ABSTRACT

Title of Dissertation: HARNESSING QUANTUM SYSTEMS FOR SENSING, SIMULATION, AND OPTIMIZATION

Jacob Allen Bringewatt
Doctor of Philosophy, 2024

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

Quantum information science offers a remarkable promise: by thinking practically about how quantum systems can be put to work to solve computational and information processing tasks, we gain novel insights into the foundations of quantum theory and computer science. Or, conversely, by (re)considering the fundamental physical building blocks of computers and sensors, we enable new technologies, with major impacts for computational and experimental physics.

In this dissertation, we explore these ideas through the lens of three different types of quantum hardware, each with a particular application primarily in mind: (1) networks of quantum sensors for measuring global properties of local field(s); (2) analog quantum computers for solving combinatorial optimization problems; and (3) digital quantum computers for simulating lattice (gauge) theories.

For the setting of quantum sensor networks, we derive the fundamental performance limits for the sensing task of measuring global properties of local field(s) in a variety of physical

settings (qubit sensors, Mach-Zehnder interferometers, quadrature displacements) and present explicit protocols that achieve these limits. In the process, we reveal the geometric structure of the fundamental bounds and the associated algebraic structure of the corresponding protocols. We also find limits on the resources (e.g. entanglement or number of control operations) required by such protocols.

For analog quantum computers, we focus on the possible origins of quantum advantage for solving combinatorial optimization problems with an emphasis on investigating the power of adiabatic quantum computation with so-called stoquastic Hamiltonians. Such Hamiltonians do not exhibit a sign problem when classically simulated via quantum Monte Carlo algorithms, suggesting deep connections between the sign problem, the locality of interactions, and the origins of quantum advantage. We explore these connections in detail.

Finally, for digital quantum computers, we consider the optimization of two tasks relevant for simulating lattice (gauge) theories. First, we investigate how to map fermionic systems to qubit systems in a hardware-aware manner that consequently enables an improved parallelization of Trotter-based time evolution algorithms on the qubitized Hamiltonian. Second, we investigate how to take advantage of known symmetries in lattice gauge theories to construct more efficient randomized measurement protocols for extracting purities and entanglement entropies from simulated states. We demonstrate how these protocols can be used to detect a phase transition between a trivial and a topologically ordered phase in Z_2 lattice gauge theory. Detecting this transition via these randomized methods would not otherwise be possible without relearning all symmetries.

HARNESSING QUANTUM SYSTEMS FOR SENSING, SIMULATION, AND OPTIMIZATION

by

Jacob Allen Bringewatt

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 Zohreh Davoudi, Chair
Professor Alexey V. Gorshkov, Advisor and Co-chair
Professor Andrew M. Childs, Dean's Representative
Professor Yi-Kai Liu
Professor Ronald Walsworth

© Copyright by
Jacob Allen Bringewatt
2024

Dedication

To Kara, who we lost too soon

Gonna lay down my sword and shield
Down by the riverside
Down by the riverside
Down by the riverside
Gonna lay down my sword and shield
Down by the riverside
Ain't gonna study war no more

—an old spiritual loved by an old spirit

Acknowledgments

When, as a 17-year-old kid, my mom overruled my lack of interest in filling out yet another college application and convinced me to apply to the University of Maryland for undergrad, I had no way of knowing it would become my home for the next decade. I am so grateful that it has been (mothers are always right). While it is impossible to thank every person who has helped make this dissertation possible it is an honor to take a couple of pages to highlight a few.

First and foremost, thank you to my advisor Alexey Gorshkov. Alexey is inspirational in his unflagging curiosity and his impossibly broad range of expertise. I am exceedingly grateful for how he has given me the independence to explore my interests and develop into my own researcher, while providing supportive and honest guidance, both personal and technical, whenever it is needed. I genuinely could not imagine having a better advisor.

I have also been fortunate to have had Zohreh Davoudi and Michael Jarret as mentors and collaborators. They have made both the content and general approach of my research much richer due to the unique talents and interests they have generously shared with me. Also, thank you to my friend and mentor William (Bill) Dorland, who set me off down the quantum computing road as an undergraduate, and provided the best early education in how to be a physicist I could have hoped for; and to Stephen Jordan, who sparked my interest in analog quantum computation, a topic that fills a good chunk of this thesis. I have been fortunate to get to work with a number of other excellent and inspirational faculty and senior researchers over the past six years. While not

exhaustive, a few deserve a special mention: Andrew Childs, Yi-Kai Liu, Wally Melnitchouk, and Nobuo Sato. All are admirable scientists and a pleasure to work with, and have made important contributions to how I approach my own research.

Thank you, also, to the many postdocs who have generously given their time and expertise over the years. Of particular note: Lucas Brady and Luis Pedro García-Pintos for being some of the most fun people to do science with; Igor Boettcher for teaching me how to make a good presentation; Niklas Mueller for helping me to focus on the physics when I get a little too carried away being a mathematician or computer scientist; also, sincere thanks to Kishor Bharti, James Watson, Dominik Hangleiter, and Przemek Bienias.

Of course, graduate school is impossible without the support and camaraderie of other graduate students. To begin, the deepest of thanks to Adam Ehrenberg, best friend,[1] office-mate, frequent collaborator and cursed travel companion. Along with being a co-first author on two of the papers included in this dissertation, he also gave a number of helpful comments on the introductory chapter. This Ph.D. would not have been nearly as fruitful or as fun without you. To the rest of "the boys": your friendship has truly made these past six years the best of my life. Nick Mennona, thank you for being the type of guy who will always drop everything to help a friend. Stefano Antonini for beers at Town Hall with the perfect words at the (im)perfect times. Martin Ritter for his wise insights on life and human nature, hikes in dress shoes, and for showing me my first qubit. Dhruvit Patel for slow Saturday mornings chatting and messing around (in my case) or playing beautifully (in yours) with guitars at the Pony Trail house. Zach Steffen for being the best Covid lockdown buddy, your understated humor, and beating all of Borderlands while I browsed

[1] While friends, at most, permit only a partial ordering, Adam gets this title for both sincere and meme-related reasons.

my inventory. Alex Wikner for being the most dedicated brew master and Dungeon Master. Evan Dowling, for living up to Midwestern stereotypes in the best possible way by leading us to IM victories and being one of the genuinely kindest people I know. John Armstrong for baseball, both playing and watching. And, of course, everyone in the Gorshkov and Davoudi groups (and QuICS and JQI more broadly) deserve a big thank you for always making work feel like "work." Of special note: thank you to Yusuf Alnawakhtha, Jon Kunjummen and Connor Mooney.

I am also grateful for the good friends I have been fortunate enough to have beyond the UMD community—most especially Piragash Swargaloganathan and Alekhya Chaparala. And for those from back home in North Carolina and from the undergrad years with whom I have remained close—Shreya Anand, Ashleigh Cleveland, James Rees, Jacob Renn, and Patrick Zhang—as well as those with whom I have lost touch—you all played a deeply appreciated role in putting me in a position intellectually and personally to succeed at this stage of life.

Even more foundationally, I am indescribably grateful to my family. All physicists know the importance of initial conditions, and I had an especially excellent set. Particular thanks to my parents, Kevin and Tamara; my siblings, Rachel, Kara, Marisa and Caden; my grandparents Mark and Sylvia, Ron and Jean. You have all instilled in me a curiosity for the world, a desire to do good, and an appreciation for the little moments. I will never be able to thank you all enough for your role in making me the person I am today.

Finally, and especially, thank you to my wonderful wife, Shaalini. While one can easily get lost in the daily grind, the romantic in me still views the pursuit of science as a noble search for the true, the beautiful, and the good. The one place I am completely sure I have found all three is in you. Together, I am confident that whatever good the world has in store for us will be elevated and whatever hardships we might face will be overcome.

Table of Contents

Dedication ii

Acknowledgements iii

Table of Contents vi

List of Tables xi

List of Figures xii

Citations to Previously Published Work xxi

Chapter 1: Introduction 1
1.1 Overview and Motivation 1
1.2 Outline of Dissertation 5
1.2.1 Part I: Quantum Sensing 5
1.2.2 Part II: Quantum Optimization 7
1.2.3 Part III: Quantum Simulation 11

I Quantum Sensing 14

Chapter 2: Minimum Entanglement Protocols for Function Estimation 15
2.1 Introduction 15
2.2 Problem Setup 17
2.3 Conditions for Saturable Bounds 19
2.4 A Family of Optimal Protocols 21
2.5 Minimum Entanglement Solutions 26
2.6 Average Entanglement 29
2.7 CNOT Costs of Minimum Entanglement Protocols 31
2.8 Time-Independent Protocols 36
2.9 Conclusion and Outlook 39

Chapter 3: Optimal Function Estimation with Photonic Quantum Sensor Networks 42
3.1 Introduction 42
3.2 Problem Setup 44
3.3 Lower Bounds 46

3.4 Protocols . . . 52
3.4.1 Existing Protocols . . . 52
3.4.2 Algebraic Conditions for New Protocols . . . 53
3.5 Entanglement Requirements . . . 59
3.6 Conclusion and Outlook . . . 61

Chapter 4: Optimal Measurement of Field Properties with Quantum Sensor Networks 63
4.1 Introduction . . . 63
4.2 Problem Setup . . . 66
4.3 MSE Bound . . . 67
4.4 Optimal Protocol . . . 70
4.5 Applications . . . 74
4.6 Outlook . . . 76

Chapter 5: Protocols for Estimating Multiple Functions with Quantum Sensor Networks: Geometry and Performance 79
5.1 Introduction . . . 79
5.2 Problem Setup . . . 83
5.3 The Strategies . . . 90
5.3.1 Local Strategy . . . 90
5.3.2 Signed Sensor Symmetric Strategy . . . 90
5.3.3 Naive Sequential Strategy . . . 95
5.3.4 Optimal Sequential Strategy . . . 96
5.4 Performance and Geometry . . . 98
5.4.1 Geometrically Symmetric Limit . . . 98
5.4.2 Nearly Overlapping Functions . . . 100
5.4.3 Numerical Results . . . 107
5.5 Conclusion and Outlook . . . 109

Chapter 6: Discussion and Further Directions for Part I 113

II Quantum Optimization 116

Chapter 7: Polynomial Time Algorithms for Estimating Spectra of Adiabatic Hamiltonians 117
7.1 Introduction . . . 117
7.2 Tight-binding approximation . . . 119
7.3 Numerical Considerations and Error Estimates . . . 126
7.4 Examples and Applications . . . 129
7.5 Conclusion . . . 138

Chapter 8: Effective Gaps Are Not Effective: Quasipolynomial Classical Simulation of Obstructed Stoquastic Hamiltonians 140
8.1 Introduction . . . 140
8.2 Background . . . 140
8.3 Approach . . . 142

8.4 Algebraic Graph Theory . . . 143
8.5 Mapping I: H to Γ . . . 144
8.6 The Algorithm . . . 146
8.7 Mapping II: H to G . . . 149
8.8 Discussion . . . 153
8.9 Conclusion . . . 154

Chapter 9: Simultaneous Stoquasticity 155
9.1 Introduction . . . 155
9.2 Lie Algebras and Lie Groups . . . 157
9.3 Simultaneous stoquasticity . . . 159
9.4 A No-go Result . . . 162
9.5 Bloch Vector Approach . . . 163
9.6 Conclusion and Outlook . . . 166

Chapter 10: Discussion and Further Directions for Part II 168

III Quantum Simulation 171

Chapter 11: Parallelization Techniques for Quantum Simulation of Fermionic Systems 172
11.1 Introduction . . . 172
11.2 Custom Fermion-to-Qubit Mappings . . . 176
11.2.1 Setup . . . 176
11.2.2 Local Fermion-to-Qubit Mappings . . . 177
11.2.3 Custom Fermionic Codes . . . 179
11.2.4 Prior Work on Optimizing System Graphs . . . 186
11.3 Parallelization and Path Coloring . . . 187
11.3.1 Notions of Parallelization . . . 187
11.3.2 Graph Coloring . . . 188
11.3.3 Conflict Graphs for Parallelizability . . . 190
11.4 Analytic Results . . . 192
11.4.1 The Hamiltonian . . . 192
11.4.2 Rules for Strong Coloring . . . 194
11.4.3 Limits of Weak and Strong Coloring . . . 197
11.5 Numerical Results . . . 210
11.5.1 Description of the Algorithm . . . 210
11.5.2 Analytically Solved Examples Revisited . . . 212
11.5.3 Current Architectures . . . 214
11.5.4 Local Interactions . . . 219
11.6 Conclusion and Outlook . . . 221

Chapter 12: Randomized Measurement Protocols for Lattice Gauge Theories 224
12.1 Introduction . . . 224
12.2 Symmetry-Conscious Randomization . . . 227

12.2.1 A Symmetry-Conscious Unitary k-design Example: Particle Number Symmetry . . . 229
12.2.2 Estimating k-purities from k-designs . . . 233
12.3 Lattice Gauge Theory Entanglement . . . 235
12.3.1 $\mathbb{Z}_2$ LGT in (1+1)d . . . 235
12.3.2 Measuring Distillable and Symmetry Components of Entanglement . . . 238
12.3.3 Classical Shadows . . . 240
12.4 Experimental Verification of Topological Phases . . . 243
12.4.1 Symmetry-Respecting Randomized Circuits . . . 244
12.4.2 Classical Shadows . . . 247
12.4.3 Detecting Topological Order Through Entanglement Hamiltonian Tomography . . . 247
12.5 Conclusion and Outlook . . . 250

Chapter 13: Discussion and Further Directions for Part III 254

IV Appendices 256

Appendix A: Technical details of the results reported in Chapter 2 257
A.1 A Useful Lemma Regarding Optimal Probe States . . . 257
A.2 Proof of the Optimality of Cat-State Protocols . . . 260
A.3 Review of Robust Phase Estimation . . . 262
A.4 Full Proof of the Main Theorem . . . 266
A.5 Minimum Entanglement Non-Echoed Protocols . . . 271
A.6 Relaxing the Assumption on a Single Maximum Element . . . 273
A.6.1 Generalizing Eq. (2.8) of the Main Text . . . 274
A.6.2 Generalizing the Derivation of Eq. (2.13) of the Main Text . . . 277
A.6.3 Generalizing the Proof of Theorem 2.5.1 of the Main Text . . . 279

Appendix B: Technical details of the results reported in Chapter 3 281
B.1 Bound for Local Phase Shifts . . . 281
B.2 Bound for Local Displacements . . . 286
B.2.1 Separable Bound . . . 286
B.2.2 General Function Estimation Bound . . . 289
B.3 Quantum Fisher Information Matrix Elements . . . 291
B.4 Protocols for Local Phase Shifts . . . 292
B.4.1 An Optimal Protocol for Functions with Positive Coefficients . . . 292
B.4.2 Extending the Optimal Protocol to Negative Coefficients . . . 294
B.4.3 A Family of Optimal Protocols . . . 295
B.5 Proof of Lemma 3.4.1 . . . 298
B.6 Fisher Information Matrix Conditions for Quadrature Displacements . . . 303
B.7 Approaching the Single-Shot Limit and Robust Phase Estimation . . . 305

Appendix C: Technical Details of the Results Reported in Chapter 4 306

C.1 Justification of Using Single-Parameter Bound . . . 306
C.2 Proof of Validity of the Consistency Condition . . . 309
C.3 Optimal Protocol: Case of Analytic Functions . . . 311
C.4 Proofs on Estimate Asymptotics . . . 313
C.5 Proof of Protocol Optimality . . . 323
C.6 Review of the Protocol by Eldredge *et. al.* . . . 327

Appendix D: Technical Details of the Results Reported in Chapter 5 328
D.1 Derivation of Signed Sensor Symmetric Bound . . . 328
D.2 Optimal Time Allocation . . . 332
D.3 Nearly Overlapping Functions . . . 334

Appendix E: Technical Details of the Results Reported in Chapter 7 337
E.1 Fully-Symmetric, Stoquastic Hamiltonians . . . 337
E.2 Relabeling Bases for Three or Fewer Hamming-Symmetric Wells . . . 339
E.3 Tight-Binding Matrix Elements . . . 341
E.4 Proof of Possible Subspaces for First Excited State . . . 343

Appendix F: Technical Details of the Results Reported in Chapter 8 345
F.1 Minimal Example . . . 345
F.2 Additional Gadgets . . . 353
F.3 Proofs . . . 354
F.4 Smooth Transitions . . . 356

Appendix G: Technical Details of the Results Reported in Chapter 9 358
G.1 Some Properties of the Star Product . . . 358
G.2 Proof of Sufficient Word Length for Unitary Similarity of a Pair of Hermitian Matrices . . . 360
G.3 Bound on the Length of Words . . . 362
G.4 Trace Invariants in Terms of Bloch Vectors . . . 364
G.5 Simultaneous Stoquasticizability is Rare . . . 366
G.6 Results on Simultaneous Diagonalizability . . . 372

Appendix H: Technical Details of the Results Reported in Chapter 12 374
H.1 Equivalence of Error Metrics . . . 374
H.2 Particle Number Analysis . . . 379
H.2.1 Details for Figure 3(a) . . . 381
H.2.2 Details for Figure 3(b) . . . 381
H.3 Details for the (1+1)d $\mathbb{Z}_2$ LGT Example . . . 385
H.4 Details of the (2+1)d $\mathbb{Z}_2$ LGT Example . . . 386
H.4.1 Approximate Unitary k-designs and k-purities . . . 386
H.4.2 Classical Shadow Analysis . . . 387
H.4.3 Entanglement Hamiltonian Tomography Analysis . . . 388

List of Tables

3.1 A 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. 61

5.1 Summary of figures of merit. Recall, that for all strategies other than signed sensor symmetric strategy, we have an explicit physical protocol to achieve the given figure of merit. For the signed sensor symmetric strategy, beyond $d = 2$, we are not necessarily guaranteed that a state exists that achieves the figure of merit, and therefore it is a lower bound, given the signed sensor symmetric state restriction. 98

5.2 Summary of analytic results comparing the signed sensor symmetric strategy and optimized sequential strategy. Recall that the figure of merit for the local strategy is $\mathcal{N}/t^2$. 106

7.1 First excited state energies for Example 4. First order tight-binding is needed to get the correct solution in this problem. 138

8.1 Gadgets summary. Vertex labels are as defined in the text (See Eqs. (8.9) and (8.10)). Superscripts are dropped and represented with a unique color/shape. 151

11.1 Rules for determining internal qubits used by the various terms that arise in simulating the Hamiltonian in Eq. (11.9) using a Jordan-Wigner encoding of the local Majoranas. Recall $a_u(v) := \lceil \frac{\xi_u(v)}{2} \rceil$ is the number of "active" qubits when implementing $\tilde{\gamma}_u^{\xi_u(v)}$, and v is the $\xi_u(v)$-th neighbor of u. 197

List of Figures

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. 36

4.1 At each position $\boldsymbol{x}_i$, a quantum sensor (black dots) is coupled to a field $f(\boldsymbol{x};\boldsymbol{\theta})$, whose functional form is known, but the parameters $\boldsymbol{\theta}$ are not. The protocols presented here utilize entanglement to obtain the highest accuracy allowed by quantum mechanics in estimating the quantity $q(\boldsymbol{\theta})$. One example problem is to estimate the field value $q = f(\boldsymbol{x}_0;\boldsymbol{\theta})$ at a location $\boldsymbol{x}_0$ (red cross) without a sensor. 64

5.1 The protocols for measuring $n \leq d$ linear functions $\{f_1(\boldsymbol{\theta}), \ldots, f_n(\boldsymbol{\theta})\}$ of d parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ considered in this work can be classified into three groups: (a) Local protocols do not utilize entanglement and measure the parameters locally, allowing for large parallelization. (b) Global protocols simultaneously estimate all functions. (c) Sequential protocols divide the problem into n parts, where each part is optimized to estimate a single function from the set $\{f'_1, \ldots, f'_n\}$, which may consist of linear combinations of the original set $\{f_1, \ldots, f_n\}$. 81

5.2 A visualization for $n = 2$ functions and $d = 3$ sensors of how we can optimally select a set of functions to measure whose coefficient vectors $\{\boldsymbol{\alpha}'_\ell\}$ span the same subspace as the coefficient vectors $\{\boldsymbol{\alpha}_\ell\}$ of the functions we care about. The vectors are the coefficient vectors and the planes indicate the subspace they span. The axes are labeled by standard basis unit vectors $\{e_1, e_2, e_3\}$. 87

5.3 (a) A visualization for $n = 2$ functions and $d = 3$ sensors of geometrically symmetric functions. In particular, the coefficent vectors lie near the surface of a cone centered on some $\boldsymbol{\omega}$. (b) The opening angle of the cone is given by ϕ' and the angular displacement from ϕ' for a particular $\boldsymbol{\alpha}_\ell$ is specified by $\epsilon_{\boldsymbol{\omega},\ell}$, as defined in Eq. (5.32). 99

5.4 $\mathcal{M}_{\rm ss}$ versus $\mathcal{M}_{\rm opt}$ for 1000 random samples from the positive orthant of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ with $n = 2$, $w_1 = w_2 = 1$ for different numbers of sensors d. Dashed lines correspond to $\mathcal{M}_{\rm local}$. Colors correspond to the geometry parameter for the problem instance. Observe that the signed sensor symmetric approach is never worse than the local protocol, whereas the optimized sequential protocol can be. However, as d increases the optimized sequential protocol is almost always superior. Also recall, that for $d > 2$, $\mathcal{M}_{\rm ss}$ is generically just a lower bound, and it is not guaranteed one can achieve this figure of merit with physical states. Therefore, one can think of $\mathcal{M}_{\rm ss}$ as a best case scenario for a physically realized signed sensor symmetric protocol. 107

5.5 $\mathcal{M}_{\rm opt}$ versus $\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2$ for $n = 2$ functions and $d = 2$ sensors. Note that the nearly orthogonal case ($\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 \approx 0$) implies $\mathcal{M}_{\rm opt} \approx 4$ (i.e., the worst case for the optimized sequential strategy). Comparing to the first panel of Fig. 5.4 we see that in this case $\mathcal{M}_{\rm ss} \approx \mathcal{M}_{\rm local} = 2$. 110

7.1 The key feature of our model and the tight-binding approach to analyzing it is the ability to exponentially reduce the size of the problem to an effective Hamiltonian that explicitly considers tunneling between a collection of Hamming symmetric wells. Here we diagramatically demonstrate this reduction for a collection of 3 Hamming symmetric wells of width 1 on 10 qubits. The original graph representing our Hamiltonian has a hypercube geometry, but the edges not within wells are eliminated for visual clarity. 119

7.2 (a) Exact ground state of a two well system on 10 qubits ($n_1 = 4$, $n_2 = 6$) as a function of h_1 and h_2 for s near the minimum gap. The well on the left (right) has width 0 (1) and we see that at the point of minimum gap the wave function can rapidly switch between wells. (b) Energy landscape for fixed $h_2 = 0$ (cut with minimum distance between wells). The minimum eigenvalue gap occurs at an energy such that the parts of the wave functions external to the wells have amplitudes on bit strings for which the potential energy is greater than the eigenenergy of the state. Here potential energy is shifted by $-(1-s)I$ so that the relation to tunneling in the sense that the there is wave function support on bitstrings such that the energy is less than the potential energy is readily apparent. 127

7.3 True error (tight-binding compared to direct diagonalization) versus error estimate for eigenvalue gap in zeroth order TB for a set of 2450 random runs on between 4 and 10 qubits, between 2 and 10 wells, with depths between -1.0 and -5.99 and width 0 at 17 evenly spaced s values. If $\tilde{\gamma}^{(TB)} < 1$ then tight-binding cannot distinguish between the ground state and first excited state and the corresponding point is not plotted. Circles indicate $s \in [0.15, 0.30]$, squares $s \in [0.15, 0.30]$, diamonds $s \in [0.35, 0.70]$, and stars $s \in [0.70, 0.95]$. Note for the low s points violating the upper bound estimate in the top right of the figure, with a different choice of ϵ the Fix-Heiberger algorithm could eliminate these points as unstable eigenvalues. 128

7.4 Error estimate minus true error estimated CDF for zeroth order TB for the 2450 random runs from Fig 7.3. 129

7.5 Plot of eigenvalue gap versus distance from marked item for a single width 0, depth -1 marked item for n=20 (circles), 30 (squares), 40 (diamonds), 50 (cross), 60 (five point star), 70 (triangle). Error bars overlaid on these points indicate the tight-binding results. Horizontal dotted lines show the gap for no prior and the shaded region indicates the distances where the prior offers an improvement in the gap over standard Grover. 132
7.6 Plot of probability scaled minimum eigenvalue gap vs. number of qubits. The dashed line demonstrates that we recover $O(2^{-n/2})$ scaling if the priors are randomly guessed. 132
7.7 Plot of a particular eigenvalue spectrum vs. s for the Grover with priors example. Lines indicate exact eigenvalues whereas error bars give the lowest two energy levels as determined by tight-binding. We see that tight-binding does an excellent job at finding the minimum gap in this problem with error bars on the order of the machine precision. 133
7.8 Comparison of exact adiabatically simulated ground state probability distribution to exact true ground state for the 3 qubit Ising model with $J_{ij} = 1 \forall i, j$ and $B_i = 0.015 \forall i$. The adiabatic Hamiltonian is constructed via a mapping between tight-binding with 8 wells and the 3 qubit Ising model Hamiltonian. 136
7.9 Plot of lowest two eigenvalues versus s for Example 3. We see at large s tight-binding is effective, however, a large number of wells does require that the minimum gap occur at larger s to capture it effectively with tight-binding. Lines indicate tight-binding without Fix-Heiberger indicating its usefulness to automatically remove unstable eigenvalues. 137

9.1 For a qubit, the geometric representation of `Stoq` in Bloch vector space is the $\pm\hat{\sigma}^{(z)}, -\hat{\sigma}^{(x)}$ half-plane. Observe that for two Hamiltonians H_1, H_2, represented by their Bloch vectors, there always exists a unitary which can simultaneously take both H_1', H_2' to the `Stoq` subspace. 160

11.1 (a) An overview of the full procedure of defining and solving the weak and strong coloring problems for parallelizing a Hamiltonian simulation of fermions. There are many stages for optimization: the choice of system graph, the choice of physical fermionic modes, the choice of paths linking those modes, and the coloring algorithm. While these choices are straightforward in this small example, for general problems, the design space is extremely large. This work focuses on the last two steps, which is an NP-hard optimization problem. Here, the conflict-graph vertices are labeled by the physical vertices of the system graph involved in the interaction. That is, $\tilde{A}_{uv}\tilde{B}_v$ is labeled by uv and $\tilde{B}_v$ is labeled by v for all u, v. Note that the difference between the weak and strong coloring problems in this example is in the ability of the strong coloring scheme to route through the virtual vertex e to implement the ad (and da) path simultaneously with bc (and cb) without any conflict, hence a lower chromatic number compared with the weak coloring scheme. This corresponds to enumerating the edges of vertex e as $\{ea, ed, eb, ec\} \mapsto \{1, 2, 3, 4\}$. (b) Corresponding circuit diagrams for ordering the Pauli strings according to the sequential strategy and via the weak and strong coloring problems. Here e_L and e_R label the left and right internal qubits of vertex e of the system graph, respectively. Colors match those in the corresponding conflict graphs and gates of the same color are implemented simultaneously. . . . 193

11.2 The central gray vertex is u and black vertices are its neighbors. This is just notation and no formal coloring has been performed yet. When constructing the conflict graph for strong coloring under a Jordan-Wigner encoding of the local Majoranas, it is useful to think of each vertex $u \in V_\Sigma$ as being expanded to a line graph of $n_u = \lceil d(u)/2 \rceil$ internal vertices, each connected to two of the original edges of u, where $d(u)$ is the degree of u. Different interaction types on vertex u induce different Jordan-Wigner strings on these internal vertices as summarized in Tab. 11.1 and depicted in Fig. 11.3. 195

11.3 Examples of the internal qubits of vertex u (enumerated top to bottom as u_1, u_2, u_3) that are required to implement the operators (a) $\tilde{B}_u$, (b) $\tilde{A}_{ux}$, (c) $\tilde{A}_{xu}\tilde{A}_{uy}$, and (d) $\tilde{A}_{xu}\tilde{B}_u$. Active qubits and input/output system-graph vertices are marked with red. The dashed lines denote internal edges. Again, the coloring is the notation and no formal graph coloring is assumed here. The choice of edge enumeration is marked. 197

11.4 Examples of the optimal procedure to parallelize two-mode interactions via strong coloring for the star graph for (a) $N = 8$ and (b) $N = 7$. In the top of the figure, the internal vertices of the central virtual vertex u are shown. Edges between internal vertices are represented by dashed lines. Red (gray) vertices are active (inactive) with red (gray) lines indicating induced (no) Jordan-Wigner strings. Terms are grouped as they appear in the respective sums over μ in Eqs. (11.28)-(11.29). Observe in the case of N odd, this means that the Jordan-Wigner strings of a given type are split in the different groupings (the locations of such splits are marked by one, two, or three stars). The step counts underneath each group give the number of steps to implement all interactions that induce that set of Jordan-Wigner strings. This gives a total of 42 and 34 steps for $N = 8$ and $N = 7$, respectively. The counts associated with each individual Jordan-Wigner string give the number of interactions corresponding to that Jordan-Wigner string. This gives a total of $8 \times (8 - 1) = 56$ and $7 \times (7 - 1) = 42$ interactions for $N = 8$ and $N = 7$, respectively. Representative examples of the types of interactions between physical vertices that induce the different Jordan-Wigner strings are shown in the bottom of the figure. Observe in (b) that for N odd, the last internal vertex has only one physical vertex as a neighbor. This is responsible for the different procedure for optimal parallelization. 202

11.5 An illustration of the conflict graph for the weak-coloring problem with a system graph K_N with $N = 4$. The vertices of the system graph are labeled as $\{a, b, c, d\}$ (e.g., as in Fig. 11.6), and the vertices of the resulting conflict graph are labeled by the corresponding interaction. The conflict graph consists of $N = 4$ interlocking complete subgraphs each associated with one of the system-graph vertices, as described in the main text. One such complete subgraph is shown. 204

11.6 A minimal example of K_4 that shows a complete system graph can have its edges enumerated such that a Hamiltonian cycle of two-mode interactions can be implemented simultaneously via strong coloring. The dashed lines denote internal edges, red denotes active qubits, and red arrows from u to v for $u, v \in V_\Sigma$ denote the implementation of an interaction of the type $\tilde{A}_{uv}\tilde{B}_v$. 205

11.7 An example of the system graph $\Sigma_{\text{bottleneck}}$ that allows for a linear-in-N scaling of the ratio of the chromatic numbers of the conflict graphs with weak and strong coloring. The physical vertices are divided into two complete subgraphs $K_{N/2}$ separated by a single vertex bottleneck, which allows only one interaction at a time between the two subgraphs in the weak coloring problem. For visual clarity, the edges between vertices of the complete subgraphs are not shown. For strong coloring, the extra layer of virtual vertices between each subgraph and the central bottleneck vertex allows any disjoint set of $\frac{N}{2}$ interactions between the two subgraphs to be implemented in a single step. 208

11.8 Numerical results for the chromatic number from the weak (blue circles) and strong coloring (red diamonds) problems for (a) star and (b) complete system graphs with N physical fermionic modes. For the non-asymptotic analytic results, see Eqs. (11.27)-(11.29) and Eq. (11.30), respectively. For the complete graph, the numerical algorithm fails to achieve the analytically determined bounds as obtaining these results requires a highly fine-tuned vertex ordering for the greedy coloring algorithm on the corresponding conflict graph. 214
11.9 Mappings from architecture graphs to system graphs. On the architecture graph, qubits are represented by black dots and are grouped into system-graph vertices as denoted by the gray shading. On the system graphs, square vertices denote virtual vertices and red circular vertices denote physical vertices. For the numerics, N physical vertices are chosen randomly from the 49 total fermionic modes. 217
11.10Numerical results for the chromatic number of the conflict graph as a function of the number of qubits N for (a) the heavy-hexagon system graph and (b) the triangular-lattice system graph compared with the scaling of the sequential implementation of the Hamiltonian terms. 218
11.11Numerical results on the amount of parallelization for various system graphs. (a) shows the improvement in the number of steps for strong coloring versus weak coloring. (b) shows the improvement for strong coloring over a sequential implementation of the interactions in the interaction list. (c) and (d) show the same as (a) and (b), respectively, but weighted by the number of qubits in the system graph. When comparing the different system graphs, recall the size of the complete graph and star graph scale with the number of physical vertices N whereas the heavy hexagon and triangular lattice do not. 218
11.12Numerical results for the chromatic number of the conflict graph of the system graphs noted corresponding to a square lattice with nearest-neighbor hopping for weak and strong coloring problems. 221

12.1 (a) Illustration of our proposed symmetry-conscious randomized measurement scheme, which preserves the symmetry structure of states, compared to symmetry-blind approaches. (b) Applications and benefits include cost reduction when measuring entanglement and finding classical representations of quantum states, allowing a rudimentary symmetry-based error mitigation scheme, and enabling symmetry-resolved measurement of entanglement structure. 225
12.2 Random circuit scheme for systems with particle number symmetry. Blue squares connected by black lines represent the unitary u in Eq. (12.3). 230
12.3 (a) Particle number-respecting random circuits $N_{\mathcal{E}}$ required to approximate a 2-design with precision better than $\epsilon = 10^{-2}$ in every symmetry sector ($\ell = 128$) for a total number of sites $N = 4, 6, 8, 10$. Inset: Relative sampling cost reduction compared with a symmetry-ignorant scheme. (b) Measurements N_M required to estimate sector-wise 2-purities with precision better than $\epsilon = 0.05$, for fixed $N_{\mathcal{E}} = 1,428$ and $\ell = 128$. Here, the states considered are reduced density matrices on subsystems of dimension $N_A = N/2$ (i.e. on a bipartition of the lattice.) 232

12.4 (a) Illustration of $\mathbb{Z}_2$ LGT in $1+1$ dimensions, depicted is an "even" site, $(-1)^j = 1$, where spin up (down) is the presence (absence) of a $\mathbb{Z}_2$ charge, $\mathcal{Q}_j = +1\ (0)$. (b) Single layer of a random-measurement circuit with non-local gates $2+|j_1-j_2|$ qubit gates, an extension of the strategy in section 12.2. (c) Near-term strategy based on 3-qubit unitaries. (d) Illustration of the symmetry structure of ρ_A. 236
12.5 (a) Sector-wise k-purities of the $\mathbb{Z}_2^{1+1}$ ground state measured based on Eq. (12.8) as a function of $N_\mathcal{E}$, for $N_A = N/2 = 5$, $m \cdot a = 0.1$, $e/m = 8$ and $\ell = 32$. (b) Von Neumann Entropy from k-purities/Rényi entropies using a 4^{th} order finite-difference approximation of Eq. (12.14). (c) Symmetry- (symm.), distillable (dist.) and total entanglement entropies as a function of e/m for $N_\mathcal{E}$=2000, $m \cdot a = 0.1$ and $\ell = 32$. Symbols represent random measurement results, dotted lines represent approximating S_E by (exact) Rényi-entropies up to $k = 4$, solid lines are exact results. 237
12.6 (a) Bottom: Shadow-reconstructed symmetry-resolved Schmidt spectrum, $m \cdot a = 0.05$, $e/m = 1$, $\ell = 32$ and $N_\mathcal{S} = 2^{17}$. Inset: Full spectrum. Top: Probability per sector p_s. (b) Sector-wise relative entropy $S(\bar{\rho}_{A,s}||\bar{\sigma}_{A,s})$ between (normalized) exact $\bar{\rho}_{A,s}$ and shadow-reconstructed reduced density matrices $\bar{\sigma}_{A,s}$, as a function of shadow number $N_\mathcal{S}$ with $m \cdot a = 0.05$, $e/m = 1$ and $\ell = 32$. 241
12.7 (a) Illustration of ${\mathbb{Z}_2}^{2+1}$ LGT, including Hilbert space and Gauss law constraints G_j. (b) Random measurement circuits: We work with an even-odd alternating layers consisting of randomly placed 'electric' rotations $R_z(\alpha)$, $R_z(\gamma)$ and plaquette rotations $U_\square(\beta)$ (orange squares), approximating k- designs for sufficient circuit depth. (c) The symmetry structure of ρ_A originates from Gauss laws at entanglement boundaries and a non-local 'ribbon' operator V_x^A spanning the two entanglement cuts. 242
12.8 (a) Sector-wise k-purities of ρ_A of the $\mathbb{Z}_2^{2+1}$ ground state, from Eq. (12.8), for $N_x \times N_y = (3+5) \times 2$, ϵ=0.1 and ℓ=64 layers and fixed BC in y. (b) Bottom: Shadow- ($N_\mathcal{S} = 2^{16}$) versus BW-EHT-reconstructed ($N_\mathcal{E} = 50$, $N_M = 1024$) Schmidt spectrum $P_{s,\lambda}$; with $N_x \times N_y = (3+5) \times 2$, ϵ=0.2 and ℓ=64. Top: probability per sector p_s; inset: entanglement spectrum $\xi_{s,\lambda} = -\log(P_{s,\lambda})$. (c) Entanglement gap Δ_ξ between 'low- and high-energy' parts of the ES as a function of ϵ, reconstructed using the BW-EHT scheme, for $N_x \times N_y = (3+3) \times 2$ (PBC in y), $N_{\text{BW}} = 50$, $N_{\text{shots}} = 1024$, and $\ell = 64$. A horizontal red line is the infinite volume limit, $\epsilon_c = 0.33 \pm 0.01$. Right: Symmetry-resolved Entanglement spectra $\xi_{s,\lambda}$ for $\epsilon = 0.075, 03, 0.5$. 246

F.1 Γ for H_{010}. Each vertex is labeled by a member of X_b. The disconnected edges connect to the boundary vertex ∞ and are labeled by their weights. 346
F.2 Full example of $M_1(E_\Gamma)$ for H_{010}. 348
F.3 Some clausal theory graphs used in the minimal example. Comparing Fig. F.3a to Fig. F.3b shows $G(X_{100}) \not\simeq G(X_{000}) \iff X_{100} \not\equiv X_{000}$. Comparing Fig. F.3a to Fig. F.3c shows $G(X_{100}) \simeq G(X_{001}) \iff X_{100} \equiv X_{001}$. 350

F.4 Visited vertices of Γ in FINDEFFECTIVEVERTICES for Hamming symmetric example on a hypercube for $n = 5, 6, 7$ as in Eq. (F.1). Green vertices are effective vertices, red vertices are checked, but not added to the set of effective vertices, and black vertices are unchecked. Columns of vertices are in Hamming weight order. Note the number of vertices grows exponentially in n, but the number of checked vertices grows only polynomially. 351
F.5 Full example of $M_1(E_\Gamma)$ for $H = -X_{001} - Y_{110} - Z_{100} + Z_{011}$. Note, H is not stoquastic as to have a legible graph with all relevant types of vertices/edges. . . . 353
F.6 Construction of a composite gadget for a Hamiltonian term of the form $X_b Y_{b'} Z_{b''}$. $G_1(b), G_2(b')$, and $G_3(b'')$ are as defined in the main text (see Table 1). 354

H.1 Ratio of our error metric ϵ [given in Eq. (H.1)] to the more standard frame potential error metric ϵ_f [given in Eq. (H.2)], for an approximate unitary 2-design per symmetry sector (labeled by particle number n_f), normalized by d_s, versus number of unitaries $N_\mathcal{E}$ sampled, indicating that the two error metrics are equivalent up to a rescaling of the approximation ratio for sufficiently large $N_\mathcal{E}$. Results are for a system of $N = 10$ sites with particle-number symmetry as described in Sect. 12.2. For both error metrics, we sample directly from the CUE within each symmetry sector. Our error metric is computed by averaging over $|S| = 2000$ non-zero matrix elements. 379
H.2 (a) Error relative to an exact unitary 2-design ϵ (Eq. (12.5)) per particle number symmetry sector (labeled by n_f) versus number of unitaries $N_\mathcal{E}$. For clarity only particle number sectors $n_f \leq 5$ are shown; as they are of equivalent dimension, particle number sectors for $n_f > 5$ have the same behavior as the particle number sector $(10 - n_f)$. The ensemble $\mathcal{E}$ is sampled both from the particle number symmetry-respecting random circuits with different numbers of layers $\ell \in \{4, 8, 16, 128\}$ (triangles) and sampled directly from the CUE within each symmetry sector (squares). Results shown are for $n = 10$ sites and the error is averaged over 2000 non-zero matrix elements. Extrapolation of fits to the error as a function of $N_\mathcal{E}$ at $\ell = 128$ layers for the data pictured here and equivalent data for $N = 4, 6, 8$ sites is used to produce Fig. 12.3a of the main text. (b) The difference between the error ϵ obtained by taking $N_\mathcal{E} = 8192$ samples from the random circuits and taking $N_\mathcal{E} = 8192$ samples directly from the CUE versus number of layers used in the circuits. By $\ell \approx 15$, the difference reaches a floor set by $N_\mathcal{E}$ (i.e. the standard sampling error) with a small dependence on the dimension d_s of the corresponding sector. 382
H.3 (a) The number of ensembles $N_\mathcal{E}$ required, in the infinite shot limit, to estimate the 2-purity and 3-purity of a single block to 5 percent error. (b) Actual k-purities for reduced density matrices on subsystems of size $N_A = N/2$ for each particle number symmetry block for the (normalized) states used in Fig. 12.3(b). 384
H.4 Average error ϵ in 2-design matrix elements (Eq. (12.5)), normalized by $d_s(d_s^2 - 1)$, versus number of random unitaries $N_\mathcal{E}$ sampled from either the symmetry-respecting random circuits of Section 12.3.1 of depth $\ell = 32$ (triangles) or from direct sampling from the CUE within each symmetry sector (squares) for $\mathbb{Z}_2^{1+1}$ LGT for a subsystem of size $N_A = 5$ ($N = 10$). Number of indices sampled is 900. 385

H.5 Average error ϵ in 2-design matrix elements (Eq. (12.5)), normalized by $d_s(d_s^2 - 1)$, versus number of random unitaries $N_{\mathcal{E}}$ sampled from either symmetry-respecting random circuits of depth $\ell = 64$ (triangles) or from direct sampling from the CUE within each symmetry sector (squares) for $\mathbb{Z}_2^{2+1}$ LGT for a subsystem of size 3×2 (with fixed boundary conditions in y). Number of indices sampled is 900. 387
H.6 Relative entropy between exact and shadow-reconstructed symmetry-resolved $\rho_{A,s}$ for the $\mathbb{Z}_2^{2+1}$ LGT ground state at $\epsilon = 0.2$, with $N_x \times N_y = (3 + 5) \times 2$ and fixed BC in y, $\ell = 64$ circuit layers. Data for $N_{\mathcal{S}} = 2^{16}$ are shown in Fig. 12.8(b) of the main text. 388
H.7 Left column: Symmetry resolved Schmidt spectrum $P_{s,\lambda}$, reconstructed using BW-EHT, for $\epsilon = 0.075, 0.3, 0.5$ and with $N_x \times N_y = (3 + 3) \times 2$ and periodic boundary conditions in y (and x), $\ell = 64$, $N_{\mathcal{E}} = 50$, $N_M = 1024$. Right column: Symmetry resolved Entanglement spectrum. 389

Citations to Previously Published Work

Most of the work appearing in this dissertation is either published or has appeared on the preprint server arXiv. The associated chapters and appendices reproduce the previously published work essentially verbatim with only minor typographical edits to ensure a uniformity of structure. For instance, sections, equations, figures, and references are renumbered and possibly resized. For papers that were published in Physical Review Letters the associated Supplemental Material has been included as an appendix in this dissertation. These appendices are now referred to as such in the corresponding chapter. Similarly, one will find the substitution "In this paper" $\rightarrow$ "In this chapter". Such changes are noted in the usual way by brackets around the changed word(s). In a few places where additional commentary felt necessary upon revisiting a work, those comments are made as footnotes, labeled clearly as "*Comment:* [...]". There may be additional minor variation due to grammatical edits made during final typesetting of the journal versions of the included work.

Below, I cite the previously published work associated with each chapter of this dissertation, as well as a few related papers that were not included. An asterix denotes co-first authorship. While significant contributions were made by all co-authors, for the papers I chose to include in this dissertation I led the project and made particularly important contributions.

Part I: Quantum Sensing

- Chapter 2 was originally published as:

A Ehrenberg*, J Bringewatt*, A V Gorshkov. "Minimum entanglement protocols for function estimation." Phys. Rev. Research 5, 033228 (2023)

- Chapter 3 was orginally published as:

  J Bringewatt*, A Ehrenberg*, T Goel*, A V Gorshkov. "Optimal function estimation with photonic quantum sensor networks." Phys. Rev. Research 6, 013246 (2024)

- Chapter 4 was orginally published as:

  T Qian, J Bringewatt, I Boettcher, P Bienias, A V Gorshkov. "Optimal measurement of field properties with quantum sensor networks." Phys. Rev. A (Letter) 103, L030601. (2021)

- Chapter 5 was originally published as:

  J Bringewatt, I Boettcher, P Niroula, P Bienias, A V Gorshkov. "Protocols for estimating multiple functions with quantum sensor networks: geometry and performance." Phys. Rev. Research 3, 033011. (2021)

Related works, not included in this disseration, include:

- L P García-Pintos, K Bharti, J Bringewatt, H Dehghani, A Ehrenberg, N Y Halpern, A V Gorshkov. "Estimation of Hamiltonian parameters from thermal states." Preprint. (2024) [arXiv:2401.10343]

- P Niroula, J Dolde, X Zheng, J Bringewatt, A Ehrenberg, K Cox, J Thompson, M Gullans, S Kolkowitz, A V Gorshkov. "Quantum sensing with erasure qubits." Preprint. (2023) [arXiv:2310.01512]

Part II: Quantum Optimization

- Chapter 7 was originally published as:

  J Bringewatt, W Dorland, S P Jordan. "Polynomial time algorithms for estimating spectra of adiabatic Hamiltonians." Phys. Rev. A 100 (3), 032336 (2019)

- Chapter 8 was orginally published as:

  J Bringewatt*, M Jarret*. "Effective gaps are not effective: quasipolynomial classical simulation of obstructed stoquastic Hamiltonians." Phys. Rev. Lett. 125, 170504 (2020)

- Chapter 9 was orginally published as:

  J Bringewatt, L T Brady. "'Simultaneous stoquasticity." Phys. Rev. A 105, 062601 (2022)

Related works, not included in this dissertion, include:

- J Bringewatt, M Jarret, T C Mooney. "On the stability of solutions to Schrodinger's equation short of the adiabatic limit." Preprint. (2023) [arXiv:2303.13478]

- L P García-Pintos, L T Brady, J Bringewatt, Y-K Liu. "Lower bounds on quantum annealing times." Phys. Rev. Lett. 130, 140601 (2023) [arXiv:2210.15687]

- T C Mooney, J Bringewatt, N C Warrington, L T Brady. "Lefschetz thimble quantum Monte Carlo for spin systems." Phys. Rev. B 106, 214416 (2022) [arXiv:2110.10699]

Part III: Quantum Simulation

- Chapter 11 was originally published as:

  J Bringewatt, Z Davoudi. "Parallelization techniques for quantum simulation of fermionic systems." Quantum 7, 975 (2023)

- Chapter 12 was orginally published as:

  J Bringewatt, J Kunjummen, N Mueller. "Randomized measurement protocols for lattice gauge theories." Quantum 8, 1300 (2024)

Related works, not included in this dissertation, include:

- J D Watson, J Bringewatt, A F Shaw, A M Childs, A V Gorshkov, Z Davoudi. "Quantum algorithms for simulating nuclear effective field theories." Preprint. (2023) [arXiv:2312.05344]

- J Bringewatt, N Sato, W Melnitchouk, J Qiu, F Steffens, M Constantinou. "Confronting lattice parton distributions with global QCD analysis." Phys. Rev. D. 103, 016003 (2021) [arXiv:2010.00548]

Chapter 1: Introduction

> If we suppose that we know all the physical laws perfectly, of course we don't have to pay any attention to computers.[1]
>
> Richard Feynman

1.1 Overview and Motivation

Traditionally, our understanding of the structure of the physical universe has come from the symbiotic interplay of experiments and theory, with the former providing observations of the physical world and the latter providing a cohesive mathematical description of those observations in such a way that suggests both the design and interpretation of future experiments. In the past century, however, digital computers have provided a third leg to the pursuit of formulating and understanding physical laws. As the field of computational physics has matured, it has extended, to great effect, V. I. Arnold's domain of "cheap experiments" from analytic mathematics to numerical "experiments."[2]

From the very first computers—whose early development proceeded in tandem with the

[1] This quote comes from the same lecture [1] as a much more popular quotation, which has been used to kick off many a quantum computing talk. Perhaps due to its novelty, I find this quote more compelling.

[2] The full quote [2]: "Mathematics is a part of physics. Physics is an experimental science, a part of natural science. Mathematics is the part of physics where experiments are cheap." V. I. Arnold is perhaps best known to physicists for a superb textbook on classical mechanics, but his writings on the teaching of mathematics this quote is taken from are also entertaining and thought-provoking reading.

Manhattan project and was accelerated by the subsequent need for numerical simulations of nuclear processes for the design of the hydrogen bomb [3]—each advance in computational technology has gone hand-in-hand with novel scientific discoveries and their application to new tasks. Continuing this trend,[3] in the past several decades, the theoretical, experimental, and numerical groundwork has been laid for a new class of computational technologies—namely, quantum computers, as well as related quantum devices, such as sensors and simulators. Such devices are exciting because they differ not just in degree, but also in kind from traditional, classical technologies. Of course, quantum physics has been behind many important technological advances before, but these new devices promise to make use of the exotic and "genuinely quantum" features of many-body quantum systems, such as entanglement, to further extend the purview of computational and experimental physics to new, previously inaccessible, regimes.

This dissertation is organized around three important classes of such quantum technologies—quantum sensors, analog quantum computers, and digital quantum computers. Within each of these classes, we focus primarily on how to optimally realize a particular set of applications of those technologies—estimation of global properties of fields, combinatorial optimization, and simulation of lattice gauge theories, respectively. Thus, one might reasonably suspect that my driving motivation[4] for the work contained in this dissertation is guided by a very practical question: how can we get quantum technology out into the "real world" to measure gravitational or electromagnetic fields, solve important optimization problems, and teach us about physical phenomena hidden beyond the computational horizon of classical computers? Certainly, achieving

[3] Although, I sincerely hope, with less violent applications than bomb design. The evidence, so far, seems to suggest we may be safe on that particular front.

[4] This initial statement of motivations is the one place in this dissertation I will use "I" or "my" instead of "we" or "our" as the broad motivations I choose to emphasize are my own. My hope is to provide a cohesive understanding of why I, personally, chose to work on the wide variety of topics covered in this dissertation.

these goals is an important motivation for working on these problems, but studying the uses of quantum technology offers something much more than just technological improvements in the service of existing modes of research.

To understand the impact of this "something extra," it is helpful to take a brief journey back to the origins of modern (classical) computer science and its relationship to physics. In 1936, Alan Turing published "On Computable Numbers, with an Application to the Entscheidungsproblem,"[5] which, while not immediately obvious at the time, allowed one to construct the foundations of modern theoretical computer science around the ideas of an abstract universal computing machine, now known as a Turing machine. A Turing machine operates on an unbounded memory tape, divided into discrete cells, on which a "head" can write and erase using a finite set of symbols. The head has an internal state, also from amongst a finite set. Given some set of rules as to how the head moves, writes, erases, or updates its internal state based on its current state and the symbol at its current location, this abstract mechanical device provides a simple model for computation. Remarkably, it appears to be true that any computable function can be computed by this simple device, a statement known as the Church-Turing thesis.

In fact, as the theory of computer science developed throughout the 20th century, a stronger conjecture, known as the extended Church-Turing thesis, came into being: Any "realistic" model of computation can be *efficiently* simulated by a (probabilistic) Turing machine [emphasis added]. For our purposes, the exact details of this statement do not particularly matter:[6] the point of these ideas is that at the highest level of abstraction, one does not have to worry too much about the

[5]This problem, posed by David Hilbert and Wilhelm Ackermann in the late 1920s, asks whether there exists an "algorithm" that for every mathematical statement can decide if it is true or false. Answering this question requires a well-defined notion of what, exactly, an algorithm is. Providing a robust and general definition, and then using it to demonstrate the unsolvability of the Entscheidungsproblem is the core of Alan Turing's contributions.

[6]There are some undefined terms here—the interested reader is suggested to consult Ref. [4], for instance.

physical hardware of a computer when thinking about the broad strokes of algorithm design or the complexity of algorithmically solving particular types of problems.

Quantum computers are an interesting computational platform because they are expected to violate the extended Church-Turing thesis. That is, a computer that takes advantage of quantum mechanical effects seems to be a reasonable model of computation, and, furthermore, it appears that such devices can efficiently perform tasks, such as factoring [5], that classical, Turing-equivalent, machines cannot perform efficiently.[7] Therefore, it appears that we should, at least, care if our computer is a quantum computer or not. From the perspective of a computer scientist this fact, in and of itself, is important and somewhat surprising. From the perspective of a physicist, this fact reminds that there are deep connections between computation and physics and that today's "second quantum revolution" is not just a source of new tools—it also suggests a new, or underappreciated, perspective from which to think about the fundamental differences between classical and quantum mechanical systems.

Beyond the birds-eye view of the extended Church-Turing thesis, which draws a distinction between quantum and non-quantum, anyone familiar with classical high performance computing knows that it can be extremely important to worry about lower levels of abstraction. In broad strokes, the physical task of computing requires the process of mapping one description of a physical system to another (description and physical system). At lower levels of abstraction, we must care deeply about the precise details of such mappings. This is especially important for quantum devices, which, currently, are relatively small and error-prone. Beyond the practical benefits, an emphasis on optimizing the algorithms we employ on quantum hardware can draw attention to regimes and features of quantum physics that a physicist might not otherwise be

[7] Up to some plausible conjectures. Factoring has not been proven to be outside the complexity class P.

drawn to study.

Thus, whether quantum technologies reach their full potential as technology or not, the associated information-theoretic view of physics, obtained via an exploration of how information can be encoded and processed in a physical system, will be an important and lasting legacy of the "second quantum revolution." In this dissertation, I seek to provide a collection of specific examples to support the power of these ideas. Along the way, by considering mappings from problems to devices at various levels of abstraction, I will also lay the groundwork for some important practical applications of quantum technology in the guise of sensing protocols, optimization algorithms, and simulation algorithms.

1.2 Outline of Dissertation

1.2.1 Part I: Quantum Sensing

Quantum sensors are well known to offer greater sensitivity and improved spatial resolution over their classical counterparts. For instance, precise control of the quantum states of individual atoms form the technological basis for the most precise clocks. Their precision floor is set by the so-called standard quantum limit, which holds for unentangled sensors. The theory of how to use entanglement to surpass this limit has been well-established, but is extremely challenging to achieve in practice, with only two examples to date: the LIGO and HAYSTAC experiments which use squeezed states to enhance sensitivity to gravitational wave [6] and (possible) axion dark matter [7] signals, respectively. Therefore, enabling entanglement-enhanced metrology in other scenarios and physical settings is an important and timely area of focus.

In Part I of this dissertation, we will consider the theoretical limits in such a multi-particle

scenario: in particular, we consider the problem of measuring global functions of local parameters (i.e. local scalar fields) in a network of quantum sensors. This problem, first introduced in Refs. [8, 9], lies at the boundary of single and multiple parameter quantum metrology, and, as we shall see, inherits features from both sorts of problems. In particular, we find that one can use single parameter performance bounds to understand the precision limits of measuring a single global function of interest, subject to saturability conditions that depend on multi-parameter bounds.

Chapter 2 sharpens and clarifies this understanding of the problem for the case of measuring a linear function of local, independent parameters coupled to qubit sensors, providing a rigorous algebraic framework to design entire families of optimal sensing protocols that saturate the ultimate precision bounds. In addition, we prove several theorems regarding the minimum amount of entanglement needed by any optimal protocol for this problem.

In Chapter 3, we extend the algebraic approach of the previous chapter to understand photonic sensors—in particular, local parameters coupled via either a number operator (as in a Mach-Zehnder interferometer) or a quadrature operator. While this is a similar problem to the case of qubit sensors, in this setting the relevant resource is the (average) number of photons, as opposed to the time spent coupled to the local parameters. Critically, unlike time, photon number is not a "parallel" resource—using some photons to gain sensitivity to one parameter necessarily requires not using those same photons for another parameter. This distinction leads to different bounds and different protocols, as well as different entanglement requirements. In the process of performing this analysis, we prove a long-standing conjecture [9] about the ultimate performance limits for measuring a linear function of local phases in a network of Mach-Zehnder interferometers.

Finally, we consider certain generalizations of the function estimation problem in networks

of quantum sensors. In Chapter 4, we relax the requirement that the local parameters be independent, and, in Chapter 5, we consider protocols for estimating multiple global functions. In both cases, we see that the problem gains structure that we can exploit to better leverage entanglement for enhanced sensing.[8] Another generalization worth noting: the case of analytic functions reduces, asymptotically, to the case of linear functions primarily considered in this dissertation [10]. In these pages, this reduction is considered explicitly only in the context of dependent local parameters (Chapter 4).

In Chapter 6, we will make concluding remarks on the topics in Part I and turn our attention forward to a few open questions.

1.2.2 Part II: Quantum Optimization

Optimization problems are ubiquitous. Ranging from linear programming problems to combinatorial optimization problems, they find application in many real-world tasks like path routing, scheduling, and compilation, to list but a few examples. Given that many of the optimization problems that are of the greatest interest are NP-hard (consider, for instance, the prototypical example of the Traveling Salesman problem), in practice, one seeks effective heuristics to find approximate solutions on particular problem instances. One might hope that quantum computers could provide a boost to such heuristics, finding better solutions more efficiently than classical algorithms.

The jury is still out on whether we can expect this to be a meaningful use case for quantum computers [11], but, in pursuit of understanding this potential application, a number of interest-

[8] Chronologically, these papers were actually published before the papers in Chapters 2 and 3, but in the context of this dissertation, these chapters seem to be a better entry point.

ing questions arise regarding the possible origins of quantum advantage over classical devices. In Part II of this dissertation, we will consider the resources leading to quantum advantage for *analog* quantum computers. Specifically, we will focus, primarily, on adiabatic quantum computing [12]. One should keep in mind, however, that adiabatic algorithms can be digitized via Trotterization for use on digital quantum computers.[9]

In adiabatic quantum computation, one begins with the system in some easy-to-prepare ground state of an initial Hamiltonian H_0 and then adiabatically (slowly) changes the Hamiltonian to some H_1, so that, at the end of this process, the state of the system is the ground state of H_1. This ground state should encode the solution to the computational problem of interest. For general sparse qubit Hamiltonians, this procedure is a universal model of quantum computation [13], and, thus, if quantum computers offer any advantage over classical computers in the digital setting, they do here as well. However, the universality construction is not an immediately practical one due to its reliance on non-geometrically local 3-body terms. Thus, often one considers a reduced set of Hamiltonians tailored to a problem class of interest.

For instance, when solving a combinatorial optimization problem, H_1 is a diagonal Hamiltonian whose entries encode the cost function one seeks to minimize. Oftentimes, H_0 is taken to be a transverse field $H_0 = \sum_{j=1}^{n} X_j$ where n is the number of qubits and X_j is the Pauli-X operator. This has theoretical justification—the ground state of H_0 is then an even superposition over all bit strings and, thus, corresponds to a uniform prior of sorts[10]—and a practical one—this is exactly the sort of Hamiltonian to which the quantum computing company D-WAVE has access.[11]

[9] Trotterization will be discussed in the context of quantum simulation algorithms in Chapter 11

[10] If we truly perform adiabatic evolution the initial state is irrelevant. However, once one considers the evolution time required to guarantee negligible transitions out of the ground state the notion of a prior becomes important.

[11] D-WAVE produces a "quantum annealer," as opposed to a full-blown quantum computer. The Hamiltonians

What is not clear is whether such Hamiltonians allow for any quantum advantage. In particular, these sorts of Hamiltonians are k-local (i.e. consist of k-body interactions for some constant k), *stoquastic* Hamiltonians. Stoquastic Hamiltonians are those with real, non-positive off-diagonal elements; the k-locality constraint is a physical one, as realized, for instance, in the D-WAVE hardware. Crucially, when attempting to simulate adiabatic quantum computation with stoquastic Hamiltonians via quantum Monte Carlo (a classical algorithm) one does not face a sign problem. The sign problem, which arises when attempting to sample from a quasiprobability distribution, leads to exponential slowdowns for quantum Monte Carlo. Therefore, by avoiding this problem it is conceivable that classical algorithms can efficiently simulate adiabatic quantum computation with k-local stoquastic Hamiltonians. That is, one might suspect that non-stoquasticity is an essential ingredient for quantum advantage in adiabatic quantum computation. Definitively verifying or ruling out such suspicions is the ultimate goal of the work contained in Part II of this dissertation.

A first approach to understanding the role of stoquasticity in quantum advantage via adiabatic quantum computing (including work done by the author [14]) consisted of constructing examples of exponential separations between the performance of adiabatic quantum computation with stoquastic Hamiltonians and *particular* quantum Monte Carlo algorithms [14, 15]. Such results are fine-tuned, however, and require leveraging heavy amounts of symmetry to enable rigorous analysis. Furthermore, while suggestive, such examples are not proofs of obstructions for any classical algorithm.

that they can implement consist of a global transverse field and diagonal Ising-like interactions. Such Hamiltonians are not universal, but, optimistically, could still provide speed-ups compared to classical algorithms for solving certain optimization problems. Understanding what, if any, quantum advantage these Hamiltonians can provide has historically been influential in determining the sorts of theoretical questions the quantum annealing and adiabatic quantum computing community has sought to address.

In Chapter 7, we present work aimed at extending the class of problems whose adiabatic performance could be efficiently analyzed. Our new class of problems still rely on symmetries to accomplish this task, but they provide a much broader set of examples, including more realistic ones with many local minima.

While the new class of toy problems developed in Chapter 7 could reasonably find broader application,[12] for the purposes of understanding the quantum advantage (or lack thereof) for k-local stoquastic Hamiltonians, Chapter 7 is subsumed by the work in Chapter 8, where we prove that a broad class of examples that leverage polynomial-sized symmetric subspaces to create a potential quantum advantage for adiabatic quantum computation with k-local quantum Hamiltonians, can, in fact, be classically simulated in quasi-polynomial time. This includes all previous examples of separations and leaves us, to date, with no known examples of possible exponential separations.[13]

The proof in Chapter 8 crucially depends on k-locality as well as stoquasticity. This is particularly interesting when one compares to the result by Hastings [18], that came out shortly after the paper Chapter 8 is based on. In Ref. [18] and the follow-up work in Ref. [19], the authors prove a super-polynomial separation between adiabatic quantum computation with sparse,

[12] For instance, I expect such toy models to be worth investigating in relation to my recent work [16], not included in this dissertation, on quantum speed limit-based lower bounds on quantum annealing times. See Chapter 10 for some further discussion.

[13] The separation between Shor's factoring algorithm and the best known classical algorithm for factoring is sub-exponential time, so one might wonder if our result can similarly be interpreted as still leaving room for a relevant separation between classical and quantum algorithms for these highly symmetric examples. The answer is likely no. For one, at a technical level, there is still a separation between the sub-exponential cost of factoring, with running time $2^{o(n)}$ for problem size n, and the quasi-polynomial cost of our algorithm, with running time $O(n)2^{\mathrm{polylog}(n)}$. Second, the quasi-polynomial feature of our classical symmetry finding algorithm comes from mapping the problem to graph isomorphism. While both factoring and graph isomorphism belong to the small set of problems believed to be NP-intermediate, graph isomorphism seems to be qualitatively different from factoring from the perspective of both classical and quantum computation. For example, while graph isomorphism, like factoring, can be reduced to an instance of the Hidden Subgroup Problem, the relevant instances provably do not admit efficient solutions via quantum algorithms in the same way that the instances corresponding to factoring do [17].

stoquastic Hamiltonians and classical computation. The construction is manifestly not k-local, however, which, when coupled with our work, suggests that both k-locality and stoquasticity are important ingredients for (possibly) limiting quantum advantage.

In Chapter 9, we consider a related mathematical problem. As is clear from the definition, stoquasticity is a basis dependent feature of a Hamiltonian. Therefore, presented with a Hamiltonian one can seek to (efficiently) find a basis that makes it stoquastic. There is a rich literature on the complexity theory of this problem in the case of a single Hamiltonian. In this chapter, we consider the problem of whether a collection of Hamiltonians admits a basis such that all Hamiltonians are simultaneously stoquastic. This work is relevant to adiabatic quantum computation, where one interpolates between Hamiltonians—and possibly uses additional "catalyst" terms during the interpolation. In particular, to apply quantum Monte Carlo to simulate quantum adiabatic computation without a sign problem one, at least naively, needs the problem presented in a fixed basis such that all relevant Hamiltonians are stoquastic.

Clearly, a number of interesting questions remain open. Can we, in fact, prove that adiabatic quantum computation with k-local Hamiltonians is efficiently classically simulable? In Chapter 10, we discuss the outlook for answering this question and other directions of interest. Of particular note, we briefly summarize some promising recent work on rigorous analysis of quantum annealing beyond the adiabatic regime [16, 20].

1.2.3 Part III: Quantum Simulation

In the long term, fault-tolerant digital quantum simulation is the most exciting use case for a quantum computer—at least for a physicist. Quantum systems exhibit a number of features

that make them, generically, intractable to simulate on classical computers. For instance, (1) quantum systems have an exponentially large state space, which, at least for naive representations (and/or generic states), leads to exponential storage and simulation costs; (2) as discussed in the previous section, classically simulating quantum systems often leads to sign problems, e.g. when computing equilibrium properties of fermionic or frustrated spin systems, or when attempting to simulate time-evolution. Quantum computers offer a natural solution to such problems: simply, simulate quantum systems with other quantum systems—using either a general purpose quantum computer or a more limited quantum simulator.

In Part III of this dissertation, we consider two problems relevant for the simulation of lattice (gauge) theories on a digital quantum computer. In Chapter 11, we consider the problem of mapping fermionic operators to qubit operators, with an emphasis on the interplay between the choice of mapping and the underlying qubit architecture. In particular, we demonstrate how to select fermion-to-qubit mappings to enable parallelization of a Trotter-based Hamiltonian simulation algorithm. This is done by treating the problem of finding a good fermion-to-qubit mapping as a path coloring problem on a certain graph, which depends on both the physical hardware and the abstract fermion-to-qubit mapping. These results are particularly relevant for implementation on near-to-intermediate term digital quantum simulators where circuit depth is a key bottleneck and working at the lower levels abstraction in the computational stack are of prime importance.

In Chapter 12, we consider the problem of extracting information from the output state of a quantum simulation. While such a quantum state may encode exponential information, only a minuscule amount of this information is typically revealed by a single measurement. Thus, it is important to be able to efficiently extract relevant quantities of interest. Randomized measurement protocols, including classical shadows [21], are one approach to this problem. In this work,

we demonstrate how one can enhance the randomized measurement toolbox [22] by making use of known symmetries in the system of interest. While the approach holds for general symmetries, we especially focus on $\mathbb{Z}_2$ lattice gauge theory. A key application is the study and verification of topologically ordered phases in synthetic quantum materials.

Finally, in Chapter 13, we briefly discuss some other related work [23], and provide suggestions for future explorations.

Part I

Quantum Sensing

Chapter 2: Minimum Entanglement Protocols for Function Estimation

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 $\Delta\theta \sim 1/\sqrt{d}$ [24], which can be surpassed only if the states are prepared in an entangled state [25, 26]; if $O(d)$-partite entanglement is used, the Heisenberg limit $\Delta\theta \sim 1/d$ can be achieved [27–29]. The necessity of entanglement for optimal measurement has also been explored in numerous other contexts [30, 31]; for instance, in sequential measurement schemes (where one may apply the encoding unitary multiple times) [32, 33], in the presence of decoherence [34–37], when the coupling Hamiltonian is non-linear [38–40], or in reference to resource theories for metrology [41–44].

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 [45–48], 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 $q(\boldsymbol{\theta})$ [9, 49–57], which is a crucial element of optimal protocols for more general quantum sensor network problems (the case of measuring one or multiple analytic functions [10, 58] and the case where the parameters $\boldsymbol{\theta}$ are not independent [59] 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 [9, 49, 54].

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 [9, 49, 51–54]. 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. [49] established that the fundamental limit for the mean square error $\mathcal{M}$ of an estimator for

[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.

q is

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

where t is the total evolution time.

Eq. (2.2) is derived via the single-parameter quantum Cramér-Rao bound [38, 45–48]. 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 $\hat{g}_q$ [38]. 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) [49]. 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 [60, 61]. For pure probe states and unitary evolution for time t under the Hamiltonian in Eq. (2.1), it has matrix elements [61]

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

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 $\mathcal{F}(\boldsymbol{q}) = J^\top \mathcal{F}(\boldsymbol{\theta}) J$. 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 $\boldsymbol{q} \to \boldsymbol{\theta}$, 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 2.5.1) 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. [62]) and quantum estimation theory [63–65]. One finds similarly derived bounds in these contexts.[2] However, the protocols we now describe, and especially their entanglement features, are new to this work.

[2]For instance, the conditions in Eqs. (2.6)-(2.7) are equivalent to the so-called global parameter orthogonality condition discussed in Sect. 5.5 of Ref. [65].

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 $\hat{H}(s)$ 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.1.1 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 en-

coding 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 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 $\boldsymbol{p} = (p_1, \cdots, p_N)^\top$ 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 [69–71]. We could assume that this information is known *a priori* [49], 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. [66–68]

[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. [66–68], 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 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.

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 $|\psi(\boldsymbol{\tau}^{(\overline{N})}; qt_k)\rangle$.

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 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 2.5.1 applies to any protocol making use of a Hamiltonian described via Eq. (2.1).

Theorem 2.5.1 (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 2.5.1 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 [72, 73], 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.1.1 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 2.5.1 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 2.6.1 (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, $\operatorname{sgn}(T_{ij}) \in \{0, \operatorname{sgn}(\alpha_j)\}$.*

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) $\operatorname{sgn}(c_i) = \operatorname{sgn}(\alpha_i)$. 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 2.6.1. *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 2.5.1, 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 2.5.1.

Theorem 2.6.1. *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 2.5.1. 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 $\mathrm{sgn}(T_{ij}) = \mathrm{sgn}(\alpha_i)$ 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.[4] 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

[4] These gates are not strictly necessary. See footnote [3, above].

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. [49] 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 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 $\boldsymbol{\tau} = (1, 0, \ldots, 0)^\top$. 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. [49] 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.[5]

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 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 in-

[5]Code is available upon request.

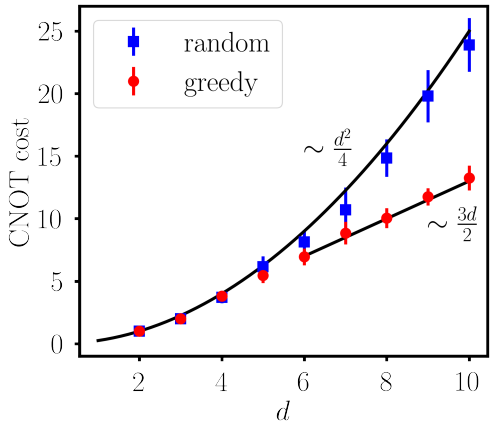


Figure 2.1: CNOT costs versus number of sensors d for minimum entanglement protocols using d optimally ordered states chosen either randomly or via the greedy algorithm described above. Twenty randomly chosen instances (that do not fail) to yield a valid protocol via the greedy algorithm. When it returns 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.

stantaneous 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) [49, 54], 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.1.1 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

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

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.1.1), 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 [49, 59], 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 [10, 58, 59].

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 work. 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. [49] 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 of itself, an important contribution of this work.

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. [49]). 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 2.5.1. 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 2.5.1 is satisfied for all r; that is, we ensure that *within* each indepen-

dent set we obtain the optimal variance for the linear function restricted to that set. The result is a variance

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

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 2.5.1. 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.

Chapter 3: Optimal Function Estimation with Photonic Quantum Sensor Networks

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 [74], the geometry of quantum state space [46], quantum speed limits [75–77], and quantum control theory [76].

Initial experimental and theoretical work on quantum sensing focused on optimizing the estimation of a single unknown parameter (see e.g. Ref. [78] for a review). More recently, the problem of distributed quantum sensing has become an area of particular interest [79]. 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 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)^T$ is particularly well-studied both theoretically [9, 49–51, 53–

55, 57, 80–85] and experimentally [86–89]. 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 sub-task of more general metrological tasks, such as measuring an analytic function of the unknown parameters [10], measuring an analytic function of dependent parameters [90, 91], or measuring multiple functions [58, 92].

For qubit sensors, the asymptotic limits on performance for these function estimation tasks are rigorously understood, and techniques for generating optimal protocols subject to various constraints, such as limited entanglement between sensors, are known [83]. However, despite extensive theoretical and experimental research on distributed quantum sensing for photonic quantum sensors (see e.g. [79, 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 (i) the number operator $\hat{n}$ or (ii) a field-quadrature operator, chosen without loss of generality to be the momentum quadrature $\hat{p} := i(\hat{a}^\dagger - \hat{a})/2$. That is, we are interested in determining a function of either unknown local phase shifts or unknown quadrature displacements. For case (i), our primary focus, we derive this bound subject to a strict constraint on photon number, proving a long-standing conjecture appearing in Ref. [9]. In case (ii), 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 [81], but, for completeness, we include derivations in this setting using an equivalent mathematical framework to the number operator case and the qubit sensor case [83]. This allows for a natural comparison of the various performance limits and resource require-

ments 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 the [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 $j \in \{1, \cdots, d\}$ 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 $j \in \{1, \cdots, M\}$ 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 is done via the unitary

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

where $V := \exp(-i\hat{\boldsymbol{g}} \cdot \boldsymbol{\theta})$ and $U^{(m)}$ for $m \in \{1, \cdots, M\}$ 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. [49, 83], 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

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

Armed with Eq. (3.6), we can write a bound on $\mathcal{M}$ in terms of a single-parameter quantum Cramér-Rao bound [25, 38, 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 $\mu = 1$.[1] 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 $\rho = |\psi\rangle\langle\psi|$.

[1] 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 [66–68]. In Appendix B.7, for completeness, we briefly summarize this approach. See also, Refs. [63, 65] and Ref. [83] for further discussion of these issues.

Ultimately, we seek a choice of new basis that yields the tightest possible bound on the quantum Fisher information $\mathcal{F}(q)$. This choice is determined by the solution to[2]

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

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

[2]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.

the other hand, we must work in the fixed photon sector, as using fixed average photon number allows for the construction of pathological probe states enabling arbitrarily precise sensing. In particular, consider the state

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

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 the [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., $\mathcal{N} = \emptyset$),

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

proving a long-standing conjecture from Ref. [9] 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:

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

Eq. (3.15) is a minor generalization of the results in Refs. [79, 81], 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. [79, 81] 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 [9]

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

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

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

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. [80] 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 $|\boldsymbol{\alpha}| \propto (1, 1, \cdots 1)^T$. 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. [49], 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. [9, 51] 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,

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

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.

Similarly, in the case of measuring a linear function of displacements using states with fixed average photon number, Ref. [80] provides a protocol that, up to small multiplicative constants, 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 only require 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 multi-mode 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. [80] and the bound for arbitrary states derived in Ref. [81] 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 [61] 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 [61].

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. [70], with some convenient, nearly optimal, protocols presented in Refs. [66–68]. 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 only be obtained up to a constant $\geq \pi^2$ [71]. 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

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

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 $\overline{N}$. Eqs. (3.24)-(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 3.4.1. *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*

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

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 3.4.1 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 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 3.4.1. 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.5.2 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 $r_n \in \{0, 1, \cdots, M\}$ 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

$$\varphi_{\text{tot}} = M\sum_n p_n(\boldsymbol{\omega}_n\cdot\boldsymbol{\theta}) \tag{3.30}$$

$$=: (W\boldsymbol{r})\cdot\boldsymbol{\theta}. \tag{3.31}$$

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. [9], 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)-(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)-(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 [94], 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 3.4.1 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 [83]. 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 inter-mode entanglement required during the sensing process. This was considered in Ref. [83] 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 2-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 only allowed 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 non-trivial mode for at least one pass. For displacement operator coupling, subject to a fixed average photon number constraint, an essentially identical argument allow 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 case, so these trivial bounds can be tight. This triviality is in contrast to the qubit case, where results analogous to Lemma 3.4.1 lead to significantly tighter constraints on the minimum amount of necessary entanglement for optimal protocols [83]. 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.

	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)	$\geq \frac{\|\boldsymbol{\alpha}\|_2^2}{t^2}$ [49]	$\geq \frac{\|\boldsymbol{\alpha}\|_{2/3}^2}{N^2t^2}$ [9]	$\geq \frac{\|\boldsymbol{\alpha}\|_1^2}{4\overline{N}t^2}$
MSE (entangled)	$\geq \frac{\|\boldsymbol{\alpha}\|_\infty^2}{t^2}$ [49]	$\geq \frac{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}^2}{N^2t^2}$	$\geq \frac{\|\boldsymbol{\alpha}\|_2^2}{4\overline{N}t^2}$ [81]
Entanglement needed (discrete controls)	$k \geq \max\left\{\left\lceil \frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty}\right\rceil, \left\lceil \frac{\|\boldsymbol{\alpha}\|_0}{M}\right\rceil\right\}$	$k > \left\lceil \frac{\|\boldsymbol{\alpha}\|_0}{M}\right\rceil$	$k \geq \left\lceil \frac{\|\boldsymbol{\alpha}\|_0}{M}\right\rceil$
Entanglement needed (arbitrary controls)	$\frac{\|\boldsymbol{\alpha}\|_1}{\|\boldsymbol{\alpha}\|_\infty} \in (k-1, k]$ [83]	$k = 2$	no entanglement
k-partite entanglement protocol always exists?	yes [83]	no	yes

Table 3.1: A 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.

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 estimation with number operator coupling [9] and showed that some of the protocols that exist in the literature [9, 51, 80], 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 [83]. 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.

Chapter 4: Optimal Measurement of Field Properties with Quantum Sensor Networks

4.1 Introduction

It is well established that entangled probes in quantum metrology can be used to obtain more accurate measurements than unentangled probes [26, 28, 34, 95–97]. In particular, while measurements of a single parameter using d unentangled probes asymptotically obtain a mean squared error (MSE) from the true value of order $\mathcal{O}(1/d)$, using d maximally entangled probes, each coupled independently to the parameter, one obtains an MSE of order $\mathcal{O}(1/d^2)$ – the so-called Heisenberg limit [95, 98]. More recently, understanding the role of entanglement and generalizing this scaling advantage to the measurement of multiple parameters at once or functions of those parameters has been an area of keen interest [8, 51–53, 80, 82, 86, 97, 99–107] due to a wide array of practical applications [108–114]. Importantly, optimal bounds and protocols have been derived for measuring analytic functions of independent parameters, each coupled to a qubit sensor in a so-called quantum sensor network [102]. The problem of directly measuring a spatially dependent field of known form, possibly with extra noise sources, has also been considered [105].

In this [chapter], we consider the following very general problem that is relevant for many

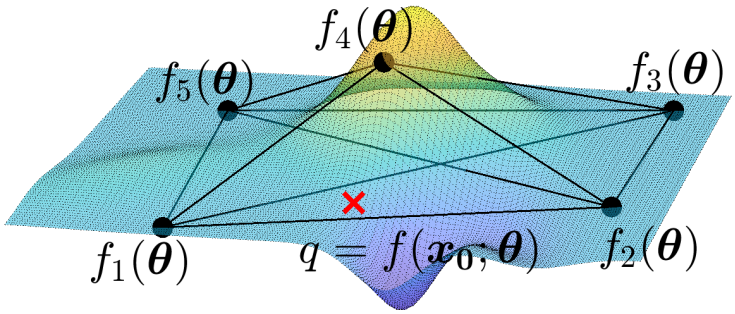


Figure 4.1: At each position $\boldsymbol{x}_i$, a quantum sensor (black dots) is coupled to a field $f(\boldsymbol{x}; \boldsymbol{\theta})$, whose functional form is known, but the parameters $\boldsymbol{\theta}$ are not. The protocols presented here utilize entanglement to obtain the highest accuracy allowed by quantum mechanics in estimating the quantity $q(\boldsymbol{\theta})$. One example problem is to estimate the field value $q = f(\boldsymbol{x}_0; \boldsymbol{\theta})$ at a location $\boldsymbol{x}_0$ (red cross) without a sensor.

technological applications of quantum sensor networks. A set of quantum sensors at positions $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d\}$ is locally probing a physical field $f(\boldsymbol{x}; \boldsymbol{\theta})$, which depends on a set of parameters $\boldsymbol{\theta} \in \mathbb{R}^k$, where we have used boldface to denote vectors. We assume that we know the functional form of $f(\boldsymbol{x}; \boldsymbol{\theta})$, but we do not know the values of the parameters $\boldsymbol{\theta}$. For instance, these parameters may be the positions of several known charges, and $f(\boldsymbol{x}; \boldsymbol{\theta})$ one of the components of the resulting electric field. Our objective is to measure a function of the parameters $q(\boldsymbol{\theta})$. This could be, for instance, the field value $q(\boldsymbol{\theta}) = f(\boldsymbol{x}_0; \boldsymbol{\theta})$ at a position $\boldsymbol{x}_0$ without sensor, or the spatial average $q(\boldsymbol{\theta}) = \int_R \mathrm{d}\boldsymbol{x}\, f(\boldsymbol{x}; \boldsymbol{\theta})$ over some region R of interest. In the following, we derive saturable bounds on the precision for measuring $q(\boldsymbol{\theta})$ using quantum entanglement. The setup is depicted in Fig. 4.1.

As a more concrete example, consider a network of three quantum sensors that are locally coupled to a field $f(\boldsymbol{x}; \theta_1, \theta_2)$ parametrized by $\boldsymbol{\theta} = (\theta_1, \theta_2)$. The field amplitudes at the positions of the sensors shall be $f_1(\boldsymbol{\theta}) = \theta_1$, $f_2(\boldsymbol{\theta}) = \theta_2$, $f_3(\boldsymbol{\theta}) = \theta_1 + \theta_2$, respectively, where we have introduced the shorthand notation $f_i(\boldsymbol{\theta}) = f(\boldsymbol{x_i}; \boldsymbol{\theta})$. Assume we want to measure the value of $q(\theta_1, \theta_2) = \theta_1$. One possible strategy is to simply use the first sensor to measure $f_1(\boldsymbol{\theta})$. On

the other hand, we could also measure $\frac{1}{2}(f_1(\boldsymbol{\theta}) - f_2(\boldsymbol{\theta}) + f_3(\boldsymbol{\theta}))$, thereby potentially gaining accuracy by harnessing entanglement between the individual sensors. In fact, there are infinitely many variations of the second strategy, and we eventually expect some of them to be superior to the first strategy.

In contrast to previous work [102], where one considers estimating a given function $F(f_1(\boldsymbol{\theta}), \ldots, f_d(\boldsymbol{\theta}))$ of independent local field amplitudes $f_1(\boldsymbol{\theta}), \ldots, f_d(\boldsymbol{\theta})$, we consider here the problem of estimating a function of the parameters, $q(\theta_1, \ldots, \theta_k)$, instead. Due to the correlation of the local field amplitudes, there are many measurement strategies that need to be considered and compared in terms of accuracy. In this [chapter], we determine the optimal protocol for this very general setup.

In applications, one often measures field amplitudes that depend on the same set of parameters. Therefore, by allowing for the estimation of quantities that depend on measurements of correlated field amplitudes, this work addresses many problems of practical interest, left unsolved by previous work. These applications include optimal spatial sensor placement and field interpolation. As a physically motivated example, we explicitly demonstrate how our protocol may be applied to a toy version of the field interpolation problem.[1] In addition to finding the optimal attainable variance and a corresponding protocol for a wide class of problems of practical significance, another primary contribution of our work is the use of optimization duality theorems in the derivation of quantum Cramér–Rao bounds, a technical approach we anticipate being of use beyond the scope of this specific problem.

[1] Ref. [102] discusses the field interpolation problem as well, but only solves the problem for the case $k = d$.

4.2 Problem Setup

We formally consider a quantum sensor network as a collection of d quantum subsystems, called sensors, each associated with a Hilbert space $\mathcal{H}_i$ [9, 51]. The full Hilbert space is $\mathcal{H} = \bigotimes_{i=1}^{d} \mathcal{H}_i$. We imprint a collection of field amplitudes $\boldsymbol{f}(\boldsymbol{\theta}) = (f_1(\boldsymbol{\theta}), \ldots, f_d(\boldsymbol{\theta}))^T$ onto a quantum state, represented by an initial density matrix ρ_{in}, through the unitary evolution $\rho_{\text{f}} = U(\boldsymbol{f})\rho_{\text{in}}U(\boldsymbol{f})^\dagger$. Here, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$ is a set of independent unknown parameters. To be specific, we consider qubit sensors and a unitary evolution generated by the Hamiltonian

$$\hat{H} = \hat{H}_{\text{c}}(t) + \sum_{i=1}^{d} \frac{1}{2} f_i(\boldsymbol{\theta}) \hat{\sigma}_i^z, \tag{4.1}$$

with $\hat{\sigma}_i^{x,y,z}$ the Pauli operators acting on qubit i and $f_i(\boldsymbol{\theta}) = f(\boldsymbol{x}_i, \boldsymbol{\theta})$ the local field amplitude at the position of the i^{th} sensor. Our results apply to more general quantum sensor networks (see Outlook). The term $\hat{H}_{\text{c}}(t)$ is a time-dependent control Hamiltonian that we choose, which may include coupling to ancilla qubits. This time-dependent control is not necessary to achieve an optimal protocol [8, 115], but one may use such control to design optimal protocols with simpler requirements on the choice of input state ρ_{in} [8].

Our goal is to estimate a given function of the parameters $q(\boldsymbol{\theta})$ at their (unknown) true value, which we denote as $\boldsymbol{\theta}'$. The estimate of this quantity $q(\boldsymbol{\theta}')$ is based on measurements of the final state ρ_{f}, specified by a set of operators $\{\hat{\Pi}_\xi\}$ that constitute a positive operator-valued measure (POVM) with $\int \mathrm{d}\xi\, \hat{\Pi}_\xi = 1$. We repeat this experiment many times and estimate $q(\boldsymbol{\theta}')$ via an estimator $\tilde{q}$ obtained from the data. On a more technical level, we assume that the sensor placements allow us to obtain an estimate of $\boldsymbol{\theta}'$, which ensures the problem is solvable

[2]. This assumption implies that the number d of quantum sensors should be larger than k. (See Outlook for cases where we can violate this assumption.) The choice of initial state ρ_{in}, control Hamiltonian $\hat{H}_{\text{c}}(t)$, and POVM $\{\hat{\Pi}_\xi\}$ defines a *protocol to estimate* $q(\boldsymbol{\theta}')$.

Before proceeding, let us fix some notation. We emphasize that $\boldsymbol{\theta}$ is treated as a variable, with unknown true value (given by the physical fields) denoted $\boldsymbol{\theta}'$. Thus $q(\boldsymbol{\theta})$ is a function, whereas $q(\boldsymbol{\theta}')$ is a specific number obtained by evaluating the function at the true value $\boldsymbol{\theta}'$. We derive our bounds as functions of this general $\boldsymbol{\theta}$ for wherever $q(\boldsymbol{\theta})$ is analytic, but importantly the ultimate bound depends on evaluation at the true value $\boldsymbol{\theta}'$. We use indices $i, j = 1, \ldots, d$ to label quantum sensors and $m, n = 1, \ldots, k$ to label parameters.

The MSE of the estimate $\tilde{q}$ from the true value $q(\boldsymbol{\theta}')$ is given by

$$\mathcal{M} = \mathbb{E}\left[\left(\tilde{q} - q(\boldsymbol{\theta}')\right)^2\right] = \operatorname{Var}\tilde{q} + \left(\mathbb{E}\left[\tilde{q}\right] - q(\boldsymbol{\theta}')\right)^2, \tag{4.2}$$

where the first and second terms are the variance and estimate bias, respectively. We define the optimal protocol to measure $q(\boldsymbol{\theta}')$ as the one that minimizes $\mathcal{M}$ given a fixed amount of total time t. To determine the optimal protocol, we first derive lower bounds on $\mathcal{M}$ using techniques from quantum information theory. We then construct specific protocols that saturate these bounds.

4.3 MSE Bound

In this section, we derive a saturable lower bound on $\mathcal{M}$ that can be achieved in time t.[3] To derive our bound, we begin with the following result on single-parameter estimation from

[2]Formally, we assume the ability to make an asymptotically (in time t per measurement run and in the number of measurement runs μ) unbiased, arbitrarily-small-variance estimate. See [Appendix C] for detailed definitions

[3]Technically, to saturate our bounds, one requires μ measurements and thus a total time of μt over all experimental runs. However, we avoid this technicality for notational clarity.

Ref. [115]. If the unitary evolution of the quantum state is controlled by a single parameter q, then

$$\mathcal{M} \geq \frac{1}{\mathcal{F}_Q} \geq \frac{1}{t^2 ||\hat{h}_q||_{\mathrm{s}}^2}, \tag{4.3}$$

where $\mathcal{F}_Q$ is the quantum Fisher information, $\hat{h}_q = \partial \hat{H}/\partial q$ is the generator with $\gamma_{\max}$ ($\gamma_{\min}$) its largest (smallest) eigenvalue, and $||\hat{h}_q||_{\mathrm{s}} = \gamma_{\max} - \gamma_{\min}$ is the seminorm of $\hat{h}_q$. The first inequality is the quantum Cramér–Rao bound [45–48].

It is not obvious that Eq. (4.3) may be applied to the problem of estimating $q(\boldsymbol{\theta})$ as we have $k > 1$ parameters controlling the evolution of the state. However, we circumvent this issue by considering an infinite set of imaginary scenarios, each corresponding to a choice of artificially fixing $k - 1$ 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, and we can apply Eq. (4.3) to the imaginary scenario under consideration.

We note that any such imaginary scenario requires giving ourselves information that we do not have in reality. However, additional information can only result in a lower value of $\mathcal{M}$. Therefore, any lower bound on $\mathcal{M}$ derived from any of the imaginary scenarios is also a lower bound for estimating the function $q(\boldsymbol{\theta})$. For a bound derived this way to be saturable, there must be some choice(s) of artificially fixing $k - 1$ degrees of freedom that does not give us *any* useful information about $q(\boldsymbol{\theta})$, and thus yields the sharpest possible bound. This is, in fact, the case. In our analysis below, the existence of such a choice becomes self-evident since we present a protocol that achieves the tightest bound. However, in the [Appendices], we prove that such a choice exists purely on information theoretic grounds.

More formally, consider a basis $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_k\}$ such that, without loss of general-

ity, $\boldsymbol{\alpha}_1 = \nabla q(\boldsymbol{\theta}') =: \boldsymbol{\alpha}$. We then consider any choice of the remaining basis vectors. For any such choice, let $\boldsymbol{\alpha}_n$ correspond to a function $q_n(\boldsymbol{\theta}) = \boldsymbol{\alpha}_n \cdot \boldsymbol{\theta}$. Therefore, if we consider a particular choice of basis, we are also considering a corresponding set of functions $\{q_1(\boldsymbol{\theta}) = q(\boldsymbol{\theta}), q_2(\boldsymbol{\theta}), \cdots, q_k(\boldsymbol{\theta})\}$. We suppose we are given the values $\{q_n(\boldsymbol{\theta}')\}_{n\geq 2}$, fixing $k-1$ degrees of freedom. The resulting problem is now determined by a single parameter, and Eq. (4.3) applies.

The derivative of H with respect to q, while holding $q_2, \ldots, q_k$ fixed, is

$$\hat{h}_q = \frac{\partial \hat{H}}{\partial q}\Big|_{q_2,\ldots,q_k} = \sum_{i=1}^{d} \frac{1}{2}(\nabla f_i(\boldsymbol{\theta}') \cdot \boldsymbol{\beta})\hat{\sigma}_i^z, \tag{4.4}$$

where $\boldsymbol{\beta} = \left(\frac{\partial \theta_1}{\partial q}, \ldots, \frac{\partial \theta_k}{\partial q}\right)|_{q_2,\ldots,q_k}$. Using the chain rule, we find that $\boldsymbol{\beta}$ satisfies $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1$.

As we show formally in the [Appendices], every $\boldsymbol{\beta} \in \mathbb{R}^k$ in Eq. (4.4) corresponds to a valid choice of the $k-1$ dimensional subspace spanned by $\{\boldsymbol{\alpha}_n\}_{n\geq 2}$. Therefore, since $\hat{h}_q$ depends on $\{\boldsymbol{\alpha}_n\}_{n\geq 2}$ only through β, the tightest bound on $\mathcal{M}$ is found by optimizing over arbitrary choices of $\boldsymbol{\beta}$ subject to the constraint $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1$.

To formulate the corresponding optimization problem, define the matrix G by

$$G_{im}(\boldsymbol{\theta}') = \frac{\partial f_i}{\partial \theta_m}(\boldsymbol{\theta}'). \tag{4.5}$$

We emphasize that G depends on the true value of the parameters $\boldsymbol{\theta}'$. Utilizing $||\frac{1}{2}\hat{\sigma}_i^z||_{\mathrm{s}} = 1$, we write the seminorm of $\hat{h}_q$ as

$$||\hat{h}_q||_{\mathrm{s}} = \sum_{i=1}^{d} |\nabla f_i(\boldsymbol{\theta}') \cdot \boldsymbol{\beta}| = \|G(\boldsymbol{\theta}')\boldsymbol{\beta}\|_1, \tag{4.6}$$

with $||\boldsymbol{x}||_1 = \sum_{i=1}^{d} |x_i|$ the L^1 or Manhattan norm. Therefore, for any $\boldsymbol{\beta}$ satisfying $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1$, we have

$$\mathcal{M} \geq \frac{1}{t^2 ||\hat{h}_q||_{\mathrm{s}}^2} = \frac{1}{t^2 \|G(\boldsymbol{\theta}')\boldsymbol{\beta}\|_1^2}. \tag{4.7}$$

In order to obtain the sharpest bound, we must solve what we refer to as the bound problem for $G(\boldsymbol{\theta}')$ and $\boldsymbol{\alpha}$:

Bound problem: *Given a non-zero vector* $\boldsymbol{\alpha} \in \mathbb{R}^k$ *and a real* $d \times k$ *matrix* G*, compute* $u = \max_{\boldsymbol{\beta}} \frac{1}{\|G\boldsymbol{\beta}\|_1}$ *under the condition* $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1$.

This is a linear programming problem and can in general be solved in time that is polynomial in d and k (see, e.g., Ref. [116]). Hereafter, we refer to the resulting sharpest bound as "the bound".

4.4 Optimal Protocol

We now turn to the problem of providing a protocol that saturates this bound. For clarity of presentation, we develop this protocol in the case that both the field $\boldsymbol{f}(\boldsymbol{\theta})$ and the objective $q(\boldsymbol{\theta})$ are linear in the parameters $\boldsymbol{\theta}$; that is, $\boldsymbol{f}(\boldsymbol{\theta}) = G\boldsymbol{\theta}$, with $\boldsymbol{\theta}$-independent G, and $q(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$. However, the existence of an asymptotically optimal protocol can be proven in the more general case that $\boldsymbol{f}(\boldsymbol{\theta})$ and $q(\boldsymbol{\theta})$ are analytic in the neighborhood of the true value $\boldsymbol{\theta}'$ [See Appendix C].

Similar to the approach taken in Ref. [102], this generalization ultimately amounts to using a two-step protocol. In the first step, one spends an asymptotically negligible time t_1 estimating the values of the parameters $\boldsymbol{\theta}$. Then one linearizes $\boldsymbol{f}(\boldsymbol{\theta})$ and $q(\boldsymbol{\theta})$ about this estimate $\tilde{\boldsymbol{\theta}}$ and spends the remaining time $t_2 = t - t_1$ estimating the resulting linearized objective. (Note,

asymptotically, $t_2 \sim t$.) Therefore, while we leave the rigorous analysis of this generalization to [Appendix C], the analytic case reduces to the linear case considered here, and therefore the principle insights are made most readily apparent in this context.

For the linear case, we propose an explicit protocol to measure q and show that it saturates the bound and thus is optimal. The optimal protocol measures the linear combination

$$\lambda(\boldsymbol{f}) = \boldsymbol{w} \cdot \boldsymbol{f}, \tag{4.8}$$

where $\boldsymbol{f}$ is the vector of local field amplitudes. The vector $\boldsymbol{w} \in \mathbb{R}^d$ is chosen such that $\tilde{\lambda}(\boldsymbol{f}) = \tilde{q}(\boldsymbol{\theta})$ is an unbiased estimator of $q(\boldsymbol{\theta}')$, and will be optimized to saturate the bound. (We note that, for $d > k$, there are many choices of $\boldsymbol{w}$ that satisfy $\lambda = q$.)

For the estimator $\tilde{\lambda}$ to be unbiased, we must have $\mathbb{E}\left[\tilde{q}\right] = q(\boldsymbol{\theta}') = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}'$. This is achieved by choosing $\boldsymbol{w}$ to satisfy the *consistency condition*

$$G^T \boldsymbol{w} = \boldsymbol{\alpha}. \tag{4.9}$$

Indeed, this implies

$$\mathbb{E}\left[\tilde{q}\right] = \mathbb{E}\left[\boldsymbol{w} \cdot \boldsymbol{f}\right] = \left(G\boldsymbol{\theta}'\right)^T \boldsymbol{w} = \boldsymbol{\theta}' \cdot \left(G^T \boldsymbol{w}\right) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}'. \tag{4.10}$$

We prove in [Appendix C] that, under our assumption that we can estimate $\boldsymbol{\theta}'$, Eq. (4.9) may always be satisfied for some $\boldsymbol{w}$, and therefore our protocol is valid.

For any such choice of $\boldsymbol{w}$, we use the optimal linear protocol of Ref. [8]—which for completeness, we summarize in [Appendix C]—to measure $\lambda(\boldsymbol{f})$. The variance obtained by this

protocol is

$$\operatorname{Var} \tilde{q} = \frac{\|\boldsymbol{w}\|_\infty^2}{t^2}, \tag{4.11}$$

where $||\boldsymbol{w}||_\infty = \max_i |w_i|$. Since we are dealing with an unbiased estimator, the MSE coincides with the variance of the estimator in Eq. (4.11). In order to find $\boldsymbol{w}$ with the lowest possible value of $\|\boldsymbol{w}\|_\infty$ (i.e. the smallest variance), we must solve what we refer to as the protocol problem:

Protocol problem: *Given a non-zero vector $\boldsymbol{\alpha} \in \mathbb{R}^k$ and a real $d \times k$ matrix G, compute $u' = \min_{\boldsymbol{w}} \|\boldsymbol{w}\|_\infty$ under the condition $G^T \boldsymbol{w} = \boldsymbol{\alpha}$.*

This, again, can be efficiently solved by generic linear programming algorithms [116, 117] or special-purpose algorithms [118–120].

To show that the optimal protocol from solving this problem saturates the bound, we now show that the bound problem and protocol problem are equivalent in that $u = u'$. For this, we utilize the strong duality theorem for linear programming [118, 121].[4] It states that, for linear programming problems like the protocol problem, there is a dual problem whose solution is identical to the original problem. In our case, we have the following dual problem:

Dual protocol problem: *Given a non-zero vector $\boldsymbol{\alpha} \in \mathbb{R}^k$ and a real $d \times k$ matrix G, compute $u'' = \max_{\boldsymbol{v}} \boldsymbol{\alpha} \cdot \boldsymbol{v}$ under the condition $\|G\boldsymbol{v}\|_1 \leq 1$.*

The strong duality theorem then implies $u'' = u'$. Additionally, there is a correspondence between the two solution vectors $\boldsymbol{w^0}$ and $\boldsymbol{v^0}$, so that, given the solution vector to one problem, we can find the solution vector to the other [118, 121]. We now prove the following theorem.

Theorem 4.4.1. *Let u and u' be the solutions to the bound and protocol problems, respectively.*

[4] See Ref. [104] for a quantum sensing use of this theorem in the context of evaluating the Holevo Cramér–Rao bound.

Then $u = u'$.

Proof. By the strong duality theorem, the solution of the dual protocol problem satisfies $u'' = \max_{\boldsymbol{v}} \boldsymbol{\alpha} \cdot \boldsymbol{v} = u'$. Let the corresponding solution vector of the dual protocol problem be $\boldsymbol{v^0}$. Define $\boldsymbol{\beta}^0 := \boldsymbol{v^0}/u'$. We have $\boldsymbol{\alpha} \cdot \boldsymbol{\beta^0} = u'/u' = 1$, thus $\boldsymbol{\beta^0}$ satisfies the constraint of the bound problem. To prove the theorem, we show that $u' \leq u$ and $u \leq u'$. On the one hand, provided $\|G\boldsymbol{\beta^0}\|_1 \neq 0$, the condition $\|G\boldsymbol{v^0}\|_1 \leq 1$ of the dual problem implies

$$u' \leq \frac{1}{\|G\boldsymbol{\beta^0}\|_1} \leq \max_{\boldsymbol{\beta}} \frac{1}{||G\boldsymbol{\beta}||_1} = u. \tag{4.12}$$

On the other hand, for *any* $\boldsymbol{\beta}$ satisfying the constraint $\boldsymbol{\alpha} \cdot \boldsymbol{\beta}$ of the bound problem, and for the optimal $\boldsymbol{w} = \boldsymbol{w^0}$ of the protocol problem satisfying $\|\boldsymbol{w^0}\|_\infty = u'$, Hölder's inequality yields

$$\begin{aligned} 1 = \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = (G^T \boldsymbol{w^0})^T \boldsymbol{\beta} = \boldsymbol{w^0} \cdot (G\boldsymbol{\beta}) \leq \|\boldsymbol{w^0}\|_\infty \|G\boldsymbol{\beta}\|_1 \\ \implies \frac{1}{\|G\boldsymbol{\beta}\|_1} \leq \|\boldsymbol{w^0}\|_\infty = u' \text{ for all } \boldsymbol{\beta}. \end{aligned} \tag{4.13}$$

This shows that $u' \geq 1/||G\boldsymbol{\beta}||_1$ for all $\boldsymbol{\beta}$, thus $u' \geq u$, which completes the proof. As a byproduct, we learn from Eq. (4.12) that $\boldsymbol{\beta^0}$ maximizes $1/||G\boldsymbol{\beta}||_1$, and so is the solution vector of the bound problem. □

Theorem 4.4.1 implies that the protocol measuring λ with optimal $\boldsymbol{w}$ saturates the bound.

As an instructive example, we return to the toy model presented in the introduction. Consider three sensors coupled to local field amplitudes $f_1(\boldsymbol{\theta}) = \theta_1$, $f_2(\boldsymbol{\theta}) = \theta_2$, and $f_3(\boldsymbol{\theta}) = \theta_1 + \theta_2$.

Our objective is $q(\boldsymbol{\theta}) = \theta_1$, so $\boldsymbol{\alpha} = (1, 0)^T$. We have

$$G^T = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \tag{4.14}$$

First consider the bound problem. The constraint $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1$ implies $\boldsymbol{\beta} = (1, b)^T$ with arbitrary b. The maximum of $1/\|G\boldsymbol{\beta}\|_1$ is achieved for $\boldsymbol{\beta^0} = (1, 0)^T$, yielding $u = 1/2$. For the protocol problem, the constraint in Eq. (4.9) gives $w_1 + w_2 = 1$ and $w_2 + w_3 = 0$. The corresponding minimal value of $\|\boldsymbol{w}\|_\infty$ is $u' = 1/2$ for $\boldsymbol{w^0} = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)^T$. Finally, for the dual protocol problem, the constraint $\|G\boldsymbol{v}\|_1 \leq 1$ implies $|v_1| + |v_2| + |v_1 + v_2| \leq 1$. The solution vector is $\boldsymbol{v^0} = (1/2, 0)^T$, which yields $u'' = \boldsymbol{\alpha} \cdot \boldsymbol{v^0} = 1/2$. This explicit example demonstrates that $u = u' = u''$. Furthermore, as noted in the proof of Theorem 4.4.1, $\boldsymbol{\beta^0} = \boldsymbol{v^0}/u'$.

4.5 Applications

Having derived optimal bounds and protocols saturating them, we now discuss some applications. We begin by considering the same example as above and show that, remarkably, our results in this case indicate that the best entangled and best unentangled weighting strategies need not be the same. With or without entanglement, we estimate $q(\boldsymbol{\theta}) = \theta_1$ by measuring a linear combination $\boldsymbol{w} \cdot \boldsymbol{f}$ with the constraints $w_1 + w_3 = 1$, $w_2 + w_3 = 0$. Without entanglement, our only option is to measure each component of $\boldsymbol{f}$ independently in parallel for time t, yielding a total MSE for $q(\boldsymbol{\theta})$ of $\|\boldsymbol{w}\|_2^2/t^2$. In stark contrast, for the entangled case, the MSE is given by $\|\boldsymbol{w}\|_\infty^2/t^2$. It is easy to see that minimizing the Euclidean and supremum norm of $\boldsymbol{w}$, subject to our constraints, does not yield the same result: Without entanglement, $\boldsymbol{w} = \left(\frac{2}{3}, -\frac{1}{3}, \frac{1}{3}\right)^T$ is

optimal, yielding an MSE of $\frac{2}{3t^2}$. With entanglement, $\boldsymbol{w} = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)^T$ is optimal, with MSE of $\frac{1}{4t^2}$. This simple example shows that, to achieve the optimal result with entanglement, one cannot in general use the weights $\boldsymbol{w}$ that are optimal without entanglement.

Our results are practically relevant for any situation where one knows the functional form of the field of interest $f(\boldsymbol{x}; \boldsymbol{\theta})$ and seeks to determine some quantity dependent on the parameters of that field. Examples include functionals of the form $q(\boldsymbol{\theta}) = \int_R \mathrm{d}\boldsymbol{x}\ k(\boldsymbol{x}) f(\boldsymbol{x}; \theta)$ with any kernel $k(\boldsymbol{x})$ and region of integration R. The examples from the introduction correspond to $k(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}_0)$ and $k(\boldsymbol{x}) = 1$. Since the $\boldsymbol{\theta}$-dependence of $f(\boldsymbol{x}, \boldsymbol{\theta})$ is analytic, this amounts to evaluating an analytic function $q(\boldsymbol{\theta})$.

As it is of clear physical relevance, we explicitly consider a simple, one-dimensional version of the former case with $k(x) = \delta(x - x_0)$, namely, field interpolation. Consider the situation of k particles at positions $x \in \{z_1, \cdots, z_k\}$ with unknown charges specified by the parameters $\{\theta_1, \cdots, \theta_k\}$ (and true values given by $\{\theta'_1, \cdots, \theta'_k\}$). Suppose we seek to determine the magnitude of (one component of) the electric field $q(\boldsymbol{\theta})$ at $x = x_0$ using $d \geq k$ sensors at positions $x \in \{s_1, \cdots s_d\}$. We then have

$$q(\boldsymbol{\theta}) = \frac{1}{4\pi\epsilon_0} \sum_{n=1}^{k} \frac{\theta_n}{(z_n - x_0)^2}, \tag{4.15}$$

which is linear in the unknown parameters $\{\theta_n\}$. Similarly, the fields measured by the sensors, given by $f_i(\boldsymbol{\theta}) = (1/4\pi\epsilon_0) \sum_{n=1}^{k} (\theta_n/(z_n - s_i)^2)$ are also linear in the parameters. Our protocol then applies quite simply to this situation with

$$G(\boldsymbol{\theta}')_{in} = \frac{1}{4\pi\epsilon_0 (z_n - s_i)^2} \tag{4.16}$$

and the elements of $\boldsymbol{\alpha}$ given by $\boldsymbol{\alpha}_n = 1/[4\pi\epsilon_0(z_n - x_0)^2]$. One can then straightforwardly solve the bound problem, the protocol problem, or the dual protocol problem given the particular locations of charges and sensors via analytic or numeric methods.

Our findings are also relevant for determining the optimal placement of sensors in space, i.e. determining the best locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d$ in the control space X in which they reside. For example, if the sensors are confined to a plane, then $X = \mathbb{R}^2$. This problem clearly consists of two parts: (1) evaluating the best possible MSE for any chosen set of sensor locations and (2) optimizing the result over possible locations. The MSE amounts to the cost function in usual optimization problems. Our results solve this first part as it would be used in the inner loop of a numerical optimization algorithm. The full problem, involving also the second part, is a high dimensional optimization in a space of dimension $d \times \dim(X)$. Therefore, in general, one expects that finding the global optimal placement could be quite challenging. However, even finding a local optimum in this space is clearly of practical use.

4.6 Outlook

While we assumed that we can obtain an individual estimate of the true value $\boldsymbol{\theta}'$ of the parameters, one could imagine situations where this assumption is not satisfied. Some such systems are underdetermined and not uniquely solvable, but in some cases we can reparametrize $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$ in order to satisfy the assumption. For example, if two parameters in the initial parametrization always appear as a product $\theta_1\theta_2$ in both f and q, we cannot individually estimate θ_1 or θ_2. However, we can reparametrize $\theta_1\theta_2 \to \theta_1^*$ and thus satisfy our initial assumption.

Our work applies to physical settings beyond qubit sensors—that is, any situation where

Eq. (4.3), may be applied our results should hold, provided we use the corresponding seminorm for the particular coupling. One example is using a collection of d Mach-Zehnder interferometers where the role of local fields is played by interferometer phases [101, 122–126]. Here the limiting resource is the number of photons N available to distribute among interferometers and not the total time t. We note, however, there are subtleties that we do not consider here when only the average number of photons is known [127]. The optimal variance for measuring a linear combination of local field values in this setting is conjectured in Ref. [8]. Under the assumption that this conjecture is correct, we may replace Eq. (4.11) with $\mathcal{M} = \frac{\|\boldsymbol{w}\|_\infty^2}{N^2}$[5] and otherwise everything remains the same as the qubit sensor case. One could also consider the entanglement-enhanced continuous-variable protocol of Ref. [80] for measuring linear combinations of field-quadrature displacements. A variation of this protocol has been experimentally implemented in Ref. [86]. We expect our bound and protocol could be extended to all the scenarios just described or even to the hybrid case where some local fields couple to qubits, some to Mach-Zehnder interferometers, and some to field quadratures. The ultimate attainable limit in such physical settings remains an open question, however.

One could consider the case $d < k$ provided the d sensors are not required to be at fixed locations. For instance, if one had access to continuously movable sensors in a 1D control space X, by the Riesz representation theorem [121], one could encode any linear functional of $f(x; \boldsymbol{\theta})$ by moving the sensors according to a particular corresponding velocity schedule. As a simple example, one can consider evaluating the integral of some function of (one component of) a magnetic field over one-dimensional physical space by moving a qubit sensor through the field

[5] *Comment:* This is a typo in the published version of this chapter. It should read $\|\boldsymbol{w}\|_1^2$. See Chapter 3 of this thesis for a proof (and generalization) of this conjecture.

and measuring the accumulated phase. One could also consider variations of this work in the context of semiparametric estimation [128]. We leave further exploration of such schemes to future work.

Finally, we emphasize that our protocol requires the use of highly entangled pure states (such as GHZ states) and does not consider the effects of decoherence or noise. Provided decoherence times are long, our results are applicable, but, beyond this limit, analyzing our protocols in such open systems (or designing different, more noise-robust protocols) remains an interesting and important question.

Chapter 5: Protocols for Estimating Multiple Functions with Quantum Sensor Networks: Geometry and Performance

5.1 Introduction

It is well-established that entanglement in quantum metrology often facilitates more accurate measurements compared to what is possible with unentangled probes [26, 28, 34, 95, 97]. This fact has been demonstrated exhaustively for the cases of measuring a single parameter [115] or a single analytic function of many parameters [8, 9, 51, 52, 54, 55, 59, 102] using quantum sensor networks, which are highly general models of quantum metrology. In these models, one considers an array of d quantum sensors, each coupled to a local parameter. One then seeks to optimally measure these local parameters directly (or some functions thereof) by selecting an initial state ρ_0 for the sensors, a unitary evolution U by which the local parameters are encoded in the state, and a choice of measurement specified by a positive operator-valued measure (POVM).

While measuring a single analytic function of multiple parameters in this setting is a bona fide multi-parameter problem, the fact that one seeks a single quantity makes the problem of finding the information-theoretic optimum for the variance of the desired quantity easier than a more general multi-parameter problem; in particular, one can make clever use of rigorous bounds originally derived for the single-parameter case [8, 54, 102]. However, when one genuinely seeks

to estimate multiple quantities, one must solve the general problem of designing provably optimal protocols for multi-parameter quantum estimation. This has proven to be a challenging problem, and has attracted a large amount of interest theoretically [45, 48, 52, 53, 63, 80, 92, 96, 97, 99–101, 104, 106, 109, 111, 112, 129–135] and experimentally [136–138]. Despite these extensive research efforts, the general problem has not yet been solved. Here, we consider another step towards this goal; in particular, we consider the case of measuring $n \leq d$ analytic functions with a quantum sensor network of d qubit sensors and develop a protocol that outperforms previously proposed protocols in many cases. We also emphasize the geometric aspects of this problem, meaning the orientations of vectors of coefficients associated with our functions, and how this geometry determines the protocol performance.

We begin by noting that, analogous to Ref. [102], one can reduce the problem of measuring n analytic functions of the parameters to that of measuring n linear functions. In particular, one can consider spending some asymptotically (in total time t) vanishing time t_1 measuring the local parameters to which the sensors are coupled and then the rest of the time $t_2 = t - t_1$ measuring the n linear combinations that result from a Taylor expansion of each analytic function about the true values of the local parameters estimated in the previous step. While provably optimal in the single-function case ($n = 1$), this reduction from analytic functions to linear functions is not necessarily optimal in the multi-function case. While we conjecture that the optimality of this reduction from analytic to linear functions *does* generalize to the multi-function case, as we do not claim general optimality of the protocols in this work, the reduction may be freely made without having to prove the veracity of this conjecture.

Having made this reduction to the problem of measuring multiple *linear* functions in a quantum sensor network, we can connect to previous works addressing the same problem, subject

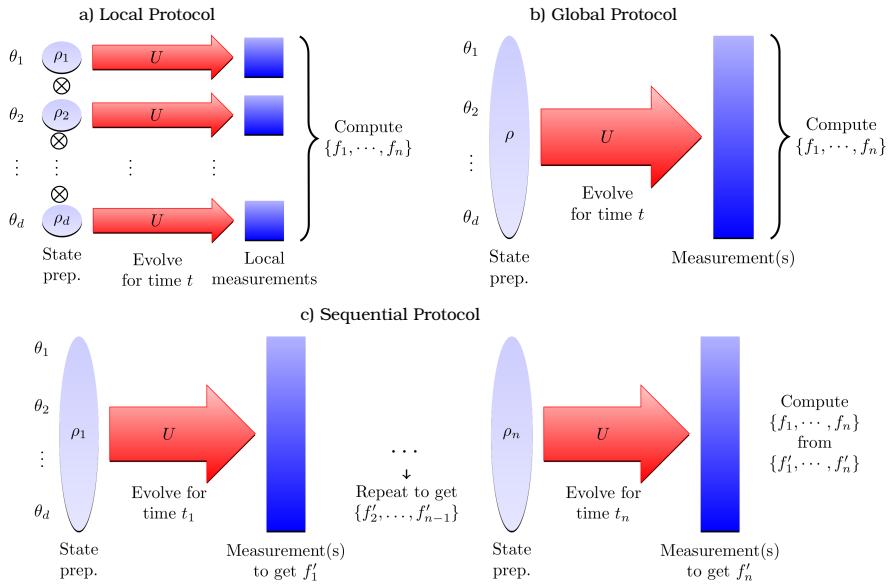


Figure 5.1: The protocols for measuring $n \leq d$ linear functions $\{f_1(\boldsymbol{\theta}), \ldots, f_n(\boldsymbol{\theta})\}$ of d parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ considered in this work can be classified into three groups: (a) Local protocols do not utilize entanglement and measure the parameters locally, allowing for large parallelization. (b) Global protocols simultaneously estimate all functions. (c) Sequential protocols divide the problem into n parts, where each part is optimized to estimate a single function from the set $\{f'_1, \ldots, f'_n\}$, which may consist of linear combinations of the original set $\{f_1, \ldots, f_n\}$.

to various simplifying constraints [9, 52, 53]. Leaving the details of these previous approaches for after we have introduced more mathematical formalism, we note that we may qualitatively divide protocols for this problem into three classes: local, global, and sequential [52]. In a local estimation protocol, one optimizes only over unentangled input states and local measurements of the sensors. In a global protocol, one simultaneously estimates all the desired functions by optimizing over all (possibly entangled) input states and all (possibly non-local) measurements. Finally, in a sequential protocol, we divide the experiment into n steps, where in each part we measure a single function (which may be a linear combination of the original set $\{f_1, \cdots, f_n\}$), preparing a new (optimal) initial state and performing a new measurement in each step. See Fig. 5.1 for diagrammatic representations of these different protocol types.

For the special case of measuring $n = d$ orthogonal, linear functions (that is, linear func-

tions such that the vectors of coefficients defining the linear functions are all mutually orthogonal), it has been known for some time that the functions can be measured optimally with a local protocol [9, 52], but for general functions, proofs of optimal protocols are lacking. In fact, the only entanglement-enhanced approach in the literature for measuring $n > 1$ general linear functions in a quantum sensor network is given in Ref. [53]. The bound on performance given there is for global protocols and is derived from the quantum Cramér-Rao bound [45–48] subject to the restriction that one considers only a special set of so-called sensor symmetric states. However, even within this restriction, beyond the case of $d = 2$, it is an open question whether the states and measurements (POVMs) required to saturate the derived bound exist for all problems.[1]

Here, we highlight a generalization of this approach, by deriving similar bounds using so-called *signed sensor symmetric states*. However, the generalized version also does not guarantee that the optimal states and measurements exist in general. Targeted at this shortcoming, we also consider an alternative, sequential protocol, subject to different restrictions, for which we can explicitly describe a protocol which achieves its theoretical performance. In addition to presenting this alternative protocol, we lay out how the precise geometric features of a given problem impact the performance of this sequential protocol compared to the signed sensor symmetric approach and the simple local protocol.

[1]The reason that there may not exist states satisfying the bound is that, as explained later in the [chapter], the bound is obtained by fixing v [defined in Eq. (5.16)] to be $t^2/4$ and then optimizing $\mathcal{J}$ [defined in Eq. (5.15)] given this restriction to obtain the best bound. In principle, we are not guaranteed a state corresponding to this pair of $4v = t^2$ and the minimizing $\mathcal{J}$, but of course the bound is still a correct lower bound if one is only allowed to use sensor symmetric states whether or not it can be saturated. Also, see Ref. [53] for further discussion.

5.2 Problem Setup

With the general approach established, we now present the rigorous formulation of the problem. We consider a quantum sensor network of d qubit sensors prepared in some initial state ρ_0. We then encode d local parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_d)^T$ into the sensors via unitary evolution under the Hamiltonian

$$\hat{H} = \hat{H}_{\mathrm{c}}(t) + \sum_{i=1}^{d} \frac{1}{2}\theta_i \hat{\sigma}_i^z, \tag{5.1}$$

with $\hat{\sigma}_i^{x,y,z}$ the Pauli operators acting on the i^{th} qubit, and θ_i the local parameter measured by the i^{th} sensor. The term $\hat{H}_{\mathrm{c}}(t)$ is a time-dependent control Hamiltonian that may include coupling to ancilla qubits. When measuring a single function, this time-dependent control is not necessary to achieve an optimal protocol [8, 115], and therefore, may freely be set to zero; however, one may use such control to design optimal protocols with simpler requirements on the choice of input state ρ_0 [8]. Using this setup, our goal is to optimally measure $n \leq d$ functions $\boldsymbol{f}(\boldsymbol{\theta}) = (f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \cdots, f_n(\boldsymbol{\theta}))^T$. In the following, we use $i, j = 1, \ldots, d$ to label qubits and $\ell, m = 1, \ldots, n$ to label functions. Boldface is used to denote vectors.

To compare the accuracy of the different approaches and to eventually optimize them, we employ a standard figure of merit, which we denote as $\mathcal{M}$, given as

$$\mathcal{M} = \sum_{\ell=1}^{n} w_\ell \mathrm{Var}\, \tilde{f}_\ell, \tag{5.2}$$

where $\tilde{\boldsymbol{f}}$ are estimators of the functions and $\boldsymbol{w} = (w_1, \cdots, w_n)^T$ is a vector of weights. Since an

accurate protocol should yield small variances, we seek to minimize $\mathcal{M}$. In this context, given a total evolution time t, a protocol is defined by choice of initial state ρ_0, control Hamiltonian $\hat{H}_c(t)$, measurements, and estimator $\tilde{\boldsymbol{f}}$ for $\boldsymbol{f}$.

The figure of merit $\mathcal{M}$ is lower bounded via the Helstrom quantum Cramér–Rao bound [45–48], which yields

$$\mathcal{M} \geq \frac{1}{N} \sum_{\ell=1}^{n} w_\ell [\mathcal{F}_Q^{-1}(\boldsymbol{f})]_{\ell\ell}, \tag{5.3}$$

where N is the number of trials (which from now on we set to one for concision and consider just the single-shot Fisher information) and $\mathcal{F}_Q(\boldsymbol{f})$ is the quantum Fisher information matrix with respect to the functions $\boldsymbol{f}$. While this bound is not generally saturable, in the setting of Eq. (5.1) it is.[2]

While saturable in the setting considered, the right hand side of Eq. (5.3) is not easily evaluated in general. However, it has been proven [8] that, if we seek to measure a *single* linear function $f(\boldsymbol{\theta}) = \boldsymbol{\alpha} \cdot \boldsymbol{\theta}$ of the parameters $\boldsymbol{\theta}$, we may evaluate this bound and obtain that the minimum (asymptotically in time t and number of trials) attainable variance of an estimator $\tilde{f}$ of $f(\boldsymbol{\theta})$ over all quantum protocols is

$$\operatorname{Var} \tilde{f} = \max_i \frac{|\alpha_i|^2}{t^2}. \tag{5.4}$$

This bound can be explicitly saturated by the protocols given in Ref. [8]. As previously described, if $f(\boldsymbol{\theta})$ is a more general analytic function, one may attain a similar bound using a two-step protocol. In the first (asymptotically negligible) step, one makes local estimates $\tilde{\boldsymbol{\theta}}$ of each of the

[2] In particular, it is saturable because the generators of translations K_i, as defined in the discussion around Eq. (5.15), commute.

parameters $\boldsymbol{\theta}$. In the second step, one uses the rest of the time to optimally measure the Taylor expansion of $f(\boldsymbol{\theta})$ about this estimate to linear order in $\boldsymbol{\theta}$ [102].

For the case of measuring multiple functions $f_1, \ldots, f_n$, we assume without loss of generality that the f_ℓ are *linear functions* in the parameters $\boldsymbol{\theta}$, because more general analytic functions could be similarly linearized in asymptotically negligible time. We parameterize the linear functions by real *coefficient vectors* $\boldsymbol{\alpha}_\ell$ such that

$$f_1(\boldsymbol{\theta}) = \boldsymbol{\alpha}_1 \cdot \boldsymbol{\theta}, \tag{5.5}$$

$$\vdots$$

$$f_n(\boldsymbol{\theta}) = \boldsymbol{\alpha}_n \cdot \boldsymbol{\theta}. \tag{5.6}$$

Defining the matrix elements $A_{\ell i} = (\partial f_\ell / \partial \theta_i)_{\tilde{\boldsymbol{\theta}}} = (\boldsymbol{\alpha}_\ell)_i$, i.e., $\boldsymbol{\alpha}_\ell^T$ is the ℓth row of A, we can phrase the problem as that of optimally measuring the n-component vector

$$A\boldsymbol{\theta} = (\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_n)^T \boldsymbol{\theta}. \tag{5.7}$$

Without loss of generality we assume normalization of the coefficient vectors,

$$||\boldsymbol{\alpha}_\ell||^2 = 1 \text{ for all } \ell, \tag{5.8}$$

because any non-unit length can be absorbed into the weights $\boldsymbol{w}$ in Eq. (5.2).

Recall, the problem of measuring $n = d$ linear functions of independent parameters with quantum sensor networks has been considered in the literature in the case where the n functions

are orthogonal (in which case local, global and sequential protocols are equivalent) [9, 52] and for general linear functions for global protocols when the input states ρ_0 are restricted to be sensor symmetric [53]. Here, we generalize the sensor symmetric approach and derive a performance bound when using so-called signed sensor symmetric input states (defined rigorously below). We refer to the variance obtained by the signed sensor symmetric protocol as $\mathcal{M}_{\text{ss}}$.

In this work, we also introduce an optimized sequential protocol for solving the n function estimation problem. We consider dividing our protocol into n sequential steps where, *within each step*, the protocol is provably information-theoretic optimal (i.e., saturates the quantum Cramér–Rao bound). In particular, for each step $\ell \in \{1, \ldots, n\}$ taking time t_ℓ, we measure a single function optimally using the protocols from Refs. [8, 102]. We cannot, however, prove that the full protocol is optimal in an information-theoretic sense. The naive version of this protocol is to measure the n given functions $\{f_1, \ldots, f_n\}$ one after another with some optimal choice of the time t_ℓ spent on each function. We denote the figure of merit of the naive sequential protocol by $\mathcal{M}_{\text{naive}}$.

However, the naive sequential protocol is not the only option for sequentially measuring multiple functions. Indeed, the coefficient vectors $\{\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n\}$ span a linear subspace of $\mathbb{R}^d$, and we may instead sequentially measure *any* set of linear functions whose vectors of coefficients $\{\boldsymbol{\alpha}'_1, \cdots, \boldsymbol{\alpha}'_n\}$ span the same subspace and then (after the measurements) calculate the original functions $\{f_1, \ldots, f_n\}$. To help understand this visually, this approach is depicted in the diagram in Fig. 5.2 for $n = 2$ functions and $d = 3$ sensors. We denote the figure of merit obtained via this method by $\mathcal{M}_{\text{opt}}$.

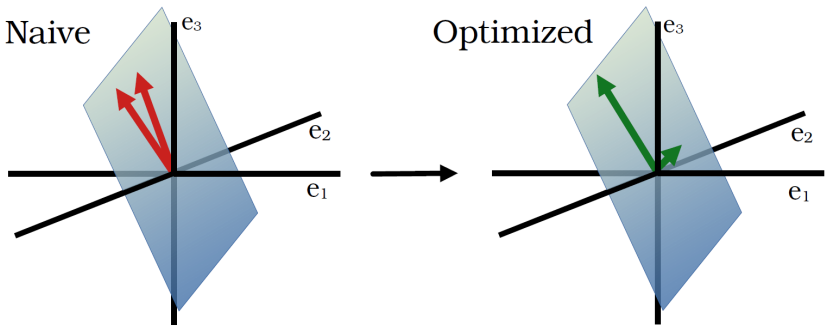


Figure 5.2: A visualization for $n = 2$ functions and $d = 3$ sensors of how we can optimally select a set of functions to measure whose coefficient vectors $\{\boldsymbol{\alpha}'_\ell\}$ span the same subspace as the coefficient vectors $\{\boldsymbol{\alpha}_\ell\}$ of the functions we care about. The vectors are the coefficient vectors and the planes indicate the subspace they span. The axes are labeled by standard basis unit vectors $\{e_1, e_2, e_3\}$.

To be explicit, define the $n \times n$ matrix C encoding the change of linear functions via

$$A = CA', \tag{5.9}$$

where $A' = (\boldsymbol{\alpha}'_1, \cdots, \boldsymbol{\alpha}'_n)^T$ is the matrix whose rows are the coefficient vectors of the new linear functions we measure. The variance of measuring any individual $\boldsymbol{\alpha}'_\ell$ is given by the optimal linear protocol [8]

$$\mathcal{M}_\ell = \frac{{\mu'_\ell}^2}{t_\ell^2}, \tag{5.10}$$

where we introduce

$$\mu'_\ell = \|\boldsymbol{\alpha}'_\ell\|_\infty = \max_j \ |\alpha'_{\ell,j}| = \max_j |\sum_{m=1}^{n} (C^{-1})_{\ell m} A_{mj}|. \tag{5.11}$$

Note that this corresponds to Eq. (5.4) for every ℓ. We denote by $\boldsymbol{\mu}'$ the vector with entries μ'_ℓ, and by $\boldsymbol{\mu}$ the analogous vector for the original functions [obtained by setting $C = I$ in Eq. (5.11)]. The figure of merit for estimating the original functions $\boldsymbol{f}$ with the optimized sequential protocol

is then formally given by

$$\mathcal{M}_{\text{opt}} = \min_{C} \min_{\{t_1,\cdots,t_n\}} \left[\sum_{\ell=1}^{n} \sum_{m=1}^{n} w_m C_{m\ell}^2 \left(\frac{\mu'_\ell}{t_\ell} \right)^2 \right], \tag{5.12}$$

which takes into account optimization over C and over the division of the total time into time steps t_ℓ; the factor $C_{m\ell}^2$ comes from the standard expression for a linear combination of variances and accounts for the linear change of functions. A more practical form of $\mathcal{M}_{\text{opt}}$ will be derived below. If the naive sequential protocol were optimal, then the minimum of $\mathcal{M}_{\text{opt}}$ would be attained at $C = I$. However, we will show in the following that choosing suitable $C \neq I$ often gives a significant improvement. This matches one's intuitive expectations — for example, if the coefficient vectors of all the functions are nearly aligned, we might expect that the optimal approach is to spend most of the time measuring a single function whose coefficient vector is in that general direction, and the rest of the time measuring functions with orthogonal coefficient vectors to distinguish the small differences in the functions we care about. We will see that this intuition is correct.

Furthermore, we note that for this approach, we do not consider taking advantage of potential parallelization that may arise for certain choices of functions to measure—in particular, those sets of functions that depend on completely disjoint sets of sensors. More formally, when one chooses functions to measure such that A' is the direct sum of matrices representing linear functions on disjoint sets of qubits, one could simultaneously measure functions that depend on disjoint sets of sensors, and thus spend more time measuring them, improving the accuracy. Therefore, purposefully choosing functions to measure that allow for such parallelization could potentially (although not necessarily) perform better than our protocol, which does not take this

possibility into account. However, improved performance via parallelization is not guaranteed as Eq. (5.12) depends on both the time t_ℓ spent measuring a function *and* the infinity-norm of the coefficient vector, $\mu'_\ell = \|\boldsymbol{\alpha}'_\ell\|_\infty$—whereas parallelization improves the former, it may worsen that latter.

We note that, when $n = d$, the local strategy is a special case of such parallelization as it consists of measuring the local parameters all in parallel, and therefore a completely diagonal A'. As another simple example, suppose $\boldsymbol{\alpha}_1 = (1,1,1)^T/\sqrt{3}$, $\boldsymbol{\alpha}_2 = (1,-1,1)^T/\sqrt{3}$, and $\boldsymbol{\alpha}_3 = (0,0,1)^T$. One way (amongst several) that this could be parallelized would be choosing to measure $\boldsymbol{\alpha}'_1 = (1,1,0)^T/\sqrt{2}$, $\boldsymbol{\alpha}'_2 = (1,-1,0)^T/\sqrt{2}$, and $\boldsymbol{\alpha}'_3 = (0,0,1)^T$; with this choice, one could, in parallel, estimate the sets of functions $\{\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2\}$ and $\{\boldsymbol{\alpha}'_3\}$.

At this point, we have commented on four approaches to our problem: (1) the local strategy with variance $\mathcal{M}_{\text{local}}$ (defined in Eq. (5.13) below), (2) the (global) signed sensor symmetric strategy generalized from Ref. [53] with variance $\mathcal{M}_{\text{ss}}$, (3) the naive sequential strategy with variance $\mathcal{M}_{\text{naive}}$, (4) the optimized sequential strategy with variance $\mathcal{M}_{\text{opt}}$. Importantly, *none* of these strategies is optimal in general. Depending on the geometry of the linear functions to be measured, each of these strategies could be the preferable one (excluding the naive strategy, which, of course, in the best case, has $\mathcal{M}_{\text{naive}} = \mathcal{M}_{\text{opt}}$). The term "geometry" here refers to the absolute and relative orientations of the coefficient vectors $\{\boldsymbol{\alpha}_\ell\}$. The question of what is the ultimate information-theoretic limit on $\mathcal{M}$ for multiple linear functions remains open. Here, we demonstrate cases in which each of these known strategies is preferable with an emphasis on the geometric interpretation. We emphasize that, in many instances, both the signed sensor symmetric and the optimized sequential strategy can out-perfom the local unentangled strategy, which is of great importance for practical applications.

5.3 The Strategies

In this section, we determine the figure of merit $\mathcal{M}$ for the four strategies considered in this work. We emphasize that while the local and sequential strategies have explicit protocols to obtain the corresponding figure of merit, the figure of merit for the signed sensor symmetric is not proven to be always be attainable beyond $d = 2$.

5.3.1 Local Strategy

First we consider the local strategy, which does not utilize entanglement. Since we can measure each local parameter θ_i simultaneously, with a variance of $1/t^2$ [98], we arrive at

$$\mathcal{M}_{\text{local}} = \sum_{\ell=1}^{n} w_\ell \frac{||\boldsymbol{\alpha}_\ell||^2}{t^2} = \frac{1}{t^2} \sum_{\ell=1}^{n} w_\ell = \frac{\mathcal{N}}{t^2}, \tag{5.13}$$

where we used the normalization of the $\boldsymbol{\alpha}_\ell$ and introduce

$$\mathcal{N} = \sum_{\ell=1}^{n} w_\ell. \tag{5.14}$$

We emphasize that the local protocol performs independently of the geometry of the measured linear functions.

5.3.2 Signed Sensor Symmetric Strategy

Next we review the results of Ref. [53] for the sensor symmetric approach, using our notation and emphasize a generalization of their approach to what we call signed sensor symmetric

states. We emphasize that, given the restriction to (signed) sensor symmetric states, this approach gives a rigorous lower bound on the figure of merit $\mathcal{M}$. However, as previously discussed, unlike the local or sequential strategies, for $d > 2$ one cannot guarantee that the figure of merit $\mathcal{M}_{\text{ss}}$ obtained via this approach is saturable [53].

Define the generators of translations in parameter space as $\boldsymbol{K} = (K_1, \cdots, K_d)^T$, where $K_i = \mathrm{i}(\partial U/\partial\theta_i)U^\dagger$ for evolution under the unitary U. Following Ref. [53], for this strategy, we specifically consider the Hamiltonian in Eq. (5.1) with $\hat{H}_c(t) = 0$, so that $U = \exp(-i\hat{H}t)$ and $K_i = \hat{\sigma}_i^z t/2$. This restriction of Eq. (5.1) to evolution under a time-independent Hamiltonian is not necessary for the sequential protocols considered later. However, the single linear function results from Ref. [8], which we use as a subroutine of our sequential protocol, presents two protocols, one that matches this restriction and one that does not (see section IV therein). Therefore, when explicitly comparing the sequential protocol to the signed sensor symmetric problem, we assume we are considering the former.

Given the generators of translations K_i, we define the inter-sensor correlations [9, 130] by

$$\mathcal{J}_{ij} = \frac{\langle K_i K_j\rangle - \langle K_i\rangle\langle K_j\rangle}{\Delta K_i \Delta K_j} \tag{5.15}$$

for $i \neq j$, where we have used $(\Delta K_i)^2 = \langle K_i^2\rangle - \langle K_i\rangle^2$. Given this definition, we define *sensor symmetric states* as those such that for all $i \neq j$, $\mathcal{J}_{ij} = \mathcal{J} = c/v$ with

$$v = \langle K_i^2\rangle - \langle K_i\rangle^2, \qquad c = \langle K_i K_j\rangle - \langle K_i\rangle\langle K_j\rangle. \tag{5.16}$$

Specifically, for evolution under the time-independent version of Eq. (5.1). we have

$$v = \frac{t^2}{4}\left(1 - \langle\sigma_i^z\rangle^2\right) \qquad\qquad c = \frac{t^2}{4}\left(\langle\sigma_i^z\sigma_j^z\rangle - \langle\sigma_i^z\rangle\langle\sigma_j^z\rangle\right) \tag{5.17}$$

for all $i \neq j$. The authors of Ref. [53] define such states in analogy with path-independent states in optical interferometry [130, 139], which, in addition to the analytic accessibility provided by such states, motivates this construction. The case of uncorrelated sensors, of course, is included for $\mathcal{J} = 0$.

Now we turn to a generalization of the sensor symmetric states considered in Ref. [53] that we call signed sensor symmetric states. This generalization is natural as the (unsigned) sensor symmetric state construction of Ref. [53] picks out functions with coefficient vectors $\boldsymbol{\alpha}$ aligned along the vector of all ones $\mathbf{1} = (1, 1, \cdots, 1)^T$ as being favorable, but we know the positive orthant is not special, and one can immediately generalize from $\mathbf{1}$ being the favorable orientation to any $\boldsymbol{\omega} \in \{-1, 1\}^d$ (of which $\mathbf{1}$ is just one example). The reason such functions are most favorable is also intuitively clear—entanglement is most helpful when one measures global, average-like quantities, which is precisely what functions with coefficient vectors aligned along some $\boldsymbol{\omega}$ are. We emphasize this generalization is very direct, as one can consider mapping any problem using a general $\boldsymbol{\omega}$ to the case of Ref. [53] merely by applying a Pauli-X operator on all qubit sensors corresponding to negative elements of $\boldsymbol{\omega}$ and correspondingly flipping the signs of all corresponding coefficients specified by $\boldsymbol{\alpha}_\ell$. However, to fairly compare to the sequential protocol, it is important we consider all such $\boldsymbol{\omega}$, as different choices can lead to an improved figure of merit. Therefore, we relax the restriction on the numerator of $\mathcal{J}_{ij}$ as presented in Ref. [53] by

defining

$$c_{ij} = \langle K_i K_j \rangle - \langle K_i \rangle \langle K_j \rangle \tag{5.18}$$

and then restrict our consideration to states such that

$$c_{ij} = c \left(\boldsymbol{\omega} \boldsymbol{\omega}^T \right)_{ij} = c \, \Omega_{ij}, \tag{5.19}$$

where $\boldsymbol{\omega} \in \{-1, 1\}^d$ is a vector with all entries ± 1 and c is a constant. The entries of Ω_{ij} are also ± 1 and so $c_{ij} = \pm c$. We keep the definition $\mathcal{J} = c/v$ for our newly defined c, but note that now $\mathcal{J}_{ij} = c_{ij}/v = \pm \mathcal{J}$.

When restricted to the (unsigned) sensor symmetric initial states, i.e. when $\boldsymbol{\omega} = \mathbf{1}$ with $\mathbf{1} = (1, \ldots, 1)^T$ the vector of all ones, the authors of Ref. [53] were able to evaluate the quantum Cramér–Rao bound and determine the minimal achievable value for $\mathcal{M}$, given the requirement of sensor symmetric input states. For the signed sensor symmetric states, the calculation is similar to that in Ref. [53], so we just state the result for our generalized approach here and present the details in Appendix D.1.

First define the $\boldsymbol{\omega}$-dependent geometry parameter $\mathcal{G}(\boldsymbol{\omega})$, which encodes the geometric relationship between the coefficient vectors $\{\boldsymbol{\alpha}_\ell\}$ of the n linear functions and the vector $\boldsymbol{\omega}$. We have

$$\mathcal{G}(\boldsymbol{\omega}) = \frac{1}{\mathcal{N}} \sum_{\ell=1}^{n} w_\ell \left(d \cos^2 \phi_{\boldsymbol{\omega},\ell} - 1 \right). \tag{5.20}$$

Here $\phi_{\boldsymbol{\omega},\ell}$ is the angle between the vectors $\boldsymbol{\alpha}_\ell$ and $\boldsymbol{\omega}$. Thus $\cos \phi_{\boldsymbol{\omega},\ell} = \boldsymbol{\alpha}_\ell \cdot \boldsymbol{\omega} / \sqrt{d}$. Note that $\mathcal{G}(\boldsymbol{\omega}) \in [-1, d-1]$. Again, we note that the relevance of this geometric quantity is intuitively

clear as entanglement provides the biggest benefit when measuring functions aligned along some $\boldsymbol{\omega}$—that is, those functions for whom $\phi_{\boldsymbol{\omega},\ell} \approx 0$. The $\boldsymbol{\omega}$-dependent lower bound on the figure of merit is found to be

$$\mathcal{M}_{\rm ss}(\boldsymbol{\omega}) = \min_{\mathcal{J}} \frac{\mathcal{N}}{t^2} \frac{1 + [d - 2 - \mathcal{G}(\boldsymbol{\omega})]\mathcal{J}}{(1-\mathcal{J})[1 + (d-1)\mathcal{J}]}, \tag{5.21}$$

where we have used $4v = t^2$ as in Ref. [53] to obtain the lowest bound. Under this condition on v, and the assumption that $\mathcal{J} \in (1/(1-d), 1)$, so that the quantum Fisher information is invertible, the minimum is attained for

$$\mathcal{J}_{\rm opt}(\boldsymbol{\omega}) = \frac{1}{\mathcal{G}(\boldsymbol{\omega}) + 2 - d} \left[1 - \sqrt{\frac{(\mathcal{G}(\boldsymbol{\omega}) + 1)[d - 1 - \mathcal{G}(\boldsymbol{\omega})]}{d-1}} \right]. \tag{5.22}$$

One can then obtain the theoretical best performance for a signed sensor symmetric strategy as

$$\mathcal{M}_{\rm ss} = \min_{\boldsymbol{\omega}} \mathcal{M}_{\rm ss}(\boldsymbol{\omega}). \tag{5.23}$$

Importantly, the obtainable accuracy is intimately related to the geometry of the linear functions we seek to measure. In particular, one finds the best performance for this strategy when $\mathcal{G}$ is approximately $d-1$; that is, when $\phi_{\boldsymbol{\omega},\ell} \approx 0$. This corresponds to the situations where the sensor symmetric states have the largest inter-sensor correlations $\mathcal{J}_{\rm opt}$ (i.e. are most entangled). We emphasize again, that there is no guarantee that this performance is always achievable, although in Ref. [53] it was proven for $d = 2$ and demonstrated for a large set of problems for $d > 2$.

5.3.3 Naive Sequential Strategy

In the naive sequential protocol, we sequentially measure the n linear functions $\{f_1, \ldots, f_n\}$ using an optimal single linear function protocol [8]. For this, we determine the optimal times t_ℓ spent to measure the ℓ^{th} function by minimizing Eq. (5.12) for $C = I$ with respect to $\{t_1, \cdots, t_n\}$ under the constraint $\sum_\ell t_\ell = t$. The solution to this Lagrange multiplier problem, presented in Appendix D.2, reads

$$\mathcal{M}_{\text{naive}} = \frac{1}{t^2}\left(\sum_{\ell=1}^{n}[w_\ell \mu_\ell^2]^{1/3}\right)^3. \tag{5.24}$$

As an important example, consider equal weights, $w_\ell \equiv \mathcal{N}/n$. Then we have

$$\frac{n^2\mathcal{N}}{dt^2} \leq \mathcal{M}_{\text{naive}} \leq \frac{n^2\mathcal{N}}{t^2}. \tag{5.25}$$

Indeed, the upper bound is obtained for unfavourable functions $\{f_\ell\}$ such that $\boldsymbol{\mu} = \mathbf{1}_n$ ("worst case"), with $\mathbf{1}_n$ the n-component vector of ones, whereas the lower bound is obtained for favourable functions $\{f_\ell\}$ with $\boldsymbol{\mu} = \mathbf{1}_n/\sqrt{d}$ ("best case"). These are the two extreme possible cases. Compared to the local protocol figure of merit of $\mathcal{N}/t^2$ for any choice of w_ℓ, we see that in the worst case, the local protocol is always superior to the naive sequential protocol. Furthermore, even in the best case, we must have $d > n^2$ to obtain an advantage from the naive sequential protocol compared to the local protocol, implying a relatively large number of sensors. This shows that the naive sequential protocol, with $C = I$, is not very competitive. On the other hand, as we show now, by optimizing over C a significant gain in accuracy over the local protocol can be achieved.

5.3.4 Optimal Sequential Strategy

Finally, we consider the optimal sequential protocol. The minimization over time in Eq. (5.26) proceeds as in the naive case but with a general C. Therefore, again leaving details to Appendix D.2, we obtain for the optimal sequential protocol that

$$\mathcal{M}_{\text{opt}} = \min_{C} \frac{1}{t^2} \left[\sum_{\ell=1}^{n} \left(\sum_{m=1}^{n} w_m C_{m\ell}^2 \right)^{\frac{1}{3}} {\mu'_\ell}^{2/3} \right]^3 , \tag{5.26}$$

with optimal time to measure the ℓth function given by

$$t_\ell = t \frac{\left(\sum_{m=1}^{n} w_m C_{m\ell}^2\right)^{1/3} {\mu'_\ell}^{2/3}}{\sum_{p=1}^{n} \left(\sum_{m=1}^{n} w_m C_{mp}^2\right)^{1/3} {\mu'_p}^{2/3}}. \tag{5.27}$$

Inserting the definition of μ'_ℓ from Eq. (5.11), we arrive at

$$\mathcal{M}_{\text{opt}} = \min_{C} \frac{1}{t^2} \left[\sum_{\ell=1}^{n} \left(\sum_{m=1}^{n} w_m C_{m\ell}^2 \right)^{\frac{1}{3}} \max_i \left| \sum_{m=1}^{n} (C^{-1})_{\ell m} A_{mi} \right|^{2/3} \right]^3 . \tag{5.28}$$

Note that due to the appearance of both C and C^{-1} in the expression with the same powers, the result is invariant under a change in the normalization of the columns of C. Therefore we may fix these column normalizations and introduce the constraint that

$$\sum_{m=1}^{n} w_m C_{m\ell}^2 = 1, \tag{5.29}$$

for each ℓ. Under this constraint, we obtain the simpler expression

$$\mathcal{M}_{\text{opt}} = \min_{C} \frac{1}{t^2} \left[\sum_{\ell=1}^{n} \max_{i} |\sum_{m=1}^{n} (C^{-1})_{\ell m} A_{mi}|^{2/3} \right]^3 , \tag{5.30}$$

with optimal time per function given by

$$t_\ell = t \frac{{\mu'_\ell}^{2/3}}{\sum_{m=1}^{n} {\mu'_m}^{2/3}}. \tag{5.31}$$

Geometrically, the constraint in Eq. (5.29) corresponds to restricting the columns of C to the surface of an $(n-1)$-dimensional ellipsoid (or $(n-1)$-sphere if $w_m = \mathcal{N}/n \; \forall \, m$). The columns of C can then be efficiently parametrized by elliptical (or spherical) coordinates, and the optimization amounts to finding the best choice of corresponding angular variables. We emphasize that this choice of normalization can be made without loss of generality.

We have now fully characterized our optimized sequential protocol. In particular, one can numerically perform the minimization over matrices C in Eq. (5.28) subject to the constraint in Eq. (5.29). However, while for practical purposes we have solved the problem, many questions of more general nature arise at this point. In particular, what kind of advantage is provided by the optimized sequential protocol over the naive one? What geometries of coefficient vectors correspond to the best performance for the sequential protocol? How does it compare to the signed sensor symmetric approach? These questions will be addressed in the following section. All of the figures of merit calculated in this section are summarized in Table 5.1.

	Local	Naive Sequential	Signed Sensor Symmetric	Optimized Sequential
$\mathcal{M}$	$\frac{\mathcal{N}}{t^2}$	$\frac{1}{t^2}\left(\sum_{\ell=1}^{n}[w_\ell \mu_\ell^2]^{1/3}\right)^3$	$\min_{\boldsymbol{\omega}} \frac{\mathcal{N}}{t^2}\frac{1+[d-2-\mathcal{G}(\boldsymbol{\omega})]\mathcal{J}_{\text{opt}}}{(1-\mathcal{J}_{\text{opt}})[1+(d-1)\mathcal{J}_{\text{opt}}]}$ $\mathcal{J}_{\text{opt}}(\boldsymbol{\omega}) = \frac{1}{\mathcal{G}(\boldsymbol{\omega})+2-d}\left[1-\sqrt{\frac{(\mathcal{G}(\boldsymbol{\omega})+1)[d-1-\mathcal{G}(\boldsymbol{\omega})]}{d-1}}\right]$ $\mathcal{G}(\boldsymbol{\omega}) = \frac{1}{\mathcal{N}}\sum_{\ell=1}^{n} w_\ell\,(d\cos^2\phi_{\boldsymbol{\omega},\ell}-1)$	$\min_{C} \frac{1}{t^2}\left[\sum_{\ell=1}^{n}\max_i\left\|\sum_{m=1}^{n}(C^{-1})_{\ell m}A_{mi}\right|^{2/3}\right]^3$ subject to $\sum_{m=1}^{n} w_m C_{m\ell}^2 = 1$

Table 5.1: Summary of figures of merit. Recall, that for all strategies other than signed sensor symmetric strategy, we have an explicit physical protocol to achieve the given figure of merit. For the signed sensor symmetric strategy, beyond $d = 2$, we are not necessarily guaranteed that a state exists that achieves the figure of merit, and therefore it is a lower bound, given the signed sensor symmetric state restriction.

5.4 Performance and Geometry

To compare the performance of the different strategies, we first study some analytically accessible limits and then turn to a numerical analysis of the related optimization problem.

5.4.1 Geometrically Symmetric Limit

We begin by considering what we refer to as the geometrically symmetric limit of the signed sensor symmetric strategy. This limit will be useful for comparing to the optimized sequential protocol in the following subsections. For this, we consider a situation where the coefficient vectors $\boldsymbol{\alpha}_\ell$ are all approximately the same angle ϕ' from some $\boldsymbol{\omega}$, which we recall is a vector with all elements ± 1. This results in a particularly useful simplification of the expression for the geometry parameter $\mathcal{G}$. We then define the parameter

$$\epsilon_{\boldsymbol{\omega},\ell} = \phi_{\boldsymbol{\omega},\ell} - \phi', \tag{5.32}$$

so that $\epsilon_{\boldsymbol{\omega},\ell}$ may be treated as a small parameter for a perturbative expansion, see Fig. 5.3

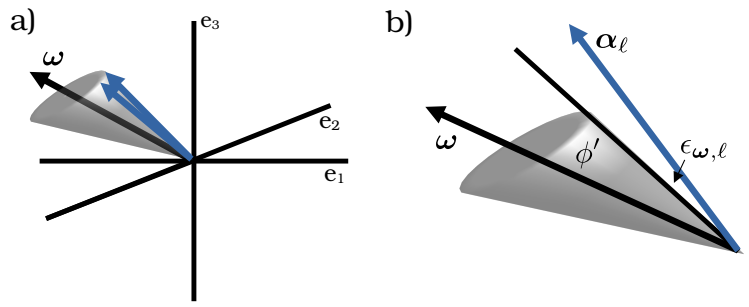


Figure 5.3: (a) A visualization for $n = 2$ functions and $d = 3$ sensors of geometrically symmetric functions. In particular, the coefficent vectors lie near the surface of a cone centered on some $\boldsymbol{\omega}$. (b) The opening angle of the cone is given by ϕ' and the angular displacement from ϕ' for a particular $\boldsymbol{\alpha}_\ell$ is specified by $\epsilon_{\boldsymbol{\omega},\ell}$, as defined in Eq. (5.32).

The geometry parameter of the signed sensor symmetric strategy then reads

$$\begin{aligned}\mathcal{G}(\boldsymbol{\omega}) &= \mathcal{G}_{\phi'}(\boldsymbol{\omega}) \\ &\quad + \frac{1}{\mathcal{N}} \sum_{\ell=1}^{n} w_\ell d\Big(-2\epsilon_{\boldsymbol{\omega},\ell} \sin\phi' \cos\phi' - \epsilon_{\boldsymbol{\omega},\ell}^2 \cos(2\phi')\Big) \\ &\quad + \mathcal{O}\left(\epsilon_{\boldsymbol{\omega},\ell}^3\right). \end{aligned} \tag{5.33}$$

Here we expand in powers of $\epsilon_{\boldsymbol{\omega},\ell}$ and define

$$\mathcal{G}_{\phi'}(\boldsymbol{\omega}) = \frac{1}{\mathcal{N}} \sum_{\ell=1}^{n} w_\ell \left(d\cos^2\phi' - 1\right) = d\cos^2\phi' - 1, \tag{5.34}$$

the geometry parameter for measuring a single function at an angle ϕ' from $\boldsymbol{\omega}$. The condition on how small $\epsilon_{\boldsymbol{\omega},\ell}$ needs to be depends on ϕ', but for any particular problem we can determine the necessary condition. In general, as long as $\epsilon_{\boldsymbol{\omega},\ell} \ll 1/\sqrt{d}$, the corrections will be negligible.

Next we consider Eq. (5.21) in the large-d limit and obtain

$$\mathcal{M}_{\text{ss}} = \frac{\mathcal{N}}{t^2}\left(1 - \frac{\mathcal{G}(\boldsymbol{\omega})}{d}\right) + \mathcal{O}\left(\frac{\mathcal{N}}{dt^2}\sqrt{\frac{(1+\mathcal{G}(\boldsymbol{\omega}))(d-\mathcal{G}(\boldsymbol{\omega})-1)}{d-1}}\right), \tag{5.35}$$

for arbitrary values of $\boldsymbol{\omega}$. We substitute Eq. (5.33) and obtain, to leading order in the geometrically symmetric limit and for large d, that

$$\mathcal{M}_{\text{ss}}(\boldsymbol{\omega}) \approx \frac{\mathcal{N}}{t^2}\left(1 - \frac{\mathcal{G}_{\phi'}(\boldsymbol{\omega})}{d}\right) \approx \frac{\mathcal{N}}{t^2}\left(\sin^2\phi' + \frac{1}{d}\right). \tag{5.36}$$

Note that, for $\phi' = 0$, i.e. when all functions are nearly aligned with $\boldsymbol{\omega}$, this reduces to the expected optimal scaling $\mathcal{N}/(t^2 d)$.

We will use these results in the following sections as we compare the signed sensor symmetric strategy to the optimized sequential strategy.

5.4.2 Nearly Overlapping Functions

Next consider the case when all the vectors $\boldsymbol{\alpha}_\ell$ are "close" in each component, i.e. we consider measuring a set of n nearly identical functions. Intuitively, one would expect the optimal sequential strategy in this case to be spending almost all the time measuring the linear combination pointing towards the average of these functions, and then spending a small amount of time measuring in other directions in order to distinguish the small variations in the functions. We find that this intuition is rigorously true. We also find that, in this case, we can analytically determine a scaling advantage (in d) for this protocol relative to the signed sensor symmetric strategy (and,

of course, the unentangled strategy). Finally, we consider a particular example from Ref. [53] and find that its implication about the role of entanglement in protocol performance—namely that it can be disadvantageous in certain circumstances—is limited to the consideration of just the (unsigned) sensor symmetric strategy and is not generally true.

To formally define what we mean by "nearly overlapping", consider angles δ_ℓ associated with each vector of coefficients $\boldsymbol{\alpha}_\ell$ as specified by

$$\cos \delta_\ell = \boldsymbol{\alpha}_\ell \cdot \bar{\boldsymbol{a}}, \tag{5.37}$$

where $\bar{\boldsymbol{a}}$ is a vector, with Euclidean norm equal to 1, chosen such that the average angle $n^{-1} \sum_{\ell=1}^{n} \delta_\ell$ is minimized. For δ_ℓ sufficiently small for all ℓ, $\boldsymbol{\alpha}_\ell \approx \bar{\boldsymbol{a}}$ for all ℓ. Furthermore,

$$\max_i A_{\ell i} = \max_i \bar{a}_i + \mathcal{O}(\delta_\ell), \tag{5.38}$$

for $A_{\ell i} = (\boldsymbol{\alpha}_\ell)_i$. Therefore, with $\delta = \max_\ell \delta_\ell$, we obtain from Eq. (5.30) that

$$\mathcal{M}_{\text{opt}} = \frac{\max_i \bar{a}_i^2 + \mathcal{O}(\delta^2)}{t^2} \min_C \left[\sum_{\ell=1}^{n} \Big| \sum_{m=1}^{n} (C^{-1})_{\ell m} \Big|^{2/3} \right]^3. \tag{5.39}$$

Leaving the somewhat tedious details to Appendix D.3, we find that this reduces to the expected result that

$$\mathcal{M}_{\text{opt}} = \frac{\mathcal{N}}{t^2} \max_i \bar{a}_i^2 + \mathcal{O}\left(\frac{\mathcal{N}\delta^2}{t^2} \right). \tag{5.40}$$

Note that, in general, $\delta \ll 1/\sqrt{d}$ ensures that this is a good leading-order approximation. This

is a reduction in the variance by a factor of approximately (to order δ^2) $\max_i \bar{a}_i^2 \in [1/d, 1]$ compared to the local protocol in Eq. (5.13), or, when compared to the naive sequential protocol in Eq. (5.24), a reduction in the variance by a factor of order $\mathcal{O}(1/n^2)$.

To compare to the signed sensor symmetric protocol, we note that this nearly overlapping case is merely a special case of the nearly geometrically symmetric case of the sensor symmetric protocol (provided δ is sufficiently small). In particular, δ is the relevant expansion parameter for our asymptotic approximations as $\epsilon_{\boldsymbol{\omega},\ell} \leq \delta$ for all ℓ. Therefore, to compare, we may simply use the previous results from Section 5.4.1 with corrections upper bounded by taking $\epsilon_{\boldsymbol{\omega},\ell} \to \delta$.

Furthermore, we note that, to leading order, $\mathcal{M}_{\text{ss}} = \mathcal{N}\mathcal{M}_{\text{ss}}^{(n=1)}$, and similarly, Eq. (5.40) also has the leading-order expression $\mathcal{M}_{\text{opt}} = \mathcal{N}\mathcal{M}_{\text{opt}}^{(n=1)}$, where the right-hand sides correspond to the accuracy $\mathcal{N}$ times the single-function estimation figure of merit. Therefore, we see that, in order to compare the accuracy of both protocols for nearly overlapping functions, it is sufficient to compare their performance for single-function estimation.

Of course, for a single function, the "sequential" strategy is provably optimal as we have reduced it to the case of Ref. [8]. So, at best, the signed sensor symmetric strategy will perform the same as the "sequential" strategy for a single function. For example, we note that for the best case for both strategies—where all functions are oriented along some $\boldsymbol{\omega}$ to order $\mathcal{O}(\delta)$—both approaches have a cost to leading order of $\mathcal{N}/(t^2 d)$, which is superior to the local protocol by $1/d$. Also, for $d = 2$, the time-independent protocol of Ref. [8] does actually utilize sensor

symmetric states, because the initial states are chosen from the set

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$$
$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right), \tag{5.41}$$

and therefore, for all choices of functions with $d = 2$ (where both approaches provide explicitly saturable bounds), the two protocols are identical and optimal.

For $d > 2$, on the other hand, as previously discussed, there may not exist physical states that obtain the figure of merit provided by the signed sensor symmetric strategy. However, even if we assume the figure of merit $\mathcal{M}_{\text{ss}}$ is attainable, we shall see that the optimized sequential strategy can often be the superior choice. In this context, we consider two examples. First, we demonstrate a scaling advantage in d for the sequential protocol in this nearly overlapping limit. Then we revisit the example from Eq. (38) of Ref. [53] and demonstrate that the implication made that entanglement can be detrimental is an artifact of the (unsigned) sensor symmetric approach and that for the better performing sequential protocol, as well as the more general signed sensor symmetric approach, entanglement is useful.

Example 1: To demonstrate an example of a scaling advantage of the sequential protocol over the signed sensor symmetric strategy, suppose we have n nearly overlapping functions such that $\delta \ll 1/\sqrt{d}$ relative to the vector of coefficients given by

$$\bar{\boldsymbol{a}} = \frac{1}{\sqrt{(x^2 - y^2)\kappa + y^2 d}}\big(\underbrace{x, \cdots, x}_{\kappa}, \underbrace{y, \cdots, y}_{d-\kappa}\big)^T, \tag{5.42}$$

where the first κ elements are (up to normalization) $x \in \mathbb{R}$ and the last $d - \kappa$ elements are $y \in \mathbb{R}$.

We assume $x, y = \mathcal{O}(1)$ and $\kappa = \mathcal{O}(d^\beta)$ for $\beta \in [0, 1)$. Without loss of generality, suppose $x > y$. In this case, the cost of the optimized sequential strategy is straightforwardly obtained from Eq. (5.40) to be

$$\begin{aligned}\mathcal{M}_{\text{opt}} &= \frac{\mathcal{N}}{t^2}\left(\frac{x^2}{(x^2-y^2)\kappa + y^2 d}\right) + \mathcal{O}\left(\frac{\mathcal{N}\delta^2}{t^2}\right) \\ &= \frac{\mathcal{N}}{t^2}\frac{x^2}{y^2 d}\left(1 - \frac{(x^2-y^2)\kappa}{y^2 d}\right) + \mathcal{O}\left[\frac{\mathcal{N}}{t^2}\left(\delta^2 + d^{2(\beta-1)}\right)\right],\end{aligned} \tag{5.43}$$

where the second line comes from expanding in powers of κ/d. For the signed sensor symmetric strategy for the same problem, we pick $\boldsymbol{\omega}$ such that $\boldsymbol{\omega}_i = \text{sgn}(\bar{a}_i)$, which minimizes the angle between $\bar{\boldsymbol{a}}$ and $\boldsymbol{\omega}$. In the large d limit, we may then use Eq. (5.36) with

$$\cos^2\phi' = \frac{(\bar{\boldsymbol{a}}\cdot\boldsymbol{\omega})^2}{d} = \frac{\left[(|x|-|y|)\kappa + |y|d\right]^2}{d\left[(x^2-y^2)\kappa + y^2 d\right]}. \tag{5.44}$$

We can perform an expansion of the numerator of Eq. (5.44) in powers of κ/d as

$$\begin{aligned}\left[(|x|-|y|)\kappa + |y|d\right]^2 &= |y|^2 d^2\left[1 + \frac{(|x|-|y|)\kappa}{|y|d}\right]^2 \\ &= |y|^2 d^2\left[1 + \frac{2(|x|-|y|)\kappa}{|y|d} + \mathcal{O}\left(\frac{\kappa^2}{d^2}\right)\right],\end{aligned} \tag{5.45}$$

and expand the denominator as

$$\begin{aligned}\frac{1}{d\left[(x^2-y^2)\kappa + y^2 d\right]} &= \frac{1}{y^2 d^2}\left[1 + \frac{(x^2-y^2)\kappa}{y^2 d}\right]^{-1} \\ &= \frac{1}{y^2 d^2}\left[1 - \frac{(x^2-y^2)\kappa}{y^2 d} + \mathcal{O}\left(\frac{\kappa^2}{d^2}\right)\right].\end{aligned} \tag{5.46}$$

We then have

$$\sin^2\phi' = 1 - \cos^2\phi' = \frac{(|x| - |y|)^2\kappa}{y^2 d} + \mathcal{O}\left(\frac{\kappa^2}{d^2}\right), \tag{5.47}$$

which we may plug into Eq. (5.36) for the signed sensor symmetric strategy

$$\mathcal{M}_{\text{ss}} = \frac{\mathcal{N}}{t^2}\frac{(|x| - |y|)^2\kappa}{y^2 d} + \mathcal{O}\left[\frac{\mathcal{N}}{t^2}\left(\delta^2 + d^{2(\beta-1)}\right)\right], \tag{5.48}$$

which demonstrates a scaling advantage by a factor of $\mathcal{O}\left(\kappa^{-1}\right) = \mathcal{O}\left(d^{-\beta}\right)$ for the optimized sequential protocol in this problem.

Example 2: Now we consider the example of a single function from Eq. (38) of Ref. [53] for $d = 3$ sensors and coefficient vector[3]

$$\boldsymbol{\alpha} = \frac{1}{\sqrt{18}}\begin{pmatrix} \sqrt{2} + \sqrt{3} + 1 \\ \sqrt{2} - \sqrt{3} + 1 \\ \sqrt{2} - 2 \end{pmatrix}. \tag{5.49}$$

The example was chosen in Ref. [53] such that for $\boldsymbol{\omega} = \mathbf{1}$, $\mathcal{G}(\boldsymbol{\omega}) = 0$, and thus $\mathcal{J}_{\text{opt}}(\boldsymbol{\omega}) = 0$, which in turn implies that the optimal (unsigned) sensor symmetric state is unentangled. Equation (5.21) then implies

$$\mathcal{M}_{\text{ss}}(\boldsymbol{\omega} = \mathbf{1}) = \frac{1}{t^2}, \tag{5.50}$$

which is larger than the true optimal figure of merit, which is obtained by the "sequential" proto-

[3] We have normalized differently by a factor of $1/\sqrt{3}$ from Eq. (38) of Ref. [53] in order to match our assumption that $||\boldsymbol{\alpha}||^2 = 1$.

Setting	Signed Sensor Symmetric	Optimized Sequential
Geometrically symmetric limit (large d)	$\mathcal{M}_{\text{ss}}(\boldsymbol{\omega}) \approx \frac{\mathcal{N}}{t^2}\left(1 - \frac{\mathcal{G}_{\phi'}(\boldsymbol{\omega})}{d}\right) \approx \frac{\mathcal{N}}{t^2}\left(\sin^2\phi' + \frac{1}{d}\right)$ $\phi' :=$ angle of functions w.r.t. $\boldsymbol{\omega}$	
Nearly overlapping limit	Same as geometrically symmetric limit	$\mathcal{M}_{\text{opt}} = \frac{\mathcal{N}}{t^2}\max_i \bar{a}_i^2 + \mathcal{O}\left(\frac{\mathcal{N}\delta^2}{t^2}\right)$ Functions aligned along $\bar{\boldsymbol{a}}$
Best Case Functions aligned along some $\boldsymbol{\omega}$	$\mathcal{M}_{\text{ss}} = \frac{\mathcal{N}}{dt^2}$	$\mathcal{M}_{\text{opt}} = \frac{\mathcal{N}}{dt^2}$
Example 1 (Scaling) Scaling advantage for $\mathcal{M}_{\text{opt}}$ Functions aligned along Eq. (5.42)	$\mathcal{M}_{\text{ss}} = \mathcal{O}\left(\frac{\mathcal{N}\kappa}{dt^2}\right) = \mathcal{O}\left(\frac{\mathcal{N}}{d^{1-\beta}t^2}\right)$ (note: $\beta \in [0,1)$)	$\mathcal{M}_{\text{opt}} = \mathcal{O}\left(\frac{\mathcal{N}}{dt^2}\right)$

Table 5.2: Summary of analytic results comparing the signed sensor symmetric strategy and optimized sequential strategy. Recall that the figure of merit for the local strategy is $\mathcal{N}/t^2$.

col:

$$\mathcal{M}_{\text{opt}} = \frac{1}{t^2}\left(\frac{\sqrt{2}+\sqrt{3}+1}{\sqrt{18}}\right)^2 \approx \frac{0.9551}{t^2}. \tag{5.51}$$

We also note that, even within the framework of sensor symmetric strategies, the result obtained from Ref. [53] is not the best one can do. If we extend to the signed sensor symmetric approach, one can consider $\boldsymbol{\omega} = (1, 1, -1)^T$ and do better. In particular, in this case, one obtains

$$\mathcal{M}_{\text{ss}}(\boldsymbol{\omega}) = \frac{0.9554}{t^2}, \tag{5.52}$$

which is only slightly worse than the true optimum, and, crucially, also involves entanglement. Therefore, from this example, we learn that (a) entanglement *is* helpful for measuring the function in Eq. (5.49), just not when we restrict to (unsigned) sensor symmetric states, and (b) accuracy is (unsurprisingly) potentially decreased when restricting ourselves to sensor symmetric states.

For convenience, we summarize the analytic results comparing the signed sensor symmetric and optimized sequential strategies in Table 5.2.

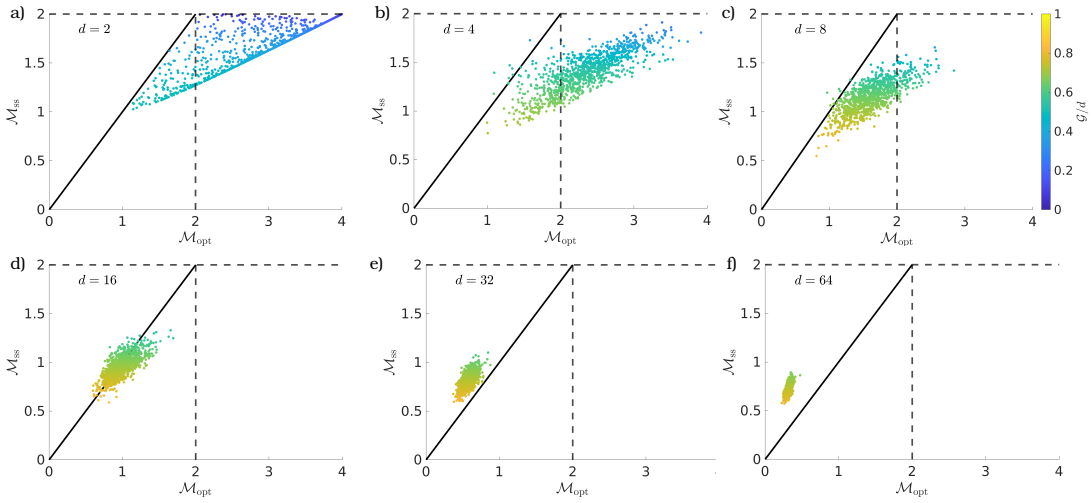


Figure 5.4: $\mathcal{M}_{\text{ss}}$ versus $\mathcal{M}_{\text{opt}}$ for 1000 random samples from the positive orthant of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ with $n = 2$, $w_1 = w_2 = 1$ for different numbers of sensors d. Dashed lines correspond to $\mathcal{M}_{\text{local}}$. Colors correspond to the geometry parameter for the problem instance. Observe that the signed sensor symmetric approach is never worse than the local protocol, whereas the optimized sequential protocol can be. However, as d increases the optimized sequential protocol is almost always superior. Also recall, that for $d > 2$, $\mathcal{M}_{\text{ss}}$ is generically just a lower bound, and it is not guaranteed one can achieve this figure of merit with physical states. Therefore, one can think of $\mathcal{M}_{\text{ss}}$ as a best case scenario for a physically realized signed sensor symmetric protocol.

5.4.3 Numerical Results

In the previous sections, we found that both the optimized sequential and signed sensor-symmetric strategies perform identically (and optimally) when measuring many functions whose coefficient vectors $\{\boldsymbol{\alpha}_\ell\}$ are aligned along a particular $\boldsymbol{\omega}$. More generally, the optimized sequential protocol always performs at least as well as, and typically outperforms the signed sensor symmetric strategy when measuring many functions with nearly overlapping coefficient vectors, and in fact, we can obtain a scaling advantage in d for certain problems (Example 1). However, while informative, the nearly overlapping limit considered above is such that the optimized sequential strategy performs its best. Therefore, it is of interest to also consider a broader class of examples and to consider where the signed sensor symmetric strategy outperforms the optimized

sequential strategy.

Unfortunately, however, a full analytic comparison between the different approaches is beyond reach as far as we know, so for a general problem, one must therefore compare the two approaches explicitly to see which one is the correct choice for a given situation. Here, to better understand the expected performance in such cases, we turn to numerics on random problem instances. Our key result is to demonstrate that generically, for large d, many problems are best approached using our optimized sequential protocol as opposed to the sensor symmetric or local strategies.

Numerically, the optimization over C in Eq. (5.30), subject to Eq. (5.29), to obtain the cost of the optimized sequential protocol can be fairly costly in terms of computation time, as the optimization is non-convex and in a high dimensional parameter space. This is not necessarily an issue for particular applications, where only a limited number of such optimizations must be performed. As an example, consider $n = 2$ functions, $d \geq n$ sensors, and equal weights in the figure of merit ($w_1 = w_2 = 1$). The normalization condition (5.29) implies that the columns of the 2×2 matrix C have unit length. We can parametrize this by two angles via

$$C = \begin{pmatrix} \cos\varphi & \cos\varphi' \\ \sin\varphi & \sin\varphi' \end{pmatrix}. \tag{5.53}$$

Given the coefficient vectors $\boldsymbol{\alpha}_{1,2}$ of the two functions to be estimated, the numerical optimization over φ, φ' is accomplished straightforwardly. For $n = 3$ functions, six angles $\varphi_1, \ldots, \varphi_6$ are needed, making the optimization more challenging for larger n.

The two functions, represented by the two normalized coefficient vectors $\boldsymbol{\alpha}_{1,2}$, depend on

$2(d-1)$ real parameters. In this context, we randomly sample coefficients for the two functions from a uniform distribution and calculate the cost of the signed sensor symmetric strategy and the optimized sequential strategy. For $d = 2^k$ for $k \in [1, 6]$, we consider 1000 such problems where for simplicity we assume that $\boldsymbol{\alpha}_{1,2}$ are sampled from the positive orthant so that the optimal $\boldsymbol{\omega}$ is necessarily $\mathbf{1}$ and plot the results in Fig. 5.4.

We observe that the signed sensor symmetric strategy is never worse than the local protocol, whereas the optimized sequential protocol can be at small d. In the particular case of $n = d = 2$, the sequential strategy is never better than the signed sensor symmetric strategy. As previously mentioned, it is well known that, for this problem, when the two functions are orthogonal, a local protocol obtains the optimal variance (that is, $\mathcal{M} = \mathcal{N}/t^2$ is optimal) [9, 52]. In this case, as demonstrated in Ref. [53], the sensor symmetric strategy matches this known optimal result. In particular, the sensor symmetric strategy predicts an optimal geometry parameter $\mathcal{G}(\omega) = 0$, corresponding to no inter-sensor correlations and, therefore, a local protocol. We observe this behavior in panel a) of Fig. 5.4 where the $\mathcal{G} = 0$ points correspond to $\mathcal{M}_{\text{ss}} = \mathcal{M}_{\text{local}} = 2$. Note that cases of $\mathcal{G} \approx 0$ that correspond to nearly orthogonal coefficient vectors are only those points where $\mathcal{M}_{\text{opt}} \approx 4$, as can be concluded from Fig. 5.5 where we plot $\mathcal{M}_{\text{opt}}$ versus $\boldsymbol{\alpha_1} \cdot \boldsymbol{\alpha_2}$. As d increases, however, the optimized sequential protocol is almost always superior to both the local and signed sensor symmetric strategies for these randomized problem instances.

5.5 Conclusion and Outlook

In this work, we explored the potential of sequential protocols to measure multiple functions with quantum sensor networks. We highlighted both analytical and numerical aspects, and

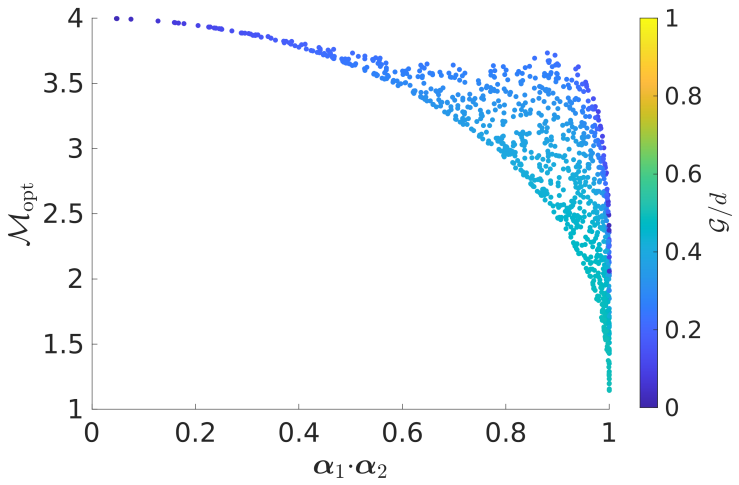


Figure 5.5: $\mathcal{M}_{\rm opt}$ versus $\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2$ for $n = 2$ functions and $d = 2$ sensors. Note that the nearly orthogonal case ($\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 \approx 0$) implies $\mathcal{M}_{\rm opt} \approx 4$ (i.e., the worst case for the optimized sequential strategy). Comparing to the first panel of Fig. 5.4 we see that in this case $\mathcal{M}_{\rm ss} \approx \mathcal{M}_{\rm local} = 2$.

compared the protocol to a generalized version of the sensor symmetric bounds for the same problem from Ref. [53]. We find that, when d is large, the sequential protocol is typically superior for generic problem instances. The sequential strategy also has the advantage of having an explicit protocol to obtain its given performance, whereas beyond $d = 2$, while shown to be saturable in certain cases [53], the lower bound when restricted to signed sensor symmetric states is not guaranteed to always be attainable. However, for a particular problem, one should compare both strategies, as neither is always superior.

Our results, together with those in Ref. [53], point to an intriguing interplay between the geometric configuration of the functions to be measured and the performance of various protocols. In particular, our optimized sequential protocol performs best with nearly overlapping functions; the signed sensor symmetric approach performs best when the set $\{\boldsymbol{\alpha}_\ell\}$ is nearly aligned along some $\boldsymbol{\omega}$. Beyond carefully tuned examples, we note that for most problems where we seek to estimate a collection of analytic functions of local field amplitudes, our protocol is the best known

choice, especially with more than a small number of sensors d.

Our sequential protocol could directly be extended to the case where the sensors are each coupled to correlated field amplitudes as in the recent work by some of the authors [59]; that is, instead of considering independent field amplitudes θ_i coupled to the sensors, one could consider the case where $\boldsymbol{\theta}$ is specified by a known analytic parameterization by some set of $k \leq d$ parameters.

Our sequential protocol could also be extended to other physical settings beyond qubit sensors—namely, for any quantum sensor network where one may measure a single linear combination of field amplitudes, one can apply our sequential approach. For example, a collection of d Mach-Zehnder interferometers could replace the qubit sensors, where the role of the local fields is played by interferometer phases [101, 122–126]. Here, the limiting resource is the number of photons N available to distribute among interferometers as opposed to the total time t. In this context, it was conjectured in Ref. [51] that one could measure a single function with variance $\mathcal{M} = \frac{\|\boldsymbol{w}\|_1^2}{N^2}$—this replaces Eq. (5.4), and otherwise everything remains the same.[4] However, there are subtleties in the case where the average number of photons is not known [127], which we do not consider here. Another relevant setting is the measurement of linear combinations of field-quadrature displacements as considered using an entanglement-enhanced continuous-variable protocol in Ref. [80]. A variation of this protocol was experimentally implemented in Ref. [86]. One could also consider a combination of these settings where some field amplitudes are coupled to qubits, some to Mach-Zehnder interferometers, and some to field-quadrature displacements.

While the importance of geometry is striking, the general question of the information-

[4] *Comment:* Note this conjecture is proven in Chapter 3 of this thesis.

theoretic optimal strategy that minimizes the quantum Fisher information for this problem remains a pressing open question. Additionally, our results are asymptotic and ignore the potential effects of decoherence. Understanding the performance of the sequential protocol in the non-asymptotic regime (i.e., via Bayesian analysis as considered in Ref. [53]) and under the effects of decoherence remains a question of great importance. These limitations aside, our findings advance the understanding of measuring multiple functions with quantum sensor networks and provides an alternative protocol that practically performs better than previously considered schemes in many instances.

Chapter 6: Discussion and Further Directions for Part I

Chapters 2-5 tell a fairly complete story about the asymptotic limits of function estimation in quantum sensor networks and the protocols required to reach those limits. For both qubit-based and photonic sensors, we have shown a tight algebraic connection between fundamental precision bounds and the resources and protocols required to achieve them. We have also emphasized how settings that allow for freedom in choosing the function to estimate enable one to make better use of entanglement—namely, by measuring more global functions for which one can obtain the greatest precision gain over unentangled protocols.

However, a number of interesting questions beyond immediate applications and extensions remain. A couple directions seem particularly compelling:

- *Secure Networked Quantum Sensing:* As pointed out in Ref. [8], for protocols using local measurements, linear function estimation is, in a weak sense, secure against sensor nodes being compromised by malicious parties. In particular, from any strict subset of local measurements one cannot infer the value of the function being estimated. Stricter forms of security against attack—both internal and external to the network—are of clear interest, however. For instance, one can consider situations where some nodes are compromised, communication channels between nodes are compromised, or a situation where one would like to delegate the whole operation of the quantum network to a third party without re-

vealing information about what function of the unknown parameters one wants to extract. Some initial work along these lines has been done [140–142], but there is much left to explore. The key challenge in this setting, compared to similar questions related to secure or delegated quantum computing, is the existence of unknown parameters, which intrinsically limits the sorts of attacks that can be defended against; for instance, a malicious party controlling one node of the sensor network could always measure a field they are not supposed to and, if the correct field is not known, this could never be detected. We hope that lessons from cryptography or delegated quantum computing can help circumvent certain aspects of this problem.

- *Multiparameter Quantum Sensing:* The work contained in this dissertation depended heavily on the fact that, while "under the hood" the problems considered are multiparameter problems, a function of those parameters is a single quantity. This allowed us to make clever use of single parameter Cramér-Rao bounds, subject to some constraints that arose from the multiparameter nature of the problem. While we extended beyond this setting somewhat in Chapter 5, in that chapter we focused on protocols, not their ultimate optimality. To fully understand the resource requirements and landscape of optimal protocols for multiple function estimation other techniques are needed. A key tool in this problem will be the more general Holevo Cramér-Rao bound [48, 104], which unlike the Helstrom version used in this dissertation, is saturable even in the multiparameter setting—although only with multi-copy measurements.

  Relatedly, the choice of cost function in a genuinely multiparameter problem makes for a rich optimization landscape. For instance, instead of the standard trace of the covariance

matrix considered in Chapter 5 and in the Holevo Cramér-Rao bound, one could consider other norms on the covariance matrix. Analysis of alternative choices has a long history in the classical parameter estimation literature (see, for instance, [143]), as well as in the related area of optimal design of experiments [144], but is relatively unexplored in the quantum setting.

Part II

Quantum Optimization

Chapter 7: Polynomial Time Algorithms for Estimating Spectra of Adiabatic Hamiltonians

7.1 Introduction

Since their introduction in [145], so-called *stoquastic* Hamiltonians, those with real non-positive off-diagonal matrix elements, have been a major point of focus for research regarding adiabatic quantum computation (AQC)[146]. Stoquastic Hamiltonians have real, non-negative ground states, and therefore, a question of particular interest is whether AQC with stoquastic Hamiltonians is capable of exponential speedup over classical computation. We note that the computational cost for an AQC problem is determined via adiabatic theorems which upper bound the runtime as the inverse of the eigenvalue gap between the lowest two eigenvalues squared [147, 148]. As a complexity theory question, the computational power of stoquastic AQC is still unknown, but for specific classical algorithms such as path integral [15] or diffusion Monte Carlo (MC) [14, 149], examples have been presented where exponential speedup *over the specific classical algorithm* is indeed still possible with stoquastic Hamiltonians.

However such finely tuned examples raise new questions as to whether such obstructions to classical simulation are typical of more general stoquastic Hamiltonians. The diffusion MC examples and others, such as the well studied "spike" example [14, 150–154] take advantage

of heavy symmetry with a potential that is a function of Hamming weight to allow for efficient analytic and computational analysis. It has been shown that such problems are always efficiently classically simulatable by path integral quantum Monte Carlo (QMC) [155], which suggests somewhat more complicated models would be helpful for addressing these questions.

Here we consider such a model designed to be both efficiently analyzable and somewhat more realistic than purely Hamming symmetric problems. In particular, we expect realistic cost functions to have many local minima; therefore we consider a collection of $K = \mathrm{poly}(n)$ individually Hamming symmetric wells. While the full Hamiltonian in such a model is no longer Hamming symmetric, enough symmetry remains to allow for a similar reduction to an effective polynomial sized subspace. We show that this reduction can always be exact for $K = 2, 3$.

For larger K we introduce an approximate tight-binding scheme for analyzing the model. The reduction this model affords is represented diagramatically in Fig. 7.1 for three Hamming symmetric potential balls on 10 qubits. In addition to being computationally efficient, this tight-binding model makes the effects of tunneling readily apparent; in the tight-binding model, the minimum eigenvalue gap is in many cases dictated by the matrix element between ground states of neighboring potential wells, which in turn is dominated by the "tunneling" part of the wavefunctions, i.e. the amplitudes on bit strings for which the potential energy is greater than the eigenenergy of the state. As interference is not manifest in the ground state of stoquastic Hamiltonians, it is expected that if AQC with such Hamiltonians is to provide advantages over classical computation these advantages should lie in the power afforded by tunneling effects between local minima of the cost function; therefore even if the model presented here also proves to be efficiently simulatable classically, it still provides a useful new tool set for addressing and understanding the performance of AQC with more realistic stoquastic Hamiltonians.

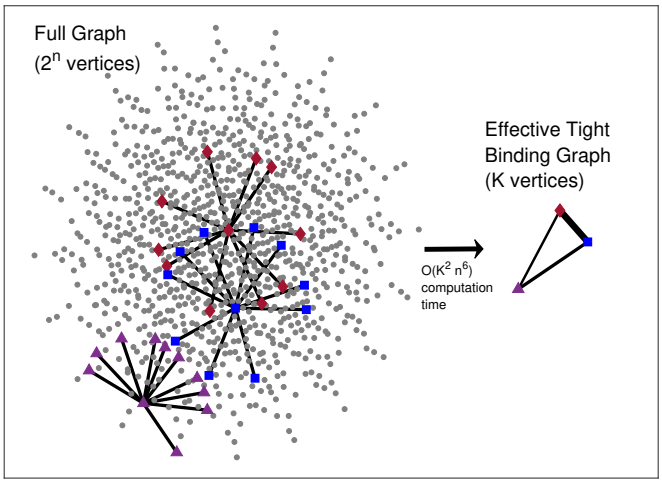


Figure 7.1: The key feature of our model and the tight-binding approach to analyzing it is the ability to exponentially reduce the size of the problem to an effective Hamiltonian that explicitly considers tunneling between a collection of Hamming symmetric wells. Here we diagramatically demonstrate this reduction for a collection of 3 Hamming symmetric wells of width 1 on 10 qubits. The original graph representing our Hamiltonian has a hypercube geometry, but the edges not within wells are eliminated for visual clarity.

In this [chapter] we present this model and our exact and approximate algorithms for analyzing it, along with a collection of examples designed to highlight their strengths and limitations.

7.2 Tight-binding approximation

Hamiltonians with $-\sum_j X_j$ driving terms and Hamming-symmetric potentials comprise a common class of Hamiltonians considered in AQC [14, 149–154], where Hamming weight is defined as the number of ones in a bit string corresponding to a basis state of the Hilbert space. Such Hamiltonians are often used due to the ability to block diagonalize the Hamiltonian into smaller subspaces: if we think of our qubits as spin-1/2 particles each block corresponds to a possible total angular momentum j and we can parameterize within each block by the z-projection of the angular momentum m and a parameter γ labeling the degeneracies of the j, m representations.

Here we introduce a reparameterization $\sigma = n/2 - j \in [0, n/2]$, $w = n/2 + m \in [\sigma, n - \sigma]$ which indexes the permutation symmetric subspace as the $\sigma = 0$ subspace and defines m in terms of Hamming weight. The ground state and thus the relevant spectral gap of such Hamiltonians is guaranteed by the Perron-Frobenius theorem to lie in the exponentially-reduced permutation symmetric $\sigma = 0, \gamma = 0$ subspace.

As we use the Perron-Frobenius theorem several times throughout this [chapter], we restate it here for convenience: Let A be a matrix with all real and non-negative entries. Then A has a unique leading eigenvalue with corresponding eigenvector with all elements strictly positive.

In the context of stoquastic AQC, we have a Hamiltonian with all non-positive matrix elements so $\exp(-\beta H) \simeq 1 - \beta H$ is nonnegative for sufficiently small $\beta > 0$ and the Perron-Frobenius theorem applies which guarantees a nondegenerate, real, nonnegative ground state. Furthermore, as the ground state is nondegenerate, it is guaranteed to transform in a one dimensional representation of any symmetry group of the Hamiltonian. In particular, Hamming-symmetric Hamiltonians have the symmetry group S_n, which has only two one-dimensional representations: the trivial representation and the sign representation. By the positivity of the amplitudes of the ground state, it cannot transform according to the sign representation and therefore is invariant under all permutations of the qubits.

Therefore, analyzing the Hamming symmetric problem in this subspace which has a basis parameterized by the Hamming weight w enables efficient computation of the spectral gap. A detailed review of this reduction is presented in Appendix E.1.

Here, we consider a generalization of the standard problem. Instead of a fully Hamming-symmetric Hamiltonian, we specify K bit strings, each with a Hamming-symmetric potential

"well" about it. This Hamiltonian is of the form

$$H(s) = -\frac{a(s)}{n}\sum_j X_j + \sum_{k=0}^{K-1} b_k(s) V_k\Big(\sum_j X^{\bar{k}}\bar{Z}_j X^{\bar{k}}\Big) \tag{7.1}$$

where $X^{\bar{k}} = \bigotimes_{i=1}^{n} X^{\bar{k}_i}$ is a bit shift operator shifting a bit string from $\bar{k} \in \{0,1\}^n$ to the all zeros bit string and $\bar{Z} = (I - Z)/2 = |1\rangle\langle 1|$ is the Hamming weight operator. The remaining (standard) notation is introduced and defined in Appendix E.1. Note, for example, that the Grover Hamiltonian with K marked items is a special (simple) case of this Hamiltonian, with $V_k = -1$ for $\sum_j X^{\bar{k}}\bar{Z}_j X^{\bar{k}}|x\rangle = 0$ and 0 otherwise.

Despite the loss of full Hamming symmetry, a similar reduction of Hamiltonians of this form exists, making it possible to calculate the spectral gap of the full problem efficiently. Relabeling symmetries that make this calculation exact for $K = 2, 3$ are described in Appendix (E.2). Here we simply indicate a key notation from these exact results for the case of $K = 2$: in analogy with the fully Hamming symmetric case we now label our subspaces by a coordinate pair (σ_1, σ_2). Basis states are further parameterized by two integers $h_1 \in [0, n_1]$ and $h_2 \in [0, n - n_1]$ where n_1 is the Hamming distance between the two wells and a pair (γ_1, γ_2) labeling degeneracies of representation. The ground state of the Hamiltonian is guaranteed to be in the $(\sigma_1, \sigma_2) = (0, 0)$ subspace. The details are left to the appendix. For $K > 3$, we introduce a tight-binding approximation, to which we now turn.

We first consider each of the K wells individually. The eigenstates for each well can be directly and efficiently calculated, as long as one ignores the existence of the other wells. We denote the ground state of the k^{th} isolated well by $|\psi_k^{(0)}\rangle$, and the set of such ground states by $T^{(0)} \equiv \{\psi_k^{(0)}\}$. Similarly, we denote $T^{(1)} \equiv T^{(0)} \cup \{\psi_k^{(1)}\}$ as the set of ground states and (possibly

degenerate) first excited states of the individual wells.

Our zeroth (first) order tight-binding model ansatz consists of the assumption that the ground state $|\phi^{(0)}\rangle$ and first excited state $|\phi^{(1)}\rangle$ of the full Hamiltonian exist in the span of the elements of $T^{(0)}$ ($T^{(1)}$). Therefore starting with the eigenvalue equation $H|\phi\rangle = E|\phi\rangle$ we can insert the tight-binding ansatz $|\phi\rangle = \sum_j c_j|\psi_j\rangle$ for some coefficients c_j to give for the lowest two energy states

$$H\sum_{j=0}^{K-1} c_j|\psi_j\rangle = E\sum_{j=0}^{K-1} c_j|\psi_j\rangle. \tag{7.2}$$

Then multiplying through by the complete set $T^{(0)}$ (or $T^{(1)}$) of basis states gives the generalized eigenproblem

$$\sum_{i,j} c_j H_{ij}^{(\mathrm{TB})}|\psi_i\rangle = E\sum_{i,j} c_j S_{ij}|\psi_i\rangle \tag{7.3}$$

where $H_{ij}^{(\mathrm{TB})} = \langle\psi_i|H|\psi_j\rangle$ and $S_{ij} = \langle\psi_i|\psi_j\rangle$. Solving this generalized eigenproblem gives a variational solution for the lowest two energy states.

To calculate the elements $H_{ij}^{(\mathrm{TB})}$ and S_{ij} we use the exponentially reduced subspaces corresponding to the pair of wells i, j (as described in detail in Appendix E.2) to calculate $|\psi_i\rangle$ and $|\psi_j\rangle$. Once we have the basis states $|\psi_i\rangle$ and $|\psi_j\rangle$, calculating the overlap S_{ij} is then self explanatory. To calculate $H_{ij}^{(\mathrm{TB})} = \langle\psi_i|H|\psi_j\rangle$ in this subspace we can exactly write the driver part of the Hamiltonian and the diagonal term corresponding to the wells i, j and then add a correction term to exactly include the effects of the other wells in this matrix element. In particular we can write

$$H_{ij}^{(TB)} = \langle\psi_i|H_d + \sum_{h_1,h_2}[V_i(h_1,h_2) + V_j(h_1,h_2) + V_c(h_1,h_2)]|\psi_j\rangle \tag{7.4}$$

where H_d is the driver part of the Hamiltonian in the appropriate basis V_i and V_j are the diagonal

potential terms corresponding to the i^{th} and j^{th} wells and

$$\text{diag}(V_c) = \frac{\sum_{k\neq i,j}\sum_{r_k=0}^{n} N(h_1, h_2, n_1, R_{ik}, R_{jk}, r_k)V_k(r_k)}{\sqrt{\binom{n_1}{h_1}\binom{n-n_2}{h_2}}} \tag{7.5}$$

where n_1, R_{ik} and R_{jk} are the Hamming distance between the i^{th} and j^{th} wells, the i^{th} and k^{th} wells and the j^{th} and k^{th} wells, respectively, r_k is the distance from the k^{th} well and $V_k(r_k)$ is the potential due to the k^{th} well at distance r_k. The function $N(h_1, h_2, n_1, R_{ik}, R_{jk}, r_k)$ gives the number of points of intersection between Hamming spheres of radius $r_i = h_1 + h_2$ and $r_j = (n_1 - h_1) + h_2$ centered on the i^{th} and j^{th} wells respectively and the Hamming sphere of radius r_k centered on the k^{th} well. For details on how to calculate N see Appendix E.3.

Note that calculating the matrix elements $H_{ij}^{(TB)}$ and S_{ij} is only efficient if we can calculate them only by considering a constant (or polynomial) set of reduced subspaces (fixed or bounded $(\sigma_1^{(ij)}, \sigma_2^{(ij)})$) of the ij^{th} basis. By the Perron-Frobenius theorem and symmetry, the ground state of any well is guaranteed to be in the $(\sigma_1^{(ij)}, \sigma_2^{(ij)}) = (0, 0)$ subspace. So if we just consider zeroth order tight-binding we must only compute individual ground states in this subspace. If we want to include first excited states as in first order tight-binding, however, we must consider the possibility that those states exist in $(\sigma_1^{(ij)}, \sigma_2^{(ij)}) > (0, 0)$ subspaces.

In Appendix E.4 we prove that the first excited state for a given well is guaranteed to exist in one of the (0,0), (1,0), or (0,1) subspaces and thus limits us to a constant set of polynomially-sized subspaces we must diagonalize for first order tight-binding. Here we give a sketch of the proof and the motivating ideas. To simplify things we note that a single well Hamiltonian can also be written in terms of the standard Hamming symmetric subspaces labeled by σ and $\sigma_1^{(ij)} + \sigma_2^{(ij)} \leq \sigma$. That is we can show this result by demonstrating that the first excited state of

well belongs in either the $\sigma = 0$ or $\sigma = 1$ subspace.

Start by considering just the driving term of the Hamiltonian $H_d \propto -\sum_j X_j$. The ground state of H_d is therefore proportional to $|+\rangle^{\otimes n}$. The first excited state is n fold degenerate where one of the n bits is flipped to a $|-\rangle$ state. An equal superposition of these states is a permutation symmetric ($\sigma = 0$) eigenstate, leaving $n - 1$ states in the $\sigma = 1$ subspace, each labeled by a different γ in our $|w\sigma\gamma\rangle$ basis. We note that while within the $\sigma = 0$ subspace this first excited eigenstate is the second lowest eigenvalue, in the $\sigma = 1$ subspaces for fixed $\gamma = \gamma'$ each of these eigenstates corresponds to the smallest unique eigenvalue *within these subspaces*. Additionally, as each of these fixed $\gamma = \gamma'$, $\sigma = 1$ subspaces is itself a stoquastic matrix, by the Perron-Frobenius theorem each of these candidate first excited states is real and non-negative in its respective subspace.

If we add a Hamming symmetric well, the degeneracy in the first excited state is broken between the $\sigma = 0$ state and the $\sigma = 1$ states. Which of these is energetically favored depends on the potential and the relative strength of the driving and potential terms, but the $\sigma > 1$ eigenstates can never have lower energy than these states even following the breaking of the degeneracy. To see this we consider an eigenstate $|\psi\rangle = \sum_{w\sigma\gamma} \alpha(w, \sigma, \gamma)|w\sigma\gamma\rangle$ with corresponding energy (independent of w for $\alpha(w, \sigma, \gamma) \neq 0$)

$$E(s) = -\frac{1-s}{n}\left(r^+C^+ + r^-C^-\right) + sV(w) \tag{7.6}$$

where $r^\pm = \frac{\alpha(w\pm1,\sigma,\gamma)}{\alpha(w,\sigma,\gamma)}$ for all w, σ, γ and $C^\pm$ are standard spin-1/2 raising and lowering coefficients (and functions of w and σ). Note that the potential term is independent of σ and thus does not affect which subspace is energetically favored. However, for a subspace σ and a subspace

$\sigma' > \sigma$ at fixed w, $C^{\pm}(\sigma' > \sigma) < C^{\pm}(\sigma)$ so if $r^{\pm}$ were independent of σ then the $\sigma = 0$ subspace would always be favored. Consider the energy difference between the candidate first excited states

$$\Delta E(s) = E_{\sigma'}(s) - E_{\sigma}(s) = \frac{1-s}{n}[r_{\sigma}^{+}C_{\sigma}^{+} + r_{\sigma}^{-}C_{\sigma}^{-} - r_{\sigma'}^{+}C_{\sigma'}^{+} - r_{\sigma'}^{-}C_{\sigma'}^{-}]. \tag{7.7}$$

This equation is independent of w so we take $w = \sigma'$ so that $C_{\sigma'}^{-}(w = \sigma') = 0$ to eliminate one term. For $\sigma > 0$, $r^{\pm}$ must be nonnegative (by the Perron-Frobenius theorem) so $\Delta E(s)$ is nonnegative unless $r_{\sigma'}^{+}$ is large relative to $r_{\sigma}^{\pm}$. By analyzing the eigenvector equation in both subspaces, however, and using the fact that $C_{\sigma'}^{+} < C_{\sigma}$ we obtain

$$\frac{E_{\sigma'} - sV}{E_{\sigma} - sV} > \frac{r_{\sigma'}^{+}}{r_{\sigma}^{+}}. \tag{7.8}$$

Both sides are positive definite for $\sigma, \sigma' > 0$. And as sV is the same in both subspaces, if $E_{\sigma'} < E_{\sigma}$ then $r_{\sigma'}^{+} < r_{\sigma}^{+}$ but this contradicts that $r_{\sigma}^{+} < r_{\sigma'}^{+}$. Therefore the first excited state must always exist either in the $\sigma = 0$ or $\sigma = 1$ subspaces.

Additionally, we can see how the $\sigma = 1$ subspace may be energetically favored over the $\sigma = 0$ subspace: the $\sigma = 0$ subspace does not have the positive definite restriction on $r^{\pm}$, so therefore if there is a sufficiently rapid sign change in the first excited state wavefunction in the $\sigma = 0$ subspace as we may see in a bound state of a well, then the $\sigma = 1$ subspace is energetically favored.

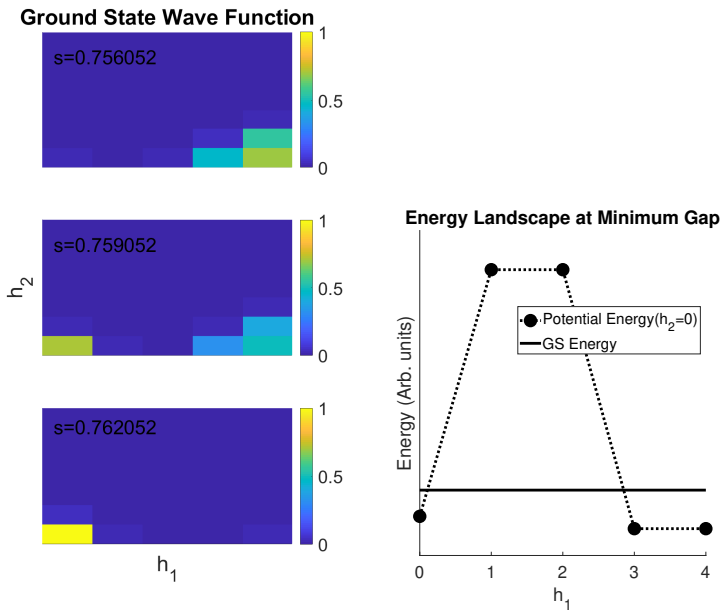
Thus, to perform first order tight-binding we must simply check the $\sigma = 0$ and $\sigma = 1$ subspaces for the first excited state for each well. If the $\sigma = 1$ states are energetically favored we

use all $n-1$ degenerate first excited states for that well in the tight-binding calculation. Therefore the tight-binding matrices for first order tight-binding can be up to $nK \times nK$ in size.

Before discussing numerical details of the practical implementation of this tight-binding framework, note that this framework explicitly considers the tunneling between potential wells. In particular non-zero off diagonal elements of the matrices $H^{(TB)}$ and S are due to evanescent portions of the bound state wavefunctions that extend beyond their respective wells. When the gap is small the ground state wave function can vary rapidly with s as depicted in Figure 7.2. To understand how this relates to the standard conception of tunneling note that we can make the driver term correspond to our intuitive conception of a kinetic energy term by pulling out an s dependent diagonal term $\propto I$ from the second term in 7.1 to obtain the standard normalized graph Laplacian for this system $H_k = (1-s)[I - \frac{1}{n}\sum_j X_j]$ which can be considered a kinetic energy term [149]. The remaining diagonal potential term is the corresponding potential energy.

7.3 Numerical Considerations and Error Estimates

Once we construct the tight-binding matrices $H^{(TB)}$ and S, we must solve the generalized eigenproblem given in 7.3. This is complicated by the fact that S may be ill-conditioned, leading to numerical instabilities. We address this by using the Fix-Heiberger reduction algorithm for solving symmetric ill-conditioned generalized eigenproblems [156]. In our code, we used a Lapack-style implementation of this algorithm from [157]. This algorithm works for real symmetric matrices with S positive definite with respect to some user defined tolerance $0 < \epsilon \ll 1$. Essentially this algorithm finds the eigenvalues of H and S which are zero with respect to ϵ and discards them before diagonalizing the remaining blocks of these matrices to solve the general-



(a) (b)

Figure 7.2: (a) Exact ground state of a two well system on 10 qubits ($n_1 = 4$, $n_2 = 6$) as a function of h_1 and h_2 for s near the minimum gap. The well on the left (right) has width 0 (1) and we see that at the point of minimum gap the wave function can rapidly switch between wells. (b) Energy landscape for fixed $h_2 = 0$ (cut with minimum distance between wells). The minimum eigenvalue gap occurs at an energy such that the parts of the wave functions external to the wells have amplitudes on bit strings for which the potential energy is greater than the eigenenergy of the state. Here potential energy is shifted by $-(1-s)I$ so that the relation to tunneling in the sense that the there is wave function support on bitstrings such that the energy is less than the potential energy is readily apparent.

ized eigenproblem.

We expect the approximation to be good when the wave functions are "tightly bound". Therefore we propose an approximate upper bound on the error based on first order pertubations of the diagonal elements of $H^{(TB)}$: $\left[\sum_i |\langle\psi_i|\bar{H}^{(i)}|\psi_i\rangle|^2\right]^{1/2}$ where $\bar{H}^{(i)}$ is the Hamiltonian excluding the potential from the i^{th} well. We test this upper bound on a set of 2450 random runs on between 4 and 10 qubits, between 2 and 10 wells, with depths between -1.0 and -5.99 (arbitrary units) and width 0. Our intuition is confirmed and we see that the relative error in the energy gap for zeroth order TB at 17 evenly spaced s values is with high probability upper bounded by our

proposed error estimate as shown in Fig. 7.3. We estimate the relative error as

$$\Big[\sum_i |\langle\psi_i|\bar{H}^{(i)}|\psi_i\rangle|^2\Big]^{1/2}/\tilde{\gamma}^{(TB)}$$

where

$$\tilde{\gamma}^{(TB)} := E_1^{(TB)} - E_0^{(TB)} - \sqrt{2}\Big[\sum_i |\langle\psi_i|\bar{H}^{(i)}|\psi_i\rangle|^2\Big]^{1/2}$$

is the minimal possible gap within the absolute error estimates. Figure 7.4 shows the empirical CDF from the data, further indicating that this is with high probability a good upper bound. In particular problems, some of the points violating this upper bound can be eliminated as unstable eigenvalues using a case by case choice of ϵ in the Fix-Heiberger algorithm as for these runs ϵ was fixed at 0.1 in all cases.

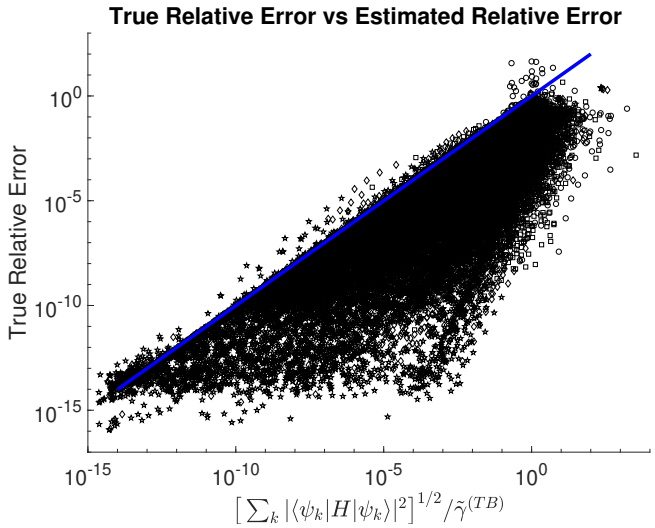


Figure 7.3: True error (tight-binding compared to direct diagonalization) versus error estimate for eigenvalue gap in zeroth order TB for a set of 2450 random runs on between 4 and 10 qubits, between 2 and 10 wells, with depths between -1.0 and -5.99 and width 0 at 17 evenly spaced s values. If $\tilde{\gamma}^{(TB)} < 1$ then tight-binding cannot distinguish between the ground state and first excited state and the corresponding point is not plotted. Circles indicate $s \in [0.15, 0.30]$, squares $s \in [0.15, 0.30]$, diamonds $s \in [0.35, 0.70]$, and stars $s \in [0.70, 0.95]$. Note for the low s points violating the upper bound estimate in the top right of the figure, with a different choice of ϵ the Fix-Heiberger algorithm could eliminate these points as unstable eigenvalues.

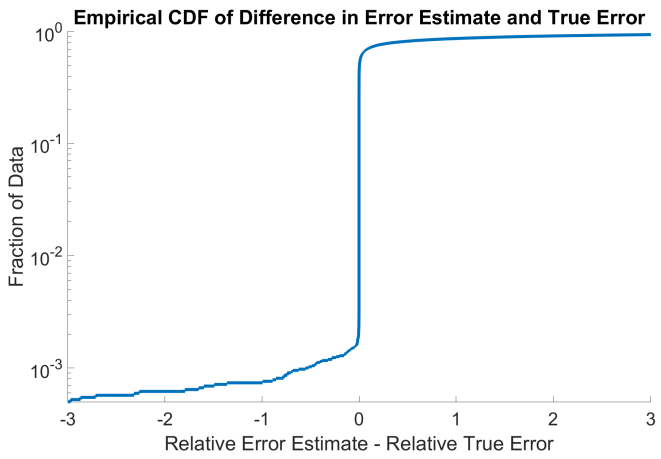


Figure 7.4: Error estimate minus true error estimated CDF for zeroth order TB for the 2450 random runs from Fig 7.3.

Finally note that the scaling of the zeroth order tight-binding algorithm is $O(T(K^2n^6 + K^3))$ where T is the number of s values for which the eigenspectrum is computed. The dominant factor comes from computing the $O(K^2)$ matrix elements of the tight-binding Hamiltonian, each of which requires diagonalizing a dimension $O(n^2)$ square matrix. Our codes for the exact solvers for $K \leq 3$ and for tight-binding along with input files for all the examples next presented are located on GitHub [158].

7.4 Examples and Applications

We now lay out a series of examples and possible applications aimed at revealing the power (and limits) of these algorithms. Example 1 gives the application of unstructured search with a prior guess, each represented as a single well. We solve this problem using both the exact two well algorithm and the tight-binding algorithm for up to a fairly large ($n = 70$) number of qubits, demonstrating both of their effectiveness. Example 2 draws attention to how the tight-binding framework generates an effective graph (Hamiltonian) describing the problem, but it looks at this

correspondence in reverse, by considering simulating an Ising model ground state adiabatically, using tight-binding as a tool for mapping the two systems to one another. The particular example presented has no practical implementation and serves merely as an interesting example of tight-binding with multiple wells, but we suggest generalizations beyond the scope of this [chapter] that could prove useful. Example 3 demonstrates the effectiveness of tight-binding for a large number of wells ($n = 10$, $K = 50$). Finally, Example 4 highlights a situation where first order tight-binding is needed.

Example 1: Unstructured Search with Priors

Unsurprisingly, as AQC is equivalent to the standard circuit model of quantum computation with polynomial overhead [13] an AQC version of Grover's algorithm for unstructured search demonstrates an equivalent speedup [159]. The speedup requires having an optimized adiabatic schedule where the adiabatic condition is obeyed locally, speeding up when the eigenvalue gap between the ground state and first excited state is large and slowing down when the gap is small. If one ran the adiabatic algorithm purely at the rate prescribed by the minimum gap, the cost would be equivalent to classical unstructured search. Such an optimized schedule depends on having knowledge of the gap structure at $O(\log(1/\gamma))$ points throughout the evolution at a precision also $O(\log(1/\gamma))$ [160], which is potentially a challenge for general problems. Questions have also been raised as to how robust this highly optimized evolution is to noise [161], although static noise and time-dependent noise small relative to the gap can be handled [160].

Here we set aside these questions and the process of integrating over local gaps and simply investigate the minimum gap while searching for a single marked item as in Grover search

but with a prior guess to the location of the marked item. If these estimates of the location of the marked item are good, then we expect the eigenvalue gap to be correspondingly wider, thus making it easier to find the marked item. This situation fits neatly into our model. We simply start at $s = 0$ with a potential well representing our guess for the marked item and evolve to the potential giving the marked state. In principle, this "guess" well could be tuned to have a functional form such that the initial wave function corresponds to a particular probability distribution corresponding precisely to our confidence in our initial guess. For simplicity, however, we shall treat our guess simply as a constant potential Hamming ball of some radius. This setup is given by the Hamiltonian

$$H(s) = (1-s)\Big[-\frac{1}{n}\sum_j X_j + V_p\Theta\Big(r_p - \sum_j X^{\bar{k}}\bar{Z}_j X^{\bar{k}}\Big)\Big] - s|m\rangle\langle m| \tag{7.9}$$

where $V_p < 0$, r_p is the depth and Hamming radius of the prior well p, respectively, and $|m\rangle$ is the marked item.

With a single bit string prior plus the marked item we can use the exact two well solver and find exact solutions, so for our example we compare both this exact solution and the tight-binding solution. For simplicity, we consider point-like wells ($r_p = 0$) so we only have to worry about the $\sigma = 0$ subspace and zeroth order tight-binding. Figure 7.5 shows how the eigenvalue gap scales with distance of one such prior from the marked item. Both the exact (two well subspace) and tight-binding solutions are shown, as further evidence of the accuracy of tight-binding for larger numbers of qubits. We see from these results that if the prior is close to the true marked item we do see an increase in the gap relative to standard Grover with no priors. However, if the prior is far from the true marked item, the gap shrinks relative to Grover with no prior.

We should expect that randomly guessing will not provide any advantage (i.e. the prior must actually come from some prior knowledge about the problem). As shown in Figure 7.6 which plots the gap times the probability of randomly guessing a prior at the Hamming distance R versus n that with random guessing we recover a $O(2^{-n/2})$ scaling which we expect for unstructured search.

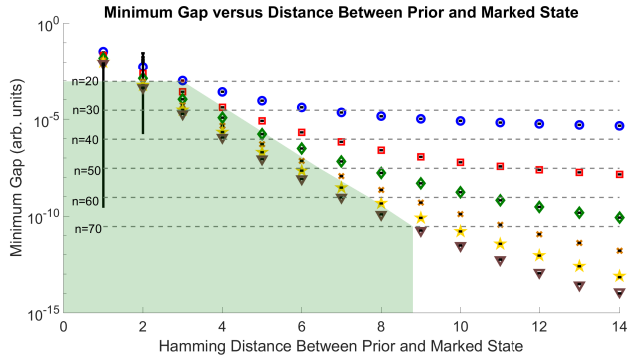


Figure 7.5: Plot of eigenvalue gap versus distance from marked item for a single width 0, depth -1 marked item for n=20 (circles), 30 (squares), 40 (diamonds), 50 (cross), 60 (five point star), 70 (triangle). Error bars overlaid on these points indicate the tight-binding results. Horizontal dotted lines show the gap for no prior and the shaded region indicates the distances where the prior offers an improvement in the gap over standard Grover.

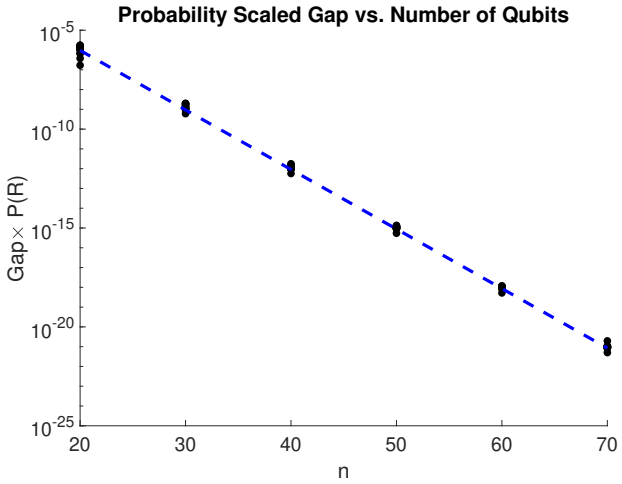


Figure 7.6: Plot of probability scaled minimum eigenvalue gap vs. number of qubits. The dashed line demonstrates that we recover $O(2^{-n/2})$ scaling if the priors are randomly guessed.

In Figure 7.7 we show a particular energy spectrum ($n = 30$, $r_p = 5$), demonstrating

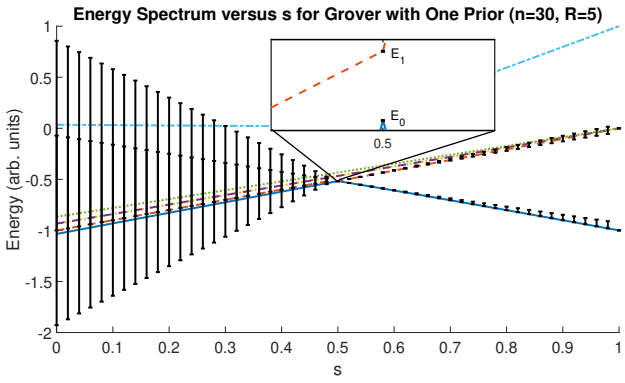


Figure 7.7: Plot of a particular eigenvalue spectrum vs. s for the Grover with priors example. Lines indicate exact eigenvalues whereas error bars give the lowest two energy levels as determined by tight-binding. We see that tight-binding does an excellent job at finding the minimum gap in this problem with error bars on the order of the machine precision.

the size of the error bars in tight-binding as a function of s. We see that tight-binding does an excellent job of identifying the minimum gap with small error (at $s = 0.5$). Also note from the exact solution that a number of other energy eigenvalues are all clustered around this point (many of which are highly degenerate). Such a feature could make these problems difficult for standard power iteration type methods for finding principal eigenvalues.

The next natural cases to consider are prior probability distributions favoring multiple bit strings and/or priors with $r_p > 0$. However, we find that for this particular example tight-binding is a relatively poor approximation due to strong overlaps, as indicated by our error estimate. While these issues could be ameliorated by using deeper wells, to explore cases with multiple wells or wells with $r_p > 0$ we turn to other examples.

Example 2: Approximate Ground State of a System of Strongly Interacting Spins

A key feature of tight-binding is that it generates a greatly reduced effective graph (representing a Hamiltonian) with edge weights determined by the tunneling matrix elements between

wells. One could imagine being given a real world Hamiltonian and then trying to identify these key features and approximately modeling it with tight-binding. This is a hard problem, however, so we leave this aside and work with a simpler example designed to demonstrate this key point. In particular, we work in reverse: starting with a small collection of strongly interacting spins we use the tight-binding framework to simulate the ground state of this Hamiltonian using the adiabatic Hamiltonian with potential wells considered in this [chapter]. As an explicit case consider an Ising model on L quantum spins of the form

$$H_I = -\sum_{i<j}^{L} J_{ij} Z_i Z_j - \sum_{i=1}^{L} B_i X_i - \alpha \mathbb{1} \tag{7.10}$$

where the last term is simply a potential shift chosen to make the diagonal terms strictly negative. If we can choose a collection of 2^L wells on some set of n qubits with tight-binding Hamiltonian $H^{(\mathrm{TB})}(s^*)$ and overlap matrix $S(s^*)$ such that $[S^{-1}H^{(TB)}](s^*) \approx H_I$ then by evolving adiabatically from s to s^* we can approximately sample the ground state of H_I.

Such a framework may have practical applicability, although understanding the extent of generality to a broad class of underlying Hamiltonians would require a thorough investigation of which interactions can be modeled using potential wells on a hypercube lattice. Namely we are limited to tunneling between wells, which enforces a fairly restrictive geometric dependence on the various matrix elements of the tight-binding Hamiltonian. There is still a number of available degrees of freedom, however, including the number of qubits in the host system, the shape and structure of the tight-binding wells, and global potential terms that could assist in tunneling between certain wells. Therefore, at least for certain problems, we conjecture that such a procedure could be practically useful in two cases: (1) if one has a large number of qubits that

can be evolved adiabatically but a limited set of available controls (namely individual control in the computational basis but only a global X term, as in the DWave machine [162]), it would allow one to simulate more complicated interactions; (2) depending on how robust the tight-binding framework is to noise in the underlying system of qubits this scheme could serve as a method for fault-tolerant simulation on a large number of noisy qubits. The Davis-Kahan theorem suggests that our framework is indeed robust to moderate noise [163]. We note that the Hamming ball wells used here are not easily generated on current hardware as they require n-body interactions. However, such potential wells are a simplification for ease of testing our code. One could make $K = 2^L$ wells with an K-local potential using a degree 2^K polynomial in the distances to each of the desired wells.

Here we leave these more general questions open and raise them purely as a possible motivation for a particular example which demonstrates the effectiveness of tight-binding. For simplicity assume that $B_i \ll J_{ij} \,\forall\, i, j$, which means that $S \approx \mathbb{1}$ and our mapping is simply $H^{(\mathrm{TB})}(s^*) = H_I$ for some s^*. Then consider a system of 3 spins with interactions $J_{ij} = 1 \,\forall\, i, j$ and $B_i = 0.015 \,\forall\, i$ (arbitrary units). Using $\alpha = 30$ and $s^* = 0.95$, this Hamiltonian can be mapped to a tight-binding Hamiltonian of 8 hyperspherical wells (2 with $s^* V_p = -33.5$ and 6 with $s^* V_p = -29.5$, $r = 0$) on 10 qubits where we consider terms in the tight-binding Hamiltonian smaller than $O(10^{-5})$ as essentially zero. In Figure 7.8, we compare the exact (full diagonalization) adiabatic ground state probability distribution at s^* as determined via the tight-binding mapping and the exact Ising model ground state probability distribution. The adiabatic probability is determined by the normalized probability of sampling within the well corresponding to a given Ising model basis state. We see that the mapping is a good one: the adiabatic version samples with probabilities within $O(10^{-5})$ of the true probability.

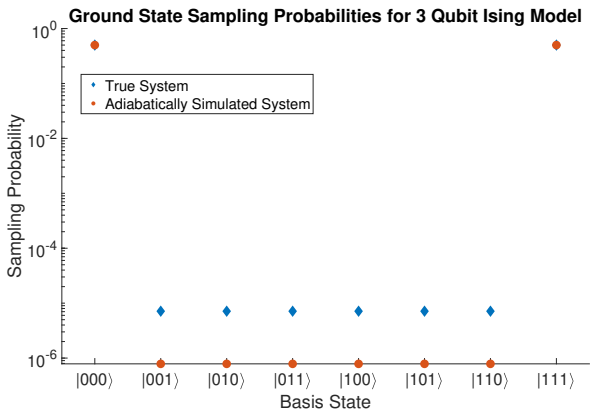


Figure 7.8: Comparison of exact adiabatically simulated ground state probability distribution to exact true ground state for the 3 qubit Ising model with $J_{ij} = 1 \forall i, j$ and $B_i = 0.015 \forall i$. The adiabatic Hamiltonian is constructed via a mapping between tight-binding with 8 wells and the 3 qubit Ising model Hamiltonian.

Example 3: Large Set of Wells

As a final test for a large set of wells, we compared exact diagonalization and tight-binding for a set of 50 wells with depths between -1 and -1.13 with $r = 0$ on 10 qubits, as depicted in Fig. 7.9. We found that tight-binding was effective in this case, however, the point in the evolution in s where the wave function is tightly bound enough for the errors to be small enough to identify the gap is later than in problems with fewer wells. This suggests that for large sets of wells tight-binding is only useful at identifying the minimum gap when the minimum gap occurs late in the evolution, as in the example here, where the gap is at $s = 1$. This brings attention to one key limitation of tight-binding: it does not let us know if the minimum gap is prior to the point in the schedule where the tight-binding errors decrease to the point that we can resolve the gap. Figure 7.9 also depicts the results for tight-binding without the use of the Fix-Heiberger algorithm. We see that due to numerical instability the eigenvalues diverge and Fix-Heiberger allows us to be control this divergence and to automatically eliminate unstable eigenvalues with

appropriate choice of ϵ.

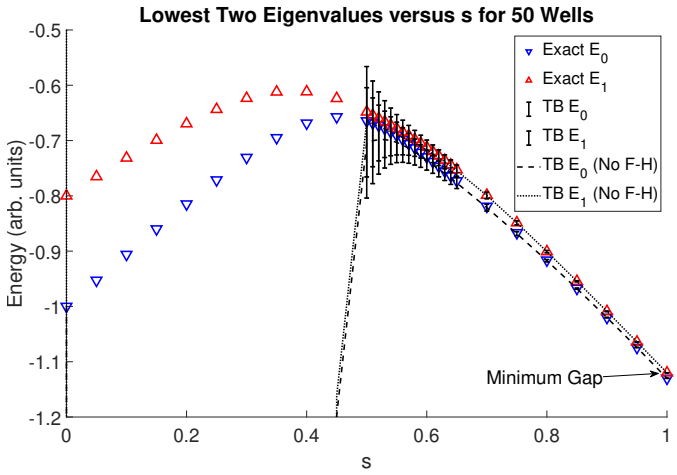


Figure 7.9: Plot of lowest two eigenvalues versus s for Example 3. We see at large s tight-binding is effective, however, a large number of wells does require that the minimum gap occur at larger s to capture it effectively with tight-binding. Lines indicate tight-binding without Fix-Heiberger indicating its usefulness to automatically remove unstable eigenvalues.

Example 4: First Order Tight-binding

Problems involving wells with $r > 0$ possibly requires first order tight-binding in order to capture internal structure of the wider wells and to check whether or not these states are relevant. Such problems are not quite as easily automated as zeroth order tight-binding and tend to require more fine tuning of the ϵ parameter in the Fix-Heiberger reduction algorithm, as well as comparison between zeroth and first order solutions to correctly ascertain the appropriate gap. Here we consider an example designed to demonstrate the need for first order tight-binding. In particular, we consider two Hamming spherical potential wells

$$H(s) = -\frac{1-s}{n}\sum_j X_j - s\sum_{i\in\{a,b\}} V_i \Theta\Big(r_i - \sum_j X^{\bar{i}}\bar{Z}_j X^{\bar{i}}\Big) \tag{7.11}$$

s	E_1 (Exact)	E_1 (TB0)	E_1 (TB1)
0.20	-1.05367	-1.04988	-1.05350
0.30	-1.52649	-1.50398	-1.52649
0.40	-2.01448	-1.73993	-2.01448
0.50	-2.50802	-2.46024	-2.50802
0.60	-3.00427	-2.94545	-3.00427
0.70	-3.50206	-3.43262	-3.50206
0.80	-4.00080	-3.92102	-4.00080
0.90	-4.50018	-4.41022	-4.50018

Table 7.1: First excited state energies for Example 4. First order tight-binding is needed to get the correct solution in this problem.

with $V_a = -5$, $V_b = -4.9$, $r_a = 1$, $r_b = 0$, $|\bar{a}| - |\bar{b}| = 6$, and $n = 10$. In this case, due to well a being energetically favored for all s the full first excited state wave function is constructed using the first excited state of well a. Zeroth order tight-binding doesn't capture this behavior. This is shown in Table 7.1. Note that in this calculation we did not use the first excited state of the width 0 well, as this state is unnecessary and not tightly bound and thus introduces avoidable error into the approximation.

In most problems with Hamming symmetric wells, this is not an issue and even with wider wells one can get the correct first excited state purely with zeroth order tight-binding, but it is important to have these tools available to have a complete understanding of the energy spectrum.

7.5 Conclusion

We provide a set of algorithms for efficiently analyzing the performance of AQC for problems that can be expressed in terms of a set of individually Hamming symmetric wells. These problems, while still highly symmetric, are more complex than the well studied Hamming symmetric example and should provide a new testbed for study of AQC. In particular, the tight-binding approach for studying this model highlights the effects of tunneling, which must be the

source of the quantum speedup if one is afforded by AQC with stoquastic Hamiltonians We also provide several examples demonstrating the effectiveness of tight-binding as a tool for studying this toy model.

Chapter 8: Effective Gaps Are Not Effective: Quasipolynomial Classical Simulation of Obstructed Stoquastic Hamiltonians

8.1 Introduction

The power of adiabatic quantum computation (AQC) with stoquastic Hamiltonians (StoqAQC), formally introduced in [164], remains difficult to understand. While we know AQC with general Hamiltonians is universal [13], one might reasonably expect that stoquastic Hamiltonians – those that have a known representation with real, non-positive off-diagonal matrix elements – are more efficiently classically simulable. In this work, we provide a classical, quasipolynomially-efficient algorithm for sampling from eigenstates of k-local, stoquastic Hamiltonians which are otherwise widely believed to "obstruct" classical simulation algorithms [14, 18, 149, 165].

8.2 Background

AQC interpolates over a one-parameter family of Hamiltonians $H(s)$ to produce a quantum state close to the ground state of $H(s_f)$. The computational cost of this process is usually bounded by an adiabatic theorem scaling inversely in the minimal eigenvalue gap $\gamma_{\min} = \min_s \gamma\left(H(s)\right)$ between the two lowest eigenvalues $\gamma(H(s)) = \lambda_1(H(s)) - \lambda_0(H(s))$ of $H(s)$ [147, 148]. Thus,

an efficient simulation algorithm must scale at most sub-exponentially with $\gamma_{\min}^{-1}$.

While most researchers do not expect StoqAQC to be capable of universal quantum computing (as evidenced by i.e. [164, 166–169]), the primary techniques for simulating these processes rely on Monte Carlo (MC) methods, and tend to focus on estimating certain properties of thermal states. For a full simulation of a StoqAQC process, however, we would like to be able to reproduce ground-state measurement statistics at zero temperature. When comparing AQC to particular MC-based algorithms for this task, there exist a number of "obstructions" that do yield exponential separations [14, 18, 149, 165]. In fact, results published shortly after the first version of this manuscript exploit these obstructions to prove a superpolynomial oracle separation between classical algorithms and StoqAQC with non-k-local, but sparse Hamiltonians [18].

Many classical methods are limited to classically sampling from a statistical distribution proportionate to probability amplitudes of a quantum state ϕ, rather than probabilities of its measurement outcomes. One can immediately see a fundamental obstruction to this approach in its most abstract—the divergence between $\|\phi\|_1$ and $\|\phi\|_2$ can greatly impact sampling statistics. For our purposes, let ϕ be the ground state of $H(s)$ and let $\phi(i) \equiv \langle i|\phi\rangle$. Now, suppose that there exists some m such that $|\phi(m)|^2/\|\phi\|_2^2 = \Omega(1/\mathsf{poly}(n))$, where n is the number of qubits. An efficient quantum process capable of producing the state ϕ will take only $O\left(\mathsf{poly}(n)\right)$ measurements of ϕ in the basis $\{|i\rangle\} \ni |m\rangle$ to reliably return m. Alternatively, suppose one can produce samples of a random variable X taking values in $\{i\}$ with probability mass function (PMF) $\Pr\left(X = i\right) = |\phi(i)|/\|\phi\|_1$. We have that $\frac{|\phi(m)|}{\|\phi\|_1} = \frac{|\phi(m)|}{\|\phi\|_2}\frac{\|\phi\|_2}{\|\phi\|_1} \geq 2^{-n/2}\frac{|\phi(m)|}{\|\phi\|_2}$. When this inequality is nearly achieved, one requires exponentially many samples of X before one expects to return m. (For an explicit example, see [149, Example 0].) Thus, even an efficient classical process for perfectly producing samples of X may be exponentially slower than its quantum

counterpart. Even in an idealized case, it seems one might require a classical process capable of directly sampling from the PMF $\Pr(X = i) = |\phi(i)|^2/\|\phi\|_2$, which can be rather difficult to obtain.

8.3 Approach

Current examples of k-local Hamiltonians where the inequality above is nearly achieved for the ground state of $H(s)$ rely on symmetries maintained by $H(s)$, constraining adiabatic dynamics to a polynomially-sized effective subspaces [150–154, 170]. This raises a natural question: is it possible to efficiently classically reproduce the quantum statistics of the ground state of $H(s)$ without knowing its (near) symmetries *a priori*? In this [chapter], we answer in the affirmative, thereby providing an algorithm capable of simulating eigenstates previously deemed "obstructed" [14, 149, 165, 170].

In particular, we consider symmetries such that the ground state ϕ of $H(s)$ is preserved under permutations π of computational basis states $\{|i\rangle\}$, such that $|\phi(m)| = |\phi(\pi(m))|$ for all $m \in \{i\}$. As elaborated later, this set of symmetries can be formally described by the automorphism group of an appropriate graph. For simplicity, in this [chapter], we restrict attention to symmetries generated by terms with the same number of interacting qubits. This includes all k-local obstructions previously discovered, but more general constructions are possible.[1] Here, we introduce a classical algorithm that discovers and leverages such symmetries in quasi-polynomial time. The algorithm is upper bounded in its complexity by the greater of graph isomorphism (GI) on graphs with $\mathsf{poly}(n)$ vertices and $\mathsf{poly}(|S|)$, where S is an irreducible set of equivalence classes between computational basis states. Since GI is solvable in quasi-polynomial time [171, 172],

[1] See [Appendix F] for further discussion.

our algorithm scales quasi-polynomially in n whenever $|S|$ is quasi-polynomial in n. This rules out exponential separations between AQC and classical algorithms for k-local, stoquastic Hamiltonians with large automorphism groups.

Our graph constructions are general, however our restriction to stoquastic Hamiltonians simplifies our study of symmetries. The Perron-Frobenius theorem guarantees that the ground state ϕ of a stoquastic Hamiltonian can be written with all non-negative amplitudes in the computational basis (e.g. [173]).[2] Automorphisms π can be expressed as tensor products of Pauli-X operators, and the non-negativity of the ground state means that we can ignore sign changes and only study the case that $\phi(m) = \phi(\pi(m))$. (Note that there exists a bitstring $b \in \{0,1\}^n$ such that $\bigotimes_i X^{b_i}|m\rangle = |l\rangle$ for all $l, m \in \{0,1\}^n$.) That is, the lack of sign change means that we do not need to consider the possibility of first conjugating $H(s)$ by unitaries that map X terms to Y terms and vice versa. Although we exploit this simplification, we anticipate these constructions can be generalized to study properties of general k-local Hamiltonians.

8.4 Algebraic Graph Theory

The primary contribution of this [chapter] is the formal construction of bijective mappings from H to a pair of graphs, which allows us to reduce the problem of simulation to that of GI. The two mappings are bijective, in the sense that the Hamiltonian can be recovered from each graph. The first, $H \mapsto \Gamma$, takes H to an exponentially-sized, undirected graph Γ with spectral properties consistent with H itself. The second, $H \mapsto G$, maps H to a vertex-colored, directed graph $G = (V_G, E_G)$ which incorporates all relevant automorphisms of Γ. In a sense, the latter

[2] Actually, discrete nodal domain theorems are generally more useful here, but they too depend upon Perron-Frobenius theorem [174].

is the compact, graph representation of Γ in the same way that H, written in terms of Pauli matrices, is the compact representation of H as a matrix. G can be used to efficiently reconstruct and determine equivalent vertices of Γ via GI and, thus, determine the effective subspace of H. We first construct Γ, then present the algorithm in detail in which we treat the construction of G as a black box. Then, we explicitly provide a construction of G. We refer the reader to [Appendix F] for a complete, minimal example.

8.5 Mapping I: H to Γ

We consider the weighted graph representation, $\Gamma = (V_\Gamma, E_\Gamma, w_\Gamma)$, of a stoquastic Hamiltonian where $V_\Gamma = \left\{X_b = \bigotimes_i X^{b_i}\right\}_{b\in\{0,1\}^n} \cup \{\infty\}$ [175]. That is, we label each vertex usually associated with computational basis state $|b\rangle$ by X_b, as $X_b|r\rangle = |b \oplus r\rangle$ for any $r \in \{0,1\}^n$, and we seek an r-independent construction. (We also define $Y_b = \bigotimes_i Y^{b_i}$ and $Z_b = \bigotimes_i Z^{b_i}$.) We assume that we are presented with a k-local stoquastic Hamiltonian $H \in \mathbb{R}^{|V_\Gamma^*|\times|V_\Gamma^*|}$, where $V_\Gamma^* = V_\Gamma \setminus \{\infty\}$. Specifically,

$$H = -\sum_{\|b\|_H\leq k} \alpha_b X_b - \sum_{\substack{\|b\|_H\leq k \\ \|b\|_H\in 2\mathbb{Z}}} \beta_b Y_b + \sum_{\|b\|_H\leq k} \kappa_b Z_b, \tag{8.1}$$

where $\|b\|_H$ is the Hamming weight of the bit string b and $|\beta_b| \leq \alpha_b \; \forall \; b$.[3] From H, we identify the set of *edge generators* $K = \{X_b \mid \alpha_b \neq 0\}$. Let $H_X = \sum_{\|b\|_H\leq k} \alpha_b X_b$ and $H_Y = \sum_{\substack{\|b\|_H\leq k \\ \|b\|_H\in 2\mathbb{Z}}} \beta_b Y_b$.

[3] For simplicity and to avoid too much notation, we consider only Hamiltonians that can be written as these combinations. However our construction will generalize to situations with mixed terms (i.e. XYY), by introducing a new gadget that connects together an X- and YY-gadget.

Now,

$$\begin{aligned}\langle b'|(H_X + H_Y)X_b|b'\rangle &= \alpha_b + \mathrm{i}^{-\|b\|_H}\beta_b\langle b'|Z_b|b'\rangle \\ &= \alpha_b + \mathrm{i}^{-\|b\|_H}(-1)^{b\cdot b'}\beta_b \\ &= \alpha_b + \mathrm{i}^{2b\cdot b' - \|b\|_H}\beta_b. \end{aligned} \tag{8.2}$$

We let $w(u,v) = w(v,u)$ and define edge weights,

$$w\left(X_{b'}, v\right) = \begin{cases} \alpha_b + \mathrm{i}^{2b\cdot b' - \|b\|_H}\beta_b & \text{if } v = X_{b'\oplus b} \\ \\ \displaystyle\sum_{\|b\|_H \leq k} (-1)^{b\cdot b'}\kappa_b & v = \infty, \end{cases} \tag{8.3}$$

and edges $E_\Gamma = \{\{u,v\} \mid w(u,v) \neq 0\}$. The eigenvectors of H satisfy

$$(w(u,\infty) - \lambda_i)\phi_i(u) = \sum_{v \in V_\Gamma^*} w(u,v)\phi_i(v), \tag{8.4}$$

where $u \in V_\Gamma^*$, $\phi(\infty) = 0$, and (ϕ_i, λ_i) is the ith eigenvector-eigenvalue pair. In order to identify symmetric subspaces of H, we consider identifying all vertices of Γ that are equivalent under an edge-weight preserving automorphism $f : V_\Gamma \longrightarrow V_\Gamma$ of Γ. We call the set of all such automorphisms $\mathrm{Aut}(\Gamma)$. Now, we sum over equivalence classes $[\![u]\!] = \{f(u)\}_{f\in\mathrm{Aut}(\Gamma)}$:

$$\sum_{u' \in [\![u]\!]} (w(u',\infty) - \lambda_i)\phi_i(u') = \sum_{u' \in [\![u]\!]} \sum_v w(u',v)\phi_i(v),$$

or

$$(w(u,\infty) - \lambda_i)\,\phi_i(u) = \sum_{[\![v]\!]} \omega_{u[\![v]\!]}\phi_i(v), \tag{8.5}$$

where $\omega_{u[\![v]\!]} = \sum_{v\in[\![v]\!]} w(u,v)$.

This defines our effective Hamiltonian $H' : [\![V_\Gamma^*]\!] \times [\![V_\Gamma^*]\!] \longrightarrow \mathbb{R}^+$ on the space of effective vertices $[\![V_\Gamma^*]\!] = \{[\![u]\!]\}_{u\in V_\Gamma^*}$:

$$H'([\![u]\!],[\![v]\!]) = \begin{cases} w(u,\infty) & \text{if } [\![u]\!] = [\![v]\!] \\ -\omega_{u[\![v]\!]} & \text{otherwise.} \end{cases} \tag{8.6}$$

Note that $\omega_{v[\![u]\!]} \neq \omega_{u[\![v]\!]}$, but rather $|[\![v]\!]|\omega_{v[\![u]\!]} = |[\![u]\!]|\omega_{u[\![v]\!]}$. By Eq. (8.5), the right eigenvector of H' corresponding to eigenvalue λ_0 is proportional to the eigenvector of H corresponding to eigenvalue λ_0.

8.6 The Algorithm

Assume that we can map our Hamiltonian to a graph Γ as described above. Our goal is to find an effective graph Γ' with vertex set $[\![V_\Gamma^*]\!] \cup \{\infty\}$, whose ground state corresponds to that of Γ.

For clarity, we break the classical algorithm into two parts: (1) FINDEFFECTIVEVERTICES, which recursively searches Γ to return V' such that $u \in V' \iff V' \cap [\![u]\!] = \{u\}$; and (2) FINDEFFECTIVEGRAPH, which takes as input V' and returns Γ'. Both routines assume the existence of an ancillary algorithm FINDREPRESENTATIVE$(u, V') = v \in V' \cap [\![u]\!]$, whose existence we will later justify. For now, we treat it as an oracle with runtime quasi-polynomial in n,

$O\left(|V'|\,\mathrm{QP}(n)\right)$, where $\mathrm{QP}(n)$ matches the runtime of the best-known GI algorithm [171, 172].

Algorithm 1 Find Effective Vertices

```
1:  function FindEffectiveVertices(u, V')
2:      if u = ∅ then
3:          u ← Random(V*_Γ)
4:          Add u to V'
5:      end if
6:      for v ∈ N(u) do
7:          if FindRepresentative(v, V') = ∅ then
8:              Add v to V'
9:              V' ← FindEffectiveVertices(v, V')
10:         end if
11:     end for
12:     return V'
13: end function
```

Algorithm 1 returns a set of vertices such that each vertex is distinct and the entire routine, including the FindRepresentative subroutine, takes time $O\left(\Delta(\Gamma)|V'|^2\,\mathrm{QP}(n)\right)$ where $\Delta(\Gamma)$ is the maximum degree of Γ. Since V' includes precisely one representative of each equivalence class in the connected component of Γ, the following routine generates the effective graph Γ'.

Algorithm 2 Find Effective Graph

```
1:  function FindEffectiveGraph(Γ)
2:      V' ← FindEffectiveVertices(∅)
3:      Ω_uv ← 0 for all u, v ∈ V'
4:      for u ∈ V' do
5:          for v ∈ N(u) do
6:              v ← FindRepresentative(v, V')
7:              Ω_uv ← Ω_uv + w(u, v)
8:          end for
9:      end for
10:     return (V', Ω)
11: end function
```

The primary loop of FindEffectiveGraph (Line 4) takes time $O\left(\Delta(\Gamma)|V'|^2\,\mathrm{QP}(n)\right)$, and therefore the total time to obtain the effective graph is also $O\left(\Delta(\Gamma)|V'|^2\,\mathrm{QP}(n)\right)$.

We can now obtain H' and sample from its eigenstates. For $u, v \in V'$, $\Omega_{uv} = \omega_{u[\![v]\!]} = \sum_{v_0 \in [\![v]\!]} w(u, v_0)$. Thus, Eq. (8.6) is well-defined and the operator H' known, even if each entire equivalence class $[\![u]\!]$ is not.

We know that existing methods, such as the power iteration method, can produce the ground state ϕ' of H' with error ϵ in time $O\left(\log(\epsilon^{-1})/\log(\lambda_1/\lambda_0)\right)$. Therefore, we can sample the ground state of the full Hamiltonian H in time $O\left(\log(\epsilon^{-1})/\log(\lambda_1/\lambda_0) + \Delta(\Gamma)|V'|^2\,\mathrm{QP}(n)\right)$.

We note that we cannot simply normalize ϕ' and expect to obtain appropriate statistics; rather, each $u \in V' \cap [\![u]\!]$ represents a sample of the class itself. Thus, we need to sample $[\![u]\!]$ with probability $|[\![u]\!]|\phi(u)^2$, where ϕ is the appropriately normalized ground state of H. By Eq. (8.5), H' has a ground state ϕ' that preserves relative amplitudes $\frac{\phi(u)}{\phi(v)} = \frac{\phi'(u)}{\phi'(v)}$ for all $u, v \in V'$.

Now, we use ϕ' and $|[\![u]\!]|$ to sample $u \in [\![u]\!] \cap V'$ with probabilities according to ϕ,

$$\Pr\left([\![u]\!]\right) = |[\![u]\!]|\phi(u)^2 = \frac{|[\![u]\!]|\phi'([\![u]\!])^2}{\sum_{v \in V'}|[\![v]\!]|\phi'([\![v]\!])^2}. \tag{8.7}$$

Note that for $\omega_{v[\![u]\!]} \neq 0$, $\frac{|[\![v]\!]|}{|[\![u]\!]|} = \frac{\omega_{u[\![v]\!]}}{\omega_{v[\![u]\!]}}$. Therefore,

$$\frac{|[\![u]\!]|}{\sum_{v \in V'}|[\![v]\!]|} = \left(\sum_{v \in V'} \prod_{e \in P(u,v)} \frac{\omega_{e_0[\![e_1]\!]}}{\omega_{e_1[\![e_0]\!]}}\right)^{-1}, \tag{8.8}$$

where $P(u, v) \subseteq E_{G'}$ is any directed path connecting $u, v \in V'$. Up to a factor constant for all u, v, Eq. (8.8) determines $|[\![u]\!]|$ and, thus, fully determines Eq. (8.7).

Repeating this process initializes a new seed in Algorithm 1 Line 3, and we return each member of $[\![u]\!]$ with equal probability. Furthermore, the random seed guarantees that a sample from a connected set of vertices V_C of Γ is returned with probability $|V_C|/|V_\Gamma|$, as expected.

8.7 Mapping II: H to G

Now, we explicitly give an implementation of FINDREPRESENTATIVE. It is helpful to keep in mind that while this construction is unavoidably definition-heavy, the construction naturally reduces our problem to GI. In fact, our approach is somewhat similar to Luks' reduction of graph automorphism to GI [176] or Crawford's formalism of symmetries in clausal theories [177], applied to the study of $\mathrm{Aut}(\Gamma)$. Inspired by the latter, we build what we abusively call a *clausal theory graph* G and our goal is to define an invertible map M such that $M(\Gamma) = M_0[V_\Gamma] \cup M_1[E_\Gamma] = G$. We do so by introducing *gadgets*, smaller graphs that allow us to separately map each $v \in V_\Gamma$ and $e \in E_\Gamma$ to specific vertex-colored, directed graphs. The union of these gadgets forms G. We will introduce a number of different types of vertices, where each type is assigned a unique color represented by a superscript (E.g. $\ell^{(a)}$). In the following $\{a, b, c, d\}$ represent fixed, unique colors and $\{x_b, y_b, z_b\}$ represent a distinct set of variable colors which are assigned based on the particular Hamiltonian. Two vertices are identical only if they both have the same color and label. Furthermore, for simplicity, we will abusively write $\{u, v\}$ for an undirected edge and (u, v) for a directed edge. (Thus, $\{u, v\} \in E$ can be read as $\{(u, v), (v, u)\} \subset E$.)

First, we define a set of *literals* $L = \left\{Z_{\boldsymbol{i}}^{(a)}\right\}_{i=0}^{n-1}$ and their negations $-L = \left\{-Z_{\boldsymbol{i}}^{(a)}\right\}_{i=0}^{n-1}$, where $\boldsymbol{i} = (\delta_{ij})_{j=0}^{n-1}$. We label each vertex $X_b \in V_\Gamma^*$ by a set of literals $A(X_b)^{(b)} = \left\{(-1)^{b_i} Z_{\boldsymbol{i}}^{(a)}\right\}_{i=0}^{n-1}$. We call $A(X_b)^{(b)}$ an *assignment*. Each X_b corresponds to a gadget, the vertex-colored star graph $M_0(X_b)$ with edge set $E_{M_0(X_b)} = \left\{\left\{\ell^{(a)}, A(X_b)^{(b)}\right\}\right\}_{\ell^{(a)} \in A(X_b)^{(b)}}$. Furthermore, $M_0 : V_\Gamma \longrightarrow [L \cup -L] \cup \left\{A(X_b))^{(b)}\right\}_{b \in \{0,1\}^n}$ is bijective and hence invertible. Thus, $M_0^{-1}(M_0(X_b)) = X_b$.

For each edge generator $X_b \in K$, we construct the graph $G_1(b)$ specified by edge set $E_{G_1(b)} = \bigcup_{b_i \neq 0} \left\{\{Z_{\boldsymbol{i}}^{(a)}, X_b^{(x_b)}\}, \{X_b^{(x_b)}, -Z_{\boldsymbol{i}}^{(a)}\}\right\}$. Here, $x_b = \alpha_b$ (i.e. we assign these vertices

a color based on the corresponding coefficient in the Hamiltonian) and we name such vertices *generator vertices*.

Each $G_1(b)$ only captures weights α_b corresponding to edges generated by $X_b \in K$. We still require gadgets that incorporate β_b, so that we can extract edge weights consistent with Eq. (8.3) from G. Define

$$U_b = \left\{ \left\{ (-1)^{b'_i} Z_{\boldsymbol{i}}^{(a)} \right\}_{b_i \neq 0}^{(c)} | \mathrm{i}^{2b\cdot b' - \|b\|_H} \beta_b < 0 \right\}_{b' \in \{0,1\}^n}. \tag{8.9}$$

Note that when $\beta_b = 0$, $U_b = \emptyset$.

To specify the gadget, we construct the directed vertex-colored graph $G_2(b) = \bigcup_{u_b^{(c)} \in U_b} g(u_b^{(c)})$, where each $g(u_b^{(c)})$ is the star graph with edge set

$$E_{g(u_b^{(c)})} = \left\{ (\ell^{(a)}, u_b^{(c)}), (u_b^{(c)}, -\ell^{(a)}), \{u_b^{(c)}, Y_b^{(y_b)}\} \right\}_{\ell^{(a)} \in u_b^{(c)}}.$$

We name the $u_b^{(c)}$s and the $Y_b^{(y_b)}$s *weight generator* and *weight generator cluster* vertices, respectively. Here, $y_b = \max_{b'} \alpha_{b'} + |\beta_b|$ is the color representing the cluster that allows us to extract edge weights of Γ.

Finally, we build a graph from the term $\sum_{\|b\|_H \leq k} \kappa_b Z_b$, where it helps to write $\kappa_b Z_b = |\kappa_b| C_b$. We call each C_b a *clause*, and we identify the set of assignments that "satisfy" the clause. For a choice of b,

$$\mathscr{C}_b = \left\{ \left\{ (-1)^{b'_i} Z_{\boldsymbol{i}}^{(a)} \right\}_{b_i \neq 0}^{(d)} | (-1)^{b\cdot b'} = \mathrm{sign}(\kappa_b) \right\}_{b' \in \{0,1\}^n}. \tag{8.10}$$

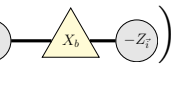
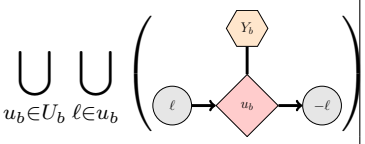
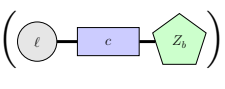
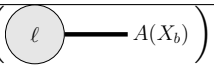
Hamiltonian Term	Gadget Name	Gadget Implementation
$X_b = \bigotimes_i X^{b_i}$	$G_1(b)$	
$Y_b = \bigotimes_i Y^{b_i}$, $\|b\|_H \in 2\mathbb{Z}$	$G_2(b)$	
$Z_b = \bigotimes_i Z^{b_i}$	$G_3(b)$	
Γ Vertex label		
X_b	$M_0(X_b)$	

Table 8.1: Gadgets summary. Vertex labels are as defined in the text (See Eqs. (8.9) and (8.10)). Superscripts are dropped and represented with a unique color/shape.

As H is k-local, $|\mathscr{C}_b| = 2^{|b|-1} \leq 2^{k-1}$. Now, we construct the edge set

$$E_{G_3(b)} = \left\{ \left\{ c^{(d)}, Z_b^{(z_b)} \right\} \cup \left\{ c^{(d)}, \ell^{(a)} \right\}_{\ell^{(a)} \in c^{(d)}} \right\}_{c^{(d)} \in \mathscr{C}_b}.$$

We name $c^{(d)}$'s and $Z_b^{(z_b)}$'s *clause* and *clause cluster* vertices, respectively. Here, $z_b = \max_{b'} \alpha_{b'} + \max_{b'} |\beta_{b'}| + |\kappa_b|$ is the color representing the cluster of satisfying assignments, which allows us to extract edge weights of Γ. These gadgets are summarized in Table 8.1.[4]

Given these constructions we define the direct mapping $G = M(\Gamma) = M_0[V_\Gamma] \cup M_1[E_\Gamma]$ as follows:

1. $M_0(u \in V_\Gamma)$ as defined above,

2. $M_1(\{u, X_b u\} \in E_\Gamma) = G_1(b) \cup G_2(b)$, and

3. $M_1(\{u, \infty\} \in E_\Gamma) = \bigcup_{\kappa_b \neq 0} G_3(b)$.

[4] As previously mentioned, for full generality one must consider composite gadgets for Hamiltonians with mixed terms (i.e. XYY or XYZ). See [Appendix F] for details.

Note that $M_1[E_\Gamma]$ contains all relevant information about $\mathrm{Aut}(\Gamma)$ and, thus, we can study $G' = M_1[E_\Gamma] \cup (L \cup \neg L, \emptyset)$. Note that when Γ is connected, $G' = M_1[E_\Gamma]$. Importantly, by construction, $|V_{G'}| = O(\mathsf{poly}(n))$. M also has the following useful property.

Theorem 8.7.1. *The function $M : \Gamma \mapsto G$ is bijective.*

Theorem 8.7.1 is true by construction, but we include proof in [Appendix F] for completeness. Now, we define $G(u) = M_0(u) \cup M_1[E_\Gamma]$ and can state the following theorem.

Theorem 8.7.2. *There exists a color-preserving isomorphism $G(u) \simeq G(v)$ if and only if $u \equiv v$.*

Theorem 8.7.2 is also true by construction, and explicit proof can be found in [Appendix F]. By exploiting the k-local structure of the Hamiltonian in the form of a compact G, we are able to reduce our problem from deciding whether $f \in \mathrm{Aut}\,\Gamma$ for an exponentially-sized graph Γ to deciding isomorphism $G(u) \simeq G(f(u))$ of polynomially-sized graphs $G(u), G(f(u))$.

Armed with this construction and the above theorems we can give the algorithm for FIND-REPRESENTATIVE.

Algorithm 3 Check equivalent vertices

```
function FindRepresentative(u, V')
    𝔾 ← G(u)
    for v ∈ V' do
        𝔾' ← G(v)
        if 𝔾 ≃ 𝔾' then return v
        end if
    end for
    return ∅
end function
```

We note that $G(u)$ can be constructed in time $O(\mathsf{poly}(n))$ and that color-preserving GI on bipartite, directed graphs is GI-complete [178]. Additionally, Algorithm 3 with stoquastic k-local Hamiltonians is GI-complete. To see this, for any two graphs S, S', label vertices such

that $V_S \cap V_{S'} = \emptyset$ and let $H = \sum_{\{i,j\} \in E_S} Z_{\boldsymbol{i}} Z_{\boldsymbol{j}} + \sum_{\{i,j\} \in E_{S'}} Z_{\boldsymbol{i}} Z_{\boldsymbol{j}}$. Then, $G\left(X_{\bigoplus_{i \in V_S} \boldsymbol{i}}\right) \simeq G\left(X_{\bigoplus_{i \in V_{S'}} \boldsymbol{i}}\right)$ iff $S \simeq S'$. In the other direction, Algorithm 3 uses GI as a subroutine. The best known algorithm for GI takes time $\mathrm{QP}(n) = 2^{O\left(\log(n)^{O(1)}\right)}$ [171, 172], and therefore the entire routine takes $O\left(|V'| \, \mathrm{QP}(n)\right)$.

8.8 Discussion

Our results can be extended to near-symmetries via a straightforward application of the Davis-Kahan $\sin \Theta$ theorem [179, 180]. In particular, let H, Δ be Hamiltonians where H has ground state density matrix ρ and $H + \Delta$ has ground state density matrix ρ_Δ from which we would like to sample. Then,

$$\sqrt{1 - F(\rho, \rho_\Delta)} \leq \frac{\pi}{2} \frac{\|(I - \rho_\Delta) \, \Delta \rho\|_F}{\lambda_1(H) - \lambda_0(H + \Delta)} \leq \frac{\pi}{2} \tan^2 \Phi$$

where $F(\rho, \rho_\Delta) = \|\rho \rho_\Delta\|_F^2$ is the fidelity, $\tan^2 \Phi = \frac{|\langle \Delta \rangle_\rho|}{\gamma(H) - \langle \Delta \rangle_\rho}$, and $\gamma(H)$ is the eigenvalue gap of H. Thus, $F(\rho, \rho_\Delta) \geq 1 - \frac{\pi^2}{4} \tan^4 \Phi$.

If one has *any* procedure for producing a guess ρ, one can later check that $\langle \Delta \rangle_\rho$ is small enough to guarantee $\|\rho - \rho_\Delta\| \leq \epsilon$. As a limited example, suppose one perturbs each $\alpha_b, \beta_b, \kappa_b$ in H by at most δ. Then, $\|\Delta\|_F \leq \delta \|H\|_F$. Therefore, provided that $\delta \leq \epsilon \frac{\gamma}{\|H\|_F}$, $\tan^2 \Phi \leq \frac{\epsilon}{1 - \epsilon}$. Hence, we can achieve arbitrary precision ϵ while perturbing each of $\alpha_b, \beta_b, \kappa_b$ by $\delta = O\left(\epsilon \gamma / \|H\|_F\right)$, where we have assumed $\gamma / \|H\|_F = \Omega\left(\mathrm{poly}^{-1} n\right)$ throughout.

Alternatively, an approximate GI algorithm [181] might suffice to implement FINDREPRESENTATIVE; more general approximation algorithms are left for future work.

8.9 Conclusion

Our algorithm rules out the existence of an exponential separation between classical algorithms and StoqAQC using Hamiltonians with effective subspaces with size $|V'|$ scaling subexponentially in n, a class containing all previously known k-local obstructions [14, 149, 165].

Beyond these symmetric and approximately symmetric problems, whether all k-local stoquastic Hamiltonians are quasi-polynomially simulable remains an open question. We conjecture that families of Hamiltonians that lack near-symmetries typically have exponentially small gaps, suggesting that they are difficult for AQC.[5] This conjecture is largely motivated by the fact that avoiding exponentially small gaps requires pathologically smooth transitions, as explained in [Appendix F].[6] Proving this, combined with our results here and a better understanding of those near-symmetries that we can efficiently approximate, would reduce understanding the simulability of StoqAQC to better understanding the significance of the gap in both classical and quantum cases.

[5] A large spectral gap is not *necessary* for successful AQC. It is not too difficult to construct examples where StoqAQC succeeds despite a small gap, though they often appear pathological and one might expect classical methods to be similarly successful.

[6] As noted in the [Appendix F], this result is similar to, but in terms of gap-analysis, stronger than that in [182]

Chapter 9: Simultaneous Stoquasticity

9.1 Introduction

The efficient simulation of quantum phenomena is essential to understanding chemistry, materials, and physics, and similarly the lack of efficient classical simulation is critical to the long-term applicability of quantum computing. One of the key properties that can make a Hamiltonian easy to simulate classically is stoquasticity [145], a basis dependent property where the off-diagonal matrix elements are real and non-positive.[1]. Such stoquastic Hamiltonians do not suffer from the sign problem allowing classical simulation of their ground state properties via Monte Carlo techniques [183, 184]

Stoquastic Hamiltonians have been especially important in the development of quantum annealing [185] and quantum adiabatic computation [146]. Adiabatic quantum computing is quantum universal [13], but the proof relies on non-stoquastic Hamiltonians. There is growing evidence that adiabatic computing with stoquastic Hamiltonians is no more powerful than classical computing [184, 186–189] except in contrived highly non-local settings [18]. In complexity theory, stoquastic Hamiltonians appear in the definition of the complexity class `StoqMA`, which characterizes the computational hardness of the local Hamiltonian problem for stoquastic Hamiltonians [190].

[1] In the mathematics literature, such matrices are called Z-matrices or negative Metzler matrices.

A large body of literature has been built up around the problem of finding stoquastic [191, 192] or nearly stoquastic [193, 194] bases for Hamiltonians. The corresponding unitary basis change is said to "cure" the non-stoquastic Hamiltonian. While the existence of such a basis is guaranteed by the diagonalizability of Hermitian matrices (the locality of such a basis is not guaranteed), finding such a basis change is an NP-hard problem [195–197]. However, this literature has mostly focused on just curing a single Hamiltonian's sign problem. In order to run simulated quantum annealing [15, 187] or otherwise simulate the behavior of adiabatic computation, both annealing Hamiltonians must be stoquastic. This raises the question not just of how to find a basis in which two Hamiltonians are simultaneously stoquastic but further whether such a basis even exists. Our work along this direction is complementary to the results in Ref. [196] where the authors consider the problem of stoquasticizing a local Hamiltonian consisting of a sum of local terms and showed that this Hamiltonian can be stoquasticized if and only if all terms can be simultaneously stoquasticized. Furthermore, they showed that it is NP-hard to find a basis that accomplishes this.

To formally state our problem of interest: let $\texttt{Stoq}$ be the set of all stoquastic matrices. Given a set of Hamiltonians $S = \{H_1, H_2, \cdots H_m\}$ defined on a d-dimensional Hilbert space $\mathcal{H}_d$, does there exist a single unitary U that simultaneously cures the non-stoquasticity of ("stoquasticizes") all $H_j \in S$; that is, $\exists\, U$ such that $UH_jU^\dagger \in \texttt{Stoq}$ for all j?

Using the mathematical theory of simultaneous unitary similarities [198–201], we find that the problem reduces to determining if there exists a solution to a system of polynomial (in)equalities. For $m > 2$ and/or $d > 2$, the resulting system of polynomial equations does not generically have a solution, and therefore a simultaneously stoquasticizing unitary does not always exist. In fact, we show almost every set S of Hamiltonians is not simultaneously stoquas-

tizable. By considering a generalized Bloch vector representation, we can geometrically interpret our results, connecting to the literature on the geometry of quantum states [202–207].

This result has broad implications for adiabatic quantum computing, where annealing between two Hamiltonians that are not simultaneously stoquasticizable should be hard to simulate classically, independent of basis. The more general theory of simultaneous transformation of two or more Hermitian operators plays a key role in other areas of quantum physics, the most obvious being simultaneous diagonalizability governing the commutativity and compatability of observables. Similarly simultaneous unitary congruence has been used to show that quantum seperability is connected to the simultaneous hollowability of matrices [208], and simultaneous orthogonal equivalence connects to local unitary equivalence of a pair of quantum states [209].

9.2 Lie Algebras and Lie Groups

Formally, any Hamiltonian $H \in \mathbb{C}^{d\times d}$ is (up to a physically irrelevant shift by a multiple of identity) an element of the Lie algebra $\mathfrak{su}(d)$. Here, we take the usual physicist convention that $\mathfrak{su}(d)$ consists of the set of all $d \times d$ traceless, Hermitian matrices. This is known as the fundamental representation. The Lie algebra $\mathfrak{su}(d)$ has real dimension $d^2 - 1$ and, therefore, any element of the Lie algebra may be expanded in a basis of $d^2 - 1$ elements of the algebra, which we choose to obey the standard orthonormality condition

$$\mathrm{Tr}(\hat{\lambda}_i \hat{\lambda}_j) = 2\delta_{ij}. \tag{9.1}$$

We also have that

$$\hat{\lambda}_i \hat{\lambda}_j = \frac{2}{d}\delta_{ij} I + i f_{ijk} \hat{\lambda}_k + d_{ijk} \hat{\lambda}_k, \tag{9.2}$$

where I is the identity matrix and f_{ijk} and d_{ijk} are the totally anti-symmetric and symmetric structure constants, respectively. We use the convention of summing over repeated indices. A standard choice of basis is the generalized Gell-Mann basis. It is made up of $d(d-1)/2$ symmetric matrices,

$$\hat{\lambda}^{(x)}_{jk} = |j\rangle\langle k| + |k\rangle\langle j|, \quad (1 \leq j < k \leq d), \tag{9.3a}$$

$d(d-1)/2$ skew-symmetric matrices,

$$\hat{\lambda}^{(y)}_{jk} = -i|j\rangle\langle k| + i|k\rangle\langle j|, \quad (1 \leq j < k \leq d), \tag{9.3b}$$

and $d-1$ diagonal matrices,

$$\hat{\lambda}^{(\mathrm{diag})}_{j} = \sqrt{\frac{2}{j(j+1)}}\mathrm{diag}(\underbrace{1, \cdots, 1}_{j}, -j, 0, \cdots, 0), \tag{9.3c}$$

where in this final equation we have $j \in \{1, \cdots, d-1\}$. For $\mathfrak{su}(2)$, the generators defined in this way are the familiar Pauli operators, which motivates the x, y superscipts for the symmetric and skew-symmetric generalized Gell-Mann matrices, respectively. For $\mathfrak{su}(3)$, they are the Gell-Mann matrices.

We can write any traceless Hamiltonian H in this basis as

$$H = \boldsymbol{b} \cdot \hat{\boldsymbol{\lambda}}, \tag{9.4}$$

where $\hat{\boldsymbol{\lambda}}$ is a vector of basis elements and $\boldsymbol{b} \in \mathbb{R}^{d^2-1}$ is the so-called (generalized) *Bloch vector* corresponding to H. We consider grouping the components of $\hat{\boldsymbol{\lambda}}$ into subsets matching the basis elements defined in Eqs. (9.3a)-(9.3c) as follows: let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{D}$ be the sets of indices corresponding to the symmetric, skew-symmetric, and diagonal generalized Gell-Mann matrices, respectively. We have $\{\mathcal{X}, \mathcal{Y}, \mathcal{D}\} = \{1, \cdots, d^2 - 1\}$.

Therefore, there exists an isomorphism $S \cong B$ between a set $S = \{H_1, H_2, \cdots, H_m\}$ of traceless Hamiltonians and a set of corresponding Bloch vectors $B = \{\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \cdots \boldsymbol{b}^{(m)}\}$. These Bloch vectors will simplify a number of proofs and provide a valuable geometric interpretation of our results.

9.3 Simultaneous stoquasticity

Given the set $S = \{H_1, H_2, \cdots, H_m\}$ we want to solve the decision problem: Does there exist a unitary U such that $H'_j = UH_jU^\dagger \in \texttt{Stoq}$ for all $H'_j \in S' := \{UH_1U^\dagger, UH_2U^\dagger, \cdots, UH_mU^\dagger\}$?

Observe that the choice of trace of the Hamiltonians and of the Hermitian generators of U can be chosen to be zero with no physical consequence. Therefore, without loss of generality, we restrict our consideration to traceless $H_j \in S$ and to special unitaries $U \in SU(d)$.

This assumption allows us to directly describe the problem in terms of Bloch vectors as detailed in the previous section. In particular, consider the sets of Bloch vectors $B \cong S$ and $B' \cong S'$. In the space of Bloch vectors `Stoq` corresponds to the subset of Bloch vectors such

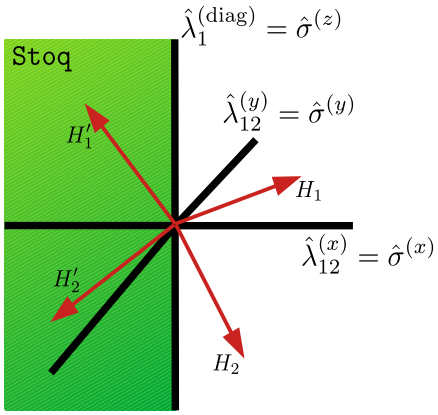


Figure 9.1: For a qubit, the geometric representation of `Stoq` in Bloch vector space is the $\pm\hat{\sigma}^{(z)}, -\hat{\sigma}^{(x)}$ half-plane. Observe that for two Hamiltonians H_1, H_2, represented by their Bloch vectors, there always exists a unitary which can simultaneously take both H_1', H_2' to the `Stoq` subspace.

that $b_j = 0$ for $j \in \mathcal{Y}$ and $b_j \leq 0$ for $j \in \mathcal{X}$. The decision problem is now that of finding whether there exists a unitary U such that the vectors $\boldsymbol{b}' \in B'$ all fall in this subspace.

When $d = 2$, the Bloch space is of dimension $d^2 - 1 = 3$ and `Stoq` is easily visualizable as the $\pm\hat{\sigma}^{(z)}, -\hat{\sigma}^{(x)}$ half-plane, as depicted in Fig. 9.1. In this case, it is well-known that $SU(2)$ is a double-cover of $SO(3)$, and therefore we can visualize the action of unitaries on S as rotations of the collection of vectors B. It is simple to observe that the answer to our decision problem is yes if and only if the vectors B all lie in a single half-plane. This plane can then be rotated via some unitary to align with the `Stoq` half-plane. This is always possible if $m \leq 2$ and $d = 2$.

We seek to generalize and formalize this geometric intuition for $d > 2$. We will make use of the mathematical theory of simultaneous unitary similarities [198–200, 210] and the related theory of simultaneous invariants [211]. Two ordered sets of m matrices S, S' are *simultaneously unitarily similar* if there exists a unitary U such that $H_j' = UH_jU^\dagger$ for all $H_j \in S$, $H_j' \in S'$. In this terminology our goal is to determine if the set S of Hamiltonians is simultaneously unitarily similar to a set $S' \in$ `Stoq`.

Define a *word* on a set T as any formal product of nonnegative powers of the elements $t_j \in T$. We then have the following theorem due to Ref. [200].

Theorem 9.3.1. *The ordered sets of Hermitian matrices* $S = \{H_1, \cdots, H_m\}$ *and* $S' = \{H'_1, \cdots, H'_m\}$ *are simultaneously unitarily similar if and only if* $\mathrm{Tr}[w(S)] = \mathrm{Tr}[w(S')]$ *for all words* w *in* S, S'.

The quantities $\mathrm{Tr}[w(S)]$ are known as *trace invariants* under simultaneous unitary similarity. Unfortunately, Theorem 9.3.1 is a practically useless condition since it requires the checking of all words in S and S'. To get around this issue, one must demonstrate that only a finite set of independent words exist [210, 212–214].

When checking the unitary similarity of a single pair of Hermitian matrices H, H', it is often quoted in the physics literature (typically in the context of density matrices) that it is sufficient to check the equivalence of the trace invariants $\mathrm{Tr}[H^k]$ and $\mathrm{Tr}[H'^k]$ for $k \in [1, d]$ [203, 215–217]. While perhaps intuitively obvious—as Hermitian matrices have d real eigenvalues—this is typically stated without proof. For completeness, we give such a proof in the Appendix.

More generally, for $m \geq 2$, we can show that it is sufficient to consider word lengths up to

$$\ell_{\max} = \min \begin{cases} \lceil \frac{(cd)^2+2}{3} \rceil \\ cd\sqrt{\frac{2(cd)^2}{cd-1} + \frac{1}{4}} + \frac{cd}{2} - 2. \end{cases} \tag{9.5}$$

where c is the minimum integer such that $(c^2-3c+2)/2 \geq m$. For instance, if $m = 2$, $c = 4$. The proof mostly follows Refs. [200, 212, 213], and the derivation of this expression is demonstrated in the Appendix.

Therefore, the decision problem of whether there exists a unitary that simultaneously stoquasticizes S is equivalent to determining if there exists a solution to the system of polynomial

(in)equalities in the matrix elements of $H' \in S'$:

$$\mathrm{Tr}[w(S)] = \mathrm{Tr}[w(S')], \qquad \forall\, |w| \leq \ell_{\max} \tag{9.6a}$$

$$\mathrm{Re}(H'_{jk}) \leq 0, \qquad \forall j \neq k, H' \in S' \tag{9.6b}$$

$$\mathrm{Im}(H'_{jk}) = 0, \qquad \forall j \neq k, H' \in S'. \tag{9.6c}$$

This amounts to a system of $\sum_{\ell=2}^{\ell_{\max}} m^\ell + md(d-1)/2 \sim m^{O((cd)^{3/2})}$ polynomial equations and $md(d-1)/2$ inequality constraints on md^2 real variables. Many of the equations for different words in Eq. (9.6a) will end up being redundant due to symmetries such as the cyclicity of the trace and algebraic dependence of the resulting trace invariants. Independent of if one can identify the minimal set of such constraints, solving the decision problem of whether or not a solution exists to this set of polynomial (in)equalitites is NP-hard and lies in PSPACE [218]. Therefore, identifying if S is simultaneously stoquasticizable is completely intractable for large problem instances.

9.4 A No-go Result

Given this computational difficulty, we also present the following no-go result.

Theorem 9.4.1. *A necessary condition for S to be simultaneously stoquasticizable is that for every eigenvalue $\lambda \neq 0$ of $i[H_i, H_j]$ there is another eigenvalue $-\lambda$ of $i[H_i, H_j]$ (paired eigenvalue condition) for all $H_i \neq H_j \in S$.*

Proof. Any two matrices $H'_i, H'_j \in \texttt{Stoq}$ have all real matrix elements. Therefore, the Hermitian matrix $C' = i[H'_i, H'_j]$ must be skew-symmetric. Skew-symmetric matrices have the property

that all eigenvalues are paired. As eigenvalues remain unchanged under action by a unitary and if we act on H_i', H_j' by a unitary, C' changes equivalently, this paired property must exist for any simultaneously stoquastizable H_i, H_j. This holds for all pairs of Hamiltonians in S. □

This theorem provides a straightforward condition to rule out if S is simultaneously stoquasticizable. However, the presence of paired eigenvalues does not guarantee simultaneous stoquasticity as: (a) stoquastic matrices must have negative, as well as real, off-diagonal elements; (b) it is possible for simultaneously non-stoquastic Hamiltonians to have a commutator with paired eigenvalues.

This condition also relates to the dynamical Lie algebra from quantum control theory [219, 220] which for simultaneously real Hamiltonians (in any basis) neatly breaks up into a Cartan decomposition with every other layer of the dynamical Lie algebra (i.e. nested commutators with even numbers of our original Hamiltonians) corresponding to purely imaginary Hamiltonians (transformed into a given basis). Therefore, finding a basis in which a set of Hamiltonians are simultaneously real, a necessary condition for simultaneous stoquasticization, is equivalent to identifying whether there is a Cartan decomposition of $\mathfrak{su}(d) = \mathfrak{p} \oplus \mathfrak{so}(d)$ where the set of Hamiltonians is contained in $\mathfrak{p}$.

9.5 Bloch Vector Approach

We now reexpress the trace invariants in terms of Bloch vectors. This provides a geometric interpretation that neatly connects back to the intuition from the one qubit example given earlier, while highlighting a number of symmetries between words that are less clear in the alternative formalism. This approach will also allow us to prove another no-go result, from which the fol-

lowing theorem establishing the rareness of simultaneous stoquasticity immediately follows.

Theorem 9.5.1. *For almost every S with $m \geq 2$, $d \geq 3$, S is not simultaneously stoquasticizable.*

"Almost every" is used in the technical sense that the set of simultaneously stoquastizable S are measure zero.

With this goal in mind, let us consider expressing the trace invariants from Theorem 9.3.1 in terms of Bloch vectors. For words of arbitrary length $|w|$, we have

$$\mathrm{Tr}[w(S)] = \mathrm{Tr}\left[\prod_{j=1}^{|w|} \sum_{\mu_j=1}^{d^2-1} b_{\mu_j}^{(w_j)} \hat{\lambda}_{\mu_j}\right], \tag{9.7}$$

where we have denoted the j-th element of w as w_j.

Now consider evaluating Eq. (9.7) explicitly for words of small length. By our assumption of tracelessness, the trace invariant for any $w(S)$ of length one is zero. A general trace invariant for $|w| = 2$ is

$$\mathrm{Tr}[H_i H_j] = \mathrm{Tr}[(\boldsymbol{b}^{(i)} \cdot \hat{\boldsymbol{\lambda}})(\boldsymbol{b}^{(j)} \cdot \hat{\boldsymbol{\lambda}})] = 2\boldsymbol{b}^{(i)} \cdot \boldsymbol{b}^{(j)}, \tag{9.8}$$

where we used Eq. (9.2) to evaluate the trace. Therefore, the lengths (from the $i = j$ case) and relative angles (from the $i \neq j$ case) of the Bloch vectors corresponding to pairs of Hamiltonians in S are simultaneous trace invariants.

This result is intuitively satisfying. For a qubit, recalling that $SU(2)$ is homomorphic to $SO(3)$, this is precisely what we would expect to be invariant for a rigid collection of vectors being rotated simultaneously about the origin. We should also expect this to be the only constraint for a qubit. This expectation is validated by computing the trace invariant for words of length

three:

$$\mathrm{Tr}[H_i H_j H_k] = 2 d_{\mu\nu\xi} b_\mu^{(i)} b_\nu^{(j)} b_\xi^{(k)} = 2(\boldsymbol{b}^{(i)} * \boldsymbol{b}^{(j)}) \cdot \boldsymbol{b}^{(k)}, \tag{9.9}$$

where we have introduced the star product, defined component-wise, using the symmetric structure constants, as,

$$(\boldsymbol{b}^{(i)} * \boldsymbol{b}^{(j)})_\xi = d_{\mu\nu\xi} b_\mu^{(i)} b_\nu^{(j)}. \tag{9.10}$$

For $\mathfrak{su}(2)$, the symmetric structure constants are all zero, so, as expected for a qubit, words of length greater than two provide no further constraints.

Various properties of the star product are detailed in the Appendix. In particular, observe that the star product is not associative and that Eq. (9.9) is completely symmetric in the input word. Similar observations allow us to show that any trace invariant can be written as $\boldsymbol{v} \cdot \boldsymbol{b}^{(i)}$ for some $i \in [1, m]$, where $\boldsymbol{v}$ is any vector in the set $\mathcal{B}$ of all possible combinations of star products between Bloch vectors in B. That is $\mathcal{B} = \{\boldsymbol{b}^{(j)}, \boldsymbol{b}^{(j)} * \boldsymbol{b}^{(k)}, (\boldsymbol{b}^{(j)} * \boldsymbol{b}^{(k)}) * \boldsymbol{b}^{(l)}, \cdots\}$. This can be verified by direct computation, but we provide explicit proof in the Appendix.

Given this formalism, we can pick a finite set of Bloch trace invariants using Eq. (9.5) and then construct an equivalent decision problem to Eqs. (9.6a)-(9.6c) to test for simultaneous stoquasticity. The stoquasticity conditions in this context are $b_j'^{(i)} = 0$ $j \in \mathcal{Y}$ and $b_j'^{(i)} \leq 0$ for $j \in \mathcal{X}$ for all i.

We also obtain the following no-go result.

Theorem 9.5.2. *Let S be a set of Hermitian matrices with corresponding Bloch vectors $B =$*

$\{\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \cdots \boldsymbol{b}^{(m)}\}$. *Let $\mathcal{B}$ be the set of all possible star products between elements of B. A necessary condition for S to be simultaneously stoquasticizable is that* $\dim(\mathrm{span}(\mathcal{B})) \leq (d^2 + d - 1)/2$.

Proof sketch. Observe that for all $H_i \in \texttt{Stoq}$, $\boldsymbol{b}_j^{(i)} = 0$ for $j \in \mathcal{Y}$. From the definition of the star product and the form of the non-zero symmetric structure constants of $\mathfrak{su}(d)$ [221], one observes that if $S \in \texttt{Stoq}$ all vectors in $\mathcal{B}$ are also in this subspace. The dimension of $\mathcal{B}$ is invariant under unitary transformations, so this is a necessary condition for simultaneous stoquasticity. □

Full details are provided in the Appendix. Most importantly, this result leads directly to Theorem 9.5.1, the proof of which we sketch below, again leaving the algebraic details to the Appendix.

Proof sketch of Theorem 9.5.1. From similar analysis of the star products, we can prove that for almost every S, $\dim(\mathcal{B}) = d^2 - 1$. That is, $\mathcal{B}$ spans the full Bloch vector space for almost every S. Combining this result with Theorem 9.5.2, Theorem 9.5.1 immediately follows. □

9.6 Conclusion and Outlook

Quantum annealing relies on the interaction of two non-commuting Hamiltonians, and there are clear connections between the power of that computation and the stoquasticity of the Hamiltonians. Our results provide proof that a general quantum annealing procedure does not possess any basis in which it can be described completely stoquastically. We know that classical computing can be described using simultaneously diagonal Hamiltonians, and the seeming power of non-stoquasticity speaks to the idea that quantum advantage might lie further past simultaneously stoquastic Hamiltonians. More work is needed to determine how tightly quantum

advantage is bound up merely with these notions of simultaneous stoquasticity and how much other factors, such as locality of the simultaneous basis, play a role.

Furthermore, our results provide a definitive set of conditions for simultaneous stoquasticity, which are, as expected, difficult to calculate in practice given the computational complexity of this problem. The commutator condition of Theorem 9.4.1, while enticing from its connections to simultaneous diagonalizability (see the Appendix for a further exploration of how these geometric ideas relate to simultaneous diagonalizability) and dynamical Lie algebras, provides only a necessary but not sufficient condition and then only on a simultaneous real basis, not specifically a simultaneous stoquastic basis. The other no-go result in Theorem 9.5.2 suffers a similar flaw.

Further work is also possible by extending these results beyond just stoquasticity to the full class of Hamiltonians lacking sign problems. General sign problem free Hamiltonians take a Vanishing Geometeric Phase (VGP) form [175, 222]. This form generalizes the notion of stoquastic Hamiltonians to all Hamiltonians generated from stoquastic Hamiltonians via diagonal unitary transformations. While this is a more general form that should be studied in the context of simultaneous transformations, it lacks linearity, meaning that linear combinations of VGP Hamiltonians are not necessarily VGP, hinting that simultaneous stoquasticizability is a more fundamental concept to consider for multiple Hamiltonians.

Chapter 10: Discussion and Further Directions for Part II

While they do not all directly address it, Chapters 7-9 are generally oriented around an essential question: Is exponential quantum advantage possible for adiabatic quantum computation with stoquastic, k-local Hamiltonians? By rigorously ruling out all known attempts at constructions that provide an exponential separation for particular examples, Chapter 8 suggests not. On the other hand, as mentioned in Chapter 1, recent results proving a separation for sparse Hamiltonians indicate that stoquasticity alone is not sufficient to constrain adiabatic quantum computation to the power of classical computation [18, 19]. However, such proofs seem unlikely to generalize to k-local Hamiltonians. Tellingly, k-locality is essential to the approach of Chapter 8.

Such results are promising, but answering the full question rigorously remains extremely challenging. One possible approach would be to prove that large amounts of symmetries of the sort considered in Chapter 7 and Chapter 8 are necessary for separations between classical algorithms and k-local, stoquastic adiabatic computation. This idea is reminiscent of digital quantum computation where sufficient structure of some sort seems to be an essential ingredient for large quantum advantages [223]. Thus, one could reasonably expect large amounts of structure of this particular kind to be necessary in this adiabatic setting, but proving it for arbitrary classical algorithms is certainly a tall order. Perhaps the geometric approaches of Chapter 9 could provide a helpful tool-set.

The challenges of a rigorous proof aside, it is generally expected that adiabatic quantum computation with k-local stoquastic Hamiltonians will not provide any particularly large speed ups over classical algorithms. Consequently, significant attention in recent years has been devoted to trying to design and understand non-adiabatic quantum annealing algorithms. In such general quantum annealing schemes, as in adiabatic quantum computation, one considers initializing a quantum system in some simple-to-prepare ground state of a Hamiltonian H_0 and then attempts to prepare the ground state of some other Hamiltonian H_1. The adiabatic theorem provides a sufficient, but not necessary, condition for successfully completing this task. Alternative schemes include algorithms like the Quantum Alternating Operator Ansatz (QAOA) [224] and other such diabatic strategies [225]. While extensively studied numerically, at the moment, little is understood rigorously about these more general quantum annealing schemes. Consequently, seeking such rigorous results is a timely direction to pursue.

We have begun to address this shortcoming using two distinct approaches in recent work. In Ref. [16] we use quantum speed limits—bounds that are essentially a re-cast form of those we leveraged for quantum metrology in Part I of this dissertation—to provide saturable lower bounds on annealing times for arbitrary annealing schedules. However, the examples of saturation involve fine-tuned highly symmetric problems similar to those in Chapter 7, and the bounds do not immediately provide guidance for designing better annealing protocols for more realistic scenarios. This is an important area for future work.

In contrast, the results of Ref. [20] are better adapted to realistic scenarios, but this comes at the cost of expressions that are significantly more challenging to evaluate than those in Ref. [16]. Here, we derive an "intermediate timescale" adiabatic theorem that provides a route towards rigorously understanding a numerically [226] and experimentally [227] observed phenomenon

where, in certain circumstances, one can go quickly through an avoided crossing, which is the bottleneck for an adiabatic algorithm, and still end up with sizeable overlap with the target state at the end of the procedure. An important piece of future work in this context is to work out these applications in detail, perhaps with the help of a loosening of our existing bound to ease its applicability.

These results and the immediate questions they raise are just the first steps towards a deeper, rigorous understanding of quantum annealing, with important practical applications for quantum optimization algorithms. Using these mathematical tools, we anticipate this being a fruitful area of research for the forseeable future.

Part III

Quantum Simulation

Chapter 11: Parallelization Techniques for Quantum Simulation of Fermionic Systems

11.1 Introduction

The ability to simulate complex fermionic systems is an important area of promise for quantum computers with applications ranging from quantum chemistry and condensed matter physics to nuclear and high-energy physics [228–231]. Before performing such a simulation, however, one must map from the fermionic operators to operators acting on the Hilbert space of the qubits of the quantum computer. A common approach to performing such a mapping is to use the Jordan-Wigner transformation [232, 233], which encodes local fermionic operators on N fermionic modes as non-local qubit operators on N qubits. This non-locality, which manifests as strings of Pauli-Z operators, is the price of obtaining the correct fermionic anti-commutation relations when using qubit operators. Unfortunately, even for physically-local fermionic interactions in higher than 1+1 dimensions, the length of these Pauli-Z strings can scale with the system size. This results in costly fermionic simulations [234, 235] on near-term quantum devices where the two-qubit entangling-gate (e.g., CNOT) count of an algorithm is expected to be the limiting factor. In particular, the Pauli weight of an operator G (the number of qubits on which it acts non-trivially) is directly related to the number of two-qubit entangling gates needed

to implement the unitary $U = \exp(-iG)$ [236]. Nonetheless, recent progress has resulted in improvement in both entangling-gate count and in circuit depth when simulating given Jordan-Wigner-transformed fermionic Hamiltonians using product formulas, resorting to e.g., suitable term ordering and nesting strategies [237] or fermionic SWAP networks [238–240].

A number of other mappings from fermions to qubits have been proposed in the literature [241–252]. Many of these proposals aim to map local fermionic operators to local qubit operators, forming a class that is called local encodings in this [chapter]. Local encodings trade operator non-locality for state non-locality as a vehicle for encoding fermionic anti-commutation relations in qubits. In particular, one finds that to preserve the appropriate anti-commutation relations via a local encoding, one must: a) increase the number of qubits, and b) restrict the state of the system to lie within some subspace of the Hilbert space—typically the logical codespace a (modified) toric code. At the price of these complications, one generally obtains lower gate counts required for simulation. The comparative analyses of various encodings given the Hamiltonian under study, the quantum resources to be optimized, and the architecture connectivity constitute an active area of research, see e.g., Refs. [240, 244, 250, 253, 254].

In this work, we explore the potential for parallelization (that is the ability to simultaneously simulate several Hamiltonian terms) in local encodings, hence reducing the circuit depth of the simulation. In this context, product formula-based Hamiltonian-simulation algorithms based on Trotter-Suzuki decomposition of the time-evolution operator [255–257] are best suited for this analysis, nonetheless, other simulation algorithms [258–265] can also benefit from the parallelization strategy explored here. We consider the parallelization problem in connection to (a slightly abstracted version of) the underlying qubit architecture of the quantum computer, and emphasize, both analytically and numerically, connections between qubit architectures, fermionic-

encoding locality, and parallelization. It is found that the problem of parallelization is equivalent to path coloring on a graph that represents the particular fermion-to-qubit mapping under consideration and the physical interactions among the fermionic modes. Consequently, heuristic classical algorithms can be used to inform efficient implementations of fermionic simulations on quantum hardware.

The particular graph-theoretic approach of this work is enabled by the strategy undertaken in Ref. [249], in which a general framework for local fermionic encodings of the sort described above is developed. In particular, it was demonstrated how to disconnect the interaction graph of the fermionic modes being simulated and the so-called *system graph*, which determines the fermionic encoding in a flexible and qubit-architecture-aware manner. This separation enables the construction of the so-called *custom fermionic codes*, which are a generalization of the Bravyi and Kiteav superfast encoding [241, 246]. The Jordan-Wigner transformation is a limiting case of such custom codes. The degree of non-locality can be reduced upon introducing further qubits and local connectivity on the system graph at will, and such choices amount to a range of custom encodings. The input to our parallelization problem is such a system graph, which fixes the encoding chosen to implement the interactions in the original fermionic Hamiltonian. The question investigated in this [chapter] is to what extent the Hamiltonian simulation can be parallelized, and whether certain system graphs are best suited for maximal parallelizability.

The structure of this [chapter] is as follows. Custom fermionic codes of Ref. [249] is reviewed in Sec. 11.2. In Sec. 11.3, we demonstrate how the problem of parallelizing product formula-based Hamiltonian-simulation algorithms maps to path coloring on the system graph, and is, therefore, a NP-hard problem. This is named the weak coloring problem. By considering the fine-grained details of the fermion-to-qubit mapping below the abstraction level of the sys-

tem graph, another path coloring problem is defined. This is called the strong coloring problem. Analytic results on the weak and strong coloring problems for a few illustrative systems are presented in Sec. 11.4. A numerical approach to heuristically solving the weak and strong coloring problems is presented Sec. 11.5. The numerical and analytic results are then compared for these system graphs, exhibiting consistency. We further numerically investigate the weak and strong coloring problems for a few realistic system graphs designed for current qubit architectures. It is found that by solving (or finding heuristics for) the more detailed strong coloring problem, one can often obtain significant gains in parallelizability compared to the more abstracted weak coloring problem. For most system graphs, these improvements are a constant factor—for instance, in the case of a star system graph or complete system graph, strong coloring asymptotically provides up to a factor of two improvement over weak coloring. However, we also construct an example for which the advantage gained grows linearly in the system size, which is the maximal possible gain from considering strong coloring. Both weak and strong coloring approaches provide large reductions in circuit depth compared to a naive sequential approach. Finally, the performance gains of strong coloring depend heavily on the choice of enumerating qubits in the encoding—a feature that is also taken advantage of in Ref. [252] to provide optimal fermion-qubit mappings. Sec. 11.6 includes a summary of the results and a discussion of possible directions for future study. The code generating the colored graph from the system graph and the corresponding physical interactions is [available online] [266].

11.2 Custom Fermion-to-Qubit Mappings

11.2.1 Setup

Consider a system of N fermionic modes. A general fermionic Hamiltonian may be written in the second quantization as

$$H = \sum_{uv} \kappa_{uv} a_u^\dagger a_v + \sum_{uvwx} \kappa_{uvwx} a_u^\dagger a_v^\dagger a_w a_x + \cdots, \tag{11.1}$$

where $a_u^\dagger$ and a_u are the fermionic creation and annihilation operators on site u, respectively, satisfying the standard fermionic anticommutation relations

$$\{a_u, a_v^\dagger\} = \delta_{uv}, \quad \{a_u, a_v\} = \{a_u^\dagger, a_v^\dagger\} = 0, \tag{11.2}$$

and $\kappa_{uv}, \kappa_{uvwx}$ are some coefficients consistent with the Hermiticity of the Hamiltonian. It is convenient to consider a Majorana basis γ, γ' for the fermionic operators as

$$\begin{aligned} \gamma_u &= a_u^\dagger + a_u, \\ \gamma'_u &= i(a_u^\dagger - a_u), \end{aligned} \tag{11.3}$$

that clearly satisfy

$$\{\gamma_u, \gamma_v\} = 2\delta_{uv}, \quad \{\gamma'_u, \gamma'_v\} = 2\delta_{uv}, \quad \{\gamma_u, \gamma'_v\} = 0. \tag{11.4}$$

To simulate a fermionic Hamiltonian on a quantum computer, one must first map from fermions to qubits while preserving these anti-commutation relations. The standard approach is the Jordan-Wigner mapping from N fermionic modes to N qubits,

$$\begin{aligned} a_u &\to \prod_{v<u} Z_v (X_u + iY_u), \\ a_u^\dagger &\to \prod_{v<u} Z_v (X_u - iY_u), \end{aligned} \tag{11.5}$$

where X_u, Y_u, and Z_u are Pauli operators on the u-th qubit. The Jordan-Wigner transformation requires non-local qubit operations whose weight scales with the system size. These high Pauli-weight operators directly translate to increased gate counts for quantum simulation, and has stimulated various strategies to alleviate the simulation cost when resorting to a Jordan-Wigner mapping [237–240].

11.2.2 Local Fermion-to-Qubit Mappings

There are many other approaches to mapping from fermions to qubits which aim to address the shortcomings of the Jordan-Wigner transformation. For instance, the Bravyi-Kitaev transformation encodes both occupation information (like the Jordan-Wigner transformation) and parity information in such a way that single fermionic operators act non-trivially on at most $\mathcal{O}(\log_2 N)$ qubits [241]. This is in contrast to the linear scaling of the Pauli weight of qubit operators in system size for the Jordan-Wigner transformation. A simpler ternary-tree-based mapping from N fermionic modes to N qubits performs even better, leading to provably minimal Pauli weights in the average case. In this case, a single fermionic operator acts on $\lceil \log_3(2N+1) \rceil$ qubits [267].

One can think of such a mapping as a generalization of the Jordan-Wigner transformation from a 1D chain to tree graphs [268].

Fully local encodings—in the sense that local fermionic operators map to local qubit operators—are possible with the addition of ancilla qubits. An important example is the Bravyi and Kitaev superfast encoding [241] and its generalizations [246, 249]. A multitude of other local mappings have also been developed, often aimed at minimizing the qubits required, while still maintaining local, low Pauli-weight qubit operators [243–245, 247, 248, 250, 251, 269, 270]. These local mappings can generally be understood as generalizations of the toric code [271]. In particular, all known local fermionic encodings are equivalent to the toric code on some lattice up to deformations by a constant-depth circuit of local Cliffords [249, 250]. Fermionic-pair excitations in the local encoding arise as freely deformable strings of Pauli operators on the lattice, and the fermionic anti-commutation relations are enforced via restriction to a particular code subspace of the ancilla-extended Hilbert space. Equivalently, one could view the ancilla qubits as being used to couple to an auxillary gauge field [248, 272]. A given local mapping, therefore, corresponds to a particular "gauge theory" and restricting the simulation to a particular subspace is equivalent to the choice of Gauss's law sector in the corresponding gauge theory. In either view, observe that local fermionic encodings of this sort require the preparation of a toric-code state. Therefore, local fermion-to-qubit mappings trade non-locality in the operators for extra qubits and non-locality in the states. To ensure the simulation proceeds in the allowed subspace of the Hilbert space—that is, that the local and non-local constraints are satisfied— strategies similar to preserving (gauge) symmetries in lattice-gauge-theory simulations [273–280] can be explored.

11.2.3 Custom Fermionic Codes

This work focuses on a particular class of fermion-to-qubit encodings—the so-called custom fermionic encodings—developed by Chien and Whitfield [249] as a generalization of the construction by Setia et al. [281]. These mappings are, in turn, a generalization of the Bravyi and Kitaev superfast encoding [241]. For our purposes, the essential feature of custom fermionic codes is that they allow for a variety of different encodings ranging from local to quasi-local to highly non-local ones. This flexibility permits trading resources like the number of qubits, qubit connectivity, and Pauli weight of simulated operators in an architecture-aware manner. In this [chapter], we will add the parallelizability of the resulting Hamiltonian-simulation algorithm to this list.

Let us briefly review how to construct a custom fermionic code. One can introduce edge operators A_{uv} and vertex operators B_u which are defined as

$$A_{uv} = -i\gamma_u\gamma_v, \tag{11.6}$$

$$B_u = -i\gamma_u\gamma'_u, . \tag{11.7}$$

These operators suffice to generate all parity-preserving fermionic operators in a Hamiltonian of the form Eq. (11.1). Therefore, the Hamiltonian with N fermionic modes

$$H_K = \sum_{uv} \kappa_{uv} a_u^\dagger a_v, \tag{11.8}$$

for some symmetric, real constants $\kappa_{uv} = \kappa_{vu}$, can be expressed as

$$H_K = -\frac{i}{2}\sum_{u<v}\kappa_{uv}(A_{uv}B_v + B_u A_{uv}) - \frac{1}{2}\sum_u \kappa_{uu}B_u \tag{11.9}$$

$$= -\frac{i}{2}\sum_{u<v}\kappa_{uv}A_{uv}(B_v - B_u) - \frac{1}{2}\sum_u \kappa_{uu}B_u, \tag{11.10}$$

up to constant terms that can be ignored.

The interaction set $\mathcal{T}$ can now be defined as the set of all terms with non-zero coefficients in the re-expressed Hamiltonian. Furthermore, an interaction graph $\Gamma = \{V_\Gamma, E_\Gamma\}$ can be defined with vertices corresponding to each fermionic mode and an edge joining any pair of vertices (u, v) such that the edge operator $A_{uv}B_v$ belongs to $\mathcal{T}$. For instance, for the Hamiltonian in Eq. (11.9) with $\kappa_{uv} \neq 0$, the interaction set is

$$\mathcal{T} = \{A_{uv}B_v\}_{u\neq v} \cup \{B_u\}_{u\in\{1,\cdots N\}}, \tag{11.11}$$

and the corresponding interaction graph is a complete graph K_N on N vertices. In what follows, it is assumed without loss of generality that Γ is connected, as if Γ is disconnected, one is dealing with two physically independent systems, and can consider the connected case on each system separately.

From Eq. (11.3) and Eqs. (11.6) and (11.7), one can show that the edge and vertex operators

obey the following relations

$$
\begin{aligned}
&B_u^\dagger = B_u, && A_{uv}^\dagger = A_{uv}, \\
&B_u^2 = A_{uv}^2 = 1, && [B_u, B_v] = 0, \\
&A_{uv} = -A_{vu}, && A_{uv} B_w = (-1)^{\delta_{uw}+\delta_{vw}} B_w A_{uv}, \\
&A_{uv} A_{wx} = (-1)^{\delta_{uw}+\delta_{ux}+\delta_{vw}+\delta_{vx}} A_{wx} A_{uv}, \\
&i^{|C|} \prod_{\nu=1}^{|C|} A_{c_\nu c_{\nu+1}} = I, && \qquad (11.12)
\end{aligned}
$$

where in the final equality C is any cycle in Γ specified via an ordered list of vertices $C = \{c_1, c_2, ..., c_{|C|}, c_1 \equiv c_{|C|+1}\}$ for $c_\nu \in V_\Gamma$ with only the final vertex repeated, and, therefore, the product is over all edge operators in the cycle.

Next, a second graph can be introduced, the so-called system graph $\Sigma = \{V_\Sigma, E_\Sigma\}$. A valid system graph is any undirected, connected graph with vertex set $V_\Sigma = V_{\text{phys}} \cup V_{\text{virt}}$ equipped with a bijective mapping $M : V_\Gamma \to V_{\text{phys}}$—that is, $|V_\Gamma| = |V_{\text{phys}}|$. The subscripts are shorthand for physical vertices and virtual vertices. The physical vertices correspond to physical fermionic modes in the interaction graph and the virtual vertices (if they exist) correspond to additional auxillary fermionic modes that can be freely introduced. One can envision constructing the system graph from the interaction graph by adding an arbitrary number of virtual vertices and adding or removing any edges so long as the final graph is connected. Note that the condition that the graph is connected implies that any two vertices that were connected before are still connected via physical or auxillary vertices. This connectivity condition is sufficient for one to implement any interaction terms in the Hamiltonian, see e.g., Eq. (11.20) below. Consequently, the edge set

of the system graph can be completely different from that of the interaction graph.

An encoding of a fermionic system on such a graph Σ consists of associating with each vertex $v \in V_\Sigma$ a set of

$$n_v := \lceil d(v)/2 \rceil \tag{11.13}$$

qubits, where $d(v)$ is the degree of vertex v. A set of $2n_v$ Pauli operators $\{\tilde{\gamma}_v^1, \tilde{\gamma}_v^2, \cdots, \tilde{\gamma}_v^{2n_v}\}$ can then be defined on these qubits. In the following, these operators are referred to as local Majoranas. Note that we have introduced the convention of using subscripts $\{u, v, w, \cdots\}$ to index vertices of Σ and superscripts $\{i, j, k, \cdots\}$ to index local quantities such as enumerations of the local Majoranas or indices of internal qubits. Furthermore, subscripts $\{\nu, \mu, \cdots\}$ are used in various places for indexing generic sets.

The local Majoranas can be any choice of operators that satisfy the following conditions:

1. They obey the Majorana-operator[1] properties including anti-commutation relations with other local Majoranas defined on the vertex. Furthermore, they must commute with the local Majoranas on other vertices. That is,

$$\tilde{\gamma}_v^{k\dagger} = \tilde{\gamma}_v^k, \quad \{\tilde{\gamma}_v^j, \tilde{\gamma}_v^k\} = 2\delta_{jk}, \quad [\tilde{\gamma}_u^j, \tilde{\gamma}_v^k] = 0 \text{ for } u \neq v. \tag{11.14}$$

2. They generate the full Pauli group on the n_v qubits associated with $v \in V_\Sigma$.

Any explicit choice for the local Majoranas can be mapped to any other via a Clifford circuit acting on the qubits associated with that vertex [249]. Most simply, one could consider encoding the local Majoranas via a Jordan-Wigner transformation. That is, given some enumeration of the

[1] Note that these correspond to both types of γ and γ' operators defined in Eq. (11.3).

qubits in a vertex v, one has

$$\{\tilde{\gamma}_v^1, \tilde{\gamma}_v^2, \tilde{\gamma}_v^3, \tilde{\gamma}_v^4, \tilde{\gamma}_v^5, \tilde{\gamma}_v^6, \cdots\} \longrightarrow \{X_v^1, Y_v^1, Z_v^1 X_v^2, Z_v^1 Y_v^2, Z_v^1 Z_v^2 X_v^3, Z_v^1 Z_v^2 Y_v^3, \cdots\}. \tag{11.15}$$

It is straightforward to verify that this choice satisfies the conditions above. One could also use other encodings—for instance, Fenwick trees [243, 281] or ternary trees [267]. This work concerns only the case of a Jordan-Wigner encoding of the local Majoranas. Note, however, that the same techniques and many of the qualitative results will apply similarly to these other choices.

Once the local Majoranas are specified, each local Majorana can be associated with an edge of that same vertex. That is, both the local Majoranas associated with a vertex and the edges connecting the vertex to its neighbors are enumerated in Σ. The j-th local Majorana corresponding to a vertex is then associated with the edge connecting it to its j-th neighbor. Therefore, each edge $e \in E_\Sigma$ has two associated local Majoranas—one at each endpoint. Given such a choice, encoded edge operators acting on qubits can be defined as

$$\tilde{A}_{uv} = \epsilon_{uv} \tilde{\gamma}_u^{\xi_u(v)} \tilde{\gamma}_v^{\xi_v(u)}, \tag{11.16}$$

where v is the $\xi_u(v)$-th neighbor of u, u is the $\xi_v(u)$-th neighbor of v, and the Levi-Civita tensor ϵ_{uv} is defined with respect to an arbitrary choice of orientation for each edge in Σ. In particular, we let $\epsilon_{uv} = 1$ if u is the head of the oriented edge (u, v) and $\epsilon_{uv} = -1$ if u is the tail. This choice of enumerating the edges of each vertex u, as specified by picking $\xi_u(v)$ for each neighbor v of u, will become important later in Sec. 11.4.3 when discussing the strong-coloring problem.

Furthermore, the vertex operator on qubits can be encoded as

$$\tilde{B}_u = (-i)^{n_u} \prod_{j=1}^{2n_u} \tilde{\gamma}_u^j. \tag{11.17}$$

One can verify that the choices of encodings in Eqs. (11.16) and (11.17) satisfy all but the final loop condition in Eq. (11.12). To satisfy the loop condition, it is necessary to restrict the simulation to the subspace of the total Hilbert space that does satisfy this condition. In the context of quantum error correction, this space is the codespace stabilized by the loop operators $\tilde{L}$ around cycles C on Γ, defined by

$$\tilde{L}(C) = i^{|C|} \prod_{j=1}^{|C|} \tilde{A}_{c_j c_{j+1}}. \tag{11.18}$$

As the encoded edge and vertex operators commute with the loop operators, once a state is initialized in the code subspace, the simulation remains in that subspace assuming no algorithmic or experimental errors. For considerations regarding boundary conditions and fermionic parity, see Refs. [246, 249]. Here, we consider only open boundary conditions for simplicity, but other boundary conditions can be analyzed within the framework of this work as well.

Once in the code subspace, the mapping from fermionic edge and vertex operators can be performed to qubit edge and vertex operators, $A_{uv} \to \tilde{A}_{uv}$ and $B_u \to \tilde{B}_u$. This completes the mapping from a fermionic Hamiltonian H to a qubit Hamiltonian $\tilde{H}$,

$$\mathcal{M} : H = \sum_\nu \kappa_\nu h_\nu \longrightarrow \tilde{H} = \sum_\nu \kappa_\nu \tilde{h}_\nu, \tag{11.19}$$

where κ_ν are constants related to the original coupling coefficients and the h_ν and $\tilde{h}_\nu$ are products of edge and vertex operators on fermionic Majorana modes and qubits, respectively.

This mapping $\mathcal{M}$ is not unique—the exact form of each $\tilde{h}_\nu$ depends not only on the particular system graph Σ that determines the mapping, but also on the paths through the system graph chosen to simulate each corresponding $\tilde{h}_\nu$, as well as the choice of encoding of the local Majoranas. In particular, one does not necessarily need to have a direct edge $(u, v) \in E_\Sigma$ to implement $\tilde{A}_{uv}$. An edge operator between two modes u and v not directly connected is given by a product of edge operators along any path $P_{uv} = \{p_1 = u, p_2, \cdots, p_{|P_{uv}|} = v\}$ connecting the two modes. That is,[2]

$$\tilde{A}_{uv} = \prod_{\nu=1}^{|P_{uv}|-1} \tilde{A}_{p_\nu p_{\nu+1}}. \tag{11.20}$$

Therefore, given the assumption that Σ is connected, all edge operators in the qubit Hamiltonian $\tilde{H}$ can be implemented by choosing any path between the relevant vertices. Again, these path choices are not unique. While the precise details depend on these choices, it is always true that each $\tilde{h}_\nu$ is a string of Pauli operators on qubits. That is $\tilde{h}_\nu = \{X, Y, Z, I\}^{\otimes n}$, where n is the number of qubits $\tilde{h}_\nu$ acts on. Importantly, whether an operator $\tilde{A}_{uv}$ can be implemented directly or must be implemented via a path of such operators through the system graph, it obeys all the same relations given in Eq. (11.12).

These choices do matter, however. In particular, recall virtual vertices are allowed in the system graph which, at the cost of more qubits, enable more choices of paths between different physical vertices. This tradeoff between more qubits and more direct (and correspondingly, more

[2]Observe that Eq. (20) slightly overloads the notation $\tilde{A}_{uv}$, as strictly speaking, the $\tilde{A}_{uv}$ operators on the left- and right-hand sides of the equation have a slightly different meaning. In particular, one should distinguish between $\tilde{A}_{uv}$ that, given the system graph, can be directly implemented as in Eq. (11.16), and those that cannot and must be implemented via a product of such operators as in Eq. (11.20). The meaning should be clear from the context. Note that tilde operators always denote those acting on qubits and not on fermionic modes.

local) paths for implementing the required Pauli operators is the essential tension in regards to optimizing a Hamiltonian-simulation algorithm for a fermionic system in this construction.

11.2.4 Prior Work on Optimizing System Graphs

Some of the tradeoffs implied by the custom fermionic encoding have already been explored. In particular, Ref. [248] discusses how the flexible framework of custom fermionic codes allows for designing fermionic encodings suited to particular qubit architectures by balancing the number of qubits required for an encoding with the Pauli weight of the resulting operators. In one limit, where the system graph is a line graph, one recovers the Jordan-Wigner transformation. By adding qubits and connectivity in the system graph, one can reduce the Pauli weight of the resulting operators, obtaining local or quasi-local encodings. This exact tradeoff was explored in detail for a variety of different system graphs in Ref. [249] for the 2-body SYK model, which has all-to-all coupled fermions.

Observe that the tradeoff between Pauli weight of operators and numbers of qubits and qubit connectivity is directly related to the properties of the system graph Σ. For instance, the number of qubits $Q(\Sigma)$ is directly determined by the degree of the vertices in Σ as

$$Q(\Sigma) = \sum_{v \in V_\Sigma} n_v = \sum_{v \in V_\Sigma} \lceil d(v)/2 \rceil. \tag{11.21}$$

As is shown in the following, this tight connection between graph-theoretic properties and resource counts holds even for more complicated properties of the fermion-to-qubit encoding and the resulting Hamiltonian simulation.

11.3 Parallelization and Path Coloring

11.3.1 Notions of Parallelization

In this work, a new possibility for optimization afforded by the flexibility of the custom fermionic codes is considered: parallelization. We use the term parallelization instead of the related concept of circuit depth because our analysis concerns a slightly higher level of abstraction than the particular circuit-level implementation of a Hamiltonian-simulation algorithm. It is assumed that provided two Pauli strings $\tilde{h}$ and $\tilde{h}'$ act non-trivially on disjoint sets of qubits, they may be implemented simultaneously in a quantum-simulation algorithm. Therefore, the goal is to minimize the number of steps required to implement the full set of Pauli operators in the interaction set $\tilde{\mathcal{T}} = \{\tilde{h}_\nu\}$. If one can choose paths on the system graph for the implementation of the required Pauli strings that minimizes collisions between those paths and orders the implementation of these operators in an optimal way, one can minimize the circuit depth for implementing the relevant operators. See Fig. 11.1b for an example. This formulation is especially relevant to quantum simulation via product formulas, in which these Pauli operators are directly implemented for each Trotter step.

It is important to note that our approach focuses solely on grouping the Pauli strings so as to minimize the number of steps to implement the full interaction set. It is well established that the choice of ordering terms can impact the Trotter error, which in turn changes the overall circuit depth of the Hamiltonian-simulation algorithm required to achieve a certain error tolerance [237, 254, 280, 282, 283]. While such effects are not considered in this work, when applying the parallelization techniques here to a particular problem of interest, one should view parallelization

of the sort considered here as one piece of a many-faceted optimization.

11.3.2 Graph Coloring

The parallelization problem defined above can be formalized using the notion of path coloring on a graph. This problem also arises in other similar networking and scheduling problems [284]. We begin by reviewing the ideas of colorings on graphs and then describe how the parallelization problem may be formulated in these terms.

Consider a graph $G = \{V, E\}$. A vertex coloring on G is a mapping $\mathcal{C} : V \rightarrow C$ where C is a set of so-called colors or wavelengths. A valid coloring $\mathcal{C}$ is one such that no adjacent vertices in G are assigned the same color. The smallest number of colors required to (vertex) color a graph is called its chromatic number, $\chi(G)$. For a general graph, finding $\chi(G)$ is NP-hard [285]. However, both bounds and effective heuristic algorithms exist. A simple and useful upper bound is

$$\chi(G) \leq \max_{v \in V} d(v) + 1, \tag{11.22}$$

where $d(v)$ is the degree of vertex v [286]. A coloring satisfying this bound can be obtained in polynomial time in the number of vertices using the greedy coloring algorithm presented below.

Algorithm 4 Greedy Coloring

```
1: function GreedyColor(G = {V, E}, C)
2:     for each v ∈ V do
3:         Assign v the first color c ∈ C not used by any of its neighbors
4:     end for
5: end function
```

If G is a simple, connected graph, but is neither a complete graph nor an odd cycle, then

this bound is improved to

$$\chi(G) \leq \max_{v \in V} d(v), \tag{11.23}$$

and the greedy coloring algorithm will still satisfy this bound [286] .

The bound on $\chi(G)$ can be still lowered by the clique number $\omega(G)$ of the graph—that is, the size of the largest clique in G, where a clique is a complete induced subgraph of G. Therefore, the size of any clique $W(G)$ is also a valid lower bound. This gives

$$\chi(G) \geq \omega(G) \geq |W(G)|. \tag{11.24}$$

A related problem to the vertex-coloring problem is the path-coloring problem. As previously described, this will be our graph-theoretic problem of interest when formalizing the problem of optimally parallelizing the implementation of the Pauli strings that result from a custom fermionic code. In this problem, given a set of paths $\mathcal{P}$ in the graph G, one seeks to color the paths such that no two paths which share a vertex in G receive the same color and that a minimum number of colors is used to color all the paths.[3]

The path coloring problem can be mapped to a vertex coloring problem on a different graph called the conflict graph $\Pi(\mathcal{P})$ of the set of paths $\mathcal{P}$. The conflict graph has a vertex set $V_{\Pi(\mathcal{P})} = \mathcal{P}$ and edge set $E_{\Pi(\mathcal{P})} = \{(q, p) \,|\, q, p \in \mathcal{P},\ q \cap p \neq \emptyset\}$. Therefore, the path coloring problem is also NP-hard.

[3]Note that typically in the literature, this problem is defined such that no paths can share an *edge* instead of a vertex. Our alternative definition is due to the particular context in which path coloring is applied.

11.3.3 Conflict Graphs for Parallelizability

Having defined the path coloring, the connection to parallelizability becomes clear. Given a system graph Σ, one seeks to efficiently implement the interactions in the interaction graph Γ as specified by the interaction set $\tilde{\mathcal{T}}$. For any interaction $\tau \in \tilde{\mathcal{T}}$, one requires a choice of path p through Σ joining the relevant vertices for the interaction τ. Choosing a particular path for each interaction gives a path set $\mathcal{P} = \{p_\tau\}_{\tau \in \tilde{\mathcal{T}}}$ with $|\mathcal{P}| = |\tilde{\mathcal{T}}|$. Given a choice of $\mathcal{P}$, one then seeks to determine the degree of parallelization via a coloring of a conflict graph $\Pi(\mathcal{P})$. We construct two different versions of the conflict graph, corresponding to what we dub the *weak coloring problem* and the *strong coloring problem*. The latter considers the internal qubit structure of the vertices of the system graph as specified by the custom fermionic encoding; the former does not. These problems can be formally specified as follows:

Definition 11.3.1 (The weak coloring problem)**.** *Given a system graph Σ and a path set $\mathcal{P}$ on Σ specifying the implementation of a set of interactions $\tilde{\mathcal{T}}$, construct a conflict graph $\Pi(\mathcal{P})$, whose vertex set is $\mathcal{P}$ and whose edge set is $E_{\Pi(\mathcal{P})} = \{(q, p) \,|\, q, p \in \mathcal{P},\, q \cap p \neq \emptyset\}$. The weak coloring problem is to optimally color $\Pi(\mathcal{P})$.*

The chromatic number χ resulting from the weak coloring problem corresponds to the minimum number of steps required to implement all the interactions $\tau \in \tilde{\mathcal{T}}$, where it is assumed that interactions that require disjoint sets of vertices of the system graph may be implemented in parallel.

Definition 11.3.2 (The strong coloring problem)**.** *Given a system graph Σ and a path set $\mathcal{P}$ on Σ specifying the implementation of a set of interactions $\tilde{\mathcal{T}}$, construct a conflict graph $\Pi(\mathcal{P})$, whose*

vertex set is $\mathcal{P}$ and whose edge set is $E_{\Pi(\mathcal{P})} = \{(q,p) \,|\, q,p \in \mathcal{P},\, Q(q) \cap Q(p) \neq \emptyset\}$, where $Q(p)$ gives the set of internal qubits required to implement to the path $p \in \mathcal{P}$. The strong coloring problem is to optimally color $\Pi(\mathcal{P})$.

Note that $Q(p)$ in the definition of the strong coloring problem depends on the local Majorana encoding (i.e., Jordan-Wigner, Fenwick trees, etc.) in the system-graph vertices. This work only considers the Jordan-Wigner encoding of local Majoranas. The following section will provide an explicit description of $Q(p)$ in this setting. Here, the resulting chromatic number χ corresponds to the minimum number of steps required to implement all the interactions $\tau \in \tilde{\mathcal{T}}$, where it is assumed that interactions that require disjoint sets of qubits may be implemented in parallel.

Compared to the weak coloring problem, the definition of parallelizability in the strong coloring problem is connected more directly to the qubit architecture and to the circuit depth of the Hamiltonian-simulation algorithm; the weak coloring problem has the advantage of being somewhat more abstracted and easier to work with. Both schemes are considered in this work. Observe also that the definitions of the weak and strong coloring problems take in both the system graph and a particular choice of path for each interaction in the interaction set. This choice of paths, as specified by the set $\mathcal{P}$, is not unique, of course, and to truly maximize the amount of parallelization, one must both pick the optimal path set $\mathcal{P}$ and optimally color the resulting conflict graph. Naturally, this is a very difficult problem. In particular, the following result can be stated:

Fact 11.3.1. *Optimally parallelizing the implementation of an interaction set $\tilde{\mathcal{T}}$—in either the weak or strong coloring sense—is NP-hard.*

Proof. Suppose there exists an oracle that, given an interaction list and a system graph, returns the solution set of paths $\mathcal{P}$ that will enable the creation of a conflict graph $\Pi(\mathcal{P})$ with the minimum chromatic number. Given $\mathcal{P}$ via this oracle, one is left with a graph coloring problem on $\Pi(\mathcal{P})$, which is known to be NP-hard [285]. □

The oracle invoked in the proof above is quite powerful in its own right. Therefore, outside some analytically accessible examples, one need to turn to heuristic algorithms to address the selection of the path set $\mathcal{P}$ and the solving of the resulting weak and strong coloring problems. The full procedure of defining and solving the weak and strong coloring problems starting from the qubit architecture is shown in Fig. 11.1 for a simple example. Each step of this process will be described in detail in the following sections.

11.4 Analytic Results

11.4.1 The Hamiltonian

For the purposes of exploring the weak and strong coloring problems for a variety of system graphs both analytically and numerically, we shall make use of an explicit choice of a fermionic Hamiltonian as a minimal example. In particular, let us consider an all-to-all Hamiltonian with two-mode interactions given by Eq. (11.8). This Hamiltonian can be expressed in terms of edge and vertex operators as in Eq. (11.9). Assuming all coefficients are non-zero, the interaction graph Γ for this problem will be the complete graph on N vertices, K_N, and the interaction set $\mathcal{T}$ is given by Eq. (11.11). Note that $|\mathcal{T}| = N^2$.

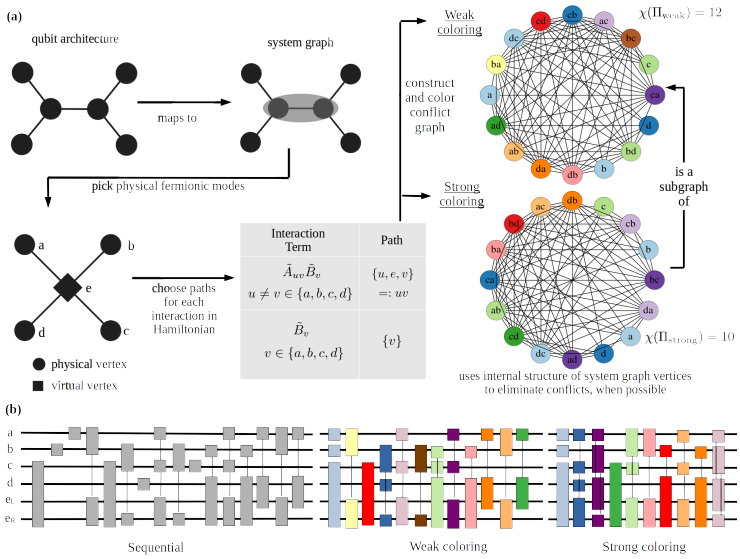


Figure 11.1: (a) An overview of the full procedure of defining and solving the weak and strong coloring problems for parallelizing a Hamiltonian simulation of fermions. There are many stages for optimization: the choice of system graph, the choice of physical fermionic modes, the choice of paths linking those modes, and the coloring algorithm. While these choices are straightforward in this small example, for general problems, the design space is extremely large. This work focuses on the last two steps, which is an NP-hard optimization problem. Here, the conflict-graph vertices are labeled by the physical vertices of the system graph involved in the interaction. That is, $\tilde{A}_{uv}\tilde{B}_v$ is labeled by uv and $\tilde{B}_v$ is labeled by v for all u, v. Note that the difference between the weak and strong coloring problems in this example is in the ability of the strong coloring scheme to route through the virtual vertex e to implement the ad (and da) path simultaneously with bc (and cb) without any conflict, hence a lower chromatic number compared with the weak coloring scheme. This corresponds to enumerating the edges of vertex e as $\{ea, ed, eb, ec\} \mapsto \{1, 2, 3, 4\}$. (b) Corresponding circuit diagrams for ordering the Pauli strings according to the sequential strategy and via the weak and strong coloring problems. Here e_L and e_R label the left and right internal qubits of vertex e of the system graph, respectively. Colors match those in the corresponding conflict graphs and gates of the same color are implemented simultaneously.

11.4.1.1 Extensions to Other Models

The Hamiltonian in Eq. (11.8) is closely related to long-range fermionic systems, such as the SYK model [287, 288]. To get exact results for specific Hamiltonians of interest (with or without long-range interactions), one can use the algorithm presented here to heuristically solve the weak and strong coloring problems for the relevant system graph. Another example of a

minimal fermionic Hamiltonian is that with only nearest-neighbor hopping on a square lattice. This case will be studied later in Sec. 11.5.4.

A generalization of our results worthy of particular emphasis is the case of Hamiltonians with k-body interactions for $k > 2$. For instance, terms such as $a_u^\dagger a_v^\dagger a_w a_x$ yield, amongst other things, terms of the form $\tilde{A}_{uv}\tilde{A}_{wx}$ when expressed as edge and vertex operators. Quite clearly, implementing such a term in terms of Pauli operators requires two simultaneous paths through Σ: one from u to v and one from w to x. The path set $\mathcal{P}$ can now be viewed as a multiset of paths, with each element of $\mathcal{P}$ (now potentially a set of paths) mapping to a vertex of the conflict graph. From there, construction of the conflict graph proceeds as usual.

11.4.2 Rules for Strong Coloring

In this section, the rules for constructing the conflict graph for the strong coloring problem given the interactions in Eq. (11.11) will be developed, under the assumption that local Majoranas are encoded via a Jordan-Wigner transformation on the internal qubits of each vertex of the system graph. This allows to abstract the problem of determining conflicts between paths to one about the properties of the system graph under consideration.

To begin, recall that each vertex $u \in V_\Sigma$ contains $n_u = \lceil d(u)/2 \rceil$ qubits. Under a Jordan-Wigner encoding, one can imagine expanding each vertex of the system graph into a line graph of n_u vertices, where each new vertex is associated with two edges of the original vertex as depicted in Fig. 11.2.[4] Any local Majorana operator on the vertex u will induce a Jordan-Wigner string on some subset of these internal vertices. The first task is to identify what precisely these strings are for the four possible (types of) operators acting on the vertex u: $\tilde{B}_u$, $\tilde{A}_{xu}$, $\tilde{A}_{xu}\tilde{A}_{uy}$, and $\tilde{A}_{xu}\tilde{B}_u$

[4]Note when $d(u)$ is odd, one of these internal vertices has only one external edge.

where $x, y \in V_\Sigma$ are arbitrary neighbors of u in Σ. Observe that determining the qubits needed within vertex u to implement the operator $\tilde{A}_{xu}\tilde{B}_u$ is equivalent to $\tilde{A}_{ux}\tilde{B}_u$ since $\tilde{A}_{ux} = -\tilde{A}_{xu}$.

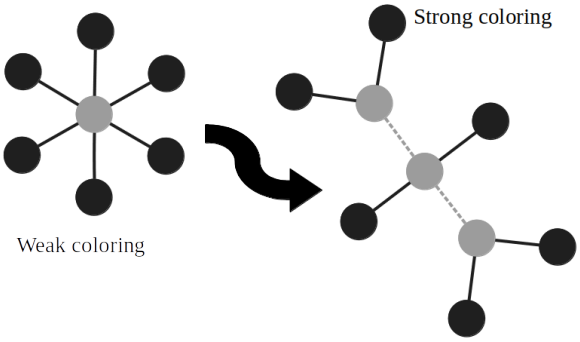


Figure 11.2: The central gray vertex is u and black vertices are its neighbors. This is just notation and no formal coloring has been performed yet. When constructing the conflict graph for strong coloring under a Jordan-Wigner encoding of the local Majoranas, it is useful to think of each vertex $u \in V_\Sigma$ as being expanded to a line graph of $n_u = \lceil d(u)/2 \rceil$ internal vertices, each connected to two of the original edges of u, where $d(u)$ is the degree of u. Different interaction types on vertex u induce different Jordan-Wigner strings on these internal vertices as summarized in Tab. 11.1 and depicted in Fig. 11.3.

First consider a vertex operator $\tilde{B}_u$. Given a Jordan-Wigner encoding of the local Majoranas, one has immediately from Eqs. (11.15) and (11.17) that in terms of Pauli operators

$$\tilde{B}_u = \bigotimes_{j=1}^{n_u} Z_u^j, \tag{11.25}$$

where Z_u^j is the Pauli-Z operator acting on qubit j of vertex u. Therefore, a vertex operator uses all qubits on that vertex (see Fig.11.3-a), affording no possibility for improved parallelization via strong coloring when implementing these terms.

On the other hand, operators of the form $\tilde{A}_{ux}$ acting on vertex u do not use all the qubits. Such operators appear when vertex u is a physical vertex and one is seeking to implement an interaction of the form $\tilde{A}_{uv}\tilde{B}_v$ between vertex u and some other vertex v via a path through Σ that starts with the edge from u to $x \in V_\Sigma$. It will be useful to introduce one more piece of

notation. In particular, define

$$a_u(x) := \left\lceil \frac{\xi_u(x)}{2} \right\rceil, \tag{11.26}$$

so that it can be compactly stated that the first $a_u(x)$ qubits of vertex u are "active" when implementing the local Majorana operator $\tilde{\gamma}_u^{\xi_u(x)}$. This follows immediately from the Jordan-Wigner encoding of these local Majoranas, where one should recall that the custom fermionic code requires an enumeration of both the internal vertices of u and of its edges. Given a fixed choice of enumeration, x is the $\xi_u(x)$-th neighbor of u. Therefore, from Eq. (11.16), one immediately finds that the operator $\tilde{A}_{ux}$ makes use of the first $a_u(x)$ qubits of vertex u (as well as the first $a_x(u)$ qubits of vertex x), see Fig. 11.3-b.

Next consider an operator of the form $\tilde{A}_{xu}\tilde{A}_{uy}$ acting on vertex u. Such operators occur when vertex u is an intermediate vertex along a path implementing an interaction between two physical fermionic modes. Just like $\tilde{A}_{ux}$, these operators also do not require the use of all qubits in u. Individually, $\tilde{A}_{xu}$ and $\tilde{A}_{uy}$ make use of the first $a_u(x)$ and the first $a_u(y)$ qubits in u, respectively. However, there are cancellations since the operators both act with Pauli-Z operators on the first $a_u(x) - 1$ and $a_u(y) - 1$ qubits, respectively. Such cancellations of Jordan-Wigner strings are reminiscent of the cancellations of such strings in sequential Trotter-Suzuki steps [237]. The net result is that only the qubits between $\min\{a_u(x), a_u(y)\}$ and $\max\{a_u(x), a_u(y)\}$ are used, see Fig. 11.3-c.

Finally, consider an operator of the form $\tilde{A}_{xu}\tilde{B}_u$ acting on vertex u. These operators arise at the starting and ending vertices of a path. Once again, there are cancellations in the Pauli-Z operators required to implement the two sub-operators. In particular, $\tilde{A}_{xu}\tilde{B}_u$ acts on the last $n_u - a_u(x) + 1$ qubits of vertex u, see Fig. 11.3-d. Tab. 11.1 summarizes the results in this section.

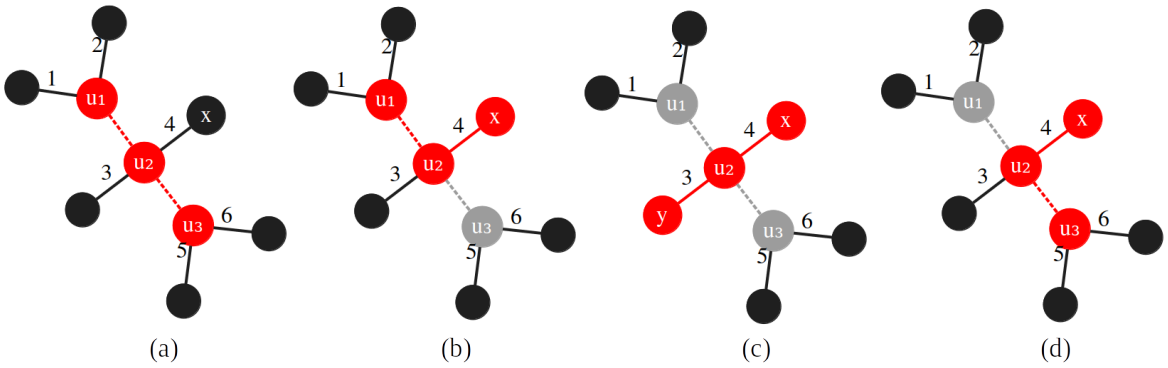


Figure 11.3: Examples of the internal qubits of vertex u (enumerated top to bottom as u_1, u_2, u_3) that are required to implement the operators (a) $\tilde{B}_u$, (b) $\tilde{A}_{ux}$, (c) $\tilde{A}_{xu}\tilde{A}_{uy}$, and (d) $\tilde{A}_{xu}\tilde{B}_u$. Active qubits and input/output system-graph vertices are marked with red. The dashed lines denote internal edges. Again, the coloring is the notation and no formal graph coloring is assumed here. The choice of edge enumeration is marked.

.

Term	Active qubits in vertex u
$\tilde{B}_u$	All n_u qubits
$\tilde{A}_{ux}$	Qubits 1 to $a_u(x)$
$\tilde{A}_{xu}\tilde{A}_{uy}$	Qubits $\min\{a_u(x), a_u(y)\}$ to $\max\{a_u(x), a_u(y)\}$
$\tilde{A}_{xu}\tilde{B}_u$	Qubits $a_u(x)$ to n_u

Table 11.1: Rules for determining internal qubits used by the various terms that arise in simulating the Hamiltonian in Eq. (11.9) using a Jordan-Wigner encoding of the local Majoranas. Recall $a_u(v) := \lceil \frac{\xi_u(v)}{2} \rceil$ is the number of "active" qubits when implementing $\tilde{\gamma}_u^{\xi_u(v)}$, and v is the $\xi_u(v)$-th neighbor of u.

11.4.3 Limits of Weak and Strong Coloring

In this section, two simple system graphs will be studied: a star graph S_N with N physical vertices all joined to a central virtual vertex, and a complete graph K_N consisting of N physical vertices. These examples are limiting cases for both the weak and strong coloring problems. In addition, they allow for straightforward analytic calculations and enable an understanding of the essential conceptual features of the two types of coloring problems. This understanding

will be leveraged to determine what properties of a system graph allow for the greatest possible improvement from using strong coloring as opposed to weak coloring. An example with such an extreme separation will be constructed at the end of the section.

11.4.3.1 Star Graph

The star graph $\Sigma = S_N$ of N physical vertices all coupled to a central virtual vertex is the worst-case limit for parallelization as there is a single bottleneck vertex through which all paths for the $N(N-1)$ two-mode interactions must pass. It helps to refer back to Fig. 11.1 to visualize the procedure for the minimal case of $\Sigma = S_4$. As seen in that figure, the corresponding conflict graph for the weak coloring problem has a complete subgraph $K_{N(N-1)}$ consisting of all vertices that correspond to two-mode interactions, which sets a lower bound on the chromatic number of the conflict graph. No additional colors are needed to color the one-mode interaction vertices as the vertex operators $\{\tilde{B}_w\}_{w \notin \{u,v\}}$ can be implemented simultaneously with the $\tilde{A}_{uv}\tilde{B}_v$ operators. Therefore, the chromatic number for weak coloring is

$$\chi_{\text{weak}}(\Pi(S_N)) = N(N-1). \tag{11.27}$$

For strong coloring, it turns out that the even and odd N cases must be addressed separately. First, consider N even. Expanding all vertices of the system graph as in Fig. 11.2, the physical vertices remain unexpanded, whereas the central virtual vertex u expands to a line graph of $n_u = N/2$ vertices $\{u_1, \cdots u_{n_u}\}$, where each $u_j \in u$ has two edges that each connect to one neighbor of u in Σ. The two-mode interactions involving vertex u induce eight Jordan-Wigner strings between each pair of these neighbors—two for each of the four choices of pairs

of physical-neighbor vertices connected to a given pair (u_μ, u_ν), $\mu \neq \nu$. They also induce single-vertex "strings" for each u_ν for the two-mode interactions between physical vertices that are both neighbors of that vertex.

Minimizing the number of steps to avoid overlaps of these strings is straightforward: starting with u_1, implement all interactions that induce Jordan-Wigner strings originating from u_1 while simultaneously implementing the interaction that induces the longest non-overlapping Jordan-Wigner string originating from u_{n_u}. This takes $8n_u = 4N$ steps. At this point, all interactions involving u_1 and u_{n_u} have been implemented. Therefore, ignore those vertices and repeat the same procedure on the remaining $n_u - 2$ internal vertices. Keep repeating this procedure until all two-mode interactions have been implemented. For $N > 2$, implementing the single-mode interactions requires no extra steps as most physical vertices are unused for any given step, giving many opportunities to implement these interactions simultaneously with a given two-mode interaction. The net result (for $N > 2$, even) is

$$\chi_{\text{strong}}(\Pi(S_N), N \text{ even}) = \begin{cases} 8\sum_{\mu=1}^{n_u/2}(2\mu) - 6 = \frac{N^2}{2} + 2N - 6, & \frac{N}{2} \text{ even}, \\ \\ 8\sum_{\mu=1}^{(n_u+1)/2}(2\mu - 1) - 6 = \frac{N^2}{2} + 2N - 4, & \frac{N}{2} \text{ odd}. \end{cases} \tag{11.28}$$

Note that number six is subtracted from the sum to correct for over-counting in the final step which only involves two-mode interactions between physical vertices that share an internal vertex. See Fig. 11.4 for an example of this construction for $N = 8$.

Now consider N odd. Expanding the vertices of the system graph, the result is identical to the even-N case except u_{n_u} has only one edge joining it to a physical vertex. This implies

that the two-mode interactions involving u_{n_u} induce only four Jordan-Wigner strings between the internal vertices of u instead of four. One can implement these interactions first while simultaneously implementing four of the eight interactions that induce the longest possible non-overlapping Jordan-Wigner strings starting from u_1. This takes $4n_u = 2(N+1)$ steps. At this point, all interactions that involve u_{n_u} are implemented, but four interactions are yet to be implemented for each induced Jordan-Wigner string involving u_1. These can be implemneted while simultaneously implementing four of the eight interactions that induce the longest possible non-overlapping Jordan-Wigner strings starting from u_{n_u-1}. This takes $4(n_u-1)$ steps. This staggered approach can be continued—implementing four of the eight interactions that induce a particular Jordan-Wigner string starting from a given internal vertex in each stage of the procedure—until all interactions are implemented. This gives

$$\chi_{\text{strong}}(\Pi(S_N), N\,\text{odd}) = 4\sum_{\mu=1}^{n_u-1}(\mu+1) - 2 = \frac{N^2}{2} + 2N - \frac{9}{2}, \tag{11.29}$$

where number two is subtracted from the sum to correct for over-counting in the final step of this procedure. See Fig. 11.4 for an example of this construction for $N = 7$.

The constructions yielding Eqs. (11.28) and (11.29) are optimal. In particular, observe that in each step of these constructions, a path is implemented that passes through the central internal vertex $u_{\lceil n_u/2 \rceil}$ of the central vertex u. The set of paths that go through this vertex form a complete induced subgraph of the conflict graph $\Pi_{\text{strong}}(S_N)$ of maximum size. That is, the induced subgraph of this set of vertices in the conflict graph form the largest-size clique. Since one of these paths is implemented in every step of the construction, the corresponding coloring of $\Pi_{\text{strong}}(S_N)$ saturates the clique-number lower bound on the chromatic number from Eq. (11.24).

Therefore, the constructions are optimal.

11.4.3.2 Complete Graph

The complete graph $\Sigma = K_N$ of N all-to-all connected physical vertices is the opposite limit of the star graph. There are no bottlenecks in implementing paths between any pair of vertices as all two-body interactions are directly implementable. Obviously, these direct connections are the optimal paths.

For the weak coloring problem, one can simultaneously implement $\lfloor N/2 \rfloor$ of the $N(N-1)$ two-mode interactions. Consequently, all two-mode interactions can be implemented in $N(N-1)/\lfloor N/2 \rfloor$ steps. In addition, there are N one-mode terms to implement. For even N, these can all be implemented in one step after doing the two-mode interactions. For odd N, there is always an unused vertex for any step where two-mode interactions are implemented, and therefore the one-mode interactions can be done while doing the two-mode interactions. Therefore, the number of steps required for each case is

$$\chi_{\text{weak}}(\Pi(K_N)) = \begin{cases} 2N-1, & N \text{ even}, \\ 2N, & N \text{ odd}. \end{cases} \tag{11.30}$$

To gain some intuition about the conflict graph, we can also arrive at Eq. (11.30) by observing that the conflict graph $\Pi_{\text{weak}}(K_N)$ consists of N pairwise-overlapping complete subgraphs K_{2N-1}, as depicted in Fig. 11.5 for the case of $N = 4$. This structure arises because for any $v \in V_\Sigma$, there are $2N-1$ interactions involving this vertex which all mutually conflict. Pairwise overlaps occur between these complete subgraphs because each vertex in $\Pi_{\text{weak}}(K_N)$ correspond-

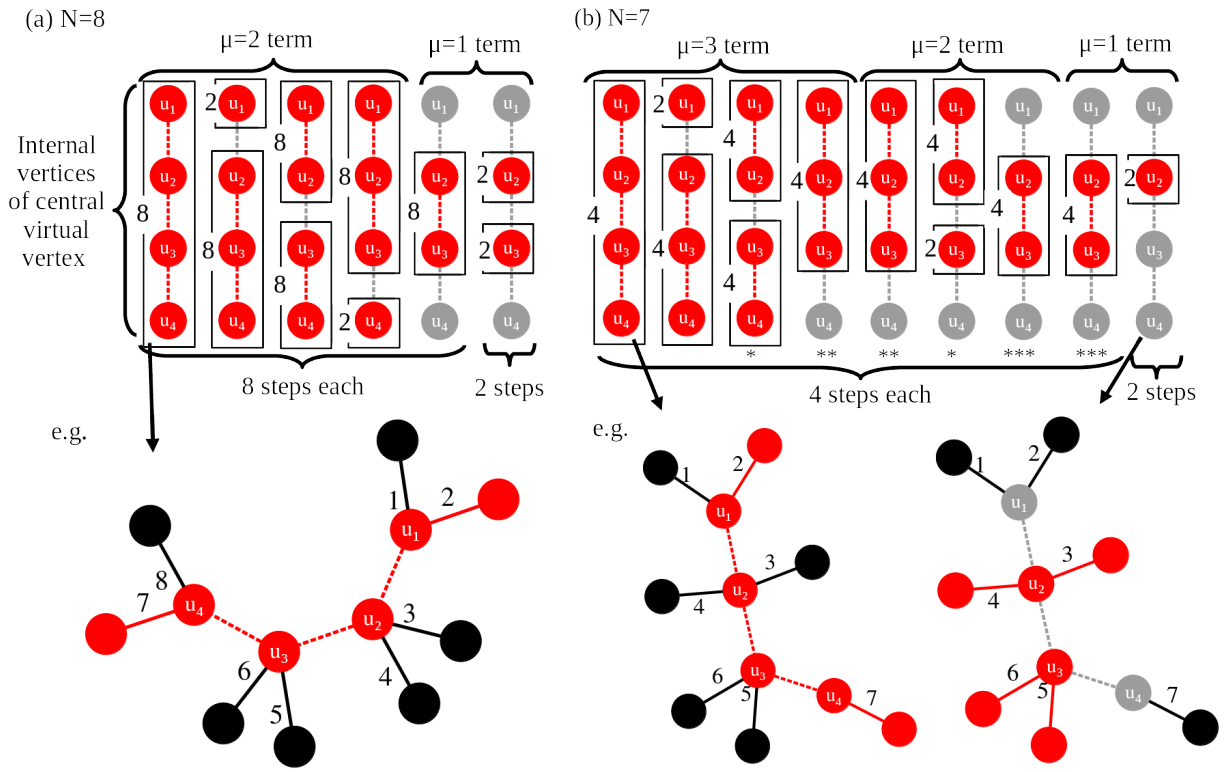


Figure 11.4: Examples of the optimal procedure to parallelize two-mode interactions via strong coloring for the star graph for (a) $N = 8$ and (b) $N = 7$. In the top of the figure, the internal vertices of the central virtual vertex u are shown. Edges between internal vertices are represented by dashed lines. Red (gray) vertices are active (inactive) with red (gray) lines indicating induced (no) Jordan-Wigner strings. Terms are grouped as they appear in the respective sums over μ in Eqs. (11.28)-(11.29). Observe in the case of N odd, this means that the Jordan-Wigner strings of a given type are split in the different groupings (the locations of such splits are marked by one, two, or three stars). The step counts underneath each group give the number of steps to implement all interactions that induce that set of Jordan-Wigner strings. This gives a total of 42 and 34 steps for $N = 8$ and $N = 7$, respectively. The counts associated with each individual Jordan-Wigner string give the number of interactions corresponding to that Jordan-Wigner string. This gives a total of $8 \times (8-1) = 56$ and $7 \times (7-1) = 42$ interactions for $N = 8$ and $N = 7$, respectively. Representative examples of the types of interactions between physical vertices that induce the different Jordan-Wigner strings are shown in the bottom of the figure. Observe in (b) that for N odd, the last internal vertex has only one physical vertex as a neighbor. This is responsible for the different procedure for optimal parallelization.

ing to a two-mode interaction between $u, v \in V_\Sigma$ is in the complete subgraph corresponding to both u and v.

Let each of the N complete subgraphs be the vertices of a new (complete) graph G with two edges between each $v \in G$. One can then map the problem of coloring the two-mode-interaction vertices to one of edge coloring G so that no edges sharing a vertex share a color. In particular, one may color the two-mode interaction vertices of $\Pi_{\text{weak}}(K_N)$ with the color of the corresponding edge of G. Edge coloring K_N takes $N-1$ colors for N even and N colors for N odd [289]. Due to the double edges between each $(u, v) \in V_G$, twice this number is required. Finally, one must consider coloring the one-mode interaction vertices in the original problem. For even N, this requires an additional color because all $2(N-1)$ colors are used in each complete subgraph. For odd N, each complete subgraph has two unused colors, and therefore, one can use one of these colors for the single-mode vertex. This recovers Eq. (11.30) for the number of steps required for weak coloring.

Consider strong coloring for this problem. As all interactions can be directly implemented along a single edge of the system graph, one is limited by the capacity of the physical vertices to have multiple inputs and outputs. In particular, referring to Tab. 11.1, it is clear that multiple "ingoing" ($\tilde{A}_{xu}\tilde{B}_u$) or "outgoing" ($\tilde{A}_{ux}$) interactions cannot be implemented simultaneously on a given vertex u. However, one can simultaneously have one "ingoing" and one "outgoing" interaction for a given vertex—that is, a term of the form $\tilde{A}_{ux}$ and a term of the form $\tilde{A}_{yu}\tilde{B}_u$ can be simultaneously implemented on vertex u provided that $a_u(x) < a_u(y)$.

A lower bound on the chromatic number of $\Pi_{\text{strong}}(K_N)$ can be obtained in terms of the

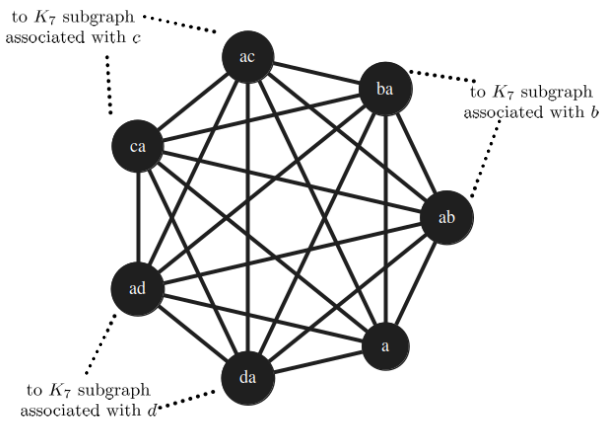


Figure 11.5: An illustration of the conflict graph for the weak-coloring problem with a system graph K_N with $N = 4$. The vertices of the system graph are labeled as $\{a, b, c, d\}$ (e.g., as in Fig. 11.6), and the vertices of the resulting conflict graph are labeled by the corresponding interaction. The conflict graph consists of $N = 4$ interlocking complete subgraphs each associated with one of the system-graph vertices, as described in the main text. One such complete subgraph is shown.

clique number of the graph (see Eq. (11.24)). In particular, one finds that

$$\chi_{\text{strong}}(\Pi(K_N)) \geq N + 2. \tag{11.31}$$

This bound is derived as follows: Given any vertex $u \in V_\Sigma$, the set of all $N-1$ interactions $\tilde{A}_{uv}\tilde{B}_v$ for all $u \neq v$, the interaction $\tilde{B}_u$, and two of the $\tilde{A}_{vu}\tilde{B}_u$ interactions all require the use of the first internal qubit u_1. These interactions form the largest complete subgraph K_{N+2} of $\Pi_{\text{strong}}(K_N)$. Coloring this complete subgraph requires $N + 2$ colors, yielding Eq. (11.31). Therefore at best, asymptotically (in N) one obtains $\chi_{\text{weak}}(\Pi(K_N))/\chi_{\text{strong}}(\Pi(K_N)) \sim 2$.

An upper bound on the chromatic number of $\Pi_{\text{strong}}(K_N)$ can be found by explicit construction. For any Hamiltonian cycle[5] on K_N, the edges in the cycle can be enumerated in such a way that all N interactions along these edges can be implemented simultaneously via strong

[5] A Hamiltonian cycle on a graph is a cycle (closed loop) through the graph that visits each vertex exactly once.

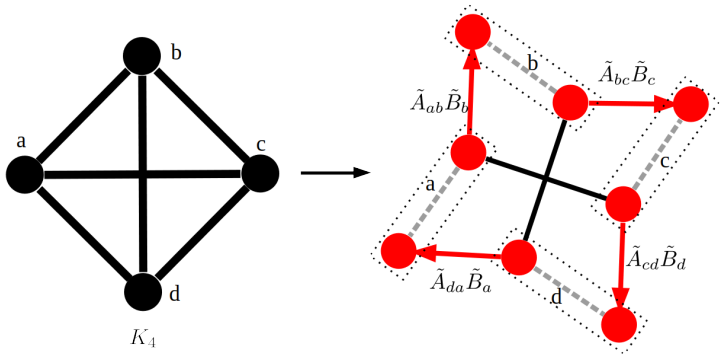


Figure 11.6: A minimal example of K_4 that shows a complete system graph can have its edges enumerated such that a Hamiltonian cycle of two-mode interactions can be implemented simultaneously via strong coloring. The dashed lines denote internal edges, red denotes active qubits, and red arrows from u to v for $u, v \in V_\Sigma$ denote the implementation of an interaction of the type $\tilde{A}_{uv}\tilde{B}_v$.

coloring. See Fig. 11.6 for an example of this for K_4. The number of edge-disjoint Hamiltonian cycles on a complete graph is $(N-1)/2$ for odd N and $(N-2)/2$ for even N [290, 291]. One can independently enumerate these edge-disjoint Hamiltonian cycles such that all interactions within each of these cycles can be implemented simultaneously. Each of these disjoint Hamiltonian cycles are then sequentially implemented. Assuming no other improvements from strong coloring over weak coloring gives an obtainable upper bound, as described below.

Let us consider odd N first. Once the interactions contained in all $(N-1)/2$ edge-disjoint Hamiltonian cycles of K_N are implemented, exactly half of the two-mode interactions are completed in $(N-1)/2$ steps and each edge is traversed exactly once. Considering the rest of the interactions in terms of weak coloring, the problem can be reduced to edge coloring a complete graph as described above. The only difference is that one no longer has parallel edges to consider since one of the two interactions along every edge $e \in E_\Sigma$ is already implemented. This gives

$$\chi_{\text{strong}}(K_N, N \text{ odd}) \leq \frac{3N-1}{2}. \tag{11.32}$$

For even N, once the interactions contained in all $(N-2)/2$ edge-disjoint Hamiltonian cycles of K_N are implemented, one is still left with some parallel edges in the edge-coloring formulation of the weak coloring problem. The extra parallel edges form a perfect matching[6] [290, 291] and therefore these "extra" interactions can be colored with a single additional color. The problem now reduces to the no-parallel edges version of the edge coloring problem on a complete graph, yielding a final upper bound of

$$\chi_{\text{strong}}(K_N, N \text{ odd}) \leq \frac{3N}{2}. \tag{11.33}$$

Given this explicit construction, the combined asymptotic bounds on the improvement from strong coloring over weak coloring is

$$\frac{3}{2} \lesssim \frac{\chi_{\text{weak}}(K_N)}{\chi_{\text{strong}}(K_N)} \lesssim 2. \tag{11.34}$$

11.4.3.3 Separating Weak and Strong Coloring

For the star-graph and the complete-graph examples, it was observed that asymptotically (in N) $\frac{\chi_{\text{weak}}}{\chi_{\text{strong}}} \lesssim 2$.

As it will be seen numerically in Sec. 11.5, such constant-factor improvements are typical for system graphs that arise from realistic qubit architectures. In the near term, eliminating such constant overheads in circuit depth is important and serves as one of the practical motivations for this work, but a more significant separation in parallelization performance between weak and strong coloring can be demonstrated. This example, while contrived, serves to show that

[6] A perfect matching is a set of pairwise non-adjacent edges that cover every vertex of the graph

polynomial separations are possible and, perhaps more importantly, highlights a key feature of system graphs which allow for a large separation between weak and strong coloring.

In particular, a necessary condition for a large improvement due to strong coloring is that the edge bottleneck(s) of the system graph for routing paths in $\tilde{\mathcal{T}}$ are significantly larger than the vertex bottleneck(s). The reason for this is clear: vertex bottlenecks are the limiting factor on parallelization for the weak coloring problem, whereas edge bottlenecks are the limiting factor for strong coloring. That is, the more vertices (edges) to route paths through in the weak (strong) coloring problems, the more room there is for parallelization. When there is a large separation between the size of the edge and vertex bottlenecks, strong coloring necessarily provides more of an advantage. The star graph is a simple example of such a large separation between edge and vertex bottlenecks. The vertex bottleneck is a single vertex, but many edges enter this vertex, suggesting a large potential improvement via strong coloring. We know analytically that this improvement is asymptotically a factor of two. Despite the large separation between edge and vertex bottlenecks, most interactions passing through the central vertex still require many of the internal qubits, hence limiting a greater potential for strong coloring. The counterexample constructed below aims to avoid this limitation.

For simplicity, let us consider $N = 4m$ for some positive integer m. Divide the physical vertices of Σ corresponding to these N fermionic modes into two disjoint sets T_1, T_2, each of size $N/2$. Consider adding edges to the system graph such that both T_1 and T_2 induce disjoint complete subgraphs. Now, consider adding $N/2$ virtual vertices to each of these complete subgraphs. Add an edge from each of these virtual vertices to all physical vertices in the subgraph, forming two bipartite subgraphs. Finally, add a single additional virtual vertex and join it to all other virtual vertices. See Fig. 11.7 for an example of the construction for $N = 8$.

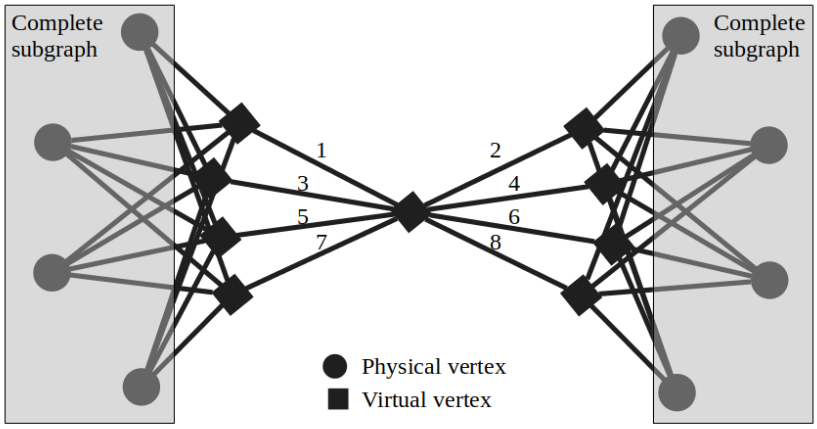


Figure 11.7: An example of the system graph $\Sigma_{\text{bottleneck}}$ that allows for a linear-in-N scaling of the ratio of the chromatic numbers of the conflict graphs with weak and strong coloring. The physical vertices are divided into two complete subgraphs $K_{N/2}$ separated by a single vertex bottleneck, which allows only one interaction at a time between the two subgraphs in the weak coloring problem. For visual clarity, the edges between vertices of the complete subgraphs are not shown. For strong coloring, the extra layer of virtual vertices between each subgraph and the central bottleneck vertex allows any disjoint set of $\frac{N}{2}$ interactions between the two subgraphs to be implemented in a single step.

.

Call this system graph $\Sigma_{\text{bottleneck}}$. $\Sigma_{\text{bottleneck}}$ has a single vertex bottleneck between its two symmetric halves. The weak coloring chromatic number can computed to be

$$\chi_{\text{weak}}(\Pi(\Sigma_{\text{bottleneck}})) = \underbrace{2\left(\frac{N}{2}\right)^2}_{\substack{\text{paths between} \\ T_1, T_2}} + \underbrace{N-1.}_{\substack{\text{paths within} \\ T_1, T_2}} \tag{11.35}$$

Now consider enumerating the edges of the central virtual vertex so that all edges going to one half of the graph are even integers and all edges going to the other half are odd integers. With this labeling, for any choice of $N/2$ interactions from one half of the graph to the other, one can route all $N/2$ interactions through the central vertex simultaneously using strong coloring. Applying

only weak coloring to implement interactions within each complete subgraphs then gives

$$\chi_{\text{strong}}(\Pi(\Sigma_{\text{bottleneck}})) \leq \underbrace{N}_{\substack{\text{paths between} \\ T_1, T_2}} + \underbrace{N-1}_{\substack{\text{paths within} \\ T_1, T_2}} = 2N-1, \tag{11.36}$$

which yields $\frac{\chi_{\text{weak}}(\Pi(\Sigma_{\text{bottleneck}}))}{\chi_{\text{strong}}(\Pi(\Sigma_{\text{bottleneck}}))} \gtrsim \frac{N}{4}$. Such linear-in-$N$ improvement from strong coloring is the best possible scaling for this ratio as the separation between sequential implementation of all interaction terms and the best possible parallel scheme is $\sim N$.

One should be cautious in interpreting this large separation. In practical settings, intelligent design of system graphs from the underlying qubit architecture will rule out such large separations. In practice, more modest, but important, constant-factor improvements between strong and weak coloring should be expected. In particular, there is no reason why the qubits in the central virtual vertex of this example should all be grouped into one system-graph vertex—there are no interactions that require more than a single qubit operator within this vertex. Therefore, a more intelligent system graph built on the same underlying qubit structure would afford the weak coloring problem access to the same performance as the strong coloring problem in our contrived example, by splitting the central virtual vertex into $N/2$ vertices.

Consequently, this example also raises the issue of intelligent system-graph design as a prerequisite to using our algorithms to greatest effect. In Sec. 11.5, we give some more examples of how system graphs may be constructed from the underlying qubit architectures.

11.5 Numerical Results

11.5.1 Description of the Algorithm

For more complicated examples, it is necessary to turn to heuristic algorithms to find good solutions to the parallelization problem in either the weak or strong coloring schemes. Any such algorithm must perform the following steps: First, it must identify a set $\mathcal{P}$ of paths between interacting vertices. Then, given $\mathcal{P}$, it must construct the corresponding conflict graph, which—in the case of strong coloring—requires a choice of enumerating the edges of the system graph. It is known from our analytic results that this choice of enumeration can have a significant impact. Finally, the algorithm must perform a vertex coloring on the resulting conflict graph, which is well-known to be an NP-hard problem in its own right.

Our algorithmic approach to these problems is largely a straightforward one. The most essential and novel aspect of the algorithm relates to choosing the enumeration of edges in an intelligent way to amplify the improvement from strong coloring over weak coloring as much as possible. This is important because it is this step that allows one to parallelize simulation of fermionic Hamiltonians in a way that is aware of the fermion-to-qubit mapping chosen. An understanding of the advantages afforded by taking this information into account is one of the primary goals of this [chapter]. An overview of the salient features of the algorithm is provided here, and the reader is referred to the `github` repository for access to the full code [158].

In either the weak or strong coloring case, to find a path set $\mathcal{P}$, one can begin by weighting the edges of Σ to penalize edges that connect to physical vertices since physical vertices are used in physical interactions and may need to be saved for the implementation of other terms that

involve them. The exact amount of this penalty is a free parameter of the algorithm. The larger the penalty, the more the algorithm will prioritize potential for parallelization over minimizing the Pauli weight of operators. This is because the use of physical vertices may provide shorter paths and, hence, shorter Pauli strings but those are penalized by the algorithm.

Next the algorithm needs to choose a random ordering of the interaction set $\tilde{\mathcal{T}}$. Given this ordering, for each interaction $\tau \in \tilde{\mathcal{T}}$, the algorithm identifies a path as the shortest distance, weighted path through Σ connecting the relevant vertices. This can be done efficiently in time $\Theta\big((|V_\Sigma| + |E_\Sigma|) \log |V_\Sigma|\big)$ via Dijkstra's algorithm [292]. Next, the weight of all edges used in this path are increased and the algorithm proceeds to finding the shortest path for the next $\tau \in \tilde{\mathcal{T}}$ on the reweighted graph. The increase in the weight of the used vertices penalizes paths that do not find "new" routes through Σ—this is advantageous since paths that traverse the same edge in Σ are guaranteed to conflict. The exact choice for this penalty is, again, a free parameter of the algorithm. Due to the sequential nature of this algorithm and the penalties for traversing previously used edges, different orderings of $\tilde{\mathcal{T}}$ will produce different path sets $\mathcal{P}$. Consequently, one needs to run the algorithm many times to sample a variety of different path sets.

Once $\mathcal{P}$ is generated, the algorithm go on to construct the corresponding conflict graph $\Pi(\mathcal{P})$. For weak coloring, this is straightforward—if two paths $p, q \in \mathcal{P}$ share any vertices, the corresponding vertices in $\Pi(\mathcal{P})$ share an edge. For strong coloring, whether or not $p, q \in \mathcal{P}$ conflict depends on the choice of edge enumeration for the vertices in the paths. This choice of enumeration is arbitrary, so a wise choice is an enumeration that attempts to minimize conflicts. In particular, whenever the algorithms finds a path $p \in \mathcal{P}$, it loops through the vertices $u \in p$ and enumerates any previously unenumerated edges according to the following rules: If v is the first vertex in the path and therefore acted on by an operator of type $\tilde{A}_{ux}$, it enumerates the outgoing

edge with the smallest available index. If u is an interior vertex along the path and therefore acted on by an operator of type $\tilde{A}_{xu}\tilde{A}_{uy}$, the algorithm enumerates the incoming vertex as the smallest available index and the outgoing edge as the next smallest available index. Finally, if u is the final vertex in the path and therefore acted on by an operator of type $\tilde{A}_{xu}\tilde{B}_{u}$, the algorithm enumerates the incoming edge with the largest available index. This method of constructing a choice of enumeration follows directly from Tab. 11.1 and minimizes conflicts between paths in a greedy manner. Like the determination of the paths, the outcome of this greedy approach depends on the initial ordering of the interaction set, so, again, it is advantageous to run the algorithm many times.

With $\Pi(\mathcal{P})$ in hand, the algorithm must vertex color it to solve the weak or strong coloring problem, that is to determine the number of steps required to implement the interactions in $\tilde{\mathcal{T}}$. In particular, one can make use of a greedy coloring algorithm as shown in Algorithm 4. The greedy coloring algorithm is guaranteed to satisfy the bound in Eq. (11.22), but its performance can be much better depending on the ordering of vertices. We take a standard approach of a largest-first ordering, where the vertices are ordered from the largest to the smallest degree [293]. For vertices of the same degree, the ordering is random as determined by the order of the initial randomized interaction list. This largest-first approach often works well in practice, but it is only one option among many [294].

11.5.2 Analytically Solved Examples Revisited

We now revisit the analytically solved examples of Sec. 11.4. Using the star graph and complete graph as examples, the algorithm described above is tested for a variety of N ranging

from 3 to 35. 1000 different random orderings of the interaction list for each N are considered. As the algorithm is deterministic once a choice of such an ordering is made, this corresponds to 1000 runs of the algorithm. The results are shown in Fig. 11.8.

For the star graph, the algorithm numerically obtains the true chromatic number for both the weak and strong coloring problem for every ordering of the interaction list. For the weak coloring problem, this is because the conflict graph is well-colored.[7] In particular, the conflict graph is a co-graph—a class whose members are known to be well-colored [295]. To demonstrate that a graph is a co-graph, it is sufficient to show that it has no length-4 paths as induced subgraphs. Such subgraphs do not exist for $\Pi_{\text{weak}}(S_N)$. Every vertex $u \in V_{\Pi_{\text{weak}}}$ is either a member of a complete subgraph $K_{N(N-1)/2}$ consisting of all vertices whose interactions correspond to two-mode interactions or its only neighbors are all contained in such a complete subgraph. Therefore, there exists no set of four vertices whose induced subgraph is a path. While the strong coloring conflict graph is no longer well-colored, the largest-first vertex ordering ensures successful greedy coloring for all interaction-list orderings.

On the other hand, for the complete graph, the algorithm fails to always produce colorings that fully achieve the analytic results. This is because the greedy coloring of the resulting conflict graphs depends heavily on the vertex ordering. Achieving the optimal coloring requires a highly fine-tuned construction. Therefore, the generic randomized greedy coloring algorithm is unlikely to obtain such a coloring as N grows large. Despite these challenges, the coloring algorithm provides almost optimal results for the complete graph for the graph sizes considered.

Fortunately for this algorithm, many realistic architectures are expected to result in system

[7] Well-colored graphs are those such that all vertex orderings produce the same number of colors for a greedy coloring.

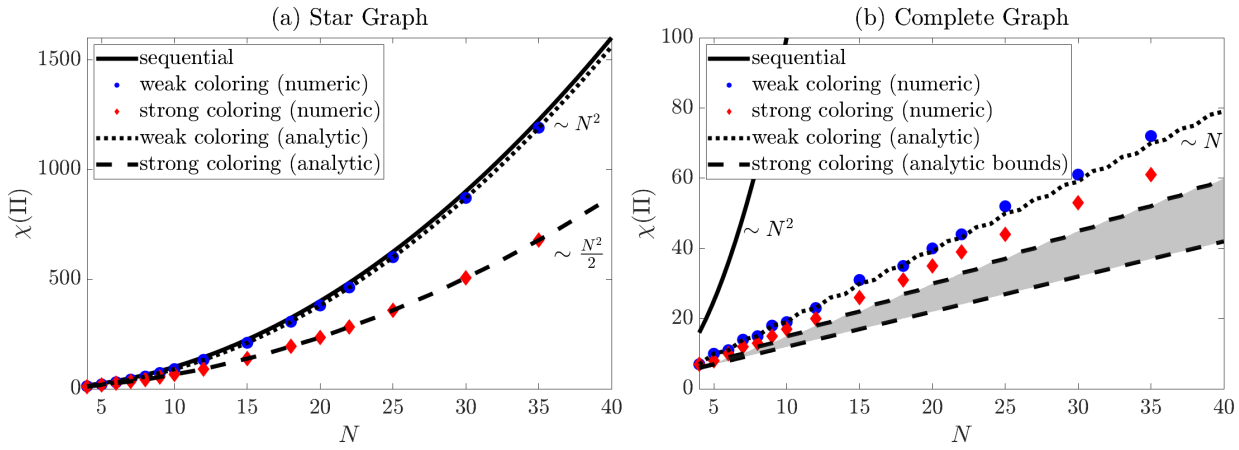


Figure 11.8: Numerical results for the chromatic number from the weak (blue circles) and strong coloring (red diamonds) problems for (a) star and (b) complete system graphs with N physical fermionic modes. For the non-asymptotic analytic results, see Eqs. (11.27)-(11.29) and Eq. (11.30), respectively. For the complete graph, the numerical algorithm fails to achieve the analytically determined bounds as obtaining these results requires a highly fine-tuned vertex ordering for the greedy coloring algorithm on the corresponding conflict graph.

graphs that are limited by vertex bottlenecks, given practical limitations on the high qubit connectivity required for producing input/output-limited system graphs like the complete graph.[8] In the next section, the weak and strong coloring problems will be investigated on two system graphs designed from such realistic architectures.

11.5.3 Current Architectures

The algorithm developed in this section can be applied to system graphs designed on examples of realistic superconducting-qubit architectures. Quantum processors built from superconducting qubits have limited connectivity and thus stand the most to gain from optimized parallelizations. The first example to be studied is a heavy-hexagon qubit architecture as used by many of IBM's quantum processors [296]. This architecture has been shown to be favorable for reducing cross-talk and frequency collisions, while allowing for error correction via a hybrid

[8] Trapped-ion systems are an exception as they provide all-to-all interactions among pairs of qubits.

surface and Bacon-Shor code [297]. The second example is a square-lattice qubit architecture similar to that used by Google's Sycamore chip [298].

Importantly, qubit architectures are distinct from the system graphs one creates on them. While the qubit architecture places constraints on the design of a system graph, one is free to create many different system graphs on a given quantum processor. In practice, this design problem can be viewed as one of optimizing the limited resources of a particular quantum processor—number of qubits, qubit connectivity, circuit depth—to extract a quality simulation of the largest possible system of fermions. Observe that at the cost of a large circuit depth and high Pauli weight operators, a simple Jordan-Wigner transformation (in the form of a system graph which is a line graph) allows one to simulate the most fermions, as no ancilla qubits are needed.

One approach to reduce circuit depths and high Pauli-weight operators is to consider more general system graphs. Here, qubit connectivity is a key limitation on designing efficient system graphs if one wants to avoid the need for SWAP operations in the circuit decomposition of the Hamiltonian-simulation algorithm. In particular, it is desirable to design a system graph so that 1) any qubits that make up a system-graph vertex have linear connectivity for the Jordan-Wigner encoding of the local Majoranas, and 2) if a pair of vertices are adjacent in the system graph, the internal qubits associated with that edge are adjacent in the architecture graph.

To apply the algorithm, let us limit ourselves to a single example of a system graph for each qubit architecture under consideration. In particular, in each case, a system with a total of 49 fermionic modes will be considered. For the heavy hexagon architecture, the system graph considered is constructed from 65 qubits and mirrors the structure of the underlying qubits. This is identical to an example considered in Ref. [249]. For the square lattice, the system graph considered is a triangular tiling of the Euclidean plane and requires 147 qubits. The precise mappings

from architecture graphs to system graphs for each of these cases are shown in Fig. 11.9.

Given these system graphs, N of the 49 vertices are randomly selected to be physical vertices for various N between 5 and 35. For each N, 50 random choices of physical vertices are considered and on each instance, the algorithm is run for 1000 different random orderings of the interaction set for both the weak and strong coloring problems. The best solution from these 1000 different random orderings is then taken. These results, along with quadratic fits are shown in Fig. 11.10.

As is seen from the plots, the improvement from strong coloring over weak coloring in the case of the heavy-hexagon graph is minimal compared to the improvement in the triangular lattice. This is consistent with the conclusions of Sect. 11.4: strong coloring provides the higher performance benefit when the size of the edge bottlenecks to routing the paths induced by interactions are much larger than the size of vertex bottlenecks. The triangular lattice has many more edges per vertex (and correspondingly more qubits) which enable greater parallelization via strong coloring.

Figure 11.11 shows tradeoffs between the number of qubits and the degree of parallelization for the various examples considered in this work: the complete graph, the star graph, the heavy-hexagon graph, and the triangular lattice. While the best balancing of these various tradeoffs depends on many variables, the triangular lattice serves as a particularly nice example of how a system graph on a realistic architecture can be subject to significant reductions in circuit depth via parallelization. While weak coloring alone offers significant performance gains over a sequential approach, considering the precise details of this mapping via the strong coloring problem is important for minimizing the circuit depth.

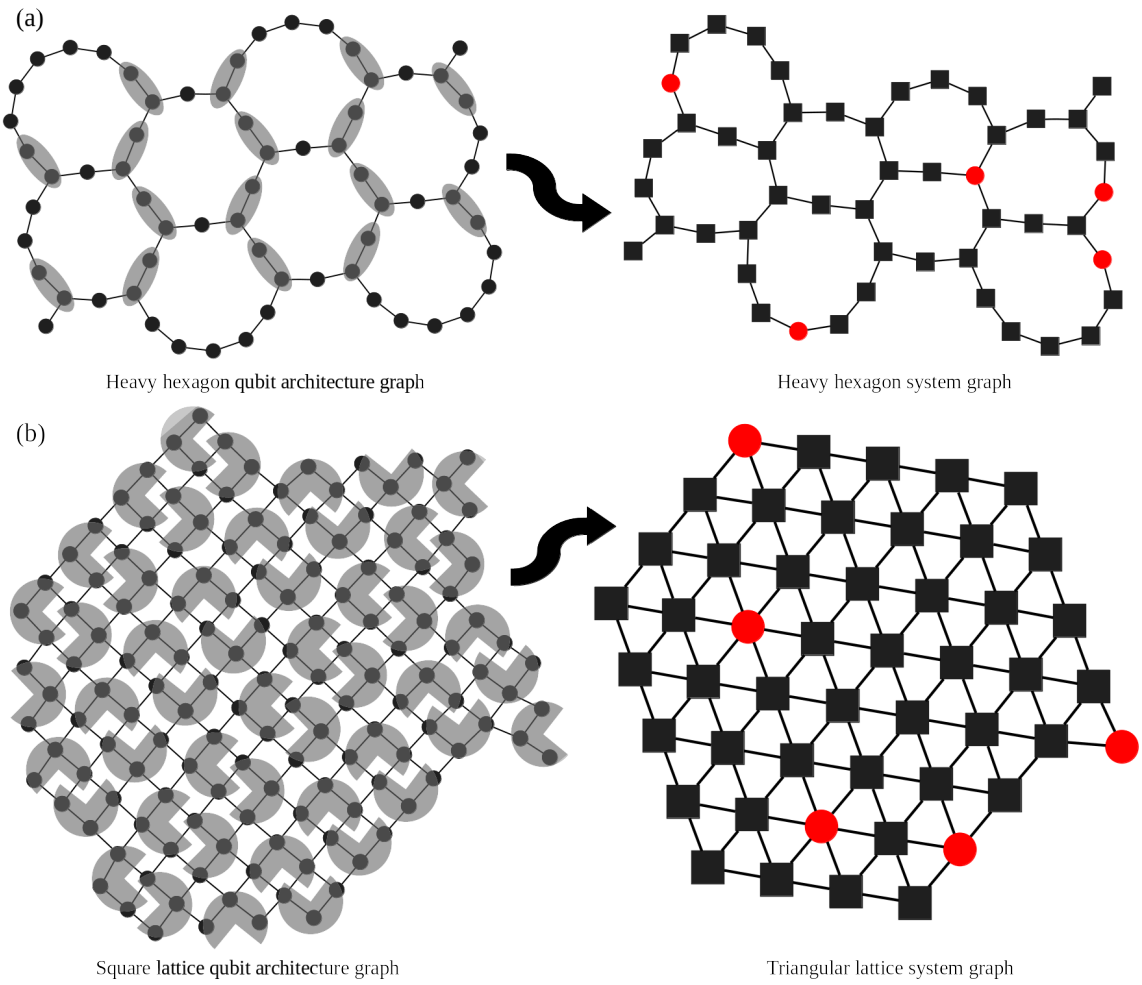


Figure 11.9: Mappings from architecture graphs to system graphs. On the architecture graph, qubits are represented by black dots and are grouped into system-graph vertices as denoted by the gray shading. On the system graphs, square vertices denote virtual vertices and red circular vertices denote physical vertices. For the numerics, N physical vertices are chosen randomly from the 49 total fermionic modes.

.

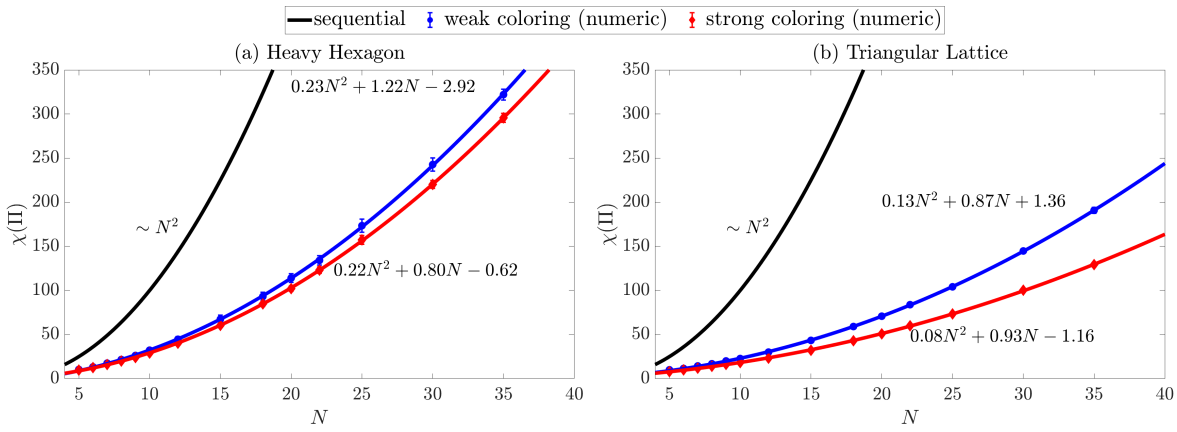


Figure 11.10: Numerical results for the chromatic number of the conflict graph as a function of the number of qubits N for (a) the heavy-hexagon system graph and (b) the triangular-lattice system graph compared with the scaling of the sequential implementation of the Hamiltonian terms.

.

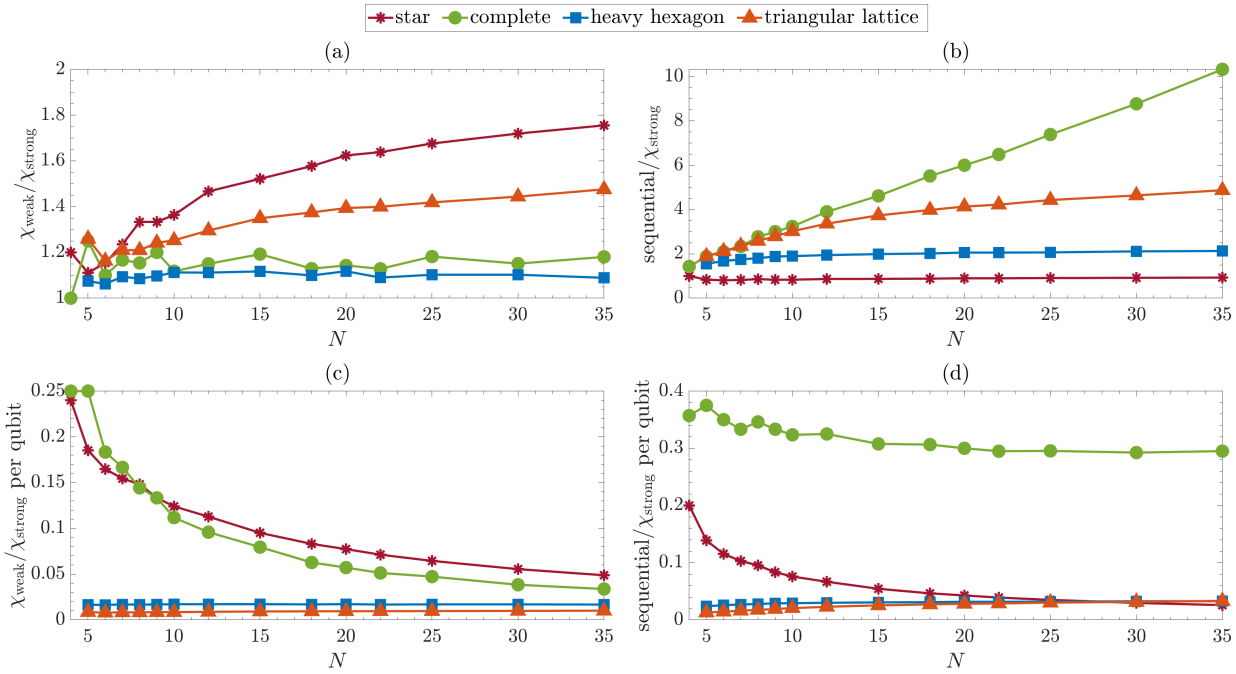


Figure 11.11: Numerical results on the amount of parallelization for various system graphs. (a) shows the improvement in the number of steps for strong coloring versus weak coloring. (b) shows the improvement for strong coloring over a sequential implementation of the interactions in the interaction list. (c) and (d) show the same as (a) and (b), respectively, but weighted by the number of qubits in the system graph. When comparing the different system graphs, recall the size of the complete graph and star graph scale with the number of physical vertices N whereas the heavy hexagon and triangular lattice do not.

11.5.4 Local Interactions

A more common fermionic interaction term is nearest-neighbor hopping on a lattice. The parallelization of this work within the framework of custom fermionic codes can be applied to optimize simulating this model as well. Consider an interaction graph in the form of a two-dimensional square lattice of physical fermionic modes with nearest-neighbor hopping and open boundary conditions,

$$H = \sum_{\langle u,v \rangle} \kappa_{uv} a_u^\dagger a_v + \sum_u \kappa_{uu} a_u^\dagger a_u, \tag{11.37}$$

for real $\kappa_{uv} = \kappa_{vu}$, where the sum is over neighbors on the square lattice. Minimizing the Pauli weight and maximizing the parallelization of such nearest-neighbor hopping terms is the limiting algorithmic factor for a variety of models of interest, such as the spinless Fermi-Hubbard model on the square lattice. This problem is well-understood analytically for a variety of specific fermion-to-qubit mappings [242, 245, 250]. While such analytic approaches to specific problems are valuable when tractable, the techniques of this work allow for an automated optimization for arbitrary Hamiltonians and arbitrary system graphs.

We consider this problem for three different system graphs. The first case is a system graph identical to the interaction graph—a two-dimensional square lattice with all physical fermionic modes. This case can be directly compared with previous work on this problem. The other two cases involve placing the physical fermionic modes in the heavy-hexagon and triangular-lattice system graphs considered above (see Fig. 11.9). The physical modes are embedded such that nearest neighbors on the square-lattice interaction graph are as close as possible on the system graph, enabling low-weight Pauli strings. The precise details of this mapping are included with

the source code as supplemental material [158].

The results for weak and strong coloring on each of these system graphs for a range of interaction-graph lattice sizes are shown in Fig. 11.12. For each lattice size $L \times L$ for $L \in \{4, 9, 16, 25, 36\}$, 1000 random orders for the interaction list are considered for both weak and strong coloring. Observe that compared to non-local models, the advantage over weak coloring due to strong coloring is minimal, independent of system graph. This is expected since all path lengths are short in this model and, similar to the complete-graph example, parallelization is limited by the input/output capacity of the physical vertices. However, parallelization provides significant gains over a naive sequential strategy which scales as $\sim 4N$.

In agreement with previous work, an $O(1)$ circuit depth is obtained with increasing lattice size for the square-lattice interaction graph, and the local fermionic interactions are mapped to local qubit interactions. The triangular lattice performs similarly although it allows for slightly improved performance, especially when small numbers of physical fermions are embedded in the system graph. This is because the triangular lattice offers more paths for implementing interactions than the square lattice. The spike in chromatic number for $N = 36$ in the triangular lattice is because a 6×6 square lattice cannot quite fit in the triangular-lattice system graph considered. Therefore, some interactions that are local in the interaction graph become non-local in the system graph. This effect is even more pronounced for the heavy-hexagon system graph which has lower connectivity than the interaction graph, and therefore cannot perform as well as the other system graphs even for small system sizes.

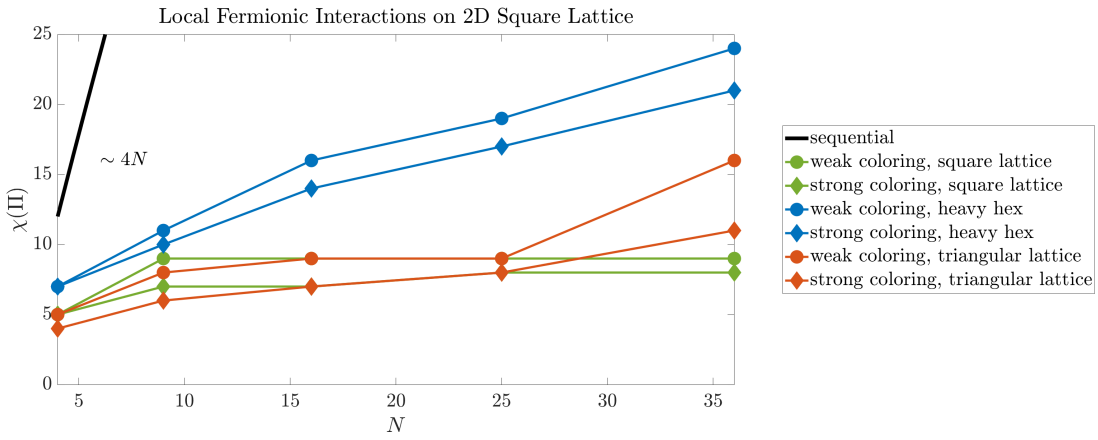


Figure 11.12: Numerical results for the chromatic number of the conflict graph of the system graphs noted corresponding to a square lattice with nearest-neighbor hopping for weak and strong coloring problems.

11.6 Conclusion and Outlook

The amount of parallelization afforded by a system graph is an important target for optimization in the quantum simulation of fermionic Hamiltonians on near-term quantum processors where circuit depth is expected to be an important limiting factor. In this work, this problem is mapped to a graph coloring problem and the relationship between parallelization and the system-graph structure for a variety of representative examples are explored both analytically and numerically. It is found that by considering the details of the fermion-to-qubit mapping, that is to seek strong coloring, one can often find constant-factor improvements in parallelizability relative to performing only weak coloring which is a more high-level approach. Both approaches enable significant reductions in circuit depth relative to a naive sequential approach. The amount of improvement of both coloring schemes compared with the sequential approach, and the strong versus weak coloring is a function of system-graph characteristics—for instance, the number of and the severity of system graph's vertex and edge bottlenecks—as well as on the choice of

enumerating edges in the system graph.

A full account of the algorithmic costs for a Hamiltonian of interest would incorporate the algorithms for parallelization presented here to design a fermion-to-qubit mapping that respects hardware-specific constraints, such as qubit connectivity, noise tolerance, and implementable circuit depths. This work considers only one approach to parallelizability offered at the level of the number of steps needed to implement the Pauli strings that result from a custom fermionic code. When attempting to fully optimize a simulation algorithm in an architecture-aware manner, our approach may further be combined with other parallelization schemes, e.g., those based on fermionic SWAP networks [238–240] or approaches that concern fine-grained details of the circuit decomposition when the Pauli strings are compiled to basic two-qubit entangling gates [237]. The strong-coloring problem, in particular, depends heavily on the choice of encoding of the local Majoranas. While this work only considers Jordan-Wigner encoding of the local Majoranas, it is known that other choices (e.g. Fenwick-tree encoding [234, 241, 246]) lead to lower Pauli weights for the local operators—potentially at the cost of reducing the possibility of parallelization via strong coloring. In fact, local Majoranas could be encoded differently on different sites to perform a full optimization at the circuit level. The problem of detailing the strong coloring rules for other (possibly mixed) choices of encoding the local Majoranas is left to for future studies.

The custom fermionic codes considered in this work, and the generalizations described above, encompass a broad range of mappings. Nonetheless, these do not exhaust the possibilities for mapping fermions to qubits. Consequently, one can imagine profitably mapping parallelization tasks to graph coloring for other encodings as well. For instance, while weak coloring allows one to parallelize a Jordan-Wigner encoding (contained in the class of encodings of this work as a system graph consisting of a line of vertices), another ancilla-free mapping, the Bravyi-Kitaev

encoding [241], does not allow for this sort of improvement. This is because the structure of the Bravyi-Kitaev encoding is given by a Fenwick tree [234], where the root qubit of the tree is non-trivially acted on for every operation, preventing parallelization of the sort we consider here.

We anticipate that applying our tool-set for analyzing parallelizability for Hamiltonian simulation in conjunction with architectural considerations will be useful for obtaining detailed simulation costs for other fermionic Hamiltonians not studied in this work. For example, local and non-local interactions involving four fermionic operators (e.g., Coulomb interactions in quantum chemistry and two-nucleon interactions in nuclear physics) and interactions involving more fermions (such as three- and higher-body interactions in nuclear physics [299]) can be incorporated in the parallelization scheme of this work, and lead to improved simulations in the near and far term. In another interesting direction, one may consider applying the strategy of this work in designing parallelized simulation steps in connection to system graph and hardware connectivity to interacting systems of fermions and bosons, such as those of relevance to lattice gauge theories [300–305]. For example, it would be interesting to thoroughly examine the simulation cost, considering parallelization potential, of fully fermionic formulations (that can be achieved only in 1+1 dimensions [306]) and fully bosonic formulations (that can be achieved for certain gauge theories [307]) of a lattice gauge theory [308]. Finally, in designing system graphs, one may need to take into consideration the entanglement structure (see e.g., Ref. [309] for a discussion in the context of quantum fields) of the resulting subgraphs and the associated computational complexity of various simulation steps, such as state preparation, that is closely tied to entanglement properties.

Chapter 12: Randomized Measurement Protocols for Lattice Gauge Theories

12.1 Introduction

Measurement in quantum mechanics reveals very limited information regarding the structure of the underlying quantum state. This has major practical implications, e.g., for variational near-term quantum-classical algorithms [310–313], the verification of quantum devices [314], or when detecting entanglement [315] in quantum simulation experiments. Randomized measurement protocols, such as randomized benchmarking [316], classical shadows [21, 317–326], and entanglement tomography [22, 327–339] are valuable techniques for addressing this problem. They allow one to estimate many observables from a few measurements [318, 319] or extract non-linear quantities, such as purities $\sim \mathrm{Tr}(\rho^k)$, $k \geq 2$ and entanglement entropies [329–332, 340], potentially without the massive overhead of traditional state tomography [341–343], see e.g. [22] for a recent overview. Many techniques are feasible on noisy, near-term quantum devices [344–346].

A key application for quantum computing and randomized measurement protocols is simulating quantum many-body systems, with digital or analog devices based on atomic, molecular and optical (AMO), and solid-state systems [347–353]. Quantum simulation promises to address long standing questions in condensed matter, high energy physics and nuclear physics. Examples include simulating non-equilibrium evolution and thermalization [354–363], thermal

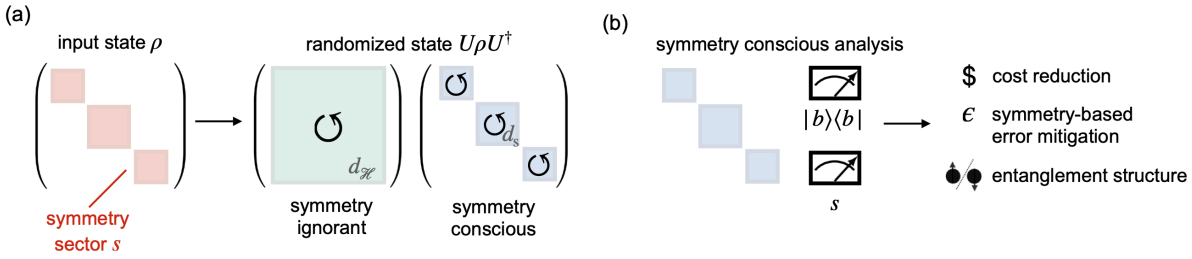


Figure 12.1: (a) Illustration of our proposed symmetry-conscious randomized measurement scheme, which preserves the symmetry structure of states, compared to symmetry-blind approaches. (b) Applications and benefits include cost reduction when measuring entanglement and finding classical representations of quantum states, allowing a rudimentary symmetry-based error mitigation scheme, and enabling symmetry-resolved measurement of entanglement structure.

systems [364, 365] and quantum phases [366–370]. One important frontier is the study of lattice gauge theories (LGTs) [230, 300–303, 305, 308, 362, 371–415] with intricate entanglement structures [363, 416–423] and, potentially, emergent topological phases that have applications in topological quantum computation [271, 424–430].

Randomized measurement protocols are based on changing the basis via unitaries U drawn from an appropriate ensemble $\mathcal{E}$, i.e. $\rho \to U\rho U^\dagger$, followed by measurement in this basis, and classical post-processing or quantum variational techniques to determine quantities of interest.[1] The effectiveness and cost of a scheme depends on the choice of $\mathcal{E}$ and the particular quantities one wants to compute. For instance, estimating expectation values, $\mathrm{Tr}[O_m\rho]$, of M q-local operators $\{O_m\}_{m=1}^M$ can be done with qubit-local random rotations, $\mathcal{O}(3^q \log(M))$ measurements, and efficient classical postprocessing [318]. For non-linear quantities, i.e. those that depend on ρ^k ($k \geq 2$), one approach is to choose an ensemble $\mathcal{E}$ that forms an (approximate) unitary k-design [329, 330].

A fundamental problem is that standard randomized measurement protocols do not account

[1] Schemes using collective measurement on many copies of ρ (e.g. shadow tomography [431, 432]) do not fit into this class of randomized measurement protocols.

for the symmetry structure of states, e.g., by randomizing unnecessarily over unphysical bases. In particular, a (classically known) symmetry $\hat{S}$ of $\rho = \bigoplus_s \rho_s$ ($[\hat{S}, \rho] = 0$), where s labels $\hat{S}$-eigensectors, is lost after randomization. Therefore, it is advantageous to perform symmetry-conscious randomized measurement by using randomizing unitaries such that $[U, \hat{S}] = 0$ for all $U \in \mathcal{E}$. This information is compactly summarized in Fig. 12.1, comparing symmetry-ignorant versus the symmetry-conscious random circuits U proposed in this manuscript. Our goal is to systematically study the construction of symmetry-conscious randomizing circuits and their use cases for exploring physical phenomena. The primary results and findings of our work are as follows

(a) In Section 12.2, we discuss a comprehensive approach to symmetry-conscious randomization for qubit-based models with inherent symmetries. We discuss one application, symmetry-conscious unitary k-designs, as an approach to global randomization. This offers several advantages, including a significant reduction in measurement complexity and allowing for symmetry-based error mitigation [395].

(b) In Section 12.3, we focus on a concrete LGT example to demonstrate the application of symmetry-conscious unitary k-design based randomization. Our main finding is that symmetry-conscious designs enable the simultaneous measurement of both the distillable and symmetry components of entanglement, a capability that symmetry-ignorant designs lack.

(c) In Section 12.4, we present our key result, a protocol designed to detect topological order (TO) experimentally by assessing the gap of the entanglement spectrum (ES). The ES is a presentation of a state in terms of an entanglement Hamiltonian in accordance with Li and

Haldane's conjecture for TO states [370]. This has attracted significant attention in recent years, but has so far remained within the realm of theoretical exploration. To address this, we leverage symmetry-conscious random measurements in combination with entanglement Hamiltonian tomography techniques. This approach represents a promising step towards experimental identification of topological phases which have already started to materialize in experiments [433, 434].

The realization of an approximate unitary k-design is not necessary for our protocol, but it is a convenient approach to illustrate the scheme given that Haar-random measurement channels can be easily inverted. Any tomographically complete symmetry-respecting scheme could serve the same purpose.

Finally, in Section 12.5, we summarize our results and further discuss applications and extensions. The manuscript is supplemented by several appendices where we discuss various details of the employed numerical techniques.

12.2 Symmetry-Conscious Randomization

We begin with a general description of symmetry-conscious unitary circuit construction—i.e., circuits that preserve the symmetry structure of an input state $\rho = \bigoplus_s \rho_s$, $U = \bigoplus_s U_s$. Our approach, partly inspired by Refs. [329, 330], uses the fact that symmetry-conscious randomization can be performed using local generators that are present in the Hamiltonian of a given system. The first step is to identify local q-qubit generators ($1 \leq q \leq m$) for symmetry-conscious Haar random unitaries from amongst the m-body terms of the relevant physical Hamiltonian H. For pragmatic reasons, such terms are good candidates for experimentally-implementable symmetry-

conscious randomization: if an experimentalist can engineer the local Hamiltonian terms for the purposes of (Trotterized) time evolution, with adequate control of coupling strengths, they can also perform randomization generated by such terms. The next step is to ensure that such terms can generate (Haar-)random unitaries over an m qubit Hilbert space within every symmetry sector. The goal is to use this m-local symmetry-conscious randomization as a building block for randomization over the full Hilbert space. General $U(2)$ rotations are in either a 'ZXZ' or 'ZYZ' generator decomposition with angles α, β and γ (and a global phase). In an m-qubit circuit one must identify the corresponding operators that act on q qubits and which 'embed' 1-qubit rotations in every symmetry sector.

Different possible families of circuits can be generated by different choices of arranging these blocks within a larger circuit; the m qubits interacting via a m-local Hamiltonian term, need not match the same (typically, geometrically local) m qubits in the corresponding Hamiltonian. While we focus primarily on generating families of unitary circuits that form symmetry-conscious unitary k-designs, one could also consider shallower circuits, e.g., for measuring local observables.

To leverage global Haar randomness, we determine if the selected local, symmetry-respecting terms are sufficient to generate global, symmetry-respecting Haar random unitaries within each symmetry sector. This is not generically the case [435]. Verifying that sufficient randomization is, indeed, possible with the selected set of terms must be done on a case-by-case basis and accounts for the primary challenge in extending this approach to new systems. Here, we will demonstrate that such local terms are sufficient to generate approximate unitary k-designs for a few different examples of interest (with an emphasis on LGTs).

12.2.1 A Symmetry-Conscious Unitary k-design Example: Particle Number Symmetry

We illustrate our approach first for spin systems with particle number symmetry, $\hat{S}_N \equiv \sum_j (\sigma_j^z + 1)/2$, where j labels a lattice site and σ_j^a ($a = x, y, z$) are Pauli matrices acting on the site. We emphasize that symmetry-conscious randomization schemes for particle number symmetry have been considered before [321, 435, 436], but it serves as a simplified setting in which to develop the particular approach we will take to symmetry-conscious randomized measurement protocols. The lessons we learn here can then be extended to LGTs.

Symmetry-conscious randomization is achieved with components consisting of (m=2)-qubit gates of the form

$$u \equiv \begin{pmatrix} e^{i\theta} & & & \\ & & & \\ & & [u_1] & \\ & & & e^{i\phi} \end{pmatrix}, \tag{12.1}$$

where the rows and columns label (from top down) the $\downarrow\downarrow$ (0-particle), $\uparrow\downarrow$, $\downarrow\uparrow$ (1-particle) and $\uparrow\uparrow$ (2-particle) sectors; $[u_1]$ indicates the 2×2 (ZYZ-decomposition) matrix structure,

$$[u_1] \equiv \begin{pmatrix} e^{i(\alpha+\gamma)}\cos(\beta) & e^{-i(\alpha-\gamma)}\sin(\beta) \\ -e^{i(\alpha-\gamma)}\sin(\beta) & e^{-i(\alpha+\gamma)}\cos(\beta) \end{pmatrix}. \tag{12.2}$$

If α, β, and γ are selected such that $[u_1]$ is drawn from a circular unitary ensemble (CUE) [332], θ and ϕ are drawn evenly from $[0, 2\pi)$, then Eq. (12.1) is a block-structured unitary acting Haar-

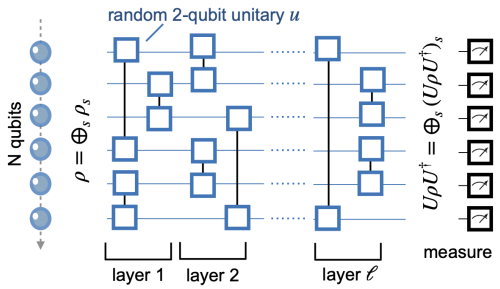
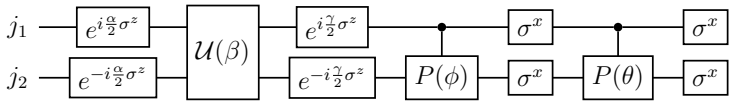


Figure 12.2: Random circuit scheme for systems with particle number symmetry. Blue squares connected by black lines represent the unitary u in Eq. (12.3).

randomly on the blocks of equal particle number.[2] It has the following circuit realization:



(12.3)

where $\mathcal{U}_{j_1,j_2}(\beta) \equiv \exp\{i\frac{\beta}{2}(\sigma^y_{j_1}\sigma^x_{j_2} - \sigma^x_{j_1}\sigma^y_{j_2})\}$ and $P(x) = \mathrm{diag}(1, \exp\{ix\})$ is the phase gate.

In line with our general approach, the intuition behind Eq. (12.3) comes from inspecting a particle number conserving Hamiltonian: $\mathcal{U}_{j_1,j_2}(\beta)$ is essentially a hopping term between j_1 and j_2 with amplitude $\beta/2$, the z-rotations before and after are density-density interactions with amplitudes $\alpha/2$ and $\gamma/2$, respectively. Together, these generate $SU(2)$ in the 1-particle block. The relative phases θ and ϕ are because we wish to embed $U(2)$, not $SU(2)$, and because the 0- and 2- particle blocks should be $U(1)$-randomized. (A global phase is irrelevant.)

As a first application, we now focus on global randomization, arranging these components into approximate unitary k-designs. To do so, we construct an ensemble $\mathcal{E}$ of $N_\mathcal{E}$ n-qubit symmetry-conscious random circuits, as shown in Fig. 12.2, consisting of ℓ layers where $n(n-1)/2$ pairs of qubits ($(n-1)(n-2)/2$ if n is odd) are randomly connected by Eq. (12.1). For sufficiently large ℓ and $N_\mathcal{E}$, this realizes an approximate unitary k-design in every symmetry

[2] An overall phase of the 2×2 CUE matrix was re-expressed in the 0-, and 2-particle sectors by partially absorbing it into a global phase.

block s.

We verify that such circuits yield an approximate k-design numerically for $k = 2$ by considering the following moments, separately for every s,

$$(\mathcal{B}^s)^{i'j'k'l'}_{ijkl} \equiv \langle U^s_{ij} U^{s*}_{i'j'} U^s_{kl} U^{s*}_{k'l'} \rangle - \frac{d_s^2}{d_s^2 - 1} \Big[(\mathcal{A}^s)^{i'j'}_{ij} (\mathcal{A}^s)^{k'l'}_{kl} + (\mathcal{A}^s)^{k'l'}_{ij} (\mathcal{A}^s)^{i'j'}_{kl} \Big] , \tag{12.4}$$

where $\langle \dots \rangle = (1/N_{\mathcal{E}}) \sum_{U \in \mathcal{E}} \dots$ is the $\mathcal{E}$-ensemble average, d_s is the dimension of sector s, and $(\mathcal{A}^s)^{kl}_{ij} \equiv \langle U^s_{ij} U^{s*}_{kl} \rangle$. For a 2-design it holds that

$$(\mathcal{B}^s)^{i'j'k'l'}_{ijkl} \overset{\text{2-des.}}{=} - \frac{\delta_{ii'}\delta_{kk'}\delta_{jl'}\delta_{lj'} + \delta_{ik'}\delta_{ki'}\delta_{jj'}\delta_{ll'}}{d_s(d_s^2 - 1)} , \tag{12.5}$$

while for a 1-design $(\mathcal{A}^s)^{kl}_{ij} = \delta_{ij}\delta_{kl}/d_s$ [437, 438]. We numerically simulate an ensemble of $N_{\mathcal{E}}$ circuits and compute their deviation from a 2-design ϵ as the absolute difference between Eq. (12.5) and Eq. (12.4), averaged over indices and multiply by $d_s(d_s^2 - 1)$ to make ϵ dimension-independent. Up to a rescaling of the approximation ratio, this choice of quantifying the error between our random ensemble and a 2-design is equivalent to other standard choices, such as the diamond norm or the the frame potential. This is demonstrated in Appendix H.1.

Fig. 12.3(a) displays the size $N_{\mathcal{E}}$ of a circuit ensemble (with $\ell = 128$ layers) to approximate a 2-design better than $\epsilon = 10^{-2}$ for all particle number blocks of $n = 4, 6, 8, 10$ qubit systems. The sampling complexity scales with the block-dimension d_s as

$$N_{\mathcal{E}} \sim d_s^{\xi} , \quad \xi = 2.1 \pm 0.2 , \tag{12.6}$$

consistent with $\xi = 2$. Error bars are found by comparing with the $\ell \to \infty$ limit (obtained by

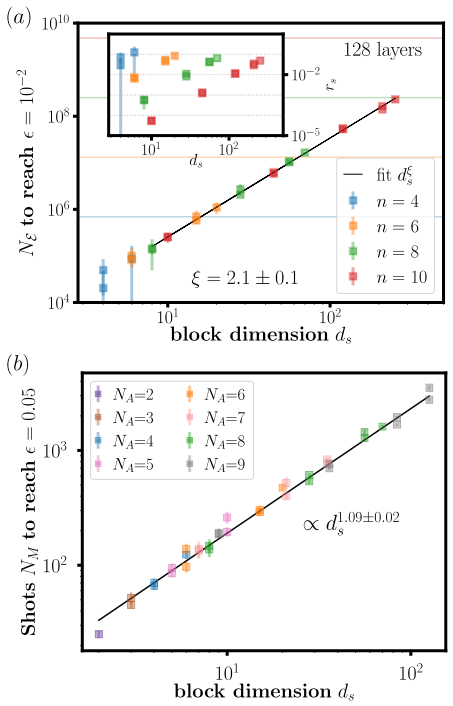


Figure 12.3: (a) Particle number-respecting random circuits $N_\mathcal{E}$ required to approximate a 2-design with precision better than $\epsilon = 10^{-2}$ in every symmetry sector ($\ell = 128$) for a total number of sites $N = 4, 6, 8, 10$. Inset: Relative sampling cost reduction compared with a symmetry-ignorant scheme. (b) Measurements N_M required to estimate sector-wise 2-purities with precision better than $\epsilon = 0.05$, for fixed $N_\mathcal{E} = 1,428$ and $\ell = 128$. Here, the states considered are reduced density matrices on subsystems of dimension $N_A = N/2$ (i.e. on a bipartition of the lattice.)

directly sampling Haar random unitaries in each symmetry block); the fit error for ξ is determined by varying the fit regime, leaving out the largest few blocks, and by varying the required ϵ by one order of magnitude. The inset of Fig. 12.3(a) shows the relative reduction r_s in sampling cost compared to using a symmetry-ignorant scheme, where $r_s := (d_s/d_\mathcal{H})^\xi$. Largest gains are found away from half filling, which can be understood by comparing $d_\mathcal{H} = 2^n$ to $d_s = \binom{n}{s} = n!/(s!(n-s)!)$ and using Stirling's approximation to find that $r_s \sim (d_s/d_\mathcal{H})^\xi \approx [2(\frac{s}{n})^{\frac{s}{n}}(1-\frac{s}{n})^{1-\frac{s}{n}}]^{-n\xi}/(2\pi n\frac{s}{n}(1-\frac{s}{n}))^{\xi/2}$, an exponential (in n) cost reduction for $s \ll n/2$ ($\gg n/2$).

A primary concern for constructing symmetry-conscious random circuits is circuit complex-

ity—in particular, the required number of layers ℓ needed to reach a given precision as a function of the system size. To investigate this, we conducted numerical simulations on systems involving up to n=10 qubits. In Appendix H.2, we provide a detailed account of the number of layers necessary to represent a k-design (specifically a 2-design), comparing our findings with exact Haar random sampling results. For the system sizes we can consider numerically this convergence is very rapid. For instance, for $n = 10$ qubits with $N_{\mathcal{E}} = 8192$ random unitaries the difference between sampling from our circuits versus sampling directly from the CUE is already very small within $\ell \approx 15$ layers for all symmetry sectors. The exact value of this error floor is set by $N_{\mathcal{E}}$, falling off with the expected $\sim 1/\sqrt{N_{\mathcal{E}}}$ scaling, see Appendix H.2. As our numerical methods are restricted to relatively small systems and considering that k-designs serve as one application rather than the core focus of our work, we refrain from asserting an exact analytical form of the error scaling that holds true for large n. Nonetheless, it seems reasonable to conclude from our data that the behavior scales as a low-degree polynomial, and that symmetry-conscious circuits require comparable depths when compared to symmetry-agnostic circuits. For convenience, we continue to work in the large layer limit, in practice $\ell = 128$, thus massively overdoing the actual number of layers needed. We will return to analytic estimates in future work.

12.2.2 Estimating k-purities from k-designs

Next, we explore measuring k-purities using a prototypical model with particle number symmetry; we consider the following spin Hamiltonian in $(1+1)$d,

$$H = \frac{1}{2a}\sum_{j=0}^{N-1}(\sigma_j^+\sigma_{j+1}^- + \text{h.c.}) + m\sum_{j=0}^{N-1}(-1)^j\sigma_j^+\sigma_j^- \,, \tag{12.7}$$

on a lattice with N sites labelled by $j \in [0, N-1]$ and with periodic boundary conditions (PBCs); $\sigma_j^{\pm} \equiv \frac{1}{2}(\sigma_j^x \pm i\sigma_j^y)$, a is the lattice spacing and m a mass term. Eq. (12.7) conserves particle number, $[H, \sum_j \frac{1}{2}(\sigma_j^z+1)] = 0$. As inputs for the particle number symmetry-conscious scheme in Fig. 12.2, we numerically determine ground states of H via exact diagonalization [439]. Because these are simple, unentangled computational basis states in the limit $1/2a \to 0$, we work in the opposite limit with $m \cdot a = 0.05$ (the ground states are half-filled). We focus on a subsystem A with $N_A = N/2$ sites. Symmetries of ρ_A are particle number $s \equiv n_A \in [0, N_A]$ in the subsystem, noting that if particle number is fixed globally, ρ_A is block-diagonal, $[\rho_A, \sum_{j=0}^{N_A-1} \frac{1}{2}(\sigma_j^z+1)] = 0$, i.e. $\rho_A = \bigoplus_s \rho_{A,s}$.

We extract k-purities, $\mathrm{Tr}(\rho_{A,s}^k)$ by measuring the probabilities $P_U(b,s)$ of bitstring b (and symmetry sector s) with N_M shots in the basis defined by the random unitary U; k-purities for $k \geq 2$ are directly related to the k-Rényi entropies $S_{A,s}^{(k)} \equiv [1/(1-k)] \log \mathrm{Tr}(\rho_{A,s}^k)$. Following the approach taken in Refs. [329, 330, 440], stochastic moments $\langle P_U(b,s)^k \rangle \equiv (1/N_{\mathcal{E}}) \sum_{U \in \mathcal{E}} P_U(b,s)^k$ are related to k-purities via [330],

$$\langle P_U(b,s)^k \rangle = \frac{1}{D_k} \sum_{\{a_i\}_k \in \mathbb{N}_0} C_{\{a_i\}_k} \prod_{j=0}^{k} \mathrm{Tr}\left[\rho_{A,s}^j\right]^{a_j} \tag{12.8}$$

where $\{a_i\}_k \equiv a_1, \ldots, a_k \in \mathbb{N}_0$ with $\sum_{j=1}^k j a_j = k$, $D_k \equiv \prod_{j=0}^{k-1}(d_s + j)$ and $C_{\{a_i\}_k} \equiv k! / \prod_{j=1}^k (j^{a_j} a_j!)$; $\langle P_U(b,s)^k \rangle$ and $\mathrm{Tr}\left[\rho_{A,s}^k\right]$ refer to the k-moments of the probabilities and k-purity per symmetry sector s, respectively, with $\sum_{b \in s} P_U(s,b) = \mathrm{Tr}[\rho_{A,s}] \equiv p_s \leq 1$ and $\sum_s p_s = 1$. We assume an ideal quantum machine, and the total measurement cost is $N_M \cdot N_{\mathcal{E}}$.

In Fig. 12.3(b), we show the required shot number N_M to measure $S_{A,s}^{(2)}$ with precision better than $\epsilon = 0.05$, for fixed $N_{\mathcal{E}} = 1{,}428$ and $\ell = 128$ layers, comparing different system

sizes and symmetry sectors. The fit reveals an approximately linear dependence on d_s, error bars indicate standard error of the mean for seven independent trials. (In Fig. H.3 of Appendix H.2 we investigate the $N_{\mathcal{E}}$ dependence in the infinite shot limit.) Together, our results show that the cost of the symmetry-conscious approach, $N_{\mathcal{E}} \cdot N_M$, is proportional to d_s instead of d_A (the size of the subsystem). In many cases $d_s \ll d_A$, yielding a significant advantage. We shall see that this advantage is exponential for lattice gauge theories.

12.3 Lattice Gauge Theory Entanglement

12.3.1 $\mathbb{Z}_2$ LGT in (1+1)d

Lattice gauge theories (LGTs) are systems with an extensive number of local constraints in the form of Gauss laws defining a physical sub-Hilbert space. One of the key applications is the simultaneous measurement of distillable and symmetry components of entanglement. This capability is exclusive to symmetry-conscious designs and is not attainable through symmetry-ignorant schemes; it would otherwise require full state tomography. We consider $\mathbb{Z}_2$ LGT coupled to staggered matter in $1+1$ dimensions, with Hamiltonian,

$$H = \frac{1}{2a}\sum_{j=0}^{N-1}(\sigma_j^+ \tilde{\sigma}^x_{j,j+1}\sigma^-_{j+1} + \text{h.c.}) + m\sum_{j=0}^{N-1}\frac{(-1)^j}{2}(1+\sigma_j^z) + e\sum_{j=0}^{N-1}\tilde{\sigma}^z_{j,j+1}, \tag{12.9}$$

Gauss laws

$$\hat{G}_j \equiv e^{i\pi \mathcal{Q}_j}\tilde{\sigma}^z_{j-1,j}\tilde{\sigma}^z_{j,j+1}, \tag{12.10}$$

and periodic boundary conditions (PBC), where j labels a site and $[H, \hat{G}_j] = 0$; σ_j^b ($\tilde{\sigma}^b_{j,j+1}$), $b = x, z$, are Pauli operators residing on the sites (links) of the lattice and representing the matter

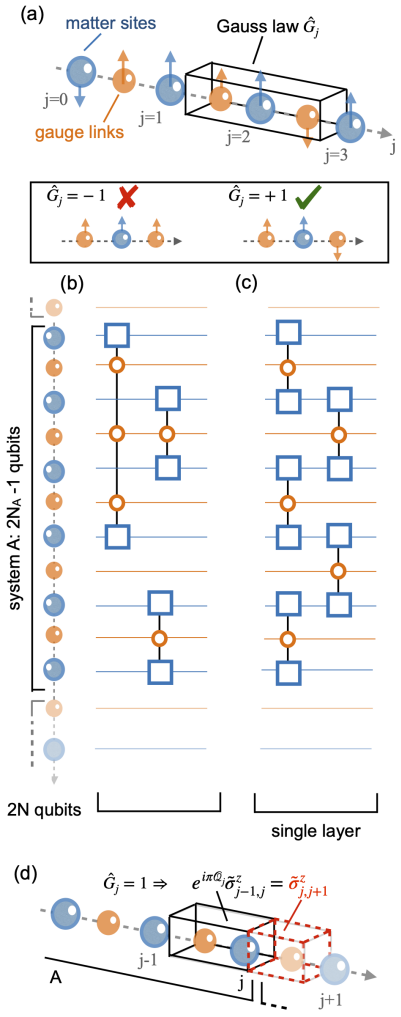


Figure 12.4: (a) Illustration of $\mathbb{Z}_2$ LGT in $1+1$ dimensions, depicted is an "even" site, $(-1)^j = 1$, where spin up (down) is the presence (absence) of a $\mathbb{Z}_2$ charge, $\mathcal{Q}_j = +1$ (0). (b) Single layer of a random-measurement circuit with non-local gates $2 + |j_1 - j_2|$ qubit gates, an extension of the strategy in section 12.2. (c) Near-term strategy based on 3-qubit unitaries. (d) Illustration of the symmetry structure of ρ_A.

(gauge) degrees of freedom of the theory; m is a mass parameter, e the $\mathbb{Z}_2$ coupling, a the lattice spacing, $\sigma_j^{\pm} \equiv (\sigma_j^x \pm i\sigma_j^y)/2$, and $\mathcal{Q}_j \equiv (\sigma_j^z + (-1)^j)/2$ is the $\mathbb{Z}_2$ charge. This is compactly summarized in Fig. 12.4(a) where Gauss law eigensector with $\hat{G}_j|\psi\rangle = +1|\psi\rangle$ are physical.

Fig. 12.4(b) and (c) depict random measurement strategies for this model, we investigate a N_A site subsystem with boundary conditions ending in a matter site. The first strategy, (b), is an

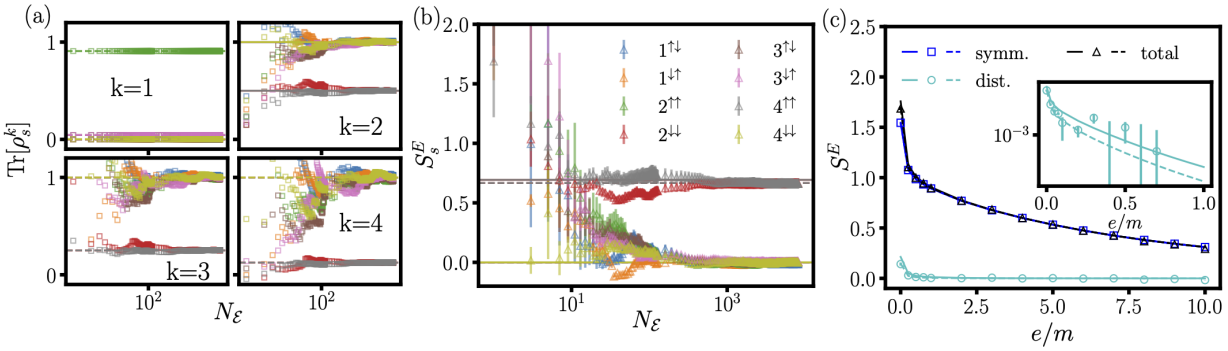
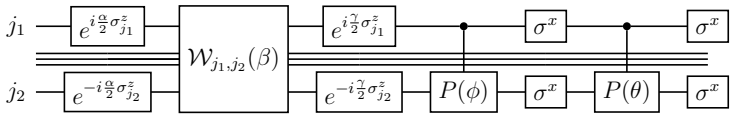


Figure 12.5: (a) Sector-wise k-purities of the $\mathbb{Z}_2^{1+1}$ ground state measured based on Eq. (12.8) as a function of $N_\mathcal{E}$, for $N_A = N/2 = 5$, $m \cdot a = 0.1$, $e/m = 8$ and $\ell = 32$. (b) Von Neumann Entropy from k-purities/Rényi entropies using a 4^{th} order finite-difference approximation of Eq. (12.14). (c) Symmetry- (symm.), distillable (dist.) and total entanglement entropies as a function of e/m for $N_\mathcal{E}$=2000, $m \cdot a = 0.1$ and $\ell = 32$. Symbols represent random measurement results, dotted lines represent approximating S_E by (exact) Rényi-entropies up to $k = 4$, solid lines are exact results.

extension of the circuit in Section 12.2, Eq. (12.3), made gauge invariant by introducing Wilson lines connecting sites j_1 and j_2,

$$W_{j_1,j_2} \equiv \prod_{j=j_1}^{j_2-1} \tilde{\sigma}^x_{j,j+1} \tag{12.11}$$

and with circuit representation,



(12.12)

where $\mathcal{W}_{j_1,j_2}(\beta) \equiv \exp\{i\frac{\beta}{2}(\sigma^y_{j_1} W_{j_1,j_2} \sigma^x_{j_2} - \sigma^x_{j_1} W_{j_1,j_2} \sigma^y_{j_2})\}$; the angles α, β, γ, ϕ and θ are randomly drawn as before. The abbreviated middle qubit bundle in Eq. (12.12) refers to qubits representing gauge links (orange in Fig. 12.4), j_1 and j_2 start and end on matter sites (blue in Fig. 12.4). This results in a $2 + |j_1 - j_2|$ qubit unitary which is not a feasible strategy near-term. Because of this we will focus on a 3-qubit unitary strategy, Fig. 12.4(c), at the cost of somewhat

deeper circuits to obtain k-designs. We demonstrate numerically that these circuit form approximate 2-designs in Appendix H.3, and focus here on measuring entanglement. We note that (b) and (c) generate k-designs in every symmetry sector, despite the fact that no explicit 2-qubit entangling operation is performed between gauge sites (orange). This is a consequence of gauge symmetry: For this simple model the gauge link degrees of freedom are not truly independent because they could have been eliminated using Gauss' law.[3]

12.3.2 Measuring Distillable and Symmetry Components of Entanglement

We now put our randomizing circuits to use by demonstrating their utility for determining the entanglement structure of ground states. In the literature, this work included, the term "entanglement structure" is used fairly ambiguously to denote anything characterizing entanglement beyond entanglement entropies. This includes the separation of entanglement entropies into distillable and symmetry entanglement, but, more generally, also includes the structure in terms of an entanglement Hamiltonian, eigenvalue spectrum (the so-called Schmidt spectrum), and symmetries of a reduced density operator. We will consider all of these in this section.

In particular, we consider a bipartition $\rho_A \equiv \mathrm{Tr}_{\bar{A}}(\rho)$ of the lattice with $N_A \equiv N/2$ sites. As before, ρ_A is block-diagonal in particle number n_A (i.e., $[\rho_A, \sum_{j=0}^{N_A-1}(1+\sigma_j^z)/2] = 0$) but additionally has symmetries beyond those of the non-gauge spin model. They are illustrated in Fig. 12.4(d): Because of Gauss' law, on the right (left) boundary $j = j_R \equiv N_A$ ($j = j_L \equiv 0$),

[3] A caveat is that with PBCs a gauge zero mode cannot be integrated out. A k-design acting on the full system, not just a subsystem, would include a modification which depends on how the remaining bosonic mode is digitized; see e.g. [441, 442] for bosonic random measurement schemes.

out-going (in-going) electric fields can be written as

$$\tilde{\sigma}^z_{j,j+1} = \exp\{i\pi\mathcal{Q}_j\}\tilde{\sigma}^z_{j-1,j}\,. \tag{12.13}$$

These operators are symmetries of ρ_A (i.e. $[\tilde{\sigma}^z_{j_R,j_R+1}, \rho_A] = [\tilde{\sigma}^z_{j_L-1,j_L}, \rho_A] = 0$) if ρ is physical (i.e. Gauss' law respecting) and result in a four-block symmetry structure. Together, we label the sectors as $s \equiv n_A^{s_L,s_R}$ where $s_{L/R} = \uparrow/\downarrow$ and $n_A \in [0, N_A]$ is particle number.

In Fig. 12.5(a), we show k-purities, $\text{Tr}[\rho^k_{A,s}]$ reconstructed by inverting Eq. (12.8), measured for $N = 10$ ($N_A = 5$), $m \cdot a = 0.1$, e/m=8 and $\ell = 32$. We plot them as a function of $N_\mathcal{E}$; measurements are obtained in the infinite shot limit ($N_M \to \infty$). The $k = 1$ result $p_s \equiv \text{Tr}[\rho_{A,s}]$ is recovered exactly by design, higher ($k \geq 2$) purities are recovered for sufficiently large $N_\mathcal{E}$ (data is shown up to $N_\mathcal{E} = 2^{14}$). Not shown is the cost in N_M to obtain constant error which, as in Sect. 12.2, scales with d_s. In Fig. 12.5(b), we estimate the von-Neumann entropy per symmetry block,

$$S_s \equiv \text{Tr}_s[\bar{\rho}_{A,s}\log(\bar{\rho}_{A,s})] = -\lim_{k\to 1^+}\frac{\mathrm{d}}{\mathrm{d}k}\text{Tr}_s[\bar{\rho}^k_{A,s}], \tag{12.14}$$

where $\bar{\rho}_{A,s} \equiv \rho_{A,s}/p_s$. We make use of a 4^{th} order finite-difference approximation of the derivative to derive Eq. (12.14) from the measured k-purities; error bars are obtained from comparing a 4^{th} and 3^{rd} order derivative. Finally, Fig. 12.5(b) shows the decomposition of the von-Neumann entropy, $S = -\text{Tr}_A[\rho_A\log(\rho_A)]$ into a symmetry- ('classical' entanglement) and distillable com-

ponent, $S^{\text{symm.}} + S^{\text{dist.}}$,

$$S^{\text{symm.}} \equiv -\sum_s p_s \log(p_s)\,, \quad S^{\text{dist.}} \equiv \sum_s p_s S_s\,, \tag{12.15}$$

as a function of e/m and for $m \cdot a = 0.1$.

Finally, we note that a difference in the number of unitaries required for approximating a 2-design versus the seemingly lower requirements for measuring 2-fidelities evident in Fig. 12.5. It is important to note that the accuracy of reproducing the latter is inherently dependent on the state itself. In contrast, the analysis we perform in section 12.2 provides an upper bound, ensuring convergence for *any* input state.

12.3.3 Classical Shadows

Next, we explore classical representations of ρ_A, starting with the classical shadow formalism of [318]. The basic idea is to randomize $\rho_A \to U\rho_A U^\dagger$ and perform a computational basis measurement yielding a bitstring b from which the sector s can be read off. A symmetry-conscious shadow is $U_s^\dagger|b,s\rangle\langle b,s|U_s$, where the subscript s indicates that one works in block s with dimension d_s. The ensemble of random rotations yields a CUE-random quantum channel $\mathcal{M}[\rho_A] = \bigoplus_s \mathcal{M}_s[\rho_{A,s}]$, and with many measurements and, consequently, many shadows one obtains classical sector-wise state representations by taking the expectation value

$$\bar{\sigma}_{A,s} \equiv \mathbb{E}\left[\mathcal{M}_s^{-1}\left(U_s^\dagger|b,s\rangle\langle b,s|U_s\right)\right]\,, \tag{12.16}$$

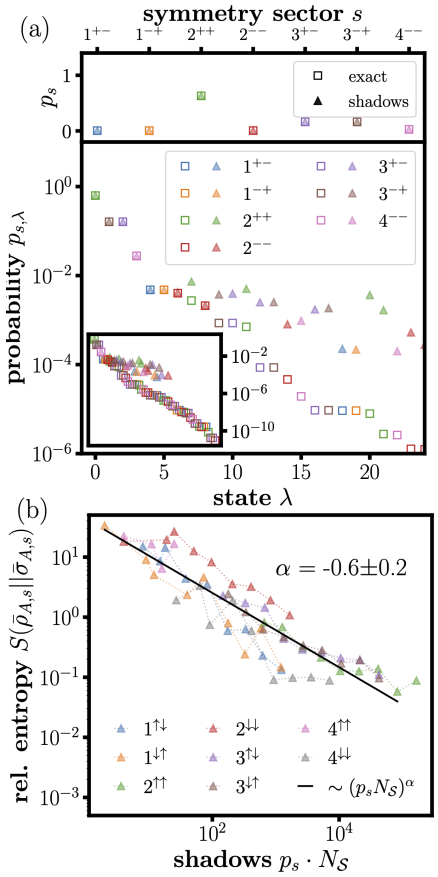


Figure 12.6: (a) Bottom: Shadow-reconstructed symmetry-resolved Schmidt spectrum, $m \cdot a = 0.05$, $e/m = 1$, $\ell = 32$ and $N_S = 2^{17}$. Inset: Full spectrum. Top: Probability per sector p_s. (b) Sector-wise relative entropy $S(\bar{\rho}_{A,s}||\bar{\sigma}_{A,s})$ between (normalized) exact $\bar{\rho}_{A,s}$ and shadow-reconstructed reduced density matrices $\bar{\sigma}_{A,s}$, as a function of shadow number N_S with $m \cdot a = 0.05$, $e/m = 1$ and $\ell = 32$.

where $\mathbb{E}[\dots] \equiv (1/N_S) \sum_{i=0}^{N_S-1} [\dots]$ is the N_S shadow average and $\mathcal{M}_s^{-1}(X) \equiv (d_s + 1)X - \mathrm{Tr}_s[X]\, \mathbb{I}_s$; Tr_s and $\mathbb{I}_s$ are the sector-wise trace and identity, respectively. The bar indicates normalization, i.e. $\bar{\sigma}_{A,s} \equiv \sigma_{A,s}/p_s$, $\mathrm{Tr}_s[\bar{\sigma}_{A,s}] = 1$, $p_s \equiv \mathrm{Tr}_s[\sigma_{A,s}]$ is simply the number of shadows measured in one sector relative to N_S.

Fig. 12.6 compactly summarizes the results of this analysis, showing the symmetry-resolved Schmidt spectrum of ρ_A in (a), comparing exact results (empty squares) versus shadows (filled triangles) for $m \cdot a = 0.05$, $e/m = 1$, $\ell = 32$ and $N_S = 2^{17}$. The eigenvalue spectrum is well reproduced down to $P_{s,\lambda} \approx 10^{-2} - 10^{-3}$; the inset shows the full spectrum and the top of (a)

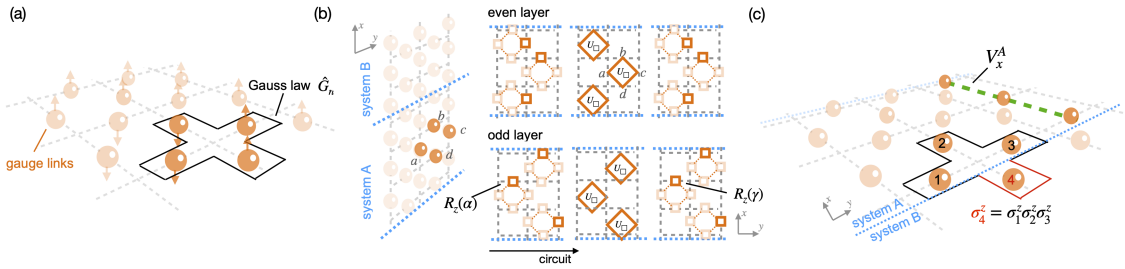


Figure 12.7: (a) Illustration of $\mathbb{Z}_2{}^{2+1}$ LGT, including Hilbert space and Gauss law constraints G_j. (b) Random measurement circuits: We work with an even-odd alternating layers consisting of randomly placed 'electric' rotations $R_z(\alpha)$, $R_z(\gamma)$ and plaquette rotations $U_\square(\beta)$ (orange squares), approximating k- designs for sufficient circuit depth. (c) The symmetry structure of ρ_A originates from Gauss laws at entanglement boundaries and a non-local 'ribbon' operator V_x^A spanning the two entanglement cuts.

the accurately recovered probability for each block, p_s. Fig. 12.6(b) shows the relative entropy, $S(\bar{\rho}_{A,s}||\bar{\sigma}_{A,s}) \equiv \mathrm{Tr}_s[\bar{\rho}_{A,s}\log(\bar{\rho}_{A,s}) - \bar{\rho}_{A,s}\log(\bar{\sigma}_{A,s})]$, between exact $\bar{\rho}_{A,s}$ and shadow-reconstructed $\bar{\sigma}_{A,s}$, as a function of shadows per sector $p_s \cdot N_{\mathcal{S}}$.[4] The fit (black curve) indicates approximate power law $\sim (p_s N_{\mathcal{E}})^{-\frac{1}{2}}$ convergence. We note that automatically fewer shadows are sampled in less important sectors.

While $\mathbb{Z}_2^{1+1}$ LGT serves as a useful case study, we next consider $2+1$ dimensions where our ability to extract LGT entanglement structure provides a window to studying topologically ordered systems. We add that symmetry-conscious shadows based on deep scrambling k-designs are not ideal for estimating local observables, a single layer scheme is better suited for this task [444].

12.4 Experimental Verification of Topological Phases

Finally, our scheme can facilitate the experimental identification of topological phases, an inherently difficult task because of the non-detectability of such phases by local measurements [433, 434]. Our proposed strategy is to combine symmetry-conscious randomization with classical shadow or entanglement Hamiltonian tomography to measure the entanglement gap of topologically ordered states, drawing on the foundational work by Li and Haldane [370].

For convenience, we will continue utilizing k-designs in a large layer limit—much larger than necessary. It is important to reiterate that these choices are not integral to our protocol but are adopted for illustration and convenience, because of the simple channel inversion associated with them for classical shadows and for easier comparison across different schemes. Any tomographically complete scheme is suitable for this task, including shallow depth circuits that are more likely to be realistically employed in near-term experiments.

We focus on a model related to the toric code, used in [433], $\mathbb{Z}_2$ LGT in (2+1)d spacetime dimensions (${\mathbb{Z}_2}^{2+1}$). This model is graphically illustrated in Fig. 12.7(a), consisting of spin 1/2 degrees of freedom placed on the links (j, b) of a two-dimensional rectangular $N_x \times N_y$ lattice, where $j = (j_x, j_y)$ and $b = x, y$ is the direction of a link. The Hamiltonian is given by

$$H = -K \sum_j \sigma^x_{j,x} \sigma^x_{j+\hat{x},y} \sigma^x_{j+\hat{y},x} \sigma^x_{j,y} - g \sum_{j,b} \sigma^z_{j,b} \,, \tag{12.17}$$

[4]Because the average shadow density matrix at finite N_S might not be positive definite, we follow the regularization strategy of [443] to project onto a positive definite matrix. The error of the scaling exponent includes an estimate of the regularization dependence, estimated by comparing with replacing $\log(\cdot) \to \frac{1}{2}\log(\cdot^2)$.

and superselection sectors are determined by Gauss laws ($[H, \hat{G}_j] = 0$),

$$\hat{G}_j \equiv \prod_{l \in +(j)} \sigma_l^z, \tag{12.18}$$

where the product is over neighboring links to each lattice site j, see Fig. 12.7(a) for illustration where l runs over links labelled (1,2,3,4); physical states obey $\hat{G}_j|\psi\rangle = +1|\psi\rangle$. We assume periodic boundary conditions along the x-direction, and fixed or periodic boundary conditions (BCs) in the y-direction; for fixed BCs, Eq. (12.18) involves three links at the y-boundary.

Our primary result for this model, and, consequently, for the [chapter], is to demonstrate that symmetry-respecting randomized measurement schemes enable an appealing and experimentally tractable route to verifying topological order. Ultimately, we shall see that by leveraging the approach known as entanglement Hamiltonian tomography (EHT) [336, 338, 339] and an ansatz for the Entanglement Hamiltonian inspired by the Bisognano-Wichmann theorem [445, 446], we can accurately extract much of the Schmidt spectrum of reduced density matrices ρ_A of the ground state in this model, allowing us to detect topological order via the presence of a so-called 'entanglement gap' in the spectrum.

12.4.1 Symmetry-Respecting Randomized Circuits

We are interested in the entanglement properties of $\rho_A \equiv \mathrm{Tr}_{\bar{A}}(\rho)$, where A is a bipartition obtained by separating the lattice along the x-direction, with entanglement cuts at $j_x = 0$ and $j_x = N_x^A - 1$ so that $N_x = N_x^A + N_x^{\bar{A}}$. The boundary is such that system A contains the y-direction links at $j_x = 0$ and $j_x = N_x^A - 1$, see Fig. 12.7(c) or [363] for more details, with $N_\square = (N_x^A - 1) \times N_y$ plaquettes ($2N_x^A N_y - N_y - N_x^A$ qubits) in A. We seek a gauge invariant,

symmetry-respecting family of random circuits that form k-designs acting on every symmetry sector of ρ_A. Our ansatz is illustrated in Fig. 12.7(b), made of alternating even-odd half-layers, consisting first of $R_z(\alpha) \equiv \exp\{i\alpha\sigma_i^z\}$ rotations randomly placed at one side $i \in \{a, b, c, d\}$ of a plaquette (orange squares), followed by $U_\square \equiv U_\square(\beta) \equiv \exp\{i\beta\sigma_a^x\sigma_b^x\sigma_c^x\sigma_d^x\}$, and again by a $R_z(\gamma) \equiv \exp\{i\gamma\sigma_i^z\}$, placed at the same random $i \in \{a, b, c, d\}$. For every plaquette, the angles α, β, γ are drawn according to a ZXZ decomposition of a CUE matrix,

$$U_{\text{CUE}} \equiv e^{i\delta} \begin{pmatrix} e^{i(\alpha+\gamma)} \cos\beta & ie^{-i(\alpha-\gamma)} \sin\beta \\ ie^{i(\alpha-\gamma)} \sin\beta & e^{-i(\alpha+\gamma)} \cos\beta \end{pmatrix}, \tag{12.19}$$

where the phase δ is irrelevant. In Appendix H.4, we verify numerically that these random circuits form an approximate unitary 2-design (k-design).

Before continuing, we discuss the symmetries of $\rho_A \equiv \text{Tr}_{\bar{A}}[\rho]$, depicted in Fig. 12.7(c). The Gauss laws at entanglement boundaries allow us to write

$$\sigma_4^z = \sigma_1^z\sigma_2^z\sigma_3^z \,. \tag{12.20}$$

Here, σ_4^z is the 'electric field' operator just outside A, and $\sigma_1^z\sigma_2^z\sigma_3^z$ is just inside A. For all $2N_y$ boundary sites, the operator $\sigma_1^z\sigma_2^z\sigma_3^z$ is a symmetry of ρ_A, i.e. $[\sigma_1^z\sigma_2^z\sigma_3^z, \rho_A] = 0$, if ρ is physical (i.e. Gauss law respecting). We label simultaneous eigensectors of all Eq. (12.20) as $s \in \{\uparrow, \downarrow\}^{2^{2N_y}}$; an example is $s = \uparrow\uparrow\downarrow\uparrow$ (for $N_y = 2$) where the first N_y bits are the eigenvalues

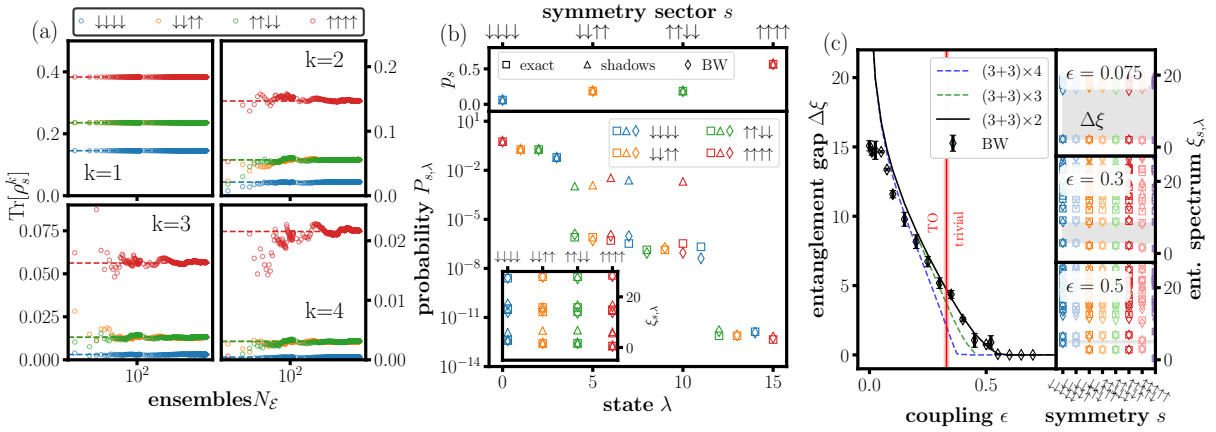


Figure 12.8: (a) Sector-wise k-purities of ρ_A of the $\mathbb{Z}_2^{2+1}$ ground state, from Eq. (12.8), for $N_x \times N_y = (3+5) \times 2$, ϵ=0.1 and ℓ=64 layers and fixed BC in y. (b) Bottom: Shadow- ($N_\mathcal{S} = 2^{16}$) versus BW-EHT-reconstructed ($N_\mathcal{E} = 50$, $N_M = 1024$) Schmidt spectrum $P_{s,\lambda}$; with $N_x \times N_y = (3+5) \times 2$, ϵ=0.2 and ℓ=64. Top: probability per sector p_s; inset: entanglement spectrum $\xi_{s,\lambda} = -\log(P_{s,\lambda})$. (c) Entanglement gap Δ_ξ between 'low- and high-energy' parts of the ES as a function of ϵ, reconstructed using the BW-EHT scheme, for $N_x \times N_y = (3+3) \times 2$ (PBC in y), $N_{\text{BW}} = 50$, $N_{\text{shots}} = 1024$, and $\ell = 64$. A horizontal red line is the infinite volume limit, $\epsilon_c = 0.33 \pm 0.01$. Right: Symmetry-resolved Entanglement spectra $\xi_{s,\lambda}$ for $\epsilon = 0.075, 03, 0.5$.

of Eq. (12.20) at $j_x = 0$ and the other N_y at $j_x = N_x^A - 1$. Additionally, a 'ribbon' operator

$$V_x^A \equiv \prod_{l \in \mathcal{C}} \sigma_l^z \tag{12.21}$$

commutes with ρ_A, $[\rho_A, V_x^A] = 0$, where $l \in \mathcal{C}$ indicates the links intersected by a contour $\mathcal{C}$ through the centers of plaquettes, from one boundary to the other in an arbitrary path, see the green dashed line in Fig. 12.7(c). For fixed y-BCs V_x^A is not independent (it is determined by fixing all sectors Eq. (12.20)), but for y-PBC its eigensectors are independent and labelled by an additional $\uparrow/\downarrow$, so that $s \in \{\uparrow, \downarrow\}^{2^{2N_y+1}}$.

12.4.2 Classical Shadows

As a warm-up, we first study again k-purities for the new model, thereby verifying correctness of the randomization, before moving on to reconstructing a classical shadow representation of the state that we will use to infer the presence or absence of TO. In Fig. 12.8(a) we display k-purities $\mathrm{Tr}[\rho_{A,s}^k]$ ($k = 1, 2, 3, 4$) of the $\mathbb{Z}_2^{2+1}$ ground state at $\epsilon \equiv g/K = 0.1$, following the approach outlined in Sect. 12.3, as a function of $N_{\mathcal{E}}$ and in the infinite shot limit ($N_M \to \infty$), with fixed y-BCs, $N_x \times N_y = (3+5) \times 2$ and $\ell = 64$. As before, the $k = 1$ results, $p_s = \mathrm{Tr}[\rho_{A,s}]$, are exact by design while $k = 2, 3, 4$-entropies are reproduced with increasing $N_{\mathcal{E}}$.

Finally, aiming at reconstructing the entanglement spectrum from classical shadows to diagnose TO, in Fig. 12.8(b), we show the Schmidt spectrum $P_{s,\lambda}$ for $\epsilon = 0.2$, comparing shadow-reconstructed ($N_{\mathcal{S}} = 2^{16} = 65,536$, triangles) versus exact results (squares). Different colors represent symmetry sectors s with weights $p_s \equiv \mathrm{Tr}_s[\rho_{A,s}]$ in the top panel. Large Schmidt values are well reproduced down to approximately 10^{-2}, beyond which we observe significant deviations that prevent us from reliably determining the entanglement gap.

12.4.3 Detecting Topological Order Through Entanglement Hamiltonian Tomography

Aiming at higher precision, we explore an alternative approach, Entanglement Hamiltonian Tomography (EHT) [336, 338, 339]. The basic idea is to parameterize the reduced density matrix

by an Entanglement Hamiltonian (EH),

$$H_A = -\log\left[\rho_A\right]. \tag{12.22}$$

A (heuristic) parameterization of ground state EHs, inspired by the Bisognano-Wichmann (BW) theorem [445, 446], is

$$H_A \equiv H_A[\{\beta_{\mathcal{O}}\}] = \sum_{\mathcal{O}} \beta_{\mathcal{O}} H_{\mathcal{O}}\,, \tag{12.23}$$

where $H_{\mathcal{O}}$ are the local operators comprising the physical Hamiltonian, Eq. (12.17), and $\beta_{\mathcal{O}}$ a local 'temperature', varying with the distance of $\mathcal{O}$ from the entanglement cut(s) (also depending on j_y if translation invariance in y is broken, i.e. for fixed y-BCs). The applicability and accuracy of this ansatz was investigated for $\mathbb{Z}_2^{2+1}$ in [363].

To extract the $\{\beta_{\mathcal{O}}\}$, we follow [336] and first measure probabilities $P_U(b,s)$ in $N_{\mathcal{E}}$ random bases and with N_M shots each. We then minimize

$$\sum_b \left\langle \left(P_U(b,s) - \mathrm{Tr}_s\left[\bar{\rho}_{A,s} U_s |b,s\rangle\langle b,s| U_s^\dagger\right]\right)^2 \right\rangle_{\mathcal{E}}, \tag{12.24}$$

with respect to $\{\beta_i\}$ via classical post-processing, where $\bar{\rho}_{A,s} \equiv \bar{\rho}_{A,s}[\{\beta_{\mathcal{O}}\}] \sim \exp\{-H_{A,s}[\{\beta_{\mathcal{O}}\}]\}$, normalized so that $\mathrm{Tr}_s[\bar{\rho}_{A,s}[\{\beta_{\mathcal{O}}\}] = p_s$. $H_{A,s}[\{\beta_{\mathcal{O}}\}] \equiv \sum_{\mathcal{O}} \beta_{\mathcal{O}} H_{s,\mathcal{O}}$ is the EH with $H_{s,\mathcal{O}}$ restricted to a symmetry sector s. Because the BW optimization is performed sector-wise, matrices of size d_s are involved, versus the dimension of A, d_A. In the infinite measurement limit, the optimization will yield one universal set $\{\beta_{\mathcal{O}}\}$ for all s, but in practice we work with normalized $\bar{\rho}_{A,s} \equiv \rho_{A,s}/p_s$ and $\bar{P}_U(b,s) = P_U(b,s)/p_s$, so that the extracted $\{\beta_{\mathcal{O},s}\}$ depend on that

normalization and differ from $\{\beta_{\mathcal{O}}\}$, see Appendix H.4 for details.

Results of the BW-EHT-optimization for the Schmidt spectrum $P_{s,\lambda}$ are displayed as diamond symbols in Fig. 12.8(b) along with shadow results, for $N_{\mathcal{E}} = 50$ and $N_M = 1024$. Despite comparable cost ($N_{\mathcal{E}} \cdot N_M = 51,200$) relative to the classical shadow approach ($N_{\mathcal{S}} = 2^{16} = 65,536$), BW-EHT reproduces the eigenvalue spectrum much more accurately; values as small as 10^{-6} are approximately recovered and even eigenvalues as small as 10^{-11} are not far off. The inset of Fig. 12.8(b) shows the entanglement spectrum (ES), i.e. the spectrum of the EH, which is also well reproduced. The apparent advantage of the BW-EHT approach comes at the expense of generality, it is tailored for ground states (it can be extended to non-equilibrium states [336, 410]) while classical shadows work regardless of the state.

Enabled by the performance of the BW-EHT optimization, we focus on a practical application: detecting topological order (TO) of quantum states. Ground states of $\mathbb{Z}_2^{2+1}$ are separated (in the infinite volume limit) into topologically ordered, $\epsilon < \epsilon_c$, and trivial states, $\epsilon > \epsilon_c$, with a phase transition at a critical coupling ϵ_c. Li and Haldane's entanglement-boundary conjecture [363, 370] asserts that TO states are 'entanglement-gapped', i.e. their ES has separated low energy (large Schmidt values) and a high energy (small Schmidt values) parts. Further, the low lying part is (up to rescaling) identical to the spectrum of a conformal field theory (CFT) describing gapless excitations at the edge of the system. We focus here on measuring the existence of an entanglement gap $\Delta\xi$ to detect TO, which has been shown as very a robust order parameter for the TO transition in this model even for very small systems [363].

Without loss of generality, to reduce finite size effects, we focus on periodic boundary conditions in y (a torus) for the BW-EHT analysis. In Fig. 12.8(c) we show the entanglement gap $\Delta\xi$ for $N_x \times N_y = (3+3) \times 2$, $N_{\mathcal{E}} = 50$, $N_M = 1024$, and $\ell = 64$ (black diamonds), compared

to exact results (black solid line). Error bars represent the combined statistical error due to finite $N_{\mathcal{E}}$ and N_M, see Appendix H.4 for details. A vertical red line indicates the infinite volume extrapolated value $\epsilon_c = 0.33 \pm 0.01$ [447]. We also show result for $N_x \times N_y = (3+3) \times 3$ (green dashed line) and $N_x \times N_y = (3+3) \times 4$ (blue dashed line), taken from [363] and approaching the infinite volume limit to within less than 10%. Side panels show the BW-EHT-reconstructed sector-wise ES for $\epsilon = 0.075, 0.3, 0.5$, demonstrating the closing of the entanglement gap (gray shaded area) at ϵ_c where our results approximately reproduce the phase transition.

We could not numerically simulate systems larger than $N_x \times N_y = (3+3) \times 2$. For example, a $N_x \times N_y = (3+3) \times 4$ lattice of 48 qubits (20 qubits in the subsystem) exhausts our classical computational resources.[5] However, the classical (shadow- or BW-EHT-) analysis is simple for such a system if it were prepared in experiment because, while $d_A = 2^{2N_x^A N_y - N_y} = 2^{20} = 1,048,576$, the analysis is restricted to symmetry blocks of only $d_s = 2^{N_x^A N_y - N_y} = 2^8 = 256$ states. This is a significant (in fact, exponential) reduction in the space over which the state is randomized, but d_s still grows exponentially with the subsystem size, albeit much slower than d_A.

12.5 Conclusion and Outlook

In this manuscript, we proposed randomized measurement protocols for lattice models that leverage symmetries, focusing primarily on LGT entanglement structure exploration. We devised deep-scrambling circuits that realize symmetry-conscious k-designs and illustrated their use in simple gauge and non-gauge model examples. Our approach is intuitive and, there-

[5] The results for $N_x \times N_y = (3+3) \times 4$ and $(3+3) \times 3$ were obtained using exact diagonalization and working with dual formulations of $\mathbb{Z}_2^{2+1}$ [363].

fore, easily generalizable: by examining the physical Hamiltonian one can readily identify basic symmetry-preserving interactions which can be used as the generators of a randomized measurement scheme. Consequently, if a particular physical Hamiltonian can be realized, so can our measurement scheme.

Symmetry-conscious randomized measurement schemes like those considered here have lower sampling costs compared to symmetry-ignorant schemes by avoiding randomizing over non-relevant Hilbert space parts. In particular, one obtains a sampling cost (to realize a 2- (k-)design) that scales with block size d_s, instead of Hilbert space dimension $d_{\mathcal{H}}$. This reduction can be exponential, e.g., for particle number conserving systems away from half-filling, or for LGTs due to randomizing only over the physical sector of Hilbert space. In constructing k-designs, while still efficient, they also incur a somewhat larger circuit complexity.

Using such symmetry-conscious randomized measurement, our primary goal was to provide a practical scheme for measuring LGT entanglement structure, a potential useful route e.g., for quantum simulating high energy and nuclear physics [230, 448–452], e.g. to understand Quantum Chromodynamics (QCD) where entanglement is largely unexplored [309, 363, 453–459], or detecting topologial order. We illustrated our approach in a simple (1+1)d LGT example, $\mathbb{Z}_2$ coupled to staggered matter, where we extracted symmetry-resolved k-purities and von Neumann entanglement entropies, and separated their symmetry- and distillable components.

We then focused on $\mathbb{Z}_2$ in $2+1$ dimensions where the intricate structure of gauge symmetric states can lead to topologically ordered (TO) phases. These are currently receiving great attention, including experimental realizations in AMO and solid-state platforms [433, 434], motivated by applications such as fractional quantum Hall effect states [460, 461] or fault-tolerant quantum computation and storage [271, 424–430]. A difficulty is that TO cannot be probed by measuring

local operators, a serious impediment for its experimental verification. Our approach to overcome this is based on measuring the entanglement structure of such systems. While the importance of entanglement as a robust indicator of topological order was realized long ago [462–464], we developed a concrete random-measurement scheme, following the logic of Li and Haldane [370], that uses a state presentation in terms of Entanglement Hamiltonians (EH) and is based on measuring entanglement gaps of their (symmetry-resolved) spectrum using a tomographic protocol based on the Bisognano-Wichmann theorem [336]. Remarkably, performing random measurements on very small subsystems as small as $N_x^A \times N_y = 3 \times 2$, we observe a relatively sharp TO-to-trivial phase transition. While our focus was on $\mathbb{Z}_2^{2+1}$, the protocol can be easily generalized to other systems.

A benefit of our approach, not explicitly explored in the main text, is that symmetry-conscious randomization allows for a rudimentary, but useful, near-term error-mitigation strategy similar to that discussed in [395]. A feature of symmetry-conscious randomization is that symmetries of the input states are not lost and can be measured. Thus, machine errors that violate those symmetries are detectable after randomization, suggesting that e.g., a postselection of measurement results can improve the computation (at the cost of reduced statistics). Another potential application is approximate Haar random state preparation for thermal state algorithms [411]. Finally, we expect symmetry-respecting randomized measurement schemes to be useful to investigate thermal systems, including e.g, systems with non-Abelian conserved charges [465–476].

There are many future extensions of our work. For example, while our approach significantly reduces algorithmic costs compared to a symmetry-ignorant scheme, extracting entanglement entropies and structure still relies on classical post-processing which ultimately scales exponentially with system size, an issue which can be addressed with quantum variational [338]

and machine learning techniques [21, 325]. Extending the robust numerical analysis performed here to provide analytical performance guarantees for the circuit depth and sampling complexity of the random circuits in this work is also of clear interest. We also emphasize that realizing an approximate k-design is a sufficient, but not necessary, condition for randomized measurement protocols, and more studies regarding optimal randomization for certain observables are needed.[6] Developing a formalism for shadow and entanglement tomography protocols that applies to systems with limited control or is independent of the circuit model would be useful for analog quantum simulation. Finally, while there are encouraging indications [344–346], the robustness of our scheme against experimental imperfections and noise should be investigated in detail.

Finally, we point out related work [321, 435], following a similar idea for fermionic systems with particle number symmetry, and also demonstrating a significant cost advantage. While not programmable enough to realize k-designs, we think this is a very useful approach for fermionic entanglement tomography. We also point out Ref. [479], proposing randomized measurement schemes that take advantage of *a priori* knowledge about observables of interest to improve sampling complexity and Refs. [480, 481], which, building on Ref. [335], propose randomized measurement schemes to extract symmetry-resolved purities. Their scheme is based on local random unitary transformations, in contrast to the symmetry-preserving unitaries we consider here.

[6] During the review process of this manuscript, some interesting analytical results along these lines for circuits with particle number symmetry were presented in [477], see also [478] for shallow depth circuits.

Chapter 13: Discussion and Further Directions for Part III

In Chapters 11-12, we considered two problems relevant for quantum simulation: (1) the problem of mapping fermions to qubits and (2) the design of symmetry-aware randomized measurement protocols for probing properties of states in lattice gauge theories. While these are distinct problems, both share a common theme of seeking to optimize computational resources in a way that respects the algebraic and/or group-theoretic structure of the underlying states.

In line with this general approach, significant work on similar optimizations for fermion-to-qubit mappings has been done since the paper that makes up Chapter 11 was published. Of particular note, Ref. [482] considers an even more general set of mappings than we do, searching over the space of mappings from Majorana monomials to Pauli operators to optimize the choice of fermion-to-qubit mapping in a hardware-aware way. While further algorithmic improvements can be expected, perhaps the most promising direction for future work is to combine these sorts of optimizations with a larger suite of hardware-aware optimizations for quantum simulation algorithms. One can also expect large improvements by targeting such optimizations to particular problems of interest. For instance, we were able to find large improvements in a nearly end-to-end cost analysis for simulating a particular set of nuclear effective field theories on the lattice compared to general purpose performance bounds [23].

On the topic of symmetry-aware randomized measurement schemes, there is still much

work to be done. Chapter 12 was largely numerical, so analytic performance guarantees for the schemes considered there are of clear interest. In addition, we focused primarily on deep randomizing circuits that produced approximate unitary k-designs. Shallow circuit versions of these symmetry-aware schemes is an important next step; such schemes should be more immediately experimentally feasible and would enable efficient estimation of local observables. Shallow randomizing circuits have recently been considered for the case of particle number symmetry in Ref. [478], but similar proofs for gauge symmetries do not yet exist. Furthermore, extending these ideas beyond Z_2 lattice gauge theory remains an important avenue for future work.

Part IV

Appendices

Appendix A: Technical details of the results reported in 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.1.1. *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. Further the probe state must be of the form*

$$|\psi\rangle = \frac{|0\rangle|\varphi_0\rangle + e^{i\phi}|1\rangle|\varphi_1\rangle}{\sqrt{2}}, \tag{A.1}$$

for all times $s \in [0, t]$*, where* $\phi, |\varphi_0\rangle, |\varphi_1\rangle$ *are arbitrary states on the* $d-1$ *remaining sensor qubits plus, potentially, the arbitrary number of ancilla—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) of the main text):

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

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) of the main text), we can write

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

$$= 4\left[\int_0^t ds \int_0^t ds' \langle\psi(0)|\hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)\hat{U}^\dagger(s')\hat{g}_1\hat{U}(s')|\psi(0)\rangle\right] - 4\left[\int_0^t ds \langle\psi(0)|\hat{U}^\dagger(s)\hat{g}_1\hat{U}(s)|\psi(0)\rangle\right]^2 \tag{A.4}$$

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

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

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

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

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

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

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

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 [38], and the final equality uses the fact that $\hat{g}_1 = \hat{\sigma}_1^z/2$ has seminorm 1.[1]

[1]Note that the above block of equations relies on the fact that we are using the fixed Hilbert space of qubit

Via Eq. (2.8) of the main text (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). Eq. (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 $|\psi(0)\rangle = \frac{1}{\sqrt{2}}\left(|\lambda_{\min}\rangle + e^{i\phi}|\lambda_{\max}\rangle\right)$, where $|\lambda_{\min}\rangle$ and $|\lambda_{\max}\rangle$ 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,

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

The statement of the lemma immediately follows. □

sensors. Were one to extend this derivation to photonic sensors with indefinite particle number, the results would not immediately follow.

Note that Lemma A.1.1 holds for any optimal protocol, not just those using our cat-like 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 the main text. In particular, we show that the Fisher information matrix condition for saturability in Eq. (2.8) of the main text is satisfied by solutions to Eq. (2.13) of the main text when we consider protocols that use $\hat{\sigma}^x$ and CNOT controls to switch between families of cat-like 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 (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.1.1, we can show that for *any* optimal protocol (i.e., not just those using

our cat-like states)

$$\mathcal{F}(\boldsymbol{\theta})_{1j} = 2\langle\{\hat{\mathcal{H}}_1, \hat{\mathcal{H}}_j\}\rangle \tag{A.15}$$

$$= 2\int_0^t ds \int_0^t ds' \langle\psi(0)|\{\hat{g}_1(s), \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{A.16}$$

$$= 2\int_0^t ds \int_0^t ds' \langle\psi(0)|\{\hat{g}_1, \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{A.17}$$

$$= 2t\int_0^t ds' \langle\psi(0)|\{\hat{g}_1, \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{A.18}$$

$$= 2t\int_0^t ds' \langle\psi(0)|\{\hat{g}_1(s'), \hat{U}^\dagger(s')\hat{g}_j\hat{U}(s')\}|\psi(0)\rangle \tag{A.19}$$

$$= 4t\int_0^t ds' \langle\psi(s')|\hat{g}_1\hat{g}_j|\psi(s')\rangle \tag{A.20}$$

$$= t\int_0^t ds' \langle\psi(s')|\hat{\sigma}_1^z\hat{\sigma}_j^z|\psi(s')\rangle. \tag{A.21}$$

The third and fifth equalities come from the argument in the proof of Lemma A.1.1 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

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

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. [66–68] described in the main text 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 pro-

tocols 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 the main text, 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 [69, 70]. The phase estimation protocols of Refs. [66–68] demonstrate how to optimize resources to deal with this issue, while still saturating the single-shot bound in Eq. (2.2) of the main text 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. Ref. [71] 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. [68], which corrects a few minor errors in Ref. [66], as noted in the corresponding erratum [67]. We refer the reader to Ref. [68] 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 j-th 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 j-th stage, using one of our protocols for a time $M_j \delta t$, we prepare $2\nu_j$ independent copies of the state

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

From now on we will drop the $d-1$ qubit sensors in the state $|0...0\rangle$, 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. [68] 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 two 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 two-fold 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 $\nu_j \to \infty$. These measurements are non-adapative, 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 2-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. Ref. [68] demonstrates rigorously that picking $M_j = 2^{j-1}$ for $j \in \{1,\cdots,K\}$ 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 [71], but with the convenient feature that the protocol uses non-adaptive measurements. We refer the interested reader to Ref. [68] for detailed derivation of the results sketched here.

It is also worth noting that other protocols are possible. For instance, in Ref. [65], 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. [66–68]) 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 2.5.1 in the main text to fully prove the result. For reference, this theorem is restated here.

Theorem A.4.1. *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) of the main text, 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 cat-like states considered in the main text, 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 [72, 73], 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 [483] is as follows:

Lemma A.4.1 (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 $1 < j < d$. Eq. (A.37) implies $\sum_{j>1} \alpha'_j y_j < -y_1$. $(T^{(k)})^\top$ has a row n^* given by $\boldsymbol{\tau}^{(n^*)} = (1, 0, \ldots, 0)$, 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 row 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 non-zero 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

$$\begin{aligned}\sum_j \left|\frac{\alpha_j}{\alpha_1} t^2\right| &= \sum_j |\mathcal{F}(\boldsymbol{\theta})_{j1}| \\ &\leq t \sum_j \int_0^t ds' E_j(s') = t \int_0^t ds' \sum_j E_j(s) \\ &\leq t \int_0^t ds'(k-1) = (k-1)t^2. \end{aligned} \tag{A.48}$$

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 cat-like 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.5.1 (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, $\operatorname{sgn}(T_{ij}) \in \{0, \operatorname{sgn}(\alpha_j)\}$.*

We now prove Theorem 2.6.1 from the main text, which we again repeat for simplicity.

Theorem A.5.1. *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 [72, 73] 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 $y_1 \geq 0$. Further, 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 $(T_+^{(k)})^\top$ 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 $(T_+^{(k)})^\top \boldsymbol{y} \geq 0$ 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 $(T_+^{(k)})^\top$ 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 $\mathrm{sgn}(\alpha_i)$ (and leave 0s untouched). □

Thus, Lemma 2.6.1 and Theorem 2.6.1 prove there exist protocols that can minimize both instantaneous entanglement (i.e., the maximum size of a cat-like 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 the main text 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) of the main text.

A.6.1 Generalizing Eq. (2.8) of the Main Text

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 the main text, 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) of the main text):

$$\mathcal{F}(\boldsymbol{q})_{11} = \frac{t^2}{\alpha_1^2}, \tag{A.58}$$

$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = 0 \quad (\forall\, i \neq 1). \tag{A.59}$$

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)-(A.59) are satisfied if and only if

$$\sum_{i \in L} \frac{\mathrm{sgn}(\alpha_1)}{\mathrm{sgn}(\alpha_i)} \mathcal{F}(\boldsymbol{\theta})_{ji} \lambda_i = \frac{\alpha_j}{\alpha_1} t^2, \tag{A.60}$$

where $\lambda_i \geq 0$ such that $\sum_i \lambda_i = 1$. If $|L| = 1$, this reduces to Eq. (2.8) of the main text.

It will be important to briefly recount how we obtain the single-parameter bound we are

trying to saturate [49, 59]. In particular, referring to Eq. (2.3) of the main text, 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 $\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(d)}\}$, where $\boldsymbol{\alpha}^{(1)} = \boldsymbol{\alpha}$. We then have that J^{-1} has rows given by these vectors. Let $\{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \cdots, \boldsymbol{\beta}^{(d)}\}$ 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

$$\hat{H} = \frac{1}{2}\boldsymbol{\theta}^\top \hat{\boldsymbol{\sigma}} + \hat{H}_c(s) = \frac{1}{2}\sum_{i=1}^{d} (\boldsymbol{\alpha}^{(i)} \cdot \boldsymbol{\theta})\boldsymbol{\beta}^{(i)} \cdot \hat{\boldsymbol{\sigma}} + \hat{H}_c(s), \tag{A.62}$$

where $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1^z, \cdots, \hat{\sigma}_d^z)^\top$. 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. [38]), we immediately have that

$$\|\hat{g}_q\|_s = \|\boldsymbol{\beta}\|_1, \tag{A.64}$$

and our tightest bound is given by

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1,$$
$$\text{s.t. } \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1. \tag{A.65}$$

Note that

$$1 = \sum_i \alpha_i \beta_i \leq \sum_i |\alpha_i||\beta_i| \leq |\alpha_1| \sum_i |\beta_i| = |\alpha_1| \|\boldsymbol{\beta}\|_1. \tag{A.66}$$

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 Eq. (A.58)-(A.59) as

$$\mathcal{F}(\boldsymbol{q})_{11} = \boldsymbol{\beta}^\top \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = \frac{t^2}{\alpha_1^2}, \tag{A.69}$$
$$\mathcal{F}(\boldsymbol{q})_{1i} = \mathcal{F}(\boldsymbol{q})_{i1} = (\boldsymbol{\beta}^{(i)})^\top \mathcal{F}(\boldsymbol{\theta}) \boldsymbol{\beta} = 0 \quad (\forall\, i \neq 1). \tag{A.70}$$

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)-(A.68) and the condition that $\mathrm{sgn}(\beta_i) = \mathrm{sgn}(\alpha_i)$ for $\beta_i \neq 0$, we write $\beta_i = \lambda_i \mathrm{sgn}(\alpha_i)/|\alpha_1|$, 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

$$\sum_{i\in L} \mathcal{F}(\boldsymbol{\theta})_{ij}\mathrm{sgn}(\alpha_i)\lambda_i = \sum_{i\in L} \mathcal{F}(\boldsymbol{\theta})_{ji}\mathrm{sgn}(\alpha_i)\lambda_i = \frac{t^2}{|\alpha_1|}\alpha_j, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0, \tag{A.72}$$

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) of the Main Text

At this point, we can generalize the derivation of Eq. (2.13) of the main text to this setting of more than one maximum element of $\boldsymbol{\alpha}$. In particular, Lemma A.1.1 can be immediately extended to the following:

Lemma A.6.1. *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*

$$|\psi\rangle = \frac{\left(\bigotimes_{j\in L}|b_j\rangle\right)|\varphi_0\rangle + e^{i\phi}\left(\bigotimes_{j\in L}|b_j+1\rangle\right)|\varphi_1\rangle}{\sqrt{2}}, \tag{A.74}$$

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 $\phi, |\varphi_0\rangle, |\varphi_1\rangle$ *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)); (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 cat-like state over computational basis states follows immediately via an identical calculation to the proof of Lemma A.1.1 for each $j \in L$. From Eq. (A.21) it follows directly that these cat-like 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.6.1, 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 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 $t_{l+1}^*-t_l^*=p_lt$. 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) of the main text 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 2.5.1 of the Main Text

Recall, we divided the proof into two parts. First, we showed the existence of an optimal protocol using k-partite entangled cat-like 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's 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 $\tau_j^{(n)}=\mathrm{sgn}(\alpha_j)/\mathrm{sgn}(\alpha_1)$ for all $j\in L$ and all $\boldsymbol{\tau}^{(n)}$, the first $|L|$ rows of $T^{(k)}$ yield redundant equations in Eq. (2.19) of the main text. 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. Further, 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

$$\frac{\alpha_j}{\alpha_1}t^2 = \frac{\operatorname{sgn}(\alpha_j)}{\operatorname{sgn}(\alpha_1)}t^2 = \sum_{i\in L}\frac{\operatorname{sgn}(\alpha_1)}{\operatorname{sgn}(\alpha_i)}\mathcal{F}(\boldsymbol{\theta})_{ij}\lambda_i, \tag{A.81}$$

which implies

$$t^2 = \sum_{i\in L}\frac{\operatorname{sgn}(\alpha_i)}{\operatorname{sgn}(\alpha_j)}\mathcal{F}(\boldsymbol{\theta})_{ij}\lambda_i. \tag{A.82}$$

This in turn implies that for $i, j \in L$

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = \frac{\operatorname{sgn}(\alpha_i)}{\operatorname{sgn}(\alpha_j)}t^2. \tag{A.83}$$

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: Technical details of the results reported in 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 [38]

$$[\Delta(\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}})_\rho]^2 \leq \frac{\|\boldsymbol{\beta} \cdot \hat{\boldsymbol{n}}\|^2_{s,N}}{4}, \tag{B.2}$$

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:

$$\begin{aligned} &\text{minimize (w.r.t. } \boldsymbol{\beta}) \ \|\boldsymbol{\beta}\cdot\hat{\boldsymbol{n}}\|_{s,N}, \\ &\text{subject to } \ \boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 1. \end{aligned} \tag{B.3}$$

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 the main text, define $\mathcal{P} := \{j \,|\, \alpha_j \geq 0\}$ and $\mathcal{N} := \{j \,|\, \alpha_j < 0\}$. We then have the following lemma.

Lemma B.1.1. *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 $\mathcal{J}_- \neq \emptyset$. 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.1.1 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 $\mathcal{P}=\emptyset$ ($\mathcal{N}=\emptyset$). A further simplification is enabled by another lemma.

Lemma B.1.2. *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.1.2 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

$$\|\hat{g}_q\|_{s,N}^2 = \frac{N^2}{\max(\sum_{i \in \mathcal{P}} \alpha_i, \sum_{i \in \mathcal{N}} \alpha_i)^2}. \tag{B.14}$$

Consequently, via the quantum Cramér-Rao bound, Eq. (3.10),

$$\mathcal{M} \geq \frac{\max\left\{\sum_{i \in \mathcal{P}} \alpha_i, \sum_{i \in \mathcal{N}} \alpha_i\right\}^2}{N^2 t^2} =: \frac{\max\left\{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}^2, \|\boldsymbol{\alpha}\|_{1,\mathcal{N}}^2\right\}}{N^2 t^2}, \tag{B.15}$$

which is Eq. (3.12) of the main text, 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., $\mathcal{N} = \emptyset$), this reduces to

$$\mathcal{M} \geq \frac{\|\boldsymbol{\alpha}\|_1^2}{N^2 t^2}, \tag{B.16}$$

which, as described in the main text, proves a conjecture from Ref. [9] 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. [79]. 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 [79], 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. [79] 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

$$
\sum_{j=1}^{d}(\Delta\hat{p}_j)^2\leq\overline{N}+\frac{d}{2}.
\tag{B.35}
$$

Eq. (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 [80] 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 IV of the main text.

B.4.1 An Optimal Protocol for Functions with Positive Coefficients

We begin by reviewing a protocol from Ref. [9] 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 the main text, this can be done via the robust phase estimation protocols of Refs. [66–68] 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. [68] and Ref. [484].

B.4.2 Extending the Optimal Protocol to Negative Coefficients

While not explicitly considered in Ref. [9], it is straightforward to extend the above protocol to the situation where $\mathcal{N} \neq \emptyset$, 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:

$$|\psi\rangle \propto \bigotimes_{j\in\mathcal{P}} |N\frac{\alpha_j}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\rangle |0\rangle^{\otimes|\mathcal{N}|}|0\rangle + |0\rangle^{\otimes|\mathcal{P}|}\bigotimes_{j\in\mathcal{N}} |N\frac{|\alpha_j|}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\rangle |N - N\frac{\|\boldsymbol{\alpha}\|_{1,\mathcal{N}}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\rangle, \tag{B.46}$$

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

$$|\psi_M\rangle \propto \bigotimes_{j\in\mathcal{P}} |N\frac{\alpha_j}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\rangle |0\rangle^{\otimes|\mathcal{N}|}|0\rangle + e^{i\boldsymbol{\alpha}\cdot\boldsymbol{\theta}\frac{NM}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}}|0\rangle^{\otimes|\mathcal{P}|}\bigotimes_{j\in\mathcal{N}} |N\frac{|\alpha_j|}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\rangle |N - N\frac{\|\boldsymbol{\alpha}\|_{1,\mathcal{N}}}{\|\boldsymbol{\alpha}\|_{1,\mathcal{P}}}\rangle. \tag{B.47}$$

This probe state is optimal in the sense of satisfying the Fisher information condition in Eq. (3.24). In the main text, 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 the main text, 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 3.4.1. 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 the main text and must obey the conditions

$$\begin{aligned} \|\boldsymbol{r}\|_1 &= M, \\ \boldsymbol{r} &\geq 0. \end{aligned} \tag{B.51}$$

Recall, W is a matrix whose columns are the vectors $\boldsymbol{\omega}_n \in \mathcal{W}$.

Here we explicitly demonstrate this. We can easily evaluate

$$\langle \hat{n}_j(m) \rangle = \langle \psi(\boldsymbol{\omega}^{(m)}; \varphi) | \hat{n}_j | \psi(\boldsymbol{\omega}^{(m)}; \varphi) \rangle = \frac{|\omega_j^{(m)}|}{2} \tag{B.52}$$

and

$$\begin{aligned} \langle \hat{n}_i(l) \hat{n}_j(m) \rangle &= \langle \psi(\boldsymbol{\omega}^{(l)}; \varphi) | \hat{n}_i U(m \leftrightarrow l) \hat{n}_j | \psi(\boldsymbol{\omega}^{(m)}; \varphi) \rangle \\ &= \frac{|\omega_i^{(l)} \omega_j^{(m)}|}{2} \langle \psi_l(\boldsymbol{\omega}^{(l)}; \varphi) | U(m \leftrightarrow l) | \psi_m(\boldsymbol{\omega}^{(m)}; \varphi) \rangle, \end{aligned} \tag{B.53}$$

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 non-zero 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

$$\sum_{j\in\mathcal{P}}\mathcal{F}(\boldsymbol{\theta})_{ij} = \mathrm{sgn}(\alpha_i)\left(\sum_{m=1}^{M}|\omega_i^{(m)}|\right)MN, \tag{B.58}$$

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 $\mathrm{sgn}(\omega_j^{(m)}) = \mathrm{sgn}(\alpha_j)$ 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 3.4.1

Here we provide a proof of Lemma 3.4.1 in the main text, restated here for convenience.

Lemma B.5.1 (Lemma 3.4.1 of main text)**.** *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

$$\begin{aligned}\sum_{i\in\mathcal{P}}\sum_{j\in\mathcal{P}}\mathcal{F}(\boldsymbol{\theta})_{ij} &= 4\sum_{l=1}^{M}\sum_{m=1}^{M}\mathrm{Cov}\left(\hat{P}(l),\hat{P}(m)\right)\\ &\leq 4\sum_{l=1}^{M}\sum_{m=1}^{M}\sqrt{\mathrm{Var}(\hat{P}(l))\mathrm{Var}(\hat{P}(m))} = 4\left(\sum_{l=1}^{M}\sqrt{\mathrm{Var}(\hat{P}(l))}\right)^2\\ &\leq 4\left(\sum_{l=1}^{M}\frac{\|\hat{P}(l)\|_{s,N}}{2}\right)^2\\ &\leq (NM)^2, \end{aligned} \tag{B.67}$$

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 $\|\hat{P}(l)\|_{s,N} \leq N$. 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,

$$\mathrm{Cov}\left(\hat{P}(l),\hat{P}(m)\right)^2 = \mathrm{Var}(\hat{P}(l))\mathrm{Var}(\hat{P}(m)), \tag{B.68}$$

$$\mathrm{Var}(\hat{P}(l)) = \frac{N^2}{4}. \tag{B.69}$$

The second condition, Eq. (B.69), means that, at all times, the state of our system must be of the form

$$\frac{|\boldsymbol{N}(l)\rangle_{\mathcal{P}}|\boldsymbol{0}\rangle_{\mathcal{NR}} + e^{i\varphi_l}|\boldsymbol{0}\rangle_{\mathcal{P}}|\boldsymbol{N'}(l)\rangle_{\mathcal{NR}}}{\sqrt{2}}, \tag{B.70}$$

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 $|\boldsymbol{N}\rangle_{\mathcal{P}}|\boldsymbol{0}\rangle_{\mathcal{NR}}$ 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, $|\boldsymbol{0}\rangle_{\mathcal{P}}|\boldsymbol{N'}(l)\rangle_{\mathcal{NR}}$ 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.5.2. *Let $|\psi_i\rangle$ be a state of the form in Lemma 3.4.1. 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

$$U_m^\dagger |\boldsymbol{N}(m)\rangle_{\mathcal{P}} |\mathbf{0}\rangle_{\mathcal{NR}} = |\boldsymbol{N}(0)\rangle_{\mathcal{P}} |\mathbf{0}\rangle_{\mathcal{NR}} \equiv |\psi_1\rangle_{\mathcal{P}}, \tag{B.80}$$

where we are defining $|\psi_1\rangle_{\mathcal{P}}, |\psi_1\rangle_{\mathcal{NR}}$ such that $|\psi_0\rangle \propto |\psi_1\rangle_{\mathcal{P}} + |\psi_1\rangle_{\mathcal{NR}}$ in the obvious way. Moving the unitary onto the right hand side of the equation yields

$$|\psi_m\rangle_{\mathcal{P}} = U_m |\psi_1\rangle_{\mathcal{P}}, \tag{B.81}$$

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 [this Appendix and the associated chapter], 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

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 this [Appendix and the associated chapter], 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 this [work] 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 [71], but the non-adaptive robust phase estimation scheme of Refs. [66–68] 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 the main text. 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 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. [66–68].

Appendix C: Technical Details of the Results Reported in Chapter 4

C.1 Justification of Using Single-Parameter Bound

In this section, we elaborate on artificially fixing $k-1$ degrees of freedom in order to use the single-parameter bound in Eq. (4.3) in the main text. We begin by showing that any choice of $\boldsymbol{\beta} \in \mathbb{R}^k$ satisfying $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1$ picks out a valid choice of a $k-1$ dimensional subspace that $\{\boldsymbol{\alpha}_n\}_{n=\{2,\ldots,k\}}$ must span such that the full set $\{\boldsymbol{\alpha}_n\}_{n=\{1,\ldots,k\}}$ is a valid basis.

We begin by noting that, formally, the basis of vectors $\{\boldsymbol{\alpha}_n\}_{n=\{1,\ldots,k\}}$ corresponds to the rows of the Jacobian matrix $J = [\boldsymbol{\alpha_1}, \ldots, \boldsymbol{\alpha_k}]^T$ of the coordinate transformation between $\boldsymbol{\theta}'$ and $\boldsymbol{q} = (q_1(\boldsymbol{\theta}'), q_2(\boldsymbol{\theta}'), \cdots, q_k(\boldsymbol{\theta}'))^T$. Further, there exists a dual basis of vectors $\{\boldsymbol{\beta}_n\}_{n=\{1,\ldots,k\}}$ corresponding to the columns of the inverse Jacobian matrix $J^{-1} = [\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \ldots, \boldsymbol{\beta_k}]$. Therefore,

$\boldsymbol{\alpha_n} \cdot \boldsymbol{\beta_m} = \delta_{nm}$ since $JJ^{-1} = I$. In particular, we have

$$\boldsymbol{\alpha_1} \cdot \boldsymbol{\beta_1} = \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 1. \tag{C.1}$$

Recall that $\boldsymbol{\alpha} = \nabla q(\boldsymbol{\theta'})$ is fixed by the quantity we desire to measure. Furthermore, we assume without loss of generality that $q_1(\boldsymbol{\theta}) = q(\boldsymbol{\theta})$. Now suppose we pick *any* $\boldsymbol{\beta} \in \mathbb{R}^k$ satisfying Eq. (C.1). If there is a valid basis $\{\boldsymbol{\alpha}_n\}_{n=\{1,\ldots,k\}}$ corresponding to this choice, we require that the $k-1$ vectors $\{\boldsymbol{\alpha}_n\}_{n=\{2,\ldots,k\}}$ span the orthogonal complement of $\boldsymbol{\beta}$. Furthermore, we require that these vectors be independent of $\boldsymbol{\alpha}$. This is clearly true for any valid basis $\{\boldsymbol{\alpha}_n\}_{n=\{2,\ldots,k\}}$ for the orthogonal complement of $\boldsymbol{\beta}$ as $\boldsymbol{\alpha}$ is not in this subspace via Eq. (C.1). Therefore, we have reduced the problem to that of picking the optimal choice of $\boldsymbol{\beta} \in \mathbb{R}^k$.

We now show via information theoretic arguments that there is such a choice of $\boldsymbol{\beta}$ that gives us no useful information about q and that therefore the sharpest bound obtained by optimizing over $\boldsymbol{\beta}$ is in fact saturable. Let $\mathcal{F}(\boldsymbol{\theta}) = (\boldsymbol{\ell_1}, \boldsymbol{\ell_2}, \cdots, \boldsymbol{\ell_k}) = (\boldsymbol{\ell_1}, \boldsymbol{\ell_2}, \cdots, \boldsymbol{\ell_k})^T$ be the Fisher information matrix with respect to the parameters $\boldsymbol{\theta}$, where we have explicitly indicated it is symmetric. Then we may use the previously defined Jacobian to obtain the Fisher information matrix with respect to $\boldsymbol{q} = (q_1(\boldsymbol{\theta'}) = q(\boldsymbol{\theta'}), q_2(\boldsymbol{\theta'}), \cdots, q_k(\boldsymbol{\theta'}))^T$:

$$\mathcal{F}(\boldsymbol{q}) = (J^{-1})^T \mathcal{F}(\boldsymbol{\theta'}) J^{-1}. \tag{C.2}$$

We note that, if $\mathcal{F}(\boldsymbol{q})_{1n} = \mathcal{F}(\boldsymbol{q})_{n1} = 0$ for all $n \neq 1$, then there is no information about the desired $q_1(\boldsymbol{\theta'})$ in the other $q_{n\neq 1}(\boldsymbol{\theta'})$. Therefore, if our bound is saturable, it must be possible to construct such an $\mathcal{F}(\boldsymbol{q})$.

Let $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha} = (a_1, \cdots, a_k)^T$. Since we know there is a protocol saturating our bound and since we know what it is, we propose the ansatz

$$\boldsymbol{\beta_1} = \left(\frac{1}{a_1}, 0, \cdots, 0\right)^T. \tag{C.3}$$

Eq. (C.3) clearly satisfies $\boldsymbol{\alpha}_1 \cdot \boldsymbol{\beta}_1 = 1$. Furthermore, we pick some choice of remaining basis vectors $\{\boldsymbol{\alpha}_n\}_{n=\{2,\ldots,k\}}$ such that

$$\boldsymbol{\alpha}_{n\neq 1}^T = (0, \boldsymbol{v_n}^T), \tag{C.4}$$

which satisfy $\boldsymbol{\alpha}_{n\neq 1} \cdot \boldsymbol{\beta}_1 = 0$. Define the $(k-1)\times(k-1)$-dimensional matrix $V = (\boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_k)^T$. Then define $U^T = V^{-1}$. Therefore, letting $U = (\boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_k)^T$, we have $\boldsymbol{u}_n \cdot \boldsymbol{v}_m = \delta_{mn}$. Defining $\boldsymbol{a} = (a_2, a_3, \cdots, a_k)^T$, we then can pick

$$\boldsymbol{\beta}_{n\neq 1}^T = \left(\frac{-\boldsymbol{u_n} \cdot \boldsymbol{a}}{a_1}, \boldsymbol{u_n}^T\right), \tag{C.5}$$

which clearly satisfies $\boldsymbol{\alpha}_n \cdot \boldsymbol{\beta}_m = \delta_{nm}$. We then have

$$\mathcal{F}(\boldsymbol{q})_{n1} = \boldsymbol{\beta}_n^T F(\boldsymbol{\theta}')\boldsymbol{\beta}_1 = \boldsymbol{\beta}_n^T \frac{\boldsymbol{\ell_1}}{a_1} = \frac{\boldsymbol{\ell_1} \cdot \boldsymbol{\beta_n}}{a_1} = \mathcal{F}(\boldsymbol{q})_{1n}. \tag{C.6}$$

The above equation implies

$$\mathcal{F}(\boldsymbol{q})_{11} = \frac{\ell_{11}}{a_1^2}, \tag{C.7}$$

where we let ℓ_{11} denote the first component of $\boldsymbol{\ell_1}$. Furthermore, if our choice of basis is to give

us no information, we must have, for $n \neq 1$,

$$\mathcal{F}(\boldsymbol{q})_{1n} = \boldsymbol{u_n} \cdot \left(-\frac{\ell_{11}}{a_1}\boldsymbol{a} + \boldsymbol{\ell_1'} \right) = 0, \tag{C.8}$$

where $\boldsymbol{\ell_1'} = (\ell_{12}, \ell_{13}, \cdots, \ell_{1k})^T$. In other words, this gives the off-diagonal elements zero. It is impossible to have k linearly independent vectors $\boldsymbol{u}_n$ all orthogonal to $\left(-\frac{\ell_{11}}{a_1}\boldsymbol{a} + \boldsymbol{\ell_1'} \right)$ in a k dimensional space, so we demand

$$-\frac{\ell_{11}}{a_1}\boldsymbol{a} + \boldsymbol{\ell_1'} = 0. \tag{C.9}$$

We note that, if $\mathcal{F}(\boldsymbol{\theta})$ is diagonal, this is impossible to satisfy. However, if $\boldsymbol{\ell_1} \propto \boldsymbol{\alpha}_1$, this is satisfied, which is in fact what is done in the linear protocol from Ref. [8] that we use as a subroutine in our protocol. Also note that a_1 must be the maximum-magnitude element of $\boldsymbol{\alpha}$ for this to be satisfiable due to the properties of the Fisher information matrix (namely $\ell_{11} \geq \ell_{1n}$ for $n \neq 1$). Without loss of generality, we let a_1 be this maximum-value element as the order of indexing our sensors is arbitrary. Therefore, we see that, by insisting that fixing $k-1$ degrees of freedom gives us no useful information, the protocol in Ref. [8] emerges naturally.

We note that one can find a somewhat related argument regarding the results of Ref. [8] in Ref. [9].

C.2 Proof of Validity of the Consistency Condition

Here we prove that, provided we can estimate $\boldsymbol{\theta}'$, the consistency condition $G^T(\boldsymbol{\theta}')\boldsymbol{w} = \boldsymbol{\alpha}$ [Eq. (4.9) in the main text] is satisfied for some $\boldsymbol{w}$. We use this result in Sec. C.4 to prove that using $\tilde{\boldsymbol{\theta}}$ (instead of $\boldsymbol{\theta}'$) in the second step of the analytic-function-case protocol of Sec. C.3

induces negligible errors.

We begin by recalling a standard definition.

Definition C.2.1. *An asymptotically unbiased estimator $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is one that asymptotically (in time t and the number of measurements μ) has $\mathrm{E}[\tilde{\boldsymbol{\theta}}] = \boldsymbol{\theta}$.*

We can now prove the following theorems.

Theorem C.2.1. *If we can make an asymptotically unbiased estimate of a function $q(\boldsymbol{\theta}')$ with arbitrarily small variance, then all $\boldsymbol{\beta}$ in the null space of $G(\boldsymbol{\theta}')$ lie in the orthogonal complement of $\boldsymbol{\alpha} = \nabla q(\boldsymbol{\theta}')$.*

Proof. Proceeding by contradiction, suppose we have a $\boldsymbol{\beta}$ satisfying $G(\boldsymbol{\theta}')\boldsymbol{\beta} = \mathbf{0}$ [*i.e.* $\boldsymbol{\beta}$ is in the null space of $G(\boldsymbol{\theta}')$] and $\boldsymbol{\alpha}\cdot\boldsymbol{\beta} \neq 0$ (*i.e.* $\boldsymbol{\beta}$ is *not* in the orthogonal complement of $\boldsymbol{\alpha}$). We can scale $\boldsymbol{\beta}$ by a constant to force $\boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 1$ and maintain $G(\boldsymbol{\theta}')\boldsymbol{\beta} = 0$. According to the bound in Eq. (4.7) of the main text, the MSE of any estimator of $q(\boldsymbol{\theta}')$ then approaches infinity. Thus we can't make an asymptotically unbiased estimate with arbitrarily small variance, a contradiction. □

Theorem C.2.2. *If we can make an asymptotically unbiased estimate of $q(\boldsymbol{\theta}')$ with arbitrarily small variance, then $G^T(\boldsymbol{\theta}')\boldsymbol{w} = \boldsymbol{\alpha} = \nabla q(\boldsymbol{\theta}')$ is consistent.*

Proof. Theorem C.2.1 implies that $\boldsymbol{\alpha}$ lies in the column space of $G^T(\boldsymbol{\theta}')$, as the null space of $G(\boldsymbol{\theta}')$ is the orthogonal complement of the column space of $G^T(\boldsymbol{\theta}')$. Thus the system $G^T(\boldsymbol{\theta}')\boldsymbol{w} = \boldsymbol{\alpha}$ is consistent. □

Corollary C.2.2.1. *If we can make asymptotically unbiased estimates of $\boldsymbol{\theta}'$ with arbitrarily small variance, then $G^T(\boldsymbol{\theta}')$ is full rank.*

Proof. If we can make an estimate of $\boldsymbol{\theta}'$, we can think of this as making an estimate of $q(\boldsymbol{\theta}') = \theta'_i$ for any i. Therefore, by Thm. C.2.1, we have that $G^T(\boldsymbol{\theta}')\boldsymbol{w} = \boldsymbol{e}_i$ is consistent for any element $\boldsymbol{e}_i$ in the standard basis of $\mathbb{R}^k$, which implies that $G^T(\boldsymbol{\theta}')$ is full rank. □

C.3 Optimal Protocol: Case of Analytic Functions

We use the results of the main text to generalize our optimal protocol to the case where both $\boldsymbol{f}(\boldsymbol{\theta})$ and $q(\boldsymbol{\theta})$ are analytic in the neighborhood of the true value $\boldsymbol{\theta}'$. Given a total time t, we consider a two-step protocol that extends the approach of Ref. [102]. In the first step, we spend time $t_1 = t^p$ with $p \in (1/2, 1)$ to obtain an initial estimate $\tilde{\boldsymbol{\theta}}$ of the true value $\boldsymbol{\theta}'$. We then linearize q about $\tilde{\boldsymbol{\theta}}$ to obtain

$$q(\boldsymbol{\theta}) \approx q(\tilde{\boldsymbol{\theta}}) + \nabla q(\tilde{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) =: \tilde{\boldsymbol{\alpha}} \cdot \boldsymbol{\theta} + K, \tag{C.10}$$

where $\tilde{\boldsymbol{\alpha}} = \nabla q(\tilde{\boldsymbol{\theta}})$ and K is a constant with respect to $\boldsymbol{\theta}$. We will show in Sec. C.5 that the error introduced by this approximation is negligible if $\tilde{\boldsymbol{\theta}}$ can be estimated with MSE $\mathcal{O}(1/t_1^2)$ in time t_1 (as can be done via phase estimation procedures like in Ref. [66]). After having obtained $\tilde{\boldsymbol{\theta}}$ in the first step, we can compute K.

In the second step, we estimate the remaining linear term $\tilde{\boldsymbol{\alpha}} \cdot \boldsymbol{\theta}$ in Eq. (C.10) in time $t_2 = t - t_1$. Define $G(\tilde{\boldsymbol{\theta}})$ as in Eq. (4.5) of the main text with $\boldsymbol{\theta}' \to \tilde{\boldsymbol{\theta}}$. Then following the procedure of the linear case protocol in the main text, we measure a linear function λ such that the corresponding estimate $\tilde{\lambda}$ is an asymptotically unbiased estimate of $\tilde{\boldsymbol{\alpha}} \cdot \boldsymbol{\theta}$. In particular, here we have $\lambda = \tilde{\boldsymbol{w}} \cdot (\boldsymbol{f}(\boldsymbol{\theta}) - \boldsymbol{C})$, where the constant vector $\boldsymbol{C}$ is chosen in such a way that $\boldsymbol{f}(\boldsymbol{\theta}) - \boldsymbol{C} = G(\tilde{\boldsymbol{\theta}})\boldsymbol{\theta} + \mathcal{O}(\boldsymbol{\Delta})$, with $\boldsymbol{\Delta} := \tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}'$ and $\tilde{\boldsymbol{w}}$ a vector of weights that we still need to

choose. With $\boldsymbol{C}$ defined in this way, we linearize $f_i(\boldsymbol{\theta})$ about $\tilde{\boldsymbol{\theta}}$ and obtain

$$\lambda \approx \tilde{\boldsymbol{w}} \cdot (G(\tilde{\boldsymbol{\theta}})\boldsymbol{\theta}) = (G(\tilde{\boldsymbol{\theta}})^T \tilde{\boldsymbol{w}}) \cdot \boldsymbol{\theta}. \tag{C.11}$$

Similar to the linear-case protocol in the main text, we ensure that $\tilde{\lambda}$ estimates q by choosing $\tilde{\boldsymbol{w}}$ to satisfy $G(\tilde{\boldsymbol{\theta}})^T \tilde{\boldsymbol{w}} = \tilde{\boldsymbol{\alpha}}$, which we show in Appendix C.4 to be a consistent system of equations. We then solve the corresponding protocol problem to obtain the optimal vector $\tilde{\boldsymbol{w}}$, given $G(\tilde{\boldsymbol{\theta}})$ and $\tilde{\boldsymbol{\alpha}}$.

Combining steps one and two yields an estimator for $q(\boldsymbol{\theta})$. In Appendix C.5, we show that the MSE for such a protocol is asymptotically equal to the linear case in Eq. (4.11). The crucial point is that the process of linearizing $\boldsymbol{f}$ and $q(\boldsymbol{\theta})$ about $\tilde{\boldsymbol{\theta}}$ introduces asymptotically negligible corrections compared to linearizing about the true value $\boldsymbol{\theta}'$. Consequently, asymptotically,

$$\mathcal{M} \sim \frac{\|\tilde{\boldsymbol{w}}\|_\infty}{t_2^2} \sim \frac{\|\boldsymbol{w}\|_\infty}{t^2}, \tag{C.12}$$

where we have used $t_2 \sim t$, and $\boldsymbol{w}$ is the optimal weight vector obtained from the protocol problem for $G = G(\boldsymbol{\theta}')$ and $\boldsymbol{\alpha} = \nabla q(\boldsymbol{\theta}')$. Referring to Eq. (4.5) and the preceding discussion in the main text, we recall that G and $\boldsymbol{\alpha}$ defined this way are precisely the appropriate input to the bound problem in order to obtain our ultimate MSE bound. As, asymptotically, our protocol for an analytic objective and field yields an MSE equivalent to the fully linear case, the same proofs as in the truly linear case hold, and the asymptotic bound obtained by solving the protocol problem in Eq. (C.12) is equivalent to the sharpest bound obtained by solving the corresponding bound problem – therefore the protocol is asymptotically optimal.

C.4 Proofs on Estimate Asymptotics

In this section, we prove that, in the two-step protocol of Appendix C.3, using the estimate $\tilde{\boldsymbol{\theta}}$ obtained from the first step of the protocol, as opposed to the unknown true value $\boldsymbol{\theta}'$, asymptotically yields negligible errors when compared to the determination of the weight vector $\boldsymbol{w}$ that is the solution of the protocol problem.

Recall that we use time $t_1 = t^p$ for $1/2 < p < 1$ on the first step of the protocol to obtain an estimate of each $\boldsymbol{\theta}_i$ with MSE $\mathcal{O}\left(\frac{1}{t_1^2}\right)$. We then spend time $t_2 = t - t_1$ estimating $q(\boldsymbol{\theta}')$ via a linearization of $q(\boldsymbol{\theta})$ about our estimate $q(\tilde{\boldsymbol{\theta}})$ with a weighted linear (in $\boldsymbol{\theta}$) protocol. We begin by assuming that our initial estimate $\tilde{\boldsymbol{\theta}}$ satisfies

$$\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}'\| = \|\boldsymbol{\Delta}\| \leq \delta \tag{C.13}$$

for some fixed positive real δ, where we defined $\boldsymbol{\Delta} := \tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}'$. From here on, norms without subscripts denote the Euclidean norm. This means we assume $\tilde{\boldsymbol{\theta}}$ lies within or on a ball of radius δ of the true value $\boldsymbol{\theta}'$ in the parameter space of $\boldsymbol{\theta}$'s. Recall that we also require that both $q(\boldsymbol{\theta})$ and $f_i(\boldsymbol{\theta})$ $\forall i$ are analytic within this ball for some δ. Crucially, asymptotically in time t, we may make δ an arbitrarily small fixed positive number. That is, as the total time $t \longrightarrow \infty$, the time spent obtaining our estimate $t_1 = t^p \longrightarrow \infty$, and therefore the MSE of our estimate $\tilde{\boldsymbol{\theta}}$ goes to zero.

We now prove the following theorem, which guarantees that, asymptotically, $G(\tilde{\boldsymbol{\theta}})$, as defined in Eq. (4.5) of the main text with $\boldsymbol{\theta}' \to \tilde{\boldsymbol{\theta}}$, has full rank.

Theorem C.4.1. *Given a $d \times k$ matrix*

$$G(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_d(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial f_d(\boldsymbol{\theta})}{\partial \theta_k} \end{pmatrix}$$

and an estimate $\tilde{\boldsymbol{\theta}}$, asymptotically for $t \longrightarrow \infty$, $G(\tilde{\boldsymbol{\theta}})$ has full rank.

Proof. From Corollary C.2.2.1, we know that all singular values of $G(\boldsymbol{\theta}')$ are nonzero, and thus the matrix has full rank. Let P be a perturbation matrix such that $G(\tilde{\boldsymbol{\theta}}) = G(\boldsymbol{\theta}') + P$. We expand $f_i(\tilde{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}'$ as

$$f_i(\tilde{\boldsymbol{\theta}}) = f_i(\boldsymbol{\theta}') + \nabla f_i(\boldsymbol{\theta}') \cdot \boldsymbol{\Delta} + \ldots \tag{C.14}$$

Thus,

$$\begin{aligned} G(\tilde{\boldsymbol{\theta}})_{in} = \frac{\partial f_i(\tilde{\boldsymbol{\theta}})}{\partial \theta_n} &= \frac{\partial f_i(\boldsymbol{\theta}')}{\partial \theta_n} + \mathcal{O}(\boldsymbol{\Delta}) \\ &= G(\boldsymbol{\theta}')_{in} + \mathcal{O}(\boldsymbol{\Delta}), \end{aligned} \tag{C.15}$$

which implies $P_{in} = \mathcal{O}(\boldsymbol{\Delta})$ and therefore, as $\dim(\boldsymbol{\Delta}) = k = \mathcal{O}(1)$, $\|P\| = \mathcal{O}(\boldsymbol{\Delta})$. It is a well-known result in matrix perturbation theory (see e.g. Ref. [485]) that, if $\|P\| < \sigma$, where σ is the minimum singular value of $G(\boldsymbol{\theta}')$, then $G(\tilde{\boldsymbol{\theta}})$ has the same rank as $G(\boldsymbol{\theta}')$, i.e. full rank. Since asymptotically we can make δ arbitrarily small in Eq. (C.13), we can also make $\|P\|$ arbitrarily small; therefore, since $\sigma > 0$, we are guaranteed to satisfy this condition asymptotically. Thus, asymptotically, all singular values of $G(\tilde{\boldsymbol{\theta}})$ are nonzero and the matrix has full rank. □

Now we consider the solutions to $G^T(\boldsymbol{\theta}')\boldsymbol{w} = \boldsymbol{\alpha}$ as compared to $G^T(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{\alpha}}$. We begin

by restating a useful result from Ref. [118], labeled there as Fundamental Theorem 2, with our notation.

Theorem C.4.2. *Given the $k \times d$ matrix G^T with rank k and the $k \times 1$ vector $\boldsymbol{\alpha}$, there exists a $k \times 1$ vector $\boldsymbol{v^0}$ such that*

$$\boldsymbol{\alpha} \cdot \boldsymbol{v^0} = \max_{\|G\boldsymbol{v}\|_1 \leq 1} \boldsymbol{\alpha} \cdot \boldsymbol{v} = \max_{\|G\boldsymbol{v}\|_1 = 1} \boldsymbol{\alpha} \cdot \boldsymbol{v},$$

and at least $k-1$ components of $G\boldsymbol{v^0}$ are zero, that is:

$$\boldsymbol{g_i} \cdot \boldsymbol{v^0} = 0 \quad \text{for} \quad i \in [i_1, i_2, \cdots, i_{k-1}] \quad \text{with} \quad 1 \leq i_\ell \leq d,$$

where $\boldsymbol{g_i}$ denotes the i^{th} column of G^T. Furthermore, the set of vectors

$$\left[\boldsymbol{g_{i_1}}, \boldsymbol{g_{i_2}}, \cdots, \boldsymbol{g_{i_{k-1}}}\right]$$

is linearly independent.

This theorem is about the protocol problem. That is, $\boldsymbol{v}^0$ is the solution vector to the protocol problem. Furthermore, we recall that, by strong duality,

$$u'' = \boldsymbol{\alpha} \cdot \boldsymbol{v^0} = \min_{G^T(\boldsymbol{\theta}')\boldsymbol{w} = \boldsymbol{\alpha}} \|\boldsymbol{w}\|_\infty = u'. \tag{C.16}$$

We now compare how the solution of the protocol problem is perturbed by considering $G^T(\tilde{\boldsymbol{\theta}})$ and $\tilde{\boldsymbol{\alpha}}$ as opposed to $G^T(\boldsymbol{\theta}')$ and $\boldsymbol{\alpha}$. To that end, we prove the following theorem.

Theorem C.4.3. *Consider the linear systems of equations $G^T(\boldsymbol{\theta}')\boldsymbol{w} = \boldsymbol{\alpha}$ and $G^T(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{\alpha}}$, where we recall $\boldsymbol{\alpha} = \nabla q(\boldsymbol{\theta}')$ and $\tilde{\boldsymbol{\alpha}} = \nabla q(\tilde{\boldsymbol{\theta}})$. Then*

$$\|\tilde{\boldsymbol{w}}\|_\infty = \|\boldsymbol{w}\|_\infty + \mathcal{O}(\boldsymbol{\Delta}),$$

where $\boldsymbol{w}$ is the solution to the protocol problem with $G(\boldsymbol{\theta}')$ and $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{w}}$ is the solution to the protocol problem with the approximate $G(\tilde{\boldsymbol{\theta}})$ and $\tilde{\boldsymbol{\alpha}}$.

Proof. As in Theorem C.4.1, write $G(\tilde{\boldsymbol{\theta}}) = G(\boldsymbol{\theta}') + P$ with perturbation matrix P. Similarly, we define a perturbation vector $\boldsymbol{p}$ such that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \boldsymbol{p}$. We expand $q(\tilde{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}'$ as

$$q(\tilde{\boldsymbol{\theta}}) = q(\boldsymbol{\theta}') + \nabla q(\boldsymbol{\theta}') \cdot \boldsymbol{\Delta} + \dots \tag{C.17}$$

We then have

$$\tilde{\boldsymbol{\alpha}} = \nabla q(\tilde{\boldsymbol{\theta}}) = \nabla q(\boldsymbol{\theta}') + \mathcal{O}(\boldsymbol{\Delta}) = \boldsymbol{\alpha} + \mathcal{O}(\boldsymbol{\Delta}). \tag{C.18}$$

Therefore we have $\boldsymbol{p} = \mathcal{O}(\boldsymbol{\Delta})$ and (from Theorem C.4.1) $P = \mathcal{O}(\boldsymbol{\Delta})$.

Now, due to strong duality, we conclude that the solution of the dual protocol problem is equal to that of the protocol problem. That is,

$$\begin{aligned} \|\boldsymbol{w}\|_\infty &= \boldsymbol{\alpha} \cdot \boldsymbol{v}^{\mathbf{0}}, \\ \|\tilde{\boldsymbol{w}}\|_\infty &= \tilde{\boldsymbol{\alpha}} \cdot (\boldsymbol{v}^{\mathbf{0}} + \boldsymbol{\epsilon}^{\mathbf{0}}), \end{aligned} \tag{C.19}$$

for the unperturbed and perturbed problems, respectively. We have introduced $\boldsymbol{\epsilon}^0$ as the pertur-

bation in $\boldsymbol{v^0}$ in the solution to the dual protocol problem in response to the perturbations P and $\boldsymbol{p}$.

Consider the solution to the unperturbed problem. We introduce the $k \times (k-1)$ matrix $M^T = \left(\boldsymbol{g_{i_1}}, \boldsymbol{g_{i_2}}, \cdots, \boldsymbol{g_{i_{k-1}}}\right)$ with columns as defined in Theorem C.4.2. By the same theorem, the solution vector satisfies

$$M\boldsymbol{v^0} = \boldsymbol{0}, \tag{C.20}$$

where we note that M is a submatrix of $G(\boldsymbol{\theta'})$. Hence the solution vector must be jointly orthogonal to all columns of M^T. Via a k-dimensional generalization of the determinant formula for a cross product (obtained from Cramer's rule), we have the unnormalized solution vector

$$\boldsymbol{v} = \det\left[\begin{pmatrix} \boldsymbol{e} \\ \\ M \end{pmatrix}\right], \tag{C.21}$$

where $\boldsymbol{e} = (\boldsymbol{e_1}, \boldsymbol{e_2}, \cdots \boldsymbol{e_k})^T$ represents a vector of the standard-basis vectors. Note that $\boldsymbol{v}$ is unique up to scalar multiplication. We then have component-wise

$$\boldsymbol{v}_n = (-1)^{n+1} \det(M_n)\boldsymbol{e_n} \tag{C.22}$$

where we naturally define M_n as the unique $(k-1) \times (k-1)$ submatrix of M that results from eliminating the first row and n^{th} column of the matrix in Eq. (C.21). We normalize the solution vector to the protocol problem so that it satisfies the condition in Theorem C.4.2 as

$$\boldsymbol{v^0} = \frac{\text{sgn}(\boldsymbol{\alpha} \cdot \boldsymbol{v})}{\|G(\boldsymbol{\theta'})\boldsymbol{v}\|_1}\boldsymbol{v} = \mathcal{N}\boldsymbol{v}, \tag{C.23}$$

where we have implicitly defined the normalization factor $\mathcal{N}$.

We now consider the perturbed problem. Introduce two $k \times (k-1)$ matrices: $\overline{M}^T = \left(\boldsymbol{g}_{j_1}, \boldsymbol{g}_{j_2}, \cdots, \boldsymbol{g}_{j_{k-1}}\right)$, a submatrix of $G^T(\boldsymbol{\theta}')$, and $\overline{Q}^T$, the corresponding submatrix of P. That is, $\overline{M} + \overline{Q}$ is a submatrix of $G(\tilde{\boldsymbol{\theta}})$. We pick indices $j_\ell \in [1, d]$ in accordance with Thm. C.4.2 such that the solution vector $\boldsymbol{v^0} + \boldsymbol{\epsilon^0}$ of the perturbed protocol problem satisfies

$$(\overline{M} + \overline{Q})(\boldsymbol{v^0} + \boldsymbol{\epsilon^0}) = \mathbf{0}. \tag{C.24}$$

Similar to the unperturbed case, Eq. (C.24) has an unnormalized solution vector given by the determinant

$$\boldsymbol{v} + \boldsymbol{\epsilon} = \det\left[\begin{pmatrix} \boldsymbol{e} \\ \\ \overline{M} + \overline{Q} \end{pmatrix}\right], \tag{C.25}$$

which component-wise reads

$$\boldsymbol{v}_n + \boldsymbol{\epsilon}_n = (-1)^{n+1} \det(\overline{M}_n + \overline{Q}_n)\boldsymbol{e_n}. \tag{C.26}$$

As in the unperturbed case, $\overline{M}_n + \overline{Q}_n$ is the submatrix of $\overline{M} + \overline{Q}$ corresponding to eliminating the first row and n^{th} column of the matrix inside the determinant of Eq. (C.25). The corresponding normalized solution vector to the perturbed protocol problem is

$$\boldsymbol{v^0} + \boldsymbol{\epsilon^0} = \frac{\operatorname{sgn}(\tilde{\boldsymbol{\alpha}} \cdot (\boldsymbol{v} + \boldsymbol{\epsilon}))}{\|(G(\boldsymbol{\theta}') + P)(\boldsymbol{v} + \boldsymbol{\epsilon})\|_1}(\boldsymbol{v} + \boldsymbol{\epsilon}) = \overline{\mathcal{N}}(\boldsymbol{v} + \boldsymbol{\epsilon}), \tag{C.27}$$

where we have implicitly defined the normalization factor $\overline{\mathcal{N}}$.

We now consider several cases for how the unperturbed and perturbed solution vectors are

related: (1) $M = \overline{M}$, (2) $M \neq \overline{M}$ and $\overline{M}$ has full rank, (3) $M \neq \overline{M}$ and $\overline{M}$ does *not* have full rank. Intuitively, case (1) corresponds to when the solution vectors of the unperturbed and perturbed protocol problems are orthogonal (via Thm. C.4.2) to the same set of columns of $G^T(\boldsymbol{\theta}')$ and $G^T(\tilde{\boldsymbol{\theta}})$, respectively. That is, the solution vector in the perturbed case is merely a perturbed version of the solution vector in the unperturbed case. In cases (2) and (3), the set of columns of G^T to which the unperturbed and perturbed solution vectors are orthogonal *do not* have the same indices. Intuitively, this means that the perturbed problem solution vector is not simply the perturbed version of the solution vector in the unperturbed problem. These cases divide into two options. In case (2), this set of columns of $G^T(\tilde{\boldsymbol{\theta}})$ in the perturbed case (given by the rows of $\overline{M} + \overline{Q}$) corresponds to a set of unperturbed columns of $G^T(\boldsymbol{\theta}')$ (given by the rows of $\overline{M}$) that are independent – i.e. $\overline{M}$ has full rank. In particular, this means that the unperturbed version of the perturbed solution vector is a *candidate* solution to the unperturbed problem. By candidate solution we refer to the fact that any choice of independent columns of $G(\boldsymbol{\theta}')$ could correspond to a *possible* solution vector according to Thm. C.4.2, in the sense that any such choice picks out a candidate, unnormalized solution vector via Eq. (C.21). There are at most $\binom{d}{k-1}$ such candidate solutions based on picking the set of $k-1$ columns that define a possible M. In case (3), $\overline{M}$ does not have full rank – i.e. the corresponding unperturbed columns are not independent and the unperturbed version of the perturbed solution vector is *not* a candidate solution vector to the unperturbed problem.

We now examine the cases one by one in detail and find that we can rule out cases (2) and (3). Starting with case (1), we may drop the bar on $\overline{M}$ as $M = \overline{M}$. We then use a bound on

determinants of perturbed matrices from Ref. [486] (see Remark 2.9 therein) and obtain

$$|\det(M_n) - \det(M_n + \overline{Q}_n)| \leq s_{k-2}\|\overline{Q}_n\|, \tag{C.28}$$

where $s_{k-2} \leq (k-1)\sigma_1 \cdots \sigma_{k-2}$ is the $(k-2)^{\text{nd}}$ elementary symmetric function in the singular values $\sigma_1 \geq \cdots \geq \sigma_{k-1}$ of M_i [486]. Importantly, $\sigma_1 = \|M_n\| \leq \sqrt{k-1}\|M_n\|_\infty = \mathcal{O}(1)$ and $\|\overline{Q}_n\| = \mathcal{O}(\boldsymbol{\Delta})$ as $\|\overline{Q}_n\| \leq \sqrt{\sum_{ab} (\overline{Q}_n)_{ab}^2}$ and all elements of $\overline{Q}_n$ are of size $\mathcal{O}(\boldsymbol{\Delta})$. Therefore,

$$|\det(M_n) - \det(M_n + \overline{Q}_n)| = \mathcal{O}(\boldsymbol{\Delta}), \tag{C.29}$$

which directly implies that $\boldsymbol{\epsilon} = \mathcal{O}(\boldsymbol{\Delta})$. Having established that $\boldsymbol{\epsilon} = \mathcal{O}(\boldsymbol{\Delta})$, we now consider the normalization factors $\mathcal{N}$ and $\overline{\mathcal{N}}$. Recall that

$$|\overline{\mathcal{N}}| = \frac{1}{\|(G(\boldsymbol{\theta}') + P)(\boldsymbol{v} + \boldsymbol{\epsilon})\|_1}. \tag{C.30}$$

By the triangle inequality,

$$\begin{aligned} \|(G(\boldsymbol{\theta}')\boldsymbol{v}\|_1 - \|(G(\boldsymbol{\theta}')\boldsymbol{\epsilon} + P(\boldsymbol{v} + \boldsymbol{\epsilon})\|_1 &\leq \|(G(\boldsymbol{\theta}') + P)(\boldsymbol{v} + \boldsymbol{\epsilon})\|_1 \\ &\leq \|(G(\boldsymbol{\theta}')\boldsymbol{v}\|_1 + \|(G(\boldsymbol{\theta}')\boldsymbol{\epsilon} + P(\boldsymbol{v} + \boldsymbol{\epsilon})\|_1. \end{aligned} \tag{C.31}$$

Then, using a binomial expansion yields

$$|\mathcal{N}| - \mathcal{O}(\boldsymbol{\Delta}) \leq |\overline{\mathcal{N}}| \leq |\mathcal{N}| + \mathcal{O}(\boldsymbol{\Delta}), \tag{C.32}$$

so $\overline{\mathcal{N}} = \mathcal{N} + \mathcal{O}(\boldsymbol{\Delta})$, where we note that $\mathcal{N} = \mathcal{O}(1)$. Therefore, for case (1), the perturbed solution vector is $\boldsymbol{v^0} + \mathcal{O}(\boldsymbol{\Delta})$, and $\boldsymbol{\epsilon^0} = \mathcal{O}(\boldsymbol{\Delta})$. Equation Eq. (C.19) then yields

$$\begin{aligned} \|\tilde{\boldsymbol{w}}\|_\infty &= (\boldsymbol{\alpha} + \boldsymbol{p}) \cdot (\boldsymbol{v^0} + \boldsymbol{\epsilon^0}) \\ &= \boldsymbol{\alpha} \cdot \boldsymbol{v^0} + \mathcal{O}(\boldsymbol{\Delta}) \\ &= \|\boldsymbol{w}\|_\infty + \mathcal{O}(\boldsymbol{\Delta}). \end{aligned} \tag{C.33}$$

We now demonstrate that neither case (2) nor case (3) can arise. Starting with case (2), we recall from the discussion above that $\overline{M}$ corresponds to a candidate solution for the original unperturbed problem. The corresponding candidate (unnormalized) solution vector to the unperturbed problem, $\overline{\boldsymbol{v}}$, can be found by an equation analogous to Eq. (C.21). By the same argument as in case (1), when we perturb $\overline{M}$ by $\overline{Q}$ to obtain the perturbed problem, the unperturbed candidate solution vector $\overline{\boldsymbol{v}}$ may only be perturbed by $\mathcal{O}(\boldsymbol{\Delta})$ and similarly the corresponding candidate solution value may also only be perturbed by $\mathcal{O}(\boldsymbol{\Delta})$. Let the difference between the candidate solution corresponding to $\overline{M}$ and the true solution to the unperturbed problem be given by r. As asymptotically we can make $\boldsymbol{\Delta}$ arbitrarily small we may always make $\|\boldsymbol{\Delta}\| \ll r$. This contradicts the fact that the solution vector $\overline{M} + \overline{Q}$ is the solution to the perturbed problem as an approach like case (1) is guaranteed to offer a better solution than case (2) for sufficiently small perturbations. Therefore case (2) cannot arise.

Similarly, we can show that case (3) cannot arise. In this case, $\overline{M}$ is rank-deficient and consequently does not correspond to a candidate solution to the original unperturbed problem. Also, due to its rank deficiency, we know there exists a linear combination of rows via $\mathcal{O}(1)$ coefficients such that we may transform $\overline{M}$ using row operations into a form that it has a row of

all zeros. Call this transformation T. We then have that $T(\overline{M} + \overline{Q})$ has a row with all elements of size $\mathcal{O}(\boldsymbol{\Delta})$. Consider $\det(\overline{M}_n + \overline{Q}_n) = \det(T(\overline{M}_n + \overline{Q}_n))$ as in Eq. (C.26), where we eliminate the n^{th} column of $T(\overline{M} + \overline{Q})$ to obtain $T(\overline{M}_n + \overline{Q}_n)$. Eliminating this column does not change the fact that $T(\overline{M}_n + \overline{Q}_n)$ has a row with all elements of size $\mathcal{O}(\boldsymbol{\Delta})$. Consequently, $\det(\overline{M}_n + \overline{Q}_n) = \det(T(\overline{M}_n + \overline{Q}_n)) = \mathcal{O}(\boldsymbol{\Delta})$ and, therefore, all components of the *unnormalized* perturbed problem solution vector must be $\mathcal{O}(\boldsymbol{\Delta})$. Let this unnormalized solution vector be $\overline{\boldsymbol{v}}$. From before, we have $\overline{\mathcal{N}} = \mathcal{N} + \mathcal{O}(\boldsymbol{\Delta})$ and $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \mathcal{O}(\boldsymbol{\Delta})$, so the solution corresponding to $\overline{\boldsymbol{v}}$ is

$$\overline{\mathcal{N}}(\tilde{\boldsymbol{\alpha}} \cdot \overline{\boldsymbol{v}}) = \mathcal{N}(\boldsymbol{\alpha} \cdot \overline{\boldsymbol{v}}) + \mathcal{O}(\boldsymbol{\Delta}) = \mathcal{O}(\boldsymbol{\Delta}). \tag{C.34}$$

In the second equality we used that $\mathcal{N} = \mathcal{O}(1)$ and $\overline{\boldsymbol{v}} = \mathcal{O}(\boldsymbol{\Delta})$. Furthermore, let $\boldsymbol{v}$ be the unnormalized solution vector for the unperturbed problem, then

$$\mathcal{N}(\boldsymbol{\alpha} \cdot \overline{\boldsymbol{v}}) \leq \mathcal{N}(\boldsymbol{\alpha} \cdot \boldsymbol{v}), \tag{C.35}$$

which implies that, asymptotically,

$$\overline{\mathcal{N}}(\tilde{\boldsymbol{\alpha}} \cdot \overline{\boldsymbol{v}}) \lesssim \mathcal{N}(\boldsymbol{\alpha} \cdot \boldsymbol{v}). \tag{C.36}$$

The right-hand side of this inequality is $\mathcal{O}(1)$, whereas the left-hand side is $\mathcal{O}(\boldsymbol{\Delta})$. Therefore, asymptotically, the perturbed solution is no longer within $\mathcal{O}(\boldsymbol{\Delta})$ of the solution to the unperturbed problem. As a result, asymptotically, an approach like case (1) is guaranteed to result in a better candidate solution vector than a candidate arising from case (3). Thus case (3) cannot lead to a solution, concluding the proof. □

Theorem C.4.3 has immediate consequences that we use in evaluating our protocol performance and in the proofs in Appendix C.5. It also implies the following useful corollary.

Corollary C.4.3.1. $\|\tilde{\boldsymbol{w}}\|$ *can, asymptotically, be upper bounded by a constant.*

Proof. $\|\boldsymbol{w}\| = \mathcal{O}(1)$, and, asymptotically, δ—which bounds $\|\boldsymbol{\Delta}\|$ (see Eq. (C.13))—can be made arbitrarily small. This directly implies the result. □

C.5 Proof of Protocol Optimality

Using the results of the previous section, we rigorously demonstrate that the two-step protocol described in Appendix C.3 is optimal. In particular, we focus on the effects of using the estimate $\tilde{\boldsymbol{\theta}}$ from step 1 of the protocol, as opposed to the true $\boldsymbol{\theta}'$, for step 2 of the protocol and demonstrate that, asymptotically, the errors introduced are negligible.

We begin by sketching how the two-step protocol saturates the MSE bound in Eq. (4.7) and, therefore, yields an optimal estimate of the function $q(\boldsymbol{\theta}')$. We then fill in the details to rigorously obtain the result. The MSE of the full protocol is given by

$$\begin{aligned}\mathcal{M} &= \mathbb{E}\left[(\tilde{q} - q(\boldsymbol{\theta}'))^2\right] \\ &= \mathcal{M}_1 + \mathcal{M}_2,\end{aligned} \tag{C.37}$$

with

$$\mathcal{M}_1 = \mathbb{E}_{\tilde{\boldsymbol{\theta}}}[\mathrm{Var}_{\tilde{\lambda}}[\tilde{\lambda}]], \tag{C.38}$$

$$\mathcal{M}_2 = \mathbb{E}_{\tilde{\boldsymbol{\theta}}}\left[(q(\tilde{\boldsymbol{\theta}}) + \lambda - \tilde{\boldsymbol{\alpha}} \cdot \tilde{\boldsymbol{\theta}} - q(\boldsymbol{\theta}'))^2\right]. \tag{C.39}$$

The variance of the estimation of λ, for fixed $\tilde{\boldsymbol{\theta}}$, is $\|\tilde{\boldsymbol{w}}\|_\infty^2/t_2^2$ [8]. We then show that

$$\mathcal{M}_1 = \frac{\|\boldsymbol{w}\|_\infty^2}{t_2^2}\left(1 + \frac{\mathcal{C}}{t_1} + \mathcal{O}\left(t_1^{-2}\right)\right), \tag{C.40}$$

with some constant $\mathcal{C}$. Given $t_1 = t^p$ with $p \in (1/2, 1)$, we conclude that $\mathcal{M}_1$ is asymptotically given by $\frac{\|\boldsymbol{w}\|_\infty}{t_2^2} \sim \frac{\|\boldsymbol{w}\|_\infty}{t^2}$. On the other hand, we also show that $\mathcal{M}_2$ is of order $\mathcal{O}(t_1^{-4})$ and is therefore asymptotically negligible. Thus, we asymptotically have

$$\mathcal{M} \sim \frac{\|\boldsymbol{w}\|_\infty}{t^2}, \tag{C.41}$$

where we recall that $\boldsymbol{w}$ is the optimal weight vector obtained from the protocol problem for $G = G(\boldsymbol{\theta}'), \boldsymbol{\alpha} = \nabla q(\boldsymbol{\theta}')$.

The point of the asymptotics is that the problem we actually solve in practice with our estimate $\tilde{\boldsymbol{\theta}}$ introduces asymptotically negligible corrections. Via the same proofs as in the linear case, we know that the protocol problem used to obtain $\boldsymbol{w}$ gives a solution equivalent to the corresponding bound problem, and therefore the protocol is asymptotically optimal. We now fill in the details and derive the asymptotic behavior of $\mathcal{M}_1$ and $\mathcal{M}_2$.

Derivation of $\mathcal{M}_1$. An immediate consequence of Theorem C.4.3 is

$$\mathbf{E}_{\tilde{\boldsymbol{\theta}}}[\mathrm{Var}_{\tilde{\lambda}}[\tilde{\lambda}]] = \frac{\mathbf{E}[\|\tilde{\boldsymbol{w}}\|_\infty^2]}{t_2^2} = \frac{\|\boldsymbol{w}\|_\infty^2}{t_2^2}\mathbf{E}\left[(1 + \mathcal{B}\boldsymbol{\Delta})^2\right] \tag{C.42}$$

for some constant $\mathcal{B}$. Note that we can expand

$$\mathbf{E}[(1+\mathcal{B}\|\boldsymbol{\Delta}\|)^2] = 1 + \mathcal{B}^2\mathbf{E}[\|\boldsymbol{\Delta}\|^2] + 2\mathcal{B}\mathbf{E}[\|\boldsymbol{\Delta}\|]. \tag{C.43}$$

Since $\mathbf{E}[\|\boldsymbol{\Delta}\|^2]$ is the sum of the squared deviations of the individual θ_i, it is $\mathcal{O}\left(\frac{1}{t_1^2}\right)$. Similarly, $\mathbf{E}[\|\boldsymbol{\Delta}\|]$ is $\mathcal{O}\left(\frac{1}{t_1}\right)$. As a result, we can expand Eq. (C.42) to

$$\mathcal{M}_1 = \mathbf{E}_{\tilde{\boldsymbol{\theta}}}[\mathrm{Var}_{\tilde{\lambda}}[\tilde{\lambda}]] = \frac{\|\boldsymbol{w}\|_\infty^2}{t_2^2}\left(1 + \frac{\mathcal{C}}{t_1} + \mathcal{O}\left(t_1^{-2}\right)\right) \tag{C.44}$$

for some constant $\mathcal{C}$, in agreement with Eq. (C.40).

Derivation of $\mathcal{M}_2$. We begin with

$$\mathcal{M}_2 = \mathbf{E}_{\tilde{\boldsymbol{\theta}}}\left[\left(q(\tilde{\boldsymbol{\theta}}) + \lambda - \tilde{\boldsymbol{\alpha}}\cdot\tilde{\boldsymbol{\theta}} - q(\boldsymbol{\theta})\right)^2\right]. \tag{C.45}$$

We define the vector $\boldsymbol{C}(\tilde{\boldsymbol{\theta}})$ to store constants which simplify our derivation through $C_i(\tilde{\boldsymbol{\theta}}) = f_i(\tilde{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\theta}}\cdot\nabla f_i(\tilde{\boldsymbol{\theta}})$. We then define

$$\lambda(\boldsymbol{f}) = \tilde{\boldsymbol{w}}\cdot(\boldsymbol{f} - \boldsymbol{C}(\tilde{\boldsymbol{\theta}})) \tag{C.46}$$

as the linear function we measure in step 2 of our protocol. Similar to Eq. (C.17), we expand $q(\boldsymbol{\theta})$ about $\tilde{\boldsymbol{\theta}}$:

$$q(\boldsymbol{\theta}) = q(\tilde{\boldsymbol{\theta}}) - \nabla q(\tilde{\boldsymbol{\theta}})\cdot\boldsymbol{\Delta} + \frac{T_2}{2} + \cdots, \tag{C.47}$$

where $T_n = \mathcal{O}(\Delta^n)$. Inserting this expansion into Eq. (C.45), where we use the definition of $\tilde{\boldsymbol{\alpha}}$, we find that $\mathcal{M}_2$ is equivalent to

$$\mathbf{E}_{\tilde{\boldsymbol{\theta}}}\left[\left(\lambda - \tilde{\boldsymbol{\alpha}}\cdot\tilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\alpha}}\cdot\boldsymbol{\Delta} - \frac{T_2}{2} + \frac{T_3}{3} - \cdots\right)^2\right] = \mathbf{E}_{\tilde{\boldsymbol{\theta}}}\left[\left(\underbrace{\lambda - \tilde{\boldsymbol{\alpha}}\cdot\boldsymbol{\theta}'}_{(*)} - \frac{T_2}{2} + \frac{T_3}{3} - \cdots\right)^2\right]. \tag{C.48}$$

To simplify the term labeled $(*)$, we insert Eq. (C.46) and use $\tilde{\boldsymbol{\alpha}}\cdot\boldsymbol{\theta}' = G^T(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{w}}\cdot\boldsymbol{\theta}' = \tilde{\boldsymbol{w}}\cdot G(\tilde{\boldsymbol{\theta}})\boldsymbol{\theta}'$. This yields

$$\lambda - \tilde{\boldsymbol{\alpha}}\cdot\boldsymbol{\theta} = \tilde{\boldsymbol{w}}\cdot\left(\boldsymbol{f} - \boldsymbol{C}(\tilde{\boldsymbol{\theta}}) - G(\tilde{\boldsymbol{\theta}})\boldsymbol{\theta}'\right). \tag{C.49}$$

Now consider the quantity $\boldsymbol{f} - C(\tilde{\boldsymbol{\theta}}) - G(\tilde{\boldsymbol{\theta}})\boldsymbol{\theta}'$. For the i^{th} component, we have

$$\boldsymbol{f}_i - C_i(\tilde{\boldsymbol{\theta}}) - [G(\tilde{\boldsymbol{\theta}})\boldsymbol{\theta}']_i = (f_i - f_i(\tilde{\boldsymbol{\theta}})) + \nabla f_i(\tilde{\boldsymbol{\theta}})\cdot\boldsymbol{\Delta}, \tag{C.50}$$

where we employed the definitions of $\boldsymbol{C}(\tilde{\boldsymbol{\theta}})$ and $G(\tilde{\boldsymbol{\theta}})$. Since $f_i(\boldsymbol{\theta})$ is an analytic function in a δ-ball around $\boldsymbol{\theta}'$, we can expanded it about $\boldsymbol{\theta}'$ as in Eq. (C.14). Substituting this expansion into Eq. (C.50) we arrive at

$$Eq.\ (\text{C.50}) = \left(\nabla f_i(\tilde{\boldsymbol{\theta}}) - \nabla f_i(\boldsymbol{\theta}')\right)\cdot\boldsymbol{\Delta} - \frac{S_2}{2} + \mathcal{O}(\boldsymbol{\Delta}^3) \tag{C.51}$$

with $S_n = \mathcal{O}(\Delta^n)$. From Eq. (C.14), we conclude that

$$\left(\nabla f_i(\tilde{\boldsymbol{\theta}}) - \nabla f_i(\boldsymbol{\theta}')\right)\cdot\boldsymbol{\Delta} = S_2 + \mathcal{O}(\boldsymbol{\Delta}^3) = \mathcal{O}(\boldsymbol{\Delta}^2). \tag{C.52}$$

Therefore, Eq. (C.50) is of order $\mathcal{O}(\boldsymbol{\Delta}^2)$. Furthermore, Corollary C.4.3.1 implies that, asymptotically, $\|\tilde{\boldsymbol{w}}\|_\infty$ can be upper bounded by a constant, i.e. the magnitude of each element of $\tilde{\boldsymbol{w}}$ is upper bounded by a constant. Combining these facts, we find that Eq. (C.49) is of order $\mathcal{O}(\boldsymbol{\Delta}^2)$. Together with Eq. (C.48), this, in turn, implies that $\mathcal{M}_2$ is $\mathcal{O}(\boldsymbol{\Delta}^4) = \mathcal{O}(t_1^{-4})$.

C.6 Review of the Protocol by Eldredge *et. al.*

In this section, we briefly summarize one of the protocols for optimal estimation of a linear combination of parameters from Eldredge *et. al.* [8], because it is as a subroutine in our protocols to obtain an estimate of Eq. (4.8) and Eq. (C.11) in the main text with variances given by Eq. (4.11) and Eq. (C.12), respectively. We seek to measure a linear combination (up to a constant shift) of the form

$$\lambda = \boldsymbol{w} \cdot \boldsymbol{f}. \tag{C.53}$$

Several specific protocols to obtain the optimal MSE estimate of such a linear combination are given in Ref. [8]. Here we present the first and simplest such protocol.

In this protocol, we suppose to have access to a time-dependent control over our evolution. We begin with a d-qubit GHZ input state of the quantum sensors given by

$$|\psi_0\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes d} + |1\rangle^{\otimes d} \right). \tag{C.54}$$

Under evolution by $\hat{\sigma}^z$ as in Eq. (4.1) in the main text, each qubit sensor accumulates a relative phase between the $|0\rangle$ and $|1\rangle$ states. We perform a partial time evolution so that each qubit sensor is evolved for a time proportional to the corresponding weight w_i on that sensor. We realize this

by applying $\hat{\sigma}_i^x$ to the i^{th} qubit at time $t_i = t(1+w_i)/2$. This results in an effective evolution of our state by the unitary

$$\hat{U}(t) = e^{-i\frac{t}{2}\sum_{i=1}^{d} w_i f_i(\boldsymbol{\theta})\hat{\sigma}_i^z}. \tag{C.55}$$

We note that this scheme assumes that $w_i \in [-1,1]$ and that the largest $|w_i|$ is equal to 1. We can always achieve this by rescaling the vector $\boldsymbol{w}$. Under this unitary evolution, the final state of the qubits is

$$|\psi_f\rangle = \frac{1}{\sqrt{2}}\left(e^{-it\lambda/2}|0\rangle^{\otimes d} + e^{it\lambda/2}|1\rangle^{\otimes d}\right). \tag{C.56}$$

We then make a measurement of the overall parity of the state using $\hat{P} = \bigotimes_{i=1}^{d}\hat{\sigma}_i^x$. Note that this measurement can be performed locally at each site. Furthermore, measurement of the expectation value $\langle\hat{P}\rangle(t)$ allows for estimation of λ with the optimal accuracy given by Eq. (4.11) [487].

Appendix D: Technical Details of the Results Reported in Chapter 5

D.1 Derivation of Signed Sensor Symmetric Bound

In this appendix, we demonstrate that the explicit calculation of the inverse of the quantum Fisher information in Ref. [53] for sensor symmetric states can be extended to the signed sensor symmetric states of Eq. (5.19). The calculation largely follows that in that reference.

Begin by defining the symmetric matrix $\Omega = \boldsymbol{\omega}\boldsymbol{\omega}^T$ for $\boldsymbol{\omega}$ a vector with all elements ± 1, as

defined in the main text. For example,

$$\Omega = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}. \tag{D.1}$$

Now, given an orthonormal basis $\{\hat{e}_i\}_{i\in[1,d]}$ for the real space where our vectors of coefficients $\{\boldsymbol{\alpha}_i\}$ are defined, we can write, for pure signed, sensor symmetric states,

$$\begin{aligned} \mathcal{F}_Q(\boldsymbol{\theta}) &= \sum_{i,j=1}^{d} t^2 \left(\langle \sigma_i^z \sigma_j^z \rangle - \langle \sigma_i^z \rangle \langle \sigma_j^z \rangle \right) \hat{e}_i \hat{e}_j^T \\ &= 4 \left(v \sum_{i=1}^{d} \hat{e}_i \hat{e}_i^T + c \sum_{i \neq j} \Omega_{ij} \hat{e}_i \hat{e}_j^T \right) \\ &= 4 \left[(v-c) I + c\Omega \right] = 4v \left[(1-\mathcal{J}) I + \mathcal{J}\Omega \right], \end{aligned} \tag{D.2}$$

where $\mathcal{F}_Q(\boldsymbol{\theta})$ is the quantum Fisher information with respect to the parameters $\boldsymbol{\theta}$ and v and c are defined as in Eqs. (5.16) and (5.19) of the main text. We note this is equivalent to Eq. (22) of Ref. [53] but with $\mathbf{1}\mathbf{1}^T \to \Omega$, where $\mathbf{1}$ is the vector of d $(+1)$s. To invert $\mathcal{F}_Q(\boldsymbol{\theta})$ and evaluate the quantum Cramér-Rao bound, we need the Fisher information matrix to be positive definite—i.e. we require its eigenvalues to be strictly positive. The characteristic equation for the eigenvalues λ of $\mathcal{F}_Q(\boldsymbol{\theta})$ is

$$\det \left[4v \left((1 - \mathcal{J} - \lambda/4v) I + \mathcal{J}\Omega \right) \right] = 0. \tag{D.3}$$

We then use the determinant identity [210]: $\det(X + \boldsymbol{x}\boldsymbol{y}^T) = (1 + \boldsymbol{x}^T X^{-1} \boldsymbol{y}) \det(X)$, with

$X = [4v(1-\mathcal{J}) - \lambda]I$, $\boldsymbol{x} = 4v\mathcal{J}\boldsymbol{\omega}$, and $\boldsymbol{y} = \boldsymbol{\omega}$. With a bit of algebra, we obtain

$$[4v(1+(d-1)\mathcal{J}) - \lambda]\,[4v(1-\mathcal{J}) - \lambda]^{d-1} = 0, \tag{D.4}$$

which is identical to Eq. (24) in [53]. Here, we used that $\omega^T\omega = d$. Therefore, the eigenvalues of $\mathcal{F}_Q(\boldsymbol{\theta})$ are $4v[1+(d-1)\mathcal{J}]$ with multiplicity one and $4v(1-\mathcal{J})$ with multiplicity $d-1$. If we insist that the eigenvalues are positive (so that $\mathcal{F}_Q(\boldsymbol{\theta})$ is invertible), we then have the condition on $\mathcal{J}$ that $\mathcal{J} \in \left(\frac{1}{1-d}, 1\right)$.

The inverse of $\mathcal{F}_Q(\boldsymbol{\theta})$ is given by

$$\mathcal{F}_Q^{-1}(\boldsymbol{\theta}) = \frac{[1+(d-1)\mathcal{J}]I - \mathcal{J}\Omega}{4v(1-\mathcal{J})[1+(d-1)\mathcal{J}]}. \tag{D.5}$$

We can verify this by computing

$$\begin{aligned}\mathcal{F}_Q^{-1}\mathcal{F}_Q &= \frac{[1+(d-1)\mathcal{J}]I - \mathcal{J}\Omega}{4v(1-\mathcal{J})[1+(d-1)\mathcal{J}]}(4v)\,[(1-\mathcal{J})I + \mathcal{J}\Omega] \\ &= I,\end{aligned} \tag{D.6}$$

where we used that $\Omega^2 = \boldsymbol{\omega}\boldsymbol{\omega}^T\boldsymbol{\omega}\boldsymbol{\omega}^T = \Omega d$. We then may evaluate the quantum Cramér-Rao bound

$$\mathcal{M} \geq \sum_{\ell=1}^{n} w_\ell (\mathcal{F}_Q^{-1}(\boldsymbol{g}))_{\ell\ell} = \sum_{\ell=1}^{n} w_\ell (A\mathcal{F}_Q^{-1}(\boldsymbol{\theta})A^T)_{\ell\ell}. \tag{D.7}$$

Plugging Eq. (D.5) into Eq. (D.7) and using $4v = t^2$ for our Hamiltonian from Eq. (5.17),

we obtain

$$\begin{aligned}\mathcal{M} &\geq \sum_{\ell=1}^{n} \frac{[1+(d-2)\mathcal{J}]w_\ell (AA^T)_{\ell\ell} - w_\ell \mathcal{J}[A(\Omega - I)A^T]_{\ell\ell}}{t^2(1-\mathcal{J})[1+(d-1)\mathcal{J}]} \\ &= \frac{[1+(d-2)\mathcal{J}]\mathcal{N} - \mathcal{N}\mathcal{J}\mathcal{G}(\boldsymbol{\omega})}{t^2(1-\mathcal{J})[1+(d-1)\mathcal{J}]} \\ &= \frac{\mathcal{N}}{t^2}\frac{1+(d-2-\mathcal{G}(\boldsymbol{\omega}))\mathcal{J}}{(1-\mathcal{J})[1+(d-1)\mathcal{J}]},\end{aligned} \tag{D.8}$$

where we introduced the generalized geometry parameter

$$\begin{aligned}\mathcal{G}(\boldsymbol{\omega}) &= \frac{1}{\mathcal{N}}\sum_{\ell=1}^{n} w_\ell [A(\Omega - I)A^T]_{\ell\ell} \\ &= \frac{1}{\mathcal{N}}\sum_{\ell=1}^{n} w_\ell \left[(A\boldsymbol{\omega}\boldsymbol{\omega}^T A^T)_{\ell\ell} - 1\right] \\ &= \frac{1}{\mathcal{N}}\sum_{\ell=1}^{n} w_\ell \left[(\boldsymbol{\alpha}_\ell \cdot \boldsymbol{\omega})^2 - 1\right] \\ &= \frac{1}{\mathcal{N}}\sum_{\ell=1}^{n} w_\ell \left(d\cos^2\phi_{\boldsymbol{\omega},\ell} - 1\right).\end{aligned} \tag{D.9}$$

Here $\mathcal{N}$ is the normalization factor as introduced in Eq. (5.14) in the main text and $\phi_{\boldsymbol{\omega},\ell}$ is the angle between the linear functions specified by $\boldsymbol{\alpha}_\ell$ and $\boldsymbol{\omega}$. Note that $\mathcal{G}(\boldsymbol{\omega}) \in [-1, d-1]$. As in Appendix C of Ref. [53], we can find the optimal $\mathcal{J}$ in Eq. (D.8), provided $\mathcal{J} \in \left(\frac{1}{1-d}, 1\right)$, to be

$$\begin{aligned}&\mathcal{J}_{\text{opt}}(\boldsymbol{\omega}) = \\ &\quad \frac{1}{\mathcal{G}(\boldsymbol{\omega}) + 2 - d}\left[1 - \sqrt{\frac{(\mathcal{G}(\boldsymbol{\omega})+1)(d-1-\mathcal{G}(\boldsymbol{\omega}))}{d-1}}\right].\end{aligned} \tag{D.10}$$

The ultimate best bound is found using

$$\mathcal{M}_{\text{ss}} = \min_{\boldsymbol{\omega}} \mathcal{M}_{\text{ss}}(\boldsymbol{\omega}). \tag{D.11}$$

D.2 Optimal Time Allocation

In this appendix, we consider the problem of optimal time division amongst the n measured functions. In particular, given some matrix C, we want to compute the optimal times $\{t_1, \ldots, t_n\}$ in

$$\begin{aligned} \mathcal{M}(C) &= \min_{\{t_1,\cdots,t_n\}} \left[\sum_{\ell=1}^{n} \sum_{m=1}^{n} w_m C_{m\ell}^2 \left(\frac{\mu'_\ell}{t_\ell} \right)^2 \right] \\ &= \min_{\{t_1,\cdots,t_n\}} \left[\sum_{\ell=1}^{n} t_\ell^{-2} \sum_{m=1}^{n} k_{m\ell} \right], \end{aligned} \tag{D.12}$$

subject to the constrain $\sum_{\ell=1}^{n} t_\ell = t$. In the second line, we define $k_{m\ell} = w_m C_{m\ell}^2 {\mu'_\ell}^2$. Introducing a Lagrange multiplier γ_0, we obtain the $n+1$ equations

$$\sum_{\ell=1}^{n} t_\ell = t, \tag{D.13a}$$

$$-\frac{2}{t_\ell^3} \sum_{m=1}^{n} k_{m\ell} = \gamma_0 \quad \forall\, \ell. \tag{D.13b}$$

Solving the latter equations for each t_ℓ and inserting the solution into the first equation yields

$$t = \left(-\frac{2}{\gamma_0} \right)^{\frac{1}{3}} \sum_{\ell=1}^{n} \left(\sum_{m=1}^{n} k_{m\ell} \right)^{\frac{1}{3}}. \tag{D.14}$$

We can rearrange this to find the Lagrange multiplier

$$\gamma_0 = -\frac{2}{t^3}\left[\sum_{\ell=1}^{n}\left(\sum_{m=1}^{n} k_{m\ell}\right)^{\frac{1}{3}}\right]^3. \tag{D.15}$$

Together with Eq. (D.13b) this gives the optimal time allocation

$$t_\ell = t\frac{\left(\sum_{m=1}^{n} k_{m\ell}\right)^{\frac{1}{3}}}{\sum_{\ell=1}^{n}\left(\sum_{m=1}^{n} k_{m\ell}\right)^{\frac{1}{3}}} \tag{D.16}$$

$$= t\frac{\left(\sum_{m=1}^{n} w_m C_{m\ell}^2\right)^{1/3} {\mu'_\ell}^{2/3}}{\sum_{k=1}^{n}\left(\sum_{m=1}^{n} w_m C_{mk}^2\right)^{1/3} {\mu'_k}^{2/3}} \tag{D.17}$$

and the time optimized figure of merit

$$\mathcal{M}(C) = \frac{1}{t^2}\left[\sum_{\ell=1}^{n}\left(\sum_{m=1}^{n} k_{m\ell}\right)^{\frac{1}{3}}\right]^3. \tag{D.18}$$

For the naive sequential protocol, we have $C = \nVdash$ and $\boldsymbol{\mu}' = \boldsymbol{\mu}$, so that

$$\mathcal{M}_{\text{naive}} = \frac{1}{t^2}\left(\sum_{\ell=1}^{n} w_\ell^{1/3}\mu_\ell^{2/3}\right)^3. \tag{D.19}$$

D.3 Nearly Overlapping Functions

Here we derive Eq. (5.40) from Eq. (5.39). For this, consider the minimization over C in Eq. (5.39). To bound the expression, first note that, for any integer $\ell \in [1, n]$, we have

$$\begin{aligned} 1 = \sum_{m=1}^{n} \delta_{\ell m} &= \sum_{m=1}^{n} \sum_{p=1}^{n} C_{\ell p} (C^{-1})_{pm} \\ &\leq \sum_{p=1}^{n} |C_{\ell p}| |\sum_{m} (C^{-1})_{pm}| \\ \implies 1 &\leq \left(\sum_{p=1}^{n} |C_{\ell p}| |\sum_{m} (C^{-1})_{pm}| \right)^2 . \end{aligned} \tag{D.20}$$

This inequality is true for all C. Also note that

$$\left(\sum_{p=1}^{n} |C_{\ell p}| |\sum_{m} (C^{-1})_{pm}| \right)^2 \geq \sum_{p=1}^{n} |C_{\ell p}|^2 |\sum_{m} (C^{-1})_{pm}|^2. \tag{D.21}$$

This inequality is an equality when $\sum_m (C^{-1})_{pm} = 0$ for all but one single $p = p^*$. When this condition is satisfied, we consequently have

$$1 \leq \sum_{p=1}^{n} |C_{\ell p}|^2 |\sum_{m} (C^{-1})_{pm}|^2. \tag{D.22}$$

Now take a weighted sum over ℓ in Eq. (D.22) and obtain

$$\begin{aligned} \sum_{\ell=1}^{n} w_\ell &\leq \sum_{\ell=1}^{n} w_\ell \sum_{p=1}^{n} |C_{\ell p}|^2 |\sum_{m} (C^{-1})_{pm}|^2 \\ &= \sum_{p=1}^{n} |\sum_{m} (C^{-1})_{pm}|^2, \end{aligned} \tag{D.23}$$

where in the second line we used the normalization from Eq. (5.29). Next use subadditivity, $\sum_p |x_p| \leq (\sum_p |x_p|^{1/3})^3$, to obtain

$$\sum_{\ell=1}^{n} w_\ell \leq \left[\sum_{p=1}^{n} |\sum_{\ell=1}^{n} (C^{-1})_{p\ell}|^{2/3} \right]^3 \tag{D.24}$$

valid for all C. The expression on the right is the one we need to minimize (over C) in Eq. (5.39). Consequently, if we can saturate the last inequality, we have found the minimum of the expression, and arrive at

$$\mathcal{M}_{\text{opt}} = \frac{\max_m \bar{a}_m^2}{t^2} \sum_{\ell=1}^{n} w_\ell + \mathcal{O}\left(\frac{\mathcal{N}\delta^2}{t^2}\right) \tag{D.25}$$

for nearly overlapping functions. We can, in fact, saturate the inequality (D.24). Recall that, in order to saturate Eq. (D.21), we require the existence of an index p^* such that

$$\sum_{\ell=1}^{n} (C^{-1})_{p\ell} = 0 \tag{D.26}$$

for $p \neq p^*$ and otherwise

$$\left(\sum_{\ell=1}^{n} C^{-1}_{p^*\ell} \right)^2 = \sum_{\ell=1}^{n} w_\ell. \tag{D.27}$$

Furthermore, we must satisfy the normalization condition in Eq. (5.29) for each column of C. Geometrically, this normalization constraint forces each column of C to be on the surface of an ellipsoid in n-dimensional space.

Suppose the row vector $C^{-1}_{p^*} = (\sqrt{\sum_{\ell=1}^{n} w_\ell}/n)\mathbf{1}^T$. This clearly satisfies Eq. (D.27). We can satisfy Eq. (D.26) by noting that Eq. (D.26) can be written as $\mathbf{1} \cdot C^{-1}_{p\neq p^*} = 0$. Therefore,

Eq. (D.26) is satisfied if the rows $p \neq p^*$ of C^{-1} are orthogonal to $C^{-1}_{p^*}$—that is, they exist in $\mathbf{1}^\perp$.

It remains to show that we can choose such a C that satisfies Eq. (5.29). We have then that the column vector of C, $C_{p^*} = (1/\sqrt{\sum_{\ell=1}^{n} w_\ell})\mathbf{1}^T$ which satisfies both $C_{p^*} \cdot C^{-1}_{p^*} = 1$ and

$$\sum_{\ell=1}^{n} w_\ell C^2_{\ell p^*} = 1. \tag{D.28}$$

The remaining columns of C exist in $\mathbf{1}^\perp$ and must exist on the n-dimensional ellipsoid specified by Eq. (5.29).

As $\mathbf{1}^\perp$ is a subspace geometrically represented as a hyperplane through the origin, it necessarily intersects the ellipsoid (centered on the origin) specified by Eq. (5.29) forming an ellipsoid of dimension $n-1$. Therefore, we can satisfy all constraints and saturate Eq. (D.21).

Furthermore, we can confirm this choice of C also saturates Eq. (D.20) as

$$\begin{aligned}\left(\sum_{p=1}^{n}|C_{\ell p}||\sum_{m}(C^{-1})_{pm}|\right)^2 &= |C_{\ell p^*}|^2|\sum_{m}(C^{-1})_{p^* m}|^2 \\ &= \frac{1}{\sum_{\ell=1}^{n} w_\ell}\sum_{\ell=1}^{n} w_\ell = 1,\end{aligned} \tag{D.29}$$

and so we have confirmed we may obtain the equality as in Eq. (D.25).

Appendix E: Technical Details of the Results Reported in Chapter 7

E.1 Fully-Symmetric, Stoquastic Hamiltonians

Consider a Hamming-symmetric, stoquastic Hamiltonian on n qubits of the form

$$H(s) = -\frac{a(s)}{n}\sum_j X_j + b(s)V\Big(\sum_j \bar{Z}_j\Big), \tag{E.1}$$

where $a(s)$ and $b(s)$ define the adiabatic schedule as a function of parameter $s \in [0,1]$ with the constraint that $a(0) = b(1) = 1$ and $a(1) = b(0) = 0$, and $\bar{Z} = (I - Z)/2 = |1\rangle\langle 1|$ is the Hamming weight operator. Since the potential V is a function only of Hamming weight, the Hamiltonian can be written in a basis such that it is block diagonal [488, 489].

A Hamiltonian of this form commutes with the operator $J^2 = J_x^2 + J_y^2 + J_z^2$ where

$$J_x = \frac{1}{2}\sum_j X_j \qquad J_y = \frac{1}{2}\sum_j Y_j \qquad J_z = \frac{1}{2}\sum_j Z_j, \tag{E.2}$$

so J^2 is a conserved quantity and we can use a basis of states $|jm\gamma\rangle : J^2|jm\gamma\rangle = j(j+1)|jm\gamma\rangle$ with total spin j and z-projections $m \in \{-j, -j+1, ..., j-1, j\}$ with γ labeling the degeneracies. Introducing a parameter $\sigma \in \{0, 1, ..., \lfloor \frac{n}{2} \rfloor\}$ to label each total spin subspace, we have $j = \frac{n}{2} - \sigma$. The degeneracy for a given total spin j and z-projection m is determined by the number of

representations of the group of permutations: $\Gamma = \binom{n}{\sigma} - \binom{n}{\sigma-1}$. Noting that m can be written in terms of Hamming weight w as $m = w - \frac{n}{2}$ we can express the basis states as

$$|w\sigma\gamma\rangle = \frac{1}{\sqrt{\binom{n}{w}}} \sum_{|x|=w} k(x)|x\rangle. \tag{E.3}$$

Here, the quantities $k(x)$ are numerical constants defining the weights of the appropriate bit strings so that we have an orthonormal basis. For the $\sigma = 0$ ($j = n/2$) subspace $k = 1 \; \forall \; x$.

Upon defining raising and lowering operators $J^{\pm} = J_x \pm iJ_y$ and noting $\sum_j X_j = J^+ + J^-$, one finds

$$H = \sum_{w,\sigma,\gamma} \Bigg[- \frac{a(s)}{n} \Big[C^+ |w, \sigma, \gamma\rangle\langle w+1, \sigma, \gamma| + C^- |w\,\sigma\,\gamma\rangle\langle w-1\,\sigma\,\gamma| \Big] + b(s) V(w) |w\,\sigma\,\gamma\rangle\langle w\,\sigma\,\gamma| \Bigg], \tag{E.4}$$

where $C^+ = \sqrt{(w-\sigma+1)(n-\sigma-w)}$ and $C^- = \sqrt{(w-\sigma)(n-\sigma-w+1)}$ are the standard raising and lowering coefficients. For a given subspace $\sigma \in \{0, 1, ..., \lfloor \frac{n}{2} \rfloor\}$, $w \in [\sigma, n-\sigma]$. This Hamiltonian is block diagonal with each block corresponding to a subspace. The $\sigma = 0$ block is a permutation-symmetric block of dimension $(n+1) \times (n+1)$. As permutation of bits is a symmetry of the Hamiltonian in E.1, and by the Perron-Frobenius theorem the ground state is non-degenerate, the ground state of the Hamiltonian must exist in this block. Therefore the only relevant gap is within this subspace, so one can directly diagonalize this permutation symmetric subspace to analyze the performance of AQC on Hamiltonians of the form of Eq. (E.1) up to a large number of qubits.

E.2 Relabeling Bases for Three or Fewer Hamming-Symmetric Wells

Consider Hamiltonians such as given by Eq. (7.1),

$$H(s) = -\frac{a(s)}{n}\sum_j X_j + \sum_{k=0}^{K} b_k(s) V_k\Big(\sum_j X^{\bar{k}}\bar{Z}_j X^{\bar{k}}\Big),$$

for the simplest non-trivial cases, $K = 2, 3$. For these two cases, it is possible to relabel the bit strings in a manner which preserves Hamming separations and makes possible efficient exact calculations of the relevant spectral gaps.

$\mathbf{K = 2}$: Given any two n-bit strings β_a and β_b we can always introduce a relabeling that preserves the Hamming distance n_1 between them, such that these two bit strings have the form

$$\begin{aligned} \beta_a &= |0..0 \;\; 0..0\rangle \\ \beta_b &= |\underbrace{1..1}_{n_1}\underbrace{0..0}_{n_2}\rangle. \end{aligned} \tag{E.5}$$

By construction, n_1 is the Hamming distance between the bit strings and $n_2 = n - n_1$. We identify the subset of the first n_1 relabeled bits as S_1 and the remaining n_2 bits as the subset $S_2 = \bar{S}_1$. For a general relabeled bit string, define h_1 to be the number of ones in S_1, and the quantity h_2 the number of ones in S_2. For each subset, the parameters $\sigma_1, \sigma_2, \gamma_1$, and γ_2 may be defined as in the Hamming symmetric case. This defines a new labeling of basis states given by $|h_1 h_2 \sigma_1 \sigma_2 \gamma_1 \gamma_2\rangle$.

From the Perron-Frobenius theorem and symmetry we know the ground state exists in the $\sigma_1 = \sigma_2 = 0$ subspace. In particular, the Perron-Frobenius theorem guarantees a non-degenerate ground state with non-negative amplitudes. Additionally, the symmetry group for this Hamil-

tonian is the direct sum of symmetric groups $S_{n_1} \oplus S_{n_2}$ acting on the sets S_1 and S_2 respectively. The trivial representation is associated with our product state basis $|h_1 h_2 \sigma_1 \sigma_2 \gamma_1 \gamma_2\rangle$ for $\sigma_1 = \sigma_2 = \gamma_1 = \gamma_2 = 0$. However we could also write a basis diagonal in the total "spin" σ which for $\sigma = 0$ is the 1D representation of the group consistent with the non-negative amplitude requirements of the Perron-Frobenius theorem. Therefore the ground state is guaranteed to transform within this one-dimensional representation group of the Hamiltonian. As $\sigma_1 + \sigma_2 \leq \sigma$ this subspace is fully within the $\sigma_1 = \sigma_2 = 0$ subspace in the product basis.

Therefore, as the ground state exists in this $\sigma_1 = \sigma_2 = 0$ subspace, the relevant gap is also in this subspace, which has dimension $(n_1 + 1)(n_2 + 1) \times (n_1 + 1)(n_2 + 1)$. The relevant gap can be exactly computed by direct diagonalization of a matrix of dimension polynomial in both K and n.

With σ and γ labels dropped for compactness the Hamiltonian in this space can be exactly written as

$$\begin{aligned} H = \sum_{h_1 h_2 \sigma_1 \sigma_2 \gamma_1 \gamma_2} \Bigg[& - \frac{a(s)}{n} \Big[C_1^+ |h_1 h_2\rangle\langle h_1 + 1\, h_2| \\ & + C_1^- |h_1 h_2\rangle\langle h_1 - 1\, h_2| + C_2^+ |h_1 h_2\rangle\langle h_1\, h_2 + 1| \\ & + C_2^- |h_1 h_2\rangle\langle h_1\, h_2 - 1| \Big] + \sum_{i \in \{1,2\}} b_i(s) V_i(h_1, h_2) \Bigg] \end{aligned} \tag{E.6}$$

where $C^{\pm}$ act just on the relevant subset.

$\mathbf{K = 3}$: For this case, the three selected bit strings are relabeled as follows:

$$\begin{aligned}
\beta_a &= |1..1\ 0..0\ 0..0\ 0..0\rangle \\
\beta_b &= |0..0\ 1..1\ 0..0\ 0..0\rangle \\
\beta_c &= |\underbrace{0..0}_{n_1}\underbrace{0..0}_{n_2}\underbrace{1..1}_{n_3}\underbrace{0..0}_{n_4}\rangle.
\end{aligned} \tag{E.7}$$

As was the case for two wells, a basis of the form $\bigotimes_{k=\{1,2,3,4\}}|h_k\sigma_k\gamma_k\rangle$ exists. Again via symmetry of the Hamiltonian, the ground state exists in the $\sigma_k = 0\,\forall\, k$ subspace, whose dimension is polynomial in K and n.

E.3 Tight-Binding Matrix Elements

Consider 7.4 reproduced here

$$\begin{aligned}
&H_{ij}^{(TB)} = \\
&\langle\psi_i|H_d + \sum_{h_1h_2}[V_i(h_1,h_2) + V_j(h_1,h_2) + V_c(h_1,h_2)]|\psi_j\rangle
\end{aligned} \tag{E.8}$$

where H_d is the driver part of the Hamiltonian in the appropriate basis V_i and V_j are the diagonal potential terms corresponding to the i^{th} and j^{th} wells and

$$V_c = \frac{\sum_{k\neq i,j}\sum_{r_k=0}^{n} N(h_1,h_2,n_1,R_{ik},R_{jk},r_k)V_k(r_k)}{\sqrt{\binom{n_1}{h_1}\binom{n-n_2}{h_2}}}. \tag{E.9}$$

Note that this is equivalent to E.6 with the additional correction factor V_c. Recall that n_1, R_{ik} and R_{jk} are the Hamming distance between the i^{th} and j^{th} wells, the i^{th} and k^{th} wells and

the j^{th} and k^{th} wells, respectively, r_k is the distance from the k^{th} well and $V_k(r_k)$ is the potential due to the k^{th} well at distance r_k. The function $N(h_1, h_2, n_1, R_{ik}, R_{jk}, r_k)$ gives the number of points of intersection between Hamming spheres of radius $r_i = h_1 + h_2$ and $r_j = h_2 + (n_1 - h_1)$ centered on the i^{th} and j^{th} wells respectively and the Hamming sphere of radius r_k centered on the k^{th} well.

To find the function $N(h_1, h_2, n_1, R_{ik}, R_{jk}, r_k)$, without loss of generality consider 3 wells i, j, k shifted so they are in the form of E.7 where we label the corresponding sets of qubits as n_1', n_2', n_3', n_4' to differentiate from n_1 and n_2 in the 2 well basis for wells i and j.

$$n_1 = n_1' + n_2' \qquad R_{ik} = n_1' + n_3' \qquad R_{jk} = n_2' + n_3'. \tag{E.10}$$

which we can solve for n_1', n_2', n_3' in terms of the input parameters.

Define h_1', h_2', h_3', h_4' as the number of ones in each of the 4 subsets of qubits. Therefore

$$\begin{aligned} r_i &= (n_1' - h_1') + h_2' + h_3' + h_4' \\ r_j &= h_1' + (n_2' - h_2') + h_3' + h_4' \\ r_k &= h_1' + h_2' + (n_3' - h_3') + h_4'. \end{aligned} \tag{E.11}$$

This is a system of linear Diophantine equations whose solutions satisfy

$$h_2' = \frac{1}{2}(n_2' + r_i - r_j - n_1') + h_1' \tag{E.12}$$

$$h_3' = \frac{1}{2}(n_3' + r_i - r_k - n_1') + h_1' \tag{E.13}$$

$$h_4' = \frac{1}{2}(n_2' + n_3' - r_j - r_k) - h_1'. \tag{E.14}$$

It is straightforward to then count the solutions for $h_1' \in [0, n_1']$. We try each possible h_1' and check that $h_2' \in [0, n_2']$ and $h_3' \in [0, n_3']$. Each solution found is then multiplied by the combinatoric factor $\binom{n_1'}{h_1'}\binom{n_2'}{h_2'}\binom{n_3'}{h_3'}\binom{n_4'}{h_4'}$. The total result is N.

E.4 Proof of Possible Subspaces for First Excited State

Here we prove that the first excited state for a single Hamming symmetric well must exist in either the $\sigma = 0$ or $\sigma = 1$ subspace. Consider an eigenstate $|\psi\rangle = \sum_{w\sigma\gamma} \alpha(w, \sigma, \gamma)|w\sigma\gamma\rangle$. Then

$$H|\psi\rangle = \sum_{w\sigma\gamma} \Bigg[-\frac{1-s}{n}\Big(\alpha(w+1,\sigma,\gamma)\sqrt{(w-\sigma+1)(n-\sigma-w)} + \alpha(w-1,\sigma,\gamma)\sqrt{(w-\sigma)(n-\sigma-w+1)}\Big) - s\alpha(w,\sigma,\gamma)V(w)\Bigg]|w\sigma\gamma\rangle \quad \text{(E.15)}$$

which implies that for $\alpha(w, \sigma, \gamma) \neq 0$ that

$$E(s) = -\frac{1-s}{n}\left(r^+C^+ + r^-C^-\right) + sV(w) \quad \text{(E.16)}$$

where $r^\pm = \frac{\alpha(w\pm1,\sigma,\gamma)}{\alpha(w,\sigma,\gamma)}$ for all w, σ, γ and $C^\pm$ are the raising and lowering coefficients. Now consider the energy difference between candidate first excited states with different σ. The potential term is independent of σ and thus does not affect which subspace is energetically favored. For a subspace σ and a subspace $\sigma' > \sigma$, $C^\pm(\sigma' > \sigma) < C^\pm(\sigma)$ so if $r^\pm$ were independent of σ then the $\sigma = 0$ subspace would always be favored. Now consider the difference in energy between the

candidate first excited states:

$$\Delta E(s) = E_{\sigma'}(s) - E_{\sigma}(s) = \frac{1-s}{n}[r_{\sigma}^{+}C_{\sigma}^{+} + r_{\sigma}^{-}C_{\sigma}^{-} - r_{\sigma'}^{+}C_{\sigma'}^{+} - r_{\sigma'}^{-}C_{\sigma'}^{-}] \tag{E.17}$$

The above equation is true for all w. We take $w = \sigma'$ so $C_{\sigma'}^{-}(w = \sigma') = 0$ eliminating one term. For $\sigma > 0$, $r^{\pm}$ must be nonnegative (by the Perron-Frobenius theorem) so $\Delta E(s)$ is nonnegative unless $r_{\sigma'}^{+}$ is large relative to $r_{\sigma}^{\pm}$. We will now show a contradiction. Consider the $w = \sigma'$ element of the eigenvector equation in both subspaces. In the σ' subspace

$$-\frac{1-s}{n}C_{\sigma'}^{+}(w = \sigma')\alpha(w = \sigma' + 1, \sigma') + sV(w = \sigma')\alpha(w = \sigma', \sigma') = E_{\sigma'}\alpha(w = \sigma', \sigma') \tag{E.18}$$

and in the σ subspace

$$\begin{aligned} -\frac{1-s}{n}\Big[C_{\sigma}^{+}(w = \sigma)\alpha(w = \sigma' + 1, \sigma) + C_{\sigma}^{-}(w = \sigma)\alpha(w = \sigma' - 1, \sigma)\Big] \\ + sV(w = \sigma')\alpha(w = \sigma', \sigma) = E_{\sigma}\alpha(w = \sigma', \sigma) \end{aligned} \tag{E.19}$$

Rearranging and dropping the arguments of the functions for compactness we obtain from the fact $C_{\sigma'}^{+} < C_{\sigma}^{+}$

$$\big(E_{\sigma'} - sV\big)\frac{1}{r_{\sigma'}^{+}} > \big(E_{\sigma} - sV\big)\frac{1}{r_{\sigma}^{+}} + \frac{1-s}{n}C_{\sigma}^{-}\frac{r_{\sigma}^{-}}{r_{\sigma}^{+}}. \tag{E.20}$$

The last term is positive definite so we can drop it and get the inequality

$$\frac{E_{\sigma'} - sV}{E_\sigma - sV} > \frac{r^+_{\sigma'}}{r^+_\sigma}. \tag{E.21}$$

Both sides are positive definite for $\sigma, \sigma' > 0$. And as sV is the same in both subspaces, if $E_{\sigma'} < E_\sigma$ then $r^+_{\sigma'} < r^+_\sigma$ but this contradicts the result from E.17 that for this to be true $r^+_\sigma < r^+_{\sigma'}$. Therefore the first excited state must always exist either in the $\sigma = 0$ or $\sigma = 1$ subspaces.

Appendix F: Technical Details of the Results Reported in Chapter 8

F.1 Minimal Example

Here we give a minimal, working example of our algorithm. We consider a Hamming-symmetric Hamiltonian

$$H = -\sum_{\|b\|_H=1} X_b + \sum_{\|b\|_H=1} Z_b, \tag{F.1}$$

and define the Hamming-symmetric Hamiltonian with base v as $H_v = X_v H X_v$, yielding

$$H_v = -\sum_{\|b\|_H=1} X_b + \sum_{\|b\|_H=1} (-1)^{v\cdot b} Z_b. \tag{F.2}$$

This corresponds to a graph Γ where V^*_Γ is a hypercube with all edge weights $\alpha_b = 1$. The vertex $\{\infty\}$ is connected to every vertex $u \in V^*_\Gamma$ by an edge weight $w(u, \infty) = \sum_{|b|=1} (-1)^{v\cdot b + u\cdot b}$. In practice, this graph Γ is too large to handle, but for illustrative purposes, we

consider only 3 qubits so that we can track the whole algorithm by hand. Furthermore, in order to be specific, we assume that $v = 010$. That is we obtain,

$$H_{010} = -X_{100} - X_{010} - X_{001} + Z_{100} - Z_{010} + Z_{001}$$

$$= -X_{\mathbf{0}} - X_{\mathbf{1}} - X_{\mathbf{2}} + Z_{\mathbf{0}} - Z_{\mathbf{1}} + Z_{\mathbf{2}}$$

which has a corresponding graph Γ depicted in Fig. F.1.

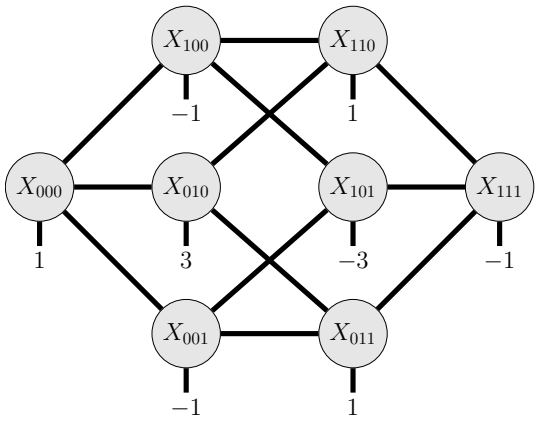


Figure F.1: Γ for H_{010}. Each vertex is labeled by a member of X_b. The disconnected edges connect to the boundary vertex ∞ and are labeled by their weights.

Below, we follow the steps in the main paper to obtain the clausal theory graph for H_{010} excluding assignments, $M_1(E_G)$ (See Fig. F.2).

In particular, as we have three qubits we introduce the six literal vertices

$$Z_{\mathbf{0}}^{(a)}, -Z_{\mathbf{0}}^{(a)}, Z_{\mathbf{1}}^{(a)}, -Z_{\mathbf{1}}^{(a)}, Z_{\mathbf{2}}^{(a)}, \text{ and } -Z_{\mathbf{2}}^{(a)}.$$

We also have three single bit flip edge generators $X_{\mathbf{0}}, X_{\mathbf{1}}$ and $X_{\mathbf{2}}$. Each generator has an associated vertex in the clausal theory graph, which joins its respective literals by an undirected edge. That is, since $X_{\mathbf{0}}$ maps the computational basis state $00\ldots0$ to $10\ldots0$, we use a gadget to

connect $Z_{\mathbf{0}}$ to $Z_{\mathbf{1}}$.

As there are no Pauli-Y terms in the Hamiltonian, we introduce no weight generators. Finally, we add clause vertices corresponding to the diagonal potential terms of the Hamiltonian. In particular, from Eq. (F.2), we introduce three clause cluster vertices $Z_{\mathbf{0}}^{(z_b=1)}, -Z_{\mathbf{1}}^{(z_b=1)}$ and $Z_{\mathbf{2}}^{(z_b=1)}$ corresponding to to their respective terms in the Hamiltonian. For each clause cluster vertex, we identify the associated set of clause vertices. Recall for a particular choice of $b \in \{100, 010, 001\}$, we have the associated clause cluster set

$$\mathscr{C}_b = \left\{ \left\{ (-1)^{b'_i} Z_{\boldsymbol{i}} \right\}_{b_i \neq 0}^{(d)} | (-1)^{b \cdot b'} = \mathrm{sign}(\kappa_b) \right\}_{b' \in \{0,1\}^n}. \tag{F.3}$$

where in our case $\kappa_{100} = \kappa_{001} = 1$ and $\kappa_{010} = -1$. Consider, for example, the $\mathscr{C}_{b=100}$ case corresponding to the clause cluster vertex $Z_{\mathbf{0}}^{(z_b=1)}$. Then, $(-1)^{b \cdot b'} = \mathrm{sign}(\kappa_b) = 1$, whenever $b' \in \{000, 001, 010, 011\}$. Hence, noting that $b_i \neq 0$ only when $i = 0$, $\mathscr{C}_{100} = \left\{ \{Z_{\mathbf{0}}\}^{(d)} \right\}$. Similarly, the clause cluster vertices $-Z_{\mathbf{1}}^{(z_b=1)}$ and $Z_{\mathbf{2}}^{(z_b=1)}$ have single associated clause vertices $\{-Z_{\mathbf{1}}\}^{(d)}$ and $\{Z_{\mathbf{2}}\}^{(d)}$ respectively. The clause cluster vertices connect their associated clause vertex to its corresponding literal vertices or, in this case, vertex.

Fig. F.2 shows the final construction. One can see that including assignments for $u, v \in \Gamma$ demonstrates $u \equiv v \iff G(u) \simeq G(v)$. This will be shown explicitly in the following walk-through of the algorithm.

We start by calling Algorithm 1 with a null argument, or FINDEFFECTIVEVERTICES($\emptyset$). Then, in Line 3, we choose a random vertex $u \in V_\Gamma^*$. Suppose that $u = X_{100}$; then Line 4 adds it to V', so that now $V' = \{X_{100}\}$. We now reach Line 5 with $N(X_{100}) = X_{000}, X_{110}, X_{101}$. The loop proceeds to recursively check each neighbor $v \in N(X_{100})$. Line 6 calls Algorithm 3 with

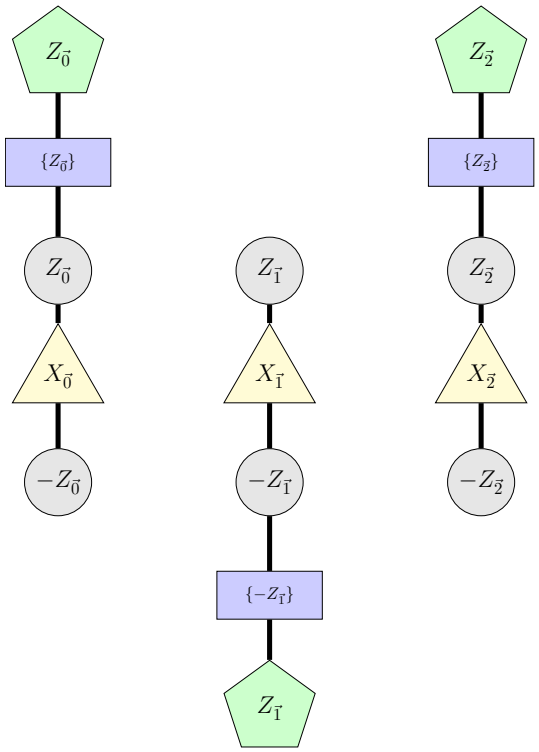


Figure F.2: Full example of $M_1(E_\Gamma)$ for H_{010}.

argument (v, V') and if v is not equivalent to an element already in V', Line 7 adds v to V'. Then, in Line 8, the Algorithm 1 calls itself with argument (v, V').

To walk through these steps, suppose that in the first iteration of the loop, the first neighbor we query is $v = X_{000}$. We generate $G(X_{100})$ and $G(X_{000})$ as shown in Fig. F.3a and Fig. F.3b, respectively. One can see that these graphs are not isomorphic, so we add X_{000} to V'. At this point, $V' = \{X_{100}, X_{000}\}$.

Having added X_{000} to V', we continue the recursion and call FINDEFFECTIVEVERTICES(X_{000}, V'). Since $v \neq \emptyset$, we immediately proceed to the loop in Line 5 and begin checking the neighbors of X_{000}. Suppose the first neighbor called is X_{001}. (See Fig. F.3c for the corresponding clausal theory graph.) Now, we see that $G(X_{001}) \simeq G(X_{100})$, so that $X_{001} \equiv X_{100}$. Because $X_{100} \in V'$, we do not add X_{001} to V', this branch of the recursion terminates, and we continue checking the

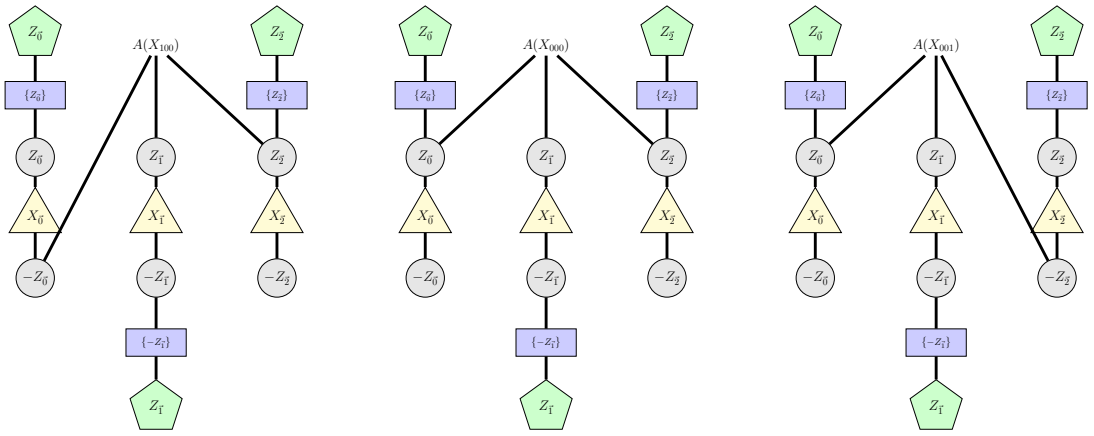
remaining neighbors of X_{000}.

Next, we check X_{010}, find that it is not equivalent to any member of V', and add to V'. Now, we call FINDEFFECTIVEVERTICES(X_{010}, V'). Note that in this call, all of the members of $N(X_{010})$ are already equivalent to members of V'. Thus, Lines 5-8 complete without adding any new members to V', and we return V' unchanged. We now return to the parent process, which had been considering members of $N(X_{000})$. Since each member of $N(X_{000})$ has been queried, Line 9 returns V' to its parent process, which was checking the members of $N(X_{100})$. We next query X_{110} and find $X_{110} \equiv X_{000}$, so X_{110} is not added to V'. In the next iteration of the loop, however, X_{101} is not equivalent to anything already in V' so it is added and we call FINDEFFECTIVEVERTICES(X_{101}, V').

The remaining unchecked neighbor of X_{101} is X_{111}, which is equivalent to X_{100}, so we return V' to its parent process and, in turn, the algorithm completes and exits, returning $V' = \{X_{100}, X_{000}, X_{010}, X_{101}\}$. Thus, as expected, we have one representative of each equivalence class of vertices of the original graph Γ in Fig. F.1.

We note that in this example every vertex is checked, so it may not be immediately obvious that this procedure is efficient in general. However, as described in the main text, this algorithm requires checking only a polynomial number of vertices (provided the symmetries lead to a polynomial sized effective subspace). See Fig. F.4 to see which vertices need to be checked by this algorithm for a Hamming symmetric example $H_v = H_{\{00\ldots0\}}$ with $5, 6,$ and 7 qubits. It is clear that while the number of vertices in the graph increases exponentially with the number of qubits, only polynomially many vertices are actually queried.

Once we have our effective set of vertices we proceed to calculate the weights Ω_{uv} in the



(a) $G(X_{100})$ (b) $G(X_{000})$ (c) $G(X_{001})$

Figure F.3: Some clausal theory graphs used in the minimal example. Comparing Fig. F.3a to Fig. F.3b shows $G(X_{100}) \not\simeq G(X_{000}) \iff X_{100} \not\equiv X_{000}$. Comparing Fig. F.3a to Fig. F.3c shows $G(X_{100}) \simeq G(X_{001}) \iff X_{100} \equiv X_{001}$.

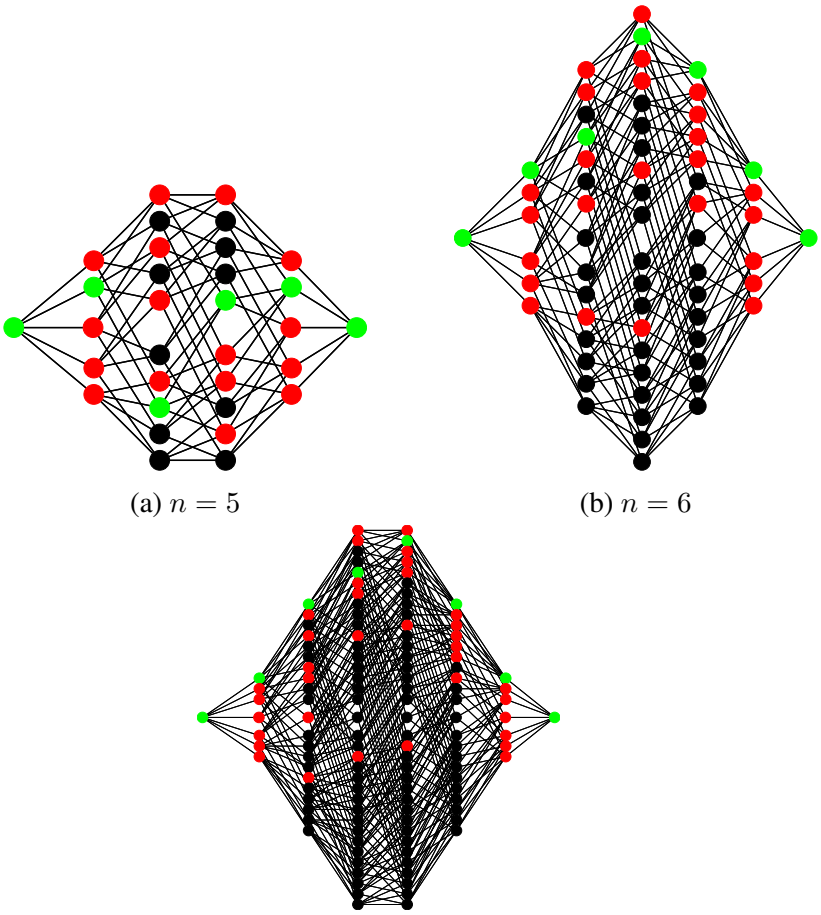
main loop of Algorithm 2. One can walk through this loop by hand with ease obtaining the matrix

$$\Omega = \begin{pmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 1 & 0 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{pmatrix}$$

where the rows and columns are in the same order as listed above in V'. Given Ω and V' we can calculate

$$H' = \begin{pmatrix} -1 & -2 & 0 & -1 \\ -2 & 1 & -1 & 0 \\ 0 & -3 & 3 & 0 \\ -3 & 0 & 0 & -3 \end{pmatrix}$$

from which we can compute the ground state in the symmetric subspace, ϕ'. We can also compute



(a) $n = 5$ (b) $n = 6$

(c) $n = 7$

Figure F.4: Visited vertices of Γ in FINDEFFECTIVEVERTICES for Hamming symmetric example on a hypercube for $n = 5, 6, 7$ as in Eq. (F.1). Green vertices are effective vertices, red vertices are checked, but not added to the set of effective vertices, and black vertices are unchecked. Columns of vertices are in Hamming weight order. Note the number of vertices grows exponentially in n, but the number of checked vertices grows only polynomially.

the size of each equivalence class using Eq. (8.8) from the main text (repeated here)

$$\frac{|[\![u]\!]|}{\sum_{v\in V'}|[\![v]\!]|} = \left(\sum_{v\in V'}\prod_{e\in P(u,v)}\frac{\omega_{e_0[\![e_1]\!]}}{\omega_{e_1[\![e_0]\!]}}\right)^{-1} \tag{F.4}$$

which yields

$$|[\![X_{100}]\!]| = |[\![X_{000}]\!]| = 3$$

$$|[\![X_{010}]\!]| = |[\![X_{101}]\!]| = 1.$$

Finding the ground state eigenvector of H' gives $\phi' = (3+2\sqrt{2}, 1+\sqrt{2}, 1, 7+5\sqrt{2})$. Now, we compute the probability of sampling each equivalence class using equation Eq. (8.7) from the main text (repeated here),

$$\Pr\left([\![u]\!]\right) = |[\![u]\!]|\phi(u)^2 = \frac{|[\![u]\!]|\phi'([\![u]\!])^2}{\sum_{v\in V'}|[\![v]\!]|\phi'([\![v]\!])^2}. \tag{F.5}$$

That is,

$$\Pr(|[\![X_{100}]\!]|) \approx 0.32 \quad \Pr(|[\![X_{000}]\!]|) \approx 0.055$$

$$\Pr(|[\![X_{010}]\!]|) \approx 0.003 \quad \Pr(|[\![X_{101}]\!]|) \approx 0.622.$$

We then return a member of V' according to the above-stated probability distribution. We can easily verify that this agrees with the probability that, upon computational-basis measurements, we return a member of the corresponding equivalence class of the full Hamiltonian.

In addition to this minimal full example we also provide a depiction of a more complicated clausal theory graph that includes all the different gadget described in the main text (see Fig. F.5). Note that this example is not stoquastic so that all the different types of gadgets can be shown clearly on one graph. Thus this example is for illustrative purposes only.

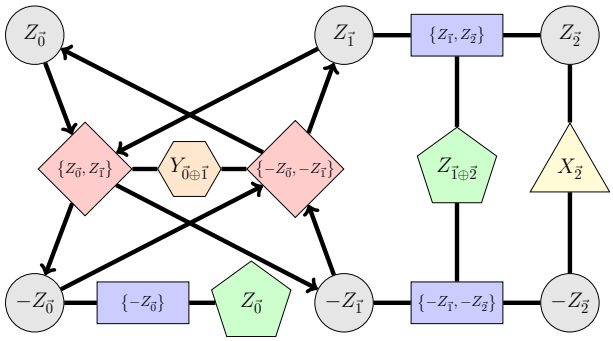


Figure F.5: Full example of $M_1(E_\Gamma)$ for $H = -X_{001} - Y_{110} - Z_{100} + Z_{011}$. Note, H is not stoquastic as to have a legible graph with all relevant types of vertices/edges.

F.2 Additional Gadgets

Generality: In addition to the gadgets corresponding to individual X_b, Y_b, Z_b terms of the Hamiltonian one must also consider additional composite gadgets to achieve full generality for stoquastic Hamiltonians with symmetries between "like" terms – i.e. those with interactions between the same number of qubits. Such a gadget is constructed as follows. We note that for each double-typed term of the form $X_b Y_{b'}, X_b Z_{b''}$, or $Y_{b'} Z_{b''}$ appearing in H, one would need to introduce re-colored versions of the gadgets $G_1(b), G_2(b')$, and $G_3(b'')$ in order to avoid conflicting with other terms that might appear independently, such as X_b. Similarly, for each term of the form $X_b Y_{b'} Z_{b''}$ one would need to introduce yet another color to avoid conflicting with single-typed and double-typed terms. Alternatively, as long as one is consistent for all terms, if $X_b Y_{b'}$ appears and $G_1(b)$ is already included in the clausal theory graph, one might just adjoin an indicator vertex with a unique color to indicate that both single- and double-typed terms exist. See Fig. F.6.

Furthermore, as noted in the main text, one could imagine similar constructions that allow for the detection of even more complicated symmetries including those between "non-like"

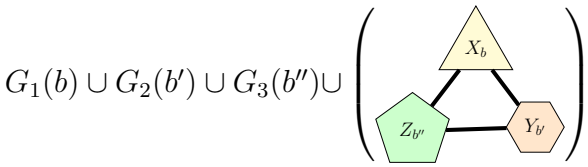


Figure F.6: Construction of a composite gadget for a Hamiltonian term of the form $X_b Y_{b'} Z_{b''}$. $G_1(b), G_2(b')$, and $G_3(b'')$ are as defined in the main text (see Table 1).

Hamiltonian terms. While the current work presented here covers all cases previously described in the literature, it is of interest to consider these more general constructions. This combined with a full accounting for algorithms to deal with approximate symmetries would serve to fully generalize the work here. We leave details of such constructions to forthcoming work.

Optimality: The gadgets in this paper were chosen for simplicity, but are suboptimal from the perspective of GI. For instance, one could combine the gadgets for XX ($G_1(11)$) and YY ($G_2(11)$) into a re-colored version of $G_2(11)$. One can similarly include ZZ terms in a gadget of this form. (See the Y_{110} and Z_{011} gadgets of Fig. F.5 and note that the structure of a potential Z_{110} gadget would also be captured by Y_{110} by deleting edges pointing towards literals $\pm Z_{\mathbf{0}}$ and $\pm Z_{\mathbf{1}}$.)

If one wishes to make the algorithm as efficient as possible, one would ideally choose such a minimalist construction.

F.3 Proofs

Proof of Theorem 8.7.1. By construction $M : \Gamma \mapsto G$ is unique, so we only need to show that the M^{-1} exists. First, note that $M_0 : V_\Gamma \longrightarrow M_0[V_\Gamma] \subset V_G$ is bijective and provides a unique mapping from vertices to assignments of literals. Hence, we only need to show how to derive $w(u, v)$ from G for all $(u, v) \in V_\Gamma \times V_\Gamma$.

For any $u \in V_\Gamma$, we calculate $w(u, \infty)$ by considering all shortest paths P_b connecting $A(u)^{(\acute{b})}$ to each clause cluster vertex $Z_b^{(z_b)}$ through G. In particular we obtain

$$w(u, \infty) = 2 \sum_{b:|P_b| \neq \|b\|_H} |\kappa_b| - \sum_b |\kappa_b| \tag{F.6}$$

where we recall that we can extract $|\kappa_b|$ from the color z_b of each clause cluster vertex.

Similarly, we can consider any pair $(u, X_b u) \in V_\Gamma \times V_\Gamma$. We consider $\mathcal{M}(u, X_b u) = (L \cup -L) \setminus (M_0(u) \cap M_0(v))$ and construct $S = \mathcal{M}(u, X_b u) \cup M_1(\{u, X_b u\})$. Let P_b be the number of shortest paths connecting $A(u)^{(\acute{b})}$ to $A(X_b u)^{(\acute{b})}$ through S. Then,

$$w(u, X_b u) = \begin{cases} \alpha_b + (-1)^{\frac{|P_b|}{\|b\|_H^2}} \beta_b & \text{if} G_1(b) \cup G_2(b) \subset G \\ 0 & \text{otherwise} \end{cases} \tag{F.7}$$

Hence, $\exists\, M^{-1} : G \mapsto \Gamma$ and M is bijective. □

Recall $G(u) = M_0(u) \cup M_1(E_\Gamma)$. We now provide a straightforward lemma that we use to prove Theorem 8.7.2. In the below proofs we drop the $(\acute{b})$ superscript on assignment vertices to avoid notational clutter.

Lemma F.3.1. *There exists a color-preserving automorphism g of G with $g \circ A(u) = A(v)$ iff $G(u) \simeq G(v)$.*

Proof. When $u = v$, this is trivial. We now assume $u \neq v$.

Suppose g is an automorphism of G with $g \circ A(u) = A(v)$. Then, clearly, $g\big|_{V_{G(u)}} : V_{G(u)} \longrightarrow V_{G(v)}$ is a color-preserving isomorphism between $G(u)$ and $G(v)$.

Conversely, suppose there exists a color-preserving isomorphism $g\big|_{V_{G(u)}} : V_{G(u)} \longrightarrow V_{G(v)}$

such that $G(u) \simeq G(v)$. We note that by construction $G(u) \cap G(v) = M_1(E_\Gamma)$ and color-consistency requires $g\big|_{V_{G(u)}} \circ A(u) = A(v)$. Therefore, $g\big|_{V_{G(u)}}$ is an automorphism of $M_1(E_\Gamma)$. Extend $g\big|_{V_{G(u)}}$ to $g : V_G \longrightarrow V_G$ such that $g \circ A(X_b) = \left\{ g\left((-1)^{b_i} Z_{\boldsymbol{i}}^{(a)} \right) \right\}_{i=0}^{n-1} = A(X_{b'})$. Then, g is an isomorphism between $M_0(X_b) \cup M_1(E_\Gamma)$ and $M_0(X_{b'}) \cup M_1(E_\Gamma)$ for any choice of b. Hence, g is an automorphism of G.

□

Now, we can prove Theorem 8.7.2.

Proof of Theorem 8.7.2. Suppose $u \equiv f(u) \in V_\Gamma$. By definition, $f \in \mathrm{Aut}(\Gamma)$. By construction, we have that g is a color-preserving isomorphism $G(u) \simeq G(f(u))$.

Now, suppose that $g\big|_{V_{G(u)}}$ is a color-preserving isomorphism $G(u) \simeq G(v)$. Then, by Lemma F.3.1, there exists some color-preserving automorphism $g : V_G \longrightarrow V_G$ of G. Define $f = A^{-1} \circ g \circ A$. Then, by Theorem 8.7.1, f is an automorphism of Γ and, thus, $u \equiv f(u)$. □

F.4 Smooth Transitions

We apply the result of [175] under very weak constraints to show that families of Hamiltonians almost invariably encounter an exponentially small gap, unless they undergo very smooth phase transitions. These results are similar to but, in terms of gap-analysis, stronger than those of [182]. Here, we reference only the behavior of the ground state and show that most phase transitions, like those we expect out of adiabatic optimization, produce exponentially small gaps. The following simple theorem is sufficiently illustrative, although its statement could be improved asymptotically and easily generalized to include more than k-local Hamiltonians.

Theorem F.4.1. *If H is a k-local Hamiltonian with ground state ϕ, $\|H\| \leq 1$, and there exists a set S_0 such that*

1. $S_0 = \left\{u \mid |\phi(u)| < 2^{-n^c}\right\}$ *with absolute constant $c > 0$,*
2. $\sum_{u \in S_0} |\phi(u)|^2, \sum_{u \notin S_0} |\phi(u)|^2 = \Omega\left(\frac{1}{\mathsf{poly}(n)}\right)$,
3. *and* $\left|\overline{S_0}\right| = O\left(\mathsf{poly}(n)\right)$*;*

then $\gamma(H) = 2^{-\Omega(n^c)}$.

Proof. Note that by [175], the weighted Cheeger constant h bounds $\gamma(H)$, as $2h \geq \gamma(H)$. In particular, $h = \min_S h_S$ where h_S is the weighted Cheeger ratio

$$h_S = \frac{\sum_{u \in S, v \notin S} \Re\left(-H_{uv}\phi(u)\phi(v)\right)}{\min\left\{\sum_{u \in S} |\phi(u)|^2, \sum_{u \notin S} |\phi(u)|^2\right\}}.$$

Now, consider S_0,

$$\begin{aligned}
h_{S_0} &\leq O\left(\frac{\max_{v_0 \in S_0} |\phi(v_0)| \sum_{u \in S_0, v \notin S_0} |-H_{uv}||\phi(u)|}{\min\left\{\sum_{u \in S_0} |\phi(u)|^2, \sum_{u \notin S_0} |\phi(u)|^2\right\}}\right) \\
&= O\left(\frac{\mathsf{poly}(n) \max_{v \in S_0} |\phi(v)|}{\min\left\{\sum_{u \in S_0} |\phi(u)|^2, \sum_{u \notin S_0} |\phi(u)|^2\right\}}\right) \\
&= O\left(\frac{\mathsf{poly}(n) 2^{-n^c}}{\min\left\{\sum_{u \in S_0} |\phi(u)|^2, \sum_{u \notin S_0} |\phi(u)|^2\right\}}\right) \\
&= 2^{-\Omega(n^c)}.
\end{aligned}$$

Since the weighted Cheeger constant $h = \min_S h_S < h_{S_0} = 2^{-\Omega(n^c)}$, $\gamma(H) = 2^{-\Omega(n^c)}$. □

Thus, if we are interpolating over a family of Hamiltonians $H(s)$ and we ever encounter a ground state ϕ such that (1) there exists a set S where we have substantial probability of re-

turning a sample from either S or $\overline{S}$, (2) for any $u \in S$ it is unlikely that we will sample u in time $O(\mathsf{poly}(n))$, and (3) $\overline{S}$ is small, we encounter an exponentially small gap. Since we typically interpolate over a family of Hamiltonians $H(s)$ such that the ground state ϕ_0 of $H(0)$ has $|\phi_0(u)|^2 = O\left(2^{-n/2}\right)$ for any u and end in a Hamiltonian $H(1)$ such that the ground state ϕ_1 satisfies $\sum_{u\in S}|\phi_1(u)|^2 = \Omega(1/\mathsf{poly}(n))$ for some small set of computational basis states $|S| = O(\mathsf{poly}(n))$, avoiding the constraints of Theorem F.4.1 with naive families $H(s)$ is unlikely.

Appendix G: Technical Details of the Results Reported in Chapter 9

G.1 Some Properties of the Star Product

Here we demonstrate a number of useful identities regarding the star product introduced in the main text. Some of these properties do not seem to be well documented in the literature on generalized Bloch vectors due to the focus on single density matrices. In these contexts, only star products between the same Bloch vector arise, which obscures some of the more general properties of the product. In particular, we emphasize that the product is non-associative.

We define the star product of two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{d^2-1}$ component-wise as

$$(\boldsymbol{a} * \boldsymbol{b})_\xi = d_{\mu\nu\xi} a_\mu b_\nu, \tag{G.1}$$

where $d_{\mu\nu\xi}$ are totally symmetric structure constants for $\mathfrak{su}(d)$. Observe that this product is basis-

dependent due to the structure constants.

The star product has the following properties which can be verified by explicit component-wise computation:

$$\boldsymbol{a} * \boldsymbol{b} = \boldsymbol{b} * \boldsymbol{a} \qquad \text{(commutative)}$$

$$(\boldsymbol{a} * \boldsymbol{b}) * \boldsymbol{c} \neq \boldsymbol{a} * (\boldsymbol{b} * \boldsymbol{c}) \qquad \text{(non-associative)}$$

$$\boldsymbol{a} * (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} * \boldsymbol{b} + \boldsymbol{a} * \boldsymbol{c} \qquad \text{(distributive)}$$

$$(\boldsymbol{a} * \boldsymbol{b}) \cdot \boldsymbol{c} = (\boldsymbol{b} * \boldsymbol{c}) \cdot \boldsymbol{a} = (\boldsymbol{a} * \boldsymbol{c}) \cdot \boldsymbol{b}.$$

The last identity can be used to show that

$$\begin{aligned}(\boldsymbol{a} * \boldsymbol{b}) \cdot (\boldsymbol{c} * \boldsymbol{d}) &= ((\boldsymbol{a} * \boldsymbol{b}) * \boldsymbol{c}) \cdot \boldsymbol{d} \\ &= ((\boldsymbol{a} * \boldsymbol{b}) * \boldsymbol{d}) \cdot \boldsymbol{c} \\ &= ((\boldsymbol{c} * \boldsymbol{d}) * \boldsymbol{a}) \cdot \boldsymbol{b} \\ &= ((\boldsymbol{c} * \boldsymbol{d}) * \boldsymbol{b}) \cdot \boldsymbol{a}. \end{aligned} \tag{G.2}$$

This result generalizes: the dot product of any combination of star products can be rearranged such that the dot product is just with a single vector at the end of the computation, provided one is careful with the non-associativity of the star product.

When it is not misleading, it can be convenient to adopt the convention that multiplication proceeds from left to right so we can drop the parenthesis and have, for instance, that

$$(((\boldsymbol{a} * \boldsymbol{b}) * \boldsymbol{c}) * \boldsymbol{d}) = \boldsymbol{a} * \boldsymbol{b} * \boldsymbol{c} * \boldsymbol{d}. \tag{G.3}$$

Finally, we introduce the notation that $\boldsymbol{b}^{*k}$ denotes the k-fold star product $\boldsymbol{b} * \boldsymbol{b} * \cdots \boldsymbol{b}$, such that $\boldsymbol{b}^{*1} = \boldsymbol{b}$, $\boldsymbol{b}^{*2} = \boldsymbol{b} * \boldsymbol{b}$ and so on.

G.2 Proof of Sufficient Word Length for Unitary Similarity of a Pair of Hermitian Matrices

Theorem G.2.1. *Two Hermitian matrices $H, H' \in \mathbb{C}^{d\times d}$ are unitarily similar if and only if $\mathrm{Tr}[H^k] = \mathrm{Tr}[H'^k]$ for $k \in [1, d]$.*

Proof. In general, from Theorem 9.3.1, it is necessary to check the trace conditions for all words H^k, H'^k. We show it is only necessary to check the first d such words. That is, $\mathrm{Tr}[H^k] = \mathrm{Tr}[H'^k]$ for $k \in [1, d]$ implies $\mathrm{Tr}[H^k] = \mathrm{Tr}[H'^k]$ for $k > d$.

Observe that traces are basis-independent so we may write the given set of equivalences for $k \in [1, d]$ as

$$\mathrm{Tr}[H^k] = \mathrm{Tr}[H'^k] \implies \sum_{j=1}^{d} \lambda_j^k = \sum_{j=1}^{d} \lambda_j'^k, \tag{G.4}$$

where $\{\lambda_j\}$, $\{\lambda_j'\}$ are the eigenvalues of H, H', respectively. Because H, H' are Hermitian these are real. The sums in Eq. (G.4) are known as the power sums $p_k(\lambda_1, \cdots, \lambda_d)$ and $p_k(\lambda_1', \cdots, \lambda_d')$. Via the Newton-Girard identities, one can explicitly write the first d elementary symmetric polynomials $e_1, \cdots e_d$ in these eigenvalues in terms of the power sums p_k for $k \in [1, d]$. By the equivalence of the power sums between the variables $\{\lambda_j\}$, $\{\lambda_j'\}$, the elementary symmetric polynomials in these two variables are also equivalent.

The elementary symmetric polynomials then allow us to write the chain of equalities via a

standard expansion of a polynomial in some variable x with roots $\{\lambda_j\}$:

$$\begin{aligned}\prod_{j=1}^{d}(x-\lambda_j) &= \sum_{k=0}^{n}(-1)^k e_k(\lambda_1,\cdots,\lambda_n)x^{n-k} \\ &= \sum_{k=0}^{n}(-1)^k e_k(\lambda'_1,\cdots,\lambda'_n)x^{n-k} = \prod_{j=1}^{d}(x-\lambda'_j). \end{aligned} \tag{G.5}$$

This implies that $\lambda_j = \lambda'_j$ for all $j \in [1,d]$. This in turn implies that $\mathrm{Tr}[H^k] = \mathrm{Tr}[H'^k]$ for $k > d$, proving the result. □

We remark that Theorem G.2.1 is often stated without proof in the context of giving the independent trace invariants of density matrices $\rho \in \mathbb{C}^{d\times d}$ [203, 216, 217]. Given that Hermitian matrices have d eigenvalues, the theorem is intuitively obvious, but we have not seen the explicit proof of this statement in the physics literature.

We also provide an alternative statement of the theorem and a corresponding proof which makes use of the Bloch vector formalism for trace invariants.

Theorem G.2.2. *Two traceless Hermitian matrices $H, H' \in \mathbb{C}^{d\times d}$ are unitarily similar if and only if $\boldsymbol{a}^{*k}\cdot\boldsymbol{a} = \boldsymbol{b}^{*k}\cdot\boldsymbol{b}$ for $k \in [1, d-1]$, where $\boldsymbol{a},\boldsymbol{b}$ are the Bloch vectors corresponding to H, H', respectively.*

Proof. Consider the infinite set of vectors $\mathcal{A} = \{\boldsymbol{a}, \boldsymbol{a}*\boldsymbol{a}, \boldsymbol{a}*\boldsymbol{a}*\boldsymbol{a}, \cdots\}$. The invariants of A under unitary transformations are $\boldsymbol{v}\cdot\boldsymbol{a}$ for any $\boldsymbol{v} \in \mathrm{span}(\mathcal{A})$.

Any Hermitian matrix can be diagonalized via a unitary matrix and the invariants are unchanged under this transformation so assume that we have diagonalized A. Suppose we are using a generalized Gell-Mann basis, as described in the main text. Then the corresponding $\boldsymbol{a}$ has $a_j = 0$ for $j \in \mathcal{X}, \mathcal{Y}$. That is, $\boldsymbol{a}$ is in a subspace T of dimension $d-1$ spanned by $\hat{\boldsymbol{\lambda}}^{(\mathrm{diag})}$, with

corresponding indices $\mathcal{D}$. We have that

$$(\boldsymbol{a} * \boldsymbol{a})_k = d_{ijk} a_i a_j, \tag{G.6}$$

where the only non-zero terms in the sum correspond to non-zero d_{ijk} with $i, j \in \mathcal{D}$. The only non-zero d_{ijk} satisfying this condition have $k \in \mathcal{D}$. Therefore, $\boldsymbol{a} * \boldsymbol{a} \in T$. This is true for all $\boldsymbol{v} \in \mathcal{A}$. As T is of dimension of $d-1$, only up to $d-1$ of the vectors in $\mathcal{A}$ are independent. Adding in the traceless condition, this makes for a maximum of d independent invariants. Therefore, to determine the simultaneous similarity of H, H' it is sufficient to check the equivalence of only the d invariants in the theorem statement. □

We remark if H, H' are not traceless one can merely check if the matrices have the same trace and then apply the theorem above.

G.3 Bound on the Length of Words

This section seeks to prove the bound on ℓ_{max}, the maximum length of word we need to check to capture all independent invariants, given in Eq. (9.5). This proof is just an application of the bound and construction provided in Ref. [200].

The idea of simultaneous unitary similarity is derived in Ref. [200], starting from whether two complex matrices are unitarily similar. It is a known result that two $n \times n$ complex matrices, A and B (with A^* and B^* denoting their complex conjugates), are unitarily similar if and only if $\text{Tr}[w(\{A, A^*\})] = \text{Tr}[w(\{B, B^*\})]$ for every word $w(\{s, t\})$ of two non-commuting matrices

whose length is less than or equal to

$$\ell'_{\max} = \min \begin{cases} \lceil \frac{n^2+2}{3} \rceil \\ n\sqrt{\frac{2n^2}{n-1} + \frac{1}{4}} + \frac{n}{2} - 2. \end{cases} \tag{G.7}$$

The $O(n^2)$ bound is due to Paz [212] and the asymptotically better $O(n^{3/2})$ bound is due to Pappacena [213]. This result holds for arbitrary complex matrices, and as we discussed in the preceding section, the bound on considered words can be much tighter if we consider just the unitary similarity of two Hermitian matrices. However, this more general bound is important in the context of the unitary similarity of sets of matrices. In Ref. [200], the authors produce an encoding of sets of matrices into two larger matrices such that if those two larger matrices are unitarily similar then all the individual pairs of matrices from the two sets must be unitarily similar under the same transformation. Furthermore, the word trace condition for unitary similarity of these larger matrices is equivalent to a word trace condition on all words of that same length compared between words made entirely of one of the sets and words made entirely from the other set.

Specifically, given two sets of matrices $S = \{s_1, s_2, \ldots, s_m\}$ and $S' = \{s'_1, s'_2, \ldots, s'_m\}$ where each matrix is of size $d \times d$, we can encode these sets into matrices A and B. A will be a block matrix constructed of $d \times d$ matrices, and the diagonal and all blocks below it are zero. Immediately above the diagonal are $d \times d$ identity matrix blocks, and into the remaining portion of the upper triangular portion, we slot the matrices from S into the blocks. In order to do this, we need m spaces remaining in the upper triangle. If the A matrix is c blocks long, then there will be $(c^2 - 3c + 2)/2$ spaces for matrices from our set S. Thus, we must choose c such that

$(c^2 - 3c + 2)/2 \geq m$. Any unused blocks are set to zero. The B matrix is constructed similarly except using S' instead of S. These matrices will both be of size $n = cd$.

In Ref. [200] it is proven that the matrices A and B are unitarily similar if and only if the sets S and S' are unitarily similar. Furthermore, the trace word conditions on A and B being unitarily similar is equivalent to the condition that all words have equal traces between sets S and S' with the lengths of these necessary words being bounded by the same length as the words necessary to check unitary similarity of A and B. Therefore, it is sufficient to check words up to length based off Eq. (G.7) with $n = cd$, recovering Eq. (9.5).

G.4 Trace Invariants in Terms of Bloch Vectors

In this section, we demonstrate the claim from the main text that all trace invariants under simultaneous unitary transformations can be expressed as linear combinations of invariants $\boldsymbol{v} \cdot \boldsymbol{b}^{(j)}$ for $\boldsymbol{v} \in \mathcal{B}$ and $\boldsymbol{b}^{(j)}$ a Bloch vector corresponding to the Hamiltonian H_j. Recall, we define $\mathcal{B}$ to be the set of all possible star products between Bloch vectors in B.

Furthermore, recall that for a trace invariant for a word of arbitrary length we have

$$\mathrm{Tr}[w(S)] = \mathrm{Tr}\left[\prod_{j=1}^{|w|} \sum_{\mu_j=1}^{d^2-1} b_{\mu_j}^{(w_j)} \hat{\lambda}_{\mu_j}\right], \tag{G.8}$$

where we have denoted the j-th element of w as w_j. We can flip the order of the product and the sum and write

$$\mathrm{Tr}[w(S)] = \sum_{\{\mu_1 \cdots \mu_{|w|}\}} \mathrm{Tr}\left[\prod_{j=1}^{|w|} b_{\mu_j}^{(w_j)} \hat{\lambda}_{\mu_j}\right], \tag{G.9}$$

where the sum is over all ordered sets (with replacement) of indices $\in [1, d^2 - 1]$. We now

make use of Eq. (9.2) to evaluate the products of basis elements of $\mathfrak{su}(d)$. Due the sum over all ordered sets the antisymmetric terms in each product of $\hat{\lambda}_{\mu_j}$ cancel and we are left to consider the identity terms and the terms with symmetric structure constants. As the basis elements are all traceless, what ultimately survives the trace once we fully expand out all products are the terms proportional to identity.

Under an expansion of the products and evalaution of the trace the term with the most symmetric structure constants is proportional to

$$\begin{aligned}(d_{\mu_1\mu_2\nu_1}d_{\mu_3\nu_1\nu_2}d_{\mu_4\nu_2\nu_3}\cdots d_{\mu_{|w|-1}\nu_{|w|-3}\nu_{|w|-2}}\delta_{\nu_{|w|-2}\mu_{|w|}})\\ \times (b_{\mu_1}^{(w_j)}\cdots b_{\mu_{|w|}}^{(w_{|w|})}),\end{aligned} \tag{G.10}$$

where we use the convention of summing over repeated indices. After staring at the proliferation of indices, one observes that this term can be compactly written as

$$(((\boldsymbol{b}^{(w_1)} * \boldsymbol{b}^{(w_2)}) * \boldsymbol{b}^{(w_3)}) * \cdots * \boldsymbol{b}^{(w_{|w|-1})}) \cdot \boldsymbol{b}^{(w_{|w|})} \tag{G.11}$$

All other terms in the expansion of the products of basis elements of $\mathfrak{su}(d)$ consist of fewer symmetric structure constants and more Kronecker deltas. These other terms amount to dot products between terms similar to this one but with smaller word length. Therefore, this term is the only one that is not dependent on invariants established from smaller length words. By the commutativity of the star product we may permute any products we like, provided we respect the lack of associativity. From this, and the fact that all words yield trace invariants, we establish the intended claim.

This result implies a nice geometric interpretation of the trace invariants. In particular, the relative angles between all vectors in $\mathcal{B}$ are the invariants under unitary transformations. This is a manifestation of the fact that $SU(d) \subset SO(d^2-1)$. Due to the asymmetry in Bloch space of the symmetric structure constants, the set of star products $\mathcal{B}$ are not, in general, rotationally invariant—the rotations where this infinite set of vectors *do* rotate rigidly picks out the rotations corresponding to $SU(d)$.

G.5 Simultaneous Stoquasticizability is Rare

For all the proofs in this section it will be necessary to identify the non-zero symmetric structure constants of $\mathfrak{su}(d)$ in the generalized Gell-Mann basis. We take the explicit form of these from Ref. [221], with some slight differences in indexing to account for our conventions differing from those used by those authors.[1] We identify these symmetric structure constants based on whether the indices correspond to symmetric (Eq. 9.3a), skew-symmetric (Eq. 9.3b), or diagonal (Eq. 9.3c) basis elements. In particular, we give a one-to-one mapping between indices $i \in [1, d^2-1]$ and indices $\mathcal{X}_{jk}, \mathcal{Y}_{jk}$ and $\mathcal{D}_j$ corresponding to the sets of symmetric,

[1] In particular, relative to the conventions of Ref. [221], we do the following: (a) we index our diagonal basis elements of the generalized Gell-Mann basis $[1, d-1]$ as opposed to to $[2, d]$; (b) we index the symmetric and anti-symmetric basis elements in increasing, rather than decreasing order. For instance, for $\hat{\lambda}_{ij}^{(x)}$ we have $i < j$, not $i > j$; (c) our basis elements are a factor of two larger and therefore our trace orthonormality condition is four times larger. That is, they have $\mathrm{Tr}(\hat{\lambda}_i \hat{\lambda}_j) = \delta_{ij}/2$, whereas we have $\mathrm{Tr}(\hat{\lambda}_i \hat{\lambda}_j) = 2\delta_{ij}$. Despite this, the normalization of the structure constants agree

skew-symmetric and diagonal basis element, respectively, as follows:

$$\mathcal{X}_{jk} = k^2 + 2(j-k) - 1 \tag{G.12}$$

$$\mathcal{Y}_{jk} = k^2 + 2(j-k) \tag{G.13}$$

$$\mathcal{D}_j = j(j+2), \tag{G.14}$$

where $1 \leq j < k \leq d$. Let $\mathcal{X} = \{\mathcal{X}_{jk}\}$, $\mathcal{Y} = \{\mathcal{Y}_{jk}\}$, and $\mathcal{D} = \{\mathcal{D}_j\}$. Given such an identification, we have the following non-zero symmetric structure constants:

$$\begin{aligned}
&d_{\mathcal{X}_{jk}\mathcal{X}_{lj}\mathcal{X}_{lk}} = d_{\mathcal{X}_{jk}\mathcal{Y}_{lj}\mathcal{Y}_{lk}} = d_{\mathcal{X}_{jk}\mathcal{Y}_{kl}\mathcal{Y}_{jl}} = -d_{\mathcal{X}_{jk}\mathcal{Y}_{jl}\mathcal{Y}_{lk}} = \frac{1}{2} \\
&d_{\mathcal{X}_{jk}\mathcal{X}_{jk}\mathcal{D}_{j-1}} = d_{\mathcal{Y}_{jk}\mathcal{Y}_{jk}\mathcal{D}_{j-1}} = -\sqrt{\frac{j-1}{2j}} \\
&d_{\mathcal{X}_{jk}\mathcal{X}_{jk}\mathcal{D}_{l-1}} = d_{\mathcal{Y}_{jk}\mathcal{Y}_{jk}\mathcal{D}_{l-1}} = \sqrt{\frac{1}{2l(l-1)}}, \quad j < l < k \\
&d_{\mathcal{X}_{jk}\mathcal{X}_{jk}\mathcal{D}_{k-1}} = d_{\mathcal{Y}_{jk}\mathcal{Y}_{jk}\mathcal{D}_{k-1}} = \frac{2-k}{\sqrt{2k(k-1)}} \\
&d_{\mathcal{X}_{jk}\mathcal{X}_{jk}\mathcal{D}_{l-1}} = d_{\mathcal{Y}_{jk}\mathcal{Y}_{jk}\mathcal{D}_{l-1}} = \sqrt{\frac{2}{l(l-1)}}, \quad k < l \\
&d_{\mathcal{D}_{j-1}\mathcal{D}_{k-1}\mathcal{D}_{k-1}} = \sqrt{\frac{2}{j(j-1)}}, \quad k < j \\
&d_{\mathcal{D}_{j-1}\mathcal{D}_{j-1}\mathcal{D}_{j-1}} = (2-j)\sqrt{\frac{2}{j(j-1)}}
\end{aligned} \tag{G.15}$$

We shall find for the following proofs that it is sufficient to observe which symmetric structure constants are non-zero, but for explicit application of these theorems these analytic expressions would be convenient.

As implied by the structure constants being considered, we will make use of the generalized

Gell-Mann basis throughout these proofs, but we observe that all trace invariants are basis independent. Consequently, the dimensions of spaces spanned by the star products of Bloch vectors are also basis independent quantities.

We begin with a simple theorem describing the maximum size of the set of star products arising from simultaneously stoquastic matrices.

Theorem G.5.1 (Theorem 4 from the main text)**.** *Let $S = \{H_1, \cdots, H_m\}$ be a set of Hermitian matrices with corresponding Bloch vectors $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots \boldsymbol{b}_m\}$. Let $\mathcal{B}$ be the set of all possible star products between elements of B. A necessary condition for the elements of S to be simultaneously stoquastizable is that* $\dim(\mathrm{span}(\mathcal{B}))) \leq (d^2 + d - 1)/2$.

Proof. Suppose $H \in \texttt{Stoq}$ for all $H \in S$. Observe from the list of non-zero symmetric structure constants in Eq. (G.15) that any star products between elements of B necessarily have all components $j \in \mathcal{Y}$ equal to zero. That is, any star product between vectors in the $\mathcal{X}, \mathcal{D}$ subspace remain in that subspace. Therefore, $\dim(\mathrm{span}(\mathcal{B})) \leq (d^2+d-1)/2$, where $(d^2+d-1)/2$ is the dimension of this subspace. The dimension of this subspace is preserved under unitary transformations as the relative angles between all elements of $\mathcal{B}$ are preserved under unitary transformations. This establishes the necessary condition for simultaneous stoquasticity in the theorem statement. □

From here, our goal will be to look at the space of star products arising from general Bloch vectors and show that this space is generally much larger than the space spanned by simultaneously stoquastic Bloch vectors. We begin with a useful lemma.

Lemma G.5.1. *Let H be a traceless Hermitian matrix with corresponding Bloch vector $\boldsymbol{b}$. Let $\mathcal{B} = \{\boldsymbol{b}^{*k} \mid k \in \mathbb{Z}^+\}$ be the (infinite) set of all possible star products of $\boldsymbol{b}$. Then* $\dim(\mathrm{span}(\mathcal{B})) \leq d - 1$

Proof. As $\dim(\mathrm{span}(\mathcal{B}))$ is a basis-independent property, assume that H is diagonal without loss of generality. Therefore, the generalized Gell-Mann basis $\boldsymbol{b}_j = 0$ for all $j \in \{\mathcal{X}, \mathcal{Y}\}$. From the definition of the star product and the form of the structure constants in Eq. (G.15) observe that the only non-zero components of the star $\boldsymbol{b} * \boldsymbol{b}$ are those with index $j \in \mathcal{D}$. This holds for the star product of any pair of vectors $\boldsymbol{u}, \boldsymbol{v}$ with $u_j, v_j = 0$ for all $j \in \{\mathcal{X}, \mathcal{Y}\}$. Therefore, all $\boldsymbol{v} \in \mathcal{B}$ are necessarily contained in the $d-1$ dimensional subspace with indices $\mathcal{D}$. □

If H in the above lemma is *generic*—i.e. has non-degenerate eigenvalues—$\dim(\mathrm{span}(\mathcal{B})) = d - 1$. For non-generic Hamiltonians, there exist additional symmetries which leads to dependence between the elements of $\mathcal{B}$ and, consequently, a reduction in the dimension of $\mathrm{span}(\mathcal{B})$. For instance, in the extreme case of H corresponding to a pure state density matrix ρ—which has zero as a $(d-1)$-fold degenerate eigenvalue—we have that $\boldsymbol{b} * \boldsymbol{b} = \boldsymbol{b}$, implying that $\dim(\mathrm{span}(\mathcal{B})) = 1$ [204].

We will also make use of the following lemma, proving a linear algebra fact that will be useful later on.

Lemma G.5.2. *Let $R \in \mathbb{R}^{2n \times 2n}$ be a diagonal matrix with all non-zero, 2-fold degenerate matrix elements such that $R_{jj} = R_{kk}$ for $k = 2j$ and $R_{jj} \neq R_{kk}$ otherwise. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{2n}$ be vectors with all unique elements such that $u_{j^*} = 0$, $v_{2j^*} = 0$ for some particular j^*. Then the vectors in the set $\{R^k \boldsymbol{u}, R^k \boldsymbol{v}\}$ for $k \in \mathbb{Z}^+$ span $\mathbb{R}^{2n}$.*

Proof. With the exception of the zero vector, $\mathrm{span}(\{R^k \boldsymbol{u}\})$ is completely disjoint from $\mathrm{span}(\{R^k \boldsymbol{v}\})$ from the uniqueness conditions on $R, \boldsymbol{u}, \boldsymbol{v}$ and the fact that $u_{j^*} = 0, v_{2j^*} = 0$. In particular, there exists no non-trivial $q_k, r_k \in \mathbb{R}$ such that $\sum_k q_k R^k \boldsymbol{u} = \sum_k r_k R^k \boldsymbol{v} \implies \sum_k q_k \boldsymbol{u} = \sum_k r_k \boldsymbol{v}$, as this would require $\sum_k q_k u_{2j^*} = 0$ and $\sum_k r_k v_{j^*} = 0$, implying $\sum_k q_k = \sum_k r_k = 0$.

Now consider the span of $\{R^k \boldsymbol{u}\}$. As the components of $\boldsymbol{u}$ are unique and R has n unique components, these vectors will span a space of dimension n. This is because for $R^k \boldsymbol{u}$ to be linearily dependent on $\{R^l \boldsymbol{u}\}$ for $l < k$ requires that there not exist constants $c_l \in \mathbb{R}$ such $R^k \boldsymbol{u} = \sum_{l=1}^{k-1} c_l R^l \boldsymbol{u}$, which, by the uniqueness of the components of $\boldsymbol{u}$ implies we require $R^k = \sum_{l=1}^{k-1} c_l R^l$ for dependence. Such constants only exist for $k \geq n$ due to the uniqueness conditions on R.

An identical argument holds for the span of $\{R^k \boldsymbol{u}\}$. As the two spans are completely disjoint, together they span the full vector space of dimension $2n$. □

Armed with the preceding two lemmas, we are now prepared to prove the following important theorem.. Here, "almost every" is used in a technical sense, in that the set of possibilities not obeying the given condition are (Lebesgue) measure zero.

Theorem G.5.2. *Let $S = \{H_1, \cdots, H_m\}$ be a set of Hermitian matrices with corresponding Bloch vectors $B = \{\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \cdots \boldsymbol{b}^{(m)}\}$. Let $\mathcal{B}$ be the (infinite) set of all possible star products between elements of B. For almost every S with $m \geq 2, d \geq 3$, $\dim(\mathrm{span}(\mathcal{B})) = d^2 - 1$. That is, $\mathcal{B}$ spans the full Bloch vector space for almost every S.*

Proof. It is sufficient to consider $m = 2$ as the dimension of the space spanned by this subset is less than or equal to that of the full set $\mathcal{B}$. Without loss of generality, assume that H_1 is diagonal so that in the generalized Gell-Mann basis $\boldsymbol{b}_j^{(1)} = 0$ for all $j \in \{\mathcal{X}, \mathcal{Y}\}$. By Lemma G.5.1 and the discussion that follows it, the star products $(\boldsymbol{b}^{(1)})^{*k}$ for $k \in \mathbb{Z}^+$ span the $d-1$ dimensional space corresponding to indices $j \in \mathcal{D}$ for any H_1 with non-degenerate eigenvalues. Hermitian matrices with degenerate eigenvalues are measure zero in the space of traceless, Hermitian matrices[2] and,

[2] In particular, Hermitian matrices with repeated eigenvalues have codimension three (c.f. [490])

therefore, almost every H_1 is such that $(\boldsymbol{b}^{(1)})^{*k}$ for $k \in \mathbb{Z}^+$ span the full $d-1$ dimensional space corresponding to indices $j \in \mathcal{D}$.

As these products of the form $(\boldsymbol{b}^{(1)})^{*k}$ span the $d-1$ dimensional space corresponding to indices $j \in \mathcal{D}$, we now seek to show that other elements of $\mathcal{B}$ span the remaining $\mathcal{X}, \mathcal{Y}$ components for almost every $\boldsymbol{b}^{(2)}$. To this end, consider the vectors $\boldsymbol{b}^{(2)}$ and $\boldsymbol{b}^{(2)} * \boldsymbol{b}^{(2)}$. Then, consider just the components of these vectors in the $\mathcal{X}, \mathcal{Y}$ subspace. Call these restricted vectors $\boldsymbol{u}, \boldsymbol{v}$. That is, $u_j = 0$ for $j \in \mathcal{D}$ and $u_j = \boldsymbol{b}^{(2)}_j$ for $j \in \{\mathcal{X}, \mathcal{Y}\}$, and similarly for $\boldsymbol{v}$. For almost all $\boldsymbol{b}^{(2)}$, the corresponding $\boldsymbol{u}, \boldsymbol{v}$ are linearily independent. Therefore, we can construct via linear combinations two new vectors $\boldsymbol{u}', \boldsymbol{v}'$ such that $u'_{\mathcal{X}_{12}} = 0, v'_{\mathcal{Y}_{12}} = 0$. Such linear combinations are, by definition, in the span of $\mathcal{B}$.

Now, consider acting from the right on this linear combinations by star products of the form $(\boldsymbol{b}^{(1)})^{*k}$—i.e. consider elements in the span of $\mathcal{B}$ of the form $((\boldsymbol{u}' * \boldsymbol{b}^{(1)}) * \boldsymbol{b}^{(1)}) * \cdots$. From the definition of the star product, the fact that $\boldsymbol{b}^{(1)}_j = 0$ for all $j \in \{\mathcal{X}, \mathcal{Y}\}$, and the form of the structure constants in Eq. (G.15) we observe that such star products by $(\boldsymbol{b}^{(1)})^{*k}$ (from the right) act to scale the $\mathcal{X}, \mathcal{Y}$ components of $\boldsymbol{u}'$, $\boldsymbol{v}'$ by symmetric structure constant-dependent factors. We can write this scaling behavior as

$$\boldsymbol{u}^{(k)} = R^k \boldsymbol{u}', \tag{G.16}$$

where R is a diagonal matrix with components given by

$$R_{jj} = \begin{cases} \sum_{i \in \mathcal{D}} d_{jji} b^{(1)}_i, & j \in \{\mathcal{X}, \mathcal{Y}\} \\ 0, & j \in \mathcal{D} \end{cases} \tag{G.17}$$

which comes from identifying the non-zero terms in the corresponding star product. There is an identical equation to Eq. (G.16) for $\boldsymbol{v}'$. Importantly, we observe that $R_{jj} = R_{kk}$ for $j = \mathcal{X}_{lm}$, $k = \mathcal{Y}_{lm}$. This can be determined by observing in Eq. (G.15) that $d_{\mathcal{X}_{lm}\mathcal{X}_{lm}j} = d_{\mathcal{Y}_{lm}\mathcal{Y}_{lm}j}$ for all $j \in \mathcal{D}$. Otherwise, for almost every H_1, $R_{jj} \neq R_{kk}$ for $j, k \in \{\mathcal{X}, \mathcal{Y}\}$. From Lemma G.5.2, $\{\boldsymbol{u}'^{(k)}, \boldsymbol{v}'^{(k)}\}$ span the $\mathcal{X}, \mathcal{Y}$ subspace of Bloch vector space for almost all H_1, H_2. □

Finally, we arrive at the following theorem which establishes that simultaneous stoquasticity is rare.

Theorem G.5.3 (Theorem 3 from main text)**.** *Let $S = \{H_1, \cdots, H_m\}$ be a set of Hermitian matrices. For almost every S with $m \geq 2$, $d \geq 3$, S is not simultaneously stoquasticizable.*

Proof. This is an immediate consequence of Theorem G.5.1 and Theorem G.5.2. □

G.6 Results on Simultaneous Diagonalizability

Similar results to those in the previous section can also be obtained for the problem of determining the simultaneous diagonalizability of a set of Hermitian matrices. This problem is, of course, well-known to be related to the problem of mutual compatibility of observables. Therefore, while taking an approach similar to that in our paper for this problem is largely overcomplicated compared to applying the simple condition that a set of Hermitian matrices are simultaneously diagonalizable if and only if they all commute, it is useful to compare the formalism established in this work to such conditions.

In particular, as discussed in the main text, we expect that a deeper understanding of how our conditions relate to commutator conditions will enable connections to the dynamical Lie

algebra of quantum optimal control theory. The case of simultaneous diagonalizability, with its well-known commutation condition provides a possible route forward.

We have the following theorem, which is analogous to Theorem G.5.1 (Theorem 9.5.2 from the main text). The logic is also similar to that of Lemma G.5.1, but extended to multiple matrices.

Theorem G.6.1. *Let $S = \{H_1, \cdots, H_m\}$ be a set of Hermitian matrices with corresponding Bloch vectors $B = \{\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \cdots \boldsymbol{b}^{(m)}\}$. Let $\mathcal{B}$ be the (infinite) set of all possible star products between elements of B . A necessary condition for the elements of S to be simultaneously diagonalizable is that* $\dim(\mathrm{span}(\mathcal{B})) \leq d - 1$.

Proof. The proof is identical to that of Theorem G.5.1, except, in this case, the star products of Bloch vectors corresponding to simultaneously diagonal Hermitian matrices are confined to the $d - 1$ dimensional subspace of Bloch vector space with all components $j \in \mathcal{X}, \mathcal{Y}$ equal to zero. Therefore, a set of simultaneously diagonal Hermitian matrices has $\dim(\mathrm{span}(\mathcal{B})) \leq d-1$. Again, the dimension of this subspace is preserved under unitary transformations, proving the result. □

Combined with Theorem G.5.2 we have the following corollary as an immediate consequence.

Corollary G.6.1.1. *Let $S = \{H_1, \cdots, H_m\}$ be a set of Hermitian matrices. For almost every S with $m \geq 2$, $d \geq 3$, S is not simultaneously diagonalizable.*

One also expects that amongst simultaneously stoquastic Hamiltonians simultaneously diagonalizable Hamiltonians are vanishingly rare. Proving this would follow a similar line of reasoning to that in the previous section. In particular, it would be sufficient to prove that for almost

every set of stoquastic Hamiltonians the inequality in Theorem G.5.1 is tight. Unfortunately, the approach in Theorem G.5.2 doesn't immediately apply here since we can't diagonalize one of the matrices in a set of stoquastic Hamiltonians and keep all Hamiltonians stoquastic.

Appendix H: Technical Details of the Results Reported in Chapter 12

H.1 Equivalence of Error Metrics

Our choice of error metric is the average difference between matrix elements of $\mathcal{B}^s$—[see] Eq. (12.4)—calculated from an ensemble of unitaries generated by random circuits and those obtained from an ensemble of unitaries that form an exact 2-design, as given in Eq. (12.5). This difference is normalized by a factor of $d_s(d_s^2 - 1)$ where d_s is the dimension of the relevant symmetry block. Because of the otherwise prohibitive cost, in practice, for large N we compute the error by only averaging over *a sample* of all matrix indices of $\mathcal{B}^s$, where we increase the number of indices sampled until we see convergence. Mathematically,

$$\epsilon = \frac{d_s(d_s^2-1)}{|S|} \sum_S \left| (\mathcal{B}^s_{\mathcal{E}})^{i'j'k'l'}_{ijkl} - (\mathcal{B}^s_{\text{2-des.}})^{i'j'k'l'}_{ijkl} \right|, \tag{H.1}$$

where S is the set of indices $i, j, k, l, i', j', k', l'$ sampled and the subscripts on $\mathcal{B}^s$ denote the ensemble with which the expectation values are taken with respect to in the definition of $\mathcal{B}^s$.

This choice of quantifying the error between our random ensemble of circuits and an exact unitary 2-design was chosen for two reasons: (1) it enables a computationally tractable approach

wherein we sample a subset of matrix elements until we see convergence; (2) once normalized (by a factor of $d_s(d_s^2-1)$), it is manifestly dimension independent once our random ensemble of circuits has converged to an approximate unitary 2-design. Despite these benefits, it is not the only reasonable choice of error metric, nor is it a standard one. Therefore, in this Appendix, we demonstrate analytically that it is equivalent to more typical definitions, up to a rescaling of the approximation ratio for the case where S is the set of all index combinations. We then numerically demonstrate that sampling (the non-zero) matrix elements is also a valid (and computationally accessible) choice of error metric.

A particularly common definition of an approximate unitary k-design states that an ensemble $\mathcal{E}$ of unitaries forms a δ-approximate unitary k-design if and only if [491]

$$\sqrt{|\mathcal{F}_{\mathcal{E}}^{(k)} - \mathcal{F}_{\text{Haar}}^{(k)}|} \leq \frac{\delta}{d_s^k}, \tag{H.2}$$

where $\mathcal{F}^{(k)}$ is the k-th frame potential for an ensemble of unitaries, defined as

$$\mathcal{F}_{\mathcal{E}}^{(k)} = \int_{U,V\in\mathcal{E}} dU dV \left|\text{Tr}(U^\dagger V)\right|^{2k}. \tag{H.3}$$

For the Haar ensemble $\mathcal{F}_{\text{Haar}}^{(k)} = k!$ for $k \leq d_s$ [492].

Ultimately, we will numerically compare our error metric to the frame potential definition, but for the purposes of analytically showing equivalence of our error metric it is convenient to use a different formulation known to be equivalent to this one up to a rescaling of the approximation factor [493]. In particular, an ensemble $\mathcal{E}$ of unitaries forms a δ-approximate k-design if and only

if for all balanced monomials $M(U)$ of degree $\leq k$ in the matrix elements of U

$$\left|\langle M(U)\rangle_{\mathcal{E}} - \langle M(U)\rangle_k\right| \leq \frac{\delta}{d_s^k}, \tag{H.4}$$

where the subscript k denotes an expectation value with respect to an exact k-design. A balanced monomial of degree k in the matrix elements of U is defined as any product in the matrix elements of the form

$$M = U_{i_1 j_1} ... U_{i_k j_k} U^*_{k_1 l_1} ... U^*_{k_k l_k}, \tag{H.5}$$

for some choice of indices.

To show the equivalence between Eq. (H.1) and Eq. (H.4) we return explicitly to the case of $k = 2$. An ensemble $\mathcal{E}$ that forms a δ-approximate 2-design necessarily also forms a δ-approximate 1-design, so in Eq. (H.4) we can restrict our attention to balanced monomials of degree 2. Consequently, for unitaries U^s acting on a symmetry sector s, and a general degree 2 balanced monomial $M = U^s_{ij} U^{s*}_{i'j'} U^s_{kl} U^{s*}_{k'l'}$, we can write

$$\begin{aligned}
&\left|(\mathcal{B}^s_{\mathcal{E}})^{i'j'k'l'}_{ijkl} - (\mathcal{B}^s_2)^{i'j'k'l'}_{ijkl}\right| = \\
&\Bigg|\langle M^{i'j'k'l'}_{ijkl}\rangle_{\mathcal{E}} - \langle M^{i'j'k'l'}_{ijkl}\rangle_2 \\
&\qquad + \frac{d_s^2}{d_s^2-1}\Big[(\mathcal{A}^s_2)^{i'j'}_{ij}(\mathcal{A}^s_2)^{k'l'}_{kl} + (\mathcal{A}^s_2)^{k'l'}_{ij}(\mathcal{A}^s_2)^{i'j'}_{kl} - (\mathcal{A}^s_{\mathcal{E}})^{i'j'}_{ij}(\mathcal{A}^s_{\mathcal{E}})^{k'l'}_{kl} - (\mathcal{A}^s_{\mathcal{E}})^{k'l'}_{ij}(\mathcal{A}^s_{\mathcal{E}})^{i'j'}_{kl}\Big]\Bigg| \\
&\leq \underbrace{\left|\langle M^{i'j'k'l'}_{ijkl}\rangle_{\mathcal{E}} - \langle M^{i'j'k'l'}_{ijkl}\rangle_2\right|}_{\text{I}} \\
&\qquad + \frac{d_s^2}{d_s^2-1}\Bigg[\underbrace{\left|(\mathcal{A}^s_2)^{i'j'}_{ij}(\mathcal{A}^s_2)^{k'l'}_{kl} - (\mathcal{A}^s_{\mathcal{E}})^{i'j'}_{ij}(\mathcal{A}^s_{\mathcal{E}})^{k'l'}_{kl}\right|}_{\text{II}} + \underbrace{\left|(\mathcal{A}^s_2)^{k'l'}_{ij}(\mathcal{A}^s_2)^{i'j'}_{kl} - (\mathcal{A}^s_{\mathcal{E}})^{k'l'}_{ij}(\mathcal{A}^s_{\mathcal{E}})^{i'j'}_{kl}\right|}_{\text{II}^*}\Bigg]
\end{aligned} \tag{H.6}$$

where the matrix elements $(\mathcal{B}^s_2)^{i'j'k'l'}_{ijkl}$ are the matrix elements for an exact unitary 2-design, as

defined in Eq. (12.5) and, as noted in the main text, $(\mathcal{A}^s)_{ij}^{kl} \equiv \langle U_{ij}^s U_{kl}^{s*} \rangle$. The subscript 2 denotes an average taken with respect to an exact unitary 2-design and the subscript $\mathcal{E}$ denotes an average taken with respect to the corresponding ensemble of unitaries.

For a δ-approximate unitary k-design the term labeled I is bounded via Eq. (H.4). The terms labeled II and II* are equivalent up to the choice of indices. Recalling that for an exact 1-design (and thus also for a 2-design) $(\mathcal{A}_2^s)_{ij}^{kl} = (\mathcal{A}_1^s)_{ij}^{kl} = \frac{\delta_{ij}\delta_{kl}}{d_s}$ and applying Eq. (H.4) for an approximate unitary 1-design, we can bound these terms as

$$\begin{aligned}
\text{II} &= \Big| (\mathcal{A}_2^s)_{ij}^{i'j'} \Big((\mathcal{A}_{\mathcal{E}}^s)_{kl}^{k'l'} - (\mathcal{A}_2^s)_{kl}^{k'l'} \Big) + (\mathcal{A}_2^s)_{kl}^{k'l'} \Big((\mathcal{A}_{\mathcal{E}}^s)_{ij}^{i'j'} - (\mathcal{A}_2^s)_{ij}^{i'j'} \Big) \\
&\qquad + \Big((\mathcal{A}_{\mathcal{E}}^s)_{ij}^{i'j'} - (\mathcal{A}_2^s)_{ij}^{i'j'} \Big) \Big((\mathcal{A}_{\mathcal{E}}^s)_{kl}^{k'l'} - (\mathcal{A}_2^s)_{kl}^{k'l'} \Big) \Big| \\
&\leq \Big| (\mathcal{A}_2^s)_{ij}^{i'j'} \Big((\mathcal{A}_{\mathcal{E}}^s)_{kl}^{k'l'} - (\mathcal{A}_2^s)_{kl}^{k'l'} \Big) \Big| + \Big| (\mathcal{A}_2^s)_{kl}^{k'l'} \Big((\mathcal{A}_{\mathcal{E}}^s)_{ij}^{i'j'} - (\mathcal{A}_2^s)_{ij}^{i'j'} \Big) \Big| \\
&\qquad + \Big| \Big((\mathcal{A}_{\mathcal{E}}^s)_{ij}^{i'j'} - (\mathcal{A}_2^s)_{ij}^{i'j'} \Big) \Big((\mathcal{A}_{\mathcal{E}}^s)_{kl}^{k'l'} - (\mathcal{A}_2^s)_{kl}^{k'l'} \Big) \Big| \\
&\leq \frac{\delta}{d_s^2} + \frac{\delta}{d_s^2} + \frac{\delta^2}{d_s^2}.
\end{aligned} \tag{H.7}$$

Plugging back into Eq. (H.6) we obtain

$$\begin{aligned}
\big| (\mathcal{B}_{\mathcal{E}}^s)_{ijkl}^{i'j'k'l'} - (\mathcal{B}_2^s)_{ijkl}^{i'j'k'l'} \big| &\leq \frac{\delta}{d_s^2} + \frac{2d_s^2}{d_s^2 - 1} \left(\frac{2\delta + \delta^2}{d_s^2} \right) \\
&= \frac{\delta}{d_s^2} \left(\frac{d_s^2(5 + 2\delta) - 1}{d_s^2 - 1} \right),
\end{aligned} \tag{H.8}$$

for any choice of indices. It is straightforward to go through the converse of this argument and show that if $|\mathcal{B}_{\mathcal{E}} - \mathcal{B}_2| < \frac{\delta}{d_s^2}$ and $|\mathcal{A}_{\mathcal{E}} - \mathcal{A}_2| < \frac{\delta}{d_s}$ (indices suppressed), then the approximate two-design property for $\langle M \rangle_{\mathcal{E}}$ holds, up to a rescaling of δ. Consequently, up to an inconsequential rescaling of the approximation factor, any norm on the difference of the tensors $(\mathcal{B}_{\mathcal{E}}^s) - (\mathcal{B}_2^s)$

is a valid choice for defining an approximate unitary 2-design consistent with the definition in Eq. (H.4).

Our error metric, if all indices are sampled, is just such a norm (namely, an element-wise 1-norm, up to dimension-dependent factors). For the sake of computational tractability, however, we randomly sample a collection of the non-zero matrix elements of $\mathcal{B}^s$ until we see convergence. If the sampled average has converged statistically, i.e. has small error bars, this is a good indication that the sampled average accurately reflects the total average, since the total is simply a sum of sample averages. Since the matrix elements of $\mathcal{B}_2^s$ can take only a handful of values and are all either zero or of order d_s^{-3}, we expect there are no "outlier" matrix elements whose errors are systematically larger than others. We have explicitly verified that the sampling scheme converges to the results obtained by evaluating all index combinations, including those where Eq. (12.5) yields zero. In all our numerics, we ensure that the number of samples taken is such that this convergence occurs.

Numerical results demonstrating the equivalence of our error metric when using finite samples of the non-zero matrix elements of $\mathcal{B}^s$, as given in Eq. (H.1), to the error metric in terms of the frame potential in Eq. (H.2) are shown in Fig. H.1. Here, we consider a system of $N = 10$ sites with particle number symmetry as presented in Sect. 12.2 of the main text. For this comparison, we compare the two error metrics as a function of the number of unitaries $N_{\mathcal{E}}$ drawn directly from CUE within each symmetry sector. In the next section, we elaborate on how we demonstrated that the random circuits we describe in the main text do, in fact, converge to sampling from the CUE for a sufficient number of layers. However, temporarily leaving this aspect aside allows us to purely compare the two error metrics. We find that the two error metrics are equivalent up to a rescaling of the approximation ratio by a factor $\sim d_s$ for a sufficiently large ensemble of random

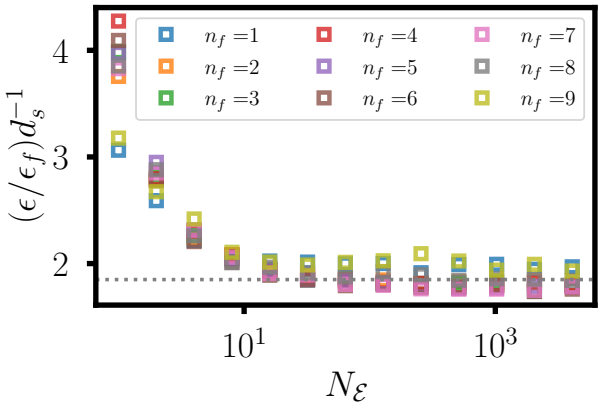


Figure H.1: Ratio of our error metric ϵ [given in Eq. (H.1)] to the more standard frame potential error metric ϵ_f [given in Eq. (H.2)], for an approximate unitary 2-design per symmetry sector (labeled by particle number n_f), normalized by d_s, versus number of unitaries $N_\mathcal{E}$ sampled, indicating that the two error metrics are equivalent up to a rescaling of the approximation ratio for sufficiently large $N_\mathcal{E}$. Results are for a system of $N = 10$ sites with particle-number symmetry as described in Sect. 12.2. For both error metrics, we sample directly from the CUE within each symmetry sector. Our error metric is computed by averaging over $|S| = 2000$ non-zero matrix elements.

unitaries. This is consistent with, but tighter, than our analytic results. Also consistent with our analytics, the numerics indicate sub-leading d_s-dependent factors in the rescaling of the error.

H.2 Particle Number Analysis

In this Appendix, we provide additional details for the particle number symmetry analysis, presented in Sect. 12.2 of the main text. In particular, we present the numerical evidence that sufficiently deep random symmetry-respecting circuits of the sort we describe there form an approximate 2-design and provide details on the numerics leading to Fig. 12.3.

For both of these purposes, we compute the error ϵ as defined in Eq. (H.1) for ensembles $\mathcal{E}$ of $N_\mathcal{E}$ random unitaries generated both by our random circuits and by drawing directly from the CUE. Note that in the latter case, the error is purely due to the fact that these ensembles have only a finite number of elements $N_\mathcal{E}$, as for $N_\mathcal{E} \to \infty$ the ensemble will be a 2-design by construction.

Consequently, as can be seen in Fig. H.2a, the error ϵ scales as $\sim 1/\sqrt{N_{\mathcal{E}}}$.

For a sufficient number of layers, the ensembles generated by our random circuits have identical error ϵ as the ensembles drawn directly from CUE, indicating that these circuits indeed generate samples that form an approximate 2-design. This is demonstrated in Fig. H.2b, where we plot the difference between the error ϵ obtained by taking $N_{\mathcal{E}} = 8192$ samples from the random circuits and taking $N_{\mathcal{E}} = 8192$ samples directly from the CUE versus number of layers used in the circuits. Within a short depth $\ell \approx 15$, the difference reaches a floor set by $N_{\mathcal{E}}$ (i.e. the standard $1/\sqrt{N_{\mathcal{E}}}$ sampling error), indicating that at this depth, drawing samples from our circuits is equivalent to drawing samples directly from CUE. All of the remaining error in approximating a 2-design is purely due to the finite number of samples $N_{\mathcal{E}}$.

While the saturation point (in terms of number of layers required) in Fig. H.2(b) shows a slight dependence on the dimension d_s and similar numerics for smaller system sizes shows a dependence on the number of sites N, the fact that our numerics are limited to around $N \leq 10$ sites prevents us from extracting an asymptotic scaling form of these dependencies. For the pragmatic approach taken in this work—namely, demonstrating that our approach is a viable one for extracting quantities of interest for specific systems with symmetries of interest, such as $\mathbb{Z}_2$ LGT, for small to moderate system sizes—these numerics are sufficient. However, a detailed analytic analysis of the scaling of the error with the number of layers for the particular models considered here (or, perhaps, generally for random symmetry-respecting circuits), remains a compelling prospect for future work. Such analyses have been done for symmetry-ignorant designs [494] and similar approaches should apply here [478].

H.2.1 Details for Figure 3(a)

Fig. H.2 demonstrates that $\ell = 128$ is well past the number of layers needed to faithfully sample from a 2-design on each of the symmetry blocks. Thus, in this large layer limit, the scaling with $N_{\mathcal{E}}$ is independent of the circuit construction, as every block simply represents a random CUE matrix in this limit. Fitting the curves in the lower right panel of Fig. H.2a ($\ell = 128$) and similar curves for $N = 4, 6, 8$ sites and, then, extrapolating to determine the number of samples $N_{\mathcal{E}}$ needed to reach an error of $\epsilon = 0.01$ leads to Fig. 12.3(a).

In the large layer limit, the scaling with Hilbert space dimension d_s observed in this figure applies equally well to symmetry-ignorant schemes if d_s is replaced with the full Hilbert space $d_{\mathcal{H}}$, i.e. the scaling observed is simply a property of sampling from any set of unitaries that forms a unitary 2-design in the $N_{\mathcal{E}} \to \infty$ limit. We use this fact and the fit in Eq. (12.6) to generate the inset of Fig. 12.3(a) showing the relative gain r_s in number of circuit samples for the symmetry-conscious over a symmetry-ignorant scheme.

H.2.2 Details for Figure 3(b)

We now turn to providing additional details for Fig. 12.3(b), which shows the sample cost scaling for estimating sector-wise k-purities for subsystems of size $N_A = N/2$ for a system in the ground state of the particle-number symmetry-preserving Hamiltonian in Eq. (12.7). Recall, that we consider periodic boundary conditions and systems of size $N = 4, 6, 8, 10, 12, 14, 16, 18$ with couplings such that $ma = 0.05$ (i.e., a parameter regime where the ground state is entangled). k-purities can be extracted via randomized measurement schemes by utilizing the identities in Eq. (12.8), which hold when the expectation values $\langle P_U(b, s)^k \rangle$ taken with respect to the ensemble

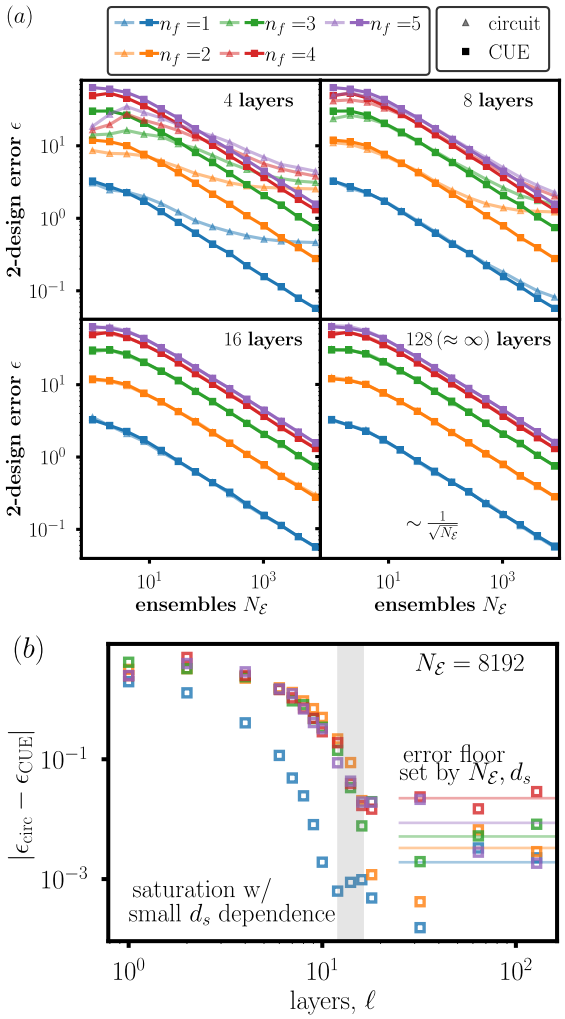


Figure H.2: (a) Error relative to an exact unitary 2-design ϵ (Eq. (12.5)) per particle number symmetry sector (labeled by n_f) versus number of unitaries $N_\mathcal{E}$. For clarity only particle number sectors $n_f \leq 5$ are shown; as they are of equivalent dimension, particle number sectors for $n_f > 5$ have the same behavior as the particle number sector $(10 - n_f)$. The ensemble $\mathcal{E}$ is sampled both from the particle number symmetry-respecting random circuits with different numbers of layers $\ell \in \{4, 8, 16, 128\}$ (triangles) and sampled directly from the CUE within each symmetry sector (squares). Results shown are for $n = 10$ sites and the error is averaged over 2000 non-zero matrix elements. Extrapolation of fits to the error as a function of $N_\mathcal{E}$ at $\ell = 128$ layers for the data pictured here and equivalent data for $N = 4, 6, 8$ sites is used to produce Fig. 12.3a of the main text. (b) The difference between the error ϵ obtained by taking $N_\mathcal{E} = 8192$ samples from the random circuits and taking $N_\mathcal{E} = 8192$ samples directly from the CUE versus number of layers used in the circuits. By $\ell \approx 15$, the difference reaches a floor set by $N_\mathcal{E}$ (i.e. the standard sampling error) with a small dependence on the dimension d_s of the corresponding sector.

of random unitaries $\mathcal{E}$ converge to the value obtained for an ensemble that is an exact unitary k-design.

A sufficient condition for this convergence is that $\mathcal{E}$ forms an approximate unitary k-design; however, note that it is possible that these expectation values converge to a fixed error for a smaller number of samples $N_{\mathcal{E}}$ than is needed to converge to the same fixed error in being a unitary k-design. This is indeed what is observed, as can be seen by comparing the number of samples needed for a convergent estimate of the k-purities, shown in Fig. H.3(a), to the number of samples needed to reach an approximate unitary k-design, shown in Fig. H.2(a). This is because an approximate unitary k-design will reproduce expectation values of all operators of degree k, whereas for this scheme, we must only reproduce the k-purities. It is also important to note that the accuracy is inherently dependent on the state under consideration; the k-design bound provides a worst case scenario.

In Fig. H.3(a), we have shown the number of ensembles $N_{\mathcal{E}}$ required to estimate the 2- and 3-purities to a relative error of $\epsilon = 0.05$ in the infinite shot limit $N_M \to \infty$. The necessary number of ensembles peaks and then begins to decrease as a function of block dimension. This trend cannot continue indefinitely (one must always implement at least one random unitary); consequently, we expect this behavior to saturate for large enough block dimension, as the variance of the infinite shot purity estimator approaches a constant in the large Hilbert space dimension limit [340].

To create Fig. 12.3(b) we fix the number of ensembles $N_{\mathcal{E}} = 1428$, well beyond the the number of ensembles needed to predict the 2-purity and 3-purity to within 5 percent for all cases considered. Since $N_{\mathcal{E}}$ is constant, this allows us to consider the scaling of the sample cost to purely depend on the number of shots per random unitary from the ensemble (i.e. the number of

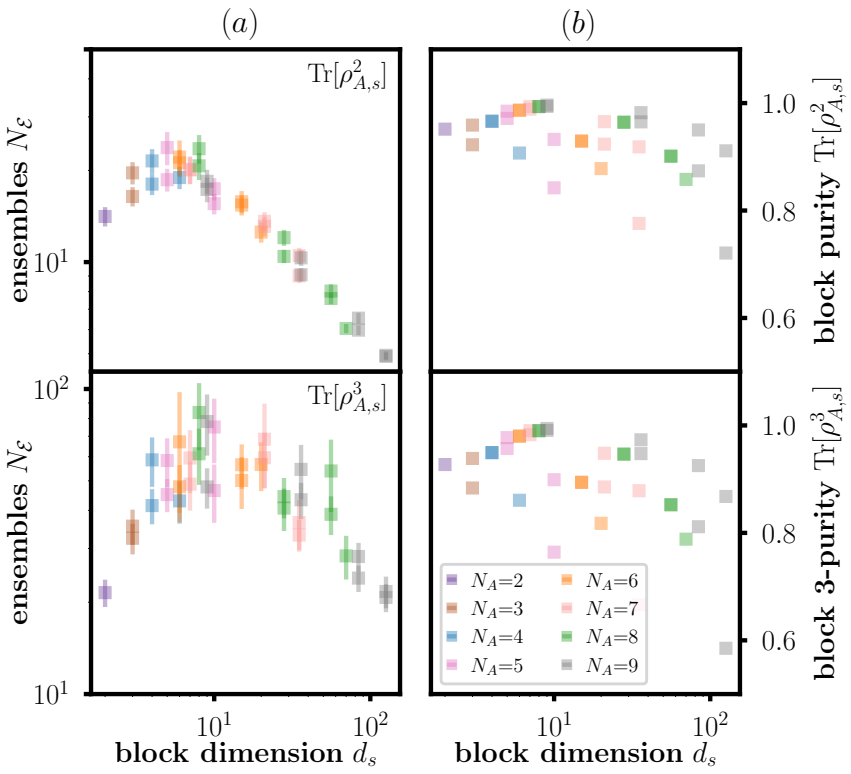


Figure H.3: (a) The number of ensembles $N_\mathcal{E}$ required, in the infinite shot limit, to estimate the 2-purity and 3-purity of a single block to 5 percent error. (b) Actual k-purities for reduced density matrices on subsystems of size $N_A = N/2$ for each particle number symmetry block for the (normalized) states used in Fig. 12.3(b).

measurements N_M made in each random basis). Therefore the cost N_M plotted in Fig. 12.3(b) is representative of the full sample cost of estimating the 2-purity of the subsystem states ρ_A.

As a sanity check on the $N_\mathcal{E}$ scaling, in Fig. H.3(b) we plot the true k-purities for the states used to create Fig. 12.3(b) and Fig. H.3(a). Note that the states, even for large Hilbert space dimension, have purities of order one. As the cost in $N_\mathcal{E}$ for purity estimation is expected to be largest in the pure state case [340], this shows that the trend in Fig. H.3(a) is not simply because states at large N (large d_s) are less pure.

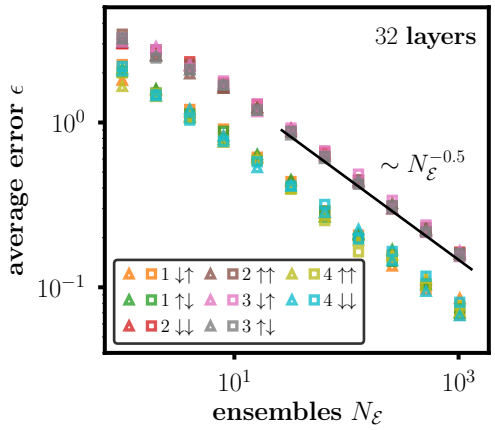


Figure H.4: Average error ϵ in 2-design matrix elements (Eq. (12.5)), normalized by $d_s(d_s^2-1)$, versus number of random unitaries $N_{\mathcal{E}}$ sampled from either the symmetry-respecting random circuits of Section 12.3.1 of depth $\ell = 32$ (triangles) or from direct sampling from the CUE within each symmetry sector (squares) for $\mathbb{Z}_2^{1+1}$ LGT for a subsystem of size $N_A = 5$ ($N = 10$). Number of indices sampled is 900.

H.3 Details for the (1+1)d $\mathbb{Z}_2$ LGT Example

In the main text, we presented circuits forming symmetry-conscious k-designs for $\mathbb{Z}_2$ LGT in (1+1)d with matter. We demonstrated that these circuits allow to measure k-purities and von Neumann entropies within each symmetry sector, as well as separately extracting the symmetry and distillable entanglement, and the symmetry-resolved Schmidt spectrum using classical shadows. A sufficient, but not necessary condition for such randomized measurement schemes to be successful is that the randomizing circuits form approximate unitary k-designs. In this Appendix, we explicitly demonstrate that the circuits in question do, in fact, form a sector-wise approximate unitary 2-design. In particular, we show that they reproduce the correct 2-design matrix elements (see Eq. (12.5)) for sufficiently deep circuits.

Representative results are shown in Fig. H.4 for a subsystem of size $N_A = 5$ (9 qubits) of a $N = 10$ (matter) site system (20 qubits). As described in the main text, Gauss laws at

the entanglement boundaries lead to symmetries of ρ_A in the subsystem. We demonstrate that, within each sector, the random circuits described in Sect. 12.3.1 form a 2-design, by computing the error defined in Eq. (H.1) with respect to $N_{\mathcal{E}}$ random circuits with $\ell = 32$ layers for all non-trivial symmetry sectors calculated.[1] To compute this error, we average over $|S| = 900$ random non-zero matrix elements. We see good agreement between sampling from our random circuits versus sampling directly from the CUE, indicating that our circuits do indeed form approximate unitary 2-designs.

H.4 Details of the (2+1)d $\mathbb{Z}_2$ LGT Example

In this Appendix, we provide details of the analysis of $\mathbb{Z}_2^{2+1}$, discussed in the main text.

H.4.1 Approximate Unitary k-designs and k-purities

The determination of k-purities follows exactly that in (1+1) spacetime dimensions. We explicitly show that the (2+1)d circuits in Fig. 12.7(b) explicitly realizes a 2-design by repeating the analysis of Section 12.2. Representative results for a subsystem of size 3×2 are summarized in Fig. H.5 showing the error Eq. (H.1), for every symmetry sector s, demonstrating agreement between sampling from our circuits for a sufficient number of layers and sampling directly from the CUE in each sector. Further, we see convergence with increasing samples $N_{\mathcal{E}}$ with the standard $1/\sqrt{N_{\mathcal{E}}}$ scaling.

[1] Filling sectors $n_A = 0$ and $n_A = N^A$ are trivial as they have unit block size and are not shown.

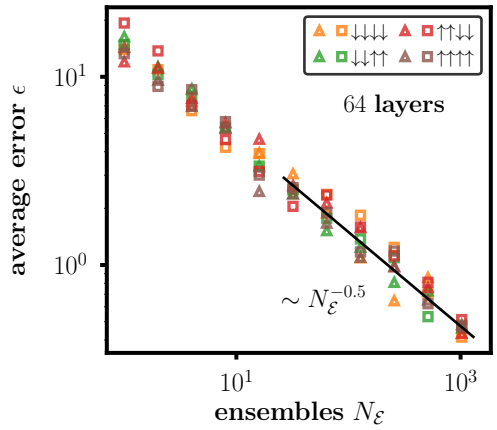


Figure H.5: Average error ϵ in 2-design matrix elements (Eq. (12.5)), normalized by $d_s(d_s^2-1)$, versus number of random unitaries $N_{\mathcal{E}}$ sampled from either symmetry-respecting random circuits of depth $\ell = 64$ (triangles) or from direct sampling from the CUE within each symmetry sector (squares) for $\mathbb{Z}_2^{2+1}$ LGT for a subsystem of size 3×2 (with fixed boundary conditions in y). Number of indices sampled is 900.

H.4.2 Classical Shadow Analysis

The classical shadow analysis for $\mathbb{Z}_2^{2+1}$ ground states follows the previously discussed (1+1)d case. Fig. H.6 shows the sectorwise relative entropy,

$$S(\bar{\rho}_{A,s}||\bar{\sigma}_{A,s}) \equiv -\text{tr}_s[\bar{\rho}_{A,s}(\log(\bar{\rho}_{A,s}) - \log(\bar{\sigma}_{A,s})] \tag{H.9}$$

where $\bar{\rho}_{A,s}$ and $\bar{\sigma}_{A,s}$ are the exact and shadow-reconstructed reduced density matrices (projected onto symmetry block s) of the $\mathbb{Z}_2^{2+1}$ ground state at $\epsilon = 0.2$, with $N_x \times N_y = (3+5) \times 2$, fixed BC in y, and $\ell = 64$ layers; bars indicate normalization i.e. $\bar{\rho}_{A,s} = \rho_{A,s}/p_s$ where $p_s = \text{tr}_s[\rho_{A,s}]$; Tr_s denotes the trace over sector $s \in \{\uparrow/\downarrow\}^{2^{2N_y}}$. An accuracy of up to $10^{-2} - 10^{-3}$ is achieved for the largest samples (where $p_s \cdot N_{\mathcal{S}} \gtrapprox 10^4$); the BW-EHT ansatz at similar cost typically reaches a precision better than $10^{-5} - 10^{-6}$ for the same configuration. Convergence with increasing shadow number $N_{\mathcal{S}}$ of the shadow-reconstructed density matrix towards the exact one is evident

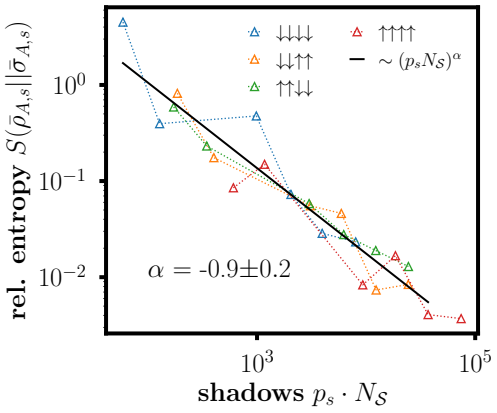


Figure H.6: Relative entropy between exact and shadow-reconstructed symmetry-resolved $\rho_{A,s}$ for the $\mathbb{Z}_2^{2+1}$ LGT ground state at $\epsilon = 0.2$, with $N_x \times N_y = (3+5) \times 2$ and fixed BC in y, $\ell = 64$ circuit layers. Data for $N_S = 2^{16}$ are shown in Fig. 12.8(b) of the main text.

and shows a power-law behavior consistent with the scaling of the (1+1)d case within error bars.

H.4.3 Entanglement Hamiltonian Tomography Analysis

Finally, our Bisognano-Wichmann theorem based entanglement Hamiltonian tomography (BW-EHT) protocol follows Ref. [336], except that we perform the optimization in every symmetry-sector s separately. The approach is based on an alternative representation of a reduced density matrix with Schmidt representation $\rho_A = \sum_\lambda P_\lambda |\lambda\rangle\langle\lambda|$. In particular, one defines an Entanglement Hamiltonian as

$$H_A = -\log[\rho_A]. \tag{H.10}$$

Eigenvalues are given by $P_\lambda = \exp(-\xi_\lambda)$, where ξ_λ are the eigenvalues of H_A.

Because ρ_A can be split into its corresponding symmetry sectors as $\rho_A = \bigoplus_s \rho_{A,s}$ it follows that also $H_A = \bigoplus_s H_{A,s}$, where s labels quantum numbers of the spectrum $\xi_{\lambda,s}$ of $H_{A,s}$. Inserting Eq. (12.23), for the Entanglement Hamiltonian we obtain the following BW-EHT ansatz for the

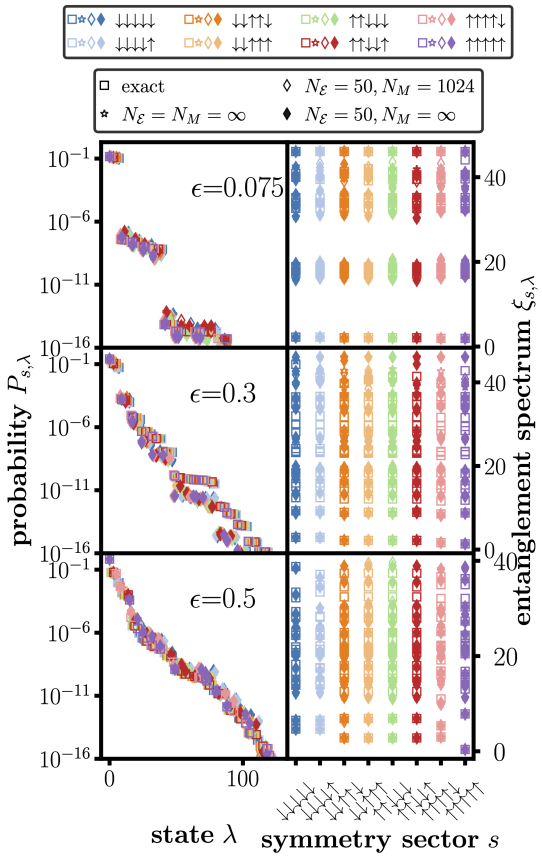


Figure H.7: Left column: Symmetry resolved Schmidt spectrum $P_{s,\lambda}$, reconstructed using BW-EHT, for $\epsilon = 0.075, 0.3, 0.5$ and with $N_x \times N_y = (3+3) \times 2$ and periodic boundary conditions in y (and x), $\ell = 64$, $N_\mathcal{E} = 50$, $N_M = 1024$. Right column: Symmetry resolved Entanglement spectrum.

state within symmetry sector labeled by quantum number s:

$$\bar{\rho}_{A,s}[\{\beta_i\}] \equiv \frac{\exp\{-H_{A,s}[\beta_i]\}}{\mathrm{Tr}_s[\exp\{-H_{A,s}[\beta_i]\}]} \tag{H.11}$$

where $H_{A,s}$ is is a deformation of the physical Hamiltonian, i.e. local couplings β_i replace the physical couplings. The ansatz is such that the state is normalized within each symmetry sector, i.e., $\mathrm{Tr}_s[\bar{\rho}_s] = \mathrm{Tr}_s[\rho]/p_s = 1$, where p_s is the probability of being in sector s. We find the optimal

couplings $\{\beta_i\}$ by minimizing the following functional

$$\sum_b \left\langle \left(P_U(b,s) - \mathrm{Tr}_s\left[\bar{\rho}_{A,s} U_s |b,s\rangle\langle b,s| U_s^\dagger\right]\right)^2 \right\rangle_{\mathcal{E}}, \tag{H.12}$$

separately for each s. Here, $\mathcal{E} = \{U_s\}$ is the ensemble of random circuits restricted to the block labeled by s, and $P_U(b,s)$ is the probability of measuring outcome bit string b, normalized such that $\sum_b P_U(b,s) = 1$ for all s. $P_U(b,s)$ is determined by classically simulating an (ideal) circuit for a given number of shots. In practice, the optimization is performed using python's simplicial homology global optimization (scipy.optimize.shgo) [495] with the following parameters

```
scipy.optimze.shgo(chi_squared, bounds,n=32,
sampling_method='sobol', options=opt_dict)
```

with sampling method 'sobol', and $n = 32$ sampling points in the construction of the simplicial complex, and very large bounds i.e. typically $\beta_i \in [-30.0, 30.0]$; all other options are set to their default values.

An example of the results of this analysis is shown in Fig. H.7, where we show the symmetry-resolved Schmidt $P_{s,\lambda}$ and entanglement spectrum $\xi_{\lambda,s}$ of the $\mathbb{Z}_2^{2+1}$ ground state at $\epsilon = 0.075, 0.3, 0.5$, for $N_x \times N_y = (3+3) \times 2$, periodic boundary conditions in y and $\ell = 64$. To estimate the error from applying a finite number of random circuits and estimating the effect of shot noise on obtaining $P_U(b,s)$, we additionally perform the following analysis: We compute the exact state using exact diagonalization. We then numerically minimize, within each symmetry sector s, the relative entropy between the exact density operator $\bar{\sigma}_s$, normalized so that $\mathrm{Tr}_s[\bar{\sigma}] = 1$, and the

BW ansatz $\bar{\rho}_s$,

$$S(\bar{\sigma}_s||\bar{\rho}_s) \equiv \mathrm{Tr}_s[\bar{\sigma}_s(\log(\bar{\sigma}_s) - \log(\bar{\rho}_s))]$$
$$= -S(\bar{\sigma}_s) + \Gamma(\bar{\rho}_s||\bar{\sigma}_s) \geq 0\,, \tag{H.13}$$

$S(\bar{\sigma}_s)$ is the exact von Neumann entropy and

$$\Gamma(\bar{\rho}_s||\bar{\sigma}_s) \equiv \log(\mathrm{Tr}_s[\bar{\rho}_{A,s}]) + \sum_i \beta_i \mathrm{Tr}_s[H_i\bar{\sigma}_s]),\,. \tag{H.14}$$

can be easily computed. Minimizing Eq. (H.13) with the exact same numerical optimization as used for Eq. (H.12) provides the BW-EHT result in the infinite measurement bases and infinite shot limit. The deviation of our circuit simulation from this result provides the error that we show in Fig. 12.8(c).

Bibliography

[1] R. P. Feynman, Int. J. Theor. Phys. **21**, 467 (1981).

[2] V. I. Arnold, On teaching mathematics (1997).

[3] G. Dyson, *Turing's cathedral: the origins of the digital universe* (Vintage Books, 2012).

[4] P. Kaye, R. Laflamme, and M. Mosca, *An introduction to quantum computing* (Oxford University Press, 2007).

[5] P. W. Shor, in *Proceedings 35th annual symposium on foundations of computer science* (IEEE, 1994) pp. 124–134.

[6] W. Jia, V. Xu, K. Kuns, M. Nakano, L. Barsotti, M. Evans, N. Mavalvala, R. Abbott, I. Abouelfettouh, R. Adhikari, *et al.*, arXiv preprint arXiv:2404.14569 (2024).

[7] K. M. Backes, D. A. Palken, S. A. Kenany, B. M. Brubaker, S. Cahn, A. Droster, G. C. Hilton, S. Ghosh, H. Jackson, S. K. Lamoreaux, *et al.*, Nature **590**, 238 (2021).

[8] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, Phys. Rev. A **97**, 042337 (2018).

[9] T. J. Proctor, P. A. Knott, and J. A. Dunningham, arXiv preprint arXiv:1702.04271 (2017).

[10] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov, Phys. Rev. A **100**, 042304 (2019).

[11] A. Abbas, A. Ambainis, B. Augustino, A. Bärtschi, H. Buhrman, C. Coffrin, G. Cortiana, V. Dunjko, D. J. Egger, B. G. Elmegreen, *et al.*, arXiv preprint arXiv:2312.02279 (2023).

[12] T. Albash and D. A. Lidar, Rev. Mod. Phys. **90**, 015002 (2018).

[13] D. Aharonov, W. v. Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, SIAM J. Comput. **37**, 166 (2008).

[14] J. Bringewatt, W. Dorland, S. P. Jordan, and A. Mink, Phys. Rev. A **97**, 022323 (2018).

[15] M. B. Hastings and M. H. Freedman, Quantum Inf. Comput. **13**, 1038 (2013), with appendix by M. H. Freedman.

[16] L. P. García-Pintos, L. T. Brady, J. Bringewatt, and Y.-K. Liu, Phys. Rev. Lett. **130**, 140601 (2023).

[17] S. Hallgren, C. Moore, M. Rötteler, A. Russell, and P. Sen, J. ACM **57**, 10.1145/1857914.1857918 (2010).

[18] M. B. Hastings, Quantum **5**, 597 (2021).

[19] A. Gilyén, M. B. Hastings, and U. Vazirani, in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (2021) pp. 1357–1369.

[20] J. Bringewatt, M. Jarret, and T. Mooney, arXiv preprint arXiv:2303.13478 (2023).

[21] H.-Y. Huang, Nat. Rev. Phys. **4** (2022).

[22] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, Nat. Rev. Phys. **5**, 9 (2023).

[23] J. D. Watson, J. Bringewatt, A. F. Shaw, A. M. Childs, A. V. Gorshkov, and Z. Davoudi, arXiv preprint arXiv:2312.05344 (2023).

[24] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[25] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[26] L. Pezzé and A. Smerzi, Phys. Rev. Lett. **102**, 100401 (2009).

[27] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi, Phys. Rev. A **85**, 022321 (2012).

[28] G. Tóth, Phys. Rev. A **85**, 022322 (2012).

[29] R. Augusiak, J. Kołodyński, A. Streltsov, M. N. Bera, A. Acín, and M. Lewenstein, Phys. Rev. A **94**, 012339 (2016).

[30] G. Tóth and I. Apellaniz, J. of Phys. A: Math. Theor. **47**, 424006 (2014).

[31] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola, Rev. Mod. Phys. **90**, 035006 (2018).

[32] A. Luis, Phys. Rev. A **65**, 025802 (2002).

[33] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature **450**, 393 (2007).

[34] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Phys. Rev. Lett. **79**, 3865 (1997).

[35] B. Escher, R. de Matos Filho, and L. Davidovich, Nat. Phys. **7**, 406 (2011).

[36] S. Boixo and C. Heunen, Phys. Rev. Lett. **108**, 120402 (2012).

[37] R. Demkowicz-Dobrzański and L. Maccone, Phys. Rev. Lett. **113**, 250801 (2014).

[38] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, Phys. Rev. Lett. **98**, 090401 (2007).

[39] S. Boixo, A. Datta, S. T. Flammia, A. Shaji, E. Bagan, and C. M. Caves, Phys. Rev. A **77**, 012317 (2008).

[40] T. Tilma, S. Hamaji, W. J. Munro, and K. Nemoto, Phys. Rev. A **81**, 022108 (2010).

[41] G. Gour and R. W. Spekkens, New J. Phys. **10**, 033023 (2008).

[42] I. Marvian and R. W. Spekkens, Nat. comm. **5**, 1 (2014).

[43] I. Marvian and R. W. Spekkens, Phys. Rev. A **94**, 052324 (2016).

[44] C. Zhang, B. Yadin, Z. Hou, H. Cao, B. Liu, Y. Huang, R. Maity, V. Vedral, C. Li, G. Guo, and D. Girolami, Phys. Rev. A **96**, 042327 (2017).

[45] C. W. Helstrom, *Quantum detection and estimation theory*, Vol. 3 (Academic press New York, 1976).

[46] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[47] S. L. Braunstein, C. M. Caves, and G. J. Milburn, Ann. Phys. (N. Y.) **247**, 135 (1996).

[48] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, Vol. 1 (Springer Science & Business Media, 2011).

[49] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, Phys. Rev. A **97**, 042337 (2018).

[50] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, Phys. Rev. Lett. **121**, 043604 (2018).

[51] T. J. Proctor, P. A. Knott, and J. A. Dunningham, Phys. Rev. Lett. **120**, 080501 (2018).

[52] S. Altenburg and S. Wölk, Phys. Scr. **94**, 014001 (2018).

[53] J. Rubio, P. A. Knott, T. J. Proctor, and J. A. Dunningham, J. Phys. A **53** (2020).

[54] J. A. Gross and C. M. Caves, arXiv preprint arXiv:2002.02898 (2020).

[55] D. Triggiani, P. Facchi, and V. Tamma, Phys. Rev. A **104**, 062603 (2021).

[56] C. Oh, L. Jiang, and C. Lee, Phys. Rev. Research **4**, 023164 (2022).

[57] M. Malitesta, A. Smerzi, and L. Pezzè, arXiv preprint arXiv:2109.09178 (2021).

[58] J. Bringewatt, I. Boettcher, P. Niroula, P. Bienias, and A. V. Gorshkov, Phys. Rev. Research **3**, 033011 (2021).

[59] T. Qian, J. Bringewatt, I. Boettcher, P. Bienias, and A. V. Gorshkov, Phys. Rev. A. **103**, L030601 (2021).

[60] A. Fujiwara, Phys. Rev. A **63**, 042304 (2001).

[61] J. Liu, H. Yuan, X. Lu, and X. Wang, J. Phys. A: Math. Theor. **53**, 023001 (2020).

[62] S.-I. Amari, *Differential-geometrical methods in statistics* (Springer, Berlin, 1985).

[63] Y. Yang, G. Chiribella, and M. Hayashi, Comm. Math. Phys. **368**, 223 (2019).

[64] J. Suzuki, J. Phys. A: Math. Theor. **53**, 264001 (2020).

[65] J. Suzuki, Y. Yang, and M. Hayashi, J. Phys. A: Math. Theor. **53**, 453001 (2020).

[66] S. Kimmel, G. H. Low, and T. J. Yoder, Phys. Rev. A **92**, 062315 (2015).

[67] S. Kimmel, G. H. Low, and T. J. Yoder, Phys. Rev. A **104**, 069901(E) (2021).

[68] F. Belliardo and V. Giovannetti, Phys. Rev. A **102**, 042613 (2020).

[69] B. Higgins, D. Berry, S. Bartlett, M. Mitchell, H. Wiseman, and G. Pryde, New J. Phys. **11**, 073023 (2009).

[70] M. Hayashi, S. Vinjanampathy, and L. Kwek, J. Phys. B: At. Mol. Opt. Phys. **52**, 015503 (2019).

[71] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry, Phys. Rev. Lett. **124**, 030501 (2020).

[72] J. Farkas, J. Reine Angew. Math. **1902**, 1 (1902).

[73] N. Dinh and V. Jeyakumar, TOP **22**, 1 (2014).

[74] M. Piani, M. Cianciaruso, T. R. Bromley, C. Napoli, N. Johnston, and G. Adesso, Phys. Rev. A **93**, 042107 (2016).

[75] M. M. Taddei, B. M. Escher, L. Davidovich, and R. L. de Matos Filho, Phys. Rev. Lett. **110**, 050402 (2013).

[76] S. Deffner and S. Campbell, J. Phys. A: Math. Theor. **50**, 453001 (2017).

[77] L. P. García-Pintos, S. B. Nicholson, J. R. Green, A. del Campo, and A. V. Gorshkov, Phys. Rev. X **12**, 011038 (2022).

[78] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Rev. Mod. Phys. **90**, 035005 (2018).

[79] Z. Zhang and Q. Zhuang, Quantum Science and Technology **6**, 043001 (2021).

[80] Q. Zhuang, Z. Zhang, and J. H. Shapiro, Phys. Rev. A **97**, 032329 (2018).

[81] Y. Xia, Q. Zhuang, W. Clark, and Z. Zhang, Phys. Rev. A **99**, 012328 (2019).

[82] C. Oh, C. Lee, S. H. Lie, and H. Jeong, Phys. Rev. **2**, 023030 (2020).

[83] A. Ehrenberg, J. Bringewatt, and A. V. Gorshkov, Phys. Rev. Research **5**, 033228 (2023).

[84] C. Oh, L. Jiang, and C. Lee, Phys. Rev. Research **4**, 023164 (2022).

[85] Y. Yang, B. Yadin, and Z.-P. Xu, arXiv preprint arXiv:2307.07758 (2023).

[86] X. Guo, C. R. Breum, J. Borregaard, S. Izumi, M. V. Larsen, T. Gehring, M. Christandl, J. S. Neergaard-Nielsen, and U. L. Andersen, Nat. Phys. **16**, 281 (2020).

[87] Y. Xia, W. Li, W. Clark, D. Hart, Q. Zhuang, and Z. Zhang, Phys. Rev. Lett. **124**, 150502 (2020).

[88] 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, *et al.*, Nat. Phot. **15**, 137 (2021).

[89] 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, J. Zhang, J. Fan, F. Xu, Q. Zhang, and J.-W. Pan, Phys. Rev. X **11**, 031009 (2021).

[90] T. Qian, J. Bringewatt, I. Boettcher, P. Bienias, and A. V. Gorshkov, Phys. Rev. A. **103**, L030601 (2021).

[91] A. Hamann, P. Sekatski, and W. Dür, Quantum Sci. Technol. **7**, 025003 (2022).

[92] J. Rubio and J. Dunningham, Phys. Rev. A **101**, 032114 (2020).

[93] E. Polino, M. Valeri, N. Spagnolo, and F. Sciarrino, AVS Quantum Science **2**, 024703 (2020).

[94] S. Arora and B. Barak, *Computational complexity: a modern approach* (Cambridge University Press, 2009).

[95] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A **54**, R4649 (1996).

[96] G. A. M. Paris, Int. J. Quantum Inf. **7**, 125 (2009).

[97] Z. Zhang and L. Duan, New J. Phys. **16**, 103037 (2014).

[98] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, Phys. Rev. A **46**, R6797 (1992).

[99] Y. Gao and H. Lee, Eur. Phys. J. D **68**, 347 (2014).

[100] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, Phys. Rev. A **94**, 052108 (2016).

[101] M. Gessner, L. Pezzè, and A. Smerzi, Phys. Rev. Lett. **121**, 130503 (2018).

[102] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov, Phys. Rev. A **100**, 042304 (2019).

[103] D. Gatto, P. Facchi, F. A. Narducci, and V. Tamma, Phys. Rev. Res. **1**, 032024 (2019).

[104] F. Albarelli, J. F. Friel, and A. Datta, Phys. Rev. Lett. **123**, 200503 (2019).

[105] P. Sekatski, S. Wölk, and W. Dür, Phys. Rev. Res. **2**, 023052 (2020).

[106] J. S. Sidhu and P. Kok, AVS Quantum Science **2**, 014701 (2020).

[107] Q. Zhuang, J. Preskill, and L. Jiang, New J. of Phys. **22**, 022001 (2020).

[108] N. Spagnolo, L. Aparo, C. Vitelli, A. Crespi, R. Ramponi, R. Osellame, P. Mataloni, and F. Sciarrino, Sci. Rep. **2**, 1 (2012).

[109] M. G. Genoni, M. G. A. Paris, G. Adesso, H. Nha, P. L. Knight, and M. S. Kim, Phys. Rev. A **87**, 012107 (2013).

[110] P. Humphreys, M. Barbieri, A. Datta, and I. A. Walmsley, Phys. Rev. Lett. **111**, 070403 (2013).

[111] J. Yue, Y. Zhang, and H. Fan, Sci. rep. **4**, 5933 (2014).

[112] T. Baumgratz and A. Datta, Phys. Rev. Lett. **116**, 030801 (2016).

[113] J. S. Sidhu and P. Kok, Phys. Rev. A **95**, 063829 (2017).

[114] P. Kok, J. Dunningham, and J. F. Ralph, Phys. Rev. A **95**, 012326 (2017).

[115] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, Phys. Rev. Lett. **98**, 090401 (2007).

[116] S. Jiang, Z. Song, O. Weinstein, and H. Zhang, arXiv preprint arXiv:2004.07470 (2020).

[117] M. J. Panik, *Linear programming: mathematics, theory and algorithms*, Vol. 2 (Springer Science & Business Media, 2013).

[118] J. Cadzow, SIAM J. Numer. Anal. **10**, 607 (1973).

[119] J. Cadzow, SIAM J. Numer. Anal. **11**, 1151 (1974).

[120] N. N. Abdelmalek, J. Approx. Theory **20**, 57 (1977).

[121] D. G. Luenberger, *Optimization by Vector Space Methods* (John Wiley & Sons, 1969) Chap. 5.

[122] M. J. Holland and K. Burnett, Phys. Rev. Lett. **71**, 1355 (1993).

[123] T. Kim, O. Pfister, M. J. Holland, J. Noh, and J. L. Hall, Phys. Rev. A **57**, 4004 (1998).

[124] A. R. U. Devi and A. K. Rajagopal, Phys. Rev. A **79**, 062320 (2009).

[125] H. T. Dinani, M. K. Gupta, J. P. Dowling, and D. W. Berry, Phys. Rev. A **93**, 063804 (2016).

[126] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, Phys. Rev. Lett. **121**, 043604 (2018).

[127] P. Hyllus, L. Pezzé, and A. Smerzi, Phys. Rev. Lett. **105**, 120501 (2010).

[128] M. Tsang, F. Albarelli, and A. Datta, Phys. Rev. X **10**, 031023 (2020).

[129] Y.-R. Zhang and H. Fan, Phys. Rev. A **90**, 043818 (2014).

[130] P. A. Knott, T. J. Proctor, A. J. Hayes, J. F. Ralph, P. Kok, and J. A. Dunningham, Phys. Rev. A **94**, 062312 (2016).

[131] M. Szczykulska, T. Baumgratz, and A. Datta, Adv. Phys. X **1**, 621 (2016), https://doi.org/10.1080/23746149.2016.1230476 .

[132] L. Pezzè, M. A. Ciampini, N. Spagnolo, P. C. Humphreys, A. Datta, I. A. Walmsley, M. Barbieri, F. Sciarrino, and A. Smerzi, Phys. Rev. Lett. **119**, 130504 (2017).

[133] M. J. W. Hall, J. Phys. A: Math. Theor. **51**, 364001 (2018).

[134] M. Tsang, arXiv preprint arXiv:1911.08359 (2019).

[135] F. Albarelli, M. Tsang, and A. Datta, arXiv preprint arXiv:1911.11036 (2019).

[136] M. D. Vidrighin, G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. S. Kim, A. Datta, M. Barbieri, and I. A. Walmsley, Nat. Commun. **5**, 1 (2014).

[137] E. Roccia, V. Cimini, M. Sbroscia, I. Gianani, L. Ruggiero, L. Mancino, M. G. Genoni, M. A. Ricci, and M. Barbieri, Optica **5**, 1171 (2018).

[138] M. Valeri, E. Polino, D. Poderini, I. Gianani, G. Corrielli, A. Crespi, R. Osellame, N. Spagnolo, and F. Sciarrino, arXiv preprint arXiv:2002.01232 (2020).

[139] H. Hofmann, Phys. Rev. A **79**, 033822 (2009).

[140] N. Shettell, E. Kashefi, and D. Markham, Phys. Rev. A **105**, L010401 (2022).

[141] H. Kasai, Y. Takeuchi, H. Hakoshima, Y. Matsuzaki, and Y. Tokura, J. Phys. Soc. Jpn. **91**, 074005 (2022).

[142] M. T. Rahim, A. Khan, U. Khalid, J. u. Rehman, H. Jung, and H. Shin, Sci. Rep. **13**, 11630 (2023).

[143] S.-I. Amari and M. Kumon, Ann. Stat. **16**, 1044 (1988).

[144] A. Atkinson, A. Donev, and R. Tobias, *Optimum experimental designs, with SAS*, Vol. 34 (OUP Oxford, 2007).

[145] S. Bravyi, D. P. DiVincenzo, R. I. Oliveira, and B. M. Terhal, Quant. Inf. Comp. **8**, 0361 (2008).

[146] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, arXiv preprint quant-ph/0001106 (2000).

[147] S. Jansen, M.-B. Ruskai, and R. Seiler, J. Math. Phys. **48**, 102111 (2007).

[148] A. Elgart and G. A. Hagedorn, J. Math. Phys. **53**, 102202 (2012).

[149] M. Jarret, S. P. Jordan, and B. Lackey, Phys. Rev. A **94**, 042318 (2016).

[150] E. Farhi, J. Goldstone, and S. Gutmann, arXiv preprint quant-ph/0201031 (2002).

[151] L. T. Brady and W. van Dam, Phys. Rev. A **94**, 032309 (2016).

[152] B. W. Reichardt, in *Proceedings of the 36th annual ACM Symposium on Theory of Computing (STOC)* (2004) pp. 502–510.

[153] E. Crosson and M. Deng, arXiv:1410.8484 (2014).

[154] S. Muthukrishnan, T. Albash, and D. Lidar, Phys. Rev. X **6**, 031010 (2016).

[155] Z. Jiang, V. N. Smelyanskiy, S. V. Isakov, S. Boixo, G. Mazzola, M. Troyer, and H. Neven, Phys. Rev. A **95**, 012322 (2017).

[156] G. Fix and R. Heiberger, SIAM Numer. Anal. **9**, 78 (1972).

[157] C. Jiang and Z. Bai (2015), http://cmjiang.cs.ucdavis.edu/publications/FHtemp.pdf.

[158] https://github.com/jbringewatt/PolyTimeAlgorithms forAQC.

[159] J. Roland and N. J. Cerf, Phys. Rev. A **65**, 042308 (2002).

[160] M. Jarret, B. Lackey, A. Liu, and K. Wan, arXiv preprint arXiv:1810.04686 (2018).

[161] M. Slutskii, T. Albash, L. Barash, and I. Hen, New J. Phys. **21**, 113025 (2019).

[162] DWave System Documentation.

[163] C. Davis and W. M. Kahan, SIAM Journal on Numerical Analysis **7**, 1 (1970).

[164] S. Bravyi, D. P. Divincenzo, R. Oliveira, and B. M. Terhal, Quantum Info. Comput. **8**, 361 (2008).

[165] M. B. Hastings, Quantum Information and Computation **13**, 1038 (2013).

[166] S. Bravyi and D. Gosset, Phys. Rev. Lett. **119**, 100503 (2017).

[167] S. Bravyi, D. P. DiVincenzo, D. Loss, and B. M. Terhal, Phys. Rev. Lett. **101**, 070503 (2008).

[168] E. Crosson and A. W. Harrow, Quantum **5**, 395 (2021).

[169] E. Crosson and S. Slezak, arXiv preprint arXiv:2002.02232 (2020).

[170] J. Bringewatt, W. Dorland, and S. P. Jordan, Phys. Rev. A **100**, 032336 (2019).

[171] L. Babai, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (2016) pp. 684–697.

[172] H. A. Helfgott, J. Bajpai, and D. Dona, arXiv preprint arXiv:1710.04574 (2017).

[173] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, 1990).

[174] E. BrianDavies, G. Gladwell, J. Leydold, and P. F. Stadler, Linear Algebra Its Appl. **336**, 51 (2001).

[175] M. Jarret, arXiv preprint arXiv:1804.06857 (2018).

[176] E. Luks, Permutation groups and polynomial-time computation (American Mathematical Society, 1993) p. 139–175.

[177] J. Crawford, in *AAAI Workshop on Tractable Reasoning* (1992) pp. 17–22.

[178] V. N. Zemlyachenko, N. M. Korneenko, and R. I. Tyshkevich, J. Math Sci. **29**, 1426 (1985).

[179] Y. Yu, T. Wang, and R. J. Samworth, Biometrika **102**, 315 (2015).

[180] R. Bhatia, *Matrix analysis*, Vol. 169 (Springer Science & Business Media, 2013).

[181] V. Arvind, J. Köbler, S. Kuhnert, and Y. Vasudev, in *International Symposium on Mathematical Foundations of Computer Science* (Springer, 2012) pp. 100–111.

[182] E. Farhi, J. Goldstone, S. Gutmann, and D. Nagaj, Int. J. Quantum Inf. **6**, 503 (2008).

[183] J. A. Barker, J. Chem. Phys. **70**, 2914 (1979).

[184] E. Crosson and A. W. Harrow, Quantum **5**, 395 (2021).

[185] T. Kadowaki and H. Nishimori, Phys. Rev. E **58**, 5355–5363 (1998).

[186] L. T. Brady and W. van Dam, Phys. Rev. A **93**, 032304 (2016).

[187] E. Crosson and A. W. Harrow, in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)* (2016) pp. 714–723.

[188] Z. Jiang, V. N. Smelyanskiy, S. V. Isakov, S. Boixo, G. Mazzola, M. Troyer, and H. Neven, Phys. Rev. A **95**, 10.1103/physreva.95.012322 (2017).

[189] J. Bringewatt and M. Jarret, Phys. Rev. Lett. **125**, 170504 (2020).

[190] S. Bravyi, A. J. Bessen, and B. M. Terhal, arXiv preprint quant-ph/0611021 (2006).

[191] Z.-X. Li, Y.-F. Jiang, and H. Yao, Phys. Rev. B **91**, 241117 (2015).

[192] I. Hen, Phys. Rev. E **99**, 033306 (2019).

[193] Z.-Q. Wan, S.-X. Zhang, and H. Yao, arXiv: 2010.01141 (2020).

[194] D. Hangleiter, I. Roth, D. Nagaj, and J. Eisert, Sci. Adv. **6**, eabb8341 (2020).

[195] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).

[196] M. Marvian, D. A. Lidar, and I. Hen, Nat. Commun. **10**, 1571 (2019).

[197] J. Klassen and B. M. Terhal, Quantum **3**, 139 (2019).

[198] W. Specht, Jahresbericht der Deutschen Mathematiker-Vereinigung **50**, 19 (1940).

[199] N. Wiegmann, Journal of the Australian Mathematical Society **2**, 122 (1961).

[200] T. G. Gerasimova, R. A. Horn, and V. V. Sergeichuk, Linear Algebra Appl. **438**, 3829 (2013).

[201] N. Jing, Linear Algebra Appl. **481**, 235 (2015).

[202] J. E. Harriman, Phys. Rev. A **17**, 1249 (1978).

[203] F. T. Hioe and J. H. Eberly, Phys. Rev. Lett. **47**, 838 (1981).

[204] M. S. Byrd and N. Khaneja, Phys. Rev. A **68**, 062322 (2003).

[205] G. Kimura, Phys. Lett. A **314**, 339 (2003).

[206] G. Kimura and A. Kossakowski, Open Syst. Inf. Dyn. **12**, 207 (2005).

[207] R. A. Bertlmann and P. Krammer, J. Phys. A: Math. Theor. **41**, 235303 (2008).

[208] A. Neven and T. Bastin, J. Phys. A: Math. Theor. **51**, 315305 (2018).

[209] N. Jing, M. Yang, and H. Zhao, J. Math Phys. **57**, 062205 (2016).

[210] R. Horn and C. R. Johnson, *Matrix analysis* (Cambridge University Press, 2013).

[211] C. Procesi, Adv. Math. **19**, 306 (1976).

[212] A. Paz, Linear and Multilinear Algebra **15**, 161 (1984).

[213] C. J. Pappacena, J. Algebra **197**, 535 (1997).

[214] D. Djoković and C. R. Johnson, Linear Algebra Appl. **421**, 63 (2007).

[215] J. Schlienz and G. Mahler, Phys. Lett. A **224**, 39 (1996).

[216] G. Mahler and V. A. Weberruß, *Quantum Networks Dynamics of Open Nanostructures* (Springer-Verlag, 1995).

[217] I. Bengtsson and K. Życzkowski, *Geometry of quantum states: An introduction to quantum entanglement* (Cambridge University Press, 2017).

[218] J. Canny, in *Proceedings of the twentieth annual ACM symposium on Theory of computing* (1988) pp. 460–467.

[219] V. Jurdjevic and H. J. Sussmann, Journal of Differential Equations **12**, 313 (1972).

[220] V. Ramakrishna, M. V. Salapaka, M. Dahleh, H. Rabitz, and A. Peirce, Phys. Rev. A **51**, 960 (1995).

[221] D. Bossion and P. Huo, arXiv preprint arXiv:2108.07219 (2021).

[222] I. Hen, Phys. Rev. Research **3**, 023080 (2021).

[223] S. Aaronson, arXiv preprint arXiv:2209.06930 (2022).

[224] E. Farhi, J. Goldstone, and S. Gutmann, arXiv preprint arXiv:1411.4028 (2014).

[225] E. Crosson and D. Lidar, Nat. Rev. Phys. **3**, 466 (2021).

[226] A. Braida, S. Martiel, and I. Todinca, Phys. Rev. A **109**, 022415 (2024).

[227] S. Ebadi, A. Keesling, M. Cain, T. T. Wang, H. Levine, D. Bluvstein, G. Semeghini, A. Omran, J.-G. Liu, R. Samajdar, *et al.*, Science **376**, 1209 (2022).

[228] I. M. Georgescu, S. Ashhab, and F. Nori, Rev. Mod. Phys. **86**, 153 (2014).

[229] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Chem. Rev. **120**, 12685 (2020).

[230] C. W. Bauer, Z. Davoudi, A. Balantekin, T. Bhattacharya, M. Carena, W. A. de Jong, P. Draper, A. El-Khadra, N. Gemelke, M. Hanada, D. Kharzeev, *et al.*, PRX Quantum **4**, 027001 (2023).

[231] *Nuclear Physics and Quantum Information Science: Report by the NSAC QIS Subcommittee*, Tech. Rep. (NSF & DOE Office of Science, 2019).

[232] P. Jordan and E. Wigner, Z. Phys. **47**, 631 (1928).

[233] G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, Phys. Rev. A **64**, 022319 (2001).

[234] V. Havlíček, M. Troyer, and J. D. Whitfield, Phys. Rev. A **95**, 032332 (2017).

[235] L. Clinton, J. Bausch, and T. Cubitt, Nat. Comm. **12**, 1 (2021).

[236] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).

[237] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, arXiv preprint arXiv:1403.1539 (2014).

[238] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, Phys. Rev. X **8**, 011044 (2018).

[239] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Phys. Rev. Lett. **120**, 110501 (2018).

[240] C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, Phys. Rev. B **102**, 235122 (2020).

[241] S. B. Bravyi and A. Y. Kitaev, Ann. of Phys. **298**, 210 (2002).

[242] F. Verstraete and J. I. Cirac, J. Stat. Mech. **2005**, P09012 (2005).

[243] J. D. Whitfield, V. Havlíček, and M. Troyer, Phys. Rev. A **94**, 030301 (2016).

[244] M. Steudtner and S. Wehner, New J. Phys. **20**, 063010 (2018).

[245] M. Steudtner and S. Wehner, Phys. Rev. A **99**, 022308 (2019).

[246] K. Setia, S. Bravyi, A. Mezzacapo, and J. D. Whitfield, Phys. Rev. **1**, 033033 (2019).

[247] Z. Jiang, J. McClean, R. Babbush, and H. Neven, Phys. Rev. Applied **12**, 064041 (2019).

[248] Y.-A. Chen, Phys. Rev. **2**, 033527 (2020).

[249] R. W. Chien and J. D. Whitfield, arXiv preprint arXiv:2009.11860 (2020).

[250] C. Derby, J. Klassen, J. Bausch, and T. Cubitt, Phys. Rev. B **104**, 035118 (2021).

[251] W. Kirby, B. Fuller, C. Hadfield, and A. Mezzacapo, PRX Quantum **3**, 020351 (2022).

[252] M. Chiew and S. Strelchuk, arXiv preprint arXiv:2110.12792 (2021).

[253] A. Tranter, S. Sofia, J. Seeley, M. Kaicher, J. McClean, R. Babbush, P. V. Coveney, F. Mintert, F. Wilhelm, and P. J. Love, Int. J. Quantum Chem. **115**, 1431 (2015).

[254] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, J. Chem. Theory Comput. **14**, 5617 (2018).

[255] M. Suzuki, J. of Math. Phys. **32**, 400 (1991).

[256] N. Wiebe, D. Berry, P. Hoyer, and B. C. Sanders, J. Phys. A **43**, 065203 (2010).

[257] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Phys. Rev. X **11**, 011020 (2021).

[258] A. M. Childs and N. Wiebe, arXiv preprint arXiv:1202.5822 (2012).

[259] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Phys. Rev. Lett. **114**, 090502 (2015).

[260] G. H. Low and I. L. Chuang, Phys. Rev. Lett. **118**, 010501 (2017).

[261] G. H. Low and I. L. Chuang, Quantum **3**, 163 (2019).

[262] S. Chakraborty, A. Gilyén, and S. Jeffery, 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019) Leibniz International Proceedings in Informatics (LIPIcs), **132**, 33:1 (2019).

[263] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019) pp. 193–204.

[264] A. Kalev and I. Hen, Quantum **5**, 426 (2021).

[265] A. Rajput, A. Roggero, and N. Wiebe, arXiv preprint arXiv:2109.03308 (2021).

[266] J. Bringewatt and Z. Davoudi, `https://github.com/jbringewatt/graph-coloring-fermion-parallelization`.

[267] Z. Jiang, A. Kalev, W. Mruczkiewicz, and H. Neven, Quantum **4**, 276 (2020).

[268] A. Y. Vlasov, Quanta **11**, 97 (2022).

[269] R. C. Ball, Phys. Rev. Lett. **95**, 176407 (2005).

[270] C. Derby and J. Klassen, arXiv preprint arXiv:2101.10735 (2021).

[271] A. Y. Kitaev, Ann. Phys. **303**, 2 (2003).

[272] Y.-A. Chen, A. Kapustin, and Đorđe Radičević, Ann. Phys. **393**, 234 (2018).

[273] J. R. Stryker, Phys. Rev. A **99**, 042301 (2019).

[274] M. C. Tran, Y. Su, D. Carney, and J. M. Taylor, PRX Quantum **2**, 010323 (2021).

[275] H. Lamm, S. Lawrence, and Y. Yamauchi, arXiv preprint arXiv:2005.12688 (2020).

[276] J. C. Halimeh, H. Lang, J. Mildenberger, Z. Jiang, and P. Hauke, PRX Quantum **2**, 040311 (2021).

[277] M. Van Damme, J. Mildenberger, F. Grusdt, P. Hauke, and J. C. Halimeh, arXiv preprint arXiv:2110.08041 (2021).

[278] K. Stannigel, P. Hauke, D. Marcos, M. Hafezi, S. Diehl, M. Dalmonte, and P. Zoller, Phys. Rev. Lett. **112**, 120406 (2014).

[279] V. Kasper, T. V. Zache, F. Jendrzejewski, M. Lewenstein, and E. Zohar, Phys. Rev. D **107**, 014506 (2023).

[280] N. H. Nguyen, M. C. Tran, Y. Zhu, A. M. Green, C. H. Alderete, Z. Davoudi, and N. M. Linke, arXiv preprint arXiv:2112.14262 (2021).

[281] K. Setia and J. D. Whitfield, J. Chem. Phys. **148**, 164104 (2018).

[282] A. M. Childs, A. Ostrander, and Y. Su, Quantum **3**, 182 (2019).

[283] A. Tranter, P. J. Love, F. Mintert, N. Wiebe, and P. V. Coveney, Entropy **21**, 10.3390/e21121218 (2019).

[284] T. Erlebach and K. Jansen, Theoretical Computer Science **255**, 33 (2001).

[285] R. M. Karp, in *Complexity of computer computations* (Springer, Boston, MA, 1972) pp. 85–103.

[286] R. L. Brooks, in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 37 (2) (Cambridge University Press, 1941) pp. 194–197.

[287] S. Sachdev and J. Ye, Phys. Rev. Lett. **70**, 3339 (1993).

[288] V. Rosenhaus, J. Phys. A **52**, 323001 (2019).

[289] A. Soifer, *The mathematical coloring book: Mathematics of coloring and the colorful life of its creators* (Springer, 2009).

[290] D. Bryant, London Mathematical Society Lecture Note Series **346**, 67 (2007).

[291] B. Alspach, Bull. Inst. Combin. Appl **52**, 7 (2008).

[292] K. Mehlhorn, P. Sanders, and P. Sanders, *Algorithms and data structures: The basic toolbox*, Vol. 55 (Springer, 2008).

[293] D. J. A. Welsh and M. B. Powell, Comput. J. **10**, 85 (1967).

[294] T. Husfeldt, in *Topics in Chromatic Graph Theory* (Cambridge University Press, 2015) pp. 277–303.

[295] A. Brandstädt, V. B. Le, and J. P. Spinrad, *Graph classes: a survey* (SIAM, 1999).

[296] IBM, `https://research.ibm.com/blog/heavy-hex-lattice`.

[297] C. Chamberland, G. Zhu, T. J. Yoder, J. B. Hertzberg, and A. W. Cross, Phys. Rev. X **10**, 011022 (2020).

[298] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Nature **574**, 505 (2019).

[299] R. Machleidt and D. R. Entem, Physics Reports **503**, 1 (2011).

[300] T. Byrnes and Y. Yamamoto, Phys. Rev. A **73**, 022328 (2006), arXiv:quant-ph/0510027 [quant-ph] .

[301] H. Lamm, S. Lawrence, Y. Yamauchi, N. Collaboration, *et al.*, Phys. Rev. D **100**, 034518 (2019).

[302] A. F. Shaw, P. Lougovski, J. R. Stryker, and N. Wiebe, Quantum **4**, 306 (2020).

[303] D. Paulson, L. Dellantonio, J. F. Haase, A. Celi, A. Kan, A. Jena, C. Kokail, R. van Bijnen, K. Jansen, P. Zoller, and C. A. Muschik, PRX Quantum **2**, 030334 (2021).

[304] A. Ciavarella, N. Klco, and M. J. Savage, Phys. Rev. D **103**, 094501 (2021).

[305] A. Kan and Y. Nam, arXiv preprint arXiv:2107.12769 (2021).

[306] C. Hamer, Z. Weihong, and J. Oitmaa, Phys. Rev. D **56**, 55 (1997).

[307] E. Zohar and J. I. Cirac, Phys. Rev. D **99**, 114511 (2019).

[308] Z. Davoudi, I. Raychowdhury, and A. Shaw, Phys. Rev. D **104**, 074505 (2021).

[309] N. Klco and M. J. Savage, Phys. Rev. D **103**, 065007 (2021).

[310] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nat. Commun. **5**, 1 (2014).

[311] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).

[312] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, *et al.*, Nature **569**, 355 (2019).

[313] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, *et al.*, Phys. Rep. **986**, 1 (2022).

[314] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Nat. Rev. Phys. **2**, 382 (2020).

[315] N. Friis, G. Vitagliano, M. Malik, and M. Huber, Nat. Rev. Phys. **1**, 72 (2019).

[316] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008).

[317] M. Paini and A. Kalev, arXiv preprint arXiv:1910.10543 (2019).

[318] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020).

[319] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **127**, 030503 (2021).

[320] H.-Y. Hu, S. Choi, and Y.-Z. You, Phys. Rev. Res. **5**, 023027 (2023).

[321] A. Zhao, N. C. Rubin, and A. Miyake, Phys. Rev. Lett. **127**, 110504 (2021).

[322] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, Phys. Rev. A **107**, 042403 (2023).

[323] R. Levy, D. Luo, and B. K. Clark, Phys. Rev. Res. **6**, 013029 (2024).

[324] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. Werner, J. Eisert, and I. Roth, Nat. Comm. **14**, 5039 (2023).

[325] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, *et al.*, Science **376**, 1182 (2022).

[326] G. Hao Low, arXiv preprint arXiv:2208.08964 (2022).

[327] H. Pichler, G. Zhu, A. Seif, P. Zoller, and M. Hafezi, Phys. Rev. X **6**, 041033 (2016).

[328] M. Dalmonte, B. Vermersch, and P. Zoller, Nat. Phys. **14**, 827 (2018).

[329] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **120**, 050406 (2018).

[330] B. Vermersch, A. Elben, M. Dalmonte, J. I. Cirac, and P. Zoller, Phys. Rev. A **97**, 023604 (2018).

[331] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Phys. Rev. A **99**, 052323 (2019).

[332] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Science **364**, 260 (2019).

[333] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, *et al.*, Phys. Rev. Lett. **125**, 200501 (2020).

[334] Y. Zhou, P. Zeng, and Z. Liu, Phys. Rev. Lett. **125**, 200502 (2020).

[335] A. Neven, J. Carrasco, V. Vitale, C. Kokail, A. Elben, M. Dalmonte, P. Calabrese, P. Zoller, B. Vermersch, R. Kueng, *et al.*, npj Quantum Inf. **7**, 1 (2021).

[336] C. Kokail, R. van Bijnen, A. Elben, B. Vermersch, and P. Zoller, Nat. Phys. **17**, 936 (2021).

[337] A. Rath, R. van Bijnen, A. Elben, P. Zoller, and B. Vermersch, Phys. Rev. Lett. **127**, 200503 (2021).

[338] C. Kokail, B. Sundar, T. V. Zache, A. Elben, B. Vermersch, M. Dalmonte, R. van Bijnen, and P. Zoller, Phys. Rev. Lett. **127**, 170501 (2021).

[339] T. V. Zache, C. Kokail, B. Sundar, and P. Zoller, Quantum **6**, 702 (2022).

[340] S. J. van Enk and C. W. Beenakker, Phys. Rev. Lett. **108**, 110503 (2012).

[341] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, New J. Phys. **14**, 095022 (2012).

[342] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (2016) pp. 913–925.

[343] R. O'Donnell and J. Wright, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (2016) pp. 899–912.

[344] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, PRX Quantum **2**, 030348 (2021).

[345] D. E. Koh and S. Grewal, Quantum **6**, 776 (2022).

[346] M. C. Tran, D. K. Mark, W. W. Ho, and S. Choi, arXiv preprint arXiv:2212.02517 (2022).

[347] R. Blatt and C. F. Roos, Nat. Phys. **8**, 277 (2012).

[348] I. Bloch, J. Dalibard, and S. Nascimbene, Nat. Phys. **8**, 267 (2012).

[349] C. Gross and I. Bloch, Science **357**, 995 (2017).

[350] F. Schäfer, T. Fukuhara, S. Sugawa, Y. Takasu, and Y. Takahashi, Nat. Rev. Phys. **2**, 411 (2020).

[351] L. Bassman, M. Urbanek, M. Metcalf, J. Carter, A. F. Kemper, and W. A. de Jong, Quantum Sci. Technol. **6**, 043002 (2021).

[352] C. Monroe, W. C. Campbell, L.-M. Duan, Z.-X. Gong, A. V. Gorshkov, P. Hess, R. Islam, K. Kim, N. M. Linke, G. Pagano, *et al.*, Rev. Mod. Phys. **93**, 025001 (2021).

[353] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Nature **607**, 667 (2022).

[354] J. M. Deutsch, Phys. Rev. A **43**, 2046 (1991).

[355] M. Srednicki, Phys. Rev. E **50**, 888 (1994).

[356] M. Rigol, V. Dunjko, and M. Olshanii, Nature **452**, 854 (2008).

[357] J. M. Deutsch, H. Li, and A. Sharma, Phys. Rev. E **87**, 042135 (2013).

[358] V. Khemani, A. Chandran, H. Kim, and S. L. Sondhi, Phys. Rev. E **90**, 052133 (2014).

[359] J. Eisert, M. Friesdorf, and C. Gogolin, Nat. Phys. **11**, 124 (2015).

[360] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Science **353**, 794 (2016).

[361] J. Berges, M. P. Heller, A. Mazeliauskas, and R. Venugopalan, Rev. Mod. Phys. **93**, 035003 (2021).

[362] Z.-Y. Zhou, G.-X. Su, J. C. Halimeh, R. Ott, H. Sun, P. Hauke, B. Yang, Z.-S. Yuan, J. Berges, and J.-W. Pan, Science **377**, 311 (2022).

[363] N. Mueller, T. V. Zache, and R. Ott, Phys. Rev. Lett. **129**, 011601 (2022).

[364] T.-C. Lu and T. Grover, Phys. Rev. Research **2**, 043345 (2020).

[365] M. Brenes, S. Pappalardi, J. Goold, and A. Silva, Phys. Rev. Lett. **124**, 040605 (2020).

[366] A. Osterloh, L. Amico, G. Falci, and R. Fazio, Nature **416**, 608 (2002).

[367] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Phys. Rev. Lett. **90**, 227902 (2003).

[368] F. Verstraete, M. Popp, and J. I. Cirac, Phys. Rev. Lett. **92**, 027901 (2004).

[369] G. Costantini, P. Facchi, G. Florio, and S. Pascazio, J. Phys. A: Math. Theor. **40**, 8009 (2007).

[370] H. Li and F. D. M. Haldane, Phys. Rev. Lett. **101**, 010504 (2008).

[371] D. Banerjee, M. Dalmonte, M. Müller, E. Rico, P. Stebler, U.-J. Wiese, and P. Zoller, Phys. Rev. Lett. **109**, 175302 (2012).

[372] E. Zohar, J. I. Cirac, and B. Reznik, Phys. Rev. Lett. **110**, 055302 (2013).

[373] E. Zohar, J. I. Cirac, and B. Reznik, Phys. Rev. A **88**, 023617 (2013).

[374] E. Zohar, J. I. Cirac, and B. Reznik, Phys. Rev. Lett. **110**, 125304 (2013).

[375] L. Tagliacozzo, A. Celi, P. Orland, M. Mitchell, and M. Lewenstein, Nat. Commun. **4**, 1 (2013).

[376] E. Zohar, J. I. Cirac, and B. Reznik, Rept. Prog. Phys. **79**, 014401 (2016), arXiv:1503.02312 [quant-ph] .

[377] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, *et al.*, Nature **534**, 516 (2016).

[378] D. Yang, G. S. Giri, M. Johanning, C. Wunderlich, P. Zoller, and P. Hauke, Phys. Rev. A **94**, 052321 (2016).

[379] T. V. Zache, F. Hebenstreit, F. Jendrzejewski, M. Oberthaler, J. Berges, and P. Hauke, Quantum Sci. Technol. (2018).

[380] N. Klco, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, and M. J. Savage, Phys. Rev. A **98**, 032331 (2018).

[381] H.-H. Lu, N. Klco, J. M. Lukens, T. D. Morris, A. Bansal, A. Ekström, G. Hagen, T. Papenbrock, A. M. Weiner, M. J. Savage, and P. Lougovski, Phys. Rev. A **100**, 012320 (2019).

[382] L. Barbiero, C. Schweizer, M. Aidelsburger, E. Demler, N. Goldman, and F. Grusdt, Sci. Adv. **5**, eaav7444 (2019).

[383] Z. Davoudi, M. Hafezi, C. Monroe, G. Pagano, A. Seif, and A. Shaw, Phys. Rev. Research **2**, 023015 (2020).

[384] F. M. Surace, P. P. Mazza, G. Giudici, A. Lerose, A. Gambassi, and M. Dalmonte, Phys. Rev. X **10**, 021041 (2020).

[385] D. Luo, J. Shen, M. Highman, B. K. Clark, B. DeMarco, A. X. El-Khadra, and B. Gadway, Phys. Rev. A **102**, 032617 (2020).

[386] M. C. Banuls, R. Blatt, J. Catani, A. Celi, J. I. Cirac, M. Dalmonte, L. Fallani, K. Jansen, M. Lewenstein, S. Montangero, *et al.*, Eur. Phys. J. D **74**, 1 (2020).

[387] A. Mil, T. V. Zache, A. Hegde, A. Xia, R. P. Bhatt, M. K. Oberthaler, P. Hauke, J. Berges, and F. Jendrzejewski, Science **367**, 1128 (2020).

[388] B. Chakraborty, M. Honda, T. Izubuchi, Y. Kikuchi, and A. Tomiya, arXiv:2001.00485 (2020).

[389] G. Magnifico, M. Dalmonte, P. Facchi, S. Pascazio, F. V. Pepe, and E. Ercolessi, Quantum **4**, 281 (2020).

[390] N. Klco, M. J. Savage, and J. R. Stryker, Phys. Rev. D **101**, 074512 (2020).

[391] N. Klco, A. Roggero, and M. J. Savage, Rept. Prog. Phys. **85**, 064301 (2022), arXiv:2107.04769 [quant-ph] .

[392] L. Homeier, C. Schweizer, M. Aidelsburger, A. Fedorov, and F. Grusdt, Phys. Rev. B **104**, 085138 (2021).

[393] G. Pederiva, A. Bazavov, B. Henke, L. Hostetler, D. Lee, H.-W. Lin, and A. Shindler, in *38th International Symposium on Lattice Field Theory* (2021).

[394] A. Rajput, A. Roggero, and N. Wiebe, Quantum **6**, 780 (2022).

[395] N. H. Nguyen, M. C. Tran, Y. Zhu, A. M. Green, C. H. Alderete, Z. Davoudi, and N. M. Linke, PRX Quantum **3**, 020324 (2022).

[396] W. A. de Jong, K. Lee, J. Mulligan, M. Płoskoń, F. Ringer, and X. Yao, Phys. Rev. D **106**, 054508 (2022).

[397] S. A Rahman, R. Lewis, E. Mendicelli, and S. Powell, Phys. Rev. D **104**, 034501 (2021).

[398] J. F. Haase, L. Dellantonio, A. Celi, D. Paulson, A. Kan, K. Jansen, and C. A. Muschik, Quantum **5**, 393 (2021).

[399] A. Ciavarella, N. Klco, and M. J. Savage, Phys. Rev. D **103**, 094501 (2021).

[400] M. S. Alam, S. Hadfield, H. Lamm, and A. C. Y. Li (SQMS Collaboration), Phys. Rev. D **105**, 114501 (2022).

[401] A. N. Ciavarella and I. A. Chernyshev, Phys. Rev. D **105**, 074504 (2022).

[402] T. D. Cohen, H. Lamm, S. Lawrence, and Y. Yamauchi (NuQS Collaboration), Phys. Rev. D **104**, 094514 (2021).

[403] D. González-Cuadra, T. V. Zache, J. Carrasco, B. Kraus, and P. Zoller, Phys. Rev. Lett. **129**, 160501 (2022).

[404] J. C. Halimeh, H. Lang, and P. Hauke, New J. Phys. **24**, 033015 (2022).

[405] B. Andrade, Z. Davoudi, T. Graß, M. Hafezi, G. Pagano, and A. Seif, Quantum Sci. Technol. **7**, 034001 (2022).

[406] Y. Y. Atas, J. F. Haase, J. Zhang, V. Wei, S. M.-L. Pfaendler, R. Lewis, and C. A. Muschik, arXiv preprint arXiv:2207.03473 (2022).

[407] R. C. Farrell, I. A. Chernyshev, S. J. M. Powell, N. A. Zemlevskiy, M. Illa, and M. J. Savage, Phys. Rev. D **107**, 054512 (2023).

[408] E. M. Murairi, M. J. Cervia, H. Kumar, P. F. Bedaque, and A. Alexandru, Phys. Rev. D **106**, 094504 (2022).

[409] G. Clemente, A. Crippa, and K. Jansen, Phys. Rev. D **106**, 114511 (2022).

[410] N. Mueller, J. A. Carolan, A. Connelly, Z. Davoudi, E. F. Dumitrescu, and K. Yeter-Aydeniz, PRX Quantum **4**, 030323 (2023).

[411] Z. Davoudi, N. Mueller, and C. Powers, Phys. Rev. Lett. **131**, 081901 (2023).

[412] C. Kane, D. M. Grabowska, B. Nachman, and C. W. Bauer, arXiv preprint arXiv:2211.10497 (2022).

[413] J. Mildenberger, W. Mruczkiewicz, J. C. Halimeh, Z. Jiang, and P. Hauke, arXiv preprint arXiv:2203.08905 (2022).

[414] E. J. Gustafson and H. Lamm, arXiv preprint arXiv:2301.10207 (2023).

[415] T. V. Zache, D. Gonzalez-Cuadra, and P. Zoller, Quantum **7**, 1140 (2023).

[416] P. Buividovich and M. Polikarpov, Phys. Lett. B **670**, 141 (2008).

[417] H. Casini, M. Huerta, and J. A. Rosabal, Phys. Rev. D **89**, 085012 (2014).

[418] S. Aoki, T. Iritani, M. Nozaki, T. Numasawa, N. Shiba, and H. Tasaki, J. High Energy Phys. **2015** (6), 1.

[419] S. Ghosh, R. M. Soni, and S. P. Trivedi, J. High Energy Phys. **2015** (9), 1.

[420] K. Van Acoleyen, N. Bultinck, J. Haegeman, M. Marien, V. B. Scholz, and F. Verstraete, Phys. Rev. Lett. **117**, 131602 (2016).

[421] J. Lin and D. Radicevic, Nucl. Phys. **958**, 115118 (2020).

[422] M. Rigobello, S. Notarnicola, G. Magnifico, and S. Montangero, Phys. Rev. D **104**, 114501 (2021).

[423] V. Panizza, R. Costa de Almeida, and P. Hauke, J. High Energy Phys. **2022** (9), 1.

[424] D. C. Tsui, H. L. Stormer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).

[425] X.-G. Wen, Int. J. Mod. Phys. A **4**, 239 (1990).

[426] A. Kitaev, Annals of Physics **321**, 2 (2006).

[427] S. Das Sarma, M. Freedman, and C. Nayak, Physics Today **59**, 32 (2006).

[428] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Rev. Mod. Phys. **80**, 1083 (2008).

[429] S. Das Sarma, M. Freedman, and C. Nayak, npj Quantum Information **1**, 1 (2015).

[430] V. Lahtinen and J. K. Pachos, SciPost Phys. **3**, 021 (2017).

[431] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (2018) pp. 325–338.

[432] S. Aaronson and G. N. Rothblum, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019) pp. 322–333.

[433] K. Satzinger, Y.-J. Liu, A. Smith, C. Knapp, M. Newman, C. Jones, Z. Chen, C. Quintana, X. Mi, A. Dunsworth, *et al.*, Science **374**, 1237 (2021).

[434] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar, A. Omran, S. Sachdev, A. Vishwanath, M. Greiner, V. Vuletić, and M. D. Lukin, Science **374**, 1242 (2021).

[435] G. H. Low, arXiv preprint arXiv:2208.08964 (2022).

[436] K. Wan, W. J. Huggins, J. Lee, and R. Babbush, Commun. Math. Phys. **404**, 629 (2023).

[437] B. Collins and P. Śniady, Commun. in Math. Phys. **264**, 773 (2006).

[438] Z. Puchała and J. Miszczak, Bull. Polish Acad. Sci. Tech. Sci. , 21 (2017).

[439] P. Weinberg and M. Bukov, SciPost Phys. **2**, 003 (2017).

[440] S. J. van Enk and C. W. J. Beenakker, Phys. Rev. Lett. **108**, 110503 (2012).

[441] S. Becker, N. Datta, L. Lami, and C. Rouzé, IEEE Transactions on Information Theory (2024).

[442] T. Gu, X. Yuan, and B. Wu, Quantum Sci. Technol. **8**, 045008 (2023).

[443] A. Acharya, S. Saha, and A. M. Sengupta, arXiv preprint arXiv:2105.05992 (2021).

[444] i. p. Bringewatt et al., .

[445] J. J. Bisognano and E. H. Wichmann, J. Math. Phys. **16**, 985 (1975).

[446] J. J. Bisognano and E. H. Wichmann, J. Math. Phys. **17**, 303 (1976).

[447] H. W. Blöte and Y. Deng, Phys. Rev. E **66**, 066110 (2002).

[448] J. Carlson, D. J. Dean, M. Hjorth-Jensen, D. Kaplan, J. Preskill, K. Roche, M. J. Savage, and M. Troyer, *Quantum Computing for Theoretical Nuclear Physics, A White Paper prepared for the US Department of Energy, Office of Science, Office of Nuclear Physics*, Tech. Rep. (USDOE Office of Science (SC)(United States), 2018).

[449] I. C. Cloët, M. R. Dietrich, J. Arrington, A. Bazavov, M. Bishof, A. Freese, A. V. Gorshkov, A. Grassellino, K. Hafidi, Z. Jacob, *et al.*, arXiv preprint arXiv:1903.05453 (2019).

[450] D. Beck *et al.*, Nuclear Physics and Quantum Information Science, Report from the NSAC QIS Subcommittee (2019).

[451] S. Catterall, R. Harnik, V. E. Hubeny, C. W. Bauer, A. Berlin, Z. Davoudi, T. Faulkner, T. Hartman, M. Headrick, Y. F. Kahn, *et al.*, arXiv preprint arXiv:2209.14839 (2022).

[452] D. Beck, J. Carlson, Z. Davoudi, J. Formaggio, S. Quaglioni, M. Savage, J. Barata, T. Bhattacharya, M. Bishof, I. Cloet, *et al.*, arXiv preprint arXiv:2303.00113 (2023).

[453] D. E. Kharzeev, Phil. Trans. R. Soc. A **380**, 20210063 (2022).

[454] A. Cervera-Lierta, J. I. Latorre, J. Rojo, and L. Rottoli, SciPost Phys. **3**, 036 (2017).

[455] S. R. Beane, D. B. Kaplan, N. Klco, and M. J. Savage, Phys. Rev. Lett. **122**, 102001 (2019).

[456] S. R. Beane and R. C. Farrell, Annals of Physics **433**, 168581 (2021).

[457] S. R. Beane, R. C. Farrell, and M. Varma, Int. J. Mod. Phys. A **36**, 2150205 (2021).

[458] N. Klco, D. H. Beck, and M. J. Savage, Phys. Rev. A **107**, 012415 (2023).

[459] N. Klco and M. J. Savage, Phys. Rev. Lett. **127**, 211602 (2021).

[460] H. L. Stormer, D. C. Tsui, and A. C. Gossard, Rev. Mod. Phys. **71**, S298 (1999).

[461] M. E. Cage, K. Klitzing, A. Chang, F. Duncan, M. Haldane, R. B. Laughlin, A. Pruisken, and D. Thouless, *The quantum Hall effect* (Springer Science & Business Media, 2012).

[462] M. A. Levin and X.-G. Wen, Phys. Rev. B **71**, 045110 (2005).

[463] M. Levin and X.-G. Wen, Phys. Rev. Lett. **96**, 110405 (2006).

[464] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006).

[465] Y. Guryanova, S. Popescu, A. J. Short, R. Silva, and P. Skrzypczyk, Nat. Commun. **7**, 12049 (2016).

[466] N. Yunger Halpern, P. Faist, J. Oppenheim, and A. Winter, Nat. Commun. **7**, 1 (2016).

[467] M. Lostaglio, D. Jennings, and T. Rudolph, New J. Phys. **19**, 043008 (2017).

[468] N. Y. Halpern, J. Phys. A: Math Theor. **51**, 094001 (2018).

[469] N. Yunger Halpern, M. E. Beverland, and A. Kalev, Phys. Rev. E **101**, 042117 (2020).

[470] K. Fukai, Y. Nozawa, K. Kawahara, and T. N. Ikeda, Phys. Rev. Res. **2**, 033403 (2020).

[471] S. Popescu, A. B. Sainz, A. J. Short, and A. Winter, Phys. Rev. Lett. **125**, 090601 (2020).

[472] N. Yunger Halpern and S. Majidy, npj Quant. Inf. **8**, 10 (2022).

[473] F. Kranzl, A. Lasek, M. K. Joshi, A. Kalev, R. Blatt, C. F. Roos, and N. Y. Halpern, arXiv preprint arXiv:2202.04652 (2022).

[474] G. Manzano, J. M. Parrondo, and G. T. Landi, PRX Quantum **3**, 010304 (2022).

[475] Y. Mitsuhashi, K. Kaneko, and T. Sagawa, Phys. Rev. X **12**, 021013 (2022).

[476] S. Majidy, A. Lasek, D. A. Huse, and N. Y. Halpern, Phys. Rev. B **107**, 045102 (2023).

[477] S. N. Hearth, M. O. Flynn, A. Chandran, and C. R. Laumann, arXiv preprint arXiv:2306.01035 (2023).

[478] S. N. Hearth, M. O. Flynn, A. Chandran, and C. R. Laumann, arXiv preprint arXiv:2311.09291 (2023).

[479] K. Van Kirk, J. Cotler, H.-Y. Huang, and M. D. Lukin, arXiv preprint arXiv:2212.06084 (2022).

[480] V. Vitale, A. Elben, R. Kueng, A. Neven, J. Carrasco, B. Kraus, P. Zoller, P. Calabrese, B. Vermersch, and M. Dalmonte, SciPost Phys. **12**, 106 (2022).

[481] A. Rath, V. Vitale, S. Murciano, M. Votto, J. Dubail, R. Kueng, C. Branciard, P. Calabrese, and B. Vermersch, PRX Quantum **4**, 010318 (2023).

[482] R. W. Chien and J. Klassen, arXiv preprint arXiv:2210.05652 (2022).

[483] S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge University Press, New York, 2004).

[484] A. Chiruvelli and H. Lee, J. Mod. Opt. **58**, 945 (2011).

[485] M. Dahleh, M. A. Dahleh, and G. Verghese, Lectures on dynamic systems and control (2004).

[486] I. C. F. Ipsen and R. Rehman, SIAM J. Matrix Anal. Appl. **30**, 762 (2008).

[487] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. J. Heinzen, Phys. Rev. A **50**, 67 (1994).

[488] K. Kechedzhi and V. N. Smelyanskiy, Phys. Rev. X **6**, 021028 (2016).

[489] S. V. Isakov, G. Mazzola, V. N. Smelyanskiy, Z. Jiang, S. Boixo, H. Neven, and M. Troyer, Phys. Rev. Lett. **117**, 180402 (2016).

[490] V. I. Arnold, Selecta Mathematica **1**, 1 (1995).

[491] N. Hunter-Jones, arXiv preprint arXiv:1905.12053 (2019).

[492] D. Gross, K. Audenaert, and J. Eisert, J. Math. Phys. **48**, 052104 (2007).

[493] R. A. Low, arXiv preprint arXiv:1006.5227 (2010).

[494] P. Dulian and A. Sawicki, IEEE Trans. Inf. Theory **70**, 2637 (2024).

[495] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.shgo.html,`.