^{(k)})_{ij} \in \{0,1\}$] with weight at most k, where $k = \lceil \|\boldsymbol{\alpha}\|_1 \rceil$. Assume a solution $\boldsymbol{y}$ exists. Noting that $(T_+^{(k)})^\top$ has a row given by $(1, 0, \ldots, 0)$, it must be that $y_1 \geq 0$. Furthermore, for $\boldsymbol{y}$ to be a valid solution, we must have

$$\langle \boldsymbol{\alpha}, \boldsymbol{y} \rangle = \alpha_1 y_1 + \sum_{i | i \neq 1, y_i \geq 0} \alpha_i y_i + \sum_{i | y_i < 0} \alpha_i y_i = y_1 + \sum_{i | i \neq 1, y_i \geq 0} \alpha_i y_i + \sum_{i | y_i < 0} \alpha_i y_i \leq 0. \tag{A.52}$$

We proceed with two cases. Suppose that at most $k - 1$ elements of $\boldsymbol{y}$ are negative. Consider the row of $(T_+^{(k)})^\intercal$ that has a 1 in the first index and exactly on the indices where $y_i < 0$ (which exists because we have sufficiently restricted the number of negative elements of $\boldsymbol{y}$). Then $(T_+^{(k)})^\intercal \boldsymbol{y} \geq 0$ implies that

$$y_1 + \sum_{i | y_i \leq 0} y_i \geq 0. \tag{A.53}$$

But because $\alpha_i < 1$, this immediately implies that

$$y_1 + \sum_{i | y_i \leq 0} \alpha_i y_i \geq 0, \tag{A.54}$$

which means that Eq. (A.52) cannot be true, yielding a contradiction.

Now suppose that there are at least k elements of $\boldsymbol{y}$ that are negative. Let S be the set of indices corresponding to the $k - 1$ largest, in magnitude, y_i. Then the row of $(T_+^{(k)})^\intercal$ with a 1 in the first index and precisely on the indices in S leads to the condition that

$$y_1 + \sum_{i \in S} y_i \geq 0. \tag{A.55}$$

However, given the constraint that $\alpha_{i>1} < 1$, we find that

$$y_1 + \sum_{i | i \neq 1, y_i \geq 0} \alpha_i y_i + \sum_{i | y_i < 0} \alpha_i y_i \geq y_1 + \sum_{i \in S} y_i \geq 0, \tag{A.56}$$

which is again a contradiction.

We briefly comment on how to lift the two assumptions we mentioned earlier. First, in the case where there exist multiple maximal elements, the same argument that generalizes the main theorem will also generalize this argument—see Appendix A.6. Second, if we allow $\alpha_i < 0$, it is simple to see that a protocol still exists; simply replace $(T_+^{(k)})_{ij} = 1$ with $\operatorname{sgn}(\alpha_i)$ (and leave 0s untouched). □

Thus, Lemma 2.1 and Theorem 2.2 prove there exist protocols that can minimize both instantaneous entanglement (i.e., the maximum size of a catlike state used in the protocol) and the average entanglement over the course of the entire protocol.

A.6 Relaxing the Assumption on a Single Maximum Element

In this Appendix, we will generalize beyond the assumption in Chapter 2 that $|\alpha_1| > |\alpha_j|$ for all $j > 1$. Conceptually, nothing is changed by relaxing the assumption, but the algebra becomes somewhat more tedious. In the process, we rigorously derive Eq. (2.2) and Eq. (2.8).

A.6.1 Generalizing Eq. (2.8)

We start with specifically generalizing Eq. (2.8). To begin, define

$$L := \{i \,|\, |\alpha_i| = |\alpha_1|\}. \tag{A.57}$$

The assumption $|\alpha_1| > |\alpha_j|$ for all $j > 1$, stated in Chapter 2, is equivalent to assuming $|L| = 1$. For arbitrary size L, we have the following set of conditions for the single-parameter bound on $q(\boldsymbol{\theta})$

to be saturable [Eqs. (2.6) and (2.7)]:

$$\mathcal{F}(\boldsymbol{q})_{11} = \frac{t^2}{\alpha_1^2}, \tag{A.58}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = 0 \quad (\forall\, i \neq 1). \tag{A.59}$$

Recall that $\mathcal{F}(\boldsymbol{q}) = J^{\top}\mathcal{F}(\boldsymbol{\theta})J$, where J is the Jacobian for the basis transformation from $\boldsymbol{\theta}$ to $\boldsymbol{q}$, $q_1 = q$ is the linear function we wish to measure, and the other q_j are some other degrees of freedom we fix. We will show that Eqs. (A.58) and (A.59) are satisfied if and only if

$$\sum_{i\in L} \frac{\operatorname{sgn}(\alpha_1)}{\operatorname{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ji}\lambda_i = \frac{\alpha_j}{\alpha_1} t^2, \tag{A.60}$$

where $\lambda_i \geq 0$ such that $\sum_i \lambda_i = 1$. If $|L| = 1$, this reduces to Eq. (2.8).

It will be important to briefly recount how we obtain the single-parameter bound we are trying to saturate [14, 64]. In particular, referring to Eq. (2.3), we seek a choice of basis that minimizes $\|\hat{g}_q\|_s^2$, which will yield the tightest possible bound on $\mathcal{M}$, the mean-square error of q. Let us formally define our basis for $\mathbb{R}^d$ as $\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \ldots, \boldsymbol{\alpha}^{(d)}\}$, where $\boldsymbol{\alpha}^{(1)} = \boldsymbol{\alpha}$. We then have that J^{-1} has rows given by these vectors. Let $\{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(d)}\}$ be the basis dual to this one. That is, these vectors form the columns of J and satisfy $\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\beta}^{(j)} = \delta_{ij}$. We can then write

$$\boldsymbol{\theta}^{\top} = (JJ^{-1}\boldsymbol{\theta})^{\top} = (J^{-1}\boldsymbol{\theta})^{\top} J^{\top}, \tag{A.61}$$

which allows us to rewrite our Hamiltonian in the convenient form

$$\hat{H} = \frac{1}{2}\boldsymbol{\theta}^{\intercal}\hat{\boldsymbol{\sigma}} + \hat{H}_c(s) = \frac{1}{2}\sum_{i=1}^{d}(\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\theta})\boldsymbol{\beta}^{(i)} \cdot \hat{\boldsymbol{\sigma}} + \hat{H}_c(s), \tag{A.62}$$

where $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1^z, \dots, \hat{\sigma}_d^z)^{\intercal}$. Then

$$\hat{g}_q(0) = \frac{\partial \hat{H}}{\partial q} = \frac{\partial \hat{H}}{\partial(\boldsymbol{\alpha}^{(1)} \cdot \boldsymbol{\theta})} = \frac{\boldsymbol{\beta} \cdot \hat{\boldsymbol{\sigma}}}{2}, \tag{A.63}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$. Because the seminorm is time independent (see Ref. [51]), we immediately have that

$$\|\hat{g}_q\|_s = \|\boldsymbol{\beta}\|_1, \tag{A.64}$$

and our tightest bound is given by

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1, \\ &\text{such that } \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1. \end{aligned} \tag{A.65}$$

Note that

$$1 = \sum_i \alpha_i \beta_i \le \sum_i |\alpha_i||\beta_i| \le |\alpha_1| \sum_i |\beta_i| = |\alpha_1| \|\boldsymbol{\beta}\|_1. \tag{A.66}$$

The first inequality is tight if either $\mathrm{sgn}(\beta_i) = \mathrm{sgn}(\alpha_i)$ or $\beta_i = 0$ for all i. The second is slightly more complicated to saturate. Recall $L = \{i \,|\, |\alpha_i| = |\alpha_1|\}$. Then the second inequality is tight if and only if

$$\beta_i = 0 \text{ for } i \notin L, \tag{A.67}$$

$$\sum_{i \in L} |\beta_i| = \frac{1}{|\alpha_1|}. \tag{A.68}$$

Any solution $\boldsymbol{\beta}$ specifies the first column of the Jacobian J and allows us to rewrite the conditions in Eqs. (A.58) and (A.59) as

$$\mathcal{F}(\boldsymbol{q})_{11} = \boldsymbol{\beta}^\intercal \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = \frac{t^2}{\alpha_1^2}, \tag{A.69}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = (\boldsymbol{\beta}^{(i)})^\intercal \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = 0 \quad (\forall\, i \neq 1). \tag{A.70}$$

As $\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\beta}^{(j)} = \delta_{ij}$, Eq. (A.70) immediately implies that the vector $\mathcal{F}(\boldsymbol{\theta})\boldsymbol{\beta}$ must be proportional to $\boldsymbol{\alpha}$ and Eq. (A.69) specifies the constant of proportionality. In particular, we require

$$\mathcal{F}(\boldsymbol{\theta})\boldsymbol{\beta} = \frac{t^2}{\alpha_1^2}\boldsymbol{\alpha}. \tag{A.71}$$

Invoking Eqs. (A.67) and (A.68) and the condition that $\operatorname{sgn}(\beta_i) = \operatorname{sgn}(\alpha_i)$ for $\beta_i \neq 0$, we write $\beta_i = \lambda_i \operatorname{sgn}(\alpha_i)/|\alpha_1|$, where $\lambda_i \geq 0$ for $i \in L$ and $\lambda_i = 0$ for $i \notin L$ such that $\sum_i \lambda_i = 1$. The individual components of Eq. (A.71) imply

$$\sum_{i \in L} \mathcal{F}(\boldsymbol{\theta})_{ij} \operatorname{sgn}(\alpha_i) \lambda_i = \sum_{i \in L} \mathcal{F}(\boldsymbol{\theta})_{ji} \operatorname{sgn}(\alpha_i) \lambda_i = \frac{t^2}{|\alpha_1|}\alpha_j, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0, \tag{A.72}$$

which, using $|\alpha_1| = \mathrm{sgn}(\alpha_1)\alpha_1$ and that $\mathrm{sgn}(\alpha_1)\mathrm{sgn}(\alpha_i) = \mathrm{sgn}(\alpha_1)/\mathrm{sgn}(\alpha_i)$ for $i \in L$, yields

$$\sum_{i\in L} \frac{\mathrm{sgn}(\alpha_1)}{\mathrm{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ij}\lambda_i = \sum_{i\in L} \frac{\mathrm{sgn}(\alpha_1)}{\mathrm{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ji}\lambda_i = \frac{\alpha_j}{\alpha_1} t^2, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0, \tag{A.73}$$

which reduces to Eq. (2.8) of the main text, when $|L| = 1$, as desired.

A.6.2 Generalizing the Derivation of Eq. (2.13)

At this point, we can generalize the derivation of Eq. (2.13) to this setting of more than one maximum element of $\boldsymbol{\alpha}$. In particular, Lemma A.1 can be immediately extended to the following:

Lemma A.3 *Any optimal protocol, independent of the choice of control, requires that* $\langle \hat{\mathcal{H}}_j(t) \rangle = 0$ *for all* $j \in L$ *and that the probe state be of the form*

$$|\psi\rangle = \frac{\left(\bigotimes_{j\in L} |b_j\rangle\right)|\chi_0\rangle + e^{i\phi}\left(\bigotimes_{j\in L} |b_j + 1\rangle\right)|\chi_1\rangle}{\sqrt{2}}, \tag{A.74}$$

for all times $s \in [0, t]$, *where*

$$b_j = \begin{cases} 0 & \text{if } \mathrm{sgn}(\alpha_j) = 1 \\ 1 & \text{if } \mathrm{sgn}(\alpha_j) = -1, \end{cases} \tag{A.75}$$

and $\phi, |\chi_0\rangle, |\chi_1\rangle$ *can be arbitrary and* s *dependent. The addition inside the second ket of Eq. (A.74) is mod 2.*

Proof. We have the following two facts: (1) $\sum_{i\in L} \lambda_i(\mathrm{sgn}(\alpha_j)/\mathrm{sgn}(\alpha_i))\mathcal{F}(\boldsymbol{\theta})_{ij} = t^2$ for all $j \in L$ [by Eq. (A.73)] and (2) $|\mathcal{F}(\boldsymbol{\theta})_{ij}| \leq \mathcal{F}(\boldsymbol{\theta})_{jj}$ for all i (by the fact that the Fisher information matrix is positive semidefinite). These facts imply that an optimal protocol must have $\mathcal{F}(\boldsymbol{\theta})_{jj} = t^2$ for all $j \in L$. The fact that $\langle \hat{\mathcal{H}}_j(t) \rangle = 0$ for all $j \in L$ and the fact that all sensors in L must be in a catlike state over computational basis states follows immediately via an identical calculation to the proof of Lemma A.1 for each $j \in L$. From Eq. (A.21) it follows directly that these catlike states over the qubit sensors in L must take the form in the theorem statement in order to achieve the correct sign on the components of $\mathcal{F}(\boldsymbol{\theta})$. □

Using Lemma A.3, it is clear that we should restrict the set $\mathcal{T}$ of states such that $\tau_j^{(n)} = \mathrm{sgn}(\alpha_j)/\mathrm{sgn}(\alpha_1)$ for all $j \in L$ and all $\boldsymbol{\tau}^{(n)}$. This is the generalization of the fact that, when $|L| = 1$, we require $\tau_1^{(n)} = 1$ for all $\boldsymbol{\tau}^{(n)}$.

In addition, given the required form of the optimal states, it is easy to generalize Eq. (A.22) to the condition that

$$\sum_{i\in L}\left[\lambda_i \int_0^t ds' \langle\psi(s')|\,\hat{\sigma}_i^z\hat{\sigma}_j^z\,|\psi(s')\rangle\right] = \frac{\alpha_j}{\alpha_1}t, \tag{A.76}$$

which implies that, for protocols switching between states in the modified $\mathcal{T}$,

$$\sum_{i\in L}\left[\lambda_i \sum_{l=0}^{n}(t_{l+1}^* - t_l^*)\tau_j^{(l)}\right] = \frac{\alpha_j}{\alpha_1}t, \tag{A.77}$$

where we assume that we switch to the state labeled by $\boldsymbol{\tau}^{(l)}$ at time t_l^*. As before, in our protocols $t_{l+1}^* - t_l^* = p_l t$. In addition, $\sum_i \lambda_i = 1$. So an optimal protocol requires

$$t\sum_{l=0}^{n} p_l \tau_j^{(l)} = \frac{\alpha_j}{\alpha_1}t \qquad \Longrightarrow \qquad T\boldsymbol{p} = \boldsymbol{\alpha}, \tag{A.78}$$

recovering Eq. (2.13) for general L, with the addition that we fix $T_{jn} = \tau_j^{(n)} = \mathrm{sgn}(\alpha_j)/\mathrm{sgn}(\alpha_1)$ for all $j \in L$ and all n.

A.6.3 Generalizing the Proof of Theorem 2.1

Recall, we divided the proof into two parts. First, we showed the existence of an optimal protocol using k-partite-entangled catlike states, subject to the upper bound of the theorem statement. Second, we showed that, subject to the lower bound of the theorem statement, there exists no optimal protocol using only $(k-1)$-partite entanglement.

Let us begin by addressing how the first part changes upon relaxing the assumption that $|\alpha_1| > |\alpha_j|$ for all $j > 1$. Note that, given our choice that $\tau_j^{(n)} = \text{sgn}(\alpha_j)/\text{sgn}(\alpha_1)$ for all $j \in L$ and all $\boldsymbol{\tau}^{(n)}$, the first $|L|$ rows of $T^{(k)}$ yield redundant equations in Eq. (2.19). Therefore, we can define $\tilde{T}^{(k)}$ as $T^{(k)}$ with all rows $j \in L \setminus \{1\}$ eliminated. Similarly, $\tilde{\boldsymbol{\alpha}}$ is $\boldsymbol{\alpha}$ with elements $j \in L \setminus \{1\}$ eliminated. Furthermore, define the new system of equations, which we call system $\tilde{A}$:

$$\tilde{T}^{(k)}\tilde{p}^{(k)} = \tilde{\boldsymbol{\alpha}}/\alpha_1, \tag{A.79}$$

$$\tilde{\boldsymbol{p}}^{(k)} \geq 0. \tag{A.80}$$

System A has a solution if and only if system $\tilde{A}$ does. We can proceed as in the proof in Appendix A.4 to show via the Farkas-Minkowski lemma that system $\tilde{A}$ has a solution if $\|\boldsymbol{\alpha}\|_1/\|\boldsymbol{\alpha}\|_\infty \leq k \implies \|\tilde{\boldsymbol{\alpha}}\|_1/\|\tilde{\boldsymbol{\alpha}}\|_\infty \leq k - |L| + 1$. The details of the proof of this part are completely identical with this substitution.

The second part of the proof can similarly be adjusted straightforwardly. In particular, to satisfy the condition of Eq. (A.73), which is the generalization of Eq. (2.8) in the main text, for

$j \in L$ we require

$$\frac{\alpha_j}{\alpha_1} t^2 = \frac{\text{sgn}(\alpha_j)}{\text{sgn}(\alpha_1)} t^2 = \sum_{i \in L} \frac{\text{sgn}(\alpha_1)}{\text{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ij} \lambda_i, \tag{A.81}$$

which implies

$$t^2 = \sum_{i \in L} \frac{\text{sgn}(\alpha_i)}{\text{sgn}(\alpha_j)} \mathcal{F}(\boldsymbol{\theta})_{ij} \lambda_i. \tag{A.82}$$

This in turn implies that for $i, j \in L$

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = \frac{\text{sgn}(\alpha_i)}{\text{sgn}(\alpha_j)} t^2. \tag{A.83}$$

Therefore, for all $i \in L$ we require $\mathcal{F}(\boldsymbol{\theta})_{ii} = t^2$. From here, arguments identical to those in Appendix A.4 apply to all $i \in L$, not just $i = 1$. That is, all the probe states must always be fully entangled on the qubits in L and matrix elements $\mathcal{F}(\boldsymbol{\theta})_{ij}$ for $i \in L$, $j \notin L$ can only accumulate magnitude if sensor j is also entangled with the qubits in L. Assuming the existence of an optimal protocol using $(k-1)$-partite entanglement, a contradiction arises in an identical way.

Appendix B: Appendices Associated with Chapter 3

B.1 Bound for Local Phase Shifts

In this appendix, we derive lower bounds for the mean square error of measuring a linear function $q(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ of local phase shifts, generated via coupling to the number operator $\hat{n}_j$, as specified by the Hamiltonian in Eq. (3.1) and Eq. (3.2a).

In particular, we seek to solve the optimization problem in Eq. (3.9), restated here for convenience:

$$\min_{\boldsymbol{\beta}} \max_{\rho} [\Delta(\boldsymbol{\beta} \cdot \hat{\boldsymbol{g}})_{\rho}]^2, \quad \text{subject to } \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1. \tag{B.1}$$

Here, $\hat{\boldsymbol{g}} = \hat{\boldsymbol{n}} = (\hat{n}_1, \hat{n}_2, \cdots, \hat{n}_d)^T$. For fixed particle number N, the Hilbert space on which possible probe states ρ are defined is finite dimensional, and it holds that [51]

$$[\Delta(\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}})_{\rho}]^2 \leq \frac{\|\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}\|_{s,N}^2}{4}, \tag{B.2}$$

where $\|\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}\|_{s,N}$ is the Fock-space-restricted seminorm of $\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}$ (defined as the difference between the maximum and minimum eigenvalues of $\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}$ restricted to the N-photon subspace). As we want to maximize the quantum Fisher information with respect to the choice of probe state ρ, and because Eq. (B.2) is saturable when ρ is an equal superposition of the eigenstates of $\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}$ with maximum and minimum eigenvalues, we can consider the following optimization problem:

$$\begin{aligned} &\text{minimize (w.r.t. } \boldsymbol{\beta}\text{)} \ \|\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}\|_{s,N}, \\ &\text{subject to } \ \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1. \end{aligned} \tag{B.3}$$

To begin, note that the largest eigenvalue of $\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}$ in the N-particle subspace is given by

$$\lambda_{\max}(\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}) = N \max\left\{ \max_j \beta_j, 0 \right\} =: N\beta_{\max}, \tag{B.4}$$

where we have implicitly defined $\beta_{\max}$. This largest eigenvalue corresponds to the eigenstate that consists of placing all photons in the mode corresponding to the largest positive β_j. If all $\beta_j \leq 0$, the largest eigenvalue is zero, obtained by any state with no particles in the sensor modes. Note that this requires the use of an extra mode (an ancilla or so-called “reference mode”) to “store” these photons, as we fix the total photon number of our state to be N.

Similarly, the smallest eigenvalue of $\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}$ in the N-particle subspace is given by

$$\lambda_{\min}(\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}) = N \min\left\{ \min_j \beta_j, 0 \right\} =: N\beta_{\min}, \tag{B.5}$$

where we have implicitly defined $\beta_{\min}$.

Using the facts above about the maximum and minimum eigenvalues of $\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}$ in the N-

particle subspace we can rewrite the optimization problem in Eq. (3.9) as

$$\begin{aligned}&\text{minimize } N\left(\beta_{\max} - \beta_{\min}\right),\\ &\text{subject to } \boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 1.\end{aligned} \tag{B.6}$$

As in Chapter 3, define $\mathcal{P} := \{j \,|\, \alpha_j \geq 0\}$ and $\mathcal{N} := \{j \,|\, \alpha_j < 0\}$. We then have the following lemma.

Lemma B.1 *The solution $\boldsymbol{\beta}^*$ to Eq. (B.6) is such that $\beta_j^* \geq 0$ for all $j \in \mathcal{P}$, and $\beta_j^* \leq 0$ for all $j \in \mathcal{N}$. That is, $\alpha_j\beta_j^* \geq 0$ for all j.*

Proof. We proceed by contradiction. Let $\mathcal{J}_- = \{j \,|\, \alpha_j\beta_j^* < 0\}$ and $\mathcal{J}_+ = \{j \,|\, \alpha_j\beta_j^* \geq 0\}$. Suppose the solution vector $\boldsymbol{\beta}^*$ to Eq. (B.6) has $\mathcal{J}_- \neq \varnothing$. We can construct an alternative candidate solution vector $\boldsymbol{\beta}'$ as follows: First, let $\boldsymbol{\beta}' = \boldsymbol{\beta}^*$. Then set $\beta_j' = 0$ for all $j \in \mathcal{J}_-$. In order to still satisfy the constraint $\boldsymbol{\alpha}\cdot\boldsymbol{\beta}' = 1$, we must reduce the values of some other components in $\boldsymbol{\beta}'$. In particular, it is simple to calculate that a valid solution is, for $j \in \mathcal{J}_+$,

$$\beta_j' = \frac{\beta_j^*}{\sum_{j\in\mathcal{J}_+}\alpha_j\beta_j^*} = \frac{\beta_j^*}{1 - \sum_{j\in\mathcal{J}_-}\alpha_j\beta_j^*}. \tag{B.7}$$

Again, when $j \in \mathcal{J}_-$, $\beta_j' = 0$.

Let $\beta_{\max}' := \max\left\{\max_j \beta_j', 0\right\}$ and $\beta_{\min}' := \max\left\{\min_j \beta_j', 0\right\}$. By construction, $\beta_{\max}' \leq \beta_{\max}^*$ and $0 = \beta_{\min}' \geq \beta_{\min}^*$. Consequently, $\boldsymbol{\beta}'$ yields a smaller solution candidate than $\boldsymbol{\beta}^*$. This contradicts the fact that $\boldsymbol{\beta}^*$ is the optimal solution. The lemma statement follows as an immediate consequence. □

Lemma B.1 allows us to rewrite the minimization problem in Eq. (B.6) once again as

$$\begin{aligned}
&\text{minimize } N\left[\max_{j\in\mathcal{P}}\beta_j - \min_{j\in\mathcal{N}}\beta_j\right], \\
&\text{where } \beta_j \geq 0 \ \forall\, j \in \mathcal{P}, \\
&\beta_j \leq 0 \ \forall\, j \in \mathcal{N}, \\
&\text{subject to } \boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 1.
\end{aligned} \tag{B.8}$$

In the above, we define $\max_{j\in\mathcal{P}}\beta_j$ ($\min_{j\in\mathcal{N}}\beta_j$) to be zero if $\mathcal{P} = \varnothing$ ($\mathcal{N} = \varnothing$). A further simplification is enabled by another lemma.

Lemma B.2 *The solution vector $\boldsymbol{\beta}^*$ to Eq. (B.8) is such that $\beta_j^* = \beta_{\max}^*$ for all $j \in \mathcal{P}$ and $\beta_j^* = \beta_{\min}^*$ for all $j \in \mathcal{N}$.*

Proof. We proceed by contradiction. Suppose the solution vector $\boldsymbol{\beta}^*$ is such that $\beta_i^* \neq \beta_j^*$ for some $i, j \in \mathcal{P}$. Then we could consider an alternative candidate solution vector $\boldsymbol{\beta}'$ where $\beta_k' = \frac{\sum_{l\in\mathcal{P}}\alpha_l\beta_l^*}{\sum_{l\in\mathcal{P}}\alpha_l}$ for all $k \in \mathcal{P}$. Similarly, if $\beta_i^* \neq \beta_j^*$ for some $i, j \in \mathcal{N}$ we could consider $\beta_k' = \frac{\sum_{l\in\mathcal{N}}\alpha_l\beta_l^*}{\sum_{l\in\mathcal{N}}\alpha_l}$ for all $k \in \mathcal{N}$. Clearly, $\boldsymbol{\beta}'$ still satisfies the constraint

$$\boldsymbol{\alpha}\cdot\boldsymbol{\beta}' = \sum_{m\in\mathcal{P}}\alpha_m\left(\frac{\sum_{l\in\mathcal{P}}\alpha_l\beta_l^*}{\sum_{l\in\mathcal{P}}\alpha_l}\right) + \sum_{m\in\mathcal{N}}\alpha_m\left(\frac{\sum_{l\in\mathcal{N}}\alpha_l\beta_l^*}{\sum_{l\in\mathcal{N}}\alpha_l}\right) = \boldsymbol{\alpha}\cdot\boldsymbol{\beta}^* = 1. \tag{B.9}$$

Additionally, β' also clearly still has $\beta_j' \geq 0$ when $j \in \mathcal{P}$ and $\beta_j' \leq 0$ when $j \in \mathcal{N}$. But, by construction (because the weighted average of a set is less than its maximum element),

$$N\left[\max_{j\in\mathcal{P}}\beta_j' - \min_{j\in\mathcal{N}}\beta_j'\right] < N\left[\max_{j\in\mathcal{P}}\beta_j^* - \min_{j\in\mathcal{N}}\beta_j^*\right]. \tag{B.10}$$

So $\boldsymbol{\beta}^*$ is not the solution vector and we have arrived at a contradiction. □

As a direct consequence of Lemma B.2 we can rewrite the optimization problem in Eq. (B.8) one last time as

$$\text{minimize (w.r.t. } \beta_{\min}, \beta_{\max}) \; N\left[\beta_{\max} - \beta_{\min}\right], \tag{B.11}$$

$$\text{subject to } \; \beta_{\max} \geq 0, \beta_{\min} \leq 0, \tag{B.12}$$

$$\beta_{\max} \sum_{j \in \mathcal{P}} \alpha_j + \beta_{\min} \sum_{j \in \mathcal{N}} \alpha_j = 1. \tag{B.13}$$

Because this is a linear objective function, the optimal solution will be one of the two boundary solutions: $\beta_{\max} = \frac{1}{\sum_{i \in \mathcal{P}} \alpha_i}, \beta_{\min} = 0$ or $\beta_{\min} = \frac{1}{\sum_{i \in \mathcal{N}} \alpha_i}, \beta_{\max} = 0$. Minimizing over these two candidate solutions, we obtain the final result

$$\left\|\hat{g}_q\right\|_{s,N}^2 = \frac{N^2}{\max(\sum_{i \in \mathcal{P}} \alpha_i, \sum_{i \in \mathcal{N}} \alpha_i)^2}. \tag{B.14}$$

Consequently, via the quantum Cramér-Rao bound, Eq. (3.10),

$$\begin{aligned} \mathcal{M} &\geq \frac{\max\left\{\sum_{i \in \mathcal{P}} \alpha_i, \sum_{i \in \mathcal{N}} \alpha_i\right\}^2}{N^2 t^2} \\ &=: \frac{\max\left\{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}^2, \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}^2\right\}}{N^2 t^2}, \end{aligned} \tag{B.15}$$

which is Eq. (3.12), and where $||\boldsymbol{\alpha}||_{1,\mathcal{P}}$ and $||\boldsymbol{\alpha}||_{1,\mathcal{N}}$ are the one-norm restricted to positive and negative values, respectively, of $\boldsymbol{\alpha}$. In the special case of all positive coefficients (i.e., $\mathcal{N} = \varnothing$), this reduces to

$$\mathcal{M} \geq \frac{\|\boldsymbol{\alpha}\|_1^2}{N^2 t^2}, \tag{B.16}$$

which, as described in Chapter 3, proves a conjecture from Ref. [13] that this is the minimum attainable variance for $\boldsymbol{\alpha} \in \mathbb{Q}^d$ with $\boldsymbol{\alpha} \geq 0$.

B.2 Bound for Local Displacements

In this Appendix, we derive Eq. (3.15) for the mean square error attainable for measuring a linear function of local displacements, restricting to probe states with fixed average photon number $\overline{N}$.

B.2.1 Separable Bound

To begin, it is helpful to present the bound for the more restricted case where we use separable input states. Begin by considering the lower bound on the variance of measuring a displacement φ coupled to a single mode via $H = \varphi \hat{p}$, following the proof sketched in Ref. [25]. The quantum Fisher information is given by

$$\mathcal{F}(\varphi) = 4[\Delta(\hat{p})_\rho]^2, \tag{B.17}$$

where ρ is the probe state, which is restricted to have an average photon number $\overline{N}$. An initial displacement does not enhance precision [25], so we can consider zero-mean displacement input states. For such probe states,

$$(\Delta \hat{p})^2 = -\frac{1}{4}\langle(\hat{a}^\dagger - \hat{a})^2\rangle = -\frac{1}{4}(\langle\hat{a}^\dagger\hat{a}^\dagger\rangle - \langle\hat{a}^\dagger\hat{a}\rangle - \langle\hat{a}\hat{a}^\dagger\rangle + \langle\hat{a}\hat{a}\rangle), \tag{B.18}$$

$$(\Delta \hat{x})^2 = \frac{1}{4}\langle(\hat{a}^\dagger + \hat{a})^2\rangle = \frac{1}{4}(\langle\hat{a}^\dagger\hat{a}^\dagger\rangle + \langle\hat{a}^\dagger\hat{a}\rangle + \langle\hat{a}\hat{a}^\dagger\rangle + \langle\hat{a}\hat{a}\rangle), \tag{B.19}$$

so that

$$\overline{N} = \langle \hat{a}^\dagger \hat{a} \rangle = (\Delta \hat{p})^2 + (\Delta \hat{x})^2 - \frac{1}{2}, \tag{B.20}$$

where we used that $\hat{a}\hat{a}^\dagger = \hat{a}^\dagger \hat{a} + 1$. We can then use the uncertainty principle

$$(\Delta \hat{p})^2 (\Delta \hat{x})^2 \geq \frac{1}{16}, \tag{B.21}$$

which follows from our definition of the quadrature operators as $\hat{x} = (\hat{a}^\dagger + \hat{a})/2$ and $\hat{p} = i(\hat{a}^\dagger - \hat{a})/2$. Therefore,

$$\xi \left(\overline{N} - \xi + \frac{1}{2} \right) \geq \frac{1}{16}, \tag{B.22}$$

where we let $\xi := (\Delta \hat{p})^2$. Then

$$-16\xi^2 + (16\overline{N} + 8)\xi - 1 \geq 0. \tag{B.23}$$

To maximize ξ, this inequality must be saturated, so we can solve the corresponding quadratic to obtain the solution

$$\xi = \frac{-8(2\overline{N} + 1) + \sqrt{64(2\overline{N} + 1)^2 - 64}}{-32} \implies 4\xi = (\sqrt{\overline{N}} + \sqrt{\overline{N} + 1})^2 \sim 4\overline{N}. \tag{B.24}$$

It is worth noting that the $\mathcal{O}(\overline{N})$ asymptotic behavior of the maximum variance of $\hat{p}$ could have been obtained with no calculation just from examining the constraint in Eq. (B.20) under the assumption that $(\Delta\hat{x})^2$ can be made negligibly small.

Putting everything back together, we have found that, optimizing over states with fixed average photon number $\overline{N}$, the following holds:

$$[\Delta(\tilde{\varphi})]^2 \geq \frac{1}{\mathcal{F}} \geq \frac{1}{t^2(\sqrt{\overline{N}} + \sqrt{\overline{N}+1})^2} = \frac{1}{4t^2\overline{N}} + \mathcal{O}\left(\frac{1}{t^2\overline{N}^2}\right). \tag{B.25}$$

Working in the asymptotic in $\overline{N}$ limit, we can use Eq. (B.25) to obtain a bound on performance for estimating a linear function $q(\boldsymbol{\theta}) = \boldsymbol{\alpha}\cdot\boldsymbol{\theta}$ with an unentangled protocol as

$$(\Delta\tilde{q})^2 \geq \frac{1}{t^2}\min_{\{\overline{N}_j\}}\sum_{j=1}^{d}\frac{|\alpha_j|^2}{4\overline{N}_j} + \mathcal{O}\left(\frac{1}{\overline{N}_j^2}\right), \tag{B.26}$$

where $\overline{N}_j = \langle\hat{a}_j^\dagger\hat{a}_j\rangle$ is the average number of photons used in mode j and $\sum_j \overline{N}_j = \overline{N}$. Assume without loss of generality that $|\alpha_j| > 0$ for all j (that is, no $\alpha_j = 0$) and independent of $\overline{N}$. Then we can optimize (at leading order in $\frac{1}{\overline{N}}$) the distribution of photons amongst the modes using the Lagrangian

$$\mathcal{L} = \sum_{j=1}^{d}\frac{|\alpha_j|^2}{4\overline{N}_j} + \gamma\left(\sum_{j=1}^{d}\overline{N}_j - \overline{N}\right), \tag{B.27}$$

where γ is a Lagrange multiplier. A bit of algebra yields that

$$\frac{\partial\mathcal{L}}{\partial\overline{N}_j} = 0 \implies \overline{N}_j = \frac{|\alpha_j|}{2\sqrt{\gamma}}. \tag{B.28}$$

This further implies that

$$\overline{N} = \sum_{j=1}^{d}\overline{N}_j = \frac{\|\boldsymbol{\alpha}\|_1}{2\sqrt{\gamma}}, \tag{B.29}$$

allowing us to obtain the optimal division of photons as

$$\overline{N}_j = \frac{|\alpha_j|}{\|\boldsymbol{\alpha}\|_1}\overline{N}. \tag{B.30}$$

We note that this solution is clearly the desired minimum of the Lagrangian, as maximizing the objective would lead to setting any $\overline{N}_j$ to 0. Plugging this back into Eq. (B.26) we obtain the (asymptotic in $\overline{N}$) separable bound

$$[\Delta\tilde{q}]^2 \geq \frac{\|\boldsymbol{\alpha}\|_1^2}{4\overline{N}t^2} + \mathcal{O}\left(\frac{1}{\overline{N}^2}\right). \tag{B.31}$$

This bound can be achieved by using the single-mode protocols in Ref. [25] for each mode and then computing the function of interest classically as a linear combination of the individual estimators.

B.2.2 General Function Estimation Bound

In this subsection, we turn to our primary task: deriving Eq. (3.15) for the mean square error attainable for measuring a linear function of local displacements, restricting to probe states with fixed average photon number $\overline{N}$.

To derive this bound, we must solve the optimization problem in Eq. (3.9) for $\hat{g}_j = \hat{p}_j$:

$$\min_{\boldsymbol{\beta}} \max_{\rho} [\Delta(\boldsymbol{\beta}\cdot\hat{\boldsymbol{p}})_\rho]^2, \quad \text{subject to } \boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 1. \tag{B.32}$$

We can write

$$\begin{aligned}
[\Delta(\boldsymbol{\beta}\cdot\hat{\boldsymbol{p}})]^2 &= \sum_{i,j=1}^{d} \beta_i\beta_j \mathrm{Cov}(\hat{p}_i,\hat{p}_j) \\
&\leq \sum_{i,j=1}^{d} \beta_i\beta_j \sqrt{(\Delta\hat{p}_i)^2(\Delta\hat{p}_j)^2} \\
&= \left[\sum_{j=1}^{d} \beta_j \Delta\hat{p}_j\right]^2 \\
&\leq \|\boldsymbol{\beta}\|_2^2 \sum_{j=1}^{d} (\Delta\hat{p}_j)^2,
\end{aligned} \tag{B.33}$$

where we applied the Cauchy-Schwarz inequality twice. Using the same assumption of zero-displacement states we made in the previous section, we can further bound $\sum_j(\Delta\hat{p}_j)^2$ using the constraint on average photon number

$$\sum_{j=1}^{d}\left[(\Delta\hat{p}_j)^2 + (\Delta\hat{x}_j)^2\right] - \frac{d}{2} = \sum_{j=1}^{d}\langle a_j^\dagger a_j\rangle = \overline{N}, \tag{B.34}$$

implying that

$$\sum_{j=1}^{d}(\Delta\hat{p}_j)^2 \leq \overline{N} + \frac{d}{2}. \tag{B.35}$$

Equation (B.35) is tight when $(\Delta\hat{x}_j)^2 = 0$ for all j. This is, of course, impossible to achieve, but can be approached asymptotically with increasing $\overline{N}$ ($\overline{N} \gg d$). Furthermore, using the fact that $\boldsymbol{\alpha}$ is dual to $\boldsymbol{\beta}$ and the Cauchy-Schwarz inequality, it holds that

$$1 = \boldsymbol{\alpha}\cdot\boldsymbol{\beta} \leq \|\boldsymbol{\beta}\|_2\|\boldsymbol{\alpha}\|_2. \tag{B.36}$$

As we want to minimize with respect to $\boldsymbol{\beta}$, we consider the case where this inequality is saturated (i.e. $\boldsymbol{\beta}^* = \frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_2^2}$). Therefore, $\|\boldsymbol{\beta}^*\|_2 = \frac{1}{\|\boldsymbol{\alpha}\|_2}$, and we obtain

$$[\Delta(\boldsymbol{\beta}\cdot\hat{\boldsymbol{p}})]^2 \leq \frac{\overline{N}}{\|\boldsymbol{\alpha}\|_2^2} + \mathcal{O}\left(\frac{d}{\|\boldsymbol{\alpha}\|_2^2}\right). \tag{B.37}$$

This yields the final bound

$$\mathcal{M} \geq \frac{\|\boldsymbol{\alpha}\|_2^2}{4\overline{N}t^2} - \mathcal{O}\left(\frac{d\|\boldsymbol{\alpha}\|_2^2}{\overline{N}^2 t^2}\right). \tag{B.38}$$

From the derivation alone, it is not obvious that this bound can be saturated, but the existence of protocols that achieve it [84] indicate that this bound is, indeed, tight asymptotically in $\overline{N}$.

B.3 Quantum Fisher Information Matrix Elements

In this Appendix, we derive the matrix elements of the quantum Fisher information matrix for generators $\hat{n}_j$ and $\hat{p}_j$ under the unitary evolution Eq. (3.4). For number operator coupling $\hat{g}_j = \hat{n}_j$,

$$\begin{aligned}\mathcal{H}_j = -iU^\dagger \partial_j U &= -\sum_{m=1}^{M}\left(\prod_{l=1}^{m-1} U^{(l)}V\right)^\dagger \hat{n}_j \left(\prod_{l=1}^{m-1} U^{(l)}V\right) \\ &=: -\sum_{m=1}^{M} \hat{n}_j(m),\end{aligned} \tag{B.39}$$

where in the second line we implicitly defined $\hat{n}_j(m)$. Consequently, we can compute the quantum Fisher information matrix elements via Eq. (3.23) to be

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4\left[\sum_{l=1}^{M}\sum_{m=1}^{M}\frac{1}{2}\langle\{\hat{n}_i(l),\hat{n}_j(m)\}\rangle - \left(\sum_{m=1}^{M}\langle\hat{n}_i(m)\rangle\right)\left(\sum_{m=1}^{M}\langle\hat{n}_j(m)\rangle\right)\right]. \tag{B.40}$$

When $\hat{U}^{(j)} = I$ for all j, this reduces to

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4M^2 \left[\langle \hat{n}_i \hat{n}_j \rangle - \langle \hat{n}_i \rangle \langle \hat{n}_j \rangle \right]. \tag{B.41}$$

For quadrature operator coupling $\hat{g}_j = \hat{p}_j$, essentially identical manipulations yield

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = 4 \left[\sum_{l=1}^{M} \sum_{m=1}^{M} \frac{1}{2} \langle \{ \hat{p}_i(l), \hat{p}_j(m) \} \rangle - \left(\sum_{m=1}^{M} \langle \hat{p}_i(m) \rangle \right) \left(\sum_{m=1}^{M} \langle \hat{p}_j(m) \rangle \right) \right], \tag{B.42}$$

where $\hat{p}_j(l)$ is defined as in Eq. (B.39) with $\hat{n}_j \to \hat{p}_j$.

B.4 Protocols for Local Phase Shifts

In this Appendix, we elaborate on the families of optimal protocols for measuring a linear function of phase shifts that we described in Section 3.4.

B.4.1 An Optimal Protocol for Functions with Positive Coefficients

We begin by reviewing a protocol from Ref. [13] for the special case of a linear function with positive coefficients (i.e., $\boldsymbol{\alpha} \geq 0$). Our results in Appendix B.1 show that, as those authors conjectured, this protocol is optimal. In particular, consider using as the probe state a so-called proportionally weighted N00N state over $d+1$ modes:

$$|\psi\rangle \propto \left| N \frac{\alpha_1}{\|\boldsymbol{\alpha}\|_1}, \cdots, N \frac{\alpha_d}{\|\boldsymbol{\alpha}\|_1}, 0 \right\rangle + \left| 0, \cdots, 0, N \right\rangle, \tag{B.43}$$

where we have expressed the state in an occupation number basis over $d+1$ modes and have dropped the normalization for concision. The last mode serves as a reference mode. Observe that, for this state to be well defined, it is essential that $\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_1} \in \mathbb{Q}^d$ and that N is such that the resulting occupation numbers are integers, which may require that N be large.

Following imprinting of the parameters $\boldsymbol{\theta}$ onto the probe state via M passes through the interferometers, one obtains

$$|\psi_M\rangle = e^{-iM\hat{\boldsymbol{n}}\cdot\boldsymbol{\theta}}|\psi\rangle \propto \left|N\frac{\alpha_1}{\|\boldsymbol{\alpha}\|_1}, \cdots, N\frac{\alpha_d}{\|\boldsymbol{\alpha}\|_1}, 0\right\rangle + e^{i\boldsymbol{\alpha}\cdot\boldsymbol{\theta}\frac{NM}{\|\boldsymbol{\alpha}\|_1}}\left|0, \cdots, 0, N\right\rangle. \tag{B.44}$$

This process allows us to saturate the bound in Eq. (3.14). In particular, using Eq. (B.40) [which reduces to Eq. (B.41) because there is no control required], it is straightforward to calculate that the quantum Fisher information matrix for the probe state is

$$\mathcal{F}(\boldsymbol{\theta}) = \frac{(MN)^2}{\|\boldsymbol{\alpha}\|_1^2}\boldsymbol{\alpha}\boldsymbol{\alpha}^T, \tag{B.45}$$

which clearly satisfies the condition in Eq. (3.24) (recalling that $||\boldsymbol{\alpha}||_1 = ||\boldsymbol{\alpha}||_{1,\mathcal{P}}$ here because we have assumed all coefficients are non-negative, and also recalling that $\Delta t = 1$ such that $M = t$).

While the conditions on the quantum Fisher information matrix for an optimal protocol are met, a full protocol requires a description of the measurements used to extract the quantity of interest from the relative phase between the branches of $|\psi_M\rangle$. As described in Chapter 3, this can be done via the robust phase estimation protocols of Refs. [71–73] with a small multiplicative constant overhead relative to the quantum Cramér-Rao bound (we also briefly discuss the idea behind robust phase estimation in Appendix B.7). The details of implementing the necessary parity measurements for N00N-like states are discussed in detail in Appendix A of Ref. [73] and Ref. [181].

B.4.2 Extending the Optimal Protocol to Negative Coefficients

While not explicitly considered in Ref. [13], it is straightforward to extend the above protocol to the situation where $\mathcal{N} \neq \varnothing$, which we do here. Without loss of generality, assume the coefficients are ordered so that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_d$. Using our standard assumption that $\|\boldsymbol{\alpha}\|_{1,\mathcal{P}} \geq \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}$, we claim that the following probe state is optimal:

$$|\psi\rangle \propto \bigotimes_{j\in\mathcal{P}} \left|N\frac{\alpha_j}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\right\rangle |0\rangle^{\otimes|\mathcal{N}|} |0\rangle + |0\rangle^{\otimes|\mathcal{P}|} \bigotimes_{j\in\mathcal{N}} \left|N\frac{|\alpha_j|}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\right\rangle \left|N - N\frac{\|\boldsymbol{\alpha}\|_{1,\mathcal{N}}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\right\rangle, \tag{B.46}$$

where, again, the last mode is a reference mode, and we have dropped the normalization of the state. Interestingly, observe that, if $\|\boldsymbol{\alpha}\|_{1,\mathcal{P}} = \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}$, the reference mode factors out and is unnecessary. Similar to the $\boldsymbol{\alpha} \geq 0$ case, for this state to be well defined, we require that $N|\alpha_j|/\|\boldsymbol{\alpha}\|_{1,\mathcal{P}} \in \mathbb{N}$ for all j, which is always true for some sufficiently large N provided $\boldsymbol{\alpha} \in \mathbb{Q}^d$.

Consider applying the encoding unitary for M passes through the interferometers. For $\|\boldsymbol{\alpha}\|_{1,\mathcal{P}} \geq \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}$, this yields

$$|\psi_M\rangle \propto \bigotimes_{j\in\mathcal{P}} \left|N\frac{\alpha_j}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\right\rangle |0\rangle^{\otimes|\mathcal{N}|} |0\rangle + e^{i\boldsymbol{\alpha}\cdot\boldsymbol{\theta}\frac{NM}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}} |0\rangle^{\otimes|\mathcal{P}|} \bigotimes_{j\in\mathcal{N}} \left|N\frac{|\alpha_j|}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\right\rangle \left|N - N\frac{\|\boldsymbol{\alpha}\|_{1,\mathcal{N}}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\right\rangle. \tag{B.47}$$

This probe state is optimal in the sense of satisfying the Fisher information condition in Eq. (3.24). In Chapter 3, we described an even more general family of protocols. Within this more general framework, we will prove this optimality.

B.4.3 A Family of Optimal Protocols

Finally, we describe a family of optimal protocols that satisfy the conditions on the quantum Fisher information matrix given in Eq. (3.24). In Chapter 3, we defined a family of optimal protocols in terms of vectors from the set

$$\mathcal{W} := \left\{ \boldsymbol{\omega} \in \mathbb{Z}^d \,\middle|\, \|\boldsymbol{\omega}\|_{1,\mathcal{P}} = N,\ \|\boldsymbol{\omega}\|_{1,\mathcal{N}} \leq N,\ \omega_j \alpha_j \geq 0\ \forall\, j \right\}. \tag{B.48}$$

In particular, from these vectors, we defined a set $\mathcal{T}$ of one-parameter families of probe states in an occupation number basis where each $|\psi(\boldsymbol{\omega};\varphi)\rangle \in \mathcal{T}$ is labeled by a particular choice of $\boldsymbol{\omega}$ such that

$$|\psi(\boldsymbol{\omega};\varphi)\rangle \propto |\boldsymbol{\omega}|_{\mathcal{P}}\rangle\,|0\rangle + e^{i\varphi}\,|-\boldsymbol{\omega}|_{\mathcal{N}}\rangle\,|N - \|\boldsymbol{\omega}|_{\mathcal{N}}\|_1\rangle\,, \tag{B.49}$$

where $\varphi \in \mathbb{R}$ is an arbitrary parameter and the last mode is a reference mode. Recall also that $\boldsymbol{\omega}_{\mathcal{P}}$ and $\boldsymbol{\omega}_{\mathcal{N}}$ are defined in Eq. (3.28) as the restriction of $\boldsymbol{\omega}$ to $j \in \mathcal{P}$ and $\mathcal{N}$, respectively (for j not in the correct set, the value is set to 0). Note that such states are of the form of those in Lemma 3.1. We claimed that, by explicitly computing the Fisher information matrix for these states, one could demonstrate that the optimality condition in Eq. (3.24) is satisfied for a protocol such that

$$W\boldsymbol{r} = NM\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}, \tag{B.50}$$

where $\boldsymbol{r} \in \mathbb{Z}^{|\mathcal{T}|}$ is as defined in Chapter 3 and must obey the conditions

$$\begin{aligned} \|\boldsymbol{r}\|_1 &= M, \\ \boldsymbol{r} &\geq 0. \end{aligned} \tag{B.51}$$

Recall that W is a matrix whose columns are the vectors $\boldsymbol{\omega}_n \in \mathcal{W}$.

Here we explicitly demonstrate this. We can easily evaluate

$$\langle \hat{n}_j(m) \rangle = \left\langle \psi(\boldsymbol{\omega}^{(m)};\varphi) \right| \hat{n}_j \left| \psi(\boldsymbol{\omega}^{(m)};\varphi) \right\rangle = \frac{|\omega_j^{(m)}|}{2} \tag{B.52}$$

and

$$\begin{aligned} \langle \hat{n}_i(l)\hat{n}_j(m) \rangle &= \left\langle \psi(\boldsymbol{\omega}^{(l)};\varphi) \right| \hat{n}_i U(m \leftrightarrow l) \hat{n}_j \left| \psi(\boldsymbol{\omega}^{(m)};\varphi) \right\rangle \\ &= \frac{|\omega_i^{(l)} \omega_j^{(m)}|}{2} \left\langle \psi_l(\boldsymbol{\omega}^{(l)};\varphi) \right| U(m \leftrightarrow l) \left| \psi_m(\boldsymbol{\omega}^{(m)};\varphi) \right\rangle, \end{aligned} \tag{B.53}$$

where $\hat{n}_j(m)$ are defined as in Eq. (B.39), and

$$U(m \leftrightarrow l) = \begin{cases} \prod_{k=m}^{l-1} U^{(k)} V, & \text{if } l \geq m \\ \prod_{k=l}^{m-1} (U^{(k)} V)^\dagger, & \text{otherwise,} \end{cases} \tag{B.54}$$

i.e., it is the unitary that converts between the m-th and l-th probe states. Additionally, $\boldsymbol{\omega}^{(m)}$ refers to the vector associated to the m-th probe state; correspondingly $|\psi_l(\boldsymbol{\omega}^{(l)};\varphi)\rangle$ is the branch of $|\psi(\boldsymbol{\omega}^{(l)};\varphi)\rangle$ with non-zero occupation number on mode l and $|\psi_m(\boldsymbol{\omega}^{(m)};\varphi)\rangle$ is the branch of $|\psi(\boldsymbol{\omega}^{(m)};\varphi)\rangle$ with non-zero occupation number on mode m. For an optimal protocol, $U(m \leftrightarrow l)$

coherently maps the first (second) branch of $|\psi(\boldsymbol{\omega}^{(l)};\varphi)\rangle$ to the first (second) branch of $|\psi(\boldsymbol{\omega}^{(m)};\varphi)\rangle$; therefore, we have that the matrix element $\langle\psi_l(\boldsymbol{\omega}^{(l)};\varphi)|\, U(m \leftrightarrow l)\, |\psi_m(\boldsymbol{\omega}^{(m)};\varphi)\rangle$ is nonzero if and only if the branches with non-zero occupation on modes l and m are the same. So we have that

$$\langle \hat{n}_i(l)\hat{n}_j(m)\rangle = \frac{|\omega_i^{(l)}\omega_j^{(m)}|}{2}\xi_{ij}, \tag{B.55}$$

where

$$\xi_{ij} := \begin{cases} 1, & \text{if } i,j \in \mathcal{P} \text{ or } i,j \in \mathcal{N} \\ 0, & \text{otherwise.} \end{cases} \tag{B.56}$$

Putting everything together we obtain that

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = (-1)^{\xi_{ij}+1}\left(\sum_{m=1}^{M}|\omega_i^{(m)}|\right)\left(\sum_{m=1}^{M}|\omega_j^{(m)}|\right). \tag{B.57}$$

To prove the protocols work, we need to show that this Fisher information matrix obeys the condition in Eq. (3.24). Without loss of generality, consider the case that $\|\boldsymbol{\alpha}\|_{1,\mathcal{P}} \geq \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}$. We have that

$$\sum_{j\in\mathcal{P}}\mathcal{F}(\boldsymbol{\theta})_{ij} = \operatorname{sgn}(\alpha_i)\left(\sum_{m=1}^{M}|\omega_i^{(m)}|\right)MN, \tag{B.58}$$

where we used that $\|\boldsymbol{\omega}\|_{1,\mathcal{P}} = N$. So, to obey the condition in Eq. (3.24), we require that

$$\sum_{m=1}^{M}|\omega_i^{(m)}| = MN\frac{|\alpha_i|}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}. \tag{B.59}$$

Or, in vector form:

$$\sum_{m=1}^{M} |\boldsymbol{\omega}^{(m)}| = MN \frac{|\boldsymbol{\alpha}|}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}. \tag{B.60}$$

Protocols in our family satisfy this condition by construction as, for any valid protocol,

$$\sum_{m=1}^{M} |\boldsymbol{\omega}^{(m)}| = |W|\boldsymbol{r}, \tag{B.61}$$

where $|W|$ denotes taking the element-wise absolute value of the elements of W. Consequently, noting that $\operatorname{sgn}(\omega_j^{(m)}) = \operatorname{sgn}(\alpha_j)$ for all m, we require

$$W\boldsymbol{r} = MN \frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}, \tag{B.62}$$

which is Eq. (B.50).

B.5 Proof of Lemma 3.1

Here we provide a proof of Lemma 3.1 in Chapter 3, restated here for convenience.

Lemma B.3 *Any optimal protocol using N photons and M passes through interferometers with a coupling as in Eq. (3.1) with $\hat{g}_j = \hat{n}_j$ requires that, for every pass m, the probe state $|\psi_m\rangle$ be of the form*

$$|\psi_m\rangle \propto |\boldsymbol{N}(m)\rangle_{\mathcal{P}} |\boldsymbol{0}\rangle_{\mathcal{NR}} + e^{i\varphi_m} |\boldsymbol{0}\rangle_{\mathcal{P}} |\boldsymbol{N}'(m)\rangle_{\mathcal{NR}}, \tag{B.63}$$

where $\mathcal{P}$, $\mathcal{N}$, and $\mathcal{R}$ represent the modes with $\alpha_j \geq 0$, $\alpha_j < 0$, and the (arbitrary number of) reference modes, respectively, $\boldsymbol{N}(m)$ and $\boldsymbol{N}'(m)$ are strings of occupation numbers such that $|\boldsymbol{N}(m)| = |\boldsymbol{N}'(m)| = N$ for all passes m. φ_m is an arbitrary phase.

Proof. The quantum Fisher information matrix elements for any protocol with $\hat{g}_j = \hat{n}_j$ are given by

$$\begin{aligned}\mathcal{F}(\boldsymbol{\theta})_{ij} &= 4\left[\sum_{l=1}^{M}\sum_{m=1}^{M}\frac{1}{2}\langle\{\hat{n}_i(l),\hat{n}_j(m)\}\rangle - \left(\sum_{m=1}^{M}\langle\hat{n}_i(m)\rangle\right)\left(\sum_{m=1}^{M}\langle\hat{n}_j(m)\rangle\right)\right] \\ &= 4\sum_{l=1}^{M}\sum_{m=1}^{M}\mathrm{Cov}\left(\hat{n}_i(l),\hat{n}_j(m)\right),\end{aligned} \tag{B.64}$$

where the expectation values are taken with respect to the initial probe state, and $\hat{n}_j(m)$ are the number operators on the j^{th} mode in the Heisenberg picture prior to the m^{th} pass, as specified in Eq. (B.39). Without loss of generality, we make the assumption that $\|\boldsymbol{\alpha}\|_{1,\mathcal{P}} \geq \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}$. Summing over $i, j \in \mathcal{P}$, we have that, for an optimal protocol,

$$\sum_{i\in\mathcal{P}}\sum_{j\in\mathcal{P}}\mathcal{F}(\boldsymbol{\theta})_{ij} = \sum_{j\in\mathcal{P}}\frac{(MN)^2}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\alpha_j = (MN)^2, \tag{B.65}$$

where we used the condition in Eq. (3.24) for an optimal protocol, and we recall that, for $j \in \mathcal{P}$, all $\alpha_j > 0$. For convenience, define

$$\hat{P}(m) := \sum_{j\in\mathcal{P}}\hat{n}_j(m). \tag{B.66}$$

Armed with this definition, we can upper bound the sum over $i, j \in \mathcal{P}$ in the explicit expression from Eq. (B.64) as

$$\begin{aligned}\sum_{i\in\mathcal{P}}\sum_{j\in\mathcal{P}}\mathcal{F}(\boldsymbol{\theta})_{ij} &= 4\sum_{l=1}^{M}\sum_{m=1}^{M}\mathrm{Cov}\left(\hat{P}(l),\hat{P}(m)\right)\\ &\leq 4\sum_{l=1}^{M}\sum_{m=1}^{M}\sqrt{\mathrm{Var}(\hat{P}(l))\mathrm{Var}(\hat{P}(m))} = 4\left(\sum_{l=1}^{M}\sqrt{\mathrm{Var}(\hat{P}(l))}\right)^2\\ &\leq 4\left(\sum_{l=1}^{M}\frac{\left\|\hat{P}(l)\right\|_{s,N}}{2}\right)^2\\ &\leq (NM)^2, \qquad \text{(B.67)}\end{aligned}$$

where in the first line we use the Cauchy-Schwarz inequality, in the second line we use that once restricted to the N-particle subspace $\mathrm{Var}(A) \leq \|A\|_{s,N}^2/4$ (where, again, $\|A\|_{s,N}$ is the seminorm restricted to the N-particle subspace) for any Hermitian operator A, and in the final line we use that $\left\|\hat{P}(l)\right\|_{s,N} \leq N$. Comparing Eq. (B.67) with Eq. (B.65), we find that, for any optimal protocol, all inequalities in Eq. (B.67) must be saturated. Specifically,

$$\mathrm{Cov}\left(\hat{P}(l),\hat{P}(m)\right)^2 = \mathrm{Var}(\hat{P}(l))\mathrm{Var}(\hat{P}(m)), \qquad \text{(B.68)}$$

$$\mathrm{Var}(\hat{P}(l)) = \frac{N^2}{4}. \qquad \text{(B.69)}$$

The second condition, Eq. (B.69), means that, at all times, the state of our system must be of the form

$$\frac{|\boldsymbol{N}(l)\rangle_{\mathcal{P}}|\mathbf{0}\rangle_{\mathcal{NR}} + e^{i\varphi_l}|\mathbf{0}\rangle_{\mathcal{P}}|\boldsymbol{N}'(l)\rangle_{\mathcal{NR}}}{\sqrt{2}}, \qquad \text{(B.70)}$$

where we are using the simplifying notation from the statement of the lemma. In particular, the subscripts $\mathcal{P},\mathcal{N},\mathcal{R}$ refer to the collection of all modes associated with $\alpha_j \geq 0, \alpha_j < 0$, and the reference modes, respectively. Therefore, the state $|\boldsymbol{N}\rangle_{\mathcal{P}}|\mathbf{0}\rangle_{\mathcal{NR}}$ means that all photons are distributed (in some potentially arbitrary way) amongst the modes with non-negative α_j, and there are no photons in the modes with negative α_j or in the reference modes. Contrastingly, $|\mathbf{0}\rangle_{\mathcal{P}} |\boldsymbol{N}'(l)\rangle_{\mathcal{N}\mathcal{R}}$ refers to a state where there are N photons in the negative and reference modes, and there are no photons in the non-negative modes. We have also shifted to the Schrödinger picture where we move the time dependence onto the state as opposed to the operators. It is simple to verify that this state satisfies Eq. (B.69), and it is also simple to verify these are the most general states that achieve this. Intuitively, $|\psi_m\rangle$ is a generalized N00N state between the positive and negative/reference modes. □

In addition, we have the following useful characterization of optimal protocols:

Lemma B.4 *Let $|\psi_i\rangle$ be a state of the form in Lemma 3.1. Refer to the first and second parts of its superposition as, respectively, the first and second or positive and non-positive branches. Let U_m be the unitary that maps the initial state $|\psi_1\rangle$ to the state just before the m-th pass, $|\psi_m\rangle$, given by*

$$U_m = \begin{cases} \prod_{i=1}^{m-1} U^{(i)} V, & M+1 \geq m \geq 2 \\ I, & m = 1. \end{cases} \tag{B.71}$$

in agreement with Eq. (3.4). Then, if U_m is part of an optimal protocol, it coherently maps the first (second) branch of $|\psi_1\rangle$ to the first (second) branch of $|\psi_m\rangle$.

Proof. We use the covariance equality in Eq. (B.68). To proceed, we evaluate the expectation value of $\hat{P}$ in the initial state. Here, we will again use the Schrödinger picture.

$$\langle\psi_1|\,\hat{P}(l)\,|\psi_1\rangle = \langle\psi_l|\,\hat{P}\,|\psi_l\rangle \tag{B.72}$$

$$= \frac{1}{2}\left(\langle\boldsymbol{N}(l)|_{\mathcal{P}}\,\langle\mathbf{0}|_{\mathcal{NR}} + e^{-i\varphi_l}\,\langle\mathbf{0}|_{\mathcal{P}}\,\langle\boldsymbol{N}'(l)|_{\mathcal{NR}}\right)\hat{P}\left(|\boldsymbol{N}(l)\rangle_{\mathcal{P}}\,|\mathbf{0}\rangle_{\mathcal{NR}} + e^{i\varphi_l}\,|\mathbf{0}\rangle_{\mathcal{P}}\,|\boldsymbol{N}'(l)\rangle_{\mathcal{NR}}\right) \tag{B.73}$$

$$= \frac{1}{2}\left(\langle\boldsymbol{N}(l)|_{\mathcal{P}}\,\langle\mathbf{0}|_{\mathcal{NR}} + e^{-i\varphi_l}\,\langle\mathbf{0}|_{\mathcal{P}}\,\langle\boldsymbol{N}'(l)|_{\mathcal{NR}}\right) N\left(|\boldsymbol{N}(l)\rangle_{\mathcal{P}}\,|\mathbf{0}\rangle_{\mathcal{NR}}\right) \tag{B.74}$$

$$= \frac{N}{2}. \tag{B.75}$$

We next evaluate the covariance:

$$\mathrm{Cov}\left(\hat{P}(l),\hat{P}(m)\right) = \langle\psi_1|\,\hat{P}(l)\hat{P}(m)\,|\psi_1\rangle - \langle\psi_1|\,\hat{P}(l)\,|\psi_1\rangle\,\langle\psi_1|\,\hat{P}(m)\,|\psi_1\rangle \tag{B.76}$$

$$= \langle\psi_l|\,\hat{P}U_lU_m^\dagger\hat{P}\,|\psi_m\rangle - \langle\psi_l|\,\hat{P}\,|\psi_l\rangle\,\langle\psi_m|\,\hat{P}\,|\psi_m\rangle \tag{B.77}$$

$$= \frac{N^2}{2}\,\langle\boldsymbol{N}(l)|_{\mathcal{P}}\,\langle\mathbf{0}|_{\mathcal{NR}}\,U_lU_m^\dagger\,|\boldsymbol{N}(m)\rangle_{\mathcal{P}}\,|\mathbf{0}\rangle_{\mathcal{NR}} - \frac{N^2}{4}, \tag{B.78}$$

where in the last line we have used the fact that $\hat{P}$ gives a factor of N when acting on the first branch of states $|\psi_l\rangle$ and $|\psi_m\rangle$, but it annihilates the second branch that has zero photons in the positive modes.

In order for Eq. (B.68) to be satisfied, and using Eq. (B.69), we therefore require that, for all pairs of passes l, m,

$$\langle\boldsymbol{N}(l)|_{\mathcal{P}}\,\langle\mathbf{0}|_{\mathcal{NR}}\,U_lU_m^\dagger\,|\boldsymbol{N}(m)\rangle_{\mathcal{P}}\,|\mathbf{0}\rangle_{\mathcal{NR}} = 1. \tag{B.79}$$

Choosing $l = 1$, this implies that we require that

$$U_m^\dagger \left|\boldsymbol{N}(m)\right\rangle_{\mathcal{P}} \left|\mathbf{0}\right\rangle_{\mathcal{NR}} = \left|\boldsymbol{N}(0)\right\rangle_{\mathcal{P}} \left|\mathbf{0}\right\rangle_{\mathcal{NR}} =: \left|\psi_1\right\rangle_{\mathcal{P}}, \tag{B.80}$$

where we are defining $\left|\psi_1\right\rangle_{\mathcal{P}}, \left|\psi_1\right\rangle_{\mathcal{NR}}$ such that $\left|\psi_0\right\rangle \propto \left|\psi_1\right\rangle_{\mathcal{P}} + \left|\psi_1\right\rangle_{\mathcal{NR}}$ in the obvious way. Moving the unitary onto the right hand side of the equation yields

$$\left|\psi_m\right\rangle_{\mathcal{P}} = U_m \left|\psi_1\right\rangle_{\mathcal{P}}, \tag{B.81}$$

which of course implies the corresponding equation for the second branch by linearity. □

B.6 Fisher Information Matrix Conditions for Quadrature Displacements

In this Appendix, we provide conditions on the quantum Fisher information matrix for an optimal protocol in the case of quadrature generators. This result yields a simpler form of the saturability condition of Eq. (3.25), although the set of states that it picks out is less clear than in the number operator case. This issue is compounded by the fact that the bound is not actually saturable (it can only be approached asymptotically as $\overline{N} \to \infty$). Regardless, it allows us to bring quadrature displacements into our general formalism and suggests a route towards designing additional optimal protocols beyond those already in the literature.

In particular, starting with the definition of $\hat{p}_i(l)$ from Eq. (B.42), we can bound the sum over the quantum Fisher information matrix elements as

$$\sum_{i=1,j=1}^{d} \mathcal{F}(\boldsymbol{\theta})_{ij} = \sum_{i=1,j=1}^{d} 4 \sum_{l=1}^{M} \sum_{m=1}^{M} \mathrm{Cov}(\hat{p}_i(l), \hat{p}_j(m)) \tag{B.82}$$

$$\leq 4 \sum_{l=1}^{M} \sum_{m=1}^{M} \sqrt{\mathrm{Var}\Big(\sum_{i=1}^{d} \hat{p}_i(l)\Big) \mathrm{Var}\Big(\sum_{i=1}^{d} \hat{p}_j(m)\Big)} \tag{B.83}$$

$$= 4 \left(\sum_{l=1}^{M} \sqrt{\mathrm{Var}\Big(\sum_{i=1}^{d} \hat{p}_i(l)\Big)} \right)^2 \tag{B.84}$$

$$\leq 4 \left(\sum_{l=1}^{M} \sqrt{\overline{N} - \frac{d}{2}} \right)^2 = 4M^2 \left(\overline{N} - \frac{d}{2} \right) \sim 4M^2 \overline{N}. \tag{B.85}$$

Above, in Eq. (B.83), we used the Cauchy-Schwarz inequality; in Eq. (B.85), we used the uncertainty relation in Eq. (B.35). Consistent with the rest of Chapter 3 and Appendix B, the $\sim$ symbol denotes asymptotically in $\overline{N}$ (for $\overline{N} \gg d$).

The saturability condition in Eq. (3.25) states that, for an optimal protocol (asymptotically in $\overline{N}$), it must hold that $\boldsymbol{\alpha}$ is an eigenvector of $\mathcal{F}(\boldsymbol{\theta})$ with eigenvalue $4M^2\overline{N}$. Thus, for an optimal protocol,

$$\mathrm{Tr}(\mathcal{F}) = \sum_{j=1}^{d} \lambda_j \gtrsim 4M^2 \overline{N}, \tag{B.86}$$

where λ_j are the eigenvalues of $\mathcal{F}$. This implies that the chain of inequalities leading to Eq. (B.85) must be saturated (asymptotically in $\overline{N}$) for an optimal protocol and that the largest eigenvalue of $\mathcal{F}$ must be $\lambda_1 \sim 4m^2\overline{N}$ with all other eigenvalues zero. It immediately follows that the saturability condition for quadrature displacements can be written as

$$\mathcal{F}(\boldsymbol{\theta})_{ij} \sim \frac{4M^2\overline{N}}{\|\boldsymbol{\alpha}\|_2^2} \alpha_i \alpha_j. \tag{B.87}$$

B.7 Approaching the Single-Shot Limit and Robust Phase Estimation[1]

As pointed out in the footnote preceding Eq. (3.8) and in the discussion of what defines an information-theoretically optimal protocol in Section 3.4.2, it is not, in practice, possible to construct an unbiased estimator achieving the single shot ($\mu = 1$) quantum Cramér-Rao bound that we analyze in Chapter 3 and Appendix B, as the quantum Cramér-Rao bound is only guaranteed to be achievable in the limit of asymptotically large amounts of data ($\mu \to \infty$). Resolving this tension while still achieving asymptotic Heisenberg scaling in the total amount of resources (here, μN photons) requires carefully designed protocols. In particular, extracting a relative phase from the probe states considered in the protocols in Chapter 3 and Appendix B requires a proper division of resources so that, asymptotically, the single-shot bound is achieved up to a small constant.

At best, this constant can be reduced to π^2 [76], but the non-adaptive robust phase estimation scheme of Refs. [71–73] provides a relatively simple-to-implement approach with a multiplicative overhead of $(24.26\pi)^2$. In brief, these protocols work by dividing the protocol into K stages where in stage j one uses N_j photons (or $\overline{N}_j$ average photons for displacement sensing). In each stage, one imprints the unknown function into the phase between two branches of a cat-like state of N_j photons and then performs a measurement, as described in Chapter 3. The experiment is performed ν_j times, allowing one to obtain an estimate of the unknown phase. This estimate is refined over the course of the K stages, with more photons used in each additional

[1] Note added: See also Appendix A.3 for a similar discussion.

stage such that the total photon resources are

$$N = \sum_{j=1}^{K} \nu_j N_j. \tag{B.88}$$

An optimal choice of ν_j and N_j ensures that, asymptotically, $N_K = \Theta(N)$ and $\nu_K = \mathcal{O}(1)$, and, thus, the asymptotic scaling of the single-shot bound is obtained up to a multiplicative constant that depends on the details of the optimization. The proof of this and the associated optimization are detailed in Refs. [71–73].

Appendix C: Appendices Associated with Chapter 4

In this Appendix, we provide details behind many of the expressions in Chapter 4. In particular, we derive expressions for the first and second moments of the output probabilities in the form of Kronecker δs and derive Eqs. (4.5), (4.8) and (4.9). In addition, we prove Theorem 4.1, Theorem 4.2, and Lemma 4.1 that are presented as building blocks toward the proof of a transition in anticoncentration. We also explain more thoroughly the connection between the hiding property and the first moment of hafnians of generalized COE matrices. We further contextualize our definition of anticoncentration with respect to the literature and show how anticoncentration of the approximate distribution of output probabilities connects to anticoncentration of the true distribution. Finally, we use Scattershot Boson Sampling as intuition for why the transition in anticoncentration in Gaussian Boson Sampling exists.

C.1 Algebraic Details of the First Moment—Derivation of Eq. (4.5)

In this Appendix, we derive Eq. (4.5), which gives an expression for the first moment of the output probabilities in terms of Kronecker δs. We use Eq. (4.2) to expand the hafnian. We then use properties of independent Gaussians to simplify the expression. Specifically,

$$\mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\left|\mathrm{Haf}(X^\intercal X)\right|^2\right] = \left(\frac{1}{2^n n!}\right)^2 \sum_{\sigma,\tau\in S_{2n}} \mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\prod_{j=1}^{n}\left(\sum_{\ell_j=1}^{k} X_{\ell_j\sigma(2j-1)}X_{\ell_j\sigma(2j)}\right)\left(\sum_{o_j=1}^{k} X^*_{o_j\tau(2j-1)}X^*_{o_j\tau(2j)}\right)\right] \tag{C.1}$$

$$= \left(\frac{1}{2^n n!}\right)^2 \sum_{\sigma,\tau\in S_{2n}} \mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\sum_{\{\ell_i,o_i\}_{i=1}^{n}=1}^{k}\left(\prod_{j=1}^{n} X_{\ell_j\sigma(2j-1)}X_{\ell_j\sigma(2j)}X^*_{o_j\tau(2j-1)}X^*_{o_j\tau(2j)}\right)\right] \tag{C.2}$$

$$= \left(\frac{1}{2^n n!}\right)^2 \sum_{\sigma,\tau\in S_{2n}} \sum_{\{\ell_i,o_i\}_{i=1}^{n}=1}^{k} \mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\left(\prod_{j=1}^{n} X_{\ell_j\sigma(2j-1)}X_{\ell_j\sigma(2j)}X^*_{o_j\tau(2j-1)}X^*_{o_j\tau(2j)}\right)\right] \tag{C.3}$$

$$= \left(\frac{1}{2^n n!}\right)^2 \sum_{\sigma,\tau\in S_{2n}} \sum_{\{\ell_i,o_i\}_{i=1}^{n}=1}^{k} \left(\prod_{j=1}^{n} \delta_{\ell_j o_{j'}}\delta_{\ell_j o_{j''}}\right), \tag{C.4}$$

where we have defined j' to be the index such that $\sigma(2j-1) = \tau(2j'-1)$ or $\tau(2j')$. Similarly, j'' is the index such that $\sigma(2j) = \tau(2j''-1)$ or $\tau(2j'')$. Observe that $j' = j''$ if $\{\sigma(2j-1), \sigma(2j)\} = \{\tau(2j-1), \tau(2j)\}$ (note that this is an equality of *sets*, meaning order does not matter). The first equation uses the definition of the hafnian, while the second follows from exchanging product and sum. The penultimate equation comes from the linearity of expectation. To get to the final equation, first recall that the X_{ij} are i.i.d. complex Gaussian random variables with mean 0 and variance 1. This means that the expectation value of a product of entries vanishes unless there are an equal number of unconjugated and conjugated copies of all indices. By the definition of j', we ensure that the entry $X_{\ell_j\sigma(2j-1)}$ is matched to one of the X^* entries as long as ℓ_j and $o_{j'}$ match, hence the first Kronecker δ. The second Kronecker δ follows similarly.

We can exactly calculate j':

$$\sigma(2j{-}1) \in \{\tau(2j'{-}1), \tau(2j')\} \iff \tau^{-1}(\sigma(2j{-}1)) \in \{2j'{-}1, 2j'\} \iff \frac{\tau^{-1}(\sigma(2j-1))}{2} \in \{j'{-}\frac{1}{2}, j'\}. \tag{C.5}$$

Thus:

$$j' = \left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2} \right\rceil. \tag{C.6}$$

Similarly:

$$j'' = \left\lceil \frac{\tau^{-1}(\sigma(2j))}{2} \right\rceil. \tag{C.7}$$

Therefore,

$$\begin{aligned}
\mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[|\mathrm{Haf}(X^\top X)|^2\right] &= \left(\frac{1}{2^n n!}\right)^2 \sum_{\sigma,\tau\in S_{2n}} \sum_{\{\ell_i,o_i\}_{i=1}^n=1}^{k} \left(\prod_{j=1}^n \delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2}\right\rceil}} \delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(\sigma(2j))}{2}\right\rceil}}\right) && \text{(C.8)} \\
&= \frac{(2n)!}{(2^n n!)^2} \sum_{\tau\in S_{2n}} \left[\sum_{\{\ell_i,o_i\}_{i=1}^n=1}^{k} \left(\prod_{j=1}^n \delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(2j-1)}{2}\right\rceil}} \delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(2j)}{2}\right\rceil}}\right)\right] && \text{(C.9)} \\
&= \frac{(2n)!}{(2^n n!)^2} \sum_{\tau\in S_{2n}} \left[\sum_{\{o_i\}_{i=1}^n=1}^{k} \left(\prod_{j=1}^n \delta_{o_{\left\lceil \frac{\tau(2j-1)}{2}\right\rceil}, o_{\left\lceil \frac{\tau(2j)}{2}\right\rceil}}\right)\right]. && \text{(C.10)}
\end{aligned}$$

In the first equality, we have used Eqs. (C.6) and (C.7). In the second, we notice that τ and σ occur only together as $\tau^{-1}\circ\sigma$, meaning we can perform a change of variables to convert our double summation over permutations in S_{2n} to a single summation over a redefined τ^{-1} while gaining a factor $(2n)!$. The third equality comes from summing over the ℓ_j indices and redefining $\tau^{-1}\to\tau$. This is Eq. (4.5).

C.2 Algebraic Details of the Second Moment—Derivation of Eqs. (4.8) and (4.9)

In this Appendix, we generalize the calculation of Appendix C.1 to the second moment of the output probabilities. The structure of the derivation is very similar, but the details are more nuanced due to the increased number of copies of X.

We again begin with some algebraic manipulations:

$$\mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\left|\mathrm{Haf}(X^\intercal X)\right|^4\right] = \left(\frac{1}{2^n n!}\right)^4 \sum_{\sigma,\tau,\alpha,\beta\in S_{2n}} \mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\prod_{j=1}^{n}\left(\sum_{\ell_j=1}^{k} X_{\ell_j\sigma(2j-1)}X_{\ell_j\sigma(2j)}\right)\left(\sum_{o_j=1}^{k} X^*_{o_j\tau(2j-1)}X^*_{o_j\tau(2j)}\right)\right.$$
$$\left.\times\left(\sum_{p_j=1}^{k} X_{p_j\alpha(2j-1)}X_{p_j\alpha(2j)}\right)\left(\sum_{q_j=1}^{k} X^*_{q_j\beta(2j-1)}X^*_{q_j\beta(2j)}\right)\right] \tag{C.11}$$

$$= \frac{1}{(2^n n!)^4}\sum_{\sigma,\tau,\alpha,\beta\in S_{2n}}\sum_{\{\ell_i,o_i,p_i,q_i\}_{i=1}^{n}=1}^{k}$$
$$\mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\prod_{j=1}^{n} X_{\ell_j\sigma(2j-1)}X_{\ell_j\sigma(2j)}X^*_{o_j\tau(2j-1)}X^*_{o_j\tau(2j)}X_{p_j\alpha(2j-1)}X_{p_j\alpha(2j)}X^*_{q_j\beta(2j-1)}X^*_{q_j\beta(2j)}\right]. \tag{C.12}$$

This first equation simply comes from the definition of the hafnian, and the second from exchanging product and sum and using the linearity of expectation. As in the proof of the first moment, we must properly match the indices of the Gaussian elements. Recall that, in order for the expectation value not to vanish, the indices i,j must show up an equal number of times in a conjugated and non-conjugated copy of X (otherwise, the expectation value of that term will vanish because our Gaussian is complex with zero mean). To proceed, first recall that permutations are bijective. Therefore, for all j and any given permutation η, there is a unique value y_j

such that $\sigma(2j-1) = \eta(2y_j - 1)$ or $\sigma(2j-1) = \eta(2y_j)$. Similarly, there is a unique value y'_j such that $\sigma(2j) = \eta(2y'_j - 1)$ or $\sigma(2j) = \eta(2y'_j)$. Using this bijectivity and the independence of matrix elements allows us to separate the single expectation value on the $8n$ matrix elements in Eq. (C.12) into a product of $2n$ expectation values of 4 elements:

$$\prod_{j=1}^{n} \mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\Big[X_{\ell_j\sigma(2j-1)}X_{p_{k_j}\sigma(2j-1)}X^*_{o_{i_j}\sigma(2j-1)}X^*_{q_{m_j}\sigma(2j-1)}\Big]\mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\Big[X_{\ell_j\sigma(2j)}X_{p_{k'_j}\sigma(2j)}X^*_{o_{i'_j}\sigma(2j)}X^*_{q_{m'_j}\sigma(2j)}\Big]. \tag{C.13}$$

To explain more thoroughly: we have defined i_j, k_j, m_j to be the indices that map to $\sigma(2j-1)$ under τ, α, β, respectively, in the sense that either $\eta(2y_j - 1) = \sigma(2j-1)$ or $\eta(2y_j) = \sigma(2j-1)$ for $\eta \in \{\tau, \alpha, \beta\}$ and $y \in \{i, k, m\}$, respectively. Because two matrix elements are necessarily independent if they do not match on the second index, we can separate all elements with $\sigma(2j-1)$ as the second element into a single expectation value, hence the first term. To get the second term, we repeat this argument where i'_j, k'_j, m'_j are the indices that map to $\sigma(2j)$ under τ, α, β, respectively, in the sense that either $\eta(2y_j - 1) = \sigma(2j)$ or $\eta(2y_j) = \sigma(2j)$ for $\eta \in \{\tau, \alpha, \beta\}$ and $y \in \{i, k, m\}$, respectively.

Now consider the first expectation value. For a nonvanishing expectation value, we must appropriately match the first indices of the matrix elements. We have three options: either all four indices can match, or the indices can be paired off in one of two ways. In the former case, the expectation value yields 2 given that the elements are complex Gaussian with mean 0 and variance 1. By the same logic, the latter two cases yield an expectation value of 1. In summary,

$$\ell_j = p_{k_j} = o_{i_j} = q_{m_j} \implies \mathbb{E} \to 2, \tag{C.14}$$

$$(\ell_j \neq p_{k_j}) \wedge (\ell_j = o_{i_j}) \wedge (p_{k_j} = q_{m_j}) \implies \mathbb{E} \to 1, \tag{C.15}$$

$$(\ell_j \neq p_{k_j}) \wedge (\ell_j = q_{m_j}) \wedge (p_{k_j} = o_{i_j}) \implies \mathbb{E} \to 1. \tag{C.16}$$

One might naively think that there should be another contribution from matching indices as

$$(\ell_j = p_{k_j}) \wedge (\ell_j \neq q_{m_j}) \wedge (q_{m_j} = o_{i_j}). \tag{C.17}$$

However, the expectation value in this case actually vanishes, as we are working with *complex* Gaussian random variables, meaning the indices need to be matched such that there are an equal number of conjugated and non-conjugated indices.

We can write this in one simple expression using Kronecker δs as

$$2\delta_{\ell_j p_{k_j} o_{i_j} q_{m_j}} + \delta_{\ell_j o_{i_j}} \delta_{p_{k_j} q_{m_j}} (1 - \delta_{\ell_j p_{k_j}}) + \delta_{\ell_j q_{m_j}} \delta_{p_{k_j} o_{i_j}} (1 - \delta_{\ell_j p_{k_j}}) = \delta_{\ell_j o_{i_j}} \delta_{p_{k_j} q_{m_j}} + \delta_{\ell_j q_{m_j}} \delta_{p_{k_j} o_{i_j}}. \tag{C.18}$$

That is,

$$\mathop{\mathbb{E}}_{X \sim \mathcal{G}^{k \times 2n}} \left[X_{\ell_j \sigma(2j-1)} X_{p_{k_j} \sigma(2j-1)} X^*_{o_{i_j} \sigma(2j-1)} X^*_{q_{m_j} \sigma(2j-1)} \right] = \delta_{\ell_j o_{i_j}} \delta_{p_{k_j} q_{m_j}} + \delta_{\ell_j q_{m_j}} \delta_{p_{k_j} o_{i_j}}, \tag{C.19}$$

which is essentially an application of Isserlis'/Wick's theorem. Equivalent calculations as those used to derive Eqs. (C.6) and (C.7) can be made to rewrite each of $o_{i_j}, p_{k_j}, q_{m_j}$ in terms of j, giving

$$\delta_{\ell_j o_{i_j}} \delta_{p_{k_j} q_{m_j}} + \delta_{\ell_j q_{m_j}} \delta_{p_{k_j} o_{i_j}} =$$

$$\delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j-1))}{2} \right\rceil} q_{\left\lceil \frac{\beta^{-1}(\sigma(2j-1))}{2} \right\rceil}} + \delta_{\ell_j q_{\left\lceil \frac{\beta^{-1}(\sigma(2j-1))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j-1))}{2} \right\rceil} o_{\left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2} \right\rceil}}. \quad \text{(C.20)}$$

Thus

$$\underset{X \sim \mathcal{G}^{k\times 2n}}{\mathbb{E}}\left[X_{\ell_j \sigma(2j-1)} X_{p_{k_j} \sigma(2j-1)} X^*_{o_{i_j} \sigma(2j-1)} X^*_{q_{m_j} \sigma(2j-1)}\right] \underset{X \sim \mathcal{G}^{k\times 2n}}{\mathbb{E}}\left[X_{\ell_j \sigma(2j)} X_{p_{k'_j} \sigma(2j)} X^*_{o_{i'_j} \sigma(2j)} X^*_{q_{m'_j} \sigma(2j)}\right] =$$
$$\left(\delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j-1))}{2} \right\rceil} q_{\left\lceil \frac{\beta^{-1}(\sigma(2j-1))}{2} \right\rceil}} + \delta_{\ell_j q_{\left\lceil \frac{\beta^{-1}(\sigma(2j-1))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j-1))}{2} \right\rceil} o_{\left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2} \right\rceil}}\right)$$
$$\times \left(\delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(\sigma(2j))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j))}{2} \right\rceil} q_{\left\lceil \frac{\beta^{-1}(\sigma(2j))}{2} \right\rceil}} + \delta_{\ell_j q_{\left\lceil \frac{\beta^{-1}(\sigma(2j))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j))}{2} \right\rceil} o_{\left\lceil \frac{\tau^{-1}(\sigma(2j))}{2} \right\rceil}}\right). \quad \text{(C.21)}$$

Therefore,

$$\underset{X \sim \mathcal{G}^{k\times 2n}}{\mathbb{E}}\left[\left|\mathrm{Haf}(X^\top X)\right|^4\right] = \left(\frac{1}{2^n n!}\right)^4 \sum_{\sigma,\tau,\alpha,\beta \in S_{2n}} \sum_{\{\ell_i, o_i, p_i, q_i\}_{i=1}^n = 1}^{k} \Bigg[\prod_{j=1}^{n}$$
$$\left(\delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j-1))}{2} \right\rceil} q_{\left\lceil \frac{\beta^{-1}(\sigma(2j-1))}{2} \right\rceil}} + \delta_{\ell_j q_{\left\lceil \frac{\beta^{-1}(\sigma(2j-1))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j-1))}{2} \right\rceil} o_{\left\lceil \frac{\tau^{-1}(\sigma(2j-1))}{2} \right\rceil}}\right)$$
$$\times \left(\delta_{\ell_j o_{\left\lceil \frac{\tau^{-1}(\sigma(2j))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j))}{2} \right\rceil} q_{\left\lceil \frac{\beta^{-1}(\sigma(2j))}{2} \right\rceil}} + \delta_{\ell_j q_{\left\lceil \frac{\beta^{-1}(\sigma(2j))}{2} \right\rceil}} \delta_{p_{\left\lceil \frac{\alpha^{-1}(\sigma(2j))}{2} \right\rceil} o_{\left\lceil \frac{\tau^{-1}(\sigma(2j))}{2} \right\rceil}}\right)\Bigg]. \quad \text{(C.22)}$$

We can again reparameterize our sums over the permutations by performing a change of variables

$(\eta^{-1} \circ \sigma) \to \eta$ for $\eta \in \{\tau, \alpha, \beta\}$. This yields

$$\underset{X \sim \mathcal{G}^{k \times 2n}}{\mathbb{E}}\left[|\mathrm{Haf}(X^\top X)|^4\right] = \left(\frac{1}{2^n n!}\right)^4 (2n)! \sum_{\tau,\alpha,\beta \in S_{2n}} \sum_{\{\ell_i, o_i, p_i, q_i\}_{i=1}^n = 1}^{k} \Bigg[\prod_{j=1}^{n} \\ \left(\delta_{\ell_j o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} + \delta_{\ell_j q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \right) \\ \times \left(\delta_{\ell_j o_{\lceil \frac{\tau(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} + \delta_{\ell_j q_{\lceil \frac{\beta(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \right) \Bigg]. \tag{C.23}$$

Expanding the product and summing over ℓ_j yields

$$\underset{X \sim \mathcal{G}^{k \times 2n}}{\mathbb{E}}\left[|\mathrm{Haf}(X^\top X)|^4\right] = \left(\frac{1}{2^n n!}\right)^4 (2n)! \sum_{\tau,\alpha,\beta \in S_{2n}} \sum_{\{o_i, p_i, q_i\}_{i=1}^n = 1}^{k} \Bigg[\prod_{j=1}^{n} \\ \Bigg(\delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} + \delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} + \\ \delta_{q_{\lceil \frac{\beta(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} + \delta_{q_{\lceil \frac{\beta(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \Bigg) \Bigg]. \tag{C.24}$$

This equation is the starting point of a new graph-theoretic approach.

As discussed in Chapter 4, we use Eq. (C.24) to define graphs, examples of which are provided in Fig. 4.2(b) and Fig. C.2(a). Specifically, we let $G_{\tau,\alpha,\beta}(z)$ be a graph on $6n$ vertices, with labels $\{O_i, P_i, Q_i\}_{i=1}^{2n}$, and z an integer from 1 to 4^n. As was the case for the proof of the first moment, we use the Kronecker δs to define black and red edges. z enumerates the different patterns of black edges, and τ, α, β determine the red edges. Specifically, there is a red edge between O_j and $O_{j'}$ if $\lceil \tau(j)/2 \rceil = \lceil \tau(j')/2 \rceil$, and similarly for the O and Q vertices using permutations α and β, respectively. However, given a choice of permutations, there are 4^n possible sets of black edges that correspond to the 4^n possible combinations of terms in Eq. (C.24). The sets of edges corresponding to each term are listed below:

$$\delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \to \{(O_{2j-1}, O_{2j}), (P_{2j-1}, Q_{2j-1}), (P_{2j}, Q_{2j})\}, \tag{C.25}$$

$$\delta_{o_{\lceil \frac{\tau(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \to \{(O_{2j-1}, Q_{2j}), (P_{2j-1}, Q_{2j-1}), (O_{2j}, P_{2j})\}, \tag{C.26}$$

$$\delta_{q_{\lceil \frac{\beta(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \to \{(O_{2j}, Q_{2j-1}), (P_{2j-1}, O_{2j-1}), (P_{2j}, Q_{2j})\}, \tag{C.27}$$

$$\delta_{q_{\lceil \frac{\beta(2j-1)}{2} \rceil} q_{\lceil \frac{\beta(2j)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j-1)}{2} \rceil} o_{\lceil \frac{\tau(2j-1)}{2} \rceil}} \delta_{p_{\lceil \frac{\alpha(2j)}{2} \rceil} o_{\lceil \frac{\tau(2j)}{2} \rceil}} \to \{(O_{2j-1}, P_{2j-1}), (O_{2j}, P_{2j}), (Q_{2j-1}, Q_{2j})\}. \tag{C.28}$$

We refer to these sets of black edges as type-1, type-2, type-3, and type-4, respectively. We take the convention that our graphs have the vertices organized into three rows and $2n$ columns. The first, second, and third rows correspond to type-O, $-P$, and $-Q$ vertices, respectively. The columns are ordered by index i. Using this convention, black edges are constrained to lie within groups of two columns $2i - 1$ and $2i$ using one of the four patterns described above. Again, see Fig. 4.2(b) and Fig. C.2(a) for examples (please note that Fig. C.2(a) is not fully general, as it only has type-1 and type-4 black edges, but it does show that patterns of black edges can repeat, and it shows how z identifies the patterns of black edges present in the graph).

We repeat the conclusion of Chapter 4, which is that we can map the number of "free indices" in the Kronecker δs to the number of connected components $C(G_{\tau,\alpha,\beta}(z))$ of the graph $G_{\tau,\alpha,\beta}(z)$. Each graph contributes $k^{C(G_{\tau,\alpha,\beta}(z))}$ to the sum, which means that the second moment can be written as

$$M_2(k,n) = \frac{(2n)!}{(2^n n!)^4} \sum_{\tau,\alpha,\beta \in S_{2n}} \sum_{z \in [4^n]} k^{C(G_{\tau,\alpha,\beta}(z))}, \tag{C.29}$$

which is Eq. (4.8). Removing the degeneracies induced by different permutations, and defining $\mathbb{G}_n^2(z)$ to be the set of graphs for the zth set of black edges and $\mathbb{G}_n^2 \coloneqq \bigcup_{z=1}^{4^n} \mathbb{G}_n^2(z)$, we get a final result of

$$M_2(k,n) = (2n-1)!! \sum_{G \in \mathbb{G}_n^2} k^{C(G)}. \tag{C.30}$$

This is Eq. (4.9).

C.3 Proofs of Theorem 4.1, Theorem 4.2, and Lemma 4.1

In this Appendix, we give the proofs of Theorem 4.1, Theorem 4.2, and Lemma 4.1 that were presented in Chapter 4. We start with a restatement and proof of Theorem 4.1, which gives the first moment of the output probabilities.

Theorem C.1. *The sum over graphs in $\mathbb{G}_n^1$ satisfies*

$$\sum_{G \in \mathbb{G}_n^1} k^{C(G)} = k(k+2)\dots(k+2n-2). \tag{C.31}$$

and hence $M_1(k,n) = (2n-1)!!(k+2n-2)!!/(k-2)!!$.

Proof. We proceed by induction on n. Let $f(k,n)$ be the LHS of Eq. (C.31). For the base case $n = 1$, there is only a single possible graph G that has a single connected component. Thus $f(k,1) = k$. For the inductive step, which is visualized in Fig. C.1, consider two subsets of $\mathbb{G}_n^1$. The first set has graphs that possess a red edge between O_1 and O_2, which means that these two vertices form their own connected component (recall that O_1 and O_2 are always connected with a black edge). Summing $k^{C(G)}$ over all graphs of this type then yields a contribution of $kf(k,n-1)$. The other subset of $\mathbb{G}_n^1$ has graphs that possess a red edge between O_1 and a vertex besides O_2, say O_x. In these graphs, the number of connected components in the graph does not change if one collapses the three vertices O_1, O_2, and O_x into a single vertex (because they are all connected by either a black or red edge). Therefore, because there are $2n-2$ choices for the vertex O_x linked to O_1 by a red edge, we get an overall contribution of $(2n-2)f(k,n-1)$ when summing $k^{C(G)}$ over these graphs.

Overall then, we find that

$$
\begin{aligned}
f(k,n) &= kf(k,n-1) + (2n-2)f(k,n-1) && \text{(C.32)}\\
&= (k+2n-2)f(k,n-1) && \text{(C.33)}\\
&= (k+2n-2)(k+2n-4)\dots(k+2)k, && \text{(C.34)}
\end{aligned}
$$

which proves the formula (and where the inductive hypothesis is used in the final equality). □

We note briefly that the structure of this proof is similar to that used in Ref. [182] to calculate $\mathop{\mathbb{E}}_{X\sim\mathcal{G}^{n\times n}}[|\mathrm{Haf}X|^4]$. The proofs are similar because there are four copies of X in each, but the proofs are not identical given that different definitions of the hafnian are used. Additionally, similar graphs, and a similar calculation involving enumerating the number of graphs of a given number of connected components, show up in the bioinformatic study of breakpoint graphs (which are a type of graph defined by two perfect matchings that show up in the theory of comparative genomics) [183].

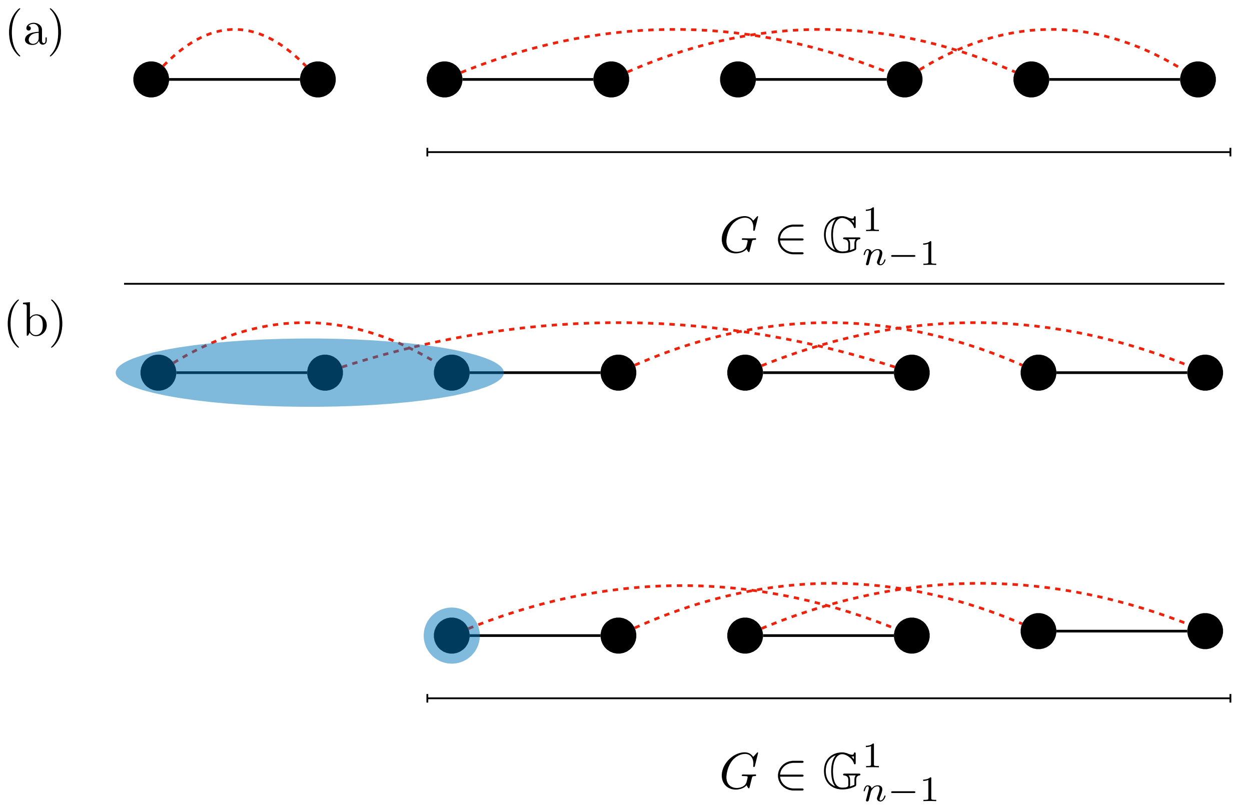

Figure C.1: Visualization of the inductive step in the proof of the first moment of the output probabilities (Theorem 4.1, restated here in this Appendix as Theorem C.1). The inductive step proceeds in two cases that are determined by the red edge that connects to the first vertex in the graph in $\mathbb{G}_n^1$. In (a), we consider the case where the first two vertices, which are linked by a black (solid) edge, are also linked by a red (dashed) edge, meaning they comprise a single connected component. This contributes a factor of k times the contribution from a graph in $\mathbb{G}_{n-1}^1$, which comes from the remaining $2n-2$ vertices and their edges. (b) considers the case where the first vertex is linked via a red edge to a different vertex $O_x \neq O_2$ for which there are $2n-2$ choices (here $a = 3$). The number of connected components does not change after identifying and combining the three vertices that are connected in this way (visualized by the blue background), meaning we again reduce down to a graph in $\mathbb{G}_{n-1}^1$, but this time without the multiplicative factor of k.

We next move on to a proof of Theorem 4.2, which gives the form of the second moment as a polynomial in k. We again restate the theorem for convenience.

Theorem C.2. *The second moment $M_2(k,n)$ is a degree-$2n$ polynomial in k and can be written as $M_2(k,n) = (2n-1)!! \sum_{i=1}^{2n} c_i k^i$, where c_i is the number of graphs $G \in \mathbb{G}_n^2$ that have i connected components.*

Proof. As mentioned in Chapter 4, once Eq. (4.9) is derived, the theorem follows after deriving the correct limits of summation. Trivially, the fewest possible number of connected components is 1. To see that the largest possible number of connected components is $2n$, we consider the four patterns of black edges that are illustrated in Fig. 4.2(b) and how many connected components can possibly occur in graphs with those different patterns. See also Fig. C.2 for a reminder of the patterns of black edges and a visual explanation of the following argument.

First note that, because all vertices are paired via black edges, every connected component has an even number of vertices. Therefore, the two smallest sizes of connected components are 2 and 4 vertices. In order to get a connected component of size 2, one must connect a pair of vertices with both a black and a red edge. Red edges are constrained to lie in a single row, meaning only type-1 and type-4 patterns of black edges, which contain a pair of vertices connected by a black edge in the same row, can yield a connect component of size 2. Pairing off the remaining vertical black edges yields connected components of size 4, the next smallest size.

Therefore, the maximum number of connected components arises from taking only type-1 and type-4 edges. This requires connecting each horizontal black edge by red edge (creating a connected component with 2 vertices) and then pairing off the vertical edges coming from the same type. This allows for the maximal 2 connected components per set of six vertices, meaning $2n$ total connected components. □

We also prove Lemma 4.1, which we again restate for convenience:

Lemma C.1 *We have that*

i. $M_2(1,n) = ((2n-1)!!)^4 4^n$

ii. $c_{2n} = (2n)!!$

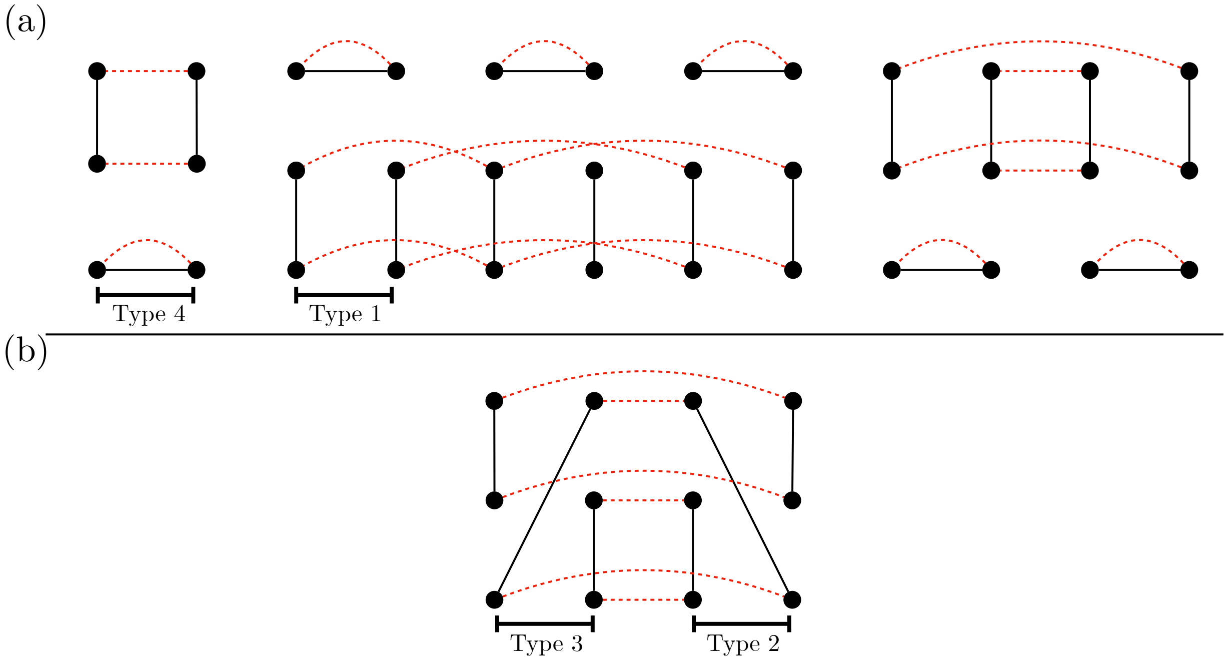

Figure C.2: (a) Example graph in $\mathbb{G}_6^2$ showing how to achieve an average of two connected components per set of six vertices using only type-1 and type-4 sets of edges. All vertices connected by horizontal black (solid) edges are also connected by red (dashed) edges. All type-1 vertical edges are paired off, as are type-4 vertical edges. Note that this graph would correspond to $z = 1 + 3 \times 4^5 + 0 \times 4^4 + 0 \times 4^3 + 0 \times 4^2 + 3 \times 4^1 + 3 \times 4^0 = 3088$. (b) Example showing how using type-2 and type-3 black edges lead to, at most, three connected components per two sets of six vertices.

Proof. Part (i): examine Eq. (C.24). Because $k = 1$, $o_i = p_i = q_i = 1$ for all i. Thus, regardless of the permutation, all Kronecker δs are always satisfied. This means that, independent of the permutation, each factor is always 4 such that the product becomes 4^n. The sum over the three copies of S_{2n} then simply yields a factor of $(2n)!^3$. The result then follows.

Part (ii): we argued in the proof of Theorem C.2 that the leading-order term in the polynomial expansion of the second moment is k^{2n}, and it comes from graphs that consist of only type-1 and type-4 black edges. Each type-1 and type-4 set of edges contains a horizontal black edge, and the two vertices linked by that black edge also must be linked by a red edge to create a 2-vertex connected component. Additionally, the vertical edges of the type-1 sets need to be paired off via red edges; similarly, the vertical edges of the type-4 sets need to be paired off. This ensures that each other connected component has exactly 4 vertices, maximizing the number of possible connected components.

Fig. C.3 visualizes how to now reduce the remaining calculation to the value of the first moment when $k = 2$. If we imagine collapsing each pair of adjacent vertical edges (i.e., those coming from the same group of 6 vertices) onto a pair of vertices connected by a black edge, we reproduce the atomic graph from the proof of the first moment. Here, by atomic graph, we mean the vertices and the fixed black edges which are shared by all graphs; the red edges are not yet included. Explicitly, there are $2n$ vertices, and vertices O_{2i-1}, O_{2i} are connected with a black edge. The black edges here act to identify that the original uncollapsed vertical edges were of the same type. Drawing red edges in the simplified graph on $2n$ vertices corresponds to pairing off vertical edges in the original graph on $6n$ vertices with red edges. Note that this also implies that red edges connect vertical edges of the same type. Therefore, a connected component in the simplified graph could correspond to two preimages in the original graph: either all type-1 vertical edges, or all type-4 vertical edges. Then, by summing over all graphs and weighting each connected component by 2, we are effectively evaluating $f(2, n)$ in Eq. (C.31), i.e.:

$$f(2,n) = \left.\frac{(k+2n-2)!!}{(k-2)!!}\right|_{k=2} = (2n)!!. \tag{C.35}$$

□

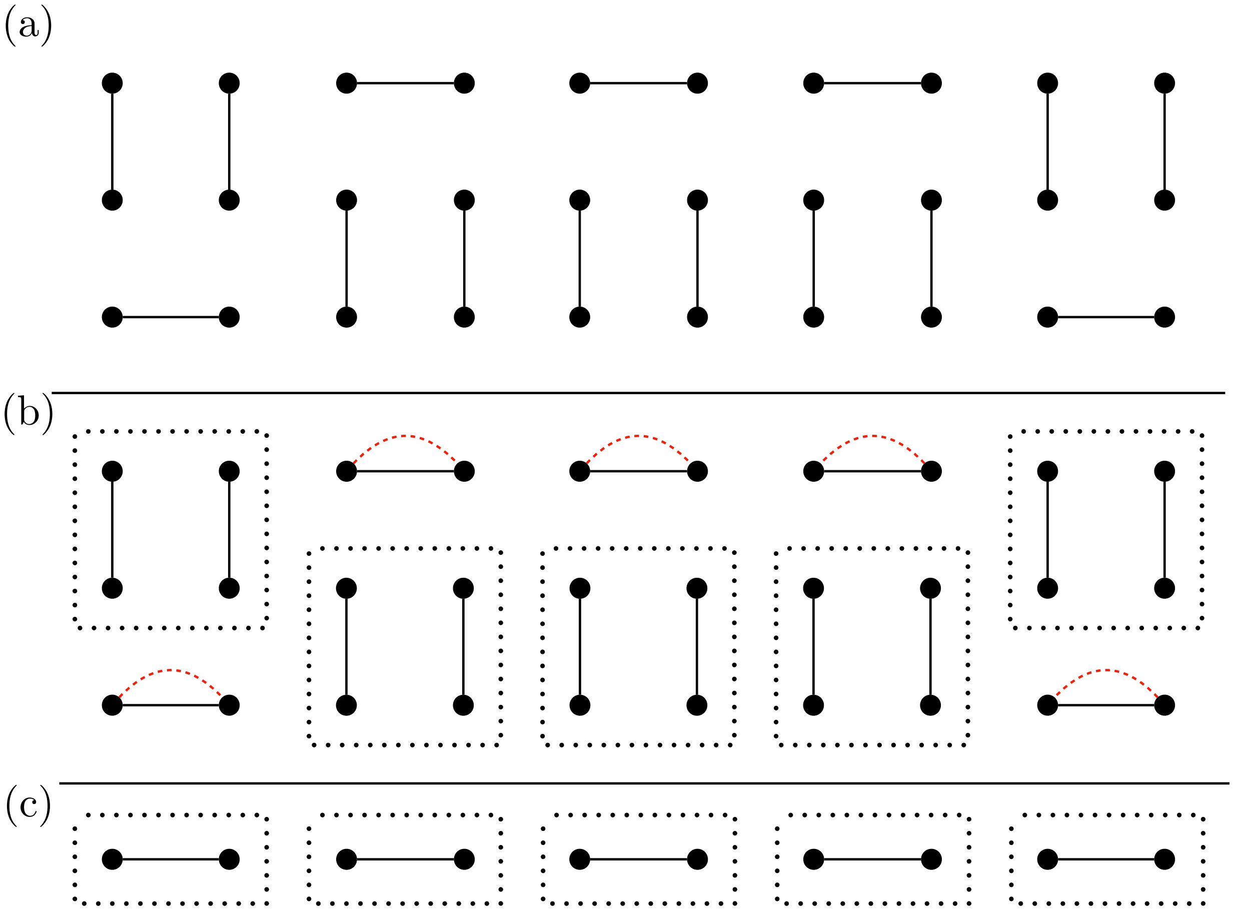

Figure C.3: Visualization of how the calculation of the coefficient of the leading-order term in the second moment can be reduced to the $k = 2$ case of the first moment. Recall that black edges are solid and red edges are dashed. (a) As proven in Theorem 4.2, graphs that maximize the number of connected components contain only type-1 and type-4 black edges. (b) To maximize the number of connected components, the horizontal black edges must form their own connected component with two vertices, meaning their vertices must be connected by a red edge. Furthermore, each vertical black edge must be paired off with exactly one other vertical black edge of the same type, forming a connected component with 4 vertices. We draw dotted boxes around the two black vertical edges to show that they come from the same type. (c) If we collapse each vertical edge onto a single vertex and then connect that vertex to the vertex stemming from its adjacent edge in the original graph (i.e. the other vertical edge from the same group of six vertices), then we reduce to the atomic graph (i.e., the graph with the fixed black edges, but without red edges) from the proof of the first moment. Red edges on this collapsed graph would then correspond to pairing off vertical edges in the original graph with red edges. Because paired edges in the original graph can only exist between edges of the same type, each connected component in the simplified graph could have come from either type-1 or type-4 vertical edges. This is equivalent to setting k, the contribution from each connected component, to 2, and then evaluating $f(2, n)$.

C.4 Approximate Hiding and Asymptotics of the First Moment

In this Appendix, we discuss more thoroughly the connection between hiding, the relevant sample space, and the first moment of squared hafnians of generalized COE matrices.

In Chapter 4, we introduce the normalized average outcome-collision probability as a measure of anticoncentration. Fixing the output state to have $2n$ photons, we write this as $|\Omega_{2n}|\mathbb{E}_{U\in U(m)}[\sum_{\boldsymbol{n}\in\Omega_{2n}} P_U(\boldsymbol{n})^2]$, where Ω_{2n} is the space of collision-free outcomes with $2n$ photons in m modes, and its size, which we write as $|\Omega_{2n}|$, is simply $\binom{m}{2n}$. We here work specifically with the non-collisional sample space because, in order for hiding to hold, collisions have to be negligible (a non-negligible likelihood of repeated columns in $U^{\top}_{1_k,\boldsymbol{n}}U_{1_k,\boldsymbol{n}}$ would prevent this distribution from being well approximated by $X^{\top}X$ with X Gaussian). And, indeed, when $n = o(\sqrt{m})$, it is easy to see that the size of the full sample space of $2n$ photons in m modes, $\binom{m+2n-1}{2n}$, approaches $|\Omega_{2n}| = \binom{m}{2n}$ when $n \gg 1$. In particular,

$$\frac{m^{2n}}{(2n)!} \leq \frac{(m+2n-1)!}{(m-1)!(2n)!} \leq \frac{(m+2n-1)^{2n}}{(2n)!} \tag{C.36}$$

and

$$\frac{\frac{(m+2n-1)^{2n}}{(2n)!}}{\frac{m^{2n}}{(2n)!}} = \left(1 + \frac{2n-1}{m}\right)^{2n} \xrightarrow{n\gg 1} 1 \tag{C.37}$$

because $(2n-1)/m = o(1/n)$. That is, Ω_{2n} is the dominant contribution to the full sample space.

We proceed to then replace $|\Omega_{2n}|$ with the expected value of the outcome probabilties, $\mathbb{E}_U[P_U(\boldsymbol{n})]$, that is, the first moment over input unitaries of a specific outcome. This holds assuming that the hiding property in Conjecture 4.1 holds. Roughly, hiding ensures that we do not preference any individual outcome, meaning we can replace the expected value over all probabilities with that over unitaries for a single probability. By linearity of expectation (and the fact that probabilities sum to unity), this expectation over unitaries should simply be the inverse of the size of the sample space of non-collisional outcomes. Finally, Conjecture 4.1 also gives us an approximate equality between $\mathbb{E}_U[P_U(\boldsymbol{n})]$ and M_1, the first moment of the squared hafnian of generalized COE matrices (properly rescaled to contain the correct prefactors).

We therefore now show that our calculation of the first moment in the hiding regime is consistent with the above discussion in the sense that $\mathbb{E}_U[P_U(\boldsymbol{n}|\sum_i n_i = 2n)] = \mathbb{E}_U[P_U(\boldsymbol{n})]/P(2n)$ is asymptotically equal to $|\Omega_{2n}|^{-1} = \binom{m}{2n}^{-1}$ assuming Conjecture 4.1. Here, $P(2n)$ is the probability that our output is in the $2n$-photon sector (i.e., the probability that Ω_{2n} is the proper sample space to consider in the first place).

Recall our input state has the first k of m modes prepared in the single-mode squeezed vacuum state with identical squeezing parameter r, and the remaining $m-k$ modes are prepared in the vacuum state. The probability of an outcome $\boldsymbol{n}$ is given by Eq. (4.1):

$$P_U(\boldsymbol{n}) = \frac{\tanh^{2n} r}{\cosh^k r}\left|\mathrm{Haf}(U_{1_k,\boldsymbol{n}}^{\top} U_{1_k,\boldsymbol{n}})\right|^2, \tag{C.38}$$

where $U_{1_k,\boldsymbol{n}}$ is the submatrix of U given by the first k rows and the columns dictated by where $\boldsymbol{n}$ is nonzero. Define $\tilde{U}_{1_k,\boldsymbol{n}} \coloneqq m U_{1_k,\boldsymbol{n}}$. Using multiplicativity of the Hafnian, one finds

$$P_U(\boldsymbol{n}) = \frac{\tanh^{2n} r}{\cosh^k r}\frac{1}{m^{2n}}\left|\mathrm{Haf}\left(\tilde{U}_{1_k,\boldsymbol{n}} \tilde{U}_{1_k,\boldsymbol{n}}^{\top}\right)\right|^2. \tag{C.39}$$

Assuming Conjecture 4.1, then $U_{1_k,\boldsymbol{n}}^{\top} U_{1_k,\boldsymbol{n}} \sim X^{\top} X$ where $X \sim \mathcal{N}(0, 1/m)_c^{k\times 2n}$, which means $\tilde{U}_{1_k,\boldsymbol{n}}^{\top} \tilde{U}_{1_k,\boldsymbol{n}} \sim X^{\top} X$, but now $X \sim \mathcal{N}(0,1)_c^{k\times 2n}$. Then, by Conjecture 4.1 and Theorem 4.2, we find

$$\mathop{\mathbb{E}}_{U\in\mathrm{U}(m)}\left[\left|\mathrm{Haf}\left(\tilde{U}^{\intercal}_{1_k,\boldsymbol{n}}\tilde{U}_{1_k,\boldsymbol{n}}\right)\right|^2\right] \approx \mathop{\mathbb{E}}_{X\in\mathcal{G}^{k\times 2n}}\left[\left|\mathrm{Haf}(X^{\intercal}X)\right|^2\right] = \frac{(2n)!}{2^n n!}\frac{(k+2n-2)!!}{(k-2)!!}, \tag{C.40}$$

where the first part of the equation is not an equality precisely because the hiding in Conjecture 4.1 is not exact. This implies that

$$\mathbb{E}_U[P_U(\boldsymbol{n})] \approx \frac{\tanh^{2n} r}{\cosh^k r}\frac{1}{m^{2n}}\frac{(2n)!}{2^n n!}\frac{(k+2n-2)!!}{(k-2)!!}. \tag{C.41}$$

Now, a single-mode squeezed vacuum state with squeezing parameter r and phase ϕ has Fock-state expansion given by

$$|\mathrm{SMSV}\rangle = \frac{1}{\sqrt{\cosh r}}\sum_{\ell=0}^{\infty}(-e^{i\phi}\tanh r)^{\ell}\frac{\sqrt{(2\ell)!}}{2^{\ell}\ell!}\,|2\ell\rangle\,. \tag{C.42}$$

Therefore, the probability of measuring 2ℓ photons is

$$|\langle 2\ell|\mathrm{SMSV}\rangle|^2 = \frac{\tanh^{2\ell r}}{\cosh r}\frac{(2\ell)!}{(2^{\ell}\ell!)^2}. \tag{C.43}$$

Given k independent single-mode squeezed vacuum states, the probability of finding $2n$ total photons is the k-fold convolution of the Fock-basis probability distribution of one single-mode squeezed vaccuum state:

$$P(2n) = \sum_{2\ell_1+\cdots+2\ell_k=2n}\prod_{i=1}^{k}\frac{\tanh^{2\ell_i r}}{\cosh r}\frac{(2\ell_i)!}{(2^{\ell_i}\ell_i!)^2} = \frac{\tanh^{2n} r}{\cosh^k r}\frac{1}{2^{2n}}\sum_{2\ell_1+\cdots+2\ell_k=2n}\prod_{i=1}^{k}\binom{2\ell_1}{\ell_1}. \tag{C.44}$$

This probability distribution is unchanged if the k independent single-mode squeezed vacuum states are acted upon by a linear-optical unitary before measurement (such a unitary does not change the photon number, only the location of the photons). The combinatorial identity at the core of this k-fold convolution has been calculated before in Refs. [184, 185]. Specifically,

$$\sum_{2\ell_1+\cdots+2\ell_k=2n}\prod_{i=1}^{k}\binom{2\ell_1}{\ell_1}=4^n\binom{n-1+k/2}{n}, \tag{C.45}$$

where we note that Eq. (C.45) holds even in the case where k is odd using a generalization of the binomial coefficients in terms of the Γ function.

The overall probability of finding $2n$ photons from k independent single-mode squeezed vacuum states, even after the application of a linear optical unitary, is therefore

$$P(2n)=\frac{\tanh^{2n}r}{\cosh^k r}\frac{1}{2^{2n}}4^n\binom{n-1+k/2}{n}=\frac{\tanh^{2n}r}{\cosh^k r}\binom{n-1+k/2}{n}. \tag{C.46}$$

We note that this expression, but not the full derivation, is also provided in Ref. [100]. A bit of algebraic manipulation reveals

$$\begin{aligned}P(2n)=\frac{\tanh^{2n}r}{\cosh^k r}\binom{n-1+k/2}{n}&=\frac{\tanh^{2n}r}{\cosh^k r}\frac{(2n-1)!!(k+2n-2)!!}{(2n)!(k-2)!!}\\&=\frac{1}{(2n)!}\frac{\tanh^{2n}r}{\cosh^k r}\mathop{\mathbb{E}}_{X\sim\mathcal{G}^{k\times 2n}}\left[\left|\mathrm{Haf}(X^\intercal X)\right|^2\right].\end{aligned} \tag{C.47}$$

According to Eq. (C.41), then

$$P(2n)\approx\frac{m^{2n}}{(2n)!}\mathbb{E}_U[P_U(\boldsymbol{n})], \tag{C.48}$$

which finally implies

$$\frac{\mathbb{E}_U[P_U(\boldsymbol{n})]}{P(2n)} \approx \frac{(2n)!}{m^{2n}} \approx \binom{m}{2n}^{-1} = |\Omega_{2n}|^{-1}, \tag{C.49}$$

where the first approximation is due to the fact that hiding is not exact, and the second approximation holds in the photon non-collisional regime.

C.5 Details on Definitions of Anticoncentration

In this Appendix, we discuss some of the details behind our definition of anticoncentration and how it relates to the standard notion of anticoncentration often used in the literature. We also discuss how these different definitions interact when it comes to showing anticoncentration holds for the exact distribution of the output probabilities of GBS given anticoncentration of the approximate distribution.

C.5.1 Definitions

We first discuss in somewhat more detail the relevance of anticoncentration to the argument for hardness of sampling from the output distribution of GBS. This argument makes use of an approximate counting algorithm due to Stockmeyer [34]. Roughly, we assume that there is an efficient sampling algorithm for GBS that, given a linear-optical unitary U, samples from a distribution Q_U which is close up to a constant $\epsilon > 0$ to the ideal GBS distribution P_U (recall Eq. (4.1)) in total-variation distance

$$\mathsf{tvd}(P_U, Q_U) \coloneqq \frac{1}{2}\sum_{\boldsymbol{n}} |P_U(\boldsymbol{n}) - Q_U(\boldsymbol{n})| \leq \epsilon. \tag{C.50}$$

Supposing such a sampling algorithm exists, and given the so-called hiding property (see Section D of Ref. [28] for details), we can use it as input to Stockmeyer's algorithm. Stockmeyer's algorithm then approximates the probability $P_U(\boldsymbol{n})$ up to an error given by

$$\varepsilon = \frac{1}{\mathrm{poly}(n)} P_U(\boldsymbol{n}) + \frac{2\epsilon}{|\Omega|\delta}\left(1 + \frac{1}{\mathrm{poly}(n)}\right), \tag{C.51}$$

with probability $1 - \delta$ over $\boldsymbol{n}$, where Ω is the sample space on which P_U is defined. If it is sufficiently hard (#P-hard, to be precise) to approximate the outcome probabilities $P_U(\boldsymbol{n})$ up to the error (C.51), on the instances on which our approximation scheme achieves this error, this rules out the approximate sampling algorithm up to very reasonable complexity-theoretic conjectures (one of which is the non-collapse of the polynomial hierarchy, a generalization of the famous $\mathsf{P} \neq \mathsf{NP}$ conjecture). The required property is thus what we call "approximate average-case hardness," that is, the statement that any algorithm which is able to compute $P_U(\boldsymbol{n})$ with probability $1 - \delta$ over the instances up to the error (C.51) is able to solve any #P-hard problem (of the same difficulty as approximating the outcome probabilities $P_U(\boldsymbol{n})$ up to the error (C.51)).

While we know average case hardness of approximating the outcome probabilities up to error $2^{-\Omega(n \log n)}$ [101, 186], it is only conjectured for the relevant approximation error given by either $c_1 P_U(\boldsymbol{n})$ or $c_2/|\Omega|$ for constants $c_1, c_2 > 0$. Anticoncentration serves as evidence for the truth of the conjecture, the idea being the following: suppose that most of the outcome probabilities are very close to zero, i.e. $\ll \epsilon/2^{-n}$, meaning only a vanishing fraction of them are relevant. Then a high approximation error on the relevant probabilities is tolerable, because we only need to distinguish between relevant and irrelevant outcomes, and a sufficiently good approximation to the irrelevant ones is zero. This is a significantly easier task than if the distribution is highly spread out and a large fraction of the probabilities is "relevant" in the sense that all of the relevant probabilities are of the same order of magnitude as the uniform distribution.

In the standard argument, this intuition is formalized as the statement

$$\Pr_{U \in \mathrm{U}(m)} \left[P_U(\boldsymbol{n}) \geq \frac{\alpha}{|\Omega|} \right] \geq \gamma(\alpha), \tag{C.52}$$

for some constants $\alpha, \gamma(\alpha) > 0$. In this formulation, we have made crucial use of the hiding property, which asserts that the distribution over circuits is invariant under a procedure by which we "hide" a particular outcome $\boldsymbol{n}$ in the probability of obtaining a different outcome $\boldsymbol{n}'$ of a random circuit. This allows us to restrict our attention to the distribution over circuits of a fixed outcome $\boldsymbol{n}$.

The anticoncentration property (C.52) implies that the error (C.51) is dominated by the first term on a $\gamma(\alpha)(1-\delta)$ fraction of the instances because with probability $\gamma(\alpha)$ we can upper bound the second term by $P_U(\boldsymbol{n})$. But, if a large fraction of the probabilities is larger than uniform, then none of them can be much larger than uniform and, hence, the approximation error needs to be exponentially small. Thus, we expect that, in the presence of anticoncentration, approximating the outcome probabilities up to the error (C.51) is much harder than without anticoncentration, lending credibility to the approximate average-case hardness conjecture.

In our definition of anticoncentration, we consider the (normalized) average collision probability

$$P_2(\mathrm{U}(m)) \coloneqq |\Omega| \sum_{\boldsymbol{n}\in\Omega} \mathbb{E}_{U\in\mathrm{U}(m)}\left[P_U(\boldsymbol{n})^2\right] \tag{C.53}$$

$$\overset{\text{hiding}}{=} |\Omega|^2 \mathbb{E}_{U\in\mathrm{U}(m)}\left[P_U(\boldsymbol{n})^2\right]. \tag{C.54}$$

The collision probability is the probability that, were one to sample the distribution twice, one would receive the same outcome both times. For very flat distributions it is very small. With the normalization, the collision probability of the uniform distribution is given by 1, which is its minimal value. On the other hand, the normalized collision probability of a fully peaked distribution with a single unit probability is given by $|\Omega|$.

The average collision probability is thus another measure of the anticoncentration of the outcome probabilities in the ensemble of linear-optical unitaries. It is a more coarse-grained measure, though, because it is only an average quantity. Indeed, a (constantly) small average collision probability implies anticoncentration in the sense of (C.52) via the Paley-Zygmund inequality as

$$\Pr_{U\in\mathrm{U}(m)}\left[P_U(\boldsymbol{n}) \geq \frac{\alpha}{|\Omega|}\right] \geq (1-\alpha)^2 \frac{1}{P_2(\mathrm{U}(m))}. \tag{C.55}$$

The relevant quantity of interest to anticoncentration is thus the inverse average collision probability $p_2(\mathrm{U}(m)) = 1/P_2(\mathrm{U}(m)$. Because by hiding the first moment $\mathbb{E}_U[P_U(\boldsymbol{n})]$ must evaluate to the inverse size of the sample space, we can rewrite p_2 for GBS as

$$p_2(\mathrm{U}(m)) = \frac{\mathbb{E}_{U\in\mathrm{U}(m)}[P_U(\boldsymbol{n})]^2}{\mathbb{E}_{U\in\mathrm{U}(m)}[P_U(\boldsymbol{n})^2]} \approx \frac{M_1(k,n)^2}{M_2(k,n)} = m_2(k,n). \tag{C.56}$$

In Chapter 4, we define various degrees of anticoncentration in terms of the inverse average collision probability p_2, which we recall here.

(A) We say that $P_U, U \in \mathrm{U}(m)$ *anticoncentrates* if $p_2 = \Omega(1)$.

(WA) We say that P_U *anticoncentrates weakly* if $p_2 = \Omega(1/n^a)$ for some $a = O(1)$.

(NA) And we say that it *does not anticoncentrate* if $p_2 = O(1/n^a)$ for any constant $a > 0$.

Here, we motivate those definitions in more detail. Clearly (A) implies anticoncentration in the sense of Eq. (C.52), hence the definition.

Lack of anticoncentration (NA) Ignoring the average over unitaries, p_2 upper-bounds the support of the distribution by $p_2|\Omega|$, as the maximum-entropy state is the uniform distribution. Let us assume for simplicity that p_2 is actually exponentially small. An exponentially small value of p_2 implies that the average support of the outcome distributions P_U is exponentially small, implying that at least a constant fraction (over U) of the distributions P_U has exponentially small support, and conversely exponentially larger than uniform probabilities on that support. At least for those distributions, this implies an exponentially larger error tolerance compared to $1/|\Omega|$. Such an exponentially larger error tolerance makes the approximate average-case hardness conjecture significantly stronger, presumably even untenable.

While it is possible that for a constant fraction of the U we are in this scenario (see Section V.C of Ref. [187] for an example), while for another constant fraction, the probabilities are highly spread out, making the anticoncentration property (C.52) true, this seems like an extremely unlikely state of affairs. Indeed, the hiding property implies that it should not matter whether we talk about the distribution over unitaries or over outcomes, which means that the situation described above is a generic feature, rendering (C.52) false in case p_2 is exponentially small.

Weak anticoncentration (WA) Our results show that weak anticoncentration holds in the regime of $k \to \infty$. But why do we think of a polynomially decaying p_2 as *weak* anticoncentration rather than no anticoncentration?

We argue that this is a meaningful regime in the sense that there is a stronger—but not inconceivable—approximate average-case hardness conjecture associated with the weak anticoncentration regime. To see this, observe that weak anticoncentration implies anticoncentration in the sense of Eq. (C.52) with $\gamma(\alpha) = \Omega(1/\text{poly}(n))$, which means that an inverse polynomial fraction of the outcome probabilities are larger than uniform. Technically, using Stockmeyer's algorithm we can thus achieve a multiplicative error for an inverse polynomial fraction of the outcome probabilities. To rule out an efficient classical sampler, we thus need to conjecture approximate average-case hardness with constant relative errors for any inverse polynomial fraction of the instances. Equivalently, we can formulate a similar conjecture for a polynomially large relative or subexponentially large additive error on a constant fraction. While clearly much stronger than the requirement of anticoncentration, this is qualitatively different from the lack of anticoncentration scenario (NA), where the difference is superpolynomial.

C.5.2 Anticoncentration of the Exact Distribution

We also need to show that our definition of anticoncentration allows us to translate between anticoncentration of the approximate distribution based on the hafnians of random Gaussian matrices, which we will refer to as $P_X(\boldsymbol{n})$, and anticoncentration of the true distribution, $P_U(\boldsymbol{n})$. For a given output $\boldsymbol{n}$, let $\mathcal{D}_U$ be the distribution of the symmetric product $U^\intercal_{1_k,\boldsymbol{n}} U_{1_k,\boldsymbol{n}}$ with $U \in \mathrm{U}(m)$. Let $\mathcal{D}_X$ be the distribution of the symmetric product $X^\intercal X$ with $X \sim \mathcal{N}(0, 1/m)_c^{k\times 2n}$. In Conjecture 4.1, we conjecture that $\mathcal{D}_U$ and $\mathcal{D}_X$ become close in total variation distance when $n = o(\sqrt{m})$. However, precisely how close these two distributions are is crucial to whether or not anticoncentration translates between the two output probabily distributions. In what follows, we will refer to anticoncentration in the sense of Eq. (C.52) as "standard" anticoncentration, and our definition of anticoncentration as "moment-based."

Ideally, we would be able to prove that statements about moment-based anticoncentration of $P_X(\boldsymbol{n})$ imply equivalent statements about moment-based anticoncentration of $P_U(\boldsymbol{n})$. However, under worst-case assumptions, we can only show that moment-based anticoncentration of $P_X(\boldsymbol{n})$ implies standard anticoncentration of $P_U(\boldsymbol{n})$. To understand this, let us fix some notation. Let also $\mathbb{1}[\cdot]$ be an indicator function which is 1 if the argument is true and 0 if it is false. Let $\mathrm{d}\mu$ be the Lebesgue measure on $\mathbb{C}^{2kn}$ (as we consider $k \times 2n$ complex matrices) and $p_U(A)$, $p_X(A)$ be the respective probabilities of generating A from $\mathcal{D}_U$ and $\mathcal{D}_X$.

Now, let the total-variation distance between $\mathcal{D}_U$ and $\mathcal{D}_X$ be less than δ. Then

$$\Pr_{U\in\mathrm{U}(m)}[P_U(\boldsymbol{n}) \geq \epsilon] = \int \mathrm{d}\mu\, p_U(A)\mathbb{1}[P_A(\boldsymbol{n}) \geq \epsilon] \tag{C.57}$$

$$= \int \mathrm{d}\mu\, (p_U(A) - p_X(A) + p_X(A))\mathbb{1}[P_A(\boldsymbol{n}) \geq \epsilon] \tag{C.58}$$

$$= \int \mathrm{d}\mu\, (p_U(A) - p_X(A))\mathbb{1}[P_A(\boldsymbol{n}) \geq \epsilon] + \int \mathrm{d}\mu\, p_X(A)\mathbb{1}[P_A(\boldsymbol{n}) \geq \epsilon] \tag{C.59}$$

$$\geq -2\delta + \Pr_{X\in\mathcal{G}}[P_X(\boldsymbol{n}) \geq \epsilon]. \tag{C.60}$$

In this calculation, we have used the Radon-Nikodym theorem [188] to express the probability measures that define $\mathcal{D}_U$ and $\mathcal{D}_X$ as $p_U(A)\mathrm{d}\mu$ and $p_X(A)\mathrm{d}\mu$, respectively. Therefore

$$\Pr_{U\in\mathrm{U}(m)}\left[P_U(\boldsymbol{n}) \geq \frac{\alpha}{|\Omega_{2n}|}\right] \geq \Pr_{X\in\mathcal{G}}\left[P_X(\boldsymbol{n}) \geq \frac{\alpha}{|\Omega_{2n}|}\right] - 2\delta \geq (1-\alpha)^2 \frac{1}{m_2(k,n)} - 2\delta. \tag{C.61}$$

The final step follows from the Paley-Zygmund inequality for the approximate distribution. This proves that we can translate statements on anticoncentration as long as 2δ is smaller than $m_2(k,n)^{-1}$, which, as we show in Chapter 4, means $\delta = o(n^{-1/2})$.

With this in mind, we can make the following more precise version of Conjecture 4.1 such that, if it holds, moment-based weak anticoncentration of the approximate distribution implies standard weak anticoncentration of the exact distribution:

Conjecture C.1 (Formal)**.** *Let $\mathcal{D}_U$ be the distribution of the symmetric product $U^\top_{1_k,\boldsymbol{n}} U_{1_k,\boldsymbol{n}}$ with U unitary and $\boldsymbol{n}$ some non-collisional outcome of a Gaussian Boson Sampling experiment. Let $\mathcal{D}_X$ be the distribution of the symmetric product $X^\top X$ with $X \sim \mathcal{N}(0,1/m)_c^{k\times 2n}$. Then, for any k such that $1 \leq k \leq m$, and for any $\delta > 0$ such that $m \geq n^2/\delta$,*

$$\mathsf{tvd}(\mathcal{D}_U, \mathcal{D}_X) = O(\delta). \tag{C.62}$$

Specifically, if $\delta = o(n^{-1/2})$, then $m \geq n^{5/2}$.

The motivation behind the choice of $m \geq n^2/\delta$ is based on the equivalent conjecture for Fock Boson Sampling in Ref. [27]. There, the authors are able to prove the equivalent result for $m \geq n^{5+\epsilon}/\delta$ (for arbitrarily small, constant ϵ), but they suspect that the result can be pushed further to $m \geq n^2/\delta$. We note that this choice makes our formal conjecture slightly stronger than the equivalent formal conjecture in Ref. [101].

As we have shown, in order to translate our results on moment-based weak anticoncentration from the approximate to the true distribution in the worst case, we require $\delta = o(n^{-1/2})$. Therefore, in order to translate statements about anticoncentration, the formal version of our conjecture requires $m \geq n^{5/2}$.

However, it is worth noting that we do not believe that this worst-case scenario truly reflects the way in which $P_X(\boldsymbol{n})$ approaches $P_U(\boldsymbol{n})$, i.e., where all of the error is concentrated on a single probability. In general, the intuition is that if hiding holds, then it is more likely that the errors are more evenly distributed amongst all of the exponentially many output probabilities. Using this intuition, each individual probability only receives an error of approximately $\delta/|\Omega_{2n}|$. If this is true, then we can show that moment-based weak anticoncentration of $P_X(\boldsymbol{n})$ does actually imply the same for $P_U(\boldsymbol{n})$. Specifically, say that $P_U(\boldsymbol{n}) \approx P_X(\boldsymbol{n}) \pm \delta/|\Omega_{2n}| \approx P_X(\boldsymbol{n}) \pm \delta\mathbb{E}[P_X(\boldsymbol{n})]$ (as per Appendix C.4). Then

$$\begin{aligned}
\frac{\mathbb{E}[P_U(\boldsymbol{n})^2]}{(\mathbb{E}[P_U(\boldsymbol{n})])^2} &\approx \frac{\mathbb{E}[(P_X(\boldsymbol{n}) \pm \delta\mathbb{E}[P_X(\boldsymbol{n})])^2]}{(\mathbb{E}[P_X(\boldsymbol{n}) \pm \delta\mathbb{E}[P_X(\boldsymbol{n})]])^2} && \text{(C.63)}\\
&= \frac{\mathbb{E}[P_X(\boldsymbol{n})^2] \pm 2\delta\mathbb{E}[P_X(\boldsymbol{n})]^2 + \delta^2\mathbb{E}[P_X(\boldsymbol{n})]^2}{(1 \pm \delta)^2\mathbb{E}[P_X(\boldsymbol{n})]^2} && \text{(C.64)}\\
&\approx \frac{1}{(1 \pm \delta)^2}\frac{\mathbb{E}[P_X(\boldsymbol{n})^2]}{(\mathbb{E}[P_X(\boldsymbol{n})])^2} + \frac{\pm 2\delta + \delta^2}{(1 \pm \delta)^2} && \text{(C.65)}\\
&= \frac{1}{(1 \pm \delta)^2}\frac{\mathbb{E}[P_X(\boldsymbol{n})^2]}{(\mathbb{E}[P_X(\boldsymbol{n})])^2} + 1 - \frac{1}{(1 \pm \delta)^2} && \text{(C.66)}\\
&\leq \frac{1}{(1 - \delta)^2}\frac{\mathbb{E}[P_X(\boldsymbol{n})^2]}{(\mathbb{E}[P_X(\boldsymbol{n})])^2} + 1. && \text{(C.67)}
\end{aligned}$$

In our case, where the normalized second moment of $P_X(\boldsymbol{n})$ scales at least polynomially in n, and δ scales inverse polynomially in n, weak anticoncentration or lack of anticoncentration of

$P_X(\boldsymbol{n})$ in terms of the normalized second moment adequately translates to $P_U(\boldsymbol{n})$ as well. Note that, in this case, we are assuming that δ, which is the total variation distance between the distributions of matrices, extends to a bound on the total variation distance between the probabilities themselves. This intuitively arises from the fact that any map from the distribution of the matrices to probabilities must be bounded, meaning we can translate the total variation distance from one to the other (however, formalizing this would require dealing with some subtleties induced by the fact that the hafnian of a product of Gaussians is not technically bounded, but any large hafnians only arise with extremely small probabilities).

C.6 Scattershot Boson Sampling Explanation of the Transition in Anticoncentration

In Scattershot Boson Sampling (SBS), the setup is as follows. $m = \omega(n^2)$ two-mode squeezed states with squeezing parameter r are prepared. The photon number distribution of the two-mode squeezed states is supported on Fock states of the form $|n\rangle |n\rangle$ for $n \in \mathbb{N}_0$. One half of each two-mode squeezed state is then measured in the Fock basis, yielding, with high probability, an outcome $n_i \in \{0, 1\}$ (assuming r is small enough). Collecting outcomes in the vector $\boldsymbol{n} = (n_1, \dots, n_m)$, the other half of the input modes is now in the postselected state $|\boldsymbol{n}\rangle = \otimes_{i=1}^m |n_i\rangle$. The outcome probabilities after passing this input state through the linear optical unitary U and measuring in the Fock basis yielding outcome $\boldsymbol{o} = (o_1, \dots, o_m)$, $o_i \in \{0, 1\}$ is then given by

$$P_U(\boldsymbol{n}, \boldsymbol{o}) = |\mathrm{Per}(U_{\boldsymbol{n},\boldsymbol{o}})|^2 \tag{C.68}$$

of the submatrix $U_{\boldsymbol{n},\boldsymbol{o}}$ in which we select the rows and columns according to the indices with nonzero entries in $\boldsymbol{n}$ and $\boldsymbol{o}$. But conditioned on input and output states being collision-free and the hiding property, the distribution of matrices $U_{\boldsymbol{n},\boldsymbol{o}}$ equals that of the Boson Sampling submatrices $U_{\boldsymbol{1}_n,\boldsymbol{o}}$, where the photons in the input state are by convention in the first n modes. The properties of Scattershot Boson Sampling postselected on collision-free outcomes in a fixed photon number sector are therefore equal to the properties of standard Boson Sampling.

We now argue that this equivalency hinges essentially on the fact that at least $\omega(n^2)$ of the input modes are squeezed. To this end, consider a modification of Scattershot Boson Sampling in which only k out of the m modes are prepared in one half of a two-mode squeezed state, while the remaining $m - k$ modes are prepared in the vacuum state. This closely resembles the GBS setting, of course. Let us also consider a squeezing parameter r of every two-mode squeezed state chosen such that the mean photon number after postselection is given by n. To achieve this, we pick the mean photon number per mode, which is given by $\sinh^2(r)$ to be equal n/k to obtain a total of $k\sinh^2 r = n$ photons on average. This ensures that in the postselection we end up with n photons with high probability.

Recall that a two-mode squeezed vacuum state with squeezing parameter r and phase ϕ has a Fock expansion given by

$$|\mathrm{TMSV}\rangle = \frac{1}{\cosh(r)} \sum_{\ell=0}^{\infty} (-e^{i\phi} \tanh r)^n |nn\rangle , \tag{C.69}$$

thus leading to a probability of measuring ℓ photons in one mode of $\tanh^{2\ell} r/\cosh^2 r$. Therefore, if the input consists of k two-mode squeezed vacuum states, then the probability that, after measuring one half of each state, one observes a collision is

$$\Pr[\text{collision}] = 1-\left(\frac{1}{\cosh^2 r} + \frac{\tanh^2 r}{\cosh^2 r}\right)^k = 1-\left(\frac{1}{1+n/k} + \frac{n/k}{(1+n/k)^2}\right)^k = 1-\left(1-\left(\frac{n/k}{1+n/k}\right)^2\right)^k . \tag{C.70}$$

We can rewrite this via Taylor series as

$$\Pr[\text{collision}] = 1 - \exp\left(-\frac{n^2}{k} + kO(n/k)^3)\right) \tag{C.71}$$

assuming $k = \omega(n)$. This collision probability remains lower bounded by a constant for $k = O(n^2)$, but vanishes for any $k = \omega(n^2)$. Thus, the probability of a collision in the input state of SBS remains high until $k = \Theta(n^2)$ and decays then. But because in SBS the roles of the (postselected) input state and the output state are symmetric, a collision implies a failure of hiding and, therefore, a failure of anticoncentration in the regime $k = O(n^2)$. Conversely, for $k = \omega(n^2)$ we believe that hiding holds [27, 108], and hence Lemma 8.8 of Aaronson and Arkhipov [27] shows weak anticoncentration for SBS with the inverse average collision probability $p_2 = 1/n$ in this regime.

This shows that generalized SBS with a variable number of input squeezed states undergoes a transition in anticoncentration as we find it here for the case of GBS. It is not at all clear that the transition in SBS implies a transition in GBS, however, as GBS does not involve postselection. Indeed, in SBS, the anticoncentration coincides with—or rather *is*—a transition in the hiding property. In GBS, in contrast, hiding is conjectured to hold for all k, while we do see the transition in anticoncentration. The situation in GBS is not immediately comparable to that in this modified SBS scenario because the input single-mode squeezed states are supported on even numbers of photons, and therefore any nonzero photon number input states are collision-full. Therefore, as mentioned in the discussion in Chapter 4, the possible connections outlined here deserve future consideration.

Appendix D: Appendices Associated with Chapter 5

In the Appendices, we provide details and derivations that supplement the discussion in Chapter 5.

- Appendix D.1: We discuss the classical complexity of evaluating the recursion and show that it is efficient (i.e., the time and space required scale polynomially) in the Fock sector n;
- Appendix D.2: We provide the graph-theoretic details for how to derive the recursion;
- Appendix D.3: We discuss how to compute individual coefficients of the polynomial expansion of the second moment. Specifically, we give one method to calculate the leading and first subleading terms in the polynomial expansion of the second moment;
- Appendix D.4: We discuss an alternative method for developing a recursion to the solve for the second moment. We also apply this alternative picture to find an expression for the constant term in the polynomial expansion of the second moment.

D.1 Classical Complexity of Evaluating the Recursion

In this Appendix, we argue that the numerical evaluation of the recursion and, hence, the second moment, is classically efficient (that is, the runtime and space used are at most polynomial) in n, which corresponds to the Fock sector of interest in the output samples.

We recall the setup of the recursion as we describe it in Chapter 5. Specifically, we define

$$g(n, a_{12}, a_{13}, a_{23}) := \sum_{\lambda \in \mathbb{G}_n^2(a_{12}, a_{13}, a_{23})} k^{C(\lambda)}. \tag{D.1}$$

$\mathbb{G}_n^2(a_{12}, a_{13}, a_{23})$ is the set of second-moment graphs of order n with a_{ij} red edges that cross between rows i and j. $C(\lambda)$ is the number of connected components of λ. The second moment is given by $(2n-1)!!g(n,0,0,0)$. We then write down the recursion using these $g(n, a_{12}, a_{13}, a_{23})$ as

$$g(n, a_{12}, a_{13}, a_{23}) = \sum_{b_{12}, b_{13}, b_{23}} c(a_{12}, a_{13}, a_{23}, b_{12}, b_{13}, b_{23}) g(n-1, b_{12}, b_{13}, b_{23}). \tag{D.2}$$

We list the following constraints on $\boldsymbol{a}$, which is shorthand for (a_{12}, a_{13}, a_{23}). First, $a_{12}+a_{13}$, $a_{12}+a_{23}$, and $a_{13}+a_{23}$ (the edges that exit the first, second, and third rows respectively) must be even. Second, $a_{12}+a_{13}, a_{12}+a_{23}, a_{13}+a_{23}$ must all be less than or equal to $2n$, as there cannot be more than $2n$ edges coming out of a row with only $2n$ vertices given that there is exactly one red edge incident on every vertex. Finally, we also add here that, clearly, a_{12}, a_{13}, a_{23} are non-negative. These constraints imply a finite number of valid vectors $\boldsymbol{a} = (a_{12}, a_{13}, a_{23})$ for a given order n, and any vector satisfying these constraints corresponds to a valid set of graphs and, therefore, a term $g(n, \boldsymbol{a})$ in the recursion. We provide an example of all possible $\boldsymbol{a}$ when $n = 4$ in Table D.1.

Clearly, as n grows, the number of possible $\boldsymbol{a}$ for which one must evaluate $g(n, \boldsymbol{a})$ also grows. However, we can bound this growth as being polynomial in n using some arguments about partitions. Recall that a partition of a positive integer m of size s is a set (i.e., order does

m	$\boldsymbol{a}$
0	$(0,0,0)$
1	$\varnothing$
2	$(2,0,0)$
3	$(1,1,1)$
4	$(2,2,0),(4,0,0)$
5	$(3,1,1)$
6	$(6,0,0),(4,2,0),(2,2,2)$
7	$(5,1,1),(3,3,1)$
8	$(8,0,0),(6,2,0),(4,4,0),(4,2,2)$
9	$(7,1,1),(5,3,1),(3,3,3)$
10	$(6,2,2),(4,2,2)$
11	$(5,3,3)$
12	$(4,4,4)$

Table D.1: All possible $\boldsymbol{a}$, up to permutations of the vector elements, for $2n = 8$. Each entry satisfies the constraints that $a_{12} + a_{13}$, $a_{12} + a_{23}$, and $a_{13} + a_{23}$ are even and less than or equal to $2n$, a_{12}, a_{13}, and a_{23} are non-negative, and $a_{12} + a_{13} + a_{23} = m$.

not matter) of s positive integers whose sum is m. A weak partition of m of size s relaxes the positivity constraint of the set such that it contains s non-negative elements (m is still positive).

Let $m \coloneqq a_{12} + a_{13} + a_{23}$. Then $m \leq 3n$, which follows from the fact that

$$2a_{12} + 2a_{13} + 2a_{23} = (a_{12} + a_{13}) + (a_{12} + a_{23}) + (a_{13} + a_{23}) \leq 6n. \tag{D.3}$$

The conditions listed above on $\boldsymbol{a}$ imply that each $\boldsymbol{a}$ is a weak partition of size 3 of $m \leq 3n$ that satisfies two further constraints: all 3 elements of the set must have the same parity as m, and no element can be larger than $2n$.

Now, the number of partitions of m of size at most 3 is $\lfloor (m+3)^2/12 \rceil$ [189] (note that $\lfloor M \rceil$ refers to the closest integer to M). Therefore, the number of partitions of m of size exactly 3, or $p_3(m)$, is bounded by this value, which implies that $\sum_{m=0}^{3n} p_3(m) = O(n^3)$. In turn, the number of $\boldsymbol{a}$, up to permutations of the elements of $\boldsymbol{a}$, is bounded by $O(n^3)$ (because they form an even more restricted class of weak permutations). We can overcount for these permutations with a simple constant multiplicative factor of $3!$ (this overcounts because, when numbers are repeated in the partition, there are fewer distinct permutations). Thus, we have a polynomial bound on the number of terms in our recursion at any Fock sector n (note that we could tighten this bound a bit by accounting more precisely for the parity constraint on the elements $\boldsymbol{a}$, but, because we are interested only in classical efficiency, this polynomial bound that arises from considering only size-3 partitions is sufficient).

To be sure that the recursion is efficiently computable, however, the actual values of the terms in the recursion must not grow too quickly. In particular, recall that each term $g(n, \boldsymbol{a})$ has a polynomial expansion in k of order at most $3n$ (this is the largest number of connected components possible when each one must have at least 2 vertices). The sum of the coefficients of $g(n, \boldsymbol{a})$ is the same as the number of graphs in $\mathbb{G}_n^2(a_{12}, a_{13}, a_{23})$, which we derived to be

$$\begin{aligned}|\mathbb{G}_n^2(a_{12}, a_{13}, a_{23})| = \binom{2n}{a_{12}}\binom{2n-a_{12}}{a_{13}}\binom{2n}{a_{12}}\binom{2n-a_{12}}{a_{23}}\binom{2n}{a_{13}}\binom{2n-a_{13}}{a_{23}}a_{12}!a_{13}!a_{23}! \\ \times (2n-a_{12}-a_{13}-1)!!(2n-a_{12}-a_{23}-1)!!(2n-a_{13}-a_{23}-1)!!4^n. \quad \text{(D.4)}\end{aligned}$$

This is, at most, factorially big in n, which means that the number of bits needed to store these numbers, and, hence, $g(n, \boldsymbol{a})$ is polynomial in n.

Therefore, we have a polynomial bound on the number of terms in the recursion, as well as on the space needed to represent each of these terms. Finally, because the actual recursion consists only of polynomial numbers of multiplication and addition, which can each be accomplished in time polynomial in the size of the inputs, the actual computation is efficient.

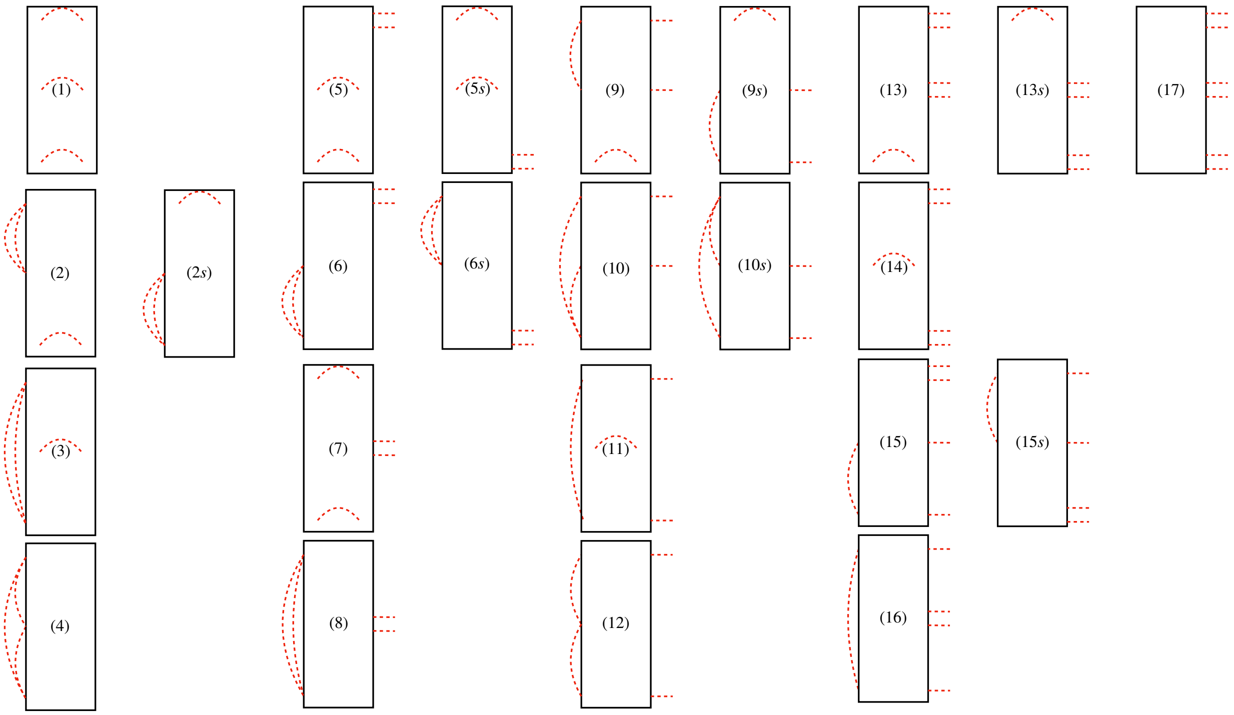

Figure D.1: Copy of Fig. 5.4. List of 17 cases (up to symmetry) for how the first two columns in a graph of order n can connect into the rest of the graph.

D.2 Building the Recursion

We now describe precisely how to derive and evaluate the recursion relation Eq. (5.25), which we copy again here for convenience:

$$g(n, a_{12}, a_{13}, a_{23}) = \sum_{b_{12}, b_{13}, b_{23}} c(a_{12}, a_{13}, a_{23}, b_{12}, b_{13}, b_{23}) g(n-1, b_{12}, b_{13}, b_{23}). \tag{D.5}$$

We note that we implement the full recursion [115] in both the Julia programming language [114] and Mathematica [190]. Recall that $g(n, a_{12}, a_{13}, a_{23})$ is a polynomial in k where the coefficient in front of k^i is the number of graphs of type $\boldsymbol{a} = (a_{12}, a_{13}, a_{23})$ that have i connected components. Again, a graph of type $\boldsymbol{a}$ has a_{ij} edges that go between rows i, j.

We first describe the base case, i.e. $g(1, a_{12}, a_{13}, a_{23})$ for all valid vectors $\boldsymbol{a} = (a_{12}, a_{13}, a_{23})$.

We then describe how to handle each of the possible 17 cases that contribute to the recursion that are depicted in Fig. 5.4, which is copied again here for convenience.

The way that we handle each case is as follows. We consider all graphs of order n such that the leftmost two columns, which, recall, we refer to as $\mathbb{C}_{1,2}$, have red edges that correspond to that case. We then "integrate out" these edges to determine how to write the contribution of that case at order n in terms of the terms at order $n - 1$. When we say integrate out, we mean that we collapse any path that goes through $\mathbb{C}_{1,2}$ into a new edge that remains entirely in the graph of order $n - 1$ by collapsing together vertices connected by these paths. In doing this, we must account for three main contributions: (1) how many loops are contained solely within $\mathbb{C}_{1,2}$—each of these loops, of course, leads to a factor of k multiplied by the contribution at order $n - 1$; (2) what edges are erased when integrating out the case, as well as what edges are created after collapsing the paths into new edges—this tells us what $\boldsymbol{b}$ at lower order contribute to $\boldsymbol{a}$ at a higher order; (3) a combinatorial factor accounting for the fact that integrating out $\mathbb{C}_{1,2}$ in multiple graphs at order n could lead to the same graph at order $n - 1$, meaning we may need to multiply the contributions at order $n - 1$ by something to get the correct final answer. The former loop calculation is usually quite simple, but the latter vectorial and combinatorial calculations require more significant casework.

In the abstract, this is quite complicated, but we explain it more thoroughly through detailed examples as we proceed. We group our analysis of these cases into four categories corresponding to the number of edges, i.e. 0, 2, 4, or 6, that protrude from the cases: (1)–(4), (5)–(12), (13)–(16), and (17), respectively. However, as mentioned, we begin with the base cases, to which we turn now.

D.2.1 Base Cases for Recursion

Here we calculate the base cases for the recursion; that is, we determine all valid $\boldsymbol{a}$ when $n = 1$, construct all graphs with each $\boldsymbol{a}$, and count their connected components. Recall that the vector $\boldsymbol{a}$ must satisfy non-negativity, pairwise sums being even, and pairwise sums being at most $2n$; should any one of these conditions not be met, then $g(n, \boldsymbol{a}) = g(n, a_{12}, a_{13}, a_{23}) = 0$. For $n = 1$, there are 5 possible options for $\boldsymbol{a}$: $(0,0,0)$, $(2,0,0)$, $(0,2,0)$, $(0,0,2)$, $(1,1,1)$. It remains then to construct the graphs and count their connected components. This is tedious, but the diagrams are shown in Figs. D.2 and D.3, and the final results are

$$g(1,0,0,0) = 2k^2 + 2k, \tag{D.6}$$

$$g(1,2,0,0) = k^3 + 3k^2 + 4k, \tag{D.7}$$

$$g(1,0,0,2) = k^3 + 3k^2 + 4k, \tag{D.8}$$

$$g(1,0,2,0) = 2k^2 + 6k, \tag{D.9}$$

$$g(1,1,1,1) = 2k^3 + 14k^2 + 16k. \tag{D.10}$$

This completes the base cases, and we now move on to the recursion.

D.2.2 Cases (1)–(4)

We now handle cases (1)–(4). There are no protruding edges, meaning many of the contributions are easy to derive because these cases are “independent” of from the lower order graph consisting of the final $n - 1$ pairs columns. Therefore, when we integrate out $\mathbb{C}_{1,2}$, none of the paths affect the graph at lower order, meaning it is much simpler to calculate their contribution.

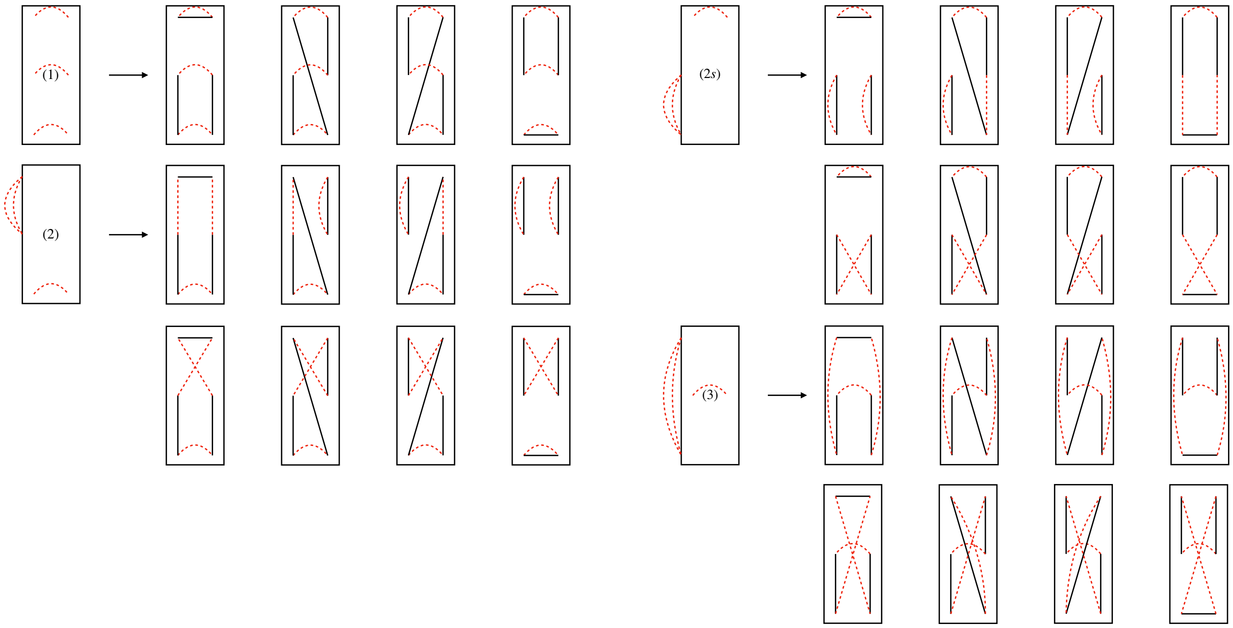

Figure D.2: Base cases corresponding to (1), (2), $(2s)$, and (3). Counting the connected components of the graphs in each case yields contributions of $2k^2 + 2k$, $k^3 + 3k^2 + 4k$, $k^3 + 3k^2 + 4k$, and $2k^2 + 6k$, respectively.

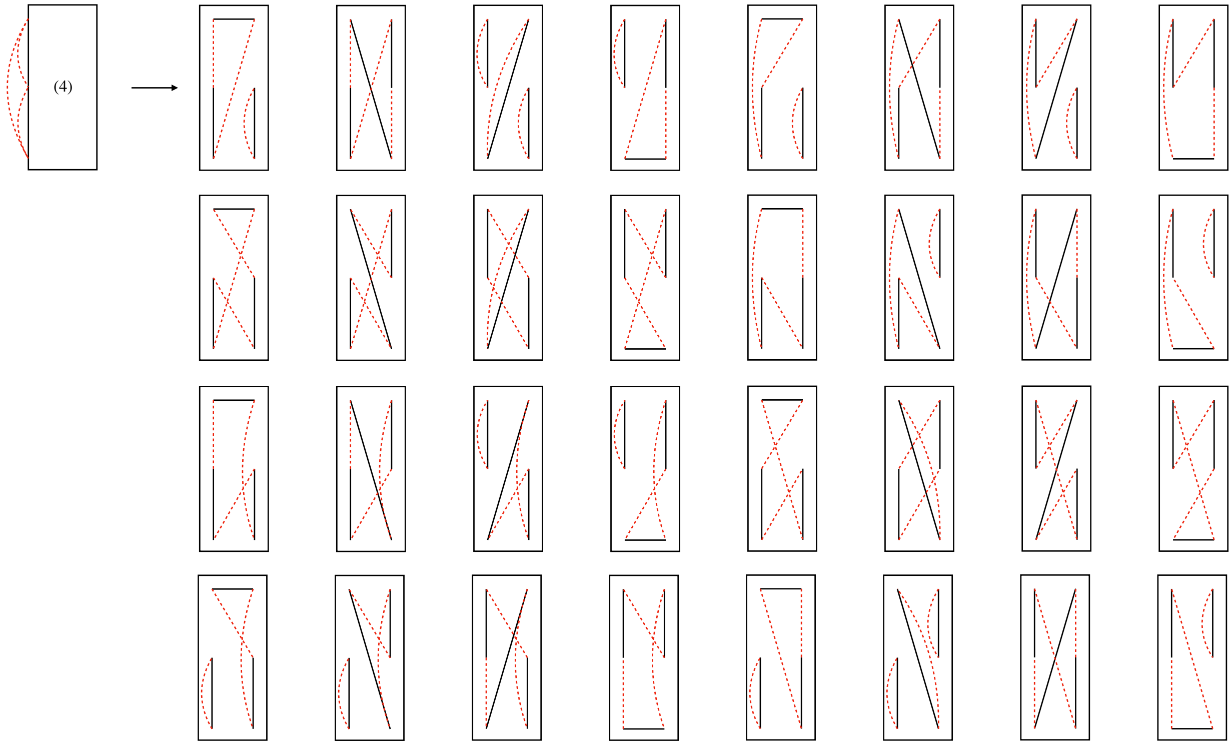

Figure D.3: Base case corresponding to (4). Counting the connected components of the graphs in each case yields $2k^3 + 14k^2 + 16k$.

In fact, it is simple to see that the evaluation of the loops mimics exactly the calculation of the base cases:

$$\text{Loop }(1) \to 2k^2 + 2k, \tag{D.11}$$

$$\text{Loop }(2) \to k^3 + 3k^2 + 4k, \tag{D.12}$$

$$\text{Loop }(2s) \to k^3 + 3k^2 + 4k, \tag{D.13}$$

$$\text{Loop }(3) \to 2k^2 + 6k, \tag{D.14}$$

$$\text{Loop }(4) \to 2k^3 + 14k^2 + 16k. \tag{D.15}$$

(D.16)

Next, examining the diagrams for each case, one can derive simple relationships between $\boldsymbol{a}$ and $\boldsymbol{b}$ that yield a nontrivial contribution in Eq. (D.5):

$$\text{Vector }(1) \to (b_{12}, b_{13}, b_{23}) = (a_{12}, a_{13}, a_{23}), \tag{D.17}$$

$$\text{Vector }(2) \to (b_{12}, b_{13}, b_{23}) = (a_{12} - 2, a_{13}, a_{23}), \tag{D.18}$$

$$\text{Vector }(2s) \to (b_{12}, b_{13}, b_{23}) = (a_{12}, a_{13}, a_{23} - 2), \tag{D.19}$$

$$\text{Vector }(3) \to (b_{12}, b_{13}, b_{23}) = (a_{12}, a_{13} - 2, a_{23}), \tag{D.20}$$

$$\text{Vector }(4) \to (b_{12}, b_{13}, b_{23}) = (a_{12} - 1, a_{13} - 1, a_{23} - 1). \tag{D.21}$$

(D.22)

These can be understood by looking at the diagram for each case and observing what kind of edges are eliminated when collapsing all of the paths that pass through the vertices in $\mathbb{C}_{1,2}$.

Finally, there are no combinatorial contributions because there are no protruding edges that have to be connected to the existing graph. That is, any graph that comes from integrating out one of these cases arises uniquely.

Therefore, we can easily combine everything to get the contributions to the recursion from each of these cases:

$$g(n, a_{12}, a_{13}, a_{23})_{\text{case}(1)} = (2k^2 + 2k)g(n-1, a_{12}, a_{13}, a_{23}), \tag{D.23}$$

$$g(n, a_{12}, a_{13}, a_{23})_{\text{case}(2)} = (k^3 + 3k^2 + 4k)g(n-1, a_{12}-2, a_{13}, a_{23}), \tag{D.24}$$

$$g(n, a_{12}, a_{13}, a_{23})_{\text{case}(2s)} = (k^3 + 3k^2 + 4k)g(n-1, a_{12}, a_{13}, a_{23}-2), \tag{D.25}$$

$$g(n, a_{12}, a_{13}, a_{23})_{\text{case}(3)} = (2k^2 + 6k)g(n-1, a_{12}, a_{13}-2, a_{23}), \tag{D.26}$$

$$g(n, a_{12}, a_{13}, a_{23})_{\text{case}(4)} = (2k^3 + 14k^2 + 16k)g(n-1, a_{12}-1, a_{13}-1, a_{23}-1). \tag{D.27}$$

Note that we have introduced a notation $g(n, a_{12}, a_{13}, a_{23})_{\text{case}(i)}$, which simply refers to the contribution to $g(n, a_{12}, a_{13}, a_{23})$ from graphs where the vertices in $\mathbb{C}_{1,2}$ and their corresponding red edges fall into case (i). That is, $g(n, \boldsymbol{a}) = \sum_{i \in \text{cases}} g(n, \boldsymbol{a})_{\text{case}(i)}$.

D.2.3 Cases (5)–(12)

We now tackle cases (5)–(12), which have two edges that protrude and attach to the rest of the graph. Because of these two protruding edges, we have to carefully derive all three of the loop, vectorial, and combinatorial contributions.

We start with the vectorial contributions, as understanding them allows us to more easily explain and derive the loop and combinatorial contributions. We start by carefully walking through case (5), which contains two edges protruding from the first row. We take an existing graph of order n where $\mathbb{C}_{1,2}$ and the respective red edges match case (5). We then count how the numbers of edges of each type change after collapsing all of the paths that pass through the vertices in $\mathbb{C}_{1,2}$ into edges that lie within the other $2(n-1)$ columns.

Now, it is crucial to observe the following extremely important fact for *all cases* (5)–(12): the two protruding edges are always part of the same path that goes through $\mathbb{C}_{1,2}$, regardless of which of the four types of black edges are present between the vertices in $\mathbb{C}_{1,2}$. Therefore, when $\mathbb{C}_{1,2}$ is integrated out in graphs that match these cases, the edge that is created in the lower order graph is simply given by the two rows upon which those protruding edges are incident. That is, if the protruding edges connected to rows i and j, then, after integrating, an edge of type ij is created.

Now, there are, of course, 6 types of edges that can be created by collapsing a path: 11, 22, 33, 12, 13, and 23. However, it is somewhat convenient to actually describe 9 possible edges, 11, 22, 33, 12, 13, 23, 21, 31, and 32. The last three are equivalent to 12, 13, and 23 edges, respectively, but we order the edges in this way to account for the two possible ways that the protruding edges can connect into the graph (that is, *which* edge connects to row i or j, for example). Note that this separation is extraneous for certain cases, i.e. those with two edges protruding from the *same* row, but it is useful when considering cases with edges protruding from different rows.

To determine the vector contribution for a graph of order n with a_{12}, a_{13}, and a_{23} edges, we consider what edges b_{12}, b_{13}, and b_{23} on the graph of order $n-1$ remain after integrating out $\mathbb{C}_{1,2}$. Case (5) has two protruding edges coming from the first row, and then additional red edges of type 22 and 33. These 22 and 33 edges do not change the 12, 13, or 23 edge counts. Therefore, the only changes come from the collapse of the path associated with the two protruding edges from row 1.

Let us say that these two protruding edges are originally incident on rows 2 and 3. In this example, this means that when integrating out $\mathbb{C}_{1,2}$, we lose one edge of type 12 and one of type 13, but we *create* one of type 23. Therefore, we must have that $b_{12} = a_{12} - 1$, $b_{13} = a_{13} - 1$, and $b_{23} = a_{23} + 1$. Or, if we define $\Delta_{ij} := b_{ij} - a_{ij}$, then $(\Delta_{12}, \Delta_{13}, \Delta_{23}) = (-1, -1, +1)$. We then consider all possible vertices that these two protruding edges could have been connected to in the remainder of the graph, and that defines all possible $g(n-1, b_{12}, b_{13}, b_{23})$ that can contribute to $g(n, a_{12}, a_{13}, a_{23})_{\text{case}(5)}$.

Now, we must also consider some combinatorial factors $\mathcal{C}$. The combinatorial factors are really just a shorthand for determining how many times a contribution $g(n-1, b_{12}, b_{13}, b_{23})$ shows up when integrating out a given case, here case (5), from all the relevant graphs of order n. This is because different graphs at order n, when appropriately collapsed, lead to the same graph at order $n-1$. The combinatorial factor, then, is just a way of encoding this information.

Say that we are again considering an example where the original protruding edges attach to vertices in rows 2 and 3. Then an edge of type 23 is created. But if we look from the perspective of the lower order graph, *any* of the 23 edges could have been the one that was generated—that is, for some graph of order n with case (5) integrated out, a different 23 edge that is present is the one generated. Therefore, when we sum up all the contribution from integrating out case (5) over all relevant graphs of order n, we get a factor of b_{23}. Note also that, as we derived above, $b_{23} = a_{23} + 1$. Also note that, were we looking at protruding edges attached to the same row, we would get an additional factor of 2 due to the ambiguity of which edge attaches to which endpoint.

Finally, we consider the loop contribution. The calculation for case (5) is a relatively straightforward diagrammatic proof, which is detailed in Appendix D.2.3. In short, we draw all possible diagrams consistent with case (5) and count up the loops that are induced. There are only four cases, as the red edges are essentially fixed and there are four possible sets of black edges. The result is a factor $2k + 2$. That is, there are two sets of black edges that lead to an internal loop, leading to an extra factor of k, and there are two sets of black edges where the protruding edges snake through all vertices in $\mathbb{C}_{1,2}$ such that collapsing them just leads to a graph of order $n - 1$ without any extra loop factors.

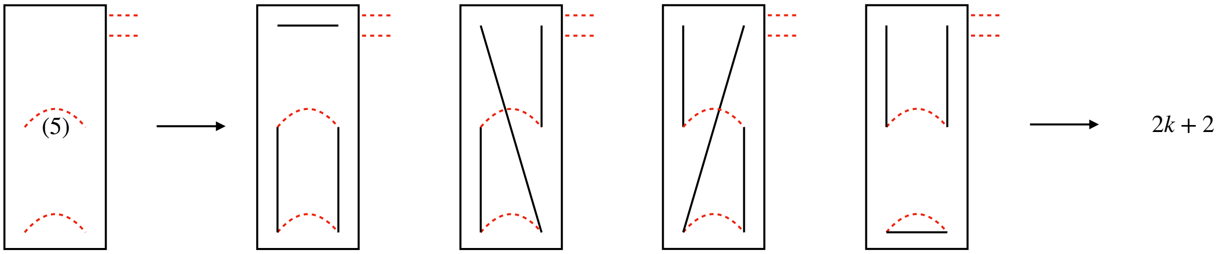

Figure D.4: Loop contribution for case (5).

So, putting all of the information together, we have that a full contribution from case (5) is

$$\begin{aligned}
g(n, a_{12}, a_{13}, a_{23})_{\text{case}(5)} &= (2k+2)[(2b_{11} + 2b_{12} + 2b_{13})g(n-1, a_{12}, a_{13}, a_{23}) \\
&\quad + 2b_{22}g(n-1, a_{12}-2, a_{13}, a_{23}) \\
&\quad + 2b_{23}g(n-1, a_{12}-1, a_{13}-1, a_{23}+1) \\
&\quad + 2b_{33}g(n-1, a_{12}, a_{13}-2, a_{23})] \\
&= (2k+2)[(2(n-1) + a_{12} + a_{13})g(n-1, a_{12}, a_{13}, a_{23}) \\
&\quad + (2(n-1) - (a_{12}-2) - a_{23})g(n-1, a_{12}-2, a_{13}, a_{23}) \\
&\quad + 2(a_{23}+1)g(n-1, a_{12}-1, a_{13}-1, a_{23}+1) \\
&\quad + (2(n-1) - (a_{13}-2) - a_{23})g(n-1, a_{12}, a_{13}-2, a_{23})].
\end{aligned} \tag{D.28}$$

This includes the loop, combinatorial, and vectorial factors. We also note that, should any of the combinatorial factors actually be negative, they should be set to 0, as that indicates that the graph that is constructed at lower order when integrating out the given case does not really exist (this is also handled by the vector input to g being negative—that is, one of the edge counts b_{12}, b_{13}, b_{23} is negative). One can get the contribution from case $(5s)$ by simply mapping $1 \leftrightarrow 3$.

We list the combinatorial and vectorial contributions for cases (5)–(8) in Table D.2 and cases (9)–(12) in Table D.3 (the main difference in the latter cases is that there is no longer a symmetry between red edges attaching to vertices ij and ji because, by convention, we attach the top protruding edge to the vertex in row i and the bottom protruding edge to the vertex in row j, which gives us different types of new edges, generically). The first column of these tables gives what kind of edge is created at order $n-1$. The second column tells us the combinatorial factor. The next four multicolumns give the vector information for each of the cases. Note that we do not give the symmetric cases, as they can be obtained by simply mapping $1 \leftrightarrow 3$.

			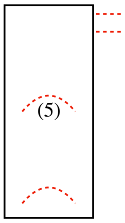			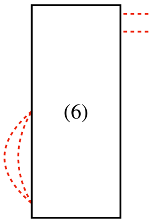			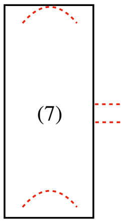			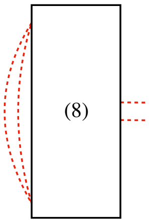	
Protruding Endpoints	$\mathcal{C}$	Δ_{12}	Δ_{13}	Δ_{23}	Δ_{12}	Δ_{13}	Δ_{23}	Δ_{12}	Δ_{13}	Δ_{23}	Δ_{12}	Δ_{13}	Δ_{23}
11	$2b_{11}$	0	0	0	0	0	-2	-2	0	0	-2	-2	0
12	b_{12}	0	0	0	0	0	-2	0	0	0	0	-2	0
13	b_{13}	0	0	0	0	0	-2	-1	+1	-1	-1	-1	-1
21	b_{12}	0	0	0	0	0	-2	0	0	0	0	-2	0
22	$2b_{22}$	-2	0	0	-2	0	-2	0	0	0	0	-2	0
23	b_{23}	-1	-1	+1	-1	-1	-1	0	0	0	0	-2	0
31	b_{13}	0	0	0	0	0	-2	-1	+1	-1	-1	-1	-1
32	b_{23}	-1	-1	+1	-1	-1	-1	0	0	0	0	-2	0
33	$2b_{33}$	0	-2	0	0	-2	-2	0	0	-2	0	-2	-2

Table D.2: Information for vectorial and combinatorial contributions to cases (5)–(8). Observe that there is a symmetry when the endpoints of the protruding edges are ij and ji. Also observe that, when the endpoints are the same, i.e. ii, there is an extra factor of 2 in the combinatorial term because of the ambiguity between how the protruding edges originally attach.

			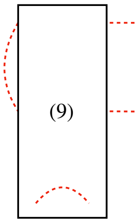			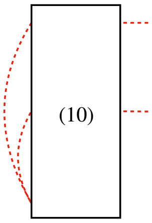			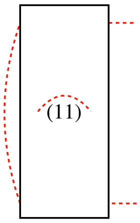			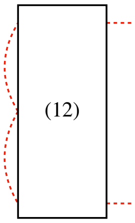	
Protruding Endpoints	$\mathcal{C}$	Δ_{12}	Δ_{13}	Δ_{23}	Δ_{12}	Δ_{13}	Δ_{23}	Δ_{12}	Δ_{13}	Δ_{23}	Δ_{12}	Δ_{13}	Δ_{23}
11	$2b_{11}$	-2	0	0	-1	-1	-1	0	-2	0	-1	-1	-1
12	b_{12}	0	0	0	+1	-1	-1	+1	-1	-1	0	0	-2
13	b_{13}	-1	+1	-1	0	0	-2	0	0	0	-1	+1	-1
21	b_{12}	-2	0	0	-1	-1	-1	0	-2	0	-1	-1	-1
22	$2b_{22}$	-2	0	0	-1	-1	-1	-1	-1	-1	-2	0	-2
23	b_{23}	-2	-0	0	-1	-1	-1	-1	-1	+1	-2	0	0
31	b_{13}	-2	0	0	-1	-1	-1	0	-2	0	-1	-1	-1
32	b_{23}	-1	-1	+1	0	-2	0	0	-2	0	-1	-1	-1
33	$2b_{33}$	-1	-1	-1	0	-2	-2	0	-2	0	-1	-1	-1

Table D.3: Information for vectorial and combinatorial contributions to cases (9)–(12). Observe that there is no longer a symmetry between ij and ji, but the ii cases still have an extra factor of 2 in the combinatorial term because the ambiguity between how the protruding edges originally attach still exists.

We also provide the loop contributions for cases (5)–(12) in Table D.4. These are derived in an analogous way to the diagrammatic approach in Appendix D.2.3, but there are many more graphs to consider. Therefore, using all of this information, we can derive an equivalent version of Eq. (D.28) for each case up to (12) (including the symmetric ones), accounting for all of their contributions.

Case	Loop contribution
(5)	$2k+2$
$(5s)$	$2k+2$
(6)	k^2+3k+4
$(6s)$	k^2+3k+4
(7)	$2k+2$
(8)	$2k+6$
(9)	$2k^2+6k+8$
$(9s)$	$2k^2+6k+8$
(10)	$2k^2+14k+16$
$(10s)$	$2k^2+14k+16$
(11)	$4k+12$
(12)	$2k^2+14k+16$

Table D.4: Loop contributions for each of the cases (5)–(12). Notice that symmetric versions of cases have the same loop contribution; only their vectorial and combinatorial contributions are different.

D.2.4 Cases (13)–(16)

We now move on to more complicated cases that have four protruding edges. The vectorial contribution is more difficult to calculate, as we must account for $3^4 = 81$ possibilities for how the protruding edges attach to the lower order graph. Furthermore, there is more interaction between the vectorial, combinatorial, and loop terms. This did not occur in the previous sets of cases because the protruding edges were always part of the same path through the black edges attached to the vertices in $\mathbb{C}_{1,2}$. However, one must now keep track of which protruding edges connect to one another through the vertices in $\mathbb{C}_{1,2}$.

For example, we look at the possibilities for case (13), shown in Fig. D.5. By convention, we take the top left vertex to row a, the top right vertex to row b, the middle left vertex to row c, and the middle right vertex to row d, where $a, b, c, d \in \{1, 2, 3\}$. We see that, when the black edges attached to the vertices in $\mathbb{C}_{1,2}$ are type-1, then the red edges that protrude from the top row are connected to one another, which means that one generates an edge of type ab when collapsing this path. However, if the black edges associated with $\mathbb{C}_{1,2}$ are type-2, then it is instead ac and bd that are connected. In total, one of the possible types of black edges connect edges ab and cd, and three connect ac and bd. In the case where ab and cd are connected, this means that we generate edges of type ab and cd but we lose edges of type $1a, 1b, 2c, 2d$. When ac and bd are connected, we of course gain edges of type ac and bd, but we still lose edges of type $1a, 1b, 2c, 2d$. We use these observations to build up the vectorial contribution of the graph by summing over all 81 possibilities of $a, b, c, d \in \{1, 2, 3\}$. This is tedious to do by hand, but simple numerically.

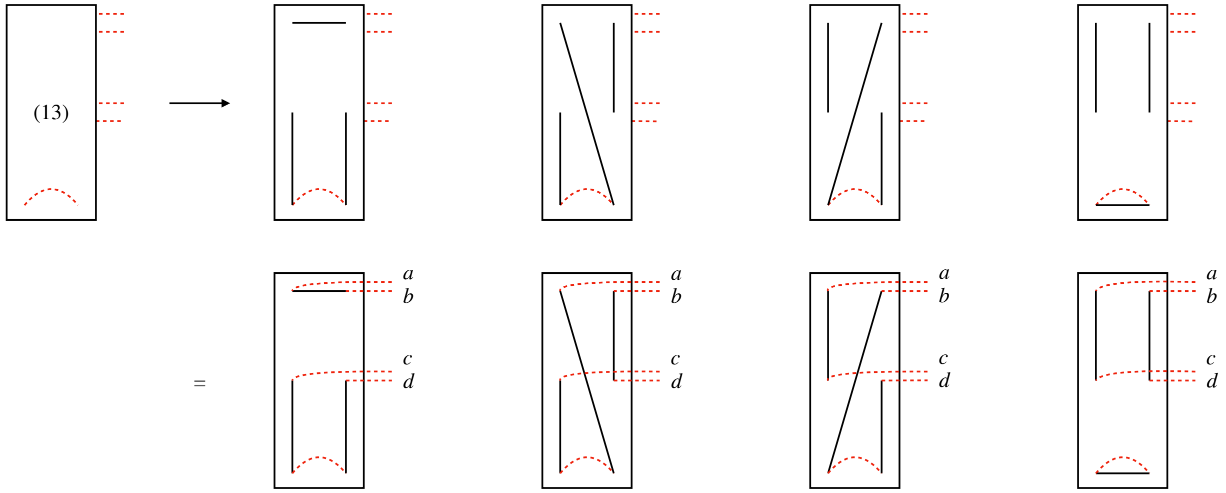

Figure D.5: Evaluation of case (13). By convention, we take the top left vertex to row a, the top right vertex to row b, the middle left vertex to row c, and the middle right vertex to row d, where $a, b, c, d \in \{1, 2, 3\}$. The types of edges that are created after integrating out the two leftmost columns are determined by the type of the black edges.

We need also account for the loop and combinatorial factors that associate to each of these vectorial contributions. Luckily, we do not need to consider 81 cases parameterized by a, b, c, d, but we must consider each of the subcases defined by the four possible sets of black edges in connecting the vertices in $\mathbb{C}_{1,2}$. Loop-wise, we simply need to count how many loops are induced. Working from the left to right in Fig. D.5, we get $0, 0, 0, 1$ loops, respectively, leading to factors of $1, 1, 1, k$, respectively. The combinatorial factor is given by

$$2^{\delta_{ab}} 2^{\delta_{cd}} \left[(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc} - \delta_{abcd}) 2 \binom{b_{ab}}{2} + (1 - (\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc} - \delta_{abcd})) b_{ab} b_{cd} \right] \tag{D.29}$$

in the case where edges ab and cd are connected. If instead ac and bd are connected, we replace each instance of ab and cd with ac and bd, respectively. We then again account for all 81 cases and attach each combinatorial factor and loop factor to its associated vectorial term.

To understand Eq. (D.29), consider the following, where we assume we are dealing with type-1 black edges so that we are creating edges ab and cd. We get a factor of 2 when a and b are the same because they correspond to protruding edges coming from the same row, meaning there is a choice of which edge to connect where. The same holds for c and d. If all four edges connect to the same row, i.e. $a = b = c = d$, then one might naively think we need to add an extra factor of 6 (to get to a total of $4!$ possible connections), but this is incorrect, as ab and cd are always paired given their connection through case (13) with black edges of type-1. Now, if $a = c$ and $b = d$ or $a = d$ and $b = c$, then the two edges ab and cd are the same type, meaning we are creating two edges of the same type in the graph of order $n - 1$. There are therefore $\binom{b_{ab}}{2}$ choices of which edges these are in the lower order graph, but we also need an extra factor of 2 to decide which one the groups of protruding edges each maps to. If ab and cd correspond to different types of edges, then we just get a factor of $b_{ab}b_{cd}$, as we simply need to account for which of these edges are generated through the integration process.

Therefore, we see that cases (13)–(16) raise substantially more complications in their evaluation. In particular, the type of black edges leads to far more interaction between the loop, vectorial, and combinatorial contributions that must be carefully combined in code to achieve the correct recursion. While we have only described case (13) in detail, cases (14)–(16) follow in the exact same manner, though there are more graphs to consider in the cases where two rows have only one protruding edge.

D.2.5 Case (17)

Case (17) raises the same issues, though there are only four graphs to consider. However, we have $243 = 3^6$ possible options for how the protruding edges may connect to the graph at lower order (this is true in general, but not all of these are possible when n is small). See Fig. D.6. We

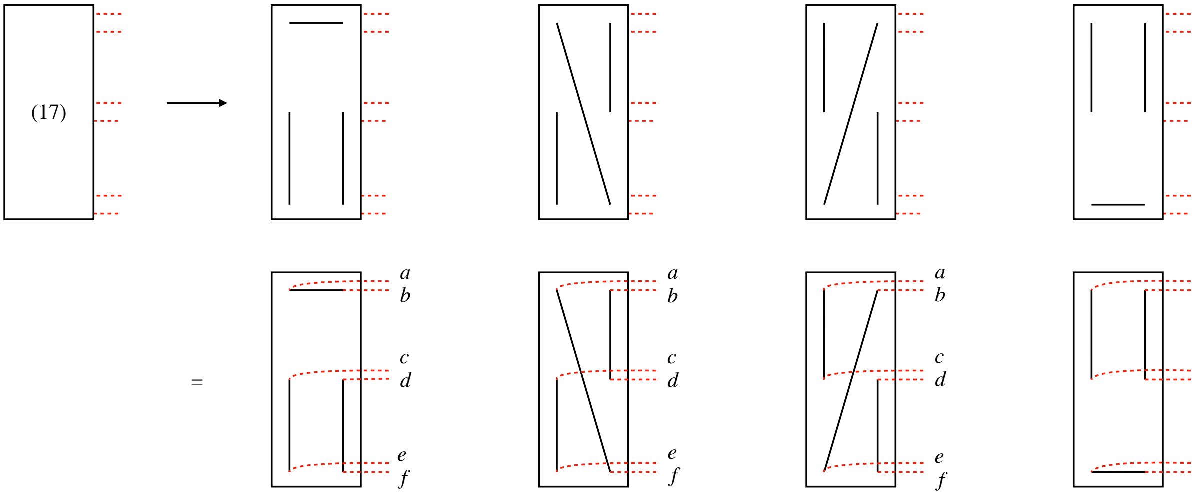

Figure D.6: Evaluation of case (17). We repeat the convention for cases (13)–(16) by taking the top left vertex to row a, the top right vertex to row b, the middle left vertex to row c, and the middle right vertex to row d, but we now also take the bottom left to e and the bottom right to f, where $a, b, c, d, e, f \in \{1, 2, 3\}$. The types of edges that are created are still determined by the type of the black edges.

repeat the convention for cases (13)–(16) by taking the top left vertex to row a, the top right vertex to row b, the middle left vertex to row c, and the middle right vertex to row d, but we now also take the bottom left to e and the bottom right to f, where $a, b, c, d, e, f \in \{1, 2, 3\}$. Now, for type-1 black edges, we create ab, ce, and df; for type-2, it is af, bd, and ce; for type-3 it is ac, be, and df; and for type-4 it is ac, bd, and ef. We always lose edges of type $1a, 1b, 2c, 2d, 3e, 3f$ regardless of the type of the black edges. Furthermore, the loop contribution is always a factor of 1, as there are no internal loops to case (17).

The combinatorial factor, however, is quite complicated. Assume for now that we are working with type-1 black edges such that ab, ce, and df are linked. The combinatorial factor is

$$(2^{\delta_{ab}})^3 3! \binom{b_{ab}}{3} \qquad \times \mathbb{1}[\{a, b\} = \{c, e\} = \{d, f\}] \tag{D.30}$$

$$+2^{\delta_{ab}} 2^{\delta_{ce}} \times 2 \binom{b_{ab}}{2} \times 2^{\delta_{df}} b_{df} \qquad \times \mathbb{1}[\{a, b\} = \{c, e\} \neq \{d, f\}] \tag{D.31}$$

$$+2^{\delta_{ab}} 2^{\delta_{df}} \times 2 \binom{b_{ab}}{2} \times 2^{\delta_{ce}} b_{ce} \qquad \times \mathbb{1}[\{a, b\} = \{d, f\} \neq \{c, e\}] \tag{D.32}$$

$$+2^{\delta_{ce}} 2^{\delta_{df}} \times 2 \binom{b_{ce}}{2} \times 2^{\delta_{ab}} b_{ab} \qquad \times \mathbb{1}[\{a, b\} \neq \{c, e\} = \{d, f\}] \tag{D.33}$$

$$+2^{\delta_{ab}} 2^{\delta_{ce}} 2^{\delta_{df}} b_{ab} b_{ce} b_{df} \qquad \times 1[\{a, b\} \neq \{c, e\} \neq \{d, f\}]. \tag{D.34}$$

Here, $\mathbb{1}[\mathrm{A}]$ is an indicator function that is 1 if statement A is true and 0 if it is false. For example, $\mathbb{1}[\{a, b\} = \{c, e\}, \{d, f\}]$ is 1 if $\{a, b\}$, $\{c, e\}$, and $\{d, f\}$ are all equal as sets (that is, order does not matter). The middle three lines [Eqs. (D.31) to (D.33)] are just repetitions of the combinatorial factors for cases (13) – (16), but accounting for which sets of four edges may be sent to the same row. The last line [Eq. (D.34)] is simple and accounts for the case where all of the edge types ab, ce, df are different. The first line [Eq. (D.30)] requires a bit of explanation. In the case where $a \neq b$, we simply have to choose three edges of type ab where the order matters (they each could have been created by integrating out different graphs at a higher order). In the case where $a = b$, this is still the case, but now we need a factor of 2 for each edge, as we can flip which vertices are connected where.

Again, it is hard to account for all of these elements by hand, but it is simple numerically. With this final case sorted out, we simply combine contributions of all of the cases $g(n, \boldsymbol{a})_{\text{case}(i)}$ to find $g(n, \boldsymbol{a})$.

D.3 Computing Individual Coefficients

In this Appendix, we discuss the various methods by which one can compute individual coefficients in the polynomial expansion of the second moment. Recall that, per Theorem 5.2, the second moment may be expanded as

$$M_2(k, n) = (2n-1)!! \sum_{i=1}^{2n} c_i k^i. \tag{D.35}$$

Ideally, one would simply be able to find a closed functional form for the right-hand side of this equation (as was possible for the equivalent definition of the first moment). But, unfortunately, such a result currently eludes us. Therefore, the best we can do is find individual coefficients. We now discuss methods of calculating c_{2n} and c_{2n-1}.

D.3.1 Leading Order Coefficient c_{2n}

We begin with the leading order coefficient c_{2n}. Recall that Lemma 5.1(ii)[1] gives that $c_{2n} = (2n)!!$. The proof of this lemma is contained in Appendix C.3, and we briefly describe that proof. However, we also provide a second technique for understanding the result that is useful to understanding the proof of the first sub-leading order term c_{2n-1}.

Recall that, in order for a graph in $\mathbb{G}_n^2 = \mathbb{G}_n^2(0,0,0)$ to have $2n$ connected components, it must possess only type-1 and type-4 black edges. The two vertices connected by each horizontal black edge must also be connected by a red edge to form a 2-vertex connected component. The remaining vertical edges from the type-1 sets of black edges are then paired off (i.e., connected via horizontal red edges) into 4-vertex connected components, and the same holds for black vertical edges from type-4 sets. This leads to $2n$ total connected components. The original proof that the total number of graphs satisfying these constraints is $(2n)!!$ proceeds by reducing these graphs to ones in $\mathbb{G}_n^1$ and then counting them (with a weight given by the number of connected components). This is evaluated by using the equation for the first moment in Theorem 5.1.

Another way to compute this coefficient is by making a combinatorial argument. As discussed, c_{2n} contains contributions only from graphs that possess solely type-1 and type-4 sets of black edges. Again, in order to create the maximal number of connected components, the horizontal black edges must also be connected by red edges to create a size-2 connected component. The remaining type-1 vertical black edges are paired off, and the type-4 vertical black edges are similarly paired off. So, for a graph of order n, say that there are p sets of type-1 black edges and, therefore, $n-p$ sets of type-4 black edges. There are $\binom{n}{p}$ sets of black edges with this type

[1] Note added: Recall that this is a restatement of Lemma 4.1, hence why the proof is listed in Appendix C.

distribution. There are then $(2p-1)!!$ ways to pair off the $2p$ vertical type-1 black edges, and $(2n-2p-1)!!$ ways to pair off the $2n-2p$ vertical type-4 black edges. Therefore, summing over $p \in \{0,1,\dots,n\}$, we get that

$$c_{2n} = \sum_{p=0}^{n} \binom{n}{p} (2p-1)!!(2n-2p-1)!!. \tag{D.36}$$

We can massage the right-hand side a bit using the fact that $(2x-1)!! = (2x)!/(2x)!! = (2x)!/(2^x x!)$. Expanding out the binomial coefficient and converting all terms to single factorials yields

$$c_{2n} = \frac{n!}{2^n} \sum_{p=0}^{n} \binom{2p}{p} \binom{2n-2p}{n-p}. \tag{D.37}$$

The summation evaluates to 4^n using the convolution of the Taylor series for $(1-4x)^{-1/2}$ [191]. Therefore,

$$c_{2n} = 2^n n! = (2n)!!, \tag{D.38}$$

which, of course, matches the known result.

D.3.2 First Subleading Coefficient c_{2n-1}

We now generalize the above combinatorial version of the c_{2n} calculation to c_{2n-1}. It is slightly more complicated, as there is a bit of casework to consider, but the general idea is the same. In particular, the key idea is that because $2n$ is the maximal number of connected components, finding a graph with $2n-1$ connected components comes down to counting the ways that one can create a "deficit" of exactly one connected component from the maximal number. There are nine ways to accomplish this.

First, consider starting with graphs with a maximal number of connected components, meaning, as per Appendix D.3.1, they have only type-1 and type-4 black edges. The connected components have either 2 vertices (red and black edge between 2 vertices in the same row) or 4 (two vertical black edges of the same type that are paired off via red edges). We refer to these as type-x 2-vertex and 4-vertex connected components, respectively (where x is either 1 or 4). One can convert these graphs with maximal connected components into graphs with a deficit of a single connected component in the following ways, all of which involve merging two connected components into a single one:

(1): merge two type-1 2-vertex connected components;

(2): merge two type-4 2-vertex connected components;

(3): merge one type-1 2-vertex connected component with one type-4 4-vertex connected component;

(4): merge one type-4 2-vertex connected component with one type-1 4-vertex connected component;

(5): merge two type-1 4-vertex connected components;

(6): merge two type-4 4-vertex connected components;

(7): merge one type-1 4-vertex connected component with one type-4 4-vertex connected component.

These options are visualized (up to the symmetry of exchanging the roles of type-1 and type-4 edges) in Fig. D.7.

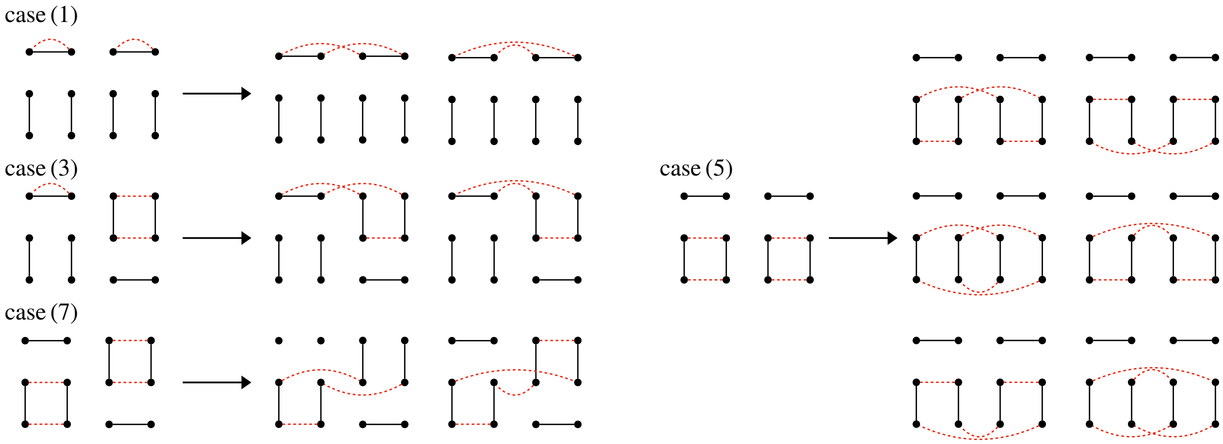

Figure D.7: Possible ways of merging type-1 and type-4 vertices to create a deficit of a single connected component. Here, we only show cases (1), (3), (5), and (7), as (2), (4), and (6) are symmetric with (1), (3), and (5) with type-1 and type-4 edges switched.

Next, we must also consider cases with type-2 and type-3 black edges. There are two options here: either the graph can have exactly one set of type-2 or type-3 edges, or it can have exactly two sets (it does not matter whether it is two type-2 sets of edges, two type-3 sets of edges, or one of each). The rest of the sets of black edges must all be of type 1 or type 4. Then, creating a deficit can be done in the following ways:

(8): connect one type-2 or type-3 edge (the edge connecting the top row to the bottom row) to one type-1 vertical edge and one type-4 vertical edge to make a 6-vertex loop;

(9): connect two type-2 or type-3 edges (again, the top-to-bottom edges) to form a 4-vertex connected component.

These are visualized in Fig. D.8. The rest of horizontal black edges must be connected with red edges to form 2-vertex connected components, and the remaining vertical edges must be appropriately paired off in order to ensure $2n - 2$ other connected components are formed.

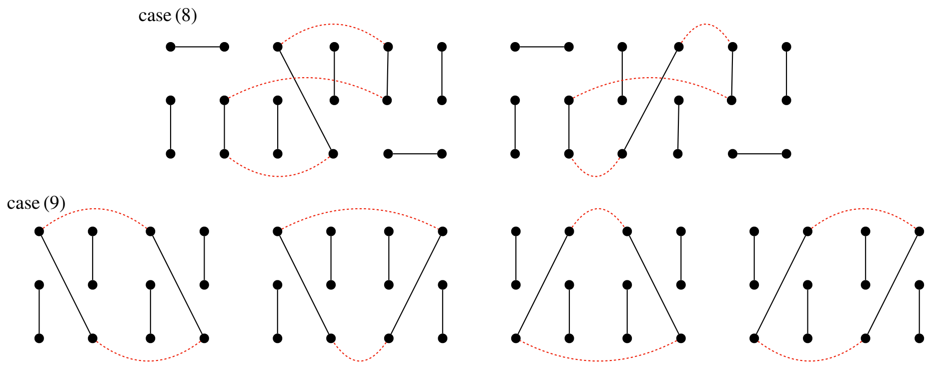

Figure D.8: Possible ways of creating a deficit of a single connected component while using type-2 and/or type-3 edges.

The end result of accounting for all of these cases is a (double) sum that computes c_{2n-1}:

$$
\begin{aligned}
c_{2n-1} = \sum_{p=0}^{n}\binom{n}{p} \times \Bigg[& \underbrace{2\binom{p}{2}(2p-1)!!(2(n-p)-1)!!}_{(1)} + \underbrace{2\binom{n-p}{2}(2p-1)!!(2(n-p)-1)!!}_{(2)} \\
& + \underbrace{2\binom{p}{1}\binom{2(n-p)}{2}(2p-1)!!(2(n-p-1)-1)!!}_{(3)} + \underbrace{2\binom{n-p}{1}\binom{2p}{2}(2(p-1)-1)!!(2(n-p)-1)!!}_{(4)} \\
& + \underbrace{6\binom{2p}{4}(2(p-2)-1)!!(2(n-p)-1)!!}_{(5)} + \underbrace{6\binom{2(n-p)}{4}(2p-1)!!(2(n-p-2)-1)!!}_{(6)} \\
& + \underbrace{2\binom{2p}{2}\binom{2(n-p)}{2}(2(p-1)-1)!!(2(n-p-1)-1)!!}_{(7)} \Bigg] \\
& + \sum_{p=0}^{n-1}\underbrace{2\binom{n}{1}\binom{n-1}{p}(2p+1)(2(n-p-1)+1)(2p-1)!!(2(n-p-1)-1)!!}_{(8)} \\
& + \sum_{p=0}^{n-2}\underbrace{4\binom{n}{2}\binom{n-2}{p}(2p+1)!!(2(n-p-2)+1)!!}_{(9)}
\end{aligned}
\tag{D.39}
$$

The last sum should be taken to be 0 when $n = 1$ and the sum is empty (this is because this case of course requires at least $n = 2$ to have two sets of type-2/3 edges). Each of these terms can be derived through a simple combinatorial argument regarding which types of edges are present and how they must be connected. For each case, say that there are p type-1 sets of black edges. This means there are $n - p$, $n - p - 1$, and $n - p - 2$ sets of type-4 black edges for cases (1)-(7), case (8), and case (9), respectively (in the latter two cases, the remaining set(s) of edges are type-2 and/or type-3). Each case then comes down to deciding how to order the sets of edges, how to choose which edges are connected together, and then pairing off the remaining edges of the same type to build the remaining 2- and 4-vertex connected components. We do not detail how to count every single case, but we discuss two examples, case (1) and case (8). The rest should be straightforward to derive by extending these arguments.

In case (1), we merge two 2-vertex connected components of type 1. First, we have a factor of $\binom{n}{p}$ to account for all ways of having p type-1 sets of edges. We then must select 2 of the p horizontal black edges to merge into a single connected component, hence the factor of $\binom{p}{2}$; see Fig. D.7. The additional factor of 2 comes from the two possible ways of merging these into a single connected component. Finally, the remaining double factorial factors are the number of ways of pairing off the vertical black edges with those of the same type. We then must sum from $p = 0$ to n to account for all possible black edge type distributions.

Case (8) proceeds similarly. First, we have a factor of $\binom{n}{1}$, or n, to choose where the type-2 or type-3 set of edges is. The factor of 2 out front now actually accounts for whether it is type 2 or type 3. Next, we have $\binom{n-1}{p}$ to account for the placement of the p type-1 sets of edges. Next, there are now $2p+1$ black edges that span the second and third rows (i.e., they are black edges that arise from type-1 sets of black edges). It is $2p + 1$ because the type-2 or type-3 set of black edges contributes 1, and the p type-1 sets contribute $2p$. Analogously, there are also $(2(n - p - 1) + 1)$ black edges spanning the first and second rows. We have to select one of each to connect to the black edge that spans the first and third rows to make a single 6-vertex connected component. The remaining factors are again the number of ways to pair off the remaining vertical black edges with those of the same type (horizontal black edges must form 2-vertex connected components to reach the required number of connected components).

It is possible, but quite tedious, to simplify this double sum by looking at each individual term and then applying a similar technique as in the evaluation of the sum for c_{2n}. That is, for each term in the sum, we use the convolution of various Taylor series and compare the coefficients of x^n. We start with the first term

$$(1) \to \sum_{p=0}^{n} \binom{n}{p} 2 \binom{p}{2} (2p-1)!!(2(n-p)-1)!! = \frac{n!}{2^n} \sum_{p=0}^{n} p(p-1) \binom{2p}{p} \binom{2n-2p}{n-p}. \tag{D.40}$$

One then has through Taylor expansion that

$$x^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} \frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty} \binom{2n}{n} n(n-1) x^n, \tag{D.41}$$

which implies that

$$12x^2 \frac{1}{(1-4x)^3} = \left(x^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} \frac{1}{\sqrt{1-4x}} \right) \frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty} \sum_{p=0}^{n} p(p-1) \binom{2p}{p} \binom{2n-2p}{n-p} x^n. \tag{D.42}$$

Using the Online Encyclopedia of Integer Sequences (OEIS), we find the three-fold convolution of powers of 4 A038845 [192] has formula $(n+2)(n+1)2^{2n-1}$, meaning

$$\sum_{n=0}^{\infty} 12(n+2)(n+1)2^{2n-1} x^{n+2} = 12x^2 \frac{1}{(1-4x)^3} = \sum_{n=0}^{\infty} \sum_{p=0}^{n} p(p-1) \binom{2p}{p} \binom{2n-2p}{n-p} x^n. \tag{D.43}$$

Therefore, comparing powers of x, we get that

$$\sum_{p=0}^{n} p(p-1)\binom{2p}{p}\binom{2n-2p}{n-p} = 12n(n-1)2^{2n-5}, \tag{D.44}$$

meaning the first term in the sum is (after some algebra)

$$\sum_{p=0}^{n}\binom{n}{p}2\binom{p}{2}(2p-1)!!(2(n-p)-1)!! = (2n)!!\frac{3n(n-1)}{8}. \tag{D.45}$$

Note also by the symmetry between p and $n-p$, the contribution of the second term is the same.

We can perform similar manipulations for the other terms. In particular,

$$(3) \to \sum_{p=0}^{n}\binom{n}{p}2p\binom{2n-2p}{2}(2p-1)!!(2(n-p-1)-1)!! = \frac{n!}{2^{n-1}}\sum_{p=0}^{n}(n-p)p\binom{2p}{p}\binom{2n-2p}{n-p}.$$

(D.46)

Instead of taking the Taylor expansion for the second derivative of $(1-4x)^{-1/2}$ and convolving it with that for $(1-4x)^{-1/2}$, we convolve the Taylor series for the first derivative with itself. That is,

$$\frac{4x^2}{(1-4x)^3} = \left(x\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sqrt{1-4x}}\right)^2 = \sum_{n=0}^{\infty}\sum_{p=0}^{n} p(n-p)\binom{2p}{p}\binom{2n-2p}{n-p}x^n, \tag{D.47}$$

which, using the same result as for (1) (just with a difference of a factor of 3), yields

$$\sum_{p=0}^{n} p(n-p)\binom{2p}{p}\binom{2n-2p}{n-p} = 4n(n-1)2^{2n-5}. \tag{D.48}$$

This means that the third term yields a contribution of

$$(3) \to \frac{n!}{2^{n-1}} 4n(n-1)2^{2n-5} = (2n)!!\frac{2n(n-1)}{8}. \tag{D.49}$$

Again, by the symmetry between n and $n-p$, the contribution from the fourth term is the same.

Next:

$$\begin{aligned}(5) \to \sum_{p=0}^{n}\binom{n}{p}6\binom{2p}{4}(2(p-2)-1)!!(2(n-p)-1)!!\\ = \frac{n!}{2^n}\sum_{p=0}^{n} p(p-1)\binom{2p}{p}\binom{2n-2p}{n-p} = (2n)!!\frac{3n(n-1)}{8}\end{aligned} \tag{D.50}$$

because this is the exact same as (1). Again, by symmetry, (6) has the same contribution.

We also have that

$$\begin{aligned}(7) \to \sum_{p=0}^{n}\binom{n}{p}2\binom{2p}{2}\binom{2(n-p)}{2}(2(p-1)-1)!!(2(n-p-1)-1)!!\\ = \frac{n!}{2^{n-1}}\sum_{p=0}^{n} p(n-p)\binom{2p}{p}\binom{2n-2p}{n-p} = (2n)!!\frac{2n(n-1)}{8},\end{aligned} \tag{D.51}$$

which follows because this term happens to be the same as (3).

We now move on to the final two cases. Again, similar manipulations yield that

$$\begin{aligned}(8) \to &\sum_{p=0}^{n-1} 2\binom{n}{1}\binom{n-1}{p}(2p+1)(2(n-p-1)+1)(2p-1)!!(2(n-p-1)-1)!! \\ &= \frac{n!}{2^{n-1}}\sum_{p=0}^{n}(2p+1)(n-p)\binom{2p}{p}\binom{2n-2p}{n-p} \\ &= \frac{n!}{2^{n-1}}2\sum_{p=0}^{n}p(n-p)\binom{2p}{p}\binom{2n-2p}{n-p} + \frac{n!}{2^{n-1}}\sum_{p=0}^{n}(n-p)\binom{2p}{p}\binom{2n-2p}{n-p}.\end{aligned} \tag{D.52}$$

We have expanded the upper limit to $p = n$ because the factor of $n - p$ sets this additional contribution to 0. The first term in the last equation is simply twice the contribution of (3), which is $(2n)!!4n(n-1)/8$. The second term requires yet another manipulation of Taylor series. By very similar arguments to the above, we have that

$$x\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty}\binom{2n}{n}nx^n, \tag{D.53}$$

which implies that

$$\frac{2x}{(1-4x)^2} = \left(x\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sqrt{1-4x}}\right)\frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty}\sum_{p=0}^{n}p\binom{2p}{p}\binom{2n-2p}{n-p}x^n, \tag{D.54}$$

which is the same as the sum we are interested in (up to the symmetry of replacing $n - p$ with p). Using OEIS sequence A002697 [192], that is, the convolution of powers of 4, we find that

$$\frac{2x}{(1-4x)^2} = \sum_{n=0}^{\infty}2(n+1)4^n x^{n+1}, \tag{D.55}$$

which means that, comparing powers of x^n,

$$\frac{n!}{2^{n-1}}\sum_{p=0}^{n}(n-p)\binom{2p}{p}\binom{2n-2p}{n-p} = \frac{n!}{2^{n-1}}2n4^{n-1} = n(2n)!!. \tag{D.56}$$

Finally, then

$$(8) \to (2n)!!\frac{4n(n-1)}{8} + (2n)!!n = \frac{4n(n+1)}{8}. \tag{D.57}$$

Last, we get that

$$\begin{aligned}(9) \to &\sum_{p=0}^{n-2}4\binom{n}{2}\binom{n-2}{p}(2p+1)!!(2(n-p-2)+1)!! \\ &= \frac{n!}{2^{n-1}}\sum_{p=0}^{n-2}\binom{2p+2}{p+1}\binom{2n-2p-2}{n-p-1}(p+1)(n-p-1) \\ &= \frac{n!}{2^{n-1}}\sum_{x=0}^{n}\binom{2x}{x}\binom{2n-2x}{n-x}(x)(n-x),\end{aligned} \tag{D.58}$$

where we have set $x = p+1$ and then expanded the limits of summation to include $x = 0$ and $x = n$ (because these terms contribute 0). Therefore, this contribution is the same as (3), (4), and (7), which is $(2n)!!2n(n-1)/8$.

Therefore, in total, we have that

$$\frac{c_{2n-1}}{(2n)!!} = 4\frac{3n(n-1)}{8} + 4\frac{2n(n-1)}{8} + \frac{4n(n-1)}{8} + n = (3n-2)n. \tag{D.59}$$

Therefore,

$$c_{2n-1} = (2n)!!(3n-2)n. \tag{D.60}$$

Numerically evaluating the sums yields the same value up to $n = 40$, and this also matches the value of c_{2n-1} computed via the recursion. We note that $(3n-2)n$ are the so-called octagonal numbers, which are OEIS entry A000567 [192]. However, we are not sure whether there is a deeper connection between these numbers and the graph theoretic problem at the core of this calculation. Additionally, while it is nice that we have been able to find an exact formula for a second coefficient, this calculation does not seem scalable, meaning other methods are likely needed to try to find the full expansion of the second moment.

D.4 Alternative method for computing coefficients c_i

In this Appendix, we present an alternative method for computing coefficients c_i in

$$M_2(k,n) = (2n-1)!! \sum_{i=1}^{2n} c_i k^i. \tag{D.61}$$

Using this method, we obtain a useful expression for c_1. We also outline how this method can be used to set up an alternative recursive code for computing the coefficients c_i for all i. While we have not implemented this code, there is a possibility it is more efficient than the recursive code discussed in the main text. It is also possible that this new method may yield other useful analytical results about c_i, including their asymptotic behavior.

We start by recalling Eq. (5.19):

$$M_2(k,n) = (2n-1)!! \sum_{G \in \mathbb{G}_n^2} k^{C(G)}, \tag{D.62}$$

where the sum goes over all graphs possessing the allowed assignments of black and red edges. The new method relies on the following key simplifying observation: for a given fixed assignment of black edges, the contribution to $M_2(k,n)$ (summed over all allowed red edge assignments) depends only on $\boldsymbol{e} = (e_{11}, e_{12}, e_{13}, e_{23}, e_{33})$, where e_{ij} is the number of black edges that connect row i to row j. In particular, the answer does not depend on what columns the black edges are connecting. The proof of this key observation is simple: for a fixed set of black edges, the contribution to $M_2(k,n)$ is summed over all possible red perfect matchings in each of the three rows. This means that we can swap any two vertices in a given row (while pulling the ends of the black edges to the new destinations) without changing the answer. This completes the proof.

Let p_1 be the number of type-1 sets of black edges, p_4 be the number of type-4 sets of black edges, and p be the combined number of type-2 and type-3 sets of black edges (type-2 and type-3 sets are equivalent as far as their contributions to e_{ij}). Then $e_{11} = p_1$, $e_{33} = p_4$, $e_{12} = p + 2p_4$, $e_{23} = p + 2p_1$, and $e_{13} = p$. We then write

$$M_2(k,n) = \sum_{p_1=0}^{n} \sum_{p_4=0}^{n-p_1} \binom{n}{p_1}\binom{n-p_1}{p_4} 2^p g(\boldsymbol{e}), \tag{D.63}$$

where $p = n - p_1 - p_4$. The combinatorial factors come from choosing p_1 sets of type-1 black edges out of n possible locations, then p_4 sets of type-4 black edges from $n - p_1$ possible locations, and finally multiplying by a factor of 2 for each choice of whether a given contribution to p is type-2 or type-3. Additionally,

$$g(\boldsymbol{e}) = \sum_{i=1}^{2n} d_i(\boldsymbol{e}) k^i, \tag{D.64}$$

where $d_i(\boldsymbol{e})$ is the number of ways (using the allowed red-edge assignments) to make i loops given the black edges specified by $\boldsymbol{e}$.

The coefficients $d_i(\boldsymbol{e})$ can then be computed with the help of the visualization shown in Fig. D.9(a). The three black dots labeled 1, 2, and 3 represent the three rows. The numbers e_{jk} on the five edges (including the two loops) show how many black edges connect row j to row k. Roughly speaking, the coefficient $d_i(\boldsymbol{e})$ is the number of ways to connect all the black edges specified by $\boldsymbol{e}$ into exactly i loops. The red edges are used to connect the black edges to each other and are taken into account automatically, which is one of the key advantages of this approach (slightly more specifically, for any two black edges that share a row, it is possible to connect them with a red edge between the vertices in that shared row). Each way of joining the edges $\boldsymbol{e}$ into loops also comes with a combinatorial factor that takes into account the fact that all edges are distinguishable and the fact that edges that stay in the same row can each be traversed in one of two directions.

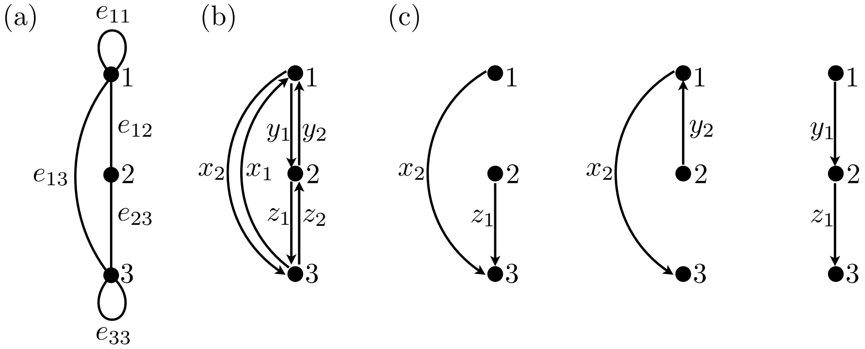

Figure D.9: Graphs useful for understanding the new method for calculating coefficients c_i. (a) Once the types of black edges are assigned, the contribution to $M_2(k,n)$ depends only on the number of black e_{ij} edges connecting row i to row j. (b) In order to compute c_1, the number of single-loop graphs contributing to M_2, we first set $e_{11} = e_{33} = 0$ (later adding in the effect of nonzero values), fix the winding number w of the loop, and break up 1-3 edges into $x_1 = (e_{13} + w)/2$ clockwise edges and $x_2 = (e_{13} - w)/2$ counterclockwise edges. We similarly break up the 1-2 and 2-3 edges. We then use the BEST theorem [193] to count the number of Eulerian circuits on this directed graph. (c) For $u = 3$, the three types of arborescences contributing to $t_u(G) = x_2 z_1 + y_2 x_2 + y_1 z_1$ for the graph G shown in (b).

The coefficients $g(\boldsymbol{e})$ can be computed using a recursive procedure. Instead of doing a recursion on n (which is what we do in the main text, with details presented in Appendix D.2), we perform the recursion on the number of black edges $e_{11} + e_{12} + e_{13} + e_{23} + e_{33}$. As in the main text, we need to define a more general function $g(\boldsymbol{e}, \sigma, c, s)$ to make the recursion work. σ is a binary variable, so that $\sigma = 1$ means we are in the process of building a loop, while $\sigma = 0$ means that we need to start a new loop. If $\sigma = 1$, we need to also specify $s \in \{1, 2, 3\}$ (standing for start) indicating the row where the current loop started and $c \in \{1, 2, 3\}$ (standing for current) indicating the row where we currently are.

As in the main text, the recursive procedure is efficient, i.e. takes polynomial time in the number of edges. We first directly compute $g(\boldsymbol{e}, \sigma, c, s)$ for small values of $e_{11}+e_{12}+e_{13}+e_{23}+e_{33}$. Then the recursive step goes as follows. If $\sigma = 0$, we can either (1) close the loop right away by reducing e_{11} or e_{33} by 1, keep $\sigma = 0$, and multiply by k, or (2) set $\sigma = 1$, start a new loop at row i, set $s = i$, reduce e_{ij} by 1 (for some j), and set $c = j$. If $\sigma = 1$, we can either (1) close the loop by reducing e_{cs} by 1, set $\sigma = 0$, and multiply by k, or (2) continue building the loop, keep $\sigma = 1$, keep s unchanged, reduce e_{cj} by 1 (for some j), and change the value of c to j. As we do these calculations, we need to also include appropriate combinatorial factors deciding which black edge to take (e.g., if we pick one of e_{ij} edges, we need to multiply by e_{ij}, and if $i = j$, we need to multiply by another factor of 2).

While we have not coded up this procedure, we believe that it offers another complementary way of understanding and analyzing the second moment.

D.4.1 Computing c_1

Again, while we have not coded up the above recursive procedure, we will show how to use the new approach to compute c_1 in Eq. (D.61), i.e., the number of ways to build a single-loop graph, which we were not able to directly compute using the original method.

To proceed, we will at first ignore the contributions of the edges e_{11} and e_{33} (effectively pretending that they are equal to zero), but we will address how to deal with them later on. We will also assign a direction to this single loop, and we will later divide the final answer by two because each loop will be counted twice (because there are two possible directions around a loop). While it may seem to make things more difficult to add directionality to a previously undirected graph, it will actually allow us to make use of known results.

To proceed, we sort the contributions to c_1 according to the winding number w of the loop around the triangle formed by rows 1, 2, and 3, which can now be well defined because we have added directionality to the edges. Once w is fixed, the total numbers of edges in the triangle of each directionality also become fixed. Specifically, as shown in Fig. D.9(b), $x_1 = (e_{13} + w)/2$ is the number of 1-3 edges traversed (i.e., directed) from 3 to 1, $x_2 = (x - w)/2$ is the number of 1-3 edges traversed from 1 to 3, $y_1 = (e_{12} + w)/2$ is the number 1-2 edges traversed from 1 to 2, $y_2 = (e_{12} - w)/2$ is the number of 1-2 edges traversed from 2 to 1, $z_1 = (e_{23} + w)/2$ is the number 2-3 edges traversed from 2 to 3, and $z_2 = (e_{23} - w)/2$ is the number 2-3 edges traversed from 3 to 2. Note that, here, we are treating a positive winding number as going *clockwise* around the graph. There will also be combinatorial factors associated with *which* edges go in which direction, but we will handle that factor later. We are now interested in the number of Eulerian circuits on the resulting directed graph G, i.e., the number of directed closed paths that visit each edge exactly once. The BEST theorem [193] says that the number of such Eulerian circuits is

$$ec(G) = t_u(G) \prod_{v \in V} (\deg(v) - 1)!, \tag{D.65}$$

where $V = \{1, 2, 3\}$ is the set of 3 vertices of our graph, $\deg(v)$ is the indegree of vertex v, and $t_u(G)$ is the number of arborescences of G with root u, i.e., the number of directed tree subgraphs of G such that, for any vertex v, there is exactly one directed path from v to u. If the graph G has an Eulerian circuit, it is known that $t_u(G)$ is independent of the choice of u. Choosing $u = 3$, the three types of trees (arborescences) contributing to $t_u(G)$ for the graph G in Fig. D.9(b) are shown in Fig. D.9(c). The result is $t_u(G) = x_2 z_1 + y_2 x_2 + y_1 z_1$. The term $x_2 z_1$ [corresponding to the first graph in Fig. D.9(c)] counts trees (arborescences) made up of a $1 \to 3$ edge and a $2 \to 3$ edge; the term $y_2 x_2$ [corresponding to the second graph in Fig. D.9(c)] counts trees made up of a $2 \to 1$ edge and a $1 \to 3$ edge; and the term $y_1 z_1$ [corresponding to the third graph in Fig. D.9(c)] counts trees made up of a $1 \to 2$ edge and a $2 \to 3$ edge. Plugging in the definitions of x_i, y_i, and z_i, we find $t_u(G) = (w^2 + e_{12}e_{13} + e_{13}e_{23} + e_{23}e_{12})/4$. Therefore,

$$\begin{aligned} ec(G) &= \frac{1}{4}(w^2 + e_{12}e_{13} + e_{13}e_{23} + e_{23}e_{12}) \left(\frac{e_{12} + e_{13}}{2} - 1\right)! \left(\frac{e_{12} + e_{23}}{2} - 1\right)! \left(\frac{e_{13} + e_{23}}{2} - 1\right)! \\ &= \frac{1}{4}\left(w^2 + 3n^2 - (p_1 - p_4)^2 - 2n(p_1 + p_4)\right)(n - p_1 - 1)!(n - 1)!(n - p_4 - 1)!. \end{aligned} \tag{D.66}$$

We now include the aformentioned combinatorial factors that account for which edges receive which directionality. When choosing which x_1 of the e_{13} edges to make into $3 \to 1$ edges, we pick up a combinatorial factor of $\binom{e_{13}}{x_1} = \binom{n-p_1-p_4}{(n-p_1-p_4+w)/2}$. Similarly for e_{12} and e_{23}: $\binom{e_{12}}{y_1} = \binom{n-p_1+p_4}{(n-p_1+p_4+w)/2}$ and $\binom{e_{23}}{z_1} = \binom{n+p_1-p_4}{(n+p_1-p_4+w)/2}$.

We can now also account for the fact that e_{11} and e_{33} may actually be nonzero. We keep G defined as before (i.e. using only 1-2, 1-3, and 2-3 edges), but we now dress the loops defined on G (and counted above) with additional 1-1 and 3-3 edges. The number of times our loop visits vertex 1 is given by $\deg(1) = n - p_1$, so we need to sort $e_{11} = p_1$ edges into $n - p_1$ buckets, which gives a factor of $\binom{e_{11}+n-p_1-1}{e_{11}} = \binom{n-1}{p_1}$ (by the standard "stars and bars" argument). Similarly, $e_{33} = p_4$ loops give $\binom{e_{33}+n-p_4-1}{e_{33}} = \binom{n-1}{p_4}$. Because all $e_{11} = p_1$ edges are distinguishable and can be traversed in two different ways, we also get a factor of $p_1!2^{p_1}$ (that is, after the bucket counts are decided, we still have to order the edges and assign each a direction). We similarly get a factor of $p_4!2^{p_4}$. Putting all these elements together, we have

$$\begin{aligned}
c_1 &= \sum_{p_1=0}^{n} \sum_{p_4=0}^{n-p_1} \binom{n}{p_1}\binom{n-p_1}{p_4} 2^p d_1(\boldsymbol{e}) \\
&= \sum_{p_1=0}^{n-1} \sum_{p_4=0}^{n-\max(p_1,1)} \binom{n}{p_1}\binom{n-p_1}{p_4} 2^p \sum_w \frac{1}{8}\left(w^2 + 3n^2 - (p_1-p_4)^2 - 2n(p_1+p_4)\right)(n-p_1-1)! \\
&\quad \times (n-1)!(n-p_4-1)!\binom{n-p_1-p_4}{(n-p_1-p_4+w)/2}\binom{n-p_1+p_4}{(n-p_1+p_4+w)/2}\binom{n+p_1-p_4}{(n+p_1-p_4+w)/2} \\
&\quad \times \binom{n-1}{p_1}\binom{n-1}{p_4} p_1!2^{p_1} p_4!2^{p_4} \\
&= n!((n-1)!)^3 2^{n-3} \sum_{p_1=0}^{n} \sum_{p_4=0}^{n-p_1} \sum_{w=-n+p_1+p_4}^{n-p_1-p_4} \\
&\quad \frac{\binom{n-p_1+p_4}{(n-p_1+p_4+w)/2}\binom{n+p_1-p_4}{(n+p_1-p_4+w)/2}(w^2 + 3n^2 - (p_1-p_4)^2 - 2n(p_1+p_4))}{p_1!p_4!((n-p_1-p_4-w)/2)!((n-p_1-p_4+w)/2)!}.
\end{aligned} \tag{D.67}$$

In the second equality, we have introduced an extra factor of $1/2$ because we counted every loop twice because of the two directions in which each loop can be traversed. In the second equality, we also excluded the cases where all black edge sets are of type-1 ($p_1 = n$) and where all black edge sets are of type-4 ($p_4 = n$), as there is no single-loop contribution in this case (allowing for $p_1 = n$ would make $\binom{n-1}{p_1}$ undefined; similarly for $p_4 = n$). In the last equality, to simplify the expression, we allow $p_1 = n$ and $p_4 = n$ because the corresponding contribution is now well-defined and vanishes anyway. In the last equality, the sum over w runs in increments of 2 due to a parity constraint (flipping the directionality of a single edge actually changes the winding number by 2). While one can evaluate the sum over w in the final expression in Eq. (D.67) in terms of hypergeometric functions, we were not able to then evaluate the remaining sums over p_4 and p_1 to obtain a closed-form expression for c_1.

Numerical evaluation of the final expression in Eq. (D.67) agrees with the evaluation of c_1 using the recursive method in the main text up to $n = 40$ (which is the largest n we apply the latter method to). The final expression in Eq. (D.67) is, however, so simple that it can easily be evaluated for much larger values of n. For example, Mathematica [190] on a personal computer evaluates it for $n = 200$ in about 15 seconds. One can also use Eq. (D.67) to study in detail the asymptotic dependence of c_1 on n. We also hope that the method introduced in this Appendix can yield other useful analytical results about c_i.

Appendix E: Appendices Associated with Chapter 6

In this Appendix, we provide more details for the algorithm simulating MBL Hamiltonians evolved only for times at most logarithmic in the system size (Appendix E.1), and we give mathematical proofs of Eqs. (6.10) to (6.12) deferred from the main text for clarity (Appendix E.2).

E.1 Logarithmic Time Simulation

In this Appendix, we give more details for a strong simulation algorithm for MBL Hamiltonians evolved for at most logarithmic times. As discussed in the main text, if the Hamiltonian H is finite-range in the physical basis, Ref. [134] provides an efficient representation of the propagator e^{-iHt} for evolution time logarithmic in the system size N:

Theorem E.1. *[Ref. [134]] Assuming H is finite-range in the physical basis, then one can construct an approximation $\tilde{U}$ to the propagator $U = e^{-iHt}$ such that $\left\|U - \tilde{U}\right\| \leq \epsilon$ and $\tilde{U}$ may be computed with classical resources that are polynomial in N and $1/\epsilon$ and exponential in $|t|$.*

We have that for some initial state $|\phi\rangle$, $\left\|U|\phi\rangle - \tilde{U}|\phi\rangle\right\|_2 = \left\|U - \tilde{U}\right\| \leq \epsilon$. Thus, approximate simulation of $U|\phi\rangle$ can be solved by exactly simulating $\tilde{U}|\phi\rangle$. As constructed in Ref. [134], $\tilde{U}$ is described in the matrix product operator formalism, which means that we have an algorithm that solves the problem of strong simulation for evolution of a product state under $\tilde{U}$. This is because products of local observables admit a trivial matrix product operator formulation (as there is no correlation between the operators, a product of local observables is a matrix product operator with zero bond dimension). Because multiplication between reasonably sized matrix product operators is efficient, it is possibly to efficiently evaluate $\left\langle e^{it\tilde{H}}(\prod_i O_i)e^{-it\tilde{H}}\right\rangle$. As described in the main text, this also implies a sampling algorithm from the approximate distribution generated by measuring the initial state evolved under $\tilde{H}$ for time t:

Corollary E.1.1. *Provided H is finite range in the physical basis, Problem 6.1 is easy for $t = \mathcal{O}(\log N)$.*

The assumption that H is finite-range in the physical basis is a technical one, but one that is reasonable, as many physical systems that are candidates for MBL, such as the disordered, short-range Ising model, fulfill such restrictions. Note, however, that finite-range Hamiltonians can also describe thermalizing systems. Thus, this result importantly establishes that there is a regime in which (many classes of) MBL systems admit sampling algorithms, but it does not use any of the salient features of MBL in order to distinguish it from the thermalizing phase.

E.2 Mathematical Details

Here we will present mathematical details deferred from the main text for clarity. Lemma E.1 bounds the difference between the full and approximate LIOMs discussed in the main text. Lemma E.2 places a bound on the sum S_{p,n_0}. Lemma E.3 [194] provides an intermediate result regarding the incomplete Gamma function that is useful in proving the bound on S_{p,n_0}. Lemma E.4 applies Lemma E.1 and Lemma E.2 in order to bound the operator norm of the difference between the full and truncated Hamiltonians.

Lemma E.1 *Let H be an MBL Hamiltonian with localization length $\xi < 1/\log 2$. Let U be a quasilocal unitary with localization length ξ as in Definition 6.1 such that U diagonalizes H, and let $\tilde{U}$ be U's truncation to constituents of range less than or equal to $r_U = 2a\xi \log N$ for some constant $a > 1$. Finally, let $\tau_i^\alpha = U\sigma_i^\alpha U^\dagger$ and $\tilde{\tau}_i^\alpha = \tilde{U}\sigma_i^\alpha \tilde{U}^\dagger$. For large enough system sizes N, it follows that*

$$\|\tau_i^z - \tilde{\tau}_i^z\| \leq 8\sqrt{q}Ne^{-\frac{r_U}{2\xi}}, \tag{E.1}$$

where $\|\cdot\|$ is the operator norm.

Proof. Let $U = U'\tilde{U}$, where

$$\tilde{U} = \prod_{n=1}^{r_U} \prod_{j=1}^{n} \prod_{i=0}^{\lfloor (N-n)/n \rfloor} U_{in+j}^{(n)}, \tag{E.2}$$

$$U' = \prod_{n=r_U+1}^{N} \prod_{j=1}^{n} \prod_{i=0}^{\lfloor (N-n)/n \rfloor} U_{in+j}^{(n)}. \tag{E.3}$$

Write $U_{in+j}^{(n)} = \mathbb{1} + \Delta_{in+j}^{(n)}$, where we use $\mathbb{1}$ to denote the identity operator on the appropriate Hilbert space. Definition 6.1 tells us that $\left\|\Delta_{in+j}^{(n)}\right\| < \sqrt{q}e^{-\frac{n-1}{2\xi}}$. Also write $\tilde{U} = \mathbb{1} + \tilde{\Delta}$ and, similarly, $U' = \mathbb{1} + \Delta'$ such that:

$$\|\Delta'\| = \left\| \prod_{n=r_U+1}^{N} \prod_{j=1}^{n} \prod_{i=0}^{\lfloor (N-n)/n \rfloor} (\mathbb{1} + \Delta_{in+j}^{(n)}) - \mathbb{1} \right\| \tag{E.4}$$

$$\|\tilde{\Delta}\| = \left\| \prod_{n=1}^{r_U} \prod_{j=1}^{n} \prod_{i=0}^{\lfloor (N-n)/n \rfloor} (\mathbb{1} + \Delta_{in+j}^{(n)}) - \mathbb{1} \right\|. \tag{E.5}$$

We now have that

$$\begin{aligned} \|\tau_i^z - \tilde{\tau}_i^z\| &= \left\|U'\tilde{\tau}_i^z(U')^\dagger - \tilde{\tau}_i^z\right\| \tag{E.6} \\ &= \left\|\Delta'\tilde{\tau}_i^z + \tilde{\tau}_i^z(\Delta')^\dagger + \Delta'\tilde{\tau}_i^z(\Delta')^\dagger\right\| \tag{E.7} \\ &\leq \|\Delta'\| + \left\|(\Delta')^\dagger\right\| + \|\Delta'\|\left\|(\Delta')^\dagger\right\|. \tag{E.8} \end{aligned}$$

Define a multi-index parameter $\alpha = (i,j,n) = (k,n)$ where $k = in + j$ specifies the left-most site of an n-site unitary. We may then rewrite Eq. (E.4):

$$\Delta' = \left[\prod_\alpha (1 + \Delta_\alpha)\right] - 1 = \left[\sum_S \prod_{\alpha\in S} (\Delta_\alpha)\right] - 1 = \sum_{S\neq\varnothing} \prod_{\alpha\in S} \Delta_\alpha, \tag{E.9}$$

where S is a subset of the possible α indices. The triangle inequality and submultiplicativity yield

$$\|\Delta'\| < \left[\sum_S \prod_{\alpha\in S} \left(\sqrt{q}e^{-\frac{n_\alpha - 1}{2\xi}}\right)\right] - 1 = \left[\sum_S \prod_\alpha \left(e^{-\left[\frac{(n_\alpha-1)}{2\xi} - \frac{\log q}{2}\right]\mathbb{I}(\alpha\in S)}\right)\right] - 1, \tag{E.10}$$

where the indicator $\mathbb{I}(x)$ is 1 (0) if x is true (false), and n_α is the size of the unitary indexed by α. To evaluate this, we switch the sum and product. In particular, instead of using the indicator and summing over subsets S, we can instead view the sum as a sum over all α where n can take either the value 0 or n_α. Define A to be the number of possible α. Then

$$\left[\sum_S \prod_\alpha \left(e^{-[\frac{(n_\alpha-1)}{2\xi}-\frac{\log q}{2}]\mathbb{I}(\alpha\in S)}\right)\right]-1=\left[\sum_{\alpha_1=\{0,n_1-1\}}\cdots\sum_{\alpha_A=\{0,n_A-1\}}\left(e^{-\frac{\alpha_1}{2\xi}+\frac{\alpha_1\log q}{2(n_1-1)}}\dots e^{-\frac{\alpha_A}{2\xi}+\frac{\alpha_A\log q}{2(n_A-1)}}\right)\right]-1 \tag{E.11}$$

$$=\left[\prod_\alpha \sum_{n=\{0,n_\alpha-1\}}\left(e^{-\frac{n}{2\xi}+\frac{n\log q}{2(n_\alpha-1)}}\right)\right]-1 \tag{E.12}$$

$$=\left[\prod_\alpha\left(1+\sqrt{q}e^{-\frac{n_\alpha-1}{2\xi}}\right)\right]-1. \tag{E.13}$$

We rewrite the infinite product as the exponential of an infinite sum:

$$\prod_\alpha\left(1+\sqrt{q}e^{-\frac{n_\alpha-1}{2\xi}}\right)-1=e^{\sum_\alpha \log\left(1+\sqrt{q}e^{-\frac{n_\alpha-1}{2\xi}}\right)}-1. \tag{E.14}$$

We now examine the sum:

$$\sum_\alpha \log\left(1+\sqrt{q}e^{-\frac{n_\alpha-1}{2\xi}}\right)=\sum_{n>r_U}\sum_k \log\left(1+\sqrt{q}e^{-\frac{n-1}{2\xi}}\right). \tag{E.15}$$

For any given n (which labels the number of sites on which the block acts nontrivially), there are $N-n+1$ possible unitaries (the left-most site can be any besides the last $n-1$). We trivially upper bound this by N such that

$$\sum_{n>r_U}\sum_k \log\left(1+\sqrt{q}e^{-\frac{n-1}{2\xi}}\right)<\sum_{n>r_U} N\sqrt{q}e^{-\frac{n-1}{2\xi}}=\frac{N\sqrt{q}}{1-e^{-\frac{1}{2\xi}}}e^{-\frac{r_U}{2\xi}}. \tag{E.16}$$

Let $r_U = 2a\xi \log N$ for some $a > 1$. Plugging back into Eq. (E.14) yields

$$\|\Delta'\| < \exp\left(\frac{\sqrt{q}}{1-e^{-\frac{1}{2\xi}}} N^{1-a}\right) - 1 < \sqrt{q}\frac{(1+o(1))}{1-\frac{1}{\sqrt{2}}} N^{1-a} \tag{E.17}$$

for large enough N. In the above, we have used $\xi < 1/\log 2$. Plugging this result back into Eq. (E.8) yields the result:

$$\|\tau_i^z - \tilde{\tau}_i^z\| \leq 8\sqrt{q}N^{1-a} = 8\sqrt{q}Ne^{-\frac{r_U}{2\xi}} \tag{E.18}$$

for large enough N. □

Lemma E.2 *Assuming $\xi < \frac{1}{\log 2}$, we may prove two bounds. First*

$$S_{p,n_0} = \sum_{n=n_0}^{\infty} \binom{n}{p} e^{-\frac{n}{\xi}} \leq C \begin{cases} e^{-\frac{n_0}{\xi}} & p = 0 \\ pe^{-ap} & n_0 < n_*, p > 0 \\ \frac{n_0^{p+1}\sqrt{p}}{p!} e^{-\frac{n_0}{\xi}} & n_0 \geq n_*, p > 0 \end{cases}, \tag{E.19}$$

where $a := \log(e^{1/\xi} - 1)$, $n_* := p\frac{e^{1/\xi}}{e^{1/\xi}-1} = p(1-e^{-1/\xi})^{-1}$, and $C = 10.8$.

And, for $0 \leq x_1 \leq x_2 \leq n_0$:

$$\sum_{p=x_1}^{x_2} S_{p,n_0} = \sum_{p=x_1}^{x_2} \sum_{n=n_0}^{\infty} \binom{n}{p} e^{-\frac{n}{\xi}} \leq \frac{1}{1-e^{-\kappa}} e^{-\kappa n_0}, \tag{E.20}$$

where $\kappa = \frac{1}{\xi} - \log 2$.

Proof. The proof of the second bound is straightforward. We simply upper bound the sum over

p of $\binom{n}{p}$ as 2^n. We then have that

$$\sum_{p=x_1}^{x_2} S_{p,n_0} \le \sum_{n=n_0}^{\infty} e^{n(-1/\xi+\log 2)}, \tag{E.21}$$

from which the result follows from exactly summing the geometric series, which converges as long as $\xi < \frac{1}{\log 2}$. We now move on to the more complicated case that retains the p-dependence.

Case 1 ($p = 0$): The $p = 0$ case is a straightforward geometric series and the constant out front can be chosen to be anything greater than $\frac{1}{1-e^{-1/\xi}} < 2$ (as $\xi < \frac{1}{\log 2}$).

Case 2 ($n_0 < n_*$): We begin with Stirling's Approximation, which says that:

$$\sqrt{\frac{2\pi}{e^4}}\sqrt{\frac{n}{p(n-p)}}\frac{n^n}{p^p(n-p)^{n-p}} \le \binom{n}{p} \le \frac{e}{2\pi}\sqrt{\frac{n}{p(n-p)}}\frac{n^n}{p^p(n-p)^{n-p}}. \tag{E.22}$$

Applying the upper bound we see that

$$\sum_{n=n_0}^{\infty}\binom{n}{p}e^{-\frac{n}{\xi}} \le \sum_{n=n_0}^{\infty}\frac{e}{2\pi}\sqrt{\frac{n}{p(n-p)}}\frac{n^n}{p^p(n-p)^{n-p}}e^{-\frac{n}{\xi}}. \tag{E.23}$$

We now note that for $n \ge p+1 > 1$, $\sqrt{\frac{n}{p(n-p)}} \le \sqrt{2}$ such that $\frac{e}{2\pi}\sqrt{\frac{n}{p(n-p)}} \le 1$. Then, for $n_0 \ge p+1$,

$$S_{p,n_0} \le \sum_{n=n_0}^{\infty}\frac{n^n}{p^p(n-p)^{n-p}}e^{-\frac{n}{\xi}} = \sum_{n=n_0}^{\infty} e^{-\frac{n}{\xi}+n\log n-(n-p)\log(n-p)-p\log p} := \sum_{n=n_0}^{\infty} e^{g(n)}. \tag{E.24}$$

We can eliminate the $n \ge p+1$ assumption by realizing that the final bound in Eq. (E.24) still holds trivially when $n = p$, as the logarithmic terms in $g(n)$ vanish. Thus, Eq. (E.24) is valid for all pairs $n_0 \ge p$, which we assume in order to make the combinatorial factor $\binom{n}{p}$ well-defined.

Maximizing the summand means maximizing $g(n)$, so we calculate:

$$\frac{\partial g}{\partial n} = -\frac{1}{\xi} + \log n - \log(n-p), \tag{E.25}$$

$$\frac{\partial^2 g}{\partial n^2} = \frac{1}{n} - \frac{1}{n-p}. \tag{E.26}$$

It is straightforward to calculate

$$g'(n) \begin{cases} = 0 & n = n_* = p\frac{e^{1/\xi}}{e^{1/\xi}-1} \\ > 0 & n < n_* \\ < 0 & n > n_* \end{cases}. \tag{E.27}$$

Furthermore, it is also straightforward to verify $g''(n) < 0$ for all $n > p$. Thus, we see that $g(n)$, and hence $e^{g(n)}$, has a single maximum on $[n_0, \infty)$; it is at n_* for $n_0 < n_*$ and n_0 for $n_0 \geq n_*$. Additionally, it will be useful to calculate that $g(n_*) = -ap$, where $a := \log(e^{1/\xi} - 1)$.

We now bound the final sum in Eq. (E.24) with an integral using a Riemann approximation. In particular, let $n_*^- = \lfloor n_* \rfloor$ and $n_*^+ = n_*^- + 1$. Then

$$\sum_{n=n_0}^{\infty} e^{g(n)} = e^{g(n_*^-)} + e^{g(n_*^+)} + \sum_{n=n_0}^{n_*^- -1} e^{g(n)} + \sum_{n=n_*^+ +1}^{\infty} e^{g(n)} \tag{E.28}$$

$$\leq 2e^{g(n_*)} + \int_{n_0}^{n_*^-} e^{g(n)} dn + \int_{n_*^+}^{\infty} e^{g(n)} dn \tag{E.29}$$

$$\leq 2e^{-ap} + \int_{n_0}^{n_*} e^{g(n)} dn + \int_{n_*}^{\infty} e^{g(n)} dn \tag{E.30}$$

$$= 2e^{-ap} + \int_{n_0}^{n_*} e^{g(n)} dn + \int_{n_*}^{2n_*} e^{g(n)} dn + \int_{2n_*}^{\infty} e^{g(n)} dn \tag{E.31}$$

$$:= 2e^{-ap} + I_< + I_{<>} + I_>, \tag{E.32}$$

Consider first $I_<$. There we can start by using that $g(n_*)$ is maximal to make the trivial bound:

$$I_< \leq (n_* - n_0)e^{g(n_*)} \leq pe^{-a(p+1)}, \tag{E.33}$$

where we have used that $n_* - n_0 \leq n_* - p = pe^{-a}$. Similarly, for $I_{<>}$, we may say that

$$I_{<>} \leq n_* e^{g(n_*)} \leq 2pe^{-ap}, \tag{E.34}$$

where we have used that $p < n_* < 2p$ because $1/\xi > \log 2$.

To bound $I_>$, we first invert the Stirling approximation from earlier and write:

$$e^{g(n)} = e^{-\frac{n}{\xi}} \frac{n^n}{p^p(n-p)^{(n-p)}} \leq \frac{e^2}{\sqrt{2\pi}} \sqrt{\frac{p(n-p)}{n}} \binom{n}{p} e^{-\frac{n}{\xi}} \leq 3\sqrt{p}\binom{n}{p}e^{-\frac{n}{\xi}} \leq 3\sqrt{p}\frac{n^p}{p!}e^{-\frac{n}{\xi}}. \tag{E.35}$$

We can thus bound

$$I_> \leq 3\frac{\sqrt{p}}{p!}\int_{2n_*}^{\infty} e^{-\frac{n}{\xi}} n^p dn. \tag{E.36}$$

Substituting $u = \frac{n}{\xi}$ and defining $u_* = \frac{n_*}{\xi}$ yield

$$I_> \leq 3\frac{\xi^{p+1}\sqrt{p}}{p!}\int_{2u_*}^{\infty} e^{-u}u^p du = 3\frac{\xi^{p+1}\sqrt{p}}{\Gamma(p+1)}\Gamma(p+1, 2u_*), \tag{E.37}$$

where $\Gamma(a)$ and $\Gamma(a, z)$ are the standard Gamma and Incomplete Gamma functions, respectively. We can bound the Incomplete Gamma Function using Lemma E.3 provided $2u_* > p$, i.e. $2\frac{e^{1/\xi}}{e^{1/\xi}-1} > \xi$. We can actually do better and show that $u_* > p$. Defining $x = 1/\xi$, we want to show $xe^x - e^x + 1 > 0$ for $x \in (0, \infty)$. At $x = 0$, the LHS is 0. Taking a derivative of the LHS with respect to x yields $xe^x > 0$ for $x \in (0, \infty)$. Thus, the LHS is 0 at $x = 0$ and increasing, which means the inequality holds. With that in mind, we apply Lemma E.3:

$$I_> \leq 3\frac{\xi^{p+1}\sqrt{p}}{p!}\int_{2u_*}^{\infty} e^{-u}u^p du \leq 3\frac{\xi^{p+1}\sqrt{p}}{p!}\frac{(2u_*)^{p+1}e^{-2u_*}}{2u_* - p} = 3(2^{p+1})\frac{n_*^{p+1}e^{-\frac{2n_*}{\xi}}}{\sqrt{p}p!}\frac{1}{2\frac{1}{\xi}(\frac{e^{1/\xi}}{e^{1/\xi}-1}) - 1}. \tag{E.38}$$

Note that

$$\begin{aligned} 2^{p+1}\frac{n_*^{p+1}e^{-\frac{2n_*}{\xi}}}{\sqrt{p}p!} = \frac{(2p)^{p+1}}{\sqrt{p}p!}\left(\frac{e^{1/\xi}}{e^{1/\xi}-1}\right)^{p+1} e^{-2\frac{p}{\xi}\frac{e^{1/\xi}}{e^{1/\xi}-1}} &\leq \frac{2}{\sqrt{2\pi}}e^{p(1+\log 2)}e^{-a(p+1)}e^{\frac{p+1}{\xi}}e^{-2\frac{p}{\xi}\frac{e^{1/\xi}}{e^{1/\xi}-1}} \\ &\leq \sqrt{\frac{2}{\pi}}e^{-ap}\underbrace{e^{-a+p(1+\log 2)+\frac{p+1}{\xi}-\frac{2p}{\xi}\frac{e^{1/\xi}}{e^{1/\xi}-1}}}_{\leq e^{\log 2}} \leq \sqrt{\frac{8}{\pi}}e^{-ap}. \end{aligned} \tag{E.39}$$

The last bound is rather involved, so we will explain the steps carefully. We want to show that

$$-a + p(1+\log 2) + \frac{p+1}{\xi} - \frac{2p}{\xi}\frac{e^{1/\xi}}{e^{1/\xi}-1} = \underbrace{p(1+\log 2) + \frac{p}{\xi} - \frac{2p}{\xi}\frac{e^{1/\xi}}{e^{1/\xi}-1}}_{\text{(A)}} + \underbrace{\frac{1}{\xi} - a}_{\text{(B)}} < \log 2. \tag{E.40}$$

We can show (A) < 0 using a strategy similar to when we proved that our bound on the incomplete gamma function was valid. In particular, first note that we can effectively cancel $\frac{p}{\xi}$ with one factor of $\frac{p}{\xi}\frac{e^{1/\xi}}{e^{1/\xi}-1}$ given that $\xi < \log 2$. We then want to show that $p(1+\log 2) - px\frac{e^x}{e^x-1} < 0$, where, again, $x = \frac{1}{\xi}$. Equivalently, we want to show that $xe^x - (1+\log 2)e^x + (1+\log 2) > 0$. Again, the LHS is 0 at $x = 0$. And, again, taking a derivative of the LHS gives us $xe^x + e^x - (1+\log 2)e^x = xe^x - \log 2e^x$, which is greater than 0 as long as $x > \log 2$, or $\xi < \frac{1}{\log 2}$. We then want to bound (B), and this is done by noting that in the limit that ξ is very small, then $a = \log(e^{1/\xi} - 1) \sim 1/\xi$ such that (B) ~ 0. In fact, the maximum of (B) is simply $\log 2$, which occurs for $\xi = \frac{1}{\log 2}$. With all of that handled, we can then say that the final bound in Eq. (E.39) is exponentially decreasing with p only if $a > 0$, which corresponds to $\xi < \frac{1}{\log 2}$ or $1/\xi > \log 2$.

We need to combine the bounds on all of the components of Eq. (E.32):

$$2e^{-ap} + I_< + I_{<>} + I_> \leq \left(2 + pe^{-a} + 2p + 3\sqrt{\frac{8}{\pi}}\frac{1}{2\frac{1}{\xi}\left(\frac{e^{1/\xi}}{e^{1/\xi}-1}\right) - 1}\right)e^{-ap} \tag{E.41}$$

$$\leq \left(\frac{2}{p} + e^{-a} + 2 + \frac{3}{p}\sqrt{\frac{8}{\pi}}\frac{1}{4\log 2 - 1}\right)pe^{-ap} \tag{E.42}$$

$$\leq C_1 pe^{-ap}, \tag{E.43}$$

where we have used the fact that $1/\xi > \log 2$ and defined

$$C_1 = 5 + 3\sqrt{\frac{8}{\pi}}\frac{1}{4\log 2 - 1} < 7.8. \tag{E.44}$$

Case 3 ($n_0 \geq n_*$): For sufficiently small ξ, we have $n_* \sim p$. Assuming $n_0 \geq p + 1$, this means that $n_0 > n_*$. However, given that situation, we can use that $g(n)$ is decreasing after n_0 to go immediately from Eq. (E.24) to

$$\sum_{n=n_0}^{\infty} e^{g(n)} \leq e^{g(n_0)} + \int_{n_0}^{\infty} e^{g(n)} dn. \tag{E.45}$$

In comparison with the case where $n_* < n_0$, the integral $I_<$ effectively does not exist here, and the bound in $I_>$ comes from simply replacing $2n_*$ with n_0 (which is now the maximal contribution) and adding on the extra term in Eq. (E.45). First, using steps nearly identical to those above (noting in particular that Lemma E.3 is valid because $n_0 > n_* > p\xi$ by the earlier proof), we can bound

$$I_> \leq \frac{3}{2\log 2 - 1}\frac{n_0^{p+1}e^{-\frac{n_0}{\xi}}}{\sqrt{p}p!}. \tag{E.46}$$

Then, the contribution from $e^{g(n_0)}$ may be bounded by inverting Stirling's approximation as in Eq. (E.35):

$$e^{g(n_0)} \leq 3\frac{n_0^p\sqrt{p}}{p!}e^{-\frac{n_0}{\xi}}. \tag{E.47}$$

Combining the two yields

$$\begin{aligned} e^{g(n_0)} + I_> &\leq \left(\frac{3}{2\log 2 - 1} + 3\right)\frac{n_0^{p+1}\sqrt{p}}{p!}e^{-\frac{n_0}{\xi}} && \text{(E.48)} \\ &= C_2\frac{n_0^{p+1}\sqrt{p}}{p!}e^{-\frac{n_0}{\xi}}, && \text{(E.49)} \end{aligned}$$

where

$$C_2 = \left(\frac{3}{2\log 2 - 1} + 3\right) < 10.8. \tag{E.50}$$

□

Lemma E.3 (Ref. [194]) *Let $\Gamma(a,z)$ be the Incomplete Gamma Function defined in the standard way:*

$$\Gamma(a,z) = \int_z^\infty e^{-x}x^{a-1}dx. \tag{E.51}$$

Let $z \in \mathbb{R} > (a-1)$. Then

$$\Gamma(a,z) \leq \frac{z^a e^{-z}}{z-(a-1)}. \tag{E.52}$$

Proof. Make the substitution $s = \frac{x}{z} - 1$. Then

$$\Gamma(a,z) = \int_0^\infty e^{-(s+1)z}z^a(1+s)^{a-1}ds = z^a e^{-z}\int_0^\infty e^{-sz}(1+s)^{a-1}ds. \tag{E.53}$$

From here, $(1+s) \leq e^s$ implies that

$$\Gamma(a,z) \leq z^a e^{-z} \int_0^\infty e^{-sz} e^{(a-1)s} ds = \frac{z^a e^{-z}}{-z+a-1} e^{-(z-(a-1))s} \Big|_{s=0}^{\infty} = \frac{z^a e^{-z}}{z-(a-1)}, \tag{E.54}$$

as long as $z > a-1$ so that the upper limit actually vanishes. □

Lemma E.4 *The difference between the truncated and true Hamiltonian obeys*

$$\left\| H - \tilde{H} \right\| \leq C_U N^2 e^{-\frac{r_U}{2\xi}} + C_J N r_J e^{-k r_J}. \tag{E.55}$$

Proof. A straightforward application of the triangle inequality yields

$$\left\| H - \tilde{H} \right\| \leq \sum_I \left| (J_I - \tilde{J}_I) \right| + \left| \tilde{J}_I \right| \| (\tau_I^z - \tilde{\tau}_I^z) \|. \tag{E.56}$$

Recall that the truncated coefficients $\tilde{J}_I$ are 0 beyond range r_J). In the sum below, the symbol p represents how many sites are coupled by J. That is, the relevant term is $J_{i_1,\ldots i_p}$, a p-body term. The symbol ℓ denotes the maximum distance between any two sites coupled by a term of this form, given by $\ell = |i_1 - i_p|$. The first term of Eq. (E.56) may be bounded as follows:

$$
\begin{aligned}
\sum_I \left| J_I - \tilde{J}_I \right| &\le \sum_{p=2}^{r_J} N \sum_{\ell=r_J}^{\infty} \binom{\ell-1}{p-2} e^{-\frac{\ell}{\xi}} + \sum_{p=r_J+1}^{\infty} N \sum_{\ell=p-1}^{\infty} \binom{\ell-1}{p-2} e^{-\frac{\ell}{\xi}} && \text{(E.57)}\\
&\le N \sum_{p=0}^{r_J-2} S_{p,r_J-1} e^{-1/\xi} + N \sum_{p=r_J-1}^{\infty} S_{p,p} e^{-1/\xi} && \text{(E.58)}\\
&\le \frac{Ne^{-1/\xi}}{1-e^{-\kappa}} e^{-\kappa(r_J-1)} + CNe^{-1/\xi} \sum_{p=r_J-1}^{\infty} pe^{-ap} && \text{(E.59)}\\
&\le \frac{Ne^{-1/\xi}}{1-e^{-\kappa}} e^{-\kappa(r_J-1)} + CNe^{-1/\xi} \left[\frac{(r_J-1)e^{-a(r_J-1)}}{1-e^{-a}} + \frac{e^{-a-a(r_J-1)}}{(1-e^{-a})^2} \right] && \text{(E.60)}\\
&= \frac{Ne^{-1/\xi}}{1-e^{-\kappa}} e^{-\kappa(r_J-1)} + CNe^{-1/\xi} \frac{e^{-ar_J}}{(1-e^{-a})^2} \times \left(e^a(r_J-1) - r_J + 2 \right) && \text{(E.61)}\\
&\le c_1 N e^{-\kappa r_J} + c_2 N r_J e^{-a r_J} && \text{(E.62)}\\
&\le C_J N r_J e^{-k r_J}, && \text{(E.63)}
\end{aligned}
$$

where

$$
\begin{aligned}
c_1 &= \frac{e^{\kappa} e^{-1/\xi}}{1-e^{-\kappa}} = \frac{1}{2(1-e^{-\kappa})}, && \text{(E.64)}\\
c_2 &= Ce^{-1/\xi} \frac{e^a+1}{(1-e^{-a})^2} = C \frac{1}{(1-e^{-a})^2}, && \text{(E.65)}\\
C &= 10.8. && \text{(E.66)}
\end{aligned}
$$

C_J is a constant that is independent of N but will depend on ξ (directly and through a and κ), and

$$
\begin{aligned}
\kappa &:= \frac{1}{\xi} - \log 2, && \text{(E.67)}\\
a &:= \log\left(e^{1/\xi} - 1 \right), && \text{(E.68)}\\
k &:= \min\left\{ \kappa, a \right\}. && \text{(E.69)}
\end{aligned}
$$

The requirements on both a and κ are the same, $\xi < \frac{1}{\log 2}$.

To bound the second term, we first use a telescoping sum, the triangle inequality, and unitary invariance of the operator norm to show that

$$\|(\tau_I^z - \tilde{\tau}_I^z)\| \leq \sum_{j=1}^{p} \left\|(\tau_{i_j}^z - \tilde{\tau}_{i_j}^z)\right\| \leq 8\sqrt{q}Npe^{-\frac{r_U}{2\xi}}, \tag{E.70}$$

where I is the multi-index $i_1 \dots i_p$. Plugging this back in yields

$$\begin{aligned}
\sum_I \left|\tilde{J}_I\right| \|\tau_I^z - \tilde{\tau}_I^z\| &\leq \sum_{\ell=1}^{r_J-1} N \sum_{p=2}^{\ell+1} \binom{\ell-1}{p-2} 8p\sqrt{q}Ne^{-\frac{r_U}{2\xi}} e^{-\frac{\ell}{\xi}} && \text{(E.71)}\\
&= 8\sqrt{q}e^{-\frac{1}{\xi}}N^2 e^{-\frac{r_U}{2\xi}} \sum_{\ell=0}^{r_J-2} \sum_{p=0}^{\ell} \binom{\ell}{p}(p+2)e^{-\frac{\ell}{\xi}} && \text{(E.72)}\\
&\leq 8\sqrt{q}e^{-\frac{1}{\xi}}N^2 e^{-\frac{r_U}{2\xi}} \sum_{p=0}^{r_J-2} (p+2) \sum_{\ell=p}^{r_J-2} \binom{\ell}{p} e^{-\frac{\ell}{\xi}} && \text{(E.73)}\\
&\leq 8\sqrt{q}e^{-\frac{1}{\xi}}N^2 e^{-\frac{r_U}{2\xi}} \sum_{p=0}^{r_J-2} (p+2)S_{p,p} && \text{(E.74)}\\
&\leq 8\sqrt{q}e^{-\frac{1}{\xi}}N^2 e^{-\frac{r_U}{2\xi}} \sum_{p=0}^{r_J-2} C(p+2)pe^{-ap} && \text{(E.75)}\\
&\leq C_U N^2 e^{-\frac{r_U}{2\xi}} && \text{(E.76)}
\end{aligned}$$

for some constant C_U. In the second-to-last line, we have bounded $S_{p,p}$ using Lemma E.2.

Thus, altogether, we have that:

$$\|\Delta H\| \leq C_U N^2 e^{-\frac{r_U}{2\xi}} + C_J N r_J e^{-kr_J}. \tag{E.77}$$

□

Bibliography

[1] A. Ehrenberg, J. Bringewatt, and A. V. Gorshkov. Minimum-Entanglement Protocols for Function Estimation. *Phys. Rev. Res.*, 5:033228, 2023.

[2] J. Bringewatt, A. Ehrenberg, T. Goel, and A. V. Gorshkov. Optimal Function Estimation with Photonic Quantum Sensor Networks. *Phys. Rev. Res.*, 6:013246, 2024.

[3] P. W. Shor. Algorithms for Quantum Computation: Discrete Logarithms and Factoring. In *Proc. – 35th Annu. IEEE Symp. Found. Comput. Sci. FOCS*, pages 124–134, 1994.

[4] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, Cambridge, 2009.

[5] A. K. Lenstra and H. W. Lenstra, editors. *The Development of the Number Field Sieve*, volume 1554 of *Lect. Notes Math.* Springer, Berlin, Heidelberg, 1993.

[6] J. Watrous. Guest Column: An Introduction to Quantum Information and Quantum Circuits 1. *ACM SIGACT News*, 42(2):52–67, 2011.

[7] E. Bernstein and U. Vazirani. Quantum Complexity Theory. *SIAM J. Comput.*, 26(5):1411–1473, 1997.

[8] R. L. Rivest, A. Shamir, and L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Commun. ACM*, 21(2):120–126, 1978.

[9] Information Technology Laboratory Computer Security Division. Call for Proposals - Post-Quantum Cryptography | CSRC | CSRC. https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-cryptography-standardization/call-for-proposals, 2017.

[10] E. Knill, R. Laflamme, and W. H. Zurek. Resilient Quantum Computation: Error Models and Thresholds. *Proc. R. Soc. Lond.*, 454(1969):365–384, 1998.

[11] D. Aharonov and M. Ben-Or. Fault-Tolerant Quantum Computation With Constant Error Rate. *arXiv:quant-ph/9906129*, 1999.

[12] A. Yu. Kitaev. Fault-Tolerant Quantum Computation by Anyons. *Ann. Phys. (N. Y.)*, 303(1):2–30, 2003.

[13] T. J. Proctor, P. A. Knott, and J. A. Dunningham. Networked Quantum Sensing. *arXiv:1702.04271*, 2017.

[14] 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):042337, 2018.

[15] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig. Distributed Quantum Metrology with Linear Networks and Separable Inputs. *Phys. Rev. Lett.*, 121(4):043604, 2018.

[16] T. J. Proctor, P. A. Knott, and J. A. Dunningham. Multiparameter Estimation in Networked Quantum Sensors. *Phys. Rev. Lett.*, 120(8):080501, 2018.

[17] S. Altenburg and S. Wölk. Multi-Parameter Estimation: Global, Local and Sequential Strategies. *Phys. Scr.*, 94(1):014001, 2019.

[18] J. Rubio, P. A. Knott, T. J. Proctor, and J. A. Dunningham. Quantum Sensing Networks for the Estimation of Linear Functions. *J. Phys. A*, 53(34):344001, 2020.

[19] J. A. Gross and C. M. Caves. One from Many: Estimating a Function of Many Parameters. *J. Phys. A: Math. Theor.*, 54(1):014001, 2021.

[20] D. Triggiani, P. Facchi, and V. Tamma. Heisenberg Scaling Precision in the Estimation of Functions of Parameters in Linear Optical Networks. *Phys. Rev. A*, 104(6):062603, 2021.

[21] C. Oh, L. Jiang, and C. Lee. Distributed Quantum Phase Sensing for Arbitrary Positive and Negative Weights. *Phys. Rev. Res.*, 4(2):023164, 2022.

[22] M. Malitesta, A. Smerzi, and L. Pezzè. Distributed Quantum Sensing with Squeezed-Vacuum Light in a Configurable Array of Mach-Zehnder Interferometers. *Phys. Rev. A*, 108(3):032621, 2023.

[23] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum Metrology. *Phys. Rev. Lett.*, 96(1):010401, 2006.

[24] L. Pezzé and A. Smerzi. Entanglement, Nonlinear Dynamics, and the Heisenberg Limit. *Phys. Rev. Lett.*, 102(10):100401, 2009.

[25] Z. Zhang and Q. Zhuang. Distributed Quantum Sensing. *Quantum Sci. Technol.*, 6(4):043001, 2021.

[26] Y. Xia, Q. Zhuang, W. Clark, and Z. Zhang. Repeater-Enhanced Distributed Quantum Sensing Based on Continuous-Variable Multipartite Entanglement. *Phys. Rev. A*, 99(1):012328, 2019.

[27] S. Aaronson and A. Arkhipov. The Computational Complexity of Linear Optics. *Theory Comput.*, 9(4):143–252, 2013.

[28] D. Hangleiter and J. Eisert. Computational Advantage of Quantum Random Sampling. *Rev. Mod. Phys.*, 95(3):035001, 2023.

[29] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, et al. Quantum Supremacy Using a Programmable Superconducting Processor. *Nature*, 574(7779):505–510, 2019.

[30] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, et al. Quantum Computational Advantage Using Photons. *Science*, 370(6523):1460–1463, 2020.

[31] H.-S. Zhong, Y.-H. Deng, J. Qin, H. Wang, M.-C. Chen, L.-C. Peng, Y.-H. Luo, D. Wu, S.-Q. Gong, H. Su, et al. Phase-Programmable Gaussian Boson Sampling Using Stimulated Squeezed Light. *Phys. Rev. Lett.*, 127(18):180502, 2021.

[32] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, et al. Quantum Computational Advantage with a Programmable Photonic Processor. *Nature*, 606(7912):75–81, 2022.

[33] Y.-H. Deng, Y.-C. Gu, H.-L. Liu, S.-Q. Gong, H. Su, Z.-J. Zhang, H.-Y. Tang, M.-H. Jia, J.-M. Xu, M.-C. Chen, et al. Gaussian Boson Sampling with Pseudo-Photon-Number-Resolving Detectors and Quantum Computational Advantage. *Phys. Rev. Lett.*, 131(15):150601, 2023.

[34] L. Stockmeyer. The Complexity of Approximate Counting. In *Proc. – 15th Annu. ACM Symp. Theory Comput.*, pages 118–126. ACM, 1983.

[35] P. W. Anderson. Absence of Diffusion in Certain Random Lattices. *Phys. Rev.*, 109(5):1492–1505, 1958.

[36] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn. Colloquium: Many-Body Localization, Thermalization, and Entanglement. *Rev. Mod. Phys.*, 91(2):021001, 2019.

[37] J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert. Architectures for Quantum Simulation Showing a Quantum Speedup. *Phys. Rev. X*, 8(2):021010, 2018.

[38] J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert, and N. Yunger Halpern. Linear Growth of Quantum Circuit Complexity. *Nat. Phys.*, 18(5):528–532, 2022.

[39] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum-Enhanced Measurements: Beating the Standard Quantum Limit. *Science*, 306(5700):1330–1336, 2004.

[40] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi. Fisher Information and Multiparticle Entanglement. *Phys. Rev. A*, 85(2):022321, 2012.

[41] G. Tóth. Multipartite Entanglement and High-Precision metrology. *Phys. Rev. A*, 85(2):022322, 2012.

[42] R. Augusiak, J. Kołodyński, A. Streltsov, M. N. Bera, A. Acín, and M. Lewenstein. Asymptotic Role of Entanglement in Quantum Metrology. *Phys. Rev. A*, 94(1):012339, 2016.

[43] G. Tóth and I. Apellaniz. Quantum Metrology from a Quantum Information Science Perspective. *J. of Phys. A: Math. Theor.*, 47(42):424006, 2014.

[44] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola. Quantum-Enhanced Measurements without Entanglement. *Rev. Mod. Phys.*, 90(3):035006, 2018.

[45] A. Luis. Phase-Shift Amplification for Precision Measurements without Nonclassical States. *Phys. Rev. A*, 65(2):025802, 2002.

[46] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J Pryde. Entanglement-Free Heisenberg-Limited Phase Estimation. *Nature*, 450(7168):393–396, 2007.

[47] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac. Improvement of Frequency Standards with Quantum Entanglement. *Phys. Rev. Lett.*, 79(20):3865–3868, 1997.

[48] 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, 2011.

[49] S. Boixo and C. Heunen. Entangled and Sequential Quantum Protocols with Dephasing. *Phys. Rev. Lett.*, 108(12):120402, 2012.

[50] R. Demkowicz-Dobrzański and L. Maccone. Using Entanglement Against Noise in Quantum Metrology. *Phys. Rev. Lett.*, 113(25):250801, 2014.

[51] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia. Generalized Limits for Single-Parameter Quantum Estimation. *Phys. Rev. Lett.*, 98(9):090401, 2007.

[52] S. Boixo, A. Datta, S. T. Flammia, A. Shaji, E. Bagan, and C. M. Caves. Quantum-Limited Metrology with Product States. *Phys. Rev. A*, 77(1):012317, 2008.

[53] T. Tilma, S. Hamaji, W. J. Munro, and K. Nemoto. Entanglement is not a Critical Resource for Quantum Metrology. *Phys. Rev. A*, 81(2):022108, 2010.

[54] G. Gour and R. W. Spekkens. The Resource Theory of Quantum Reference Frames: Manipulations and Monotones. *New J. Phys.*, 10(3):033023, 2008.

[55] I. Marvian and R. W. Spekkens. Extending Noether's Theorem by Quantifying the Asymmetry of Quantum States. *Nat. Comm.*, 5(1):3821, 2014.

[56] I. Marvian and R. W. Spekkens. How to Quantify Coherence: Distinguishing Speakable and Unspeakable Notions. *Phys. Rev. A*, 94(5):052324, 2016.

[57] C. Zhang, B. Yadin, Z.-B. Hou, H. Cao, B.-H. Liu, Y.-F. Huang, R. Maity, V. Vedral, C.-F. Li, G.-C Guo, and D. Girolami. Detecting Metrologically Useful Asymmetry and Entanglement by a Few Local Measurements. *Phys. Rev. A*, 96(4):042327, 2017.

[58] C. W. Helstrom. *Quantum Detection and Estimation Theory*, volume 3. Academic Press, New York, 1976.

[59] S. L. Braunstein and C. M. Caves. Statistical Distance and the Geometry of Quantum States. *Phys. Rev. Lett.*, 72(22):3439, 1994.

[60] S. L. Braunstein, C. M. Caves, and G. J. Milburn. Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance. *Ann. Phys. (N. Y.)*, 247(1):135–173, 1996.

[61] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*, volume 1. Springer, Berlin, 2011.

[62] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov. Heisenberg-Scaling Measurement Protocol for Analytic Functions with Quantum Sensor Networks. *Phys. Rev. A*, 100(4):042304, 2019.

[63] J. Bringewatt, I. Boettcher, P. Niroula, P. Bienias, and A. V. Gorshkov. Protocols for Estimating Multiple Functions with Quantum Sensor Networks: Geometry and Performance. *Phys. Rev. Research*, 3(3):033011, 2021.

[64] T. Qian, J. Bringewatt, I. Boettcher, P. Bienias, and A. V. Gorshkov. Optimal Measurement of Field Properties with Quantum Sensor Networks. *Phys. Rev. A.*, 103(3):L030601, 2021.

[65] A. Fujiwara. Quantum Channel Identification Problem. *Phys. Rev. A*, 63(4):042304, 2001.

[66] J. Liu, H. Yuan, X.-M. Lu, and X. Wang. Quantum Fisher Information Matrix and Multiparameter Estimation. *J. Phys. A: Math. Theor.*, 53(2):023001, 2020.

[67] S.-I. Amari. *Differential-Geometrical Methods in Statistics*. Springer, Berlin, 1985.

[68] Y. Yang, G. Chiribella, and M. Hayashi. Attaining the Ultimate Precision Limit in Quantum State Estimation. *Comm. Math. Phys.*, 368(1):223–293, 2019.

[69] J. Suzuki. Nuisance Parameter Problem in Quantum Estimation Theory: Tradeoff Relation and Qubit Examples. *J. Phys. A: Math. Theor.*, 53(26):264001, 2020.

[70] J. Suzuki, Y. Yang, and M. Hayashi. Quantum State Estimation with Nuisance Parameters. *J. Phys. A: Math. Theor.*, 53(45):453001, 2020.

[71] S. Kimmel, G. H. Low, and T. J. Yoder. Robust Calibration of a Universal Single-Qubit Gate Set via Robust Phase Estimation. *Phys. Rev. A*, 92(6):062315, 2015.

[72] S. Kimmel, G. H. Low, and T. J. Yoder. Erratum: Robust Calibration of a Universal Single-Qubit Gate Set via Robust Phase Estimation [Phys. Rev. A **92**, 062315 (2015)]. *Phys. Rev. A*, 104(6):069901(E), 2021.

[73] F. Belliardo and V. Giovannetti. Achieving Heisenberg Scaling with Maximally Entangled States: An Analytic Upper Bound for the Attainable Root-Mean-Square Error. *Phys. Rev. A*, 102(4):042613, 2020.

[74] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde. Demonstrating Heisenberg-Limited Unambiguous Phase Estimation without Adaptive Measurements. *New J. Phys.*, 11(7):073023, 2009.

[75] M. Hayashi, S. Vinjanampathy, and L. C. Kwek. Resolving Unattainable Cramer–Rao Bounds for Quantum Sensors. *J. Phys. B: At., Mol. Opt. Phys.*, 52(1):015503, 2019.

[76] W. Górecki, R. Demkowicz-Dobrzański, H. M Wiseman, and D. W. Berry. π-Corrected Heisenberg Limit. *Phys. Rev. Lett.*, 124(3):030501, 2020.

[77] J. Farkas. Theorie der Einfachen Ungleichungen.:. *J. Reine Angew. Math.*, 1902(124):1–27, 1902.

[78] N. Dinh and V. Jeyakumar. Farkas' Lemma: Three Decades of Generalizations for Mathematical Optimization. *TOP*, 22(1):1–22, 2014.

[79] M. Piani, M. Cianciaruso, T. R. Bromley, C. Napoli, N. Johnston, and G. Adesso. Robustness of Asymmetry and Coherence of Quantum States. *Phys. Rev. A*, 93(4):042107, 2016.

[80] M. M. Taddei, B. M. Escher, L. Davidovich, and R. L. de Matos Filho. Quantum Speed Limit for Physical Processes. *Phys. Rev. Lett.*, 110(5):050402, 2013.

[81] S. Deffner and S. Campbell. Quantum Speed Limits: From Heisenberg's Uncertainty Principle to Optimal Quantum Control. *J. Phys. A: Math. Theor.*, 50(45):453001, 2017.

[82] L. P. García-Pintos, S. B. Nicholson, J. R. Green, A. del Campo, and A. V. Gorshkov. Unifying Quantum and Classical Speed Limits on Observables. *Phys. Rev. X*, 12(1):011038, 2022.

[83] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein. Quantum Metrology with Nonclassical States of Atomic Ensembles. *Rev. Mod. Phys.*, 90(3):035005, 2018.

[84] Q. Zhuang, Z. Zhang, and J. H. Shapiro. Distributed Quantum Sensing Using Continuous-Variable Multipartite Entanglement. *Phys. Rev. A*, 97(3):032329, 2018.

[85] C. Oh, C. Lee, S. H. Lie, and H. Jeong. Optimal Distributed Quantum Sensing Using Gaussian States. *Phys. Rev. Research*, 2(2):023030, 2020.

[86] Y. Yang, B. Yadin, and Z.-P. Xu. Quantum-Enhanced Metrology with Network States. *arXiv:2307.07758*, 2023.

[87] X. Guo, C. R. Breum, J. Borregaard, S. Izumi, M. V. Larsen, T. Gehring, M. Christandl, J. S. Neergaard-Nielsen, and U. L. Andersen. Distributed Quantum Sensing in a Continuous-Variable Entangled Network. *Nat. Phys.*, 16(3):281–284, 2020.

[88] Y. Xia, W. Li, W. Clark, D. Hart, Q. Zhuang, and Z. Zhang. Demonstration of a Reconfigurable Entangled Radio-Frequency Photonic Sensor Network. *Phys. Rev. Lett.*, 124(15):150502, 2020.

[89] L.-Z. Liu, Y.-Z. Zhang, Z.-D. Li, R. Zhang, X.-F. Yin, Y.-Y. Fei, L. Li, N.-L. Liu, F. Xu, Y.-A. Chen, and J.-W. Pan. Distributed Quantum Phase Estimation with Entangled Photons. *Nat. Phot.*, 15(2):137–142, 2021.

[90] S.-R. Zhao, Y.-Z. Zhang, W.-Z. Liu, J.-Y. Guan, W. Zhang, C.-L. Li, B. Bai, M.-H. Li, Y. Liu, L. You, et al. Field Demonstration of Distributed Quantum Sensing without Post-Selection. *Phys. Rev. X*, 11(3):031009, 2021.

[91] A. Hamann, P. Sekatski, and W. Dür. Approximate Decoherence Free Subspaces for Distributed Sensing. *Quantum Sci. Technol.*, 7(2):025003, 2022.

[92] J. Rubio and J. Dunningham. Bayesian Multiparameter Quantum Metrology with Limited Data. *Phys. Rev. A*, 101(3):032114, 2020.

[93] E. Polino, M. Valeri, N. Spagnolo, and F. Sciarrino. Photonic Quantum Metrology. *AVS Quantum Sci.*, 2(2):024703, 2020.

[94] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani. Experimental Realization of Any Discrete Unitary Operator. *Phys. Rev. Lett.*, 73(1):58–61, 1994.

[95] H. Wang, J. Qin, X. Ding, M.-C. Chen, S. Chen, X. You, Y.-M. He, X. Jiang, L. You, Z. Wang, et al. Boson Sampling with 20 Input Photons and a 60-Mode Interferometer in a 10^{14}-Dimensional Hilbert Space. *Phys. Rev. Lett.*, 123(25):250503, 2019.

[96] D. J. Brod, E. F. Galvão, A. Crespi, R. Osellame, N. Spagnolo, and F. Sciarrino. Photonic Implementation of Boson Sampling: A Review. *Adv. Photon.*, 1(3):034001, 2019.

[97] A. P. Lund, A. Laing, S. Rahimi-Keshari, T. Rudolph, J. L. O'Brien, and T. C. Ralph. Boson Sampling from a Gaussian State. *Phys. Rev. Lett.*, 113(10):100502, 2014.

[98] S. Rahimi-Keshari, A. P. Lund, and T. C. Ralph. What Can Quantum Optics Say about Computational Complexity Theory? *Phys. Rev. Lett.*, 114(6):060501, 2015.

[99] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex. Gaussian Boson Sampling. *Phys. Rev. Lett.*, 119(17):170501, 2017.

[100] R. Kruse, C. S. Hamilton, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex. Detailed Study of Gaussian Boson Sampling. *Phys. Rev. A*, 100(3):032326, 2019.

[101] A. Deshpande, A. Mehta, T. Vincent, N. Quesada, M. Hinsche, M. Ioannou, L. Madsen, J. Lavoie, H. Qi, J. Eisert, D. Hangleiter, B. Fefferman, and I. Dhand. Quantum Computational Advantage via High-Dimensional Gaussian Boson Sampling. *Sci. Adv.*, 8(1):eabi7894, 2022.

[102] D. Grier, D. J. Brod, J. M. Arrazola, M. B. d. A. Alonso, and N. Quesada. The Complexity of Bipartite Gaussian Boson Sampling. *Quantum*, 6:863, 2022.

[103] U. Chabaud and M. Walschaers. Resources for Bosonic Quantum Computational Advantage. *Phys. Rev. Lett.*, 130(9):090602, 2023.

[104] L. G. Valiant. The Complexity of Computing the Permanent. *Theor. Comput. Sci*, 8(2):189–201, 1979.

[105] A. Barvinok. *Combinatorics and Complexity of Partition Functions*, volume 30 of *Algorithms and Combinatorics*. Springer, Cham, 2016.

[106] A. Ehrenberg, J. T. Iosue, A. Deshpande, D. Hangleiter, and A. V. Gorshkov. The Second Moment of Hafnians in Gaussian Boson Sampling. *arXiv:2403.13878*, 2024.

[107] T. Jiang. The Entries of Circular Orthogonal Ensembles. *J. Math. Phys.*, 50(6):063302, 2009.

[108] T. Jiang. How Many Entries of a Typical Orthogonal Matrix can be Approximated by Independent Normals? *Ann. Probab.*, 34(4):1497–1529, 2006.

[109] A. Björklund, B. Gupt, and N. Quesada. A Faster Hafnian Formula for Complex Matrices and Its Benchmarking on a Supercomputer. *J. Exp. Algor.*, 24(1.11):1–17, 2019.

[110] A. Ehrenberg, J. T. Iosue, A. Deshpande, D. Hangleiter, and A. V. Gorshkov. Transition of Anticoncentration in Gaussian Boson Sampling. *arXiv:2312.08433*, 2023.

[111] A. Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, Boca Raton, 2017.

[112] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, A. Megrant, B. Chiaro, A. Dunsworth, K. Arya, et al. A Blueprint for Demonstrating Quantum Supremacy with Superconducting Qubits. *Science*, 360(6385):195–199, 2018.

[113] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven. Characterizing Quantum Supremacy in Near-Term Devices. *Nat. Phys.*, 14(6):595–600, 2018.

[114] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.*, 59(1):65–98, 2017.

[115] J. T. Iosue and A. Ehrenberg. jtiosue/LXEB GitHub repository, 2024. `https://github.com/jtiosue/LXEB`.

[116] B. Gupt, J. Izaac, and N. Quesada. The Walrus: a Library for the Calculation of Hafnians, Hermite Polynomials and Gaussian Boson Sampling. *J. Open Source Softw.*, 4(44):1705, 2019.

[117] J. Shi and T. Byrnes. Effect of Partial Distinguishability on Quantum Supremacy in Gaussian Boson Sampling. *npj Quantum Inf.*, 8(1):54, 2022.

[118] N. Quesada, J. M. Arrazola, and N. Killoran. Gaussian Boson Sampling Using Threshold Detectors. *Phys. Rev. A*, 98(6):062322, 2018.

[119] A. Deshpande, B. Fefferman, M. C. Tran, M. Foss-Feig, and A. V. Gorshkov. Dynamical Phase Transitions in Sampling Complexity. *Phys. Rev. Lett.*, 121(3):030501, 2018.

[120] G. Muraleedharan, A. Miyake, and I. H. Deutsch. Quantum Computational Supremacy in the Sampling of Bosonic Random Walkers on a One-Dimensional Lattice. *New J. Phys.*, 21:055003, 2018.

[121] N. Maskara, A. Deshpande, A. Ehrenberg, M. C. Tran, B. Fefferman, and A. V. Gorshkov. Complexity Phase Diagram for Interacting and Long-Range Bosonic Hamiltonians. *Phys. Rev. Lett.*, 129(15):150604, 2022.

[122] D. J. Thouless. Electrons in Disordered Systems and the Theory of Localization. *Phys. Rep.*, 13(3):93 – 142, 1974.

[123] B. Kramer and A. MacKinnon. Localization: Theory and Experiment. *Rep. Prog. Phys.*, 56(12):1469–1564, 1993.

[124] J. Billy, V. Josse, Z. Zuo, A. Bernard, B. Hambrecht, P. Lugan, D. Clément, L. Sanchez-Palencia, P. Bouyer, and A. Aspect. Direct Observation of Anderson Localization of Matter Waves in a Controlled Disorder. *Nature*, 453(7197):891–894, 2008.

[125] G. Roati, C. D'Errico, L. Fallani, M. Fattori, C. Fort, M. Zaccanti, G. Modugno, M. Modugno, and M. Inguscio. Anderson Localization of a Non-Interacting Bose–Einstein Condensate. *Nature*, 453(7197):895–898, 2008.

[126] R. Nandkishore and D. A. Huse. Many-Body Localization and Thermalization in Quantum Statistical Mechanics. *Annu. Rev. Condens. Matter Phys.*, 6:15–38, 2015.

[127] D. A. Abanin and Z. Papić. Recent Progress in Many-Body Localization. *Ann. Phys.*, 529(7):1700169, 2017.

[128] M. Serbyn, Z. Papić, and D. A. Abanin. Local Conservation Laws and the Structure of the Many-Body Localized States. *Phys. Rev. Lett.*, 111(12):127201, 2013.

[129] D. A. Huse, R. Nandkishore, and V. Oganesyan. Phenomenology of Fully Many-Body-Localized systems. *Phys. Rev. B*, 90(17):174202, 2014.

[130] A. Chandran, I. H. Kim, G. Vidal, and D. A. Abanin. Constructing Local Integrals of Motion in the Many-Body Localized Phase. *Phys. Rev. B*, 91(8):085425, 2015.

[131] V. Ros, M. Müller, and A. Scardicchio. Integrals of Motion in the Many-Body Localized Phase. *Nucl. Phys. B.*, 891:420 – 465, 2015.

[132] J. Z. Imbrie. On Many-Body Localization for Quantum Spin Chains. *J. Stat. Phys.*, 163(5):998–1048, 2016.

[133] B. M. Terhal and D. P. DiVincenzo. Adaptive Quantum Computation, Constant Depth Quantum Circuits and Arthur-Merlin Games. *Quantum Inf. Comput.*, 4(2):134–145, 2002.

[134] T. J. Osborne. Efficient Approximation of the Dynamics of One-Dimensional Quantum Spin Systems. *Phys. Rev. Lett.*, 97(15):157202, 2006.

[135] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao. Holographic Complexity Equals Bulk Action? *Phys. Rev. Lett.*, 116(19):191301, 2016.

[136] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao. Complexity, Action, and Black Holes. *Phys. Rev. D*, 93(8):086006, 2016.

[137] F. G. S. L. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng, and J. Preskill. Models of Quantum Complexity Growth. *PRX Quantum*, 2(3):030316, 2021.

[138] A. Bhattacharyya, P. Nandy, and A. Sinha. Renormalized Circuit Complexity. *Phys. Rev. Lett.*, 124(10):101602, 2020.

[139] V. Balasubramanian, M. DeCross, A. Kar, and O. Parrikar. Quantum Complexity of Time Evolution with Chaotic Hamiltonians. *J. High Energy Phys.*, 2020(134), 2020.

[140] V. Balasubramanian, M. DeCross, A. Kar, Y. Li, and O. Parrikar. Complexity Growth in Integrable and Chaotic Models. *J. High Energy Phys.*, 2021(11), 2021.

[141] A. Kar, L. Lamprou, M. Rozali, and J. Sully. Random Matrix Theory for Complexity Growth and Black Hole Interiors. *J. High Energy Phys.*, 2022(16), 2022.

[142] Y. Atia and D. Aharonov. Fast-Forwarding of Hamiltonians and Exponentially Precise Measurements. *Nat. Commun.*, 8(1):1572, 2017.

[143] S. A. Weidinger, S. Gopalakrishnan, and M. Knap. Self-Consistent Hartree-Fock Approach to Many-Body Localization. *Phys. Rev. B*, 98(22):224205, 2018.

[144] G. De Tomasi, F. Pollmann, and M. Heyl. Efficiently Solving the Dynamics of Many-Body Localized Systems at Strong Disorder. *Phys. Rev. B*, 99(24):241114(R), 2019.

[145] A. Chandran, J. Carrasquilla, I. H. Kim, D. A. Abanin, and G. Vidal. Spectral Tensor Networks for Many-Body Localization. *Phys. Rev. B*, 92(2):024201, 2015.

[146] F. Pollmann, V. Khemani, J. I. Cirac, and S. L. Sondhi. Efficient Variational Diagonalization of Fully Many-Body Localized Hamiltonians. *Phys. Rev. B*, 94(4):041116(R), 2016.

[147] T. B. Wahl, A. Pal, and S. H. Simon. Efficient Representation of Fully Many-Body Localized Systems Using Tensor Networks. *Phys. Rev. X*, 7(2):021018, 2017.

[148] A. K. Kulshreshtha, A. Pal, T. B. Wahl, and S. H. Simon. Approximating Observables on Eigenstates of Large Many-Body Localized Systems. *Phys. Rev. B*, 99(10):104201, 2019.

[149] E. Chertkov, B. Villalonga, and B. K. Clark. Numerical Evidence for Many-Body Localization in Two and Three Dimensions. *Phys. Rev. Lett.*, 126(18):180602, 2021.

[150] A. Arkhipov. Boson Sampling is Robust against Small Errors in the Network Matrix. *Phys. Rev. A*, 92(6):062326, 2015.

[151] M. J. Bremner, R. Jozsa, and D. J. Shepherd. Classical Simulation of Commuting Quantum Computations Implies Collapse of the Polynomial Hierarchy. *Proc. R. Soc. Math. Phys. Eng. Sci.*, 467(2126):459–472, 2011.

[152] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert. Anticoncentration Theorems for Schemes Showing a Quantum Speedup. *Quantum*, 2:65, 2018.

[153] M. J. Bremner, A. Montanaro, and D. J. Shepherd. Average-Case Complexity versus Approximate Simulation of Commuting Quantum Computations. *Phys. Rev. Lett.*, 117(8):080501, 2016.

[154] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani. On the Complexity and Verification of Quantum Random Circuit Sampling. *Nat. Phys.*, 15(2):159–163, 2019.

[155] I. H. Kim, A. Chandran, and D. A. Abanin. Local Integrals of Motion and the Logarithmic Lightcone in Many-Body Localized Systems. *arXiv:1412.3073*, 2014.

[156] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information 10th Anniversary Edition 2011*. Cambridge University Press, Cambridge, 2011.

[157] V. Kliuchnikov, A. Bocharov, M. Roetteler, and J. Yard. A Framework for Approximating Qubit Unitaries. *arXiv:1510.03888*, 2015.

[158] W. De Roeck and F. Huveneers. Stability and Instability Towards Delocalization in Many-Body Localization Systems. *Phys. Rev. B*, 95(15):155129, 2017.

[159] P. Niroula, J. Dolde, X. Zheng, J. Bringewatt, A. Ehrenberg, K. C. Cox, J. Thompson, M. J. Gullans, S. Kolkowitz, and A. V. Gorshkov. Quantum Sensing with Erasure Qubits. *arXiv:2310.01512*, 2023.

[160] W. Górecki, S. Zhou, L. Jiang, and R. Demkowicz-Dobrzański. Optimal Probes and Error-Correction Schemes in Multi-Parameter Quantum Metrology. *Quantum*, 4:288, 2020.

[161] D. Layden, S. Zhou, P. Cappellaro, and L. Jiang. Ancilla-Free Quantum Error Correction Codes for Quantum Metrology. *Phys. Rev. Lett.*, 122(4):040502, 2019.

[162] S. Zhou, M. Zhang, J. Preskill, and L. Jiang. Achieving the Heisenberg Limit in Quantum Metrology Using Quantum Error Correction. *Nat. Commun.*, 9(1):78, 2018.

[163] S. Zhou and L. Jiang. Optimal Approximate Quantum Error Correction for Quantum Metrology. *Phys. Rev. Res.*, 2(1):013235, 2020.

[164] S. Zhou, A. G. Manes, and L. Jiang. Achieving Metrological Limits Using Ancilla-Free Quantum Error-Correcting Codes. *Phys. Rev. A*, 109(4):042406, 2024.

[165] S. Zhou. Limits of Noisy Quantum Metrology with Restricted Quantum Controls. *arXiv:2402.18765*, 2024.

[166] M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, A. Chernoguzov, J. M. Dreiling, C. Foltz, F. Frachon, J. P. Gaebler, T. M. Gatterman, L. Grans-Samuelsson, et al. Demonstration of Logical Qubits and Repeated Error Correction with Better-than-Physical Error Rates. *arXiv:2404.02280*, 2024.

[167] J. Rubio and J. Dunningham. Quantum Metrology in the Presence of Limited Data. *New J. Phys.*, 21(4):043037, 2019.

[168] J. Bringewatt, A. Ehrenberg, J. Lautier-Gaud, and A. V. Gorshkov. In preparation.

[169] C. Oh, L. Jiang, and B. Fefferman. Spoofing Cross-Entropy Measure in Boson Sampling. *Phys. Rev. Lett.*, 131(1):010401, 2023.

[170] J. T. Iosue, A. Ehrenberg, D. Hangleiter, A. Deshpande, and A. V. Gorshkov. Page Curves and Typical Entanglement in Linear Optics. *Quantum*, 7:1017, 2023.

[171] J. Youm, J. T. Iosue, A. Ehrenberg, Y.-X. Wang, and A. V. Gorshkov. Average Rényi Entanglement Entropy in Gaussian Boson Sampling. *arXiv:2403.18890*, 2024.

[172] D. N. Page. Average Entropy of a Subsystem. *Phys. Rev. Lett.*, 71(9):1291–1294, 1993.

[173] S. K. Foong and S. Kanno. Proof of Page's Conjecture on the Average Entropy of a Subsystem. *Phys. Rev. Lett.*, 72(8):1148–1151, 1994.

[174] J. Sánchez-Ruiz. Simple Proof of Page's Conjecture on the Average Entropy of a Subsystem. *Phys. Rev. E*, 52(5):5653–5655, 1995.

[175] S. Sen. Average Entropy of a Quantum Subsystem. *Phys. Rev. Lett.*, 77(1):1–3, 1996.

[176] E. Bianchi, L. Hackl, M. Kieburg, M. Rigol, and L. Vidmar. Volume-Law Entanglement Entropy of Typical Pure Quantum States. *PRX Quantum*, 3(3):030201, 2022.

[177] A Serafini, O. C. O. Dahlsten, D Gross, and M. B. Plenio. Canonical and Micro-Canonical Typical Entanglement of Continuous Variable Systems. *J. Phys. A: Math. Theor.*, 40(31):9551, 2007.

[178] A. Serafini, O. C. O. Dahlsten, and M. B. Plenio. Teleportation Fidelities of Squeezed States from Thermodynamical State Space Measures. *Phys. Rev. Lett.*, 98(17):170501, 2007.

[179] M. Fukuda and R. Koenig. Typical Entanglement for Gaussian States. *J. Math. Phys.*, 60(11):112203, 2019.

[180] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, 2004.

[181] A. Chiruvelli and H. Lee. Parity Measurements in Quantum Optical Metrology. *J. Mod. Opt.*, 58(11):945–953, 2011.

[182] J. M. Arrazola, T. R. Bromley, and P. Rebentrost. Quantum Approximate Optimization with Gaussian Boson Sampling. *Phys. Rev. A*, 98(1):012322, 2018.

[183] P. Feijão, F. V. Martinez, and A. Thévenin. On the Distribution of Cycles and Paths in Multichromosomal Breakpoint Graphs and the Expected Value of Rearrangement Distance. *BMC Bioinform.*, 16(19):S1, 2015.

[184] N. N. Li and W. Chu. Multifold Convolutions of Binomial Coefficients. *Math. Commun.*, 24(2):279–287, 2019.

[185] G. Chang and C. Xu. Generalization and Probabilistic Proof of a Combinatorial Identity. *Am. Math. Mon.*, 118(2):175–177, 2011.

[186] A. Bouland, B. Fefferman, Z. Landau, and Y. Liu. Noise and the Frontier of Quantum Supremacy. In *Proc. – 62nd Annu. IEEE Symp. Found. Comput. Sci. FOCS*, pages 1308–1317, 2022.

[187] A. Deshpande, P. Niroula, O. Shtanko, A. V. Gorshkov, B. Fefferman, and M. J. Gullans. Tight Bounds on the Convergence of Noisy Random Circuits to the Uniform Distribution. *PRX Quantum*, 3(4):040329, 2022.

[188] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley, New York, 2007.

[189] G. H. Hardy. *Some Famous Problems of the Theory of Numbers and in Particular Waring's Problem; An Inaugural Lecture Delivered before the University of Oxford*. Clarendon Press, London, 2011.

[190] Wolfram Research, Inc. Mathematica, Version 14.0, 2024. Champaign, IL, 2024.

[191] robjohn (https://math.stackexchange.com/users/13854/robjohn). Binomial Sum Gives 4^n. Mathematics Stack Exchange, 2016. URL:https://math.stackexchange.com/q/1595627 (version: 2016-01-01).

[192] OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences, 2024. Published electronically at `http://oeis.org`.

[193] T. van Aardenne-Ehrenfest and N. G. de Bruijn. Circuits and Trees in Oriented Linear Graphs. *Simon Stevin.*, 28:203–217, 1951.

[194] J. Borwein and O.-Y. Chan. Uniform Bounds for the Complementary Incomplete Gamma Function. *Math. Inequalities Appl.*, 12(1):115–121, 2009.