Accelerating Learning of Quantum Systems using Prior Information

by

Arkopal Dutt

B.Tech. (Honors) Aerospace Engineering, Indian Institute of Technology, Bombay, 2015
S.M. Mechanical Engineering, Massachusetts Institute of Technology, 2018

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2024

Authored by:	Arkopal Dutt Department of Mechanical Engineering December 22, 2023
Certified by:	Isaac L. Chuang Professor of Physics Professor of Electrical Engineering and Computer Science Thesis Supervisor
Accepted by:	Nicolas Hadjiconstantinou Professor of Mechanical Engineering Chairman, Department Committee on Graduate Theses

Accelerating Learning of Quantum Systems using Prior Information

by

Arkopal Dutt

Submitted to the Department of Mechanical Engineering
on December 22, 2023 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

Towards realizing practically useful quantum devices, the sizes of quantum devices are being scaled up. An imminent challenge to scalability is ensuring resource requirements of learning tasks that occur as part of device characterization and execution of quantum algorithms also scale favorably. Resource requirements in general grow exponentially with the system size or the number of qubits and at the standard quantum limit (SQL) with respect to the learning error. On the other hand, much is known about the quantum system, which is of known construction with prior experience. A natural question is then, can we accelerate learning of quantum systems using prior information? In this thesis, we describe how prior information can be exploited to reduce resources over current baseline methods.

In the first part, we consider the problem of quantum state tomography and identify a class of quantum states that are hard to simulate classically, to be learnable in sample complexity growing polynomially in the number of qubits. Our learning algorithm can be used to verify circuits commonly used in quantum advantage experiments. In the second part, we consider the problem of discriminating quantum channels on a physical system under experimental constraints of limited control and lack of direct readout. After introducing an ancillary measurement system that weakly interacts with the physical system, we show sequential protocols adapted to this setting outperform multi-shot and parallel protocols, achieving learning rates faster than SQL. In the third part, we consider the common recurring task of Hamiltonian learning during calibration. We introduce a batch-mode Hamiltonian active learner (HAL) that proposes informative queries adaptively during learning. In our experiments on an IBM quantum device, HAL reduced resources by 95% compared to standard methods and by 33% compared to a sequential active learner. In the fourth part, we consider the problem of estimating the expectation value of a Hamiltonian with respect to a quantum state, which features in many hybrid quantum-classical algorithms for ground state energy estimation in quantum chemistry. To guide the selection of of measurement methods designed for this problem, we propose a benchmark that assesses their performance against a set of common molecular Hamiltonians and common states. Benchmarking on IBM quantum devices reveal that decision diagrams are preferred for near-term quantum hardware. Finally in the fifth part, we propose a quantum algorithm based on molecular bootstrap embedding for ground state estimation of large molecular Hamiltonians that could potentially take advantage of access to multiple smaller quantum computers.

Thesis supervisor: Isaac L. Chuang
Title: Professor of Physics, Professor of Electrical Engineering and Computer Science

Acknowledgments

First and foremost, I feel an enormous sense of gratitude to my advisor, Professor Isaac (Ike) Chuang for his mentorship, guidance and support in both academic and personal settings! I still fondly remember my first meeting with Ike in Fall 2018. I had just returned from my summer internship at Los Alamos National Laboratory (LANL) and I was convinced that I wanted to work in the area of quantum computation for my PhD research despite not having taken a single graduate class in quantum mechanics yet. Ike encouraged me to pursue research in the field and introduced me to the *quantum drum* problem which would later become my first PhD research project with him. Thank you Ike for your patience, time and advice throughout the years. Thank you for the multiple Star Wars analogies during our discussions and adapting to how I think. At the same time, thank you for teaching me how to go a little beyond being curious about the world around us and asking the right questions. I had an incredible experience learning from you.

I would like to thank my committee for their comments and feedback during our meetings together. In particular, I would like to thank the chair of my committee, Professor Vivishek Sudhir for meeting with me repeatedly over the last two years to discuss my research work. Many thanks to Professor George Barbastathis for their insightful questions that helped me dive deeper into my research results. I owe the deepest gratitude to Dr. Sarah Sheldon who was on my committee as well as my mentor during my internships at IBM Research. Sarah, thank you for the multiple meetings we have had over the years and for sharing your unique perspective on research. I would be remiss if I did not mention my research collaborators. I am deeply indebted to Edwin Pednault, Lev Bishop, John Smolin, Chai Wah Wu, Sho Sugiura, Srinivasan Arunachalam, Ted Yoder, Sergey Bravyi, William Kirby, Antonio Mezzacapo, Yuan Liu, Zachary Chin, Max Tao, Rudy Raymond, and Charles Hadfield. Srinivasan, thank you for answering my numerous questions on computational learning theory and giving me a reason to still use Google Hangouts/Chat. Will, thank you for all your help in running experiments and increasing my knowledge of birds not inside IBM Quantum's dilution refrigerators. It was also great exploring Portugal with you and Patrick during TQC.

During my time at MIT, I have had the opportunity to be a teaching assistant (TA) for two semesters. I am grateful to Professor Alberto Rodriguez for giving me a chance to TA for Dynamics and Controls (2.003) when I was transitioning from one research lab to another. Alberto, thank you so much for providing a unique perspective on the subject and looking out for me. I also want to thank Professor Rohan Abeyaratne and Professor Julie Shah for multiple helpful conversations when I was switching research areas and labs for my PhD. I also want to convey many many thanks to Leslie Regan, Saana McDaniel, Joanna

Welch, and Janice McCarthy for helping me through various logistical challenges. Leslie, thank you for helping me overcome many obstacles in the first two years of graduate school. Saana, thank you so much for everything you do!

I strongly believe that my graduate school experience would not have been as fun as it was if not for the remarkable people I met along the way and the groups I was a part of. I was fortunate to be a part of the MIT SIAM Student Chapter. Thank you to Mo, Corbin, Saviz, Rohit, Adam, and Nisha, for making this the most fun academically oriented group I was a part of. Thank you for going all out with me in organizing seminars, lightning talks and socials. Lauren, thank you for advising us along the way. I should also mention at this time that I owe most of the numerical results in thesis to Lauren and the rest of the MIT Supercloud team for providing computational resources. Thank you to my lab - Andrew, Zane, John C, John M, Jasmine, Curtis, Gabe, Felix - for making the environment so enjoyable to work in.

Surprisingly or unsurprisingly for those that know me well, thank you to "Cafe Spice Rocks"! Ravi and Rohit, we met in the first week of graduate school and never looked back. We have had many biryani nights, Punjabi Dhaba bike rides, terrible movie outings and so many conversations sprinkled with masala over the years. To Ravi, thank you for always being there to lend a ear and telling me about the LANL opportunity all those years ago that sparked my PhD journey. Rohit, thank you for slowing down when we walk together, for appreciating the importance of chicken korma in a diet, for being my remote work buddy during the COVID-19 pandemic and keeping me up to date with Insta posts. Saviz, thank you mon ami for being my swimming (aka gossip in the P pool) buddy, for being the calm skiing instructor, introducing me to the finer things in life such as dimsum, and being so positive about where life takes us. Cindy, thank you for introducing me to the other finer thing in life, hotpot, and making sure our adventure plans are pragmatic.

Mo, my roomie, thank you for being my constant buddy throughout this PhD journey. It has been a wild roller coaster and you have been there all the way. Thank you to joining me in the crazy late night taco runs, weekend house-hackathons, and spontaneous trips. Corbin, thank you for sharing my love for fried chicken, being my other swimming buddy, for not taking me too seriously and being the second best storyteller I know. Jvo, thank you for being just a floor away until you moved incredibly far away, for all the Harry Potter marathons and getting me excited about talking about space! Sydney, thank you for introducing me to Flour early on in graduate school but more importantly, introducing me to new board games and being the most fantastic storyteller I know. Joseph, thank you for telling me how those nuts and bolts in robots work. Thank you Nigamaa, Karthik GK and Rushina for being my spontaneous lunch buddies. Thank you to Rafa, Seb and Sachin for boosting spirits in 468! I am also thankful for the friendship of the Sidpac crew: Sandra, Greg, Nicholas, AJ Fenton, Haozhe, Eric, Will, Jenny, Fabian and Player 1 Sandro. Wael and Jing, thank you both for the multiple conversations and dinner outings we have had. Shaswat, thank you for understanding it is important that Nadal wins and all the sport outings we have had over the years. Ashwin, thank you for being one quick phone call away and for supporting me throughout my undergraduate and graduate journeys.

Katie, thank you putting up with me over the last few months as I worked towards finalizing my last paper and thesis. Thank you for being my life, squash, and travel partner. We have had some cool adventures in the past year and I am very much looking forward to

what's next. Murray, you lil bugger, thank you for meowing incessantly so I never forget to feed you, being your general dorky self, and my constant mate during my research night-ins.

And, of course, I would not be where I am today without the support of my parents. Thank you, Mama and Baba; everything I have managed to accomplish so far has been in no small part due to your love, sacrifice and support! Finally, thank you to Dada and Didi for always believing in me and inspiring me.

Contents

Title page **1**

Abstract **3**

Acknowledgments **5**

List of Figures **15**

List of Tables **27**

1 Introduction **31**
1.1 Thesis motivation . . . 31
1.1.1 Thesis question . . . 33
1.2 Learning tasks across the life cycle of quantum computing . . . 34
1.2.1 Quantum system characterization . . . 35
1.2.2 Implementing quantum programs and algorithms . . . 37
1.3 Thesis contributions . . . 39
1.4 Thesis organization . . . 44

2 Efficiently learning quantum phase states **47**
2.1 Introduction . . . 47
2.2 Our contributions . . . 49
2.3 The intuition . . . 51
2.4 Preliminaries . . . 56
2.4.1 Notation. . . . 56
2.4.2 Boolean functions . . . 56
2.4.3 Phase states . . . 57
2.4.4 Useful lemmas . . . 58
2.4.5 Measurements . . . 59
2.5 Learning binary phase states . . . 61
2.5.1 Learning algorithm using separable measurements . . . 61
2.5.2 Learning using entangled measurements . . . 64
2.5.3 Lower bounds . . . 65
2.6 Learning sparse and low Fourier-degree binary phase states . . . 68
2.6.1 Sparse learning algorithm . . . 68
2.6.2 Learning low Fourier-degree phase states . . . 70

2.7 Learning generalized phase states . . . 71
2.7.1 Learning using separable measurements . . . 71
2.7.2 Learning stabilizer states . . . 76
2.8 Learning phase states under depolarization . . . 76
2.8.1 Global depolarizing noise . . . 76
2.8.2 Local depolarizing noise . . . 78
2.8.3 Local depolarizing noise and small graph degree . . . 79
2.9 Applications . . . 79
2.9.1 Learning quantum circuits producing binary phase states . . . 79
2.9.2 Learning circuits containing diagonal gates in the Clifford hierarchy . 80
2.10 Discussion . . . 83

3 Heisenberg-limited *hidden* quantum channel discrimination 85
3.1 Introduction . . . 85
3.2 Hidden binary channel discrimination (HBCD) . . . 87
3.2.1 Problem statement . . . 87
3.2.2 Protocols for HBCD . . . 87
3.2.3 Estimators . . . 89
3.3 Performance guarantees for protocols . . . 89
3.4 Heisenberg-limited learning by sequential protocol . . . 95
3.4.1 Lower bound . . . 95
3.4.2 Numerical experiments . . . 96
3.5 Practical operation . . . 97
3.6 Conclusion . . . 100

4 Active learning of quantum system Hamiltonians 101
4.1 Introduction . . . 101
4.2 Hamiltonian Learning . . . 103
4.2.1 Problem Statement . . . 103
4.2.2 Learning in the Presence of Noise . . . 109
4.2.3 Active Learning of Hamiltonians . . . 111
4.3 Hamiltonian active learning algorithms . . . 117
4.3.1 Algorithm . . . 118
4.3.2 Comment on Query Advantage and Heisenberg Limited Scaling . . . 120
4.3.3 Computational Cost and Extensions . . . 121
4.4 Hamiltonian learning for a two-qubit superconducting cross-resonance gate: Model and setup . . . 121
4.4.1 Cross-Resonance Hamiltonian . . . 122
4.4.2 Experimental Setup . . . 124
4.4.3 Estimates of Noise and Nonidealities for Experimental System . . . 125
4.4.4 Estimation Procedures for Learning CR Hamiltonians . . . 131
4.4.5 Implementation of HAL Algorithm for Learning CR Hamiltonians . . . 134
4.5 Results . . . 135
4.5.1 Data and Experiment Protocol . . . 136
4.5.2 Sequential Active Learner: Qinfer . . . 139

4.5.3 Performance of Hamiltonian Learning Methods . . . 140
4.5.4 Analysis . . . 146
4.6 Conclusion . . . 152

5 Benchmarking randomized measurements of quantum chemsitry Hamiltonians 154
5.1 Introduction . . . 154
5.2 Background . . . 158
5.2.1 Notation . . . 158
5.2.2 Observable estimation: Learning problem of measuring quantum Hamiltonians . . . 159
5.2.3 Strategy for CSHOREBench . . . 159
5.2.4 General measurement protocol . . . 160
5.2.5 Summary of performance metrics in CSHOREBench . . . 164
5.3 Estimators . . . 165
5.3.1 Monte-Carlo Estimator . . . 165
5.3.2 Weighted Monte-Carlo Estimator . . . 166
5.3.3 Bayesian Estimator . . . 167
5.4 Measurement Methods . . . 168
5.4.1 Randomized Measurements . . . 169
5.4.2 Adaptive Pauli Measurements . . . 171
5.4.3 Derandomization . . . 172
5.5 CSHOREBench: Common States and Hamiltonians for ObseRvable Estimation Benchmark . . . 174
5.5.1 Common quantum computation task and common objects . . . 175
5.5.2 Performance metrics in CSHOREBench . . . 176
5.5.3 Experimental protocol for CSHOREBench . . . 177
5.6 Results . . . 179
5.6.1 Comparison of estimators . . . 179
5.6.2 Experiments on classical simulator . . . 180
5.6.3 Experiments on quantum device . . . 186
5.7 Conclusion . . . 189

6 Quantum Bootstrap Embedding: A distributed quantum algorithm for ground state energy estimation 191
6.1 Introduction . . . 191
6.2 Ideas of Bootstrap Embedding . . . 193
6.2.1 Fragmentation and Embedding Hamiltonians . . . 193
6.2.2 Matching Electron Densities: an Optimization Problem . . . 197
6.2.3 Resource Requirement and Typical Behavior of BE on Classical Computers . . . 199
6.2.4 The Quest for BE on Quantum Computers . . . 201
6.3 Quantum Bootstrap Embedding Methods . . . 202
6.3.1 Fermion-Qubit Mapping - Global Symmetry vs. Locality . . . 203
6.3.2 Naive RDM Linear Matching and its Disadvantage . . . 203

6.3.3 Coherent Quantum Matching from SWAP Test . . . 204
6.3.4 Optimization Using the Quadratic Penalty Method . . . 206
6.4 Quantum Bootstrap Embedding Algorithms . . . 208
6.4.1 The Algorithm . . . 208
6.4.2 Eigensolver Subroutines and Sampling Complexity . . . 210
6.4.3 Additional Quadratic Speedup . . . 211
6.5 Results and Discussions . . . 212
6.5.1 Convergence of QBE in Infinite Sampling Limit . . . 213
6.5.2 Sampling Advantage of Coherent Quantum Matching . . . 214
6.6 Conclusion . . . 217

7 Conclusion and Outlook 220

A Appendix to Chapter 3 223
A.1 Perspective on measurement protocols for conventional QCD . . . 223
A.2 Relation to Quantum Signal Processing . . . 224
A.3 Numerical Experiments . . . 224

B Appendix to Chapter 4 227
B.1 Details of Cross-Resonance Hamiltonian . . . 227
B.1.1 Description of IBM Quantum Devices . . . 227
B.1.2 Experimental Implementation of Query Space . . . 228
B.1.3 Modeling Pulse Shapes . . . 229
B.1.4 Likelihood Function and Fisher Information Matrix for the CR Hamiltonian . . . 231
B.2 Computational Details of Query Optimization . . . 234
B.2.1 Different Query Optimizations and Strategies for Handling Query Constraints . . . 234
B.2.2 Uncertainty Filtering of Query Space . . . 237
B.3 Estimation Procedure for Learning Cross-Resonance Hamiltonians . . . 238
B.3.1 Regression . . . 239
B.3.2 Maximum Likelihood Estimation . . . 240
B.3.3 Energy Landscapes of Negative Log-Likelihood Loss for Cross-Resonance Hamiltonian . . . 241
B.3.4 Incorporating uncertainty from shot noise . . . 242
B.4 Learned Hamiltonian Parameters and Learning Error on IBM Quantum Devices 243
B.4.1 Summary of Model Parameters on IBM Quantum Devices . . . 243
B.4.2 Expected trends of learning error . . . 245
B.4.3 Sparse Query Distributions . . . 247
B.5 Heisenberg Limited Scaling in Cross-Resonance Type Hamiltonians . . . 247
B.5.1 Examples . . . 248
B.5.2 Examples of non-HLS scaling during Hamiltonian learning . . . 250

C Appendix to Chapter 5 253
C.1 Details of derandomizing decision diagrams . . . 253

C.2 Details of molecules and measurement methods . . . 257
C.2.1 Pauli weight distributions . . . 257
C.2.2 Query distributions for different molecules . . . 258
C.2.3 Convergence behavior of Adaptive Pauli Shadows . . . 259

D Appendix to Chapter 6 **260**
D.1 Construction of the Embedding Hamiltonians . . . 260
D.2 Basis Transformation Unitary from Localized to Canonical Molecular Orbitals 260
D.3 Proof of Equivalence of the Linear and Quadratic Constraint . . . 262
D.4 Estimating Quadratic Penalty from Subsystem SWAP Test . . . 263
D.4.1 Quantum Circuit of the SWAP Test . . . 263
D.4.2 Ancilla Measurement Probability . . . 263
D.4.3 Connection to Quadratic Penalty . . . 265
D.5 Eigenvalue Equations for the Quadratic Penalty Method . . . 265
D.5.1 Derivative $\frac{\partial \mathcal{L}_A}{\partial C_{ai\mu}}$ . . . 266
D.5.2 Derivative $\frac{\partial C_{ai\mu}}{\delta V_{\mathrm{BE}}}$ . . . 268
D.5.3 Gradient of Cost Function versus BE Potential . . . 270
D.6 Sample Complexity for Estimating the Overlap from Tomography . . . 270
D.7 Details of Quantum Amplitude Estimation and Quadratic Speedup . . . 272
D.7.1 Amplitude Amplification . . . 273
D.7.2 Estimate the Amplitude from Binary Search . . . 275
D.7.3 Quadratic Speedup . . . 276
D.8 QBE Algorithm Using Naive RDM Linear Matching . . . 277
D.9 Computational Details . . . 279
D.9.1 FCI and QPE Eigensolver Runtime Benchmark . . . 279
D.9.2 Classical Bootstrap Embedding with VMC and FCI as Eigensolver . . . 280
D.9.3 SWAP Test Circuit in Quantum Bootstrap Embedding . . . 281
D.9.4 Quantum Bootstrap Embedding Calculation . . . 281
D.9.5 Details of VQE Eigensolver . . . 284

References **288**

List of Figures

1.1 Trend of size of quantum computers (in number of qubits) with year. Source of numbers: IBM Quantum roadmap [22a], Google Quantum AI roadmap [22b] and IonQ roadmap [20]. 32

1.2 Overview of different learning tasks as we move forward in the life cycle of quantum computing starting from development of devices to deployment in useful applications. 34

1.3 Connectivity map of 1121-qubit IBM quantum device Condor with nodes denoting different qubits and edges between pairs of qubits indicating direct connections. Edges directly correspond to presence of cross-resonance Hamiltonian interaction between pairs of qubits. Figure adapted from [Pat+23]. . 42

3.1 Comparison between conventional quantum channel discrimination (QCD) and hidden quantum channel discrimination (HQCD). Here the black boxes indicate the unknown channels and state. In both cases, the action of unknown channel C is inferred by selecting an input state ρ_m, applying controlled V_n operations ($n = 1, \ldots, N$), and measuring with M to minimize the error probability. (a) Conventional QCD, involving the direct manipulation and measurement of the system. (b) HQCD, where the physical system $\mathcal{H}$ and measurement system $\mathcal{M}$ are explicitly distinguished. 86

3.2 The query and sequential/multi-shot/parallel protocols. (a) Query Q_n (Def. 4) with phases ψ_n and ϕ_n specified independently. The upper (hidden) qubit undergoes unitary evolution every round and at the end we measure the lower (measurement) qubit. (b-d) Discrimination protocol S: (b) Sequential protocol, (c) Multi-shot protocol with depth $d = 2$, (d) Parallel protocol. ρ_m can be a highly-entangled state for the parallel protocol. 88

3.3 Number of queries N sufficient for solving $\mathsf{HBCD}(\alpha, \epsilon, \rho_h = \mathbb{I}/2)$. For the sequential protocol (Σ_S), we measure only once and use a phase sequence Φ of length N. For the multi-shot protocol (Σ_M), we use queries of depth $d = 4$ and measure multiple times. Trends for different values of $\epsilon \in \{0.005, 0.025, 0.05\}$ are shown for Σ_S and $\epsilon \in \{0.025, 0.005\}$ for Σ_M. 96

3.4 Quantum detector operating characteristics (QDOC) of the sequential protocol in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$ with increasing length N of phase sequence Φ. 98

3.5 Quantum detector operating characteristics (QDOC) of the multi-shot protocol with a fixed query depth of $d = 8$ in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$ with increasing number of shots m. (a) Trend of P_D with P_F considering linear scales on both x-axis and y-axis. (b) Trend of P_D with P_F for intermediate operating points obtained by each protocol considering a log-scale on the x-axis to illustrate that P_F remains orders of magnitude below 1 for all values of P_D. 99

3.6 Quantum detector operating characteristics (QDOC) of a sequential protocol (Σ_S) and multi-shot protocol (Σ_M) with a fixed budget of $N = 16$ in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$. For the multi-shot protocol, we show OCs with increasing query depth d and decreasing number of shots m such that $N = d \cdot m = 16$. (a) Trend of P_D with P_F considering linear scales on both x-axis and y-axis. (b) Trend of P_D with P_F for intermediate operating points obtained by each protocol considering a log-scale on the x-axis to illustrate that P_F remains orders of magnitude below 1 for all values of P_D. 100

4.1 Quantum circuit picture of how the query $x = (M, U, t)$ is used. The input to the quantum circuit is the zero state $|0\rangle^{\otimes n}$. Application of the preparation operator U on $|0\rangle^{\otimes n}$ is used to create an initial state before interacting with the system Hamiltonian H for time t. Finally, a measurement operator M is applied to the evolved state before measuring in the usual computational basis denoted by the meter. The output is a single shot of qubits being readout. 105

4.2 Schematic of Hamiltonian learning with an active learner: The *oracle* constitutes the unknown Hamiltonian and noise sources, with the true parameter vector of $\boldsymbol{\theta}^\star$ which is unknown and needs to be learned. Here, we show the device coupling map of the 20-qubit IBM Quantum device *ibmq_ boeblingen*. Learning is carried out on training examples of the form (x, y) where x are queries inputted to the oracle specifying the measurement observable M, preparation operator U and system time evolution t, and y are the corresponding measurement outcomes outputted by the oracle. An illustrative quantum circuit picture of the oracle in the noiseless case is shown in Figure 4.1. The set of queries inputted to the oracle is denoted as X and the corresponding set of measurement outcomes outputted by the oracle as Y. An estimation procedure is run on the training examples of (X, Y) to learn a model parameter estimate $\hat{\boldsymbol{\theta}}$. The complete top row corresponds to how a passive learner operates and is also called open-loop Hamiltonian learning. We add a feedback loop to introduce an active learner which uses the current estimate of model parameters $\hat{\boldsymbol{\theta}}$ to prescribe the distribution $\hat{q}$ from which queries to the oracle should be sampled from next. This process is then repeated during learning until an accurate estimate $\hat{\boldsymbol{\theta}}$ is obtained. The Hamiltonian Active Learning (HAL) algorithm introduced in this work comprises the estimation procedure and active learner shown in this schematic. The active learner is described in Section 4.2.3 and the HAL algorithm in Section 4.3. 106

4.3 Rabi oscillations obtained experimentally through queries to a CR Hamiltonian on the IBM Quantum device (*ibmq_boeblingen*) for different measurement operators (rows), preparation operators (markers) and evolution times t (x-axis). The set of measurement and preparation operators will be described in Section 4.4.2. The experimental data is indicated by markers and the model fit to the data by lines. The model fit was generated by learning the CR Hamiltonian (Eq. 4.31) from experimental data using MLE and predicting values of Rabi oscillations for the query space using the learned Hamiltonian. 124
4.4 Characterization of readout noise from calibration data of single qubit readouts considering a bit-flip channel model using (a) a trained Bayesian naive classifier, and (b) fitting Gaussian distributions. In (a)-(b), the experimental data points of the complex readout signal $\mathbf{c}$ are shown as markers. In (a), the decision boundary is shown as a line. In (b), the contours indicate the single standard deviation. 127
4.5 Examples of Rabi oscillations computed from experimental data collected from IBM Quantum device D *ibmq_boeblingen* for different measurement operators M (rows), preparation operators U (markers) and evolution times t (x-axis). These were computed (a) using a binary classifier and (b) using Gaussian distribution fits. The difference between the two approaches is negligible when the misclassification error of the binary classifier is low ($r_0, r_1 < 0.01$) as is the case here. 128
4.6 Dependence of the time offset Δt on parameters ω for IBM Quantum device D *ibmq_boeblingen*. The plotted data points correspond to driving the device under different conditions and hence different cross-resonance Hamiltonians. The imperfect pulse shaping model extracted from these experimental data points is shown by a fit and this is later used in the MLE. 129
4.7 Comparison of Rabi oscillations computed from data (46800 queries) of IBM Quantum device D *ibmq_boeblingen* from the (a) collected experimental data, and (b) simulator. In (c), we plot the difference in the model Rabi oscillations predicted from the Hamiltonians learned in (a) and (b). In the subplots of (a)-(c), we plot Rabi oscillations (or difference) for different measurement operators M (rows), preparation operators U (colors), and evolution times t (x-axis), corresponding to the query space described in Section 4.4.2. 138
4.8 Scaling of RMSE with number of queries for different learners on a simulator when MFC is satisfied. In (a), we compare the trends in RMSE of HAL-FI against the passive learner with estimations based on MLE and linear regression (Lin. Reg.). In (b), we compare the trends in RMSE of HAL-FI using the Bayesian estimator of sequential Monte Carlo (SMC) against HAL-FI using MLE, Qinfer using Bayes risk as the query criteria and passive learners equipped with different estimators. In (a)-(b), estimators of each learner are indicated in brackets. Slopes indicate the scaling of RMSE with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner. In (b), we do not show slopes and errors for learners shown earlier in (a) on the left, and their trends are shown for visual reference. 141

4.9 Scaling of testing error with number of queries for different learners on a simulator when MFC is satisfied. In (a), we compare the trends in testing error of HAL-FIR against the passive learner with estimations based on MLE and linear regression (Lin. Reg.). In (b), we compare the trends in testing error of HAL-FIR using the Bayesian estimator of sequential Monte Carlo (SMC) against HAL-FIR using MLE, and passive learners equipped with different estimators. In (a)-(b), estimators of each learner are indicated in brackets. Slopes indicate the scaling of testing error with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner. In (b), we do not show slopes and errors for learners shown earlier in (a) on the left, and their trends are shown for visual reference. 142
4.10 Scaling of RMSE with number of queries for different learners on a simulator when MFC is not satisfied. In (a), we compare the trends in RMSE of HAL-FI against the passive learner with estimations based on MLE and linear regression (Lin. Reg.). In (b), we compare the trends in RMSE of HAL-FI using the Bayesian estimator of sequential Monte Carlo (SMC) against HAL-FI using MLE, Qinfer using Bayes risk as the query criteria and passive learners equipped with different estimators. In (a)-(b), estimators of each learner are indicated in brackets. Slopes indicate the scaling of RMSE with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner. In (b), we do not show slopes and errors for learners shown earlier in (a) on the left, and their trends are shown for visual reference. 144
4.11 Scaling of RMSE for different learners on experimental data with (a) number of queries made and (b) number of mini-batches requested. The time range $\mathcal{T}$ of the query space is such that Minimum-Frequency criteria is not necessarily satisfied for frequency estimation using FFT. We compare the trends in RMSE of HAL-FI using MLE, sequential active learner of Qinfer using Bayes risk as the query criteria and passive learners equipped with different estimators. Estimators of each learner are indicated in brackets. Slopes indicate the scaling of RMSE with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner. 145
4.12 (a) Trends of RMSE with number of queries for different learners in fixed or growing query spaces (QS) on experimental data (Expt.) compared against simulator (Sim.). Slopes indicate the scaling of RMSE ϵ with number of queries N in the low query/sample and high query/sample regimes. Filled in areas indicate the respective errors on trends for each learner on the simulator. Passive learner which has similar behavior to HAL-FI in fixed query space is not shown for brevity. (b) Trends of the maximum evolution time associated with query spaces of different learners. 147

4.13 Trends of RMSE with computational cost for different learners on (a) simulator where MFC is not necessarily satisfied, and (b) experimental data. The computational cost is measured in terms of the number of likelihood evaluations during learning. 148
4.14 Query advantage (QA) of different learners on the simulator when MFC (minimum-frequency criteria) is not satisfied over the following baselines: (a) passive learner with estimation based on linear regression and (b) passive learner with the MLE estimator. In (a)-(b), the data points correspond to values of QA computed from the data points of Figure 4.10 and the lines are the fits to these values of QA. The annotated text for the different trends indicate the QA obtained for the corresponding learner over the baseline for the lowest observed value of RMSE in our numerical experiments. Filled in areas indicate the 95% confidence interval on the data points shown. 149
4.15 Query advantage (QA) of different learners over specified baselines considering the problem of (a) Hamiltonian learning without prior information and (b) Hamiltonian learning with prior information on subset of Hamiltonian parameters from previous calibrations. In (a), we plot the QA of passive learners, HAL-FI, and Qinfer on experimental data over two different baselines. In (b), we plot the QA of HAL-FI over the baseline of the passive learner on the simulator for fixed and adaptively growing query spaces (QS). In (b), the QAs obtained on the simulator closely match those obtained on the experimental data and hence the latter is not shown. In (a)-(b), the data points correspond to values of QA computed from the data points of Figures 4.11,4.12 and the lines are the fits to these values of QA. The annotated text for the different trends indicate the QA obtained for the corresponding learner over the baseline for the lowest observed value of RMSE in our numerical experiments. Filled in areas indicate the 95% confidence interval on the data points shown. 150

5.1 Prescription for benchmarking a candidate measurement protocol (measurement method Σ_M and estimator $\mathcal{E}$) on a Hamiltonian H and state ρ against measurement budget M. The benchmarking stages of setup (S), execution (E) and analysis (A) are shown for each step of a general measurement protocol. General measurement protocols, to be discussed in Section 5.2.4, can be divided into steps of classical pre-processing where measurement samples are generated, experiments on a quantum device and classical post-processing on data acquired. Each step has a benchmarking stage associated with it. As shown (left), setup defines the testing dataset, candidate algorithms and experiment design. Execution (right) defines the computational resources available for any step. Analysis (right) defines the metrics associated with each step. 157

5.2 Schematic of estimation of $\mathrm{Tr}(\rho H)$: The procedure is divided into three steps of (i) pre-processing on a classical computer (CPU), (ii) experiments on a quantum device or quantum processing unit (QPU), and (iii) post-processing on the CPU. In (i), the measurement method plays the central role of deciding which measurement bases (denoted by B here) to execute on the QPU given inputs of an n-qubit Hamiltonian and measurement budget M. In (ii), experiments are executed on the QPU using the inputs of the Pauli measurement bases output by step (i). A measurement circuit corresponding to an arbitrary basis B is shown inset (Had denotes the Hadamard gate, and $S^\dagger$ denotes the inverse phase gate, which are used to transform the local measurement basis). Finally, in (iii) estimation is carried out on measurement results of the form (B, y) where B are the Pauli measurement bases and y are the corresponding measurement outcomes from the QPU. 161
5.3 Instances of different decision diagrams for H_2 (4 qubits, sto3g basis, JW encoding). We show (a) uniform classical shadows, (b) locally biased classical shadows (LBCS) and (c) an optimized compact decision diagram. 169
5.4 Comparison of RMSE achieved by different estimators in numerical simulations on HeH^+ cation (8 qubits, 6-31g basis, JW encoding) with measurement methods of (a) classical shadows (CS), (b) locally biased classical shadows (LBCS), and decision diagrams (DD). The state is set to be the ground state of the Hamiltonian. A common legend is shown for the subfigures. Trends of RMSE achieved by Monte Carlo (MC) with Laplace smoothing of $\gamma \in \{0, 0.5, 1\}$, weighted Monte Carlo (WMC) and Bayesian estimators on the same sets of measurements collected for each measurement method are shown. Averaged values over 200 independent runs are shown. Inset shows the standard deviation or uncertainty on these expectation values for each estimator. 180
5.5 Comparison of RMSE achieved in numerical simulations by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as the ground state and H is the Hamiltonian of (a) tapered H_2 (5 qubits, 3-21g basis, JW encoding), and (b) tapered HeH^+ (6 qubits, 3-21g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator. 182
5.6 Comparison of RMSE achieved in numerical simulations by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as the ground state and H is the Hamiltonian of (a) HeH^+ (8 qubits, 6-31g basis, JW encoding), (b) LiH (12 qubits, sto-6g basis, JW encoding), and (c) N_2 (16 qubits, sto-6g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator in (a)-(b) and the MC estimator in (c). 183

5.7 Comparison of empirical RMSE achieved in experiments on quantum device by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as a randomly fixed ansatz and H is the Hamiltonian of (a) tapered H_2 (5 qubits, 3-21g basis, JW encoding), and (b) tapered HeH^+ (6 qubits, 3-21g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator. 187
5.8 Comparison of empirical RMSE achieved in experiments on a quantum device by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as a randomly fixed ansatz and H is the Hamiltonian of (a) HeH^+ (8 qubits, 6-31g basis, JW encoding), (b) LiH (12 qubits, sto-6g basis, JW encoding), and (c) N_2 (16 qubits, sto-6g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator. 188

6.1 Schematic of bootstrap embedding on classical (left, blue arrows) and quantum (right, red arrows) computers. The arrows indicate BE iterative loops that are used to optimize the corresponding objective functions. Starting from panel (i) (upper center), the original system is first broken into overlapping fragments (Fragmentation), where each fragment is solved using a classical (iic) (upper left) or quantum eigensolver (iiq) (upper right). In classical matching, the 1-electron reduced density matrices (1-RDM) on the overlapping sites of adjacent fragments are used to obtain the matching condition (iiic) (lower left), while in the quantum case a coherent matching protocol based on `SWAP` tests of overlapping sites combined with a single qubit measurement (iiiq) (lower right). The matching results are then used by classical computers to generate the bootstrap embedding potential V_{BE} (iv) (lower center) and the updated fragment embedding Hamiltonian $H_{\mathrm{emb}} + V_{\mathrm{BE}}$ (back to panel (i) in order to minimize a target objective function $\mathcal{L}$ in both classical and quantum case. 196
6.2 Typical convergence of density mismatch with respect to the number of eigensolver calls in classical bootstrap embedding with a deterministic eigensolver (FCI, black circle) and a stochastic eigensolver (VMC) with different number of samples (grey, blue, and orange solid lines). The horizontal dashed lines shows the final plateaued value of the density mismatch for VMC, while the FCI data converges to 10^{-6} after 700 eigensolver calls (not shown on the figure). The discrete jumps around 200 and 300 eigensolver calls are due to switching to the next BE iteration. The data is obtained for an H_8 linear chain under STO-3G basis. See SI Sec. D.9.2 for computational details. . . 201
6.3 Runtime (normalized) as a function of system size n for finding the ground state of a linear hydrogen chain H_n at STO-3G basis, comparing an exact classical solver (FCI, blue square) and an exact quantum solver (QPE, red circle) on real classical and quantum devices. Red (blue) dashed line shows a polynomial (exponential) fit to the QPE (FCI) runtime. Note the crossover at large system size. 202

6.4 Quantum circuit of a SWAP test between two qubits (lower, with state $|\phi\rangle$ and $|\psi\rangle$). The circuit is composed of two Hadamard gate (H), a controlled-SWAP operation in between, and a final Z-basis measurement M on an additional ancilla qubit (top), where $M = 0, 1$. 205
6.5 Convergence of the quantum bootstrap embedding algorithms on (a) density mismatch and (b) energy error for the linear constraint (pink) and quadratic penalty method (red) in the infinite sample limit for an H_8 molecule. The dashed trend lines in both panels indicate an exponential fit. 213
6.6 Sampling complexity ratio of naive density matrix tomography (TMG) and SWAP test versus number of qubits in the overlapping region for a target precision $\epsilon = 0.001$ on overlap S. The inset shows a simulated convergence of overlap (S) estimation using quantum matching (SWAP) for the case of two overlapping qubits. Data are obtained from a non-interacting chain of H_4 (see SI Sec. D.9.3 for details). 215
6.7 Number of eigensolver calls required as a function of target precision at overlap $S = 0.4$, comparing SWAP or VMC (blue) and SWAP+AE (red) estimation for the H_8 chain with STO-3G basis. The blue dashed line shows the number of samples (eigensolver calls) needed in SWAP test as derived in Eq. (6.27), while the red dashed line plots a more accurate version of Eq. (6.28) (Sec. D.7.3 of SI) with red circles highlighting a few data points spanning low to high target precisions. The blue scatter points are the number of VMC eigensolver calls required to achieve the corresponding target precision on the 1-RDM overlap estimation for the same H_8 molecule. The inset plots the number of eigensolver calls as a function of the overlap S for a fixed target precision $\epsilon = 0.001$. Note the crossover in both plots. 216
6.8 Summary of different choices of quantum eigensolvers (QES) and matching algorithms discussed in the present work, with speedup and cost labeled on each arrow accordingly. Overall, the best algorithm (QBE+SWAP+AE with exact QES) is highlighted in red. Note that approximate QES are likely to achieve exponential speedup as compared to classical FCI solver. It is however not guaranteed and depends on specific implementation and the ease of input state preparation. We therefore use "possible exponential speedup" for it. . 218

A.1 Quantum circuit used in numerical experiments for HBCD using the sequential or multi-shot protocols 225
A.2 Probability of success of the sequential protocol in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$ with increasing length N of phase sequence Φ. 226

B.1 Connectivity maps for (a) IBM Quantum device A, (b) IBM Quantum device B, (c) IBM Quantum device C, and (d) IBM Quantum device D *ibmq_boeblingen*. Each node represents a physical qubit on the chip and the presence of an edge between two nodes in the connectivity map indicates that a CR gate can be applied between these two nodes. 228

B.2 An example of a CR pulse schedule on the IBM Quantum device *ibmq_ boeblingen* considering to the query of $x = (M, U, t)$ where $M = \sigma_I \otimes \exp\left(-i\frac{\pi}{4}\sigma_X\right)$, $U = \sigma_X \otimes \sigma_I$, and time duration $t = 6 \times 10^{-7}$s. The x-axis corresponds to time normalized by $dt = 2.22 \times 10^{-10}$ (Eq. B.1). The different channels corresponding to each qubit (y-axis) are written as the type of channel (see plot legend) followed by qubit number. Qubit 0 is set to be the control qubit and 1 to be the target qubit. The envelope of the different pulses are shown in each channel. The rotations on the drive or control channels indicate virtual Z gates. An equivalent representation of the quantum circuit is shown in Fig. 4.1. 230

B.3 Hamiltonian parameters $\mathbf{J}$ as a function of the amplitude of the control pulse or drive as reported in [She+16]. 231

B.4 Visualization of queries being selected for each batch during a particular run of HAL-FI for $N_0 = 2000$ and 25 batches of size $N_b = 5000$. In (a), we plot the available number of shots for each query after batches of queries are made during active learning. In (b), we plot the query distribution for each batch during active learning. The number of shots available for each query before learning starts is $N^{(0)}_{\text{shots}}(x) = 512\,\forall x$. Half the total number of shots available in this dataset is exhausted by the end of learning. Note that the different parameter values considered for HAL-FI are for stress testing the query constraints' handling procedure and are not tuned for the HAL-FI algorithm. 237

B.5 Uncertainty filtering of $\mathcal{Q}$ as defined in Section 4.4.2 with $\mathcal{T}$ set to be a sequence of 243 linearly equispaced evolution times in $[10^{-7}, 18 \times 10^{-7}]s$. The The x-axis corresponds to the system evolution times $t \in \mathcal{T}$. The y-axis indicates the different combinations of measurement operators and preparation operators available for each query in $\mathcal{Q}$. The different preparation operators are denoted as $U_0 = \sigma_I\sigma_I$ and $U_1 = \sigma_X\sigma_I$. We consider the $\boldsymbol{\theta}^\star$ corresponding to IBM Quantum device D *ibmq_ boeblingen* under drive configuration 2, with the different noise sources of readout noise, imperfect pulse-shaping, and decoherence being accounted for. Queries colored as dark blue are retained in the filtered query space and rest is filtered out. 238

B.6 Slices of energy landscapes of the log-likelihood loss function along the different parameter components considering experimental data collected from IBM Quantum device D *ibmq_ boeblingen* under drive configuration 3 (Table 4.4) (a) using parameterization $\mathbf{J}$, and (b) using parameterization $\mathbf{\Lambda}$. In each slice of θ_i (x-axis), we fix the values of the other components as obtained through estimation and evaluate the negative log-likelihood loss (y-axis) by changing the value of θ_i. We indicate the Hamiltonian parameter estimate $\hat{\theta}_i$ as obtained through our estimation procedure by a dashed red line. 242

B.7 Dependence of the time offset Δt on parameters ω for IBM Quantum devices (a) A, (b) B, (c) C CR_{01}, and (d) C CR_{02}. The plotted data points correspond to driving the device under different conditions and hence different cross-resonance Hamiltonians. The imperfect pulse shaping model extracted from these experimental data points is shown by a fit and this is later used in the MLE. 245

B.8 Scaling of learning testing error with number of queries for Hamiltonian learning considering access to the asymptotic optimal query distribution $q(\boldsymbol{\theta}^\star)$ of HAL-FI or HAL-FIR. We show trends of (a) RMSE and (b) testing error obtained upon analysis of the lower Cramer-Rao bounds. We consider the Hamiltonian parameters of $\boldsymbol{\theta}^\star$ as determined from the experimental dataset of IBM Quantum device D *ibmq_boeblingen* under drive configuration 2. We plot the trends of learning error with number of queries for HAL-FI/HAL-FIR against the passive learner which uses the uniform distribution over $\mathcal{Q}$. . . . 246
B.9 Hamiltonian learning with acess to prior information of subset of parameters during recalibration: Scaling of RMSE with Number of Queries. We assume access to the asymptotic optimal query distribution $q(\boldsymbol{\theta}^\star)$ of HAL-FI and analysis of the Cramer-Rao Bounds. We consider the Hamiltonian parameters of $\boldsymbol{\theta}^\star$ as determined from the experimental dataset of IBM Quantum device D *ibmq_boeblingen* under drive configuration 3. 247
B.10 Asymptotic optimal query distribution $q(\boldsymbol{\theta}^\star)$ for HAL-FI with a fixed query space (Section 4.4.2) on *ibmq_boeblingen* under drive configuration 2 (Table 4.4). We consider different noise sources of readout noise, imperfect pulse-shaping, and decoherence. The y-axis indicates the different combinations of measurement operators and preparation operators available for each query in $\mathcal{Q}$. The different preparation operators are denoted as $U_0 = \sigma_I\sigma_I$ and $U_1 = \sigma_X\sigma_I$. The x-axis corresponds to $\mathcal{T}$ which is set to 81 equispaced evolution times in $[10^{-7}, 6\times10^{-7}]s$. The query distribution is color-coded according to the colormap on the right. 248

C.1 Pauli weight distributions across non-identity target Paulis in Hamiltonians of different molecules. Molecules correspond to those in Table 5.2 and as shown are (a) H_2(5 qubits, 3-21g, JW), (b) HeH^+ (6 qubits, 3-21g, JW), (c) HeH^+ (8 qubits, 6-31g, JW), (d) LiH (12 qubits, sto6g, JW), and (e) N_2 (16 qubits, sto6g, JW). 257
C.2 Query distributions in LBCS for Hamiltonians of different molecules. As the query distribution in LBCS is a product distribution, the marginal probability distribution over single-qubit Paulis are shown for each qubit. Molecules correspond to those in Table 5.2 and as shown are (a) H_2(5 qubits, 3-21g, JW), (b) HeH^+ (6 qubits, 3-21g, JW), (c) HeH^+ (8 qubits, 6-31g, JW), (d) LiH (12 qubits, sto6g, JW), and (e) N_2 (16 qubits, sto6g, JW). 258
C.3 Comparison of RMSE achieved in numerical simulations by different measurement methods including APS in estimating $\mathrm{Tr}(\rho H)$ with ρ set as the ground state and H is the Hamiltonian of (a) tapered H_2 (5 qubits, 3-21g basis, JW encoding), and (b) tapered HeH^+ (6 qubits, 3-21g basis, JW encoding). RMSE is shown with the number of samples made. The estimator for each measurement method is set to be the Bayesian estimator. 259

D.1 Subsystem `SWAP` test between the overlapping regions of fragment A and B, where a controlled `SWAP` operation is performed on A_0 and B_0. The measurement probability of the top ancilla qubit encodes information of the overlap. 263

D.2 Quantum circuit to estimate the ground state overlap between subsystems of two fragments, composed of two quantum eigensolver (QES) for two fragments ground state wave function followed by a SWAP test. The circuit in the dashed box is U in Eq. (D.73) which will be repeated multiple times during the amplitude amplification process as will be discussed in the following. 273
D.3 Quantum circuit for fixed-point oblivious amplitude amplification of the coherent quantum matching. The quantum gates in dashed box corresponds to the gates in bracket of Eq. (D.73) which needs to be repeated by $d/2$ times, where d will be determined by the slope of the amplification polynomial $\frac{1}{\Delta_k}$ (more about this in the next section). 275
D.4 Schematic for quantum phase estimation. 279
D.5 SWAP test circuit between two H_4 molecule with one overlapping site. Note that each QPE uses 8 system qubits and 5 evaluation qubits. The top qubit is the control ancilla for SWAP test. The overall circuit is composed of 27 qubits. 282
D.6 A QPE circuit with for H_4 molecule, where a Hartree-Fock initial trial state preparation circuit is also showed at the beginning. The circuit block labelled as "Hamiltonian" after the four initial Pauli X are a basis transformation unitary from canonical MO basis to localized orbitals (LOs). 283
D.7 Histogram of the measurement outcome of the 5 evaluation qubits for estimating the H_4 ground state energy with 1024 shots. Post-selection is performed on the highest peak "11011". 284
D.8 Quantum bootstrap embedding convergence of the density mismatch for the 4-qubit spin model versus the number of eigenvalue calls, comparing VQE simulation (blue cross) and classical exact diagonalization (green circle). The blue shaded area shows the standard deviation from VQE estimations. The inset shows the absolute error from the VQE estimation of the mismatch to the exact value from exact diagonalization, versus number of eigenvalue calls. The number of shots considered in each step of VQE for measuring the ansatz is fixed at 10^4. 286
D.9 Quantum bootstrap embedding on H_4 chain. (a) We compare the convergence of the density mismatch versus the number of eigenvalue calls, comparing VQE simulation (blue cross) and classical exact diagonalization (green circle). The inset shows the absolute error from the VQE estimation of the mismatch to the exact value from exact diagonalization, versus number of eigenvalue calls. The number of shots considered in each step of VQE for measuring the ansatz is fixed at 10^4. (b) We compare the difference in the magnitude of the updates of the Pauli terms (being updated or appended during QBE) in the fragment Hamiltonians between the exact solver and the VQE solver, for each pair of fragment f and neighbor nb with number of iterations. For each pair of (f, nb), we indicate the Pauli term for each trend in the legend. 287

List of Tables

2.1 Upper and lower bounds of sample complexity for exact learning of n-qubit phase states with degree-d. For precise statements of the bounds, we refer the reader to the theorem statements. 50

4.1 Query complexity of HAL-FI for Hamiltonian learning (Problem 4.2.1) under different conditions as observed on a real quantum device. The case of fixed query space corresponds to adaptivity in query space being *none*. 122

4.2 Comparison of different decoherence models in fitting Rabi oscillations inferred from experimental data collected from IBM Quantum Device D. Kullback-Leibler divergence (KL divergence) is computed as $D_{KL}(p_{\text{data}}||p_{\text{model}})$ where p_{data} is the probability inferred from data and p_{model} is that predicted from the model. 131

4.3 Summary of different learning scenarios 135

4.4 Summary of estimated CR Hamiltonian parameters for the IBM Quantum device D *ibmq_boeblingen* with different drive configurations (Config.) corresponding to amplitude (Amp.) of CR pulse. We give the Hamiltonian parameters in the parameterization $\mathbf{J}$ and the physically relevant frequency components in $\mathbf{\Lambda}$. The readout noise is defined by the parameters of r_0 and r_1 which are the conditional probabilities of bit flip given the measurement outcomes are $y = 0$ and $y = 1$ respectively (see Section 4.4.3). 137

4.5 Query advantage of various learners considering different baselines on a simulator when MFC is not satisfied. Estimators of each learner are indicated in brackets. For each combination of learner and baseline, we report query advantage at lowest value of RMSE achieved by the learner (compared to extrapolated fit of baseline) and the query advantage at lowest value of RMSE achieved by the baseline in brackets below. The errors represent one standard deviation on the value of expected query advantage. 151

4.6 Query advantage (QA) of various learners considering different baselines on experimental data. Estimators of each learner are indicated in brackets. For each combination of learner and baseline, we report query advantage at lowest value of RMSE achieved by the learner (compared to extrapolated fit of baseline) and the query advantage at lowest value of RMSE achieved by the baseline in brackets below. The errors represent one standard deviation on the value of expected query advantage and are underlined in cases where the range of query advantage should be read as [QA-error,1) instead of [QA-error, QA+error]. 151

5.1 Definition of the PEP-SEA matrix. 164
5.2 Molecules considered for benchmarking of measurement methods on the simulator and quantum device. The Jordan Wigner encoding is considered. For each Hamiltonian, the state is set to be the ground state on the simulator and a random ansatz on quantum devices. 175
5.3 Different regimes of weights for classical and quantum computers in our heuristic. 177
5.4 Comparison of resource utilization with experiments on a classical simulator by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ to achieve an accuracy of 5×10^{-3} Hartree, with H set to be different different molecular Hamiltonians (Table 5.2) and ρ as the ground state. Values shown are averages over 192 independent runs against different metrics. Metrics shown are described in Section 5.5.2. Note that the contribution to classical post-processing time is from the estimator which is set to be the Bayesian estimator for all molecules except for N_2 where it is set to be the MC estimator. Lowest values obtained for each metric against a Hamiltonian is boldfaced. 184
5.5 Comparison of predicted resource utilization on IBM Quantum devices by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ to achieve an accuracy of 5×10^{-3} Hartree, with H set to be different molecular Hamiltonians (Table 5.2) and ρ is the ground state. Predicted quantum runtime or the wall-clock that would be spent by the quantum device in executing experiments is shown for the different measurement methods based on the number of shots required to achieve 5×10^{-3} Hartree (Table 5.4) and assuming the quantum runtime would primarily be from delay between circuit executions. Estimates of resource utilization, indicated by R, here account for classical and predicted quantum computing resources via the heuristic in Eq. 5.23 and using the regimes of Table 5.3. Lowest values obtained for each metric against a Hamiltonian is boldfaced. 185

5.6 Comparison of resource utilization with experiments on IBM Quantum devices by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ to achieve an accuracy of 5×10^{-3} Hartree, with H set to be different molecular Hamiltonians (Table 5.2) and ρ is a random ansatz. Values shown are averages over 192 bootstrapped runs against different metrics. Metrics shown are described in Section 5.5.2. Number of unique circuits run contribute to classical latencies such as compilation. Number of shots per circuit (all, top 5%, bottom 5%) summarizes shot distribution across circuits and is an indication of latencies due to circuit loading on control electronics. Classical pre-processing runtime is the wall-clock time spent by measurement methods in constructing a query distribution and generating samples. Classical post-processing runtime is the wall-clock time spent by the Bayesian estimator in computing an estimate from acquired measurement outcomes. Quantum runtime is the wall-clock spent by the quantum device in executing experiments. Estimates of resource utilization, indicated by R, account for both classical and quantum computing resources via the heuristic in Eq. 5.23 and using the regimes of Table 5.3. Lowest values obtained for each metric against a Hamiltonian is boldfaced. . 189

B.1 Relevant parameters of IBM Quantum devices. The qubit used as the control or target qubit is indicated by its number in the device connectivity map shown in Fig B.1. Error per gate refers to the average infidelity of single-qubit gates implemented on that qubit. 228

B.2 Summary of estimated CR Hamiltonian parameters for the IBM Quantum devices A, B, and C. We give the Hamiltonian parameters in the parameterization $\mathbf{J}$ and the physically relevant frequency components in $\mathbf{\Lambda}$. The readout noise is defined by the parameters of r_0 and r_1 which are the conditional probabilities of bit flip given the measurement outcomes are $y = 0$ and $y = 1$ respectively (see Section 4.4.3). We show the results for CR Hamiltonians between two different pairs of qubits on Device C, specified as (control qubit, target qubit): $(0, 1)$ (CR_{01}) and $(0, 2)$ (CR_{02}). 244

C.1 Details of decision diagrams for different molecular Hamiltonians (Table 5.2). Number of paths in a decision diagram correspond to number of unique measurement circuits and contributes to classical latencies such as compilation time and circuit loading. Diagonal cost of a query distribution corresponds to the one-shot variance of the energy estimate considering ρ to be the maximally mixed state. The reported computational runtimes account for both initialization and optimization of the decision diagrams. 259

Chapter 1

Introduction

1.1 Thesis motivation

Quantum computing promises to be a revolutionizing technology with potential speedups over classical computing for simulating quantum dynamics [Fey82; Llo96; Chi+18], solving linear systems of equations [HHL09], factoring large numbers [Sho94], and optimization [WBL12; BS17]. Quantum computers could also find practical utility in scientific computation for solving classically intractable problems in computational chemistry [McA+20], accelerating drug discovery [Kas+11], and solving linear differential equations [Ber+14; Ber+17; Kro23]. In an effort to realize such a useful quantum computer, several companies have invested heavily in quantum computing [Bog+23] and the US government has increased annual funding for quantum information science [Cas20] by hundreds of millions of dollars. Just over the last decade, we have had significant advances [Bru+19; Kja+20] in qubit coherence times reaching 100μs [Che+14; Ngu+19] on superconducting architectures, gate fidelity [Bar+14; Hon+20] exceeding 0.99, device sizes [GCS17; Kel+15; Aru+19] increasing upwards of 50 qubits and higher connectivity. In Figure 1.1, we show the trends of the sizes of quantum computers in terms of the number of qubits with year. As per the roadmap of IBM [22a], we can expect quantum computers of sizes up to thousands of qubits by year 2025. These trends of current and projected quantum computer sizes suggest the emergence of a quantum Moore's law with quantum computers growing five times in size every two years.

Today's quantum computers however are marked by noise, no error correction, and limited number of sequential gates that can be applied in a circuit. This has led to the current era of quantum computing to be dubbed as that of Noisy Intermediate Scale Quantum (NISQ) computers [Pre18]. To transition from the current NISQ era to the pre-fault tolerant era of large modular quantum computers albeit still without error correction [Ang+22; 22a] and eventually to the era of quantum advantage, we will need to continue scaling the size of quantum computers which presents multiple formidable challenges. One large obstacle is overcoming experimental difficulties in controlling a large number of qubits and shielding them appropriately from the surrounding environment. Outside of fabrication and hardware design, there needs to be progress in the scalability of a dependable software stack, and quantum compilation techniques along with development of more resource efficient quantum

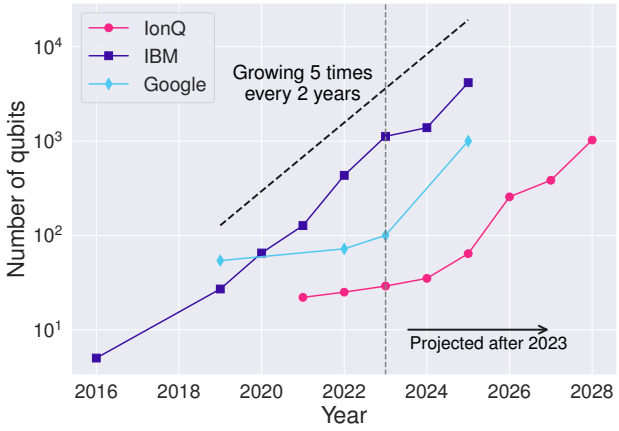


Figure 1.1: Trend of size of quantum computers (in number of qubits) with year. Source of numbers: IBM Quantum roadmap [22a], Google Quantum AI roadmap [22b] and IonQ roadmap [20].

algorithms for practical utility in the near-term.

A pressing challenge for scalability of quantum computers is that of quantum system identification which has resource requirements that scale poorly. In quantum system identification, we aim to learn features (e.g., state description, dynamics, noise) of a quantum system by carrying out experiments on the system, involving queries by applying signals and taking measurements. Unfortunately, in general, the number of such experiments or queries, also called query complexity, scales exponentially with the number of qubits [Haa+17; Leu00; NC10]. This has implications for a host of learning tasks across quantum computation where quantum system identification is carried out. For example, at the very start of the life of a quantum computer after hardware design and manufacturing, we carry out characterization to learn the quality of qubits implemented and learn the underlying Hamiltonian dictating the dynamics of qubits [She+16]. This can then be used to devise strategies for driving the Hamiltonian and implementing gates. Once a set of universal gates can be executed on the device, it is important to carry out benchmarking to obtain device level metrics of quality [Cro+19], speed [Wac+21] and reliability [McK+23]. This can then be used to suggest the types of quantum algorithms that could be implemented on the quantum computer. Moreover, quantum algorithms typically have learning tasks associated with them. Quantum algorithms output quantum state encoded solutions but the goal may be to obtain relevant classical information by performing measurements on the state. The learning task is then to learn properties of the quantum state.

Quantum system identification is thus an outstanding problem that appears across the operational life cycle of quantum computing from conceptualization of the quantum device to its deployment for useful applications. Applicability of the quantum computer depending on its power and reliability then has implications for the designs of future quantum computers. While this is similar to the design process of classical computing as well, the main distinction is that quantum devices are short-lived compared to classical devices. Due to noise from the environment, quantum devices maintain their quantum behavior only for a certain time duration during which we can execute quantum algorithms before quantum decoherence kicks

in. Moreover, the state of a quantum system collapses after measurement. The functioning of quantum devices is thus inherently cyclic, with the quantum computer having to be reset after execution of an experiment and measurement. Quantum system identification is also more challenging than classical system identification where complexity generally scales polynomially with the underlying dimension.

To see how we could accelerate learning of quantum systems by reducing resource requirements, we take inspiration from the classical setting. Just as in the case of quantum computers (Figure 1.1), sizes of classical computers have been growing at the Moore's law [Moo98] over the past decades with the number transistors in an integrated circuit doubling every two years. One relevant example of a classical system identification task that grows exponentially with the size of classical computers is that of experimental verification of transistor-level circuits [KSL95]. This task involves verification of the encoded Boolean logic and the resource requirements grows exponentially with the size of circuits. However, resource requirements of verification are brought down by using prior information available due to design choices, access to structural information of the encoded Boolean logic and symmetry of transistor-level circuits across the processor.

There are many such examples in classical system identification where resource requirements of learning algorithms have been reduced through the incorporation of different prior information [Lin94; SM71; ÅE71; Joh96; Wah91] available. The prior information could be knowledge of model structure, valid range of system parameters, physical constraints on the system from conservation laws, and the system's behavior in the frequency domain. In learning classical systems from observations of its dynamics or behavior, prior information about the expected structure of the system model is often imposed as constraints. This could include information about the presence of certain dynamics (e.g., delays), and order of interactions or even known interconnections between subsystems. Prior knowledge about the range or approximate values of system parameters can be used to initialize the parameter estimation in optimization algorithms to help them converge more quickly and find better local minima. Further, one can use available knowledge from the underlying physics of the system to restrict the feasible parameter space. For example, constraints on the positivity of certain parameters (e.g., masses or damping coefficients in mechanical systems) or adherence to physical laws (e.g., mass conservation, energy conservation, boundary conditions, etc.) can be imposed. Finally, prior information about the system's behavior in the frequency domain can be utilized (e.g., dominant frequencies, resonant peaks, etc.). Algorithmically, Bayesian methods have been used to incorporate prior information in a probabilistic framework and regularization methods such as Tikhonov regularization have been used during learning to incorporate prior beliefs about the simplicity of the system and improve robustness of the learning algorithms.

1.1.1 Thesis question

This motivates the following thesis question which we will tackle throughout this thesis:

How can we exploit prior information to accelerate learning of quantum systems across the life cycle of quantum computation?

Let us parse the above question by first describing what prior information may be available and then what is meant to accelerate learning of quantum systems. As for classical systems, prior information on quantum systems may be presented to us, the learners, in different ways and at different stages of the life cycle of quantum computing. We may have prior information on the structure of the quantum system, space of system parameters, and physical constraints (e.g., symmetry, invariance, etc.). During the implementation of quantum algorithms, we often need to learn properties of the outputted quantum state via measurements of some observables. The prior information in such situations may be the structure of these observables themselves.

To accomplish any learning task, we need to expend resources including the number of queries to the quantum system (also called sample complexity) and classical processing (also called computational complexity). To accelerate learning of quantum systems, the goal is to then reduce the resource requirements of different learning tasks. This can be achieved by improving the corresponding sample or computational complexities. Additionally, queries to the quantum system may involve highly entangling operations which may be beneficial for learning in terms of reduced sample complexity but generally comes at the cost of increased classical computational complexity. This implies the appearance of quantum and classical trade-offs. Lastly, even if prior information is not initially available during learning, we might be able to build information about the system in real-time during learning and take advantage of adaptivity as part of the learning algorithm to reduce resource requirements.

We now define the learning tasks across the life cycle of quantum computing that will be of interest to us. We then describe the scope, methodology and contributions of this thesis.

1.2 Learning tasks across the life cycle of quantum computing

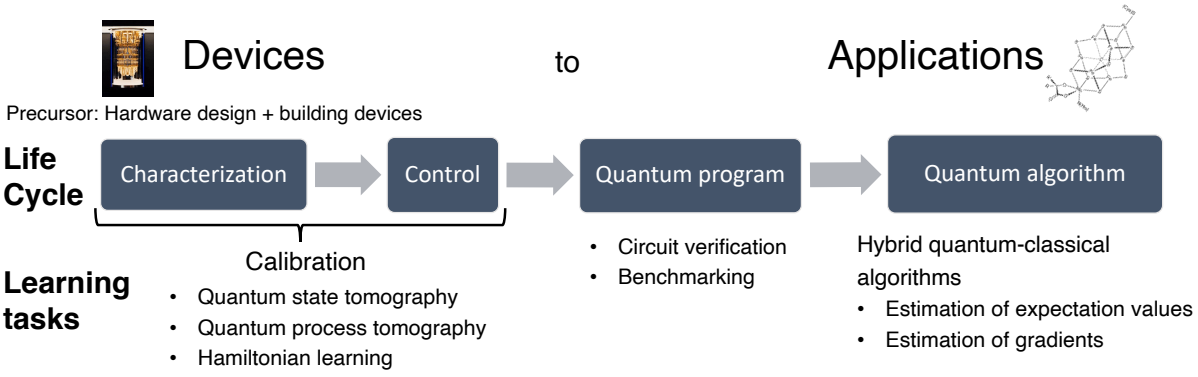


Figure 1.2: Overview of different learning tasks as we move forward in the life cycle of quantum computing starting from development of devices to deployment in useful applications.

In this section, we describe the life cycle of quantum computing alluded to earlier. The life cycle of quantum computing involves several stages from conception to development,

deployment and eventual decommissioning. In Figure 1.2, we show the stages of the life cycle involved after development of the device and up to deployment in useful applications. After building a quantum computer, it must be characterized to develop a description of low-level operations. Learning tasks in this stage include quantum state tomography, quantum process tomography and Hamiltonian learning which give us access to description of the state, process and dynamics respectively. As part of the characterization step, we also seek to learn local noise sources along with the underlying Hamiltonian and inform the next stage of quantum control. As part of quantum control, we learn how to drive or modulate the Hamiltonian evolution to implement various multi-qubit gates. The stages of characterization and control may be executed alternately multiple times and we refer to these stages together as calibration in Figure 1.2.

After having accomplished the step of calibration, the next stage of the quantum computing life cycle is the implementation of quantum programs or subroutines. At this stage, it may be desirable to perform benchmarks to ascertain the power or reliability of the device as well as verification to assess the correctness of implemented circuits. As tomography protocols scale exponentially with the number of qubits, they cannot be used for these tasks in general and different learning algorithms have to be developed. Finally, we will compose different quantum programs to implement quantum algorithms. As part of these quantum algorithms, we may need to carry out other learning tasks such as estimation of certain expectation values and gradients in variational quantum eigensolvers (VQE) [Kan+17; McC+16]. Other learning tasks may involve learning properties of the prepared quantum state without carrying out full tomography.

The list of learning tasks mentioned is not exhaustive but we have attempted to highlight the most relevant ones. We note that in many texts, quantum system characterization encompasses all protocols used for learning any feature of the quantum system or testing a property of the system. Here, we present the operational viewpoint of system characterization and where it features in the life cycle of quantum computing.

We are now in a position to delve deeper into each of these aforementioned learning tasks at different stages of the life cycle and evaluate the types of prior information that may be available for each learning task.

1.2.1 Quantum system characterization

In this section, we go through the different learning tasks typically considered as part of characterization and calibration. We also give examples of quantum systems where prior information was useful in accelerating the learning task.

Quantum state tomography. Quantum state tomography (QST) is the problem of learning a classical description $\hat{\rho}$ of an unknown n-qubit quantum state ρ drawn from a specified class of states by performing measurements on multiple copies of ρ, such that $|\hat{\rho}-\rho|\leq \epsilon$ under some metric $|\cdot|$ (e.g., infidelity or trace distance). The main figure of merit of a tomography protocol is its *sample complexity* or the number of copies of ρ used to learn $\hat{\rho}$. It is well known that the number of copies required to learn general unknown quantum states grows as $\Theta(\exp(n))$ [Haa+17; OW16].

QST is ubiquitous throughout system characterization and appears as a subroutine for

other learning tasks. It can be used to assess the quality of quantum gates and the fidelity of quantum operations. It is crucial for verifying quantum circuits and validating the performance of quantum algorithms. Partial state tomography where one performs QST on a subset of qubits appears as a subroutine in numerous quantum algorithms [BBO20; RBM18].

Given the prevalence of QST across the life cycle of quantum computation, it is imperative to accelerate QST. One way to do this would be by exploiting prior information on the structure of the unknown quantum state or the class of unknown quantum states. Known examples of such n-qubit classes for which QST can be performed efficiently, i.e., the sample complexity grows polynomially in n, include Matrix Product States [Cra+10], stabilizer states [Mon17a], symmetric states invariant under qubit permutations [Tót+10], and Gibbs states of local Hamiltonians [Ans+21; HKT21].

Quantum process tomography. Quantum process tomography (QPT) [CN97] is the problem of learning the dynamics or operations that transform one quantum state into another. Formally in QPT of n-qubit quantum unitary operators, given access to an unknown unitary U, we aim to output a classical description $\hat{U}$ that is ϵ-close to the unknown unitary under some norm (e.g., diamond norm, Frobenius distance, etc.). QPT has applications in the step of characterization to learn the quantum process being implemented by a control and assess the fidelity of quantum gates. It is thus a useful tool in the calibration of quantum control parameters for implementing gates. QPT is also used for verifying quantum circuits implemented as part of quantum algorithms.

One way to solve this problem is via QST [NC10, Chapter 8.4.2]: prepare a basis of quantum states, apply the unknown unitary U and perform QST on the results. More sophisticated strategies including use of ancilla and adaptivity have been developed [MRL08; Leu00; Car22; YRC20; Bis+10]. However, as was the case for QST, it has been established that the query complexity (or the number of queries to U) of QPT is $\Theta(\exp(n))$ [Leu00; Haa+23] which is prohibitive in practice.

We can ask here again as we did for QST if we can do better given prior information on the structure of U? Indeed, we can. Examples of efficiently learnable quantum unitaries with query complexity scaling polynomially in n include Clifford circuits [Low09], Clifford circuits with one layer of the non-Clifford $T = \mathrm{diag}(1, \exp(i\pi/4))$ gates [LC22], and $O(\mathrm{poly}(\log n))$-junta quantum unitaries [CNY] i.e., n-qubit quantum unitaries that act non-trivially on a subset of qubits.

In many cases such as circuit verification and quantum sensing, further prior information may be available. We may know ahead of time that the unknown quantum process or channel belongs to a finite ensemble $\mathcal{E} = \{C_1, C_2, \ldots, C_M\}$. The problem of estimating a quantum channel becomes that of quantum channel discrimination (QCD) where the goal is to discriminate the unknown channel C in $\mathcal{E}$ given query access to C. Notably, the sample complexity of QCD is now $\tilde{O}(\log M))$ [Har+10; HW12] where we have hidden the dependence on minimum distance between any pair of channels in $\mathcal{E}$.

Prior information can be thus be useful in accelerating QPT and we will study another scenario in this thesis.

Hamiltonian learning. In Hamiltonian learning (HL), the goal is to learn the Hamiltonian governing system dynamics through experimental measurements and application on prepared states. Formally, given access to the quantum unitary $U = \exp(-iHt)$ corresponds to a unitary time evolution of an unknown Hamiltonian H for evolution time t, the goal is to learn an estimate $\hat{H}$ that is close to H under some norm (e.g., Frobenius norm). If we have access to a parametrized Hamiltonian model, Hamiltonian learning becomes the problem of Hamiltonian parameters' estimation where we seek to learn an estimate $\hat{\theta}$ of the true unkown parameters θ under the ℓ_2 norm. Hamiltonian learning is an important procedure in calibration of quantum computers where it is used to characterize the underlying Hamiltonian of the dynamics of a set of qubits and the estimated Hamiltonian can then be used to learn how to drive it to implement multi-qubit gates [Inn+20; She+16]. Hamiltonian learning is thus used around the clock for the successful operation of quantum computers and plays a pivotal step in recalibrations to avoid drift.

The query complexity of HL, perhaps unsurprisingly at this stage, is $\Theta(\exp(n))$ [MRL08; Car22]. Further, HL requires $O(\epsilon^{-2})$ queries in achieving learning error ϵ due to the standard quantum limit (SQL) [GLM04a; GLM11a]. Given prior information on the structure of the Hamiltonian, the query complexity can be reduced substantially for some instances and learning can be made efficient. For example, we can learn k-local n-qubit Hamiltonians i.e., Hamiltonians involving interactions between at most k qubits $O(\text{poly}(n)$ queries given access to Gibbs states [Aru+23; HKT22; Bak+23] and short-time unitary evolution [GCC22; Fra+22; Hua+23] (as considered here). Hamiltonians which are sparse in the sense that they contain at most s terms in a certain basis can also be learned efficiently in $O(sn)$ queries [Sha+11] if the basis is known and $O(sn/\epsilon^4)$ queries [Yu+23] if the basis is unknown but with strong assumptions on the support and magnitude of the Hamiltonian parameters.

All the HL approaches given prior information mentioned so far have query complexity of SQL ($O(1/\epsilon^2)$) or worse except for [Hua+23] which is able to achieve the so called Heisenberg limited scaling $O(1/\epsilon)$. However, this latter approach requires the ability to implement interleaved single-qubit gates as part of the HL protocol which is not usually available in practice as HL itself is typically used for gate design. When such control is not available during HL, it has been shown that Heisenberg limited scaling cannot be achieved for general n-qubit Hamiltonians [DOS23]. This does not rule out particular instances of Hamiltonian models though where Heisenberg limited scaling could be achieved and we show how it may be achieved in some regimes for a quantum hardware relevant Hamiltonian in this thesis.

1.2.2 Implementing quantum programs and algorithms

In the life cycle of quantum computing after system characterization and gate design as part of quantum control, we are in a position to implement quantum programs i.e., a sequence of layers of quantum gates, corresponding to building blocks (also called subroutines) of quantum algorithms. In this section, we now consider some learning tasks that appear in the stages of implementing quantum programs and executing quantum algorithms.

Benchmarking. After implementing a quantum program, we might want to learn the various errors that accumulate during its runtime. This has implications for the types of quantum algorithms (or compositions of different quantum programs) we may afford to

run. The tomography protocols from each of the stages - characterization, calibration and control - that have come before in the life cycle reveal very fine-grained information about the quantum device and give us a complete description of quantum processes but they can come at prohibitive cost for large-scale systems. They are thus only practical in general for low-level operations such as assessing quality of qubits and gates over few qubits. It is then desirable to have protocols that give us access to partial descriptions of error rates without losing too much information, at a lower resource cost and higher scalability. Such protocols include direct fidelity estimation [FL11], randomized benchmarking [MGE11], and cycle benchmarking [Erh+19]. If we wanted to summarize the performance of the quantum device as a whole, several metrics and approaches have been developed such as quantum volume (QV) [Cro+19], volumetric benchmarks [BY20], circuit layer operations per second (CLOPS) [Wac+21], and use of specific certain applications as estimation of performance e.g., quantum chemistry [McC+19].

Quantum verification. Quantum verification [Eis+20] is the task of checking whether a quantum device or implemented circuits behave as intended and produce the correct output for a given set of inputs. Unlike benchmarking, the goal is to then verify correctness of the functioning of the quantum device rather than assign a reproducible performance measure to the quantum device. As current quantum computers are noisy and inherently error-prone, errors can occur during the execution of circuits and verification aids in identifying them.

One way to accomplish different verification tasks would be to carry out tomography but due to its impracticality on large-scale systems, various approaches have been developed including cross-platform verification of quantum states [Car+21], blind quantum computation [GKK19], and self-testing protocols [ŠB20]. Consider the particular verification task of *quantum state certification*, where the goal is to check if the state $\tilde{\rho}$ produced by a quantum device is ϵ-close to a given quantum state ρ in infidelity. This could be accomplished via QST but would be prohibitively expensive for large-scale systems. Instead, another approach is to carry out direct fidelity estimation [KR21] and whose sample complexity can be further improved for well-conditioned states.

Estimation of observables and state properties. We now discuss a learning task encountered in quantum algorithms. In many hybrid quantum-classical algorithms and even fault-tolerant algorithms, the output on the quantum computer is a quantum state encoded solution but we need access to certain properties of this state or classical relevant information. Let us describe two use cases. Consider the scenario of solving linear partial differential equation (PDE) e.g., the heat equation [LMS22] using a quantum computer. One approach is to use the HHL algorithm [HHL09] on the linear system of equations obtained after discretizing the linear PDE spatially and temporally. This offers a potential exponential speedup in the terms of the spatial resolution. However, the output of the HHL algorithm will be quantum state encoding the solution to the PDE whereas classically relevant information are the flux and energy over small regions in space. We could obtain this information via QST but this would come at an exponential cost and we would lose any speedup. Given prior knowledge of the target properties, [LMS22] shows that one can apply a quantum protocol for numerical integration incorporating amplitude estimation and retain a quadratic speedup

over the best classical algorithm for solving the PDE.

As a second illustrative scenario, we consider the application of ground state energy estimation of a many-body molecular Hamiltonian using variational quantum eigensolvers (VQE) [McA+20; McC+16]. In VQE, quantum states produced by parameterized quantum circuits undergo frequent measurements with respect to H. The learning problem is that of estimating $\mathrm{Tr}(\rho H)$ efficiently. This can be accomplished by noting the Pauli decomposition of H and performing suitable single-qubit Pauli measurements P. We then estimate $\mathrm{Tr}(\rho H)$ via estimation of local observables $\mathrm{Tr}(\rho P)$. Prior information that may be available to us in such a scenario is the number of Pauli terms in the Pauli decomposition of H and the highest weight (number of non-identity single-qubit Paulis in any Pauli term of H). For example, qubit Hamiltonians of molecules have $O(n^4)$ Paulis [McA+20] in its decomposition and thus this learning task can be performed in $O(\mathrm{poly}(n))$ measurements. If we are given prior information on the highest weight w, we can utilize classical shadows [HKP20; Elb+23] to accomplish this in $O(3^w \log n)$ Pauli measurements, which is efficient for $w = O(\mathrm{poly}(\log n))$.

1.3 Thesis contributions

We now discuss the contributions in this thesis towards answering the thesis question in the context of the learning tasks identified in Section 1.2. To address the challenges for each learning task, we will employ the common tools of statistical learning and asymptotic analysis, while at the same time attempting to validate our learning algorithms through numerical experiments or experimentals on quantum hardware. These contributions are based on independent research articles by the author and the involvement of the author in each of these articles is also mentioned simultaneously.

Quantum state tomography. We address the thesis question in the context of learning quantum states in Chapter 2. This contribution is based on the paper [Aru+23] for which the thesis author was the main contributing author (listed alphabetically).

Assuming that the prior information available is regarding the structure or the class of unknown quantum states available to us, we ask the question: What are other classes of n-qubit quantum states that can be learned efficiently, given access to copies of the unknown quantum state? Particularly, could we learn states that cannot be efficiently simulated classically?

It is well-known that Clifford circuits and states produced by these circuits (i.e., stabilizer states) are learnable in polynomial time. Stabilizer states and Clifford circuits are also known to be classically simulatable using the Gottesman-Knill framework which is crucial in these learning procedures. To obtain a candidate of a class of states that are not classically simulatable, we have to look no further than states produced by IQP circuits (Instantaneous Quantum Polynomial-time) which are prevalent in quantum-advantage experiments [BJS11; BMS17]. Specifically, we are interested in the question of efficient learnability of a subclass of IQP states, which correspond to $|\psi_f\rangle = 2^{-n/2}\sum_{x\in\{0,1\}^n}(-1)^{f(x)}|x\rangle$ and where f is a degree-3 Boolean polynomial [Mon17b]. These states are produced by n-qubit circuits with the structure $\mathsf{H}^{\otimes n}\mathsf{V}$ containing a layers of Hadamard H gates across all qubits followed by circuit V containing internal gates in $\{\mathsf{Z}, \mathsf{CZ}, \mathsf{CCZ}\}$ (where CZ is the controlled-Z gate). We go a step

further and consider learning *binary phase states* produced by circuits with internal gates in $\{\mathsf{Z}, \mathsf{CZ}, \mathsf{CCZ}, \ldots, \mathsf{C^{d-1}Z}\}$ where $\mathsf{C^{d-1}Z}$ is a controlled-Z with controls over $d-1$ qubits. Binary phase states can be expressed as $|\psi_f\rangle = 2^{-n/2}\sum_{x\in\{0,1\}^n}(-1)^{f(x)}|x\rangle$ where f is now a degree-d Boolean polynomial. We also consider learning *generalized phase states* produced by the d-th level of the diagonal Clifford hierarchy and can be expressed as $|\psi_f\rangle = 2^{-n/2}\sum_x \omega_q^{f(x)}|x\rangle$ where $q = 2^d$, ω_q is qth root of unity, and f is a degree-d Boolean polynomial. The goal of this work is to then show that phase states can be learned efficiently in $O(\text{poly}(n))$ samples by exploiting the structure of these states and their correspondence to Boolean polynomials.

Apart from being natural states to learn, phase states have appeared in several recent works: [JLS18; BS19] showed phase states are efficiently preparable and statistically indistinguishable from a Haar random state (for a polynomial-time quantum algorithm), subsequently there have been followup works using phase states for cryptosystems [AQY21]; Irani et al. [Ira+21] showed that in order to construct the witness to a QMA complete problem, say the ground state $|\phi\rangle$ to a local-Hamiltonian problem, it suffices to consider a phase state which has a good overlap to $|\phi\rangle$; level-3 phase states are universal for measurement based quantum computing [Ros+13; TMH19].

It is widely open how to learn states or circuits beyond the Clifford group. In this thesis, we give *optimal* algorithms for learning degree-d binary phase states using separable measurements in $O(n^d)$ sample complexity and using entangled measurements in $O(n^{d-1})$ sample complexity. Our learning algorithms are optimal with regards to the dependence of the sample complexity achieved on d and this is proved via lower bounds. These state learning algorithms can in turn be used to learn the corresponding IQP like circuits and diagonal unitaries of the Clifford hierarchy given query access in $O(\text{poly}(n))$ samples. Moreover, we find that the sample complexity of learning binary phase states can be reduced by using entangled measurements instead of separable measurements but this comes at an increase in computational complexity of $O(\text{poly}(n))$ to $O(\exp(n))$. While this is specific for our learning algorithms, this showcases an instance of a trade-off in quantum resources and classical computational runtime.

Quantum process tomography. We next address the thesis question in the context of learning quantum channels in Chapter 3. This contribution is based on [Sug+23] for which the thesis author was the second main contributing author.

In Section 1.2.1, we introduced the learning problem of quantum channel discrimination (QCD) where we have prior information of the ensemble of valid unknown quantum channels. Various protocols to solve QCD have been designed including sequential protocols, multi-shot protocols and entanglement-enhanced parallel protocols [Har+10; DCS17]. However, these protocols cannot be directly applied to many physical systems on which an unknown channel is acting. Consider naturally occurring quantum systems such as those undergoing chemical reactions [Kat+22]. Directly measuring the physical system of interest in this case would disrupt the dynamics. There also have been proposals of engineered quantum systems based on superconducting quantum circuits [Pec+21] to prevent overcrowding of qubits, where qubits lack dedicated readout lines due to limited on-chip routing capabilities and to ensure the number of control lines is not a limiting factor for scaling the system size.

Common restrictions on these quantum systems on which a channel is acting are: ar-

bitrary control of the system is not possible, direct measurement is not allowed and initial state preparation is unreliable. We call quantum systems with these restrictions as *hidden* quantum systems and the associated channel discrimination problem as hidden quantum channel discrimination (HQCD). We investigate if HQCD can be accomplished successful by adapting protocols from conventional QCD and with fewer queries than classical methods.

In this thesis, we formally pose the problem of HQCD and study the limits of different protocols in solving HQCD by considering the minimal two-qubit example of hidden binary channel discrimination (HBCD) where the quantum channel C acting on a hidden qubit is a single-qubit rotation $C = \exp(i\theta_C \sigma_x)$ with the parameter θ_C being unknown and to be discriminated. The initial state of the hidden qubit is assumed to be the maximally mixed state. The true value of θ_C is either 0 or a given $\alpha \in (0, 2\pi)$ with equal probability. The hidden qubit is accessed by a measurement apparatus consisting of a single (target) qubit through a query that cannot directly manipulate the hidden system. A query involves N serial applications of C on the hidden qubit, a tunable controlled-rotation gate on to the target qubit, and single-qubit rotations on the target qubit. The figure of merit of a protocol is the query complexity N or the number of interactions with the hidden system required for accomplishing HBCD with error probability $\epsilon \in [0, 1/2)$. We then ask: Can we discriminate the hidden quantum channel at the Heisenberg limited scaling $O(\epsilon^{-1})$ rather than that allowed by the central limit theorem $O(\epsilon^{-2})$ and thereby accelerate learning?

In this work, we compare the performance of sequential, multi-shot and parallel protocols for HBCD. In the sequential protocol, a query of length N is applied and the target qubit is only measured once. In the multi-shot protocol, queries of fixed length d are applied and the target qubit is measured multiple times. We show that (i) it is impossible to accomplish HBCD using a non-sequential protocol (e.g., multi-shot and parallel protocols with queries of length one) when the initial state of the hidden system is maximally mixed, and (ii) the sequential protocol achieves Heisenberg limited learning $N \sim O(\alpha^{-1})$ in HBCD whereas a multi-shot protocol is only able to achieve the standard quantum limit (SQL) of $N \sim O(\alpha^{-2})$. We thus propose a general framework for learning physical processes on hidden quantum systems.

Hamiltonian learning. We address the thesis question in the context of Hamiltonian learning (HL) in Chapter 4. This contribution is based on [Dut+23a] for which the thesis author was the first and main contributing author.

Current state-of-art HL approaches aim to accelerate learning by improving estimation methods compatible with any set of queries [EHF19; Aru+20a], proposing engineered set or sequence of queries to be made [Aru+20a; Fra+22; Hua+23] and introducing adaptivity [Gra+12; Gra+17]. The strategies proposed in [Gra+12; Gra+17] are not practical on current and near-term quantum hardware as they propose new queries to be made one shot at a time, requiring multiple accesses to the quantum computer. In this thesis, our goal is to accelerate HL by introducing adaptivity in selection of queries (or experiments) but respect limitations of access to quantum devices and latenices on current quantum hardware in the NISQ era. At this same time, we want our resulting learning algorithm to be compatible with any set of experiments and estimation strategies, so we can take advantage of advances in estimation or experiment design. Towards this end, we introduce active learning, a machine

learning tool for proposing informative queries, that operates in batch-mode and could be used to achieve resource reduction in practice. The active learner can be viewed as a learning algorithm that develops and utilizes prior information in real-time. Finally, we ask: Can we discover regimes where we can beat standard quantum limit and achieve Heisenberg limited scaling?

To obtain a reduction of quantum resources in practice, we introduce a batch-mode Hamiltonian active learner (HAL) that proposes informative queries adaptively during learning. In our experiments on a 20-qubit IBM quantum device, HAL achieved a 95% query advantage compared to standard methods, and a 33% query advantage over a sequential active learner. In practice, this leads to an order of magnitude reduction in queries made during learning of two-qubit cross-resonance gates to reach a desired accuracy during gate design or a reduction of quantum wall-clock time in running experiments for learning from around 10 minutes to 5 seconds. This has significant implications for resource requirements of learning all the two-qubit cross-resonance (CR) Hamiltonians on large devices such as the 1121-qubit IBM quantum device Condor [22a] shown in Figure 1.3. Standard methods would require nearly 17 days in sequentially learning all the CR Hamiltonians within an error of 0.05. Whereas, HAL would only need 3.5 hours. Moreover, with access to information from previous calibrations, HAL can exceed the standard quantum limit and achieve the Heisenberg limit during learning.

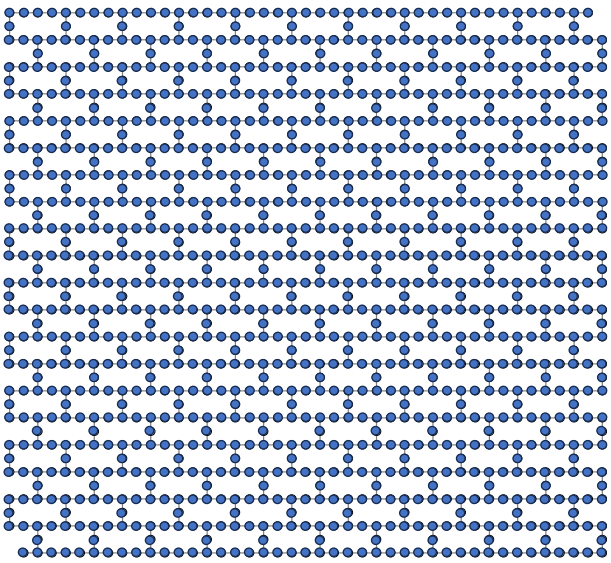


Figure 1.3: Connectivity map of 1121-qubit IBM quantum device Condor with nodes denoting different qubits and edges between pairs of qubits indicating direct connections. Edges directly correspond to presence of cross-resonance Hamiltonian interaction between pairs of qubits. Figure adapted from [Pat+23].

During operation of quantum devices, HAL could be utilized to reduce downtime and enable frequent recalibrations to mitigate the effects of drift. More broadly, by employing adaptivity in real-time for characterization and control, we could improve robustness to noise and applicability to large-scale quantum systems.

Estimation of physical observables. We consider the learning task of observable estimation for the application of ground state energy estimation using near-term hybrid quantum-classical algorithms in Chapter 5. This contribution is based on [Dut+23b] for which the thesis author is the first and main contributing author.

A leading candidate application for showing practical quantum advantage is ground state energy estimation [McA+20; Lee+23], even though a provable exponential speedup is yet to be shown. On current near-term quantum hardware which are noisy and where we are limited by circuit depth, hybrid quantum-classical algorithms such as variational quantum eigensolvers (VQE) [Per+14; Kan+17], and quantum subspace expansion (QSE) methods [McC+17; PM19] will be used for this application. A common subroutine across these algorithms is that of observable estimation or estimating $\mathrm{Tr}(\rho H)$ for an n-qubit molecular Hamiltonian H and a quantum state ρ. Particular prior knowledge of molecular n-qubit Hamiltonians H is also available: they contain $O(n^4)$ Paulis in their Pauli decomposition.

Various approaches for solving the observable estimation problem using local Pauli measurements have thus been proposed such as randomized methods [HKP20; Had+22; Hil+21; KG22; Elb+23], grouping methods [YVI20; VYI20; Wu+23; Shl+23] and derandomization [Hua+23]. However, it is relatively unknown how these measurement methods perform in practice and systematic benchmarking is lacking. The main performance metric that has been used to compare different measurement methods has been accuracy achieved against a fixed measurement budget. This does not account for the utilization of classical computational resources in generating measurement bases from these measurement methods (e.g., optimization, sampling, etc.) and post-processing of measurement outcomes acquired from experiments on the quantum device. In this thesis, we ask: How do we systematically benchmark measurement methods for the common quantum computation of estimating $\mathrm{Tr}(\rho H)$ for molecular Hamiltonians H in hybrid quantum-classical algorithms?

In this thesis, we propose the benchmarking tool CSHOREBench: Common Systems and Hamiltonians for ObseRvable Estimation Benchmark. Analogous to benchmarks in machine learning and scientific computing, we assess the performance of our candidate algorithms (measurement methods) on common states and common molecular Hamiltonians. Given a Hamiltonian, the set of common states include the Hartree-Fock state which is often used as an initialization in VQE, random quantum ansatz and the ground state. In CSHOREBench, we use a heuristic to summarize utilization of both quantum and classical resources. In our experiments on IBM Quantum devices for molecular Hamiltonians (up to 16 qubits), we find that the methods of decision diagrams and derandomization are most amenable to near-term quantum hardware.

Applications. We reconsider the application of ground state estimation but now with access to entanglement during measurements and with a different viewpoint on the prior information available on molecular Hamiltonians. This contribution is based on [Liu+23] for which the thesis author was a contributing author but did not lead the research.

As we have emphasized throughout this introductory chapter, finding the ground state of interacting fermionic systems is an outstanding challenge for quantum chemistry, material science, and condensed matter physics. Numerically solving the time-independent Schrodinger equation of a meaningfully large many-electron system in an exact fashion is

a daunting task because the dimension of the underlying Hilbert space grows exponentially as the number of electrons increases, making this classically intractable. To overcome this, quantum algorithms to be run on quantum computers have been designed for this purpose. Namely, these include fault-tolerant algorithms which require quantum error correction and high-depth quantum circuits (e.g., quantum phase estimation [Kit95]). In the NISQ era and pre-fault tolerant era, we will only be able to execute near-term hybrid quantum-classical algorithms that use short-depth quantum circuits but are affected by noise e.g., variational quantum eigensolvers (VQE) [Per+14; McC+16; Kan+17; Gri+19] and quantum subspace expansion (QSE) methods [McC+17; Col+18a; PM19; Mot+20].

While being able to run these hybrid quantum-classical algorithm may appear encouraging, we expect practical quantum advantage for chemistry Hamiltonians requiring upwards of hundreds of thousands of qubits and assuming low qubit error rates [Goi+22]. We thus still have a way to go before sizes of quantum computers reach those numbers. However, we are likely to obtain multinode quantum computers in the near-future [Ang+22] where smaller quantum computers are linked together akin to classical computing clusters (or distributed computing). How can we then take advantage of distributed quantum computing for the problem of ground estimation? We do this by using ideas from domain decomposition prevalent in classical computational science [DJN15] and exploiting prior knowledge of chemistry Hamiltonians.

To address this challenge, we present a distributed quantum algorithm called quantum bootstrap embedding (QBE). Operationally, QBE decomposes the molecular Hamiltonian for which we want to find the ground state into multiple overlapping fragment Hamiltonians and to be solved on separate quantum computers or quantum computing nodes. To obtain the ground state over the original Hamiltonian, we formulate a constraint optimization problem for a composite Lagrangian where the constraint is constructed from matching conditions on the qubit reduced density matrices and present an iterative algorithm to solve the optimization problem using a quantum subroutine as an eigensolver to solve each fragment Hamiltonian. An adaptive sampling scheduling and a quantum coherent matching algorithm based on a quantum SWAP test are designed to dramatically improve the efficiency of the algorithm as compared to the usual exponentially costly method of measuring every qubit on the fragment edge to construct the reduced density matrix. Moreover, by using amplitude amplification and a binary search algorithm, an additional quadratic speedup could be realized. Current quantum computers are small, but QBE proves a potentially generalizable strategy for harnessing such small devices, since it enables the stitching together of fragment solutions to solve a quantum chemistry problem that is much larger than current quantum computer capacities.

1.4 Thesis organization

The rest of the thesis is organized in the following chapters in the order of the learning tasks as they appear in the life cycle of quantum computing and for which we give brief summaries.

Chapter 2. This marks our point of entry into quantum system characterization. We consider the learning problem of quantum state tomography and the class of n-qubit quantum

phase states as a class of quantum states that potentially can be learned efficiently but is hard to simulate classically. A degree-d phase state is defined as a superposition of all 2^n basis vectors x with amplitudes proportional to $(-1)^{f(x)}$, where f is a degree-d Boolean polynomial over n variables. We show that the sample complexity of learning an unknown degree-d phase state is $\Theta(n^d)$ if we allow separable measurements and $\Theta(n^{d-1})$ if we allow entangled measurements. Our learning algorithm based on separable measurements has runtime $\mathsf{poly}(n)$ (for constant d) and is well-suited for near-term demonstrations as it requires only single-qubit measurements in the Pauli X and Z bases. We show similar bounds on the sample complexity for learning generalized phase states with complex-valued amplitudes. We further consider learning phase states when f has sparsity-s, degree-d in its $\mathbb{F}_2$ representation (with sample complexity $O(2^d sn)$), f has Fourier-degree-t (with sample complexity $O(2^{2t})$), and learning quadratic phase states with ε-global depolarizing noise (with sample complexity $O(n^{1+\varepsilon})$). These learning algorithms give us a procedure to learn the diagonal unitaries of the Clifford hierarchy and IQP circuits.

Chapter 3. We now turn our attention to the problem of learning quantum channels under experimental constraints. In many natural and engineered systems, unknown quantum channels act on a subsystem that cannot be directly controlled and measured, but is instead learned through a controllable subsystem that weakly interacts with it. In this chapter, we study quantum channel discrimination (QCD) under these restrictions, which we call hidden system QCD (HQCD). We find that sequential protocols achieve perfect discrimination and saturate the Heisenberg limit. In contrast, depth-1 parallel and multi-shot protocols cannot solve HQCD. This suggests that sequential protocols are superior in experimentally realistic situations.

Chapter 4. We continue our travel through the realm of quantum system characterization and consider the problem of Hamiltonian learning which is an important step in calibration and control. With the goal of efficiently and accurately estimating the Hamiltonian parameters within learning error ϵ through minimal queries, we introduce an active learner based on Fisher information that is given an initial set of training examples and the ability to interactively query the quantum system to generate new training data. To ensure applicability on near-term quantum hardware, the active learner operates in batch-mode as opposed to sequentially, proposing batches of queries to be made during learning. We formally specify and experimentally assess the performance of this Hamiltonian active learning (HAL) algorithm for learning the six parameters of a two-qubit cross-resonance Hamiltonian on four different superconducting IBM Quantum devices. Compared with standard techniques for the same problem and a specified learning error, HAL achieves more than a 95% reduction in queries required, and upwards of 33% reduction over a sequential active learner. Moreover, with access to prior information on a subset of Hamiltonian parameters and given the ability to select queries with linearly (or exponentially) longer system interaction times during learning, HAL can exceed the standard quantum limit and achieve Heisenberg (or super-Heisenberg) limited convergence rates during learning.

Chapter 5. One of the most promising applications of quantum computing is ground state energy estimation. Many near-term hybrid quantum-classical algorithms designed for this task involve estimation of the expectation value of observables through measurements on the quantum device. With the goal of minimizing the number of measurements required to compute the expectation value within some accuracy, various measurement methods have been proposed based on randomization and derandomization. However, a systematic benchmarking of these methods considering different performance metrics in addition to accuracy, such as quantum resource usage and classical computational runtime, is currently missing. In this work, we propose a simple benchmark based on a heuristic to assess and rank the performance of these measurement methods in practice. This heuristic accounts for the hybrid quantum-classical nature of these methods, weighing usage of quantum hardware, classical computation, and latencies. We apply this benchmark to a variety of measurement methods on a collection of molecular Hamiltonians (of size up to 16 qubits). Our discussion is aided by using the framework of decision diagrams which provides an efficient data structure for various randomized measurement methods and efficiently generating samples from them. In our experiments on a simulator and on IBM quantum devices, we find that the methods of compact decision diagrams and derandomization of decision diagrams are the most preferable for algorithms on near-term quantum hardware.

Chapter 6. We continue our focus on the application of ground state estimation of molecular Hamiltonians and propose a distributed quantum algorithm based on molecular bootstrap embedding, which we call quantum bootstrap embedding (QBE). QBE can be used to solve the ground state of a large molecule as an optimization problem over composite Lagrangian governing fragments of the total system, in such a way that fragment solutions can harness the capabilities of quantum computers. By employing state-of-art quantum subroutines including the quantum `SWAP` test and quantum amplitude amplification, we show how a quadratic speedup can be obtained over the classical algorithm based on bootstrap embedding, in principle. Utilization of quantum computation also allows the algorithm to match – at little additional computational cost – full density matrices at fragment boundaries, instead of being limited to 1-RDMs. Current quantum computers are small, but QBE provides a potentially generalizable strategy for harnessing such multiple small machines through quantum fragment matching.

Finally, in chapter 7, we make concluding remarks to our thesis and propose future directions of research.

Chapter 2

Efficiently learning quantum phase states

2.1 Introduction

Quantum state tomography is the problem of learning an unknown quantum state ρ drawn from a specified class of states by performing measurements on multiple copies of ρ. The preeminence of this problem in verification of quantum experiments has motivated an in-depth study of state tomography protocols and their limitations for various classes of quantum states [Haa+17; OW16; Ape+22; Yue22]. The main figure of merit characterizing a state tomography protocol is its *sample complexity* defined as the number of copies of ρ consumed by the protocol in order to learn ρ. Of particular interest are classes of n-qubit quantum states that can be learned efficiently, such that the sample complexity grows only polynomially with n. Known examples of efficiently learnable states include Matrix Product States describing weakly entangled quantum spin chains [Cra+10], output states of Clifford circuits [Mon17a], output states of Clifford circuits with a single layer of T gates [LC22], and high-temperature Gibbs states of local Hamiltonians [Ans+21; HKT21]. Apart from their potential use in experiments, efficiently learnable quantum states are of great importance for quantum algorithm design. For example, a quantum algorithm for solving the dihedral hidden subgroup problem [BCD05] can be viewed as a tomography protocol for learning so-called hidden subgroup states (although this protocol is efficient in term of its sample complexity, its runtime is believed to be super-polynomial [BCD05]).

A natural question to then ask is: What are other classes of n-qubit quantum states that are ubiquitous in quantum computing, which can be learned efficiently? In this work, we consider the problem of state tomography for *phase states* associated with (generalized) Boolean functions. Phase states are encountered in quantum information theory [HEB04], quantum algorithm design [BCD05], quantum cryptography [JLS18; BS19], and quantum-advantage experiments [BJS11; BMS17].

By definition, an n-qubit, degree-d phase state has the form

$$|\psi_f\rangle = 2^{-n/2} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle, \tag{2.1}$$

where $f : \{0,1\}^n \to \{0,1\}$ is a degree-d polynomial, that is,

$$f(x) = \sum_{J \subseteq [n], |J| \leq d} \alpha_J \prod_{j \in J} x_j \pmod 2, \tag{2.2}$$

for some coefficients $\alpha_J \in \{0,1\}$. Phase states associated with homogeneous degree-2 polynomials $f(x)$ coincide with graph states that play a prominent role in quantum information theory [HEB04]. Such states can be alternatively represented as

$$|\psi_f\rangle = \prod_{(i,j) \in E} \mathsf{CZ}_{i,j} |+\rangle^{\otimes n},$$

where n qubits live at vertices of a graph, E is the set of graph edges, $\mathsf{CZ}_{i,j}$ is the controlled-Z gate applied to qubits i, j, and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. It is known that the output state of any Clifford circuit is locally equivalent to a graph state for a suitable graph [Sch02]. Our results imply that graph states can be learned efficiently using only single-qubit gates and measurements. The best previously known protocol for learning graph states [Mon17a] requires entangled measurements across two copies of $|\psi_f\rangle$. Other examples of circuits producing phase states include measurement-based quantum computing [Ros+13] and a subclass of IQP circuits (Instantaneous Quantum Polynomial-time), which correspond to degree-3 phase states [Mon17b]. IQP circuits are prevalent in quantum-advantage experiments [BJS11; BMS17] and are believed to be hard to simulate classically.

We also consider generalized degree-d phase states

$$|\psi_f\rangle = 2^{-n/2} \sum_{x \in \{0,1\}^n} \omega_q^{f(x)} |x\rangle, \qquad \omega_q = e^{2\pi i/q} \tag{2.3}$$

where $q \geq 2$ is an even integer and $f : \{0,1\}^n \to \mathbb{Z}_q$ is a degree-d polynomial, that is,

$$f(x) = \sum_{J \subseteq [n], |J| \leq d} \alpha_J \prod_{j \in J} x_j \pmod q. \tag{2.4}$$

for coefficients $\alpha_J \in \mathbb{Z}_q = \{0, 1, \ldots, q-1\}$. It is also known that generalized degree-d phase states with $q = 2^d$ can be prepared from diagonal unitary operators [CGK17] in the d-th level of the Clifford hierarchy [GC99]. Additionally, it is known that the output state of a random n-qubit Clifford circuit is a generalized $q = 4$, degree-2 phase state with a constant probability [BG16, Appendix D]. Binary and generalized phase states have also found applications in cryptography [JLS18; BS19], and complexity theory [Ira+21] (we discuss this in the next section).

In this work, we consider learning phase states through two types of tomography protocols based on *separable* and *entangled* measurements. The former can be realized as a sequence of M independent measurements, each performed on a separate copy of $|\psi_f\rangle$ (furthermore our learning algorithms only require single *qubit* measurements). The latter performs a joint measurement on the state $|\psi_f\rangle^{\otimes M}$. Our goal is to then derive upper and lower bounds on the sample complexity M of learning f, as a function of n and d. In the next section, we state our main results. Interestingly, our protocols based on separable measurements require only single-qubit gates and single-qubit measurements making them well suited for near-term demonstrations.

2.2 Our contributions

We first introduce some notation before giving an overview of our contributions. For every n and $d \leq n/2$, let $\mathcal{P}(n,d)$ be the set of all degree-d polynomials of the form Eq. (2.2). Let $\mathcal{P}_q(n,d)$ be the set of all degree-d $\mathbb{Z}_q$-valued polynomials of the form Eq. (2.3). By definition, $\mathcal{P}_2(n,d) \equiv \mathcal{P}(n,d)$. To avoid confusion, we shall refer to states defined in Eq. (2.1) as binary phase states and in Eq. (2.3) as generalized phase states. Our learning protocol takes as input integers n, d and M copies of a degree-d phase state $|\psi_f\rangle$ with unknown $f \in \mathcal{P}(n,d)$ (or $f \in \mathcal{P}_q(n,d)$). The protocol outputs a classical description of a polynomial $g \in \mathcal{P}(n,d)$ (or $g \in \mathcal{P}_q(n,d)$) such that $f = g$ with high probability.

The main result in this work are optimal algorithms for learning phase states if the algorithm is allowed to make separable or entangled measurements. Prior to our work, we are aware of only two works in this direction (i) algorithms for efficiently learning degree-1 and degree-2 phase states; (ii) Montanaro [Mon12] considered learning multilinear polynomials f, assuming we have *query access* to f, which is a stronger learning model than the sample access model that we assume for our learning algorithm. In this work, we show that if allowed separable measurements, the *sample* complexity of learning binary phase states and generalized phase states is $O(n^d)$. If allowed entangled measurements, we obtain a sample complexity of $O(dn^{d-1})$ for learning binary phase states. We further consider settings where the unknown function f we are trying to learn is known to be sparse, has a small Fourier-degree and the setting when given noisy copies of the quantum phase state. In Table 2.1, we summarize all our main results (except the first two rows, which include the main prior work in this direction).

Before we give a proof sketch of these results and the intuition behind them, we first discuss a couple of motivations for considering the task of learning phase states and corresponding applications.

Quantum complexity. Recently, there has been a few results in quantum cryptography [JLS18; AQY21; BS19] and complexity theory [Ira+21] which used the notion of phase states.

Ji et al. [JLS18] introduced the notion of *pseudorandom quantum states* as states of the form $|\phi\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \omega_N^{F(x)} |x\rangle$ where F is a pseudorandom function.[1] Ji et al. showed that states of the form $|\phi\rangle$ are efficiently preparable and statistically indistinguishable from a Haar random state, which given as input to a polynomial-time quantum algorithm. A subsequent work of Brakerski [BS19] showed that it suffices to consider $|\phi'\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{F(x)} |x\rangle$ (where F again is a pseudorandom function) and such states are also efficiently preparable and statistically indistinguishable from Haar random states. Subsequently, these states have found applications in proposing many cryptosystems [AQY21]. Although none of these works discuss the degree of the phase function F, our result shows implicitly that when F is low-degree, then $|\phi\rangle$ is exactly learnable and hence distinguishable from Haar random states, implying that they cannot be quantum pseudorandom states. In another recent work, Irani et al. [Ira+21] considered the power of quantum witnesses in proof systems. In particular, they showed that in order to construct the witness to a QMA complete problem,

[1] We do not discuss the details of pseudorandom functions here, we refer the interested reader to [JLS18].

	Sample complexity	Time complexity	Measurements
Binary phase state $\mathbb{F}_2$-degree-1 [BV97]	$\Theta(1)$	$O(n^3)$	Separable
Binary phase state $\mathbb{F}_2$-degree-2 [Mon17a; Röt09]	$O(n)$	$O(n^3)$	Entangled
Binary phase state $\mathbb{F}_2$-degree-d	$\Theta(n^d)$ Theorem 3, 5	$O(n^{3d-2})$	Separable
Binary phase state $\mathbb{F}_2$-degree-d	$\Theta(n^{d-1})$ Theorem 4	$O(\exp(n^d \log 2))$	Entangled
Generalized phase states degree-d	$\Theta(n^d)$ Theorem 8	$O(\exp(n^d \log q))$	Separable
Sparse Binary phase state $\mathbb{F}_2$-degree-d, $\mathbb{F}_2$-sparsity s	$O(2^d sn)$ Theorem 6	$O(2^{3d} s^3 n)$	Separable
Binary phase state $\mathbb{F}_2$-degree-2 with global depolarizing noise ε	$n^{1+O(\varepsilon)}$ Theorem 9	$O(2^{n/\log n})$	Entangled
Binary phase state $\mathbb{F}_2$-degree-2 with local depolarizing noise ε	$\Theta((1-\varepsilon)^n)$ Theorem 11	$O(2^{n/\log n})$	Entangled
Binary phase state Fourier-degree-d	$O(2^{2d})$ Theorem 7	$O(\exp(n^2))$	Entangled

Table 2.1: Upper and lower bounds of sample complexity for exact learning of n-qubit phase states with degree-d. For precise statements of the bounds, we refer the reader to the theorem statements.

say the ground state $|\phi\rangle$ to a local-Hamiltonian problem, it suffices to consider a phase state $\frac{1}{\sqrt{2^n}}\sum_x (-1)^{F(x)}|x\rangle$ which has a good overlap to $|\phi\rangle$. To this end, they show a strong property that, for *every* state $|\tau\rangle$ and a random Clifford operator U (or, more generally, an element of some unitary 2-design), the state $U|\tau\rangle$ has constant overlap with a phase state [Ira+21, Lemma A.5]. Our learning result implicitly shows that, assuming $\mathsf{QMA} \neq \mathsf{QCMA}$, then the phase state that has constant overlap with the ground space energy of the local Hamiltonian problem, cannot be of low degree.

Learning quantum circuits. Given access to a quantum circuit U, the goal of this learning task is to learn a circuit representation of U. The sample complexity for learning a general n-qubit quantum circuit is known to be $2^{\Theta(n)}$ [CN97; MRL08], which is usually impractical.

If we restrict ourselves to particular classes of quantum circuits, there are some known results for efficient learnability. Low [Low09] showed that an n-qubit Clifford circuit can be learned using $O(n)$ samples. However, this result was only an existential proof and requires access to the conjugate of the circuit. Constructive algorithms were given in Low [Low09], and Lai and Cheng [LC22], both of which showed that Clifford circuits can be learned using $O(n^2)$ samples. Both these algorithms require entangled measurements with the former

algorithm using pretty-good measurement [HW12], and the latter using Bell sampling. In this work, we show that Clifford circuits producing degree-2 binary phase states, can be learned in $O(n^2)$ samples, matching their result but only using separable measurements. Moreover, Low [Low09] also gave an existential proof of algorithms for learning circuits in the d-th level of the Clifford hierarchy, using $O(n^{d-1})$ samples. In this work, we give constructive algorithms for learning the diagonal elements of the Clifford hierarchy in $O(n^d)$ samples using separable measurements. A direct result of this is that a subset of IQP circuits, which are also believed to be hard to simulate classically [BJS11; BMS16], are shown to be efficiently learnable. Our learning result thus gives an efficient method for verifying IQP circuits that may be part of quantum-advantage experiments [BMS17; NBG21].

Learning hypergraph states. We finally observe that degree-3 (and higher-degree) phase states have appeared in works [Ros+13; TMH19] on measurement-based quantum computing (MBQC), wherein they refer to these states as *hypergraph states*. These works show that single-qubit measurements in the Pauli X or Z basis performed on a suitable degree-3 hypergraph state are sufficient for universal MBQC. Our learning algorithm gives a procedure for learning these states in polynomial-time and could potentially be used as a subroutine for verifying MBQC.

2.3 The intuition

In this section, we provide intuition behind our main results and briefly sketch their proofs.

Binary phase states

As we mentioned earlier, Montanaro [Mon17a] and Roettler [Röt09] showed how to learn degree-2 phase states using $O(n)$ copies of the state. Crucial to both their learning algorithms was the following so-called Bell-sampling procedure: given two copies of $|\psi_f\rangle = \frac{1}{\sqrt{2^n}}\sum_x (-1)^{f(x)}|x\rangle$ where $f(x) = x^\top Ax$ (where $A \in \mathbb{F}_2^{n\times n}$), perform n CNOTs from the first copy to the second, and measure the second copy. One obtains a uniformly random $y \in \mathbb{F}_2^n$ and the state

$$\frac{1}{\sqrt{2^n}}\sum_x (-1)^{f(x)+f(x+y)}|x\rangle = \frac{(-1)^{y^\top Ay}}{\sqrt{2^n}}\sum_x (-1)^{x^\top(A+A^\top)\cdot y}|x\rangle.$$

Using Bernstein-Vazirani [BV97] one can apply n-qubit Hadamard transform to obtain the bit string $(A + A^\top)\cdot y$. Repeating this process $O(n\log n)$ many times, one can learn n linearly independent constraints about A, and along with Gaussian elimination, allows one to learn A.[2]

Applying this same Bell-sampling procedure to degree-3 phase states does not easily learn the phase function. In this direction, from two copies of the degree-3 phase state $|\psi_f\rangle$ one obtains a uniformly random $y \in \mathbb{F}_2^n$ and the state $|\psi_{g_y}\rangle = \frac{1}{\sqrt{2^n}}\sum_x (-1)^{g_y(x)}|x\rangle$ for a degree-2

[2]It remains to learn the diagonal elements of A, but one can learn those using an extra step, which we discuss further in Theorem 9.

polynomial $g_y(x) = f(x) + f(x+y)$. One might now hope to apply the degree-2 learning algorithm from above, but since the single copy of $|\psi_{g_y}\rangle$ was randomly generated, it takes $\Omega(\sqrt{2^n})$ copies of $|\psi_f\rangle$ to obtain enough copies of $|\psi_{g_y}\rangle$. Our main idea is to circumvent this Bell-sampling approach and instead propose two techniques that allow us to learn binary phase states using separable and entangled measurements which we discuss further below.

Separable measurements, upper bound. Our first result is that we are able to learn binary phase states using separable measurements with sample complexity $O(n^d)$. In order to prove our upper bounds of sample complexity for learning with separable measurements, we make a simple observation. Given one copy of $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{f(x)} |x\rangle$, measure qubits $2, 3, \ldots, n$ in the computational basis. Suppose the resulting string is $y \in \{0,1\}^{n-1}$. The post-measurement state of qubit 1 is then given by

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}} \left[(-1)^{f(0y)} |0\rangle + (-1)^{f(1y)} |1\rangle \right].$$

By applying a Hadamard transform to $|\psi_{f,y}\rangle$ and measuring, the algorithm obtains $p_1(y) = f(0y) + f(1y) \mod 2$, which can be viewed as the derivative of f in the first direction at point y. Furthermore observe that p_1 is a degree $\leq d-1$ polynomial over $(n-1)$ variables. Hence, the learning algorithm repeatedly measures the last $(n-1)$ qubits and obtains $y^{(1)}, \ldots, y^{(M)}$ for $M = n^{d-1}$ and obtains $(y^{(k)}, p_1(y^{(k)}))$ for all $k = 1, 2, \ldots, M$ using the procedure above, which suffices to learn p_1 completely. Then the algorithm repeats the same procedure by measuring all the qubits except the second qubit in the computational basis and learns the derivative of f in the second direction. This is repeated over all the n qubits. Through this procedure, a learning algorithm learns the partial derivatives of f in the n directions and a simple argument shows that this is sufficient to learn f completely. This gives an overall sample complexity of $O(n^d)$. The procedure above only uses single qubit measurements in the $\{X, Z\}$ basis.

Separable measurements, lower bound Given the algorithm for learning binary phase states using separable measurements, a natural question is: Is the upper bound on sample complexity we presented above tight? Furthermore, suppose the learning algorithm was allowed to make arbitrary n-qubit measurements on a single copy of $|\psi_f\rangle$, instead of *single qubit* measurements (which are weaker than single *copy* measurements), then could we potentially learn f using fewer than $O(n^d)$ copies?

Here we show that if we allowed *arbitrary* single copy measurements, then a learning algorithm needs $\Omega(n^d)$ many copies of $|\psi_f\rangle$ to learn f. In order to prove this lower bound, our main technical idea is the following. Let f be a degree-d polynomial with n variables sampled uniformly at random. Suppose a learning algorithm measures the phase state $|\psi_f\rangle$ in an arbitrary orthonormal basis $\{U|x\rangle\}_x$. We show that the distribution describing the measurement outcome x is "fairly" uniform. In particular,

$$\mathop{\mathbb{E}}_f [H(x|f)] \geq n - O(1), \tag{2.5}$$

where $H(x|f)$ is the Shannon entropy of a distribution $P(x|f) = |\langle x|U^*|\psi_f\rangle|^2$. Thus, for a typical f, measuring one copy of the phase state $|\psi_f\rangle$ provides at most $O(1)$ bits of

information about f. Since a random uniform degree-d polynomial f with n variables has entropy $\Omega(n^d)$, one has to measure $\Omega(n^d)$ copies of ψ_f in order to learn f. To prove Eq. (2.5), we first lower bound the Shannon entropy by Renyi-two entropy and bound the latter by deriving an explicit formula for $\mathbb{E}_f[|\psi_f\rangle\langle\psi_f|^{\otimes 2}]$.

Entangled measurements. After settling the sample complexity of learning binary phase states using separable measurements, one final question question remains: Do entangled measurements help in reducing the sample complexity? For the case of quadratic polynomials, we know that Bell measurements (which are entangled measurements) can be used to learn these states in sample complexity $O(n)$. However, as mentioned earlier, it is unclear how to extend the Bell measurement procedure for learning larger degree polynomials.

Here, we give a learning algorithm based on the so-called pretty-good measurements (PGM) that learns $|\psi_f\rangle$ for a degree-d polynomial f using $O(n^{d-1})$ copies of $|\psi_f\rangle$. In order to prove this bound, we follow the following three step approach: (a) we first observe that in order to learn degree-d binary phase states, the *optimal* measurement is the pretty good measurement since the ensemble $\mathcal{S} = \{|\psi_f\rangle\}_f$ is geometrically uniform. By geometrically uniform, we mean that $\mathcal{S}$ can be written as $\mathcal{S} = \{U_f|\phi\rangle\}_f$ where $\{U_f\}_f$ is an Abelian group. (b) We next observe a property about the geometrically uniform state identification problem (which is new as far as we are aware): suppose $\mathcal{S}$ is a geometrically uniform ensemble, then the success probability of the PGM in correctly identifying f, given copies of $|\psi_f\rangle$, is *independent* of f, i.e., every element of the ensemble has the same probability of being identified correctly when measured using the PGM. (c) Finally, we need one powerful tool regarding the the weight distribution of Boolean polynomials: it was shown in [ASW15] that for any degree-d polynomial f, the following relation on $\mathsf{wt}(f)$ or the fraction of strings in $\{0,1\}^n$ for which f is non-zero holds:

$$|\{f \in \mathcal{P}(n,d) : \mathsf{wt}(f) \leq (1-\varepsilon)2^{-\ell}\}| \leq (1/\varepsilon)^{C\ell^4 \cdot \binom{n-\ell}{\leq d-\ell}},$$

for every $\varepsilon \in (0, 1/2)$ and $\ell \in \{1, \ldots, d-1\}$. Using this statement, we can comment on the average inner product of $|\langle\psi_f|\psi_g\rangle|$ over all ensemble members with $f \neq g \in \mathcal{P}(n,d)$. Combining this with a well-known result of PGMs, we are able to show that, given $M = O(n^{d-1})$ copies of $|\psi_f\rangle$ for $f \in \mathcal{S}$, the PGM identifies f with probability ≥ 0.99. Combining observations (a) and (b), the PGM also has the same probability of acceptance given an arbitrary $f \in \mathcal{S}$. Hence, we get an overall upper bound of $O(n^{d-1})$ for sample complexity of learning binary phase states using entangled measurements.

The lower bound for entangled measurement setting is straightforward: each quantum sample $\frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}(-1)^{f(x)}|x\rangle$ provides n bits of information and the goal is to learn f which contains $O(n^d)$ bits of information, hence by Holevo's bound, we need at least n^{d-1} quantum samples in order to learn f with high probability.

Implications for property testing. We remark that our learning algorithm can also be used in a naive way for property testing phase states. Let $\mathcal{C} = \{|\psi_f\rangle : f \in \mathcal{P}(n,d)\}$ be the class of degree-d phase states. The property testing question is: How many copies of an unknown $|\phi\rangle$ is sufficient to decide if $|\phi\rangle \in \mathcal{C}$ or $\min_{|\psi_f\rangle\in\mathcal{C}}\||\phi\rangle - |\psi_f\rangle\|_2 \geq 1/3$? As

far as we are aware, the only prior work in this direction is when $d = 1$ (using Bernstein-Vazirani [BV97]) and $d = 2$ (using [GNW21] which shows how to solve this task using 6 copies), but for larger d it is unclear what is the sample complexity. It is also unclear how to perform the property testing task (even for $d = 2$) using just *separable measurements*. Using our learning result, we get the following: take n^d copies of $|\phi\rangle$ and run our learning procedure using separable measurements.[3] If $|\phi\rangle = |\psi_f\rangle$, then our algorithm learns f. If $|\phi\rangle$ is not a phase state, the algorithm may fail, in which case the test classifies the state as a non-phase state. The worst case is if the algorithm succeeds and learns some incorrect phase state $|\psi_f\rangle$ from the non-phase input state. So, after running the learning algorithm and obtaining $|\psi_f\rangle$, use $O(1)$ more copies of $|\phi\rangle$ and run a swap test between $|\phi\rangle$ and $|\psi_f\rangle$, which succeeds with probability 1 if $|\phi\rangle = |\psi_f\rangle$ and rejects with probability at least $\Omega(1)$ if $\min_{|\psi_f\rangle\in\mathcal{C}}\||\phi\rangle - |\psi_f\rangle\|_2 \geq 1/3$.

Generalized phase states

As far as we are aware, ours is the first work that considers the learnability of generalized phase states (using either entangled or separable measurements). The sample complexity upper bounds follow the same high-level idea as that in the binary phase state setting. However, we need a few more technical tools for the generalized setting which we discuss below.

Separable bounds. At a high-level, the learning procedure for generalized phase states is similar to the procedure for learning binary phase states with the exception of a couple of subtleties that we need to handle here. Suppose we perform the same procedure as in binary phase states by measuring the last $(n-1)$ qubits in the computational basis. We then obtain a uniformly random $y \in \mathbb{F}_2^{n-1}$, and the post-measurement state for a generalized phase state is given by

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}}(\omega_q^{f(0y)}|0\rangle + \omega_q^{f(1y)}|1\rangle).$$

This state is proportional to $(|0\rangle + \omega_q^c|1\rangle)/\sqrt{2}$, where $c = f(1y) - f(0y) \pmod q$. In the binary case, $q = 2$, the states associated with $c = 0$ and $c = 1$ are orthogonal, so that the value of c can be learned with certainty by measuring $|\psi_{f,y}\rangle$ in the Pauli X basis. However, in the generalized case, $q > 2$, the states $(|0\rangle + \omega_q^c|1\rangle)/\sqrt{2}$ with $c \in \mathbb{Z}_q$ are not pairwise orthogonal. It is then unclear how to learn c given a single copy of $|\psi_{f,y}\rangle$. However, we observe that it is still possible to obtain a value $b \in \mathbb{Z}_q$ such that $b \neq c$ with certainty. To this end, consider a POVM whose elements are given by $\mathcal{M} = \{|\phi_b\rangle\langle\phi_b|\}_{b\in\mathbb{Z}_q}$, where $|\phi_b\rangle = \frac{1}{\sqrt{2}}(|0\rangle - \omega_q^b|1\rangle)$. Applying this POVM $\mathcal{M}$ onto an unknown state $(|0\rangle + \omega_q^c|1\rangle)/\sqrt{2}$ we observe that c is the outcome with probability 0 and furthermore *every* other outcome $b \neq c$ appears with non-negligible probability $\Omega(q^{-3})$.

Hence with one copy of $\frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}\omega_q^{f(x)}|x\rangle$, we obtain uniformly random $y \in \{0,1\}^{n-1}$ and $b \in \mathbb{Z}_q$ such that $f(1y) - f(0y) \neq b$. We now repeat this process $m = O(n^{d-1})$ many times and obtain $(y^{(k)}, b^{(k)})$ for $k = 1, 2, \ldots, M$ such that $f(1y^{(k)}) - f(0y^{(k)}) \neq b^{(k)}$ for all $k \in [M]$. We next show a variant of the Schwartz-Zippel lemma in the following sense: that

[3] We could also use n^{d-1} copies of $|\phi\rangle$ and run our learning procedure using entangled measurements.

for every $f \in \mathcal{P}_q(n,d)$ and $c \in \mathbb{Z}_q$, then either f is a constant function or the fraction of $x \in \mathbb{F}_2^n$ for which $f(x) \neq c$ is at least 2^{-d}. Using this, we show that after obtaining $O(2^d n^{d-1})$ samples, we can find a polynomial $g \in \mathcal{P}_q(n-1,d-1)$ for which $f(1y) - f(0y) = g(y)$. We now repeat this protocol for n different directions (by measuring each of the n qubits in every iteration) and we learn all the n directional derivatives of f, which suffices to learn f completely.

Entangled bounds. We do not give a result on learning generalized phase states with entangled measurements. We expect the proof of the sample complexity upper bound for learning generalized phase states using entangled measurements should proceed similarly to our earlier analysis of learning binary phase states using entangled measurements. However, we need a new technical tool that generalizes the earlier work on the weight distribution [ASY20] of Boolean functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ to those of form $f : \mathbb{F}_2^n \to \mathbb{Z}_q$ with $q = 2^d$.

Learning with further constraints

Learning sparse and low-Fourier degree states. A natural constraint to put on top of having low $\mathbb{F}_2$-degree in the polynomial is the sparsity, i.e., number of monomials in the $\mathbb{F}_2$ decomposition of f. Sparse low-degree phase states appear naturally when learning circuits with few gates. In particular, suppose we are learning a quantum circuit U with s gates from $\{\mathsf{Z}, \mathsf{CZ}, \ldots, \mathsf{C}^{d-1}\mathsf{Z}\}$ (where $\mathsf{C}^m\mathsf{Z}$ is the controlled-Z gate with m controls), then the output of $U|+\rangle^{\otimes n}$ is a phase state with sparsity-s and degree-d.

One naive approach to learn sparse $\mathbb{F}_2$ polynomials is to directly apply our earlier learning algorithm for binary phase states but this ignores the $\mathbb{F}_2$-sparsity information, and doesn't improve the sample complexity. Instead, here we use ideas from compressed sensing [DM09] to propose a linear program that allows us to improve the sample complexity to $O(2^d sn)$. Finally we make an observation that, if the function has *Fourier*-degree d, then one can learn f, given only $O(2^d \log n)$ many copies of $|\psi_f\rangle$, basically using the fact that there are only 2^{2^d} many such functions, each having at least a 2^{-d} distance between them.

Learning with depolarizing noise. One motivation for learning stabilizer states was potential experimental demonstrations of the learning algorithm [Roc+19]. Here, we consider a theoretical framework in order to understand the sample complexity of learning degree-2 phase states under global and local depolarizing noise. In this direction, we present two results. Under global depolarizing noise, i.e., when we are given $\rho_f = (1-\varepsilon)|\psi_f\rangle\langle\psi_f| + \varepsilon \cdot \mathbb{I}$, then it suffices to take $O(n^{1+\varepsilon})$ many copies ρ_f in order to learn f. The crucial observation is that one can use Bell sampling to reduce learning ρ_f to learning parities with noise, which we can accomplish using $O(n^{1+\varepsilon})$ samples and in time $2^{n/(\log\log n)}$ [Lyu05]. Additionally, however, a simple argument reveals that under local depolarizing noise, the sample complexity of learning stabilizer states is exponential in n.

Organization. In Section 2.4, we introduce phase states, discuss separable and entangled measurements. In Section 2.5, we prove our upper and lower bounds for learning binary phase states with separable and entangled measurements. In Section 2.6, we prove our results on

learning sparse and low-Fourier-degree phase states. In Section 2.7, we prove our upper bound for learning generalized phase states using separable and entangled measurements. Finally, in Section 2.9 we explicitly discuss the connection between phase states, and the diagonal unitaries in the d-th level of the Clifford hierarchy and IQP circuits.

2.4 Preliminaries

2.4.1 Notation.

Let $[n] = \{1, \ldots, n\}$. Let e_i be an n-dimensional vector with 1 in the ith coordinate and 0s elsewhere. We denote the finite field with the elements $\{0, 1\}$ as $\mathbb{F}_2$ and the ring of integers modulo q as $\mathbb{Z}_q = \{0, 1, \ldots, q-1\}$ with q usually being a power of 2 in this work. For a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, the bias of f is defined as

$$\text{bias}(f) = \mathop{\mathbb{E}}_x[(-1)^{f(x)}],$$

where the expectation is over a uniformly random $x \in \{0,1\}^n$. For $g : \mathbb{F}_2^n \to \mathbb{Z}_{2^d}$, the bias of g in the coordinate $j \in \mathbb{F}_{2^d}^\star$ is defined as $\text{bias}_j(g) = \mathbb{E}_x[(\omega_{2^d})^{j \cdot g(x)}]$. For a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, $y \in \mathbb{F}_2^{n-1}$ and $k \in [n]$, we denote $(D_k f)(y) = f(y^{k=1}) + f(y^{k=0})$, where $y^{i=1}, y^{i=0} \in \mathbb{F}_2^n$ is defined as: the ith bit of $y^{i=1}$ equals 1 and $y^{i=0}$ equals 0 and otherwise equals y.

2.4.2 Boolean functions

$\mathbb{F}_2$ representation. A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ can be uniquely represented by a polynomial over $\mathbb{F}_2$ as follows:

$$f(x) = \sum_{J \subseteq [n]} \alpha_J \prod_{i \in J} x_i \pmod 2, \tag{2.6}$$

where $\alpha_J \in \{0, 1\}$. Similar to Eq. (2.6), we can write Boolean functions $f : \mathbb{F}_2^n \to \mathbb{Z}_q$ as

$$f(x) = \sum_{J \subseteq [n]} \alpha_J \prod_{i \in J} x_i \pmod q \tag{2.7}$$

for some integer coefficients $\alpha_J \in \{0, 1, \ldots, q-1\}$. Throughout this chapter, unless explicitly mentioned, we will be concerned with writing Boolean functions as a decomposition over $\mathbb{F}_2$ or $\mathbb{Z}_q$ with $q = 2^d$. The $\mathbb{F}_2$ degree of f is defined as

$$\deg(f) = \max\{|J| \colon \alpha_J \neq 0\}.$$

Similarly for polynomials over $\mathbb{Z}_{2^d}$, we can define the degree as the size of the largest monomial whose coefficient α_J is non-negative.

We will call $g : \mathbb{F}_2^n \to \mathbb{F}_2$ with $g = \prod_{i \in J} x_i$ as monic monomials over n variables of at most degree-d, characterized by set $J \subseteq [n]$, $|J| \leq d$. We will denote the set of these monic monomials by $\mathcal{M}(n, d)$. Note that $|\mathcal{M}(n, d)| = \sum_{j=0}^{d} \binom{n}{j} = O(n^d)$. We will denote the set of polynomials over n variables of $\mathbb{F}_2$-degree d as $\mathcal{P}(n, d)$. Note that these polynomials are just

linear combinations of monomials in $\mathcal{M}(n,d)$. We will denote the set of polynomials over n variables of $\mathbb{F}_2$-degree d with sparsity s as $\mathcal{P}(n,d,s)$. Similarly, we will denote $\mathcal{P}_q(n,d)$ as the set of all degree-d Boolean polynomials $f:\mathbb{F}_2^n \to \mathbb{Z}_q$ with n variables. In particular, one can specify any polynomial $f \in \mathcal{P}_q(n,d)$ by $O(dn^d)$ bits and $|\mathcal{P}_q(n,d)| \leq 2^{O(dn^d)}$.

Consider a fixed d, and any $x \in \mathbb{F}_2^n$. Let the d-evaluation of x, denoted by $\mathrm{eval}_d(x)$, be a column vector in $\mathbb{F}_2^{|\mathcal{M}(n,d)|}$ with its elements being the evaluations of x under different monomials $g \in \mathcal{M}(n,d)$. This can be expressed as follows:

$$\mathrm{eval}_d(x) = \left(\prod_{i \in J \subseteq [n], |J| \leq d} x_i \right)^\top \tag{2.8}$$

For a set of points $\mathbf{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(m)}) \in (\mathbb{F}_2^n)^m$, we will call the matrix in $\mathbb{F}_2^{|\mathcal{M}(n,d)| \times m}$ with its kth column corresponding to d-evaluations of $x^{(k)}$, as the d-evaluation matrix of $\mathbf{x}$, and denote it by $Q_{\mathbf{x}}$.

Fourier Decomposition A Boolean function $f:\mathbb{F}_2^n \to \mathbb{F}_2$ admits the following Fourier decomposition

$$f(x) = \sum_{J \subseteq [n]} f\widehat{f}_J \chi_J(x), \tag{2.9}$$

where J are subsets of $[n] = \{1,2,\ldots,n\}$ and $\chi_J(x) = (-1)^{x_J}$ where $x_J = \sum_{i \in J} x_i$. Additionally the Fourier coefficients are defined as $f\widehat{f}_J = \mathbb{E}_x[f(x)\chi_J(x)]$. The *Fourier degree* of f is defined as $\max_J\{|J|\colon f\widehat{f}_J \neq 0\}$. Note that here all arithmetic operations use the field of real numbers $\mathbb{R}$, as opposed to the modular arithmetics used in the previous subsections.

2.4.3 Phase states

Binary Phase State For a Boolean function $f:\{0,1\}^n \to \{0,1\}$, we define a binary phase state as the n-qubit state given by

$$|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} |x\rangle. \tag{2.10}$$

We use the subscript f since $|\psi_f\rangle$ is characterized by f.

Generalized Phase State We will also consider degree-d generalized phase states of the form

$$|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} \omega_q^{f(x)} |x\rangle, \tag{2.11}$$

where $\omega_q = e^{2\pi i/q}$ and $f:\mathbb{F}_2^n \to \mathbb{Z}_q$, with $q = 2^d$, is a degree-d polynomial. We consider $\mathbb{Z}_q = \{0,1,\ldots,q-1\}$ to be the ring of integers modulo q.

2.4.4 Useful lemmas

Let $e_i \in \mathbb{F}_2^n$ denote the vector of all zeros except for a 1 in the i^{th} coordinate.

Fact 1. *Let $d \in [n]$, $s \leq |\mathcal{M}(n,d)| = \sum_{k=1}^{d} \binom{n}{k}$, and $f \in \mathcal{P}(n,d,s)$. There exists $g_i \in \mathcal{P}(n,d-1,s)$ such that $g_i(x) = f(x+e_i) + f(x) \pmod 2$ for all $x \in \{0,1\}^n$.*

The proof of this fact is straightforward. Without loss of generality, consider $i = 1$. For every $f(x) = \sum_S \alpha_S \prod_{i \in S} x_i$, we can express it as

$$f(x) = x_1 p_1(x_2, \ldots, x_n) + p_2(x_2, \ldots, x_n),$$

where p_1 has degree $\leq d-1$ and p_2 has degree $\leq d$. Observe that $f(x+e_1) - f(x)$ is either $p_1(x_2, \ldots, x_n)$ or $-p_1(x_2, \ldots, x_n)$ which has degree $d-1$ and corresponds to the polynomial g_1 in the fact statements. This applies for every coordinate i.

Note that the polynomial g_i above is also often called the *directional* derivative of f in direction w and is denoted as $D_i f$.

Fact 2. *Let $N, s \geq 1$ such that $\gamma = s/N \leq 1/2$. Then we have*

$$\sum_{\ell=1}^{s} \binom{N}{\ell} \leq 2^{H_b(\gamma)N} \leq 2^{2\gamma \log(1/\gamma)}.$$

where we used above that $H_b(\gamma) = \gamma \log \frac{1}{\gamma} + (1-\gamma) \log \frac{1}{1-\gamma} \leq 2\gamma \log \frac{1}{\gamma}$ (for $\gamma \leq 1/2$).

Lemma 1 (The Schwartz-Zippel Lemma)**.** *Let $p(y_1, \ldots, y_n)$ be a nonzero polynomial on n variables with degree d. Let S be a finite subset of $\mathbb{R}$, with at least d elements in it. If we assign $y_1, \ldots, y_n$ values from S independently and uniformly at random, then*

$$\Pr[p(y_1, \ldots, y_n) = 0] \leq \frac{d}{|S|}. \tag{2.12}$$

Lemma 2 ([NS94])**.** *Let $p(x_1, \ldots, x_n)$ be a non-zero multilinear polynomial of degree d. Then*

$$\Pr_{x \in \{0,1\}^n}[p(x) = 0] \leq 1 - 2^{-d},$$

where the probability is over a uniformly random distribution on $\{0,1\}^n$.

We will also need the following structural theorem about Reed-Muller codes which comments on the weight distribution of Boolean functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$.

Theorem 1 ([ASY20, Theorem 3])**.** *Let $n \geq 1$ and $d \leq n/2$. Define $|f| = \sum_{x \in \{0,1\}^n}[f(x) = 1]$ and $\mathsf{wt}(f) = |f|/2^n$. Then, for every $\varepsilon \in (0, 1/2)$ and $\ell \in \{1, \ldots, d-1\}$, we have that*

$$|\{f \in P(n,d) : \mathsf{wt}(f) \leq (1-\varepsilon)2^{-\ell}\}| \leq (1/\varepsilon)^{C\ell^4 \cdot \binom{n-\ell}{\leq d-\ell}}.$$

Fix $w = (1-\varepsilon)2^{n-\ell}$ and we get

$$|\{f \in P(n,d) : |f| \leq w\}| \leq (1 - w/2^{n-\ell})^{-C\ell^4 \cdot \binom{n-\ell}{\leq d-\ell}}.$$

Lemma 3 (Fano's inequality)**.** *Let A and B be classical random variables taking values in $\mathcal{X}$ (with $|\mathcal{X}| = r$) and let $q = \Pr[\mathsf{A} \neq \mathsf{B}]$. Then,*

$$H(\mathsf{A}|\mathsf{B}) \leq H_b(q) + q \log(r-1),$$

where $H(\mathsf{A}|\mathsf{B})$ is the conditional entropy and $H_b(q)$ is the standard binary entropy.

2.4.5 Measurements

Throughout this chapter we will be concerned with learning algorithms that use either separable or entangled measurements. Given $|\psi_f\rangle^{\otimes k}$, a learning algorithm for f is said to use separable measurements if it only measure each copy of $|\psi_f\rangle$ separately in order to learn f. Similarly, a learning algorithm for f is said to use entangled measurements if it makes an entangled measurement on the k-fold tensor product $|\psi_f\rangle^{\otimes k}$. In this direction, we will often use two techniques which we discuss in more detail below: sampling random partial derivatives in order to learn from separable measurements and Pretty Good Measurements in order to learn from entangled measurements.

Separable Measurements

Below we discuss a subroutine that we will use often to learn properties about $f : \mathbb{F}_2^n \to \mathbb{F}_2$: given a single copy of $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle$, the subroutine produces a uniformly random $y \in \mathbb{F}_2^{n-1}$ and $f(1y) + f(0y) \pmod 2$. To this end, suppose we measure qubits $2, 3, \ldots, n$ of $|\psi_f\rangle$ in the usual Z basis. We denote the resulting string as $y \in \{0,1\}^{n-1}$. The post-measurement state of qubit 1 is then given by

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}} \left[(-1)^{f(0y)}|0\rangle + (-1)^{f(1y)}|1\rangle\right]. \tag{2.13}$$

We note that $|\psi_{f,y}\rangle$ is then an X-basis state ($|+\rangle$ or $|-\rangle$) depending on the values of $f(1y)$ and $f(0y)$. If $f(1y) = f(0y)$, then $|\psi_{f,y}\rangle = |+\rangle$ and if $f(1y) = f(0y) + 1 \pmod 2$, then $|\psi_{f,y}\rangle = |-\rangle$. Measuring qubit 1 in the X-basis and qubits $2, 3, \ldots, n$ in the Z-basis thus produces examples of the form (y, b) where $y \in \{0,1\}^{n-1}$ is uniformly random and $b = f(0y) + f(1y) \pmod 2$. Considering Fact 1 with the basis of e_1, we note that theses examples are of the form $(y, D_1 f(y))$, where $D_1 f(y) = f(1y) + f(0y) \pmod 2$ is the partial derivative of f along direction e_1. Changing the measurement basis chosen above to $ZZ \cdots X_k \cdots Z$ such that we measure all the qubits in the Z basis except for the kth qubit which is measured in the X basis, will allow us to obtain random samples of the form $(y, D_k f(y))$. Accordingly, we introduce a new subroutine.

Definition 1 (Random Partial Derivative Sampling (RPDS) along e_k)**.** *For every $k \in [n]$, measuring every qubit of $|\psi_f\rangle$ in the Z basis, except the kth qubit which is measured in the X basis, we obtain a uniformly random $y \in \mathbb{F}_2^{n-1}$ and $(D_k f)(y)$.*

Entangled Measurements

In general one could also consider a joint measurement applied to multiple copies of $|\psi_f\rangle$, which we refer to as entangled measurements. In this work, we will generally consider two types of entangled measurements, Bell sampling and the pretty-good measurement (which we discuss in more detail in the next section). Bell sampling is a procedure that involves measuring a quantum state (or in this case, two copies of a quantum state) in the Bell basis.[4] We will use the following version of Bell sampling that applies to the scenario where we are

[4]The Bell basis is the basis given by $\left\{\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle), \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)\right\}$.

given noisy copies of degree-2 phase states (for global depolarizing noise), and covers the core idea of standard Bell sampling.

Lemma 4 (Bell sampling)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a degree-2 polynomial, i.e., $f(x) = x^\top Ax$ (for an upper triangular $A \in \mathbb{F}_2^{n\times n}$). Using two copies of $\rho = (1-\varepsilon)|\psi_f\rangle\langle\psi_f| + \varepsilon\mathbb{I}$, there exists a procedure that outputs a uniformly random $z \in \{0,1\}^n$ and $(A + A^\top)\cdot z \in \{0,1\}^n$. Additionally, the same procedure, when given two copies of ψ_n, outputs a uniformly random $z \in \{0,1\}^n$ and $w_z \in \{0,1\}^n$ that satisfies:*

$$w_z = \begin{cases} (A + A^\top)\cdot z & \text{w.p. } (1-\varepsilon)^2 \\ \text{uniformly random } v & \text{w.p. } 1-(1-\varepsilon)^2. \end{cases}$$

Proof. We first consider the case where $\varepsilon = 0$. Then, the following procedure produces $(z, (A + A^\top)\cdot z)$ for a uniformly random z. Take two copies of $|\psi\rangle$

$$\begin{aligned} |\psi_f\rangle \otimes |\psi_f\rangle &= \sum_{x,y\in\mathbb{F}_2^n} (-1)^{x^\top Ax + y^\top Ay} |x, y\rangle \\ &\xrightarrow{\mathsf{CNOT}} \sum_{x,y} (-1)^{x^\top Ax + y^\top Ay} |x, x+y\rangle \\ &= \sum_{x,z} (-1)^{x^\top Ax + (x+z)^\top A(x+z)} |x, z\rangle = \sum_{x,z} (-1)^{x^\top (A+A^\top) z + z^\top Az} |x, z\rangle \end{aligned}$$

Measure the second register and suppose we obtain $\tilde{z}$, resulting state is

$$(-1)^{\tilde{z}^\top A\tilde{z}} \Big(\sum_x (-1)^{x^\top (A+A^\top)\tilde{z}} |x\rangle \Big) |\tilde{z}\rangle \xrightarrow{\mathsf{H}^n} |(A + A^\top)\cdot\tilde{z}\rangle|\tilde{z}\rangle,$$

where H^n is the n-qubit Hadamard transform. Let us now consider the case where $\varepsilon > 0$. Given $\rho^{\otimes 2}$, with probability $(1-\varepsilon)^2$, we obtain $(A + A^\top)\cdot\tilde{z}$ and it is not hard to see that on input $\mathbb{I}^{\otimes 2}$ or $|\psi_f\rangle\langle\psi_f| \otimes \mathbb{I}$, the output of the procedure above produces a uniformly random bit string $v \in \mathbb{F}_2^n$. □

Pretty Good Measurements. Consider an ensemble of quantum states, $\mathcal{E} = \{(p_i, |\psi_i\rangle)\}_{i\in[m]}$, where $p = \{p_1, \ldots, p_m\}$ is a probability distribution. In the quantum state identification problem, a learning algorithm is given an unknown quantum state $|\psi_i\rangle \in \mathcal{E}$ sampled according to the distribution p and the learning algorithm needs to identity i with probability $\geq 2/3$. In this direction, we are interested in maximizing the average probability of success to identify i. For a POVM specified by positive semidefinite matrices $\mathcal{M} = \{M_i\}_{i\in[m]}$, the probability of obtaining outcome j equals $\langle\psi_i|M_j|\psi_i\rangle$ and the average success probability is given by

$$P_{\mathcal{M}}(\mathcal{E}) = \sum_{i=1}^m p_i \langle\psi_i|M_i|\psi_i\rangle.$$

Let $P^{opt}(\mathcal{E}) = \max_{\mathcal{M}} P_{\mathcal{M}}(\mathcal{E})$ denote the optimal average success probability of $\mathcal{E}$, where the maximization is over the set of valid m-outcome POVMs. For every ensemble $\mathcal{E}$, the so-called

Pretty Good Measurement (PGM) is a specific POVM (depending on the ensemble $\mathcal{E}$) that does *reasonably* well against $\mathcal{E}$. In particular, it is well-known that

$$P^{opt}(\mathcal{E})^2 \leq P^{PGM}(\mathcal{E}) \leq P^{opt}(\mathcal{E}).$$

We now define the POVM elements of the pretty-good measurement. Let $|\psi_i'\rangle = \sqrt{p_i}|\psi_i\rangle$, and $\mathcal{E}' = \{|\psi_i'\rangle : i \in [m]\}$ be the set of states in $\mathcal{E}$, renormalized to reflect their probabilities. Define $\rho = \sum_{i\in[m]} |\psi_i'\rangle\langle\psi_i'|$. The PGM is defined as the set of measurement operators $\{|\nu_i\rangle\langle\nu_i|\}_{i\in[m]}$ where $|\nu_i\rangle = \rho^{-1/2}|\psi_i'\rangle$ (the inverse square root of ρ is taken over its non-zero eigenvalues). We will use the properties of these POVM elements later on and will also need the following theorems about PGMs.

Theorem 2 ([HW12]). *Let $\mathcal{S} = \{\rho_1, \ldots, \rho_m\}$. Suppose $\rho \in \mathcal{S}$ is an unknown quantum state picked from $\mathcal{S}$. Let $\max_{i\neq j} \|\sqrt{\rho_i}\sqrt{\rho_j}\|_1 \leq F$. Then, given*

$$M = O((\log(m/\delta))/\log(1/F))$$

copies of ρ, the Pretty good measurement identifies ρ with probability at least $1-\delta$.

The above theorem in fact implies the following stronger statement immediately (also stated in [BK02a]) that we use here.

Corollary 1. *Let $\mathcal{S} = \{\rho_1, \ldots, \rho_m\}$. Suppose $\rho \in \mathcal{S}$ is an unknown quantum state picked uniformly from $\mathcal{S}$. Suppose there exists k such that*

$$\frac{1}{m}\sum_{i\neq j} \|\sqrt{\rho_i^{\otimes k}}\sqrt{\rho_j^{\otimes k}}\|_1 \leq \delta,$$

then given k copies of ρ, the Pretty good measurement identifies ρ with probability at least $1-\delta$.

2.5 Learning binary phase states

In this section, we consider the problem of learning binary phase states as given by Eq. (2.10), assuming that f is a Boolean polynomial of $\mathbb{F}_2$-degree d.

2.5.1 Learning algorithm using separable measurements

We now describe our learning algorithm for learning binary phase states $|\psi_f\rangle$ when f has $\mathbb{F}_2$-degree d, using separable measurements. We carry out our algorithm in n rounds, which we index by t. In the t-th round, we perform RPDS along e_t (Def. 1) in order to obtain samples of the form $(y, D_t f(y))$ where $y \in \{0,1\}^{n-1}$. For an $m \geq 1$ to be fixed later, we use RPDS on m copies of $|\psi_f\rangle$ to obtain $\{\left(y^{(k)}, D_t f(y^{(k)})\right)\}_{k\in[m]}$ where $y^{(k)} \in \{0,1\}^{n-1}$ is uniformly random. We now describe how to learn $D_t f$ using these m samples.

Using Fact 1, we know that $D_t f \in \mathcal{P}(n-1, d-1)$. Thus, there are at most $N = |\mathcal{M}(n-1, d-1)| = \sum_{k=1}^{d-1} \binom{n}{k} = n^{O(d)}$ monomials in the $\mathbb{F}_2$ representation of $D_t f$. Let $A_t \in \mathbb{F}_2^{m\times N}$ be the transpose of the $(d-1)$-evaluation matrix (defined in Eq. (2.8)), such that the kth row of A_t corresponds to the evaluations of $y^{(k)}$ under all monomials in $\mathcal{M}(n-1, d-1)$,

i.e., $(y_S^{(k)})_{|S|\leq d-1}$, where $y_S^{(k)} = \prod_{j\in S} y_j^{(k)}$, and let $\beta_t = (\alpha_S)_{|S|\leq d-1}$ be the vector of unknown coefficients. Obtaining $\{(y^{(k)}, D_t f(y^{(k)}))\}_{k\in[m]}$, allows one to solve $A_t\beta_t = D_t f(\mathbf{y})$ for β_t (where $\mathbf{y} = (y^{(1)}, \ldots, y^{(m)})$ and $(D_t f(\mathbf{y}))_k = D_t f(y^{(k)})$) and learn the $\mathbb{F}_2$-representation of $D_t f$ completely. Over n rounds, one then learns $D_1 f, D_2 f, \ldots, D_n f$. The $\mathbb{F}_2$-representations of these partial derivatives can then be used to learn f completely, as show in Fact 3. This procedure is shown in Algorithm 1.

Fact 3. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be such that $f \in \mathcal{P}(n, d)$. Learning $D_1 f, \ldots, D_n f$ suffices to learn f.*

Proof. Let the $\mathbb{F}_2$-representation of the unknown f be

$$f(x) = \sum_{J\subseteq[n],|J|\leq d} \alpha_J \prod_{i\in J} x_i. \tag{2.14}$$

The $\mathbb{F}_2$-representation of $D_t f$ for any $t \in \{1, 2, \ldots, n\}$ is then given by

$$D_t f(x) = \sum_{\substack{J\subseteq[n]:\\ t\in J,|J|\leq d}} \alpha_J \prod_{i\in J\setminus t} x_i, \tag{2.15}$$

where we notice that $D_t f$ only contains those monomials that correspond to sets J containing the component x_t. Let the $\mathbb{F}_2$-representation of $D_t f$ with the coefficient vector β_t be given by

$$D_t f(x) = \sum_{S\subset[n],|S|\leq d-1} (\beta_t)_S \prod_{i\in S} x_i. \tag{2.16}$$

Suppose an algorithm learns $D_1 f, \ldots, D_n f$. In order to learn f, we must retrieve the coefficients α_J from the learned coefficients $\{\beta_t\}_{t\in\{1,2,\ldots,n\}}$. We accomplish this by noting that $(\beta_t)_S = \alpha_{S\cup t}$ or in other words, $\alpha_J = \{\beta_t\}_{J\setminus t}$, $t \in J$. However, there may be multiple values of t that will allow us retrieve α_J. For example, suppose f contains the monomial term $x_1x_2x_3$ (i.e., $J = \{1, 2, 3\}$) then $\alpha_{\{1,2,3\}}$ could be retrieved from $(\beta_1)_{\{2,3\}}$, $(\beta_2)_{\{1,3\}}$, or $(\beta_3)_{\{1,2\}}$. When $D_t f$ (or β_t) for all t is learned with zero error, all these values coincide and it doesn't matter which learned coefficient is used. When there may be error in learning $D_t f$ (or β_t), we can carry out a majority vote: $\alpha_J = \text{Majority}(\{(\beta_t)_{J\setminus t} | t \in J\})$ for all $J \subseteq [n], |J| \leq d$. The majority vote is guaranteed to succeed as long as there is no error in at least half of the contributing β_t (which is the case in our learning algorithm). □

We now prove the correctness of this algorithm.

Theorem 3. *Let $n \geq 2, d \leq n/2$. Algorithm 1 uses $M = O(2^d n^d)$ copies of an unknown $|\psi_f\rangle$ for $f \in \mathcal{P}(n, d)$ and with high probability identifies f using single qubit X, Z measurements.*

Proof. Algorithm 1 learns f by learning $D_1 f, \ldots, D_n f$ and thereby learns f completely. Here we prove that each $D_t f$ can be learned with $m = O(2^d n^{d-1})$ copies of $|\psi_f\rangle$ and an exponentially small probability of error. This results in an overall sample complexity of $O(2^d n^d)$ for learning f and hence $|\psi_f\rangle$. Let us consider round t in Algorithm 1. We generate m constraints $\{\left(y^k, (D_t f)(y^{(k)})\right)\}_{k\in[m]}$ where $y^{(k)} \in \mathbb{F}_2^{n-1}$ by carrying out RPDS along e_t on m copies of $|\psi_f\rangle$.

We learn the $\mathbb{F}_2$-representation of $D_t f$ by setting up a linear system of equations using these m samples: $A_t\beta_t = D_t f(\mathbf{y})$, where A_t is the transposed $(d-1)$-evaluation matrix

Algorithm 1 Learning binary phase states through separable measurements

Input: Given $M = O((2n)^d)$ copies of $|\psi_f\rangle$ where $f \in \mathcal{P}(n,d)$

1: **for** qubit $t = 1, \ldots, n$ **do**
2: Set $m = M/n$
3: Perform RPDS along e_t to obtain $\{(y^{(k)}, D_t f(y^{(k)})\}_{k\in[m]}$ by measuring m copies of $|\psi_f\rangle$.
4: Solve the linear system of equations $A_t \cdot \beta_t = D_t f(\mathbf{y})$ to learn $D_t f$ explicitly.
5: **end for**
6: Use Fact 3 to learn f using $D_1 f, \ldots, D_n f$ (let $\tilde{f}$ be the output).

Output: Output $\tilde{f}$

in round t, evaluated over $\mathbf{y} = (y^{(1)}, y^{(2)}, \ldots y^{(m)})$, and $\beta_t \in \mathbb{F}_2^{|\mathcal{M}(n-1,d-1)|}$ is the collective vector of coefficients corresponding to the monomials in $\mathcal{M}(n-1, d-1)$. By construction, this system has at least one solution. If there is exactly one solution, then we are done. Otherwise, the corresponding system has a non-zero solution, that is, there exists a non-zero degree-$(d-1)$ polynomial $g : \mathbb{F}_2^{n-1} \to \mathbb{F}_2$ such that $g(y^{(j)}) = 0$ for all $j = 1, 2, \ldots, m$.

Below we prove that the probability of this bad event can be bounded through the Schwartz-Zippel lemma. Applying Lemma 2 and by noting that $y^j \in \mathbb{F}_2^{(n-1)}$ are independent and uniformly distributed, we have that

$$\Pr[g(y^{(1)}) = g(y^{(2)}) = \cdots = g(y^{(m)}) = 0] \leq (1 - 2^{-d})^m \leq e^{-m2^{-d}} \tag{2.17}$$

Let $\mathcal{P}_{\text{nnz}}(n,d)$ be the set of all degree-d polynomials $g : \mathbb{F}_2^n \to \mathbb{F}_2$ which are not identically zero. Define event

$$\mathsf{BAD}(y^1, \ldots, y^m) = [\exists g \in \mathcal{P}_{\text{nnz}}(n-1, d-1) \,:\, g(y^1) = \ldots = g(y^m) = 0 \pmod 2]. \tag{2.18}$$

We note that $|\mathcal{P}_{\text{nnz}}(n-1,d-1)| \leq 2^N$ where $N = O(n^{d-1})$. By union bound and Eq. (2.17), we have

$$\Pr[\mathsf{BAD}(y^{(1)}, \ldots, y^{(m)})] \leq |\mathcal{P}_{\text{nnz}}(n-1,d-1)| \cdot (1-2^{-d})^m \leq 2^{n^{d-1} - m2^{-d}(\ln 2)}. \tag{2.19}$$

Thus choosing $m = O((2n)^{d-1})$ is enough to learn all coefficients $\{\alpha_J\}_{t\in J}$ (through β_t) in the $\mathbb{F}_2$ representation of f with an exponentially small probability of error. We need to repeat this over all the n qubits in order to learn $D_1 f, \ldots, D_n f$ and then use Fact 3 to learn f completely. This gives an overall sample complexity of $O((2n)^d)$ for learning binary phase states. Observe that the only measurements that we needed in this algorithm were single qubit $\{X, Z\}$ measurements. □

Corollary 2. *An n-qubit state $|\psi_f\rangle$ with the unknown Boolean function f of given Fourier-sparsity s can be learned with Algorithm 1 that consumes M copies of $|\psi_f\rangle$ with probability $1 - 2^{-\Omega(n)}$ provided that $M \geq O(sn^{\log s})$.*

The proof of this corollary simply follows from the following: for a Boolean function, the Fourier sparsity s of f is related to the $\mathbb{F}_2$-degree d of f [BC99] as $d \leq \log s$. Along with Theorem 3 we obtain the corollary.

2.5.2 Learning using entangled measurements

We now consider the problem of learning binary phase states using entangled measurements. We have the following result.

Theorem 4. *Let $n \geq 2, d \leq n/2$. There exists an algorithm that uses $M = O((2n)^{d-1})$ copies of an unknown $|\psi_f\rangle$ for $f \in \mathcal{P}(n,d)$ and identifies f using entangled measurements with probability $\geq 2/3$. There is also a lower bound of $\Omega(n^{d-1})$ for learning these states.*

Proof. In order to prove this theorem, we follow the following steps. We first observe that the optimal measurement for our state distinguishing problem is the pretty good measurement (PGM). Second we observe that the success probability of the PGM is the same for *every* concept in the ensemble. We bound the success probability of the PGM using Corollary 1 we get our upper bound.

For $f \in \mathcal{P}(n,d)$, let U_f be the unitary defined as $U_f = \mathsf{diag}(\{(-1)^{f(x)}\}_x)$, that satisfies $U_f|+\rangle^n = |\psi_f\rangle$. Observe that the set $\{U_f\}_{f\in\mathcal{P}(n,d)}$ is an Abelian group. The ensemble we are interested in is $\mathcal{S} = \{U_f|+\rangle^n\}_{f\in\mathcal{P}(n,d)}$ and such an ensemble is called *geometrically uniform* if the $\{U_f\}$ is an Abelian group. A well-known result of Eldar and Forney [EF01] showed that the optimal measurement for state distinguishing a geometrically uniform (in particular $\mathcal{S}$) is the pretty-good measurement. We now show that the success probability of the PGM is the same for every state in the ensemble. In this direction, for $M \geq 1$, let $\sigma_f = |\psi_f\rangle\langle\psi_f|^{\otimes M}$. The POVM elements of the pretty good measurement $\{E_f : f \in \mathcal{P}(n,d)\}$ is given by the POVM elements $E_f = S^{-1/2}\sigma_f S^{-1/2}$ where $S = \sum_{f\in\mathcal{P}(n,d)} \sigma_f$. The probability that the PGM identifies the unknown σ_f is given by

$$\Pr(f) = \mathrm{Tr}(\sigma_f E_f) = \langle\psi_f^{\otimes M}|S^{-1/2}|\psi_f^{\otimes M}\rangle^2.$$

Our claim is that $\Pr(f)$ is the same for every $f \in \mathcal{P}(n,d)$. Using the Abelian property of the unitaries $\{U_f\}_f$, observe that $U_f|\psi_g\rangle = |\psi_{f\oplus g}\rangle$ for every $f, g \in \mathcal{P}(n,d)$. Thus, we have that $(U_f^{\otimes M})^\dagger S U_f^{\otimes M} = S$, which implies that$(U_f^{\otimes M})^\dagger S^{-1/2} U_f^{\otimes M} = S^{-1/2}$. Hence it follows that

$$\Pr(f) = (\langle+|^{\otimes M}(U_f^{\otimes M})^\dagger S^{-1/2} U_f^{\otimes M}|+\rangle^{\otimes M})^2 = (\langle+|^{\otimes M} S^{-1/2}|+\rangle^{\otimes M})^2 = \Pr(0),$$

for every $f \in \mathcal{P}(n,d)$. Finally, observe that $\langle\psi_f|\psi_g\rangle = \mathbb{E}_x\left[(-1)^{f(x)+g(x)}\right] = 1 - 2\Pr_x[f(x) \neq g(x)]$. Let $\mathcal{P}^*(n,d)$ be the set of non-constant polynomials in $\mathcal{P}(n,d)$. We now have the following

$$\frac{1}{2^{\binom{n}{\leq d}}} \sum_{\substack{f\neq g:\\ f,g\in P(n,d)}} \|\sqrt{\rho_f^{\otimes k}}\sqrt{\rho_g^{\otimes k}}\|_1 = \sum_{g\in P^*(n,d)} (1-2\Pr_x[g(x)=1])^{2k}$$

$$= \sum_{g\in P^*(n,d)} (1-2\mathsf{wt}(g))^{2k}$$

$$= \sum_{\ell=1}^{d-1} \sum_{g\in P^*(n,d)} (1-2|g|/2^n)^{2k} \cdot \left[|g|\in [2^{n-\ell-1}, 2^{n-\ell}-1]\right]$$

$$= \sum_{g\in P^*(n,d)} (1-2|g|/2^n)^{2k} \cdot \left[|g|\in [2^{n-2}, 2^{n-1}-1]\right]$$

$$+ \sum_{\ell=2}^{d-1} \sum_{g\in P^*(n,d)} (1-2|g|/2^n)^{2k} \cdot \left[|g|\in [2^{n-\ell-1}, 2^{n-\ell}-1]\right]$$

$$\leq 2^{n-1}2^{-2k+C\binom{n-1}{\leq d-1}} + \sum_{\ell=2}^{d-1} (1-\frac{1}{2^\ell})^{2k} \sum_{g\in P^*(n,d)} \left[|g|\leq 2^{n-\ell}\right],$$

where the first equality used that the PGM has the same success probability for every $f, g \in \mathcal{P}(n,d)$, third equality used that $|g| \geq 2^{n-d}$ for any non-zero polynomial $g \in P(n,d)$ [MS77] and last inequality used Theorem 1. For $k = O(n^{d-1})$ (by picking a sufficiently large constant in $O(\cdot)$), the first term is at most $\leq 1/100$. To bound the second term, using Theorem 1 we have

$$\sum_{\ell=2}^{d-1} (1-\frac{1}{2^\ell})^{2k} \sum_{g\in P^*(n,d)} \left[|g|\leq 2^{n-\ell}\right] \leq \sum_{\ell=2}^{d-1} 2^{n-\ell} \exp(-2k/2^\ell + (n-\ell)\ell^4 \binom{n-\ell}{\leq d-\ell}).$$

Each term is $\exp(-n^{d-1})$ for $k = O(n^{d-1})$, so the overall sum is $\leq 1/100$. Corollary 1 implies our desired upper bound.

In order to see the lower bound, observe that each state $|\psi_f\rangle$ contains n bits of information and the goal of the learning algorithm is to learn an unknown f, i.e., obtain $O(n^d)$ bits of information. Hence by Holevo's theorem [Hol73], one requires $\Omega(n^{d-1})$ copies of the unknown state for state identification.[5] □

2.5.3 Lower bounds

In the last section we saw that $\Theta(n^{d-1})$ many copies of $|\psi_f\rangle$ with degree-d are necessary and sufficient to learn f if we allowed only entangled measurements. Earlier we saw that $O(n^d)$ many copies of $|\psi_f\rangle$ sufficed to learn f using separable measurements. A natural question is: Can we learn f using fewer copies if we are restricted to using only separable measurements? In the theorem below, we provide a lower bound that complements our upper bound, thereby showing $\Theta(n^d)$ copies are necessary and sufficient to learn f using separable measurements.

[5] We refer the reader to Montanaro [Mon12, Proposition 1] for a detailed exposition of this lower bound proof.

Theorem 5. *Let $2 \leq d \leq n/2$. Suppose there exists an algorithm that with probability $\geq 1/10$, learns an n-variate polynomial $f \in \mathcal{P}(n,d)$, given M copies of the phase state $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle$, measuring each copy in an arbitrary orthonormal basis, and performing an arbitrary classical processing. Then*

$$M = \Omega(\log|\mathcal{P}(n,d)|) = \Omega(n^d). \tag{2.20}$$

Proof. Let U be an n-qubit unitary operator. Define the probability distribution

$$\Pr_U(x|f) = |\langle x|U|\psi_f\rangle|^2. \tag{2.21}$$

Let $H_U(x|f)$ be the Shannon entropy of x sampled from $\Pr_U(x|f)$, i.e.,

$$H_U(x|f) = -\sum_x \Pr_U(x|f) \log \Pr_U(x|f). \tag{2.22}$$

Below we prove the following.

Lemma 5. *Suppose $d \geq 2$ and U is an n-qubit unitary. Then*

$$\mathbb{E}_f[H_U(x|f)] \geq n - 2, \tag{2.23}$$

where the expectation is over a uniformly random $f \in \mathcal{P}(n,d)$

We will assume the lemma now and prove the theorem statement. Below we assume that $f \in \mathcal{P}(n,d)$ is picked uniformly at random. Suppose we measure the j-th copy of $|\psi_f\rangle$ in a basis $\{U_j^\dagger|x\rangle\}_x$ for some n-qubit unitary U_j. Let $x^1, x^2, \ldots, x^M \in \mathbb{F}_2^n$ be the measured bit strings. The joint probability distribution of f and x is given by

$$\Pr(f,x) = \frac{1}{|\mathcal{P}(n,d)|} \prod_{j=1}^M \Pr_{U_j}(x^j|f). \tag{2.24}$$

The conditional entropy of x given f is

$$H(x|f) = \frac{1}{|\mathcal{P}(n,d)|} \sum_{f\in\mathcal{P}(n,d)} \sum_{j=1}^M H_{U_j}(x^j|f) = \sum_{j=1}^M \mathbb{E}_f H_{U_j}(x^j|f) \geq M(n-2), \tag{2.25}$$

where the inequality used Lemma 5. It follows that the conditional entropy of f given x obeys

$$H(f|x) = H(x|f) - H(x) + H(f) \geq M(n-2) - H(x) + H(f). \tag{2.26}$$

Since $H(x) \leq nM$ and $H(f) = \log|\mathcal{P}(n,d)|$, we get

$$H(f|x) \geq \log|\mathcal{P}(n,d)| - 2M. \tag{2.27}$$

Assuming there exists a learning algorithm that given x learns f with probability $\geq 1/10$, by Fano's inequality (Lemma 3), we know that $H(f|x) \leq H_b(1/10) + (1/10)\cdot\log|\mathcal{P}(n,d)|$. It remains to prove Lemma 5.

Proof of Lemma 5. It is known that the Shannon entropy of a distribution is lower bounded by its Renyi entropy of order two. Thus we have

$$H_U(x|f) \geq -\log\left[R_U(x|f)\right], \tag{2.28}$$

where

$$R_U(x|f) := \sum_x (\Pr_U(x|f))^2 = \sum_x \langle x,x|U^{\otimes 2}|\psi_f\rangle\langle\psi_f|^{\otimes 2}(U^\dagger)^{\otimes 2}|x,x\rangle. \tag{2.29}$$

Taking the expected value of Eq. (2.28) and noting that the function $-\log(\cdot)$ is convex, one gets

$$\mathbb{E}_f[H_U(x|f)] \geq -\mathbb{E}_f\left[\log R_U(x|f)\right] \geq -\log \mathbb{E}_f\left[R_U(x|f)\right]. \tag{2.30}$$

Below we prove

Proposition 1. *Let $f \in \mathcal{P}(n,d)$ be a uniformly random degree-d polynomial with $d \geq 2$. Then*

$$\mathbb{E}_f[|\psi_f\rangle\langle\psi_f|^{\otimes 2}] = \frac{1}{4^n}(\mathbb{I} + \mathrm{SWAP}) + \frac{1}{2^n}|\Phi^+\rangle\langle\Phi^+| - \frac{2}{4^n}\sum_x |x,x\rangle\langle x,x|, \tag{2.31}$$

where SWAP *swaps two n-qubit registers and $|\Phi^+\rangle = 2^{-n/2}\sum_x |x,x\rangle$ is the EPR state of $2n$ qubits.*

Combining the proposition and the bound $|\langle\Phi^+|U^{\otimes 2}|x,x\rangle|^2 \leq 2^{-n}$ gives

$$\mathbb{E}_f[R_U(x|f)] \leq \frac{3}{2^n} \leq \frac{1}{2^{n-2}}. \tag{2.32}$$

Substituting this into Eq. (2.30) completes the proof. We now prove the proposition.

Proof of Proposition 1. Let

$$Q = \mathbb{E}_f[|\psi_f\rangle\langle\psi_f|^{\otimes 2}] = \frac{1}{4^n}\sum_{w,x,y,z} E(w,x,y,z)|w,x\rangle\langle y,z|,$$

where

$$E(w,x,y,z) = \mathbb{E}_f[(-1)^{f(w)+f(x)+f(y)+f(z)}].$$

Our proof strategy uses a couple of lemmas from [Bra+19, Proposition 5] and [BMS16, Lemma 11].

Claim 1. *$E(w,x,y,z) = 0$ unless $w+x+y+z = 0^n$ and at least two of the strings w,x,y coincide.*

Proof. We can write

$$f(v) = \sum_{p=1}^n A_p v_p + \sum_{1\leq p<q\leq n} A_{p,q} v_p v_q + \ldots \pmod 2$$

where $A_p \in \{0,1\}$ and $A_{p,q} \in \{0,1\}$ are picked uniformly at random and dots represents higher order terms. Taking the expectation value over A_p gives

$$\mathbb{E}_{A_p}\left[(-1)^{A_p(w_p+x_p+y_p+z_p)}\right] = 0 \quad \text{unless} \quad w_p+x_p+y_p+z_p = 0 \pmod 2.$$

This proves the first part of the claim. Taking the expectation value over $A_{p,q}$ gives

$$\mathbb{E}_{A_{p,q}}\left[(-1)^{A_{p,q}(w_pw_q+x_px_q+y_py_q+z_pz_q)}\right] = 0 \quad \text{unless} \quad w_pw_q + x_px_q + y_py_q + z_pz_q = 0 \pmod 2.$$

Substituting $z_p = x_p+y_p+w_p \pmod 2$ and $z_q = x_q+y_q+w_q \pmod 2$ in the above expression one concludes that $E(w,x,y,z) = 0$ unless $w + x + y + z = 0^n \pmod 2$ and

$$(w_px_q + x_pw_q) + (x_py_q + y_px_q) + (w_py_q + y_pw_q) = 0 \pmod 2. \tag{2.33}$$

If $w = x = y$ then we are done. Otherwise there exists an index $p \in [n]$ such that exactly two of the variables w_p, x_p, y_p coincide. Since Eq. (2.33) is symmetric under permutations of w, x, y we can assume wlog that $x_p = y_p \neq w_p$. Consider two cases:
Case 1: $x_p = y_p = 0$ and $w_p = 1$. Then Eq. (2.33) gives $x_q = y_q$ for all $q \neq p$. Thus $x = y$.
Case 2: $x_p = y_p = 1$ and $w_p = 0$. Then Eq. (2.33) gives $w_q + y_q + x_q + w_q = 0 \pmod 2$ for all $q \neq p$. Thus $x_q = y_q$ for all $q \neq p$, that is, $x = y$. □

Note that $E(w,x,y,z) = 1$ whenever $w+x+y+z = 0^n \pmod 2$ and at least two of the strings w, x, y coincide. For example, if $w = x$ then one must have $y = z$ and thus the sum $f(w) + f(x) + f(y) + f(z)$ is zero modulo two. This leads to

$$Q = \frac{1}{4^n}\sum_{w,x,y,z} E(w,x,y,z)|w,x\rangle\langle y,z| = \frac{1}{4^n}(I + \text{SWAP}) + \frac{1}{2^n}|\Phi^+\rangle\langle\Phi^+| - \frac{2}{4^n}\sum_x |x,x\rangle\langle x,x|.$$

Here the last term is introduced to avoid overcounting. □

This concludes the proof of the proposition and Lemma 5. □

The proof of this lemma concludes the proof of Theorem 5. □

2.6 Learning sparse and low Fourier-degree binary phase states

In this section, we first consider the problem of learning binary phase states (Eq. (2.10)) using separable measurements under the assumption that f is an s-sparse $\mathbb{F}_2$-degree d Boolean function written as

$$f(x) = \sum_{J\subseteq[n]} \alpha_J \prod_{i\in J} x_i \pmod 2 \tag{2.34}$$

where $|\{J : \alpha_J \neq 0\}| = s$, i.e., there are s terms in the $\mathbb{F}_2$ representation of f.

2.6.1 Sparse learning algorithm

Our algorithm for learning sparse binary phase states and analysis of its sample complexity is similar to that in Section 2.5.1. Similar to Algorithm 1 in the t-th round, for an $m \geq 1$ to be fixed later (where m is the sample complexity), we use RPDS along e_t on m copies of $|\psi_f\rangle$ to obtain m samples $\{(y^{(k)}, D_tf(y^{(k)}))\}_{k\in[m]}$ where $y^{(k)} \in \{0,1\}^{n-1}$ is uniformly random. We now describe how to learn D_tf using these m examples.

Let $A_t \in \mathbb{F}_2^{m \times |\mathcal{M}(n-1,d-1)|}$ be the transposed $(d-1)$-evaluation matrix (defined in Eq. (2.8)) such that the kth row of A_t is given by the vector $(y_S^{(k)})_{|S| \leq d-1}$, where $y_S^{(k)} = \prod_{j \in S} y_j^{(k)}$. We can then write a system of linear equations $A_t \beta_t = D_t f(\mathbf{y})$ where $\beta_t \in \mathbb{F}_2^{|\mathcal{M}(n-1,d-1)|}$ is the vector of coefficients corresponding to monomials in $\mathcal{M}(n-1, d-1)$ and $(D_t f(\mathbf{y}))_k = D_t f(y^{(k)})$. Instead of explicitly solving $A_t \beta_t = D_t f(\mathbf{y})$ (which we did in Section 2.5.1), we propose to estimate the unknown coefficients vector β_t by solving the following linear program by drawing connection to compressed sensing [DM09].

$$\hat{\beta}_t \in \operatorname{argmin} \|\beta\|_1 \text{ such that } A_t \beta = D_t f(\mathbf{y}) \pmod 2 \quad \text{over } \beta \in \mathbb{F}_2^{|\mathcal{M}(n-1,d-1)|}. \tag{2.35}$$

The solution vector $\hat{\beta}_t$ produced by the above linear program corresponds to solving the subset of coefficients α_J in the $\mathbb{F}_2$-representation of f corresponding to sets J which contain t. Like in Section 2.5.1, we repeat the above procedure over n rounds to learn $D_1 f, D_2 f, \ldots, D_n f$ and then eventually learn f using Fact 3. We give details of this algorithm in Algorithm 2.

Algorithm 2 Learning sparse binary phase states through separable measurements

Input: Access to $M = O(2^d sdn \log n)$ copies of $|\psi_f\rangle$ where $f \in \mathcal{P}(n, d, s)$

1: **for** qubit $t = 1, \ldots, n$ **do**
2: Set $m = M/n$
3: Perform RPDS to obtain $\{(y^{(k)}, D_t f(y^{(k)})\}_{k \in [m]}$ by measuring m copies of $|\psi_f\rangle$.
4: Solve linear program $\hat{\beta}_t \in \operatorname{argmin} \|\beta\|_1$ s.t. $A_t \beta = D_t f(\mathbf{y})$ for $\beta \in \mathbb{F}_2^{|\mathcal{M}(n-1,d-1)|}$ (Eq. (2.35))
5: **end for**
6: Use Fact 3 to learn f using $D_1 f, \ldots, D_n f$ (let $\tilde{f}$ be the output).

Output: Output $\tilde{f}$

We now argue the correctness of the algorithm.

Theorem 6. *An n-qubit state $|\psi_f\rangle$ with the unknown Boolean function f of given $\mathbb{F}_2$-degree d and $\mathbb{F}_2$-sparsity $s \leq |\mathcal{M}(n-1, d-1)|/2$ can be learned with an algorithm that consumes M copies of $|\psi\rangle$ with probability $1 - 2^{-\Omega(n)}$ provided that $M = O(2^d sdn \log n)$. Moreover the algorithm only uses $\{X, Z\}$ single-qubit measurements.*

Proof. Algorithm 2 learns f by learning $D_1 f, \ldots, D_n f$ and thereby learns f completely. Here we prove that each $D_t f$ can be learned with with $m = O(2^d sd \log n)$ copies of $|\psi_f\rangle$ and an exponentially small probability of error. This results in an overall sample complexity of $O(2^d sdn \log n)$ for learning f and hence $|\psi_f\rangle$. Let us consider round t in Algorithm 2. We generate m samples $\{y^{(k)}, D_t f(y^{(k)})\}$ through RPDS. Using these m samples, we can solve Eq. (2.35) and obtain the solution $\hat{\beta}_t$. An error occurs when $\hat{\beta}_t \neq \beta_t^\star$ where we have denoted the true solution by $\beta_t^\star$.[6] Probability of this error occurring is then given by

$$\Pr[\hat{\beta}_t \neq \beta_t^\star] = \Pr[\exists \beta \in \{0,1\}^N, \beta \neq \beta_t^\star \mid A_t \beta = D_t f(\mathbf{y}) \cap \|\beta\|_1 \leq \|\beta_t^\star\|_1] \tag{2.36}$$

[6]Note that the true coefficients $(\beta_t^\star)_S = \alpha_{S \cup t}$ where α_J is the true coefficient in the $\mathbb{F}_2$-representation of f, corresponding to set J.

where we have denoted $N = |\mathcal{M}(n-1, d-1)|$. Below we prove that the probability of this bad event can be bounded through Schwartz-Zippel (Lemma 1). Define event

$$\mathsf{BAD}(y^{(1)}, \ldots, y^{(m)}) = [\exists \beta \in \{0,1\}^N, \beta \neq \beta_t^\star \mid A_t \beta = D_t f(\mathbf{y}) \cap \|\beta\|_1 \leq \|\beta_t^\star\|_1]. \tag{2.37}$$

Let us consider the kth row of A_t. We note that the corresponding equation can be rewritten as

$$(A_t)_k \cdot \beta = (D_t f(\mathbf{y}))_k \pmod 2 \tag{2.38}$$

$$(A_t)_k \cdot \beta = (A_t)_k \beta_t^\star \pmod 2 \tag{2.39}$$

$$(A_t)_k \cdot (\beta - \beta_t^\star) = 0 \pmod 2 \tag{2.40}$$

As $\beta \neq \beta_t^\star$, this means there exists a non-zero polynomial $g \in \mathcal{P}(n-1, d-1)$ corresponding to the coefficients $\beta - \beta_t^\star \pmod 2$. Applying Lemma 2 and by noting that $y^{(j)} \in \mathbb{F}_2^{(n-1)}$ are independent and uniformly distributed, we have that

$$\Pr[g(y^{(1)}) = g(y^{(2)}) = \cdots = g(y^{(m)}) = 0] \leq (1 - 2^{-d})^m \leq e^{-m2^{-d}} \tag{2.41}$$

Let $\mathcal{P}_{\mathrm{nnz}}(n, d, s)$ be the set of all degree-d polynomials $g : \mathbb{F}_2^n \to \mathbb{F}_2$ with sparsity s which are not identically zero. By union bound and Eq. (2.41), we have that

$$\Pr[\mathsf{BAD}(y^1, \ldots, y^m)] \leq |\mathcal{P}_{\mathrm{nnz}}(n-1, d-1, s)| \cdot (1 - 2^{-d})^m \tag{2.42}$$

$$= \sum_{\substack{\{\beta \in \{0,1\}^N, \beta \neq \beta_t^\star : \\ \|\beta\|_1 \leq \|\beta_t^\star\|_1\}}} (1 - 2^{-d})^m \tag{2.43}$$

$$\leq \sum_{\ell=1}^{\|\beta_t^\star\|_1} |\{\beta \in \{0,1\}^N \,:\, \|\beta\|_1 = \ell\}| 2^{-m2^{-d}(\ln 2)} \tag{2.44}$$

$$= \sum_{\ell=1}^{s} \binom{N}{\ell} 2^{-m2^{-d}(\ln 2)} \leq 2^{2s \log(N/s) - m2^{-d}(\ln 2)}, \tag{2.45}$$

where the final inequality used Fact 2. We can thus learn all the coefficients $\beta_t^\star$ with an exponentially small probability of error by choosing $m = O(2^d s \log N)$. We need to repeat this over all the n qubits, giving an overall sample complexity of $O(2^d sdn \log n)$ (by noting that $N = \mathcal{M}(n-1, d-1) = O(n^d)$) of learning sparse binary phase states using only separable measurements. Using Fact 3, we can completely learn f. Observe that Algorithm 2 uses only single qubit $\{X, Z\}$ measurements. □

2.6.2 Learning low Fourier-degree phase states

We conclude this section with a theorem about learning low Fourier-degree phase states.

Theorem 7. *Consider binary phase states $|\psi\rangle = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{f(x)} |x\rangle$ where $f : \{0,1\}^n \to \{0,1\}$ has Fourier-degree d. Then $O\widetilde{O}(2^{2d})$ copies of $|\psi\rangle$ are sufficient to identify f with probability $\geq 2/3$.*

Proof. The proof is a simple application of couple of results. Observe that the number of degree-d Boolean functions is 2^{2d2^d}. To see this, observe that [ODo14] for a degree-d Boolean function, all the Fourier coefficients are integer multiple of 2^{-d}, and since by Parseval's theorem $\sum_S f\widehat{f}(S)^2 \leq 1$, the number of non-zero Fourier coefficients is $\leq 2^{2d}$. Hence the number of degree-d Boolean functions is $(2^d)^{2^{2d}} = 2^{2d2^d}$. Additionally, for degree-d functions f, g we have that

$$\langle \psi_f | \psi_g \rangle = \mathop{\mathbb{E}}_x[(-1)^{f(x)+g(x)}] = 1 - 2\Pr_x[f(x) \neq g(x)] \leq 1 - 2^{-d+1}$$

where the final inequality uses the Schwartz-Zippel lemma (Lemma 2) and the fact that $f-g$ has degree at most d. Now, putting this together with Theorem 1 in [Mon19], we get that the number of copies of $|\psi_f\rangle$ sufficient to learn f is given by $O((\log 2^{2d2^d})/2^{-d}) = O\widetilde{O}(2^{2d})$. □

2.7 Learning generalized phase states

In this section, we consider the problem of learning generalized phase states $|\psi_f\rangle$ as given by Eq. (2.11), assuming that f is a degree-d $\mathbb{Z}_q$-valued polynomial, $f \in \mathcal{P}_q(n,d)$. Note that since our goal is to learn $|\psi_f\rangle$ up to an overall phase, we shall identify polynomials which differ only by a constant shift.

Definition 2. *Polynomials $f, g \in \mathcal{P}_q(n,d)$ are equivalent if $f(x)-g(x)$ is a constant function.*

To simplify notation, here and below we omit modulo operations keeping in mind that degree-d polynomials take values in the ring $\mathbb{Z}_q$. Thus all equal or not-equal constraints that involve a polynomial's value are modulo q.

2.7.1 Learning using separable measurements

Let $q \geq 2$ and $d \geq 1$ be integers. For technical reasons, we shall assume that q is even. Let $\omega_q = e^{2\pi i/q}$. Our main result is as follows.

Theorem 8. *Let $d \leq n/2$. There exists an algorithm that uses $M = O(2^d q^3 n^d \log q) = O(n^d)$ copies of a generalized phase state $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \omega_q^{f(x)} |x\rangle$ with an unknown polynomial $f \in \mathcal{P}_q(n,d)$ and outputs a polynomial $g \in \mathcal{P}_q(n,d)$ such that g is equivalent to f with the probability at least $1 - 2^{-\Omega(n)}$. The quantum part of the algorithm requires only single-qubit unitary gates and measurements in the standard basis.*

Moreover, suppose there exists an algorithm that with probability $\geq 1/10$, learns an n-variate polynomial $f \in \mathcal{P}_q(n,d)$, given k copies of the phase state $|\psi_f\rangle$, measuring each copy in an arbitrary orthonormal basis, and performing an arbitrary classical processing. Then

$$M = \Omega(n^d). \tag{2.46}$$

Before stating our learning algorithm and sample complexity, we need the following lemmas.

Lemma 6. *Choose any $f \in \mathcal{P}_q(n,d)$ and $c \in \mathbb{Z}_q$. Then either $f(x)$ is a constant function or the fraction of inputs $x \in \{0,1\}^n$ such that $f(x) \neq c$ is at least $1/2^d$.*

Proof. We shall use the following simple fact.

Proposition 2. *Consider a function $f : \{0,1\}^n \to \mathbb{Z}_q$ specified as a polynomial*

$$f(x) = \sum_{J \subseteq [n]} \alpha_J \prod_{j \in J} x_j \pmod{q}. \tag{2.47}$$

Here $\alpha_J \in \mathbb{Z}_q$ are coefficients. The function f is constant if and only if $\alpha_J = 0 \pmod{q}$ for all non-empty subsets $J \subseteq [n]$.

Proof. If $\alpha_J = 0 \pmod{q}$ for all non-empty subsets J then $f(x) = f(0^n) \pmod{q}$ for all x, that is, f is constant. Conversely, suppose f is constant. Choose a subset $J \subseteq [n]$. We can consider J as an n-bit string with the Hamming weight $|J|$ such that $J_i = 1$ if $i \in J$ and $J_i = 0$ otherwise. If $|J| = 1$ then $f(J) = f(0^n) + \alpha_J \pmod{q}$ and thus $\alpha_J = 0 \pmod{q}$ for all subsets J with $|J| = 1$. Suppose we have already proved that $\alpha_J = 0 \pmod{q}$ for any subset with $1 \leq |J| \leq w$. If $|J| = w + 1$ then $f(J) = f(0^n) + \alpha_J \pmod{q}$ and thus $\alpha_J = 0 \pmod{q}$ for all subsets J with $|J| = w + 1$. Proceeding inductively proves the claim. □

We shall prove Lemma 6 by induction in n. The base case of induction is $n = d$. Clearly, a non-constant function $f : \{0,1\}^d \to \mathbb{Z}_q$ takes a value different from c at least one time, that is, the fraction of inputs $x \in \{0,1\}^d$ such that $f(x) \neq c$ is at least $1/2^d$.

Suppose $n > d$ and $f \in \mathcal{P}_q(n,d)$ is not a constant function. Let d' be the maximum degree of non-zero monomials in f. Clearly $1 \leq d' \leq d$. Suppose f contains a monomial $\alpha_S \prod_{j \in S} x_j$ where $\alpha_S \in \mathbb{Z}_q \setminus \{0\}$ and $|S| = d'$. Since $|S| < n$, one can choose a variable x_i with $i \in [n] \setminus S$. Let $g_a : \{0,1\}^{n-1} \to \mathbb{Z}_q$ be a function obtained from f by setting the variable x_i to a constant value $a \in \{0,1\}$. Clearly, $g_a \in \mathcal{P}_q(n-1,d)$. The coefficients of the monomial $\prod_{j \in S} x_j$ in g_0 and g_1 are α_S and $\alpha_S + \alpha_{S \cup \{i\}} \pmod{q}$ respectively. However, $\alpha_{S \cup \{i\}} = 0 \pmod{q}$ since otherwise f would contain a monomial $x_i \prod_{j \in S} x_j$ of degree larger than d'. We conclude that both g_0 and g_1 contain a non-zero monomial $\alpha_S \prod_{j \in S} x_j$. By Proposition 2, g_0 and g_1 are not constant functions. Since g_0 and g_1 are degree-d polynomials in $n-1$ variables, the induction hypothesis gives

$$\Pr_y[g_a(y) \neq c] \geq \frac{1}{2^d}. \tag{2.48}$$

Here $y \in \{0,1\}^{n-1}$ is picked uniformly at random. Thus

$$\Pr_x[f(x) \neq c] = \frac{1}{2}\left[\Pr_y[g_0(y) \neq c] + \Pr_y[g_1(y) \neq c]\right] \geq \frac{1}{2^d}. \tag{2.49}$$

Here $x \in \{0,1\}^n$ is picked uniformly at random. This proves the induction step. □

With this lemma, we are now ready to prove Theorem 8. In the section below we first describe our learning algorithm and in the next section we prove the theorem by proving the sample complexity upper bound.

Learning Algorithm in Theorem 8

We are now ready to state our learning algorithm. As in Section 2.5.1 for learning binary phase states with separable measurements, we learn generalized phase states through examples containing information about the derivatives of $f(x)$. The crucial difference between the binary phase state learning algorithm and the generalized setting is, in the binary case, we

obtained a measurement outcome b_y that corresponded to $b_y = f(0y) - f(1y)$, however in the generalized scenario, we obtain a measurement outcome b'_y that satisfies $f(0y) - f(1y) \neq b'_y$. Nevertheless, we are able to still learn f using such measurement outcomes which we describe in the rest of the section.

We now describe the learning algorithm. We carry out the algorithm in n rounds, which we index by t. For simplicity, we describe the procedure for the first round. Suppose we measure qubits $2, 3, \ldots, n$ of the state $|\psi_f\rangle$ in the Z-basis. Let $y \in \{0,1\}^{n-1}$ be the measured bit string. Note that the probability distribution of y is uniform. The post-measurement state of qubit 1 is

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}}(\omega_q^{f(0y)}|0\rangle + \omega_q^{f(1y)}|1\rangle) \tag{2.50}$$

For each $b \in \mathbb{Z}_q$ define a single-qubit state

$$|\phi_b\rangle = \frac{1}{\sqrt{2}}(|0\rangle - \omega_q^b|1\rangle) \tag{2.51}$$

Using the identity $\sum_{b\in\mathbb{Z}_q} \omega_q^b = 0$ one gets

$$I = \frac{2}{q}\sum_{b\in\mathbb{Z}_q} |\phi_b\rangle\langle\phi_b| \tag{2.52}$$

One can view Eq. (2.52) as a single-qubit POVM with q elements $(2/q)|\phi_b\rangle\langle\phi_b|$. Let $\mathcal{M}$ be the single-qubit measurement described by this POVM. Applying $\mathcal{M}$ to the state $|\psi_{f,y}\rangle$ returns an outcome $b \in \mathbb{Z}_q$ with the probability

$$\Pr(b|y) := \frac{2}{q}|\langle\phi_b|\psi_{f,y}\rangle|^2 = \frac{1}{2q}\left|1 - \omega_q^{f(1y)-f(0y)-b}\right|^2. \tag{2.53}$$

Clearly, $\Pr(b|y)$ is a normalized probability distribution, $\sum_{b\in\mathbb{Z}_q} \Pr(b|y) = 1$. Furthermore,

$$f(1y) - f(0y) = b \quad \text{implies} \quad \Pr(b|y) = 0, \tag{2.54}$$

$$f(1y) - f(0y) \neq b \quad \text{implies} \quad \Pr(b|y) \geq \frac{2}{q}\sin^2(\pi/q) = \Omega(1/q^3). \tag{2.55}$$

To conclude, the combined n-qubit measurement consumes one copy of the state $|\psi_f\rangle$ and returns a pair $(y, b) \in \{0,1\}^{n-1} \times \mathbb{Z}_q$ such that

$$f(1y) - f(0y) \neq b \tag{2.56}$$

with certainty and all outcomes b satisfying Eq. (2.56) appear with a non-negligible probability. Define a function $g : \{0,1\}^{n-1} \to \mathbb{Z}_q$ such that

$$g(y) = f(1y) - f(0y). \tag{2.57}$$

We claim that g is a degree-$(d-1)$ polynomial, that is, $g \in \mathcal{P}_q(n-1, d-1)$. Indeed, it is clear that $g(y)$ is a degree-d polynomial. Moreover, all degree-d monomials in $f(x)$ that do not contain the variable x_1 appear in $f(1y)$ and $f(0y)$ with the same coefficient. Such

monomials do not contribute to $g(y)$. A degree-d monomial in $f(x)$ that contains the variable x_1 contributes a degree-$(d-1)$ monomial to $g(y)$. Thus $g \in \mathcal{P}_q(n-1, d-1)$, as claimed.

From Eq. (2.56) one infers a constraint

$$g(y) \neq b \tag{2.58}$$

whenever the combined n-qubit measurement of $|\psi_f\rangle$ returns an outcome (y, b). Suppose we repeat the above process m times obtaining constraints

$$g(y^{(k)}) \neq b^{(k)}, \qquad k = 1, 2, \ldots, m. \tag{2.59}$$

This consumes m copies of $|\psi_f\rangle$. We claim that the probability of having more than one polynomial $g \in \mathcal{P}_q(n-1, d-1)$ satisfying the constraints Eq. (2.59) is exponentially small if we choose

$$m = O(q^3 \log(q) 2^d n^{d-1}). \tag{2.60}$$

Sample Complexity bound in Theorem 8

Define a probability distribution $\pi(\vec{y}, \vec{b})$ where

$$\vec{z} = (y^{(1)}, \ldots, y^{(m)}) \in \{0,1\}^{(n-1)m} \quad \text{and} \qquad \vec{b} = (b^{(1)}, \ldots, b^{(m)}) \in (\mathbb{Z}_q)^{\times m} \tag{2.61}$$

such that $y^{(j)}$ are picked uniformly at random and $b^{(k)}$ are sampled from the distribution $\Pr(b^{(k)}|y^{(k)})$ defined in Eq. (2.53). For each polynomial $h \in \mathcal{P}_q(n-1, d-1)$ define an event

$$\mathsf{BAD}(h) = \{(\vec{y}, \vec{b}) \ : \ h(y^{(k)}) \neq b^{(k)} \quad \text{for all} \quad k \in [m]\}. \tag{2.62}$$

We claim that

$$\Pr[\mathsf{BAD}(h)] := \sum_{(\vec{y},\vec{b}) \in \mathsf{BAD}(h)} \pi(\vec{y}, \vec{b}) \leq \left[1 - \Omega(2^{-d} q^{-3})\right]^m \tag{2.63}$$

for any $h \neq g$. Indeed, consider some fixed $k \in [m]$. The event $b^{(k)} \neq h(y^{(k)})$ occurs automatically if $h(y^{(k)}) = g(y^{(k)})$. Otherwise, if $h(y^{(k)}) \neq g(y^{(k)})$, the event $b^{(k)} \neq h(y^{(k)})$ occurs with the probability at most $1 - \Omega(1/q^3)$ since $b^{(k)} = h(y^{(k)})$ with the probability at least $\Omega(1/q^3)$ due to Eq. (2.55). It follows that

$$\Pr_{y^{(k)}, b^{(k)}}[h(y^{(k)}) \neq b^{(k)}] \leq \Pr_{y^{(k)}}[h(y^{(k)}) = g(y^{(k)})] + \Pr_{y^{(k)}}[h(y^{(k)}) \neq g(y^{(k)})] \left(1 - \Omega(1/q^3)\right) \tag{2.64}$$

$$= 1 - \Pr_{y^{(k)}}[h(y^{(k)}) \neq g(y^{(k)})] \cdot \Omega(1/q^3). \tag{2.65}$$

If h and g are equivalent then $h(y) = g(y) + c$ for some constant $c \in \mathbb{Z}_q$. Note that $c \neq 0$ since we assumed $h \neq g$. In this case

$$\Pr_{y^{(k)}}[h(y^{(k)}) \neq g(y^{(k)})] = 1. \tag{2.66}$$

If h and g are non-equivalent, apply Lemma 6 to a non-constant degree-$(d-1)$ polynomial $h - g$. It gives

$$\Pr_{y^{(k)}}[h(y^{(k)}) \neq g(y^{(k)})] \geq \frac{1}{2^{d-1}}. \tag{2.67}$$

In both cases we get

$$\Pr_{y^{(k)},b^{(k)}}[h(y^{(k)}) \neq b^{(k)}] \leq 1 - \Omega(2^{-d}q^{-3}), \tag{2.68}$$

which proves Eq. (2.63) since the pairs $(y^{(k)}, b^{(k)})$ are i.i.d. random variables.

As noted earlier in the preliminaries, observe that $|\mathcal{P}_q(n-1,d-1)| \leq q^{O(n^{d-1})} = 2^{O(\log(q)n^{d-1})}$. By the union bound, one can choose $m = O(2^d q^3 \log(q) n^{d-1})$ such that

$$\Pr\left[\bigcup_{h\in\mathcal{P}_q(n-1,d-1)\setminus g} \mathsf{BAD}(h)\right] \leq 2^{O(\log(q)n^{d-1})}\left[1-\Omega(2^{-d}q^{-3})\right]^m \leq 2^{-\Omega(n)}. \tag{2.69}$$

In other words, the probability that g is the unique element of $\mathcal{P}_q(n-1,d-1)$ satisfying all the constraints Eq. (2.59) is at least $1 - 2^{-\Omega(n)}$. One can identify such polynomial g by checking the constraints Eq. (2.59) for every $g \in \mathcal{P}_q(n-1,d-1)$. If the constraints are satisfied for more than one polynomial, declare a failure.

At this point we have learned a polynomial $g \in \mathcal{P}_q(n-1,d-1)$ such that $f(1y) - f(0y) = g(y)$ for all $y \in \{0,1\}^{n-1}$. For simplicity, we ignore the exponentially small failure probability. Applying the same protocol n times to copies of the quantum state $|\psi_f\rangle$ by a cyclic shift of qubits, one can learn polynomials $g_0, g_1, \ldots, g_{n-1} \in \mathcal{P}_q(n-1,d-1)$ such that

$$f(C^i(1y)) - f(C^i(0y)) = g_i(y) \quad \text{for all} \quad i \in [n] \quad \text{and} \quad y \in \{0,1\}^{n-1}, \tag{2.70}$$

where C is the cyclic shift of n bits. This consumes $M = O(nm) = O(2^d q^3 \log(q) n^d)$ copies of the state $|\psi_f\rangle$. We can assume wlog that $f(0^n) = 0$ since our goal is to learn $f(x)$ modulo a constant shift. Suppose we have already learned values of $f(x)$ for all bit strings x with the Hamming weight $|x| \leq w$ (initially $w = 0$). Any bit string x with $|x| = w+1$ can be represented as $x = C^i(1y)$ for some $y \in \{0,1\}^{n-1}$ such that $|y| = w$. Now Eq. (2.70) determines $f(x)$ since $|C^i(0y)| = |y| = w$ so that $f(C^i(0y))$ is already known and the polynomial $g_i(y)$ has been learned. Proceeding inductively one can learn $f(x)$ for all x.

It remains to note that the POVM Eq. (2.52) is a probabilistic mixture of projective single-qubit measurements whenever q is even. Indeed, in this case the states $|\phi_b\rangle$ and $|\phi_{b+q/2}\rangle = Z|\phi_b\rangle$ form an orthonormal basis of a qubit, see Eq. (2.51). Thus the POVM defined in Eq. (2.52) can be implemented by picking a random uniform $b \in \mathbb{Z}_q$ and measuring a qubit in the basis $\{|\phi_b\rangle, Z|\phi_b\rangle\}$. Thus the learning protocol only requires single-qubit unitary gates and measurements in the standard basis.

The lower bound in the proof of Theorem 8 follows in a straightforward manner from the lower bound for binary phase states. Indeed, suppose

$$f'(x) = \sum_{J\in[n]} \alpha_J \prod_{j\in J} x_j \pmod 2$$

is an $\mathbb{F}_2$-valued degree-d polynomial, $f' \in \mathcal{P}(n,d)$. Suppose $q = 2r$ for some integer r. Define a polynomial

$$f(x) = r f'(x) \pmod q.$$

Clearly $f \in \mathcal{P}_q(n,d)$ and $\omega_q^{f(x)} = (-1)^{f'(x)}$ for all x, that is the binary phase state corresponding to f' coincides with the generalized phase state corresponding to f. Using Theorem 5, we obtain a lower bound of $M = \log|\mathcal{P}(n,d)| = \Omega(n^d)$ for learning ψ_f. This concludes the proof of Theorem 8.

2.7.2 Learning stabilizer states

We now describe how the algorithm stated in Theorem 8 could be used to learn any n-qubit stabilizer state (produced by a Clifford circuit applied to $|0^n\rangle$ state) using separable measurements. Note that we can learn a subclass of stabilizer states called graph states (which are simply binary phase states with $d = 2$) using Algorithm 1 with the sample complexity of $O(n^2)$ (as shown in Theorem 3).

From a result in [DD03], we know that a stabilizer state can be represented as follows

$$|\psi\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{\ell(x)} (-1)^{q(x)} |x\rangle, \tag{2.71}$$

where A is an affine subspace of $\mathbb{F}_2^n$, $\ell : \mathbb{F}_2^n \to \mathbb{F}_2$ is a linear function and $q : \mathbb{F}_2^n \to \mathbb{F}_2$ is quadratic function. Clearly, an alternate form is a generalized phase state with degree-2

$$|\psi_f\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{f(x)} |x\rangle \tag{2.72}$$

where the summation is over A instead of the entire $\mathbb{F}_2^n$, and the function $f : \mathbb{F}_2^n \to \mathbb{Z}_4$ has its coefficients corresponding to the quadratic monomials take values in $\{0, 2\}$. We can now learn this using separable measurements as stated in the following statement as opposed to entangled measurements as required by Bell sampling [Mon17a].

Corollary 3. *There exists an algorithm that uses $M = O(n^2)$ copies of a stabilizer state $|\psi_f\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{f(x)} |x\rangle$ with an unknown polynomial $f \in \mathcal{P}_4(n, 2)$ and outputs a polynomial $g \in \mathcal{P}_4(n, 2)$ such that g is equivalent to f with the probability at least $1 - 2^{-\Omega(n)}$. The quantum part of the algorithm requires only single-qubit unitary gates and measurements in the standard basis.*

Proof. The subspace A of an unknown stabilizer state can be denoted as $a + S_A$ where $a \in \mathbb{F}_2^n$ is a translation vector and S_A is a linear subspace of $\mathbb{F}_2^n$. To learn a and a basis of the subspace S_A, it is enough to measure $O(n \log n)$ copies of $|\psi_f\rangle$ in the computational basis. This in turn defines a subset of the n directions $\{e_i\}$ along which we need to search for non-zero monomials in the partial derivatives of f. We can now use the learning algorithm in Theorem 8 to learn the unknown stabilizer state using $O(n^2)$ copies with the desired probability. □

2.8 Learning phase states under depolarization

In this section, we consider learning algorithms in the presence of noise (in particular we consider global depolarizing noise, local depolarizing noise and local depolarizing noise when the phase state has additional graph structure).

2.8.1 Global depolarizing noise

Let

$$|\psi_f\rangle = \frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} |x\rangle$$

where $f : \{0,1\}^n \to \{0,1\}$ is a degree-2 polynomial in $\mathbb{F}_2$. For simplicity, we assume $f(x) = x^\top Ax$ (where $A \in \mathbb{F}_2^{n\times n}$ is upper triangular). Suppose we are given noisy copies of $|\psi\rangle$ of the form

$$\psi_f = (1-\varepsilon)\cdot|\psi_f\rangle\langle\psi_f| + \varepsilon\cdot\mathbb{I}/2^n$$

for some $\varepsilon > 0$, then how many copies of ψ_f are necessary and sufficient to learn f?

Below, we observe the following theorem:

Theorem 9. *Let $\varepsilon > 0$ be a constant. Given $2n^{1+\delta}$ copies of ψ_f (with error ε) and $n \cdot 2^{O(n/\log\log n)}$ time (for some constant $\delta \in (0,1)$ dependent on ε), there exists a procedure to learn A.*

Our argument crucially uses the result of Lyubashevsky.

Theorem 10 ([Lyu05])**.** *We are given $n^{1+\delta}$ ordered pairs (a_i, ℓ_i) where a_i are chosen uniformly and independently at random from the set $\{0,1\}^n$ and for some $c \in \{0,1\}^n$,*

$$\ell_i = \begin{cases} c\cdot a_i \pmod 2 & \text{w.p. } 1/2+\eta \\ 1 + c\cdot a_i \pmod 2 & \text{w.p. } 1/2-\eta \end{cases}$$

If $\eta > 2^{-(\log n)^\delta}$ for constant $\delta < 1$, then there is an algorithm that can recover c in time $2^{O(n/\log\log n)}$ with high probability.

We also use the following simple lemma. The procedure above is an application of Bell sampling [Mon17a] to the pure state $|\psi_f\rangle^{\otimes 2}$ and mixed state $\psi_f^{\otimes 2}$. We now prove our main theorem statement.

Proof of Theorem 9. For simplicity let $B = A + A^\top$. One way to view Lemma 4 is that, it uses two copies of ψ_n and produces a $(z, w_z) \in \{0,1\}^{2n}$ such that $(w_z)_i = B^i \cdot z$ (where B^i is the i row of B) with probability $(1-\varepsilon)^2$ and is a uniformly random bit $b \in \{0,1\}$ with probability $1-(1-\varepsilon)^2$. In particular,

$$(w_z)_i = \begin{cases} B^i\cdot z & \text{w.p. } 1/2 + (1-\varepsilon)^2/2 \\ 1 + B^i\cdot z & \text{w.p. } 1/2 - (1-\varepsilon)^2/2. \end{cases}$$

Hence two copies of ψ_f can be used to obtain $(z,(w_z)_1),\ldots,(z,(w_z)_n)$. So the learning algorithm first uses $T = n^{1+\delta}$ many copies of ψ_f and produces

$$\begin{gathered} (z^1,(w_{z^1})_1),\ldots,(z^1,(w_{z^1})_n) \\ (z^2,(w_{z^2})_1),\ldots,(z^2,(w_{z^2})_n) \\ \vdots \\ (z^T,(w_{z^T})_1),\ldots,(z^T,(w_{z^T})_n). \end{gathered}$$

Each column above (i.e., $(z^1,(w_{z^1})_1),\ldots,(z^T,(w_{z^T})_1)$) can be now be given as input to the algorithm of Theorem 10 where $\eta = (1-\varepsilon)^2/2$ is a constant (and $\delta > 0$ is also a tiny constant), which produces B^i with high probability.[7] Hence feeding all the n different columns to Theorem 10 allows the algorithm to learn $B^1,\ldots,B^n$ explicitly. The overall

[7] The high probability in Theorem 10 is in fact inverse exponential in n.

sample complexity is $2n^{1+\delta}$ and time complexity is $n \cdot 2^{O(n/\log\log n)}$. Once we learn the off-diagonal elements of A (since above we only obtain information of $A + A^\top$ which zeroes the diagonal entries of A), a learning algorithm, applies the operation $|x\rangle \to (-1)^{x_{ij}}|x\rangle$ if $A_{ij} = 1$ for $i \neq j$. Repeating this for all the $n(n-1)/2$ different $i \neq j$, the resulting quantum state is $\sum_x (-1)^{\sum_i x_i A_{ii}}|x\rangle$ which is a linear phase state, and we can learn using Bernstein-Vazirani algorithm. □

2.8.2 Local depolarizing noise

Let us now show that learning phase states subject to a local depolarizing noise has an exponential sampling complexity in the worst case.

Theorem 11. *For every $\varepsilon > 0$, learning degree-2 phase states with ε-local depolarizing noise has sample complexity $\Omega((1-\varepsilon)^n)$.*

Proof. Let $\mathcal{D}_1$ be a single-qubit depolarizing channel that implements the identity with probability $1-\varepsilon$ and outputs a maximally mixed state with probability ε,

$$\mathcal{D}_1(\rho) = (1-\varepsilon)\rho + \varepsilon \mathrm{Tr}(\rho)\frac{I}{2}.$$

Let $\mathcal{D} = \mathcal{D}_1^{\otimes n}$ be the n-qubit depolarizing channel. Consider n-qubit GHZ-like states

$$|\phi^\pm\rangle = (|0^n\rangle \pm |1^n\rangle)/\sqrt{2}.$$

Using the identity $\mathcal{D}_1(|0\rangle\langle 1|) = (1-\varepsilon)|0\rangle\langle 1|$ one gets

$$\mathcal{D}(|\phi^+\rangle\langle\phi^+|) - \mathcal{D}(|\phi^-\rangle\langle\phi^-|) = (1-\varepsilon)^n(|\phi^+\rangle\langle\phi^+| - |\phi^-\rangle\langle\phi^-|)$$

which implies

$$\| \mathcal{D}(|\phi^+\rangle\langle\phi^+|) - \mathcal{D}(|\phi^-\rangle\langle\phi^-|) \|_1 \leq 2(1-\varepsilon)^n. \tag{2.73}$$

It follows that the trace distance between k copies of the states $\mathcal{D}(|\phi^+\rangle\langle\phi^+|)$ and $\mathcal{D}(|\phi^-\rangle\langle\phi^-|)$ is at most $2k(1-\varepsilon)^n$. By Helstrom theorem, these states cannot be distinguished reliably unless $k = \Omega((1-\varepsilon)^{-n})$. Next we observe that $|\phi^\pm\rangle$ are degree-two phase states modulo single-qubit rotations $R_x = e^{i(\pi/4)X}$. Indeed, suppose $n = 1 \pmod 4$. Then a simple algebra shows that

$$R_x^{\otimes n}|\phi^+\rangle = e^{i\pi/4}|\psi_f\rangle \quad \text{and} \quad R_x^{\otimes n}|\phi^-\rangle = e^{-i\pi/4}|\psi_g\rangle, \tag{2.74}$$

where ψ_f and ψ_g are n-qubit phase states associated with degree-two polynomials

$$f(x) = \sum_{1\leq i<j\leq n} x_i x_j \pmod 2 \quad \text{and} \quad g(x) = f(x) + \sum_{i=1}^n x_i \pmod 2.$$

Since the depolarazing channel $\mathcal{D}$ commutes with single-qubit unitary operators, Eqs. (2.73,2.74) give

$$\| \mathcal{D}(|\psi_f\rangle\langle\psi_f|) - \mathcal{D}(|\psi_g\rangle\langle\psi_g|) \|_1 \leq 2(1-\varepsilon)^n.$$

Thus k copies of the noisy phase states $\mathcal{D}(|\psi_f\rangle\langle\psi_f|)$ and $\mathcal{D}(|\psi_g\rangle\langle\psi_g|)$ cannot be distinguished reliably unless $k = \Omega((1-\varepsilon)^{-n})$. We conclude that the sampling complexity of learning phase states subject to local ε-depolarizing noise is at least $\Omega((1-\varepsilon)^{-n})$, which is exponentially large in n for any constant error rate $\varepsilon > 0$. □

2.8.3 Local depolarizing noise and small graph degree

Although learning phase states with local depolarizing is hard in general, a restricted class of states is again easy to learn using the same technique from Section 2.8.1. Recall from that section the notation $f(x) = x^\top Ax$ and $B = A + A^\top$, where A is upper triangular and B is symmetric. Interpreting B as the adjacency matrix of a graph, we define the graph-degree of f to be $\mathrm{gd}(f) = \max_i |B^i|$, where B^i is the i^{th} row of B. The graph-degree of f is also one less than the maximum stabilizer weight of the stabilizer state $|\psi_f\rangle$.

It is possible to learn phase states suffering from local depolarizing using only a few copies if their graph degree is promised to be small.

Theorem 12. *Let $\varepsilon > 0$ be a constant, $\mathcal{D}$ be local depolarizing noise on n qubits with strength ε, and $f(x)$ be a degree-2 polynomial with $gd(f) < (\log n)^{\delta'}$ for some constant δ'. Given $2n^{1+\delta}$ copies of $\mathcal{D}(\psi_f)$ and $n2^{O(n/\log\log n)}$ time (for some constant $\delta \in (0,1)$ dependent on ε and δ'), there exists a procedure to learn A.*

Proof. Suppose we apply Bell-sampling, Lemma 4, to phase states suffering from local depolarizing noise $\mathcal{D}$. Recall that this involves measuring the two-body operators $Z \otimes Z$ and $X \otimes X$ on corresponding pairs of qubits from two-copies of the state. Since these are two-qubit operators, they are randomized by the noise with probability $1 - (1-\varepsilon)^2$.

One use of Bell-sampling on two copies of the state gives $(z, w_z) \in \{0,1\}^{2n}$, where each of the $2n$ bits is correct (i.e. is the same as we would get without noise) with probability $(1-\varepsilon)^2$ and uniformly random with probability $1 - (1-\varepsilon)^2$. Therefore, $(w_z)_i + B^i \cdot z$, which is a sum of $|B^i|+1$ bits, is 0 (mod 2) with probability $(1-\varepsilon)^{2(|B^i|+1)}$ and uniformly random otherwise. Just as in the proof of Theorem 9, we can apply Theorem 10, now with $\eta = (1-\varepsilon)^{2(\mathrm{gd}(f)+1)} > (1-\varepsilon)^{2((\log n)^{\delta'}+1)}$, to learn B. Once we learn B, we can learn the diagonal elements of A, using the same procedure as in the proof of Theorem 9. □

2.9 Applications

In this section, we describe how the algorithms for learning phase states (see Table 2.1) can be used to learn the quantum circuits that produce binary phase states in Section 2.9.1 and generalized phase states in Section 2.9.2. For each, we firstly describe the quantum circuits that produce the phase states of interest followed by our results for learning these quantum circuits.

2.9.1 Learning quantum circuits producing binary phase states

We consider a n-qubit circuit C produced from the set of gates $S = \{\mathsf{H}, \mathsf{Z}, \mathsf{CZ}, \mathsf{CCZ}, \ldots, \mathsf{C^{d-1}Z}\}$ where H is the Hadamard gate and $\mathsf{C^{d-1}Z}$ denotes the controlled-Z gate with $(d-1)$ control qubits. We will actually restrict ourselves to circuits C which start and end with a column of Hadamard gates over all n-qubits, with its internal part C' containing gates from $S \setminus \mathsf{H}$. We then have the following statement regarding the states produced by C

Proposition 3. *Let C be an n-qubit quantum circuit, starting and ending with a column of*

Hadamard gates, with its internal part C' *only containing* s *gates from* $\{\mathsf{Z}, \mathsf{CZ}, \mathsf{CCZ}, \ldots, \mathsf{C^{d-1}Z}\}$*, then*

$$|\psi_f\rangle = \mathsf{C}'|+\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle, \tag{2.75}$$

where the corresponding Boolean function $f \in \mathcal{P}(n, d, s)$*.*

Proof. We follow a proof strategy similar to that in [Mon07, Prop. 1] and [BMS16, Appendix B], which treated the case of $d = 3$. Let Z_{i_1} be the Z gate acting on the i_1th qubit, CZ_{i_1,i_2} be the controlled-Z gate with i_1th qubit as the control, and similarly $\mathsf{C^{d-1}Z}_{i_1,i_2,\ldots,i_d}$ with controls on $(i_1, \ldots, i_{d-1})$ qubits. We note that for any $x \in \mathbb{F}_2^n$,

$$\langle x|\mathsf{Z}_{i_1}|x\rangle = (-1)^{x_{i_1}},\ \langle x|\mathsf{CZ}_{i_1,i_2}|x\rangle = (-1)^{x_{i_1}x_{i_2}},\ \langle x|\mathsf{C^{d-1}Z}_{i_1,i_2,\ldots,i_d}|x\rangle = (-1)^{x_{i_1}x_{i_2}\cdots x_{i_d}}. \tag{2.76}$$

As all these gates are diagonal, we can obtain an expression for $\langle x|\mathsf{C}'|x\rangle$ by simply multiplying the expressions of $\langle x|G|x\rangle$ for the different gates G in C'. To complete the proof, we note that the $\mathbb{F}_2$-degree of f is k if and only if $\mathsf{C^{k-1}Z}$ is the gate with highest controls present in C' and the number of terms in f is at most the number of gates applied in C'. □

Note that the states produced in Proposition 3 are exactly the binary phase states corresponding to Boolean functions f given by Eq. (2.6). Some special classes of circuits included in the above statement are Clifford circuits which produce graph states for $d = 2$, and IQP circuits for $d = 3$. We observe from the above proposition that there can be more than one quantum circuit C corresponding to a given polynomial $f \in \mathcal{P}(n, d, s)$. As the internal gates of C' in S commute, these gates can be reordered arbitrarily while still producing the same Boolean function f.

To learn a circuit representation of C from samples, we have the following result.

Theorem 13. *Let* C *be an unknown* n*-qubit quantum circuit, starting and ending with a column of Hadamard gates, with its internal part* C' *only containing gates from* $\{\mathsf{Z}, \mathsf{CZ}, \mathsf{CCZ}, \ldots, \mathsf{C^{d-1}Z}\}$*. A circuit representation of* C *can then be learned through* $O(n^d)$ *queries to* C *and using only separable measurements. This can be improved to* $O(n^{d-1})$ *queries to* C *and using entangled measurements.*

Proof. From Proposition 3, we note the correspondence between C' and the binary phase state $|\psi_f\rangle$. From Theorem 3, we have that we can learn the $\mathbb{F}_2$ representation of f corresponding to such a state, using $O(n^d)$ separable measurements. Given $O(n^d)$ uses of the unknown C, we thus learn f from samples generated by applying $\mathsf{H}^{\otimes n}\mathsf{C}$ on $|0\rangle^{\otimes n}$ followed by separable measurements. We obtain a circuit representation of C' (and hence C which is $\mathsf{H}^{\otimes n}\mathsf{C}'\mathsf{H}^{\otimes n}$) by inserting gates $\mathsf{C^{|J|-1}Z}_{i_1,i_2,\ldots,i_{|J|}}$ (where $i_1, i_2, \ldots, i_{|J|} \in J$) for each monomial $\prod_{i\in J\subseteq[n]} x_i$, characterized by set J, present in f. The result for entangled measurements is obtained through application of Theorem 4. □

2.9.2 Learning circuits containing diagonal gates in the Clifford hierarchy

For any two n-qubit unitaries $U, V \in \mathcal{U}(n)$, let $[U, V] = UVU^\dagger V^\dagger$ denote the group commutator, and let $P(n) = \langle iI, X_j, Y_j, Z_j : j \in [n]\rangle$ be the n-qubit Pauli group. The d^{th} level of

the Clifford hierarchy on n-qubits, denoted $C_d(n)$, is defined inductively

$$C_1(n) = P(n), \tag{2.77}$$
$$C_d(n) = \{U \in \mathcal{U}(n) : [U, p] \in C_{d-1}(n), \forall p \in P(n)\}. \tag{2.78}$$

The second level $C_2(n)$ is the n-qubit Clifford group, while higher levels are not groups at all, as they fail to be closed. In general, $C_d(n)$ includes $C_{d-1}(n)$ and more – for instance, some gates that are in the set $C_d(n)$ and not in any lower level of the hierarchy are $Z^{1/2^{d-1}}$ and the controlled-Z gate with $d-1$ control qubits.

Let $D_d(n)$ denote the subset of diagonal unitaries in $C_d(n)$. In fact, $D_d(n)$ are groups for all d and n. Moreover, $D_d(n)$ can be generated by $Z^{1/2^{d-1}}$ and $C^{(i)}Z^{1/2^j}$ with $i + j = d - 1$ [ZCC08]. In [CGK17], the authors characterize unitaries in the diagonal Clifford hierarchies for qudits with prime power dimension. We reproduce one of these results for qubits.

Theorem 14. *For $d > 1$, $V \in D_d(n)$ if and only if, up to a global phase, V takes the form*

$$\exp\left(i\frac{\pi}{2^d}\sum_{\substack{S\subseteq[n]\\S\neq\emptyset}} a_S Z^S\right) \tag{2.79}$$

with $Z^S = \prod_{j\in S} Z_j$ and $a_S \in \mathbb{Z}$ for all $S \subseteq [n]$. We also have $V \notin D_{d-1}(n)$ if and only if at least one a_S is odd.

Proof. The reverse direction, that V in the form of Eq. (2.79) is in $D_d(n)$, is easy to show inductively.

The forward direction is a simple proof via contradiction. Suppose $V \in D_d(n)$. Any diagonal unitary can be written in the form of Eq. (2.79) if we allow the a_S to be real numbers.[8] So assume that for some $S_0 \subseteq [n]$, $S_0 \neq \emptyset$ we have $a_{S_0} \notin \mathbb{Z}$. Let $i \in S_0$ and define a Clifford unitary C to be a circuit of CX gates that maps Z^{S_0} to $Z^{(i)} = Z_i$ and any other Z^S, $S \neq S_0$, to some $Z^{S'}$ with $i \notin S'$.

Now, define $K_0 = CVC^\dagger$, $K_j = [K_{j-1}, X_i]$. Since $V \in C_d(n)$, also $CVC^\dagger \in C_d(n)$, and we must have $K_d = \pm I$. Calculating K_d however we have

$$K_1 = [CVC^\dagger, X_i] = \exp\left(-i\frac{\pi}{2^{d-1}}a_{S_0}Z_i\right), \tag{2.80}$$
$$K_j = K_{j-1}^2, \quad j \geq 2, \tag{2.81}$$
$$K_d = \exp\left(-i\pi a_{S_0}Z_i\right). \tag{2.82}$$

We thus realize that K_d can only be proportional to identity if a_{S_0} is an integer, from which we get our contradiction. □

We then have the following statement regarding the states produced by circuits $V \in D_d(n)$.

[8] Suppose $V = \sum_T e^{-i\pi\phi_T/2^n}|T\rangle\langle T|$. Then we can choose the phases in Eq. (2.79) to be $a_S = \frac{1}{2^N}\sum_T(-1)^{S\cdot T}\phi_T$ (treating $S, T \in \{0,1\}^n$ as bit strings). The inverse is of course $\phi_T = \sum_S(-1)^{S\cdot T}a_S$.

Proposition 4. *Let V be an n-qubit quantum circuit belonging to $D_d(n)$, the subgroup of diagonal unitaries in the d-th level of the Clifford hierarchy $C_d(n)$. The state produced by the action of V on $|+\rangle^{\otimes n} = \mathsf{H}^{\otimes n}|0\rangle^{\otimes n}$ (up to a global phase) is*

$$|\psi_f\rangle = V|+\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \omega_q^{f(x)}|x\rangle, \tag{2.83}$$

for some $f \in \mathcal{P}_q(n,d)$ with $q = 2^d$, with the $\mathbb{F}_2$-representation

$$f(x) = \sum_{\substack{T\subseteq[n]\\ 1\le|T|\le d}} c_T \prod_{j\in T} x_j \pmod{2^d}, \quad c_T \in 2^{|T|-1}\mathbb{Z}_{2^{d+1-|T|}}. \tag{2.84}$$

Proof. Applying a unitary from $D_d(n)$ to the state $|+\rangle^{\otimes n}$, we obtain a generalized phase state

$$V|+\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \omega_{2^{d+1}}^{g(x)}|x\rangle, \tag{2.85}$$

$$g(x) = \sum_{\substack{S\subseteq[n]\\ S\neq\emptyset}} a_S \prod_{i\in S}(-1)^{x_i}, \quad a_S \in \mathbb{Z}. \tag{2.86}$$

We can also understand this phase state by converting $g(x)$ to its $\mathbb{F}_2$-representation using $(-1)^{x_j} = 1-2x_j$ (as $x \in \{0,1\}^n$). Since $g(x)$ can be evaluated modulo 2^{d+1}, monomials with degree greater than d can be removed. We find

$$g(x) = \sum_{\substack{T\subseteq[n]\\ |T|\le d}} b_T \prod_{j\in T} x_j \pmod{2^{d+1}}, \quad b_T = (-2)^{|T|} \sum_{\substack{S\supseteq T\\ S\neq\emptyset}} a_S \in 2^{|T|}\mathbb{Z}. \tag{2.87}$$

We note, from the $b_\emptyset$ term, that V introduces a phase of $\omega_{2^{d+1}}^{b_\emptyset}$ to the basis state $|0\rangle^{\otimes n}$. Removing this, we obtain $\tilde{g}(x) = g(x) - b_\emptyset$, which is divisible by 2, i.e. $f(x) = \tilde{g}(x)/2$ is a polynomial, because all b_T for $T \neq \emptyset$ are even. Therefore, Eq. (2.85) becomes

$$\omega_{2^{d+1}}^{-b_\emptyset} V|+\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \omega_{2^d}^{f(x)}|x\rangle, \tag{2.88}$$

which are exactly the states from Eq. (2.11), since $f(x)$ is the degree-d polynomial

$$f(x) = \sum_{\substack{T\subseteq[n]\\ 1\le|T|\le d}} c_T \prod_{j\in T} x_j \pmod{2^d}, \quad c_T = b_T/2 \in 2^{|T|-1}\mathbb{Z}. \tag{2.89}$$

□

To learn a circuit representation of V from samples, we have the following result.

Theorem 15. *Let V be an unknown n-qubit quantum circuit in $D_d(n)$, the group of diagonal unitaries in the d-th level of the Clifford hierarchy. A circuit representation of V can then be learned through $O(n^d)$ queries to V and using only separable measurements.*

Proof. From Proposition 4, we note the correspondence between V and the generalized phase state $|\psi_f\rangle$. Using Theorem 8, we can learn the multi-linear representation of f (Eq. (2.84)) corresponding to such a state, using separable measurements on $O(n^d)$ copies of $V\mathsf{H}^{\otimes n}|0\rangle^{\otimes n}$. We obtain a circuit representation of V by inserting appropriate gates corresponding to monomials $\prod_{i\in T\subseteq[n]} x_i$, characterized by set T, present in f. We now define these gates with respect to the state $|x\rangle$ where $x \in \{0,1\}^n$. For $|T|\geq 2$, we insert a controlled-diagonal gate over qubits in T, that puts a phase of $\exp(i\pi c_T/2^{d-1})$ if $x_j = 1$, $\forall j \in T$ and no phase otherwise. For monomials corresponding to singletons x_j for $j \in [n]$ (i.e., $|T|= 1$), we insert the phase gate $Z^{c_T/2^{d-1}} = |0\rangle\langle 0| + \exp(i\pi c_T/2^{d-1})|1\rangle\langle 1|$ on qubit j. □

2.10 Discussion

Our work leaves open a few interesting questions, some of which will be mentioned again in Chap. 7.

Improving runtime. While our algorithms for learning phase states are optimal in terms of the sample complexity, their runtime scales polynomially with the number of qubits only in the case of binary phase states and separable measurements. It remains to be seen whether a polynomial runtime can be achieved in the remaining cases of learning binary phase states with entangled measurements and generalized phase states with separable measurements.

Quantum advantage. Suppose U is a polynomial size quantum circuit such that $U|0^n\rangle$ is a low-degree phase state associated with some Boolean function $f : \{0,1\}^n \to \{0,1\}$. Our results imply there exists an efficient quantum algorithm that learns f given a classical description of U. An interesting open question is whether the problem of learning f given a description of U is classically hard. If this is the case, our results would imply a quantum advantage for the considered learning task.

Property testing. What is the sample complexity of property testing phase states? Given M copies of $|\phi\rangle$ with the promise that either $|\phi\rangle$ is a degree-d phase state or ε-far from the set of degree-d phase states, what is an upper and lower bound on M? For $d = 1$, we can learn the entire state using $M = 1$ copy and for $d = 2$, Gross et al. [GNW21] showed that $M = 6$ copies suffice for this testing question. For larger d, understanding the complexity of testing phase states is an intriguing open question left open by our work, in particular does the sample complexity of testing n-qubit degree-d phase states scale as n^{d-2} (for $d \geq 2$) or does it scale as $\mathsf{poly}(c^d, n)$ for some $c > 1$?

Learning more expressive quantum states. We leave as an open question whether our learning algorithms can be extended to binary phase states with a small *algebraic degree*. Such states have amplitudes proportional to $(-1)^{\mathsf{tr}F(x)}$, where $F(x) = \sum_{i=0}^{d} a_i x^i$ is a degree-d polynomial with coefficients $a_i \in \mathbb{F}_{2^n}$ and $\mathsf{tr} : \mathbb{F}_{2^n} \to \mathbb{F}_2$ is the trace function defined as $\mathsf{tr}(x) = \sum_{j=0}^{n-1} x^{2^j}$. Here all arithmetic operations use the field $\mathbb{F}_{2^n}$. What is the sample complexity of learning n-qubit states produced by circuits containing *non-diagonal* unitaries in the k-th level in the Clifford hierarchy, on the $|+\rangle^n$ input? Similarly, what is the complexity of learning a state which has stabilizer rank k?[9] Similarly can we PAC learn these classes of

[9] We know how to learn stabilizer states and stabilizer-rank 2 states in polynomial time, what is the complexity as a function of rank-k?

quantum states in polynomial time?[10]

[10]For stabilizer circuits, we have both positive and negative results in this direction [Roc18; Lia23] but for more generalized circuits, it remains an open question.

Chapter 3

Heisenberg-limited *hidden* quantum channel discrimination

3.1 Introduction

Discriminating between physical operations, often called quantum channel discrimination (QCD) in quantum information science, is a fundamental task in experiments [Aćí01; DFY07; Cal+08; ZP20; Pir+19; WW19]. In QCD, an unknown physical operation is modeled as a quantum channel C through a completely-positive trace-preserving map acting on the system of interest [NC11]. The goal is to identify C from known alternatives using a discrimination protocol. Discrimination protocols are considered efficient (1) when a desired error probability is achieved with fewer queries compared to classical methods [GLM04b; Bra+18] or particularly successful (2) when the error probability is zero [Aćí01; DFY07; DFY09]. For example, sequential protocols [DFY07] involve an initial state ρ_m, and a positive operator-valued measurement (POVM) M, and N queries, each consisting of the unknown channel C and tunable unitary operations V_n $(n = 1, \ldots, N)$, as shown in Fig. 3.1(a). Protocols including sequential and parallel protocols are able to achieve (1) and (2) when arbitrary operations of V_n and measurements M are allowed on the system [Har+10; Pez+18a; Bra+18].

While conventional QCD considers a fully controllable system, experimental systems often consist of a fully-controllable subsystem, which we call the measurement system $\mathcal{M}$, and an uncontrollable subsystem, which we call the channel system $\mathcal{H}$ [IHY85; GLP98; Sch+05; Kat+22; Pec+21; Xia+13]. Here, $\mathcal{M}$ interacts with $\mathcal{H}$ to detect the action of C on $\mathcal{H}$. Such composite systems are used in quantum non-demolition measurements [IHY85; GLP98], quantum logic detection [Sch+05; Kat+22], and occur in designs of superconducting quantum devices [Pec+21].

These experiments motivate us to consider the following restrictions on system $\mathcal{H}$ in QCD: arbitrary control of $\mathcal{H}$ is not possible, measurement on $\mathcal{H}$ is not allowed, and the initialization of $\mathcal{H}$ is unreliable. The state on $\mathcal{H}$ thus evolves only under the dynamics C on $\mathcal{H}$. The separation between $\mathcal{H}$ and $\mathcal{M}$ motivates the third restriction as one no longer has control over the state preparation on $\mathcal{H}$ and the initial state cannot be purified. We call $\mathcal{H}$ probed under these three restrictions hidden, and the associated channel discrimination problem Hidden system Quantum Channel Discrimination (HQCD). The effect of these

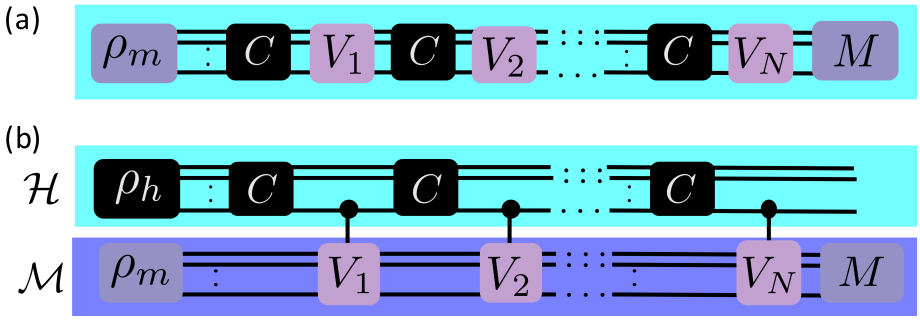


Figure 3.1: Comparison between conventional quantum channel discrimination (QCD) and hidden quantum channel discrimination (HQCD). Here the black boxes indicate the unknown channels and state. In both cases, the action of unknown channel C is inferred by selecting an input state ρ_m, applying controlled V_n operations ($n = 1, \ldots, N$), and measuring with M to minimize the error probability. (a) Conventional QCD, involving the direct manipulation and measurement of the system. (b) HQCD, where the physical system $\mathcal{H}$ and measurement system $\mathcal{M}$ are explicitly distinguished.

restrictions on a conventional sequential QCD protocol is illustrated in Fig. 3.1(b). The restrictions become crucial when the interactions between $\mathcal{H}$ and $\mathcal{M}$ have limited ability to change the state in $\mathcal{H}$ [1]. In a typical experiment, however, back-action on the channel is avoided by using a high-impedance meter. Here, we model this meter as a controlled unitary with the control on $\mathcal{H}$ and the unitary operation on $\mathcal{M}$.

It is then natural to ask: Is discrimination with zero error probability or with fewer queries than classical methods still possible under these restrictions? This is a difficult task if conventional QCD techniques are employed. For example, discrimination of a unitary channel is impossible when the input state is the maximally-mixed state in conventional QCD [Hel76]. Nevertheless, we give an affirmative answer to this question by studying Hidden Binary Channel Discrimination (HBCD), which is a minimal binary HQCD model consisting of two qubits as shown in Fig. 3.2, and by constructing concrete measurement protocols with the desired performance. The new protocols inherit ideas from conventional QCD, including sequential, parallel, and multi-shot strategies [Har+10; DCS17]. Inspired by studies showing that entanglement-free protocols can achieve performance comparable to that of parallel protocols A, in our new sequential and multi-shot protocols the input states are entanglement-free [DFY07; Bra+18; Hig+07a; Reh+18; RS21; KK22; DeB+23], while our parallel protocol can utilize an entangled initial state. Nevertheless, surprising performance differences arise.

In this chapter, we demonstrate that for the HBCD problem, sequential protocols outperform non-sequential protocols, including parallel and multi-shot protocols, in terms of the number of queries required to achieve a desired error probability. Furthermore, we prove that sequential protocols can achieve perfect discrimination with zero error. In contrast, we show a case where non-sequential protocols fail to solve HBCD when C is applied once before measurement. We extend the quantum metrology concepts of standard quantum limit

[1] When the universal gate set is available for interaction, one could use the SWAP gate between $\mathcal{H}$ and $\mathcal{M}$. The problem then becomes equivalent to conventional QCD.

(SQL) and the Heisenberg limit to HBCD.

The number of queries needed to solve the HBCD by sequential protocols is proven to be asymptotically optimal using an information-theoretic bound and saturates the Heisenberg limit, whereas non-sequential protocols achieve only the SQL. These advantages of sequential protocols over parallel protocols in QCD are reported for the first time to the best of our knowledge.

3.2 Hidden binary channel discrimination (HBCD)

In this section, we formally introduce the problem of HBCD, describe the type of queries allowed in our learning model and finally the design of protocols involving these queries.

3.2.1 Problem statement

In our HBCD problem, we consider a two qubit system composed of a one-qubit hidden system $\mathcal{H}$ on which the unknown channel C acts and a one-qubit measurement system $\mathcal{M}$ used to learn C.

Definition 3. Unknown Channel *C Let $\alpha \in (0, 2\pi)$, and θ_C be a Bernoulli random variable taking values in $\{0, \alpha\}$ with probability $P_{\theta_C}(0) = P_{\theta_C}(\alpha) = 1/2$. The unknown quantum channel acting on $\mathcal{H}$ is then $C = e^{i\theta_C \sigma_x}$.*

Definition 4. Query. *A query $Q(\psi, \phi)$ is a unitary operation that acts on the two-qubit system composed of $\mathcal{H}$ and $\mathcal{M}$, and is parametrized by a pair of phases $\{\psi, \phi\}$. The circuit of $Q(\psi, \phi)$ is depicted in Fig. 3.2(a). It involves three components: (i) the unknown channel C, (ii) a controlled rotation on $\mathcal{M}$ by ψ along the z-axis conditioned on the state of $\mathcal{H}$, and (iii) a single-qubit rotation on $\mathcal{M}$ by ϕ along the x-axis.*

The query as defined above is inspired from *quantum signal processing* (QSP) [LC17] and lends to the success of the constructed protocols. Connections to QSP are elaborated in Appendix A.2. We now define our HBCD problem.

Definition 5. HBCD Problem. *Suppose that $\epsilon \in [0, 1/2]$, and ρ_h is the initial one-qubit mixed state on $\mathcal{H}$. Let C be the unknown channel from Def. 3 with θ_C determined at the start of the experiment and which remains constant for all subsequent queries. Then $\mathsf{HBCD}(\alpha, \epsilon, \rho_h)$ defines the problem of learning an estimate $\hat{\theta}_C$ of the unknown θ_C with error probability $P(\hat{\theta}_C \neq \theta_C) \leq \epsilon$.*

3.2.2 Protocols for HBCD

We would ideally like to solve an HBCD problem using as few queries as possible. In addition to specifying these queries, we are allowed to specify the initial state ρ_m to $\mathcal{M}$ and the POVM measurement M acting on $\mathcal{M}$. Collectively, this is used to design a discrimination protocol Σ to learn the unknown θ_C. The discrimination protocols considered here involve N queries $\{Q_1, \ldots, Q_N\}$. We denote the corresponding vector of phases as $\Phi \equiv (\psi_1, \ldots, \psi_N, \phi_1, \ldots, \phi_N) \in [0, 2\pi)^{2N}$.

Definition 6. Discrimination Protocols. *Given a problem $\mathsf{HBCD}(\alpha, \epsilon, \rho_h)$, we define a discrimination protocol $\Sigma(N, d, Z, S)$ where N is the total number of the queries used, depth d*

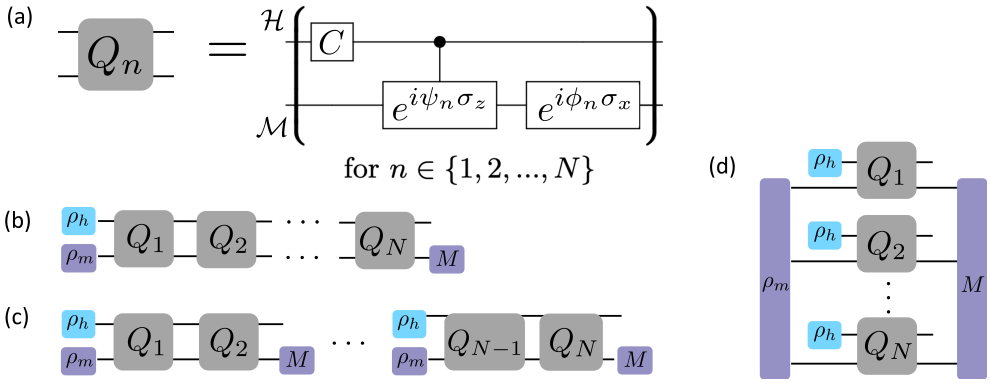


Figure 3.2: The query and sequential/multi-shot/parallel protocols. (a) Query Q_n (Def. 4) with phases ψ_n and ϕ_n specified independently. The upper (hidden) qubit undergoes unitary evolution every round and at the end we measure the lower (measurement) qubit. (b-d) Discrimination protocol S: (b) Sequential protocol, (c) Multi-shot protocol with depth $d = 2$, (d) Parallel protocol. ρ_m can be a highly-entangled state for the parallel protocol.

is the number of concatenated queries before measurement, $Z = (\rho_m, \Phi, M)$ is the collection of specified settings with Φ being the vector of phases specifying the N queries, and S defines the type of protocol which can be sequential, multi-shot or parallel. The circuit corresponding to Σ for different S is shown in Fig. 3.2(b)-(d).

Note that our discrimination protocols are designed using knowledge of ρ_h and α. The depth d takes the value of N when S is sequential, N/m when S is a multi-shot protocol using m shots and 1 when S is a parallel protocol over an N-qubit measurement system $\mathcal{M}$ interacting with N copies of $\mathcal{H}$ (see Fig. 3.2(d)). Our sequential protocol uses one probe qubit [2], which is entanglement-free and has weaker discrimination performance compared to parallel protocols in conventional QCD [PW09a; BCP19]. We compare their performances in HBCD. The multi-shot protocol allows for adaptive choice of Z, but we do not explore it in this chapter [3].

Let us now define the discrimination error associated with each protocol. Suppose $(y^1, \ldots, y^m)$ is a set of m POVM outcomes, collectively denoted by the vector $\mathbf{y} \in \{0, 1\}^m$. Given $\mathbf{y}$, an estimator $\hat{\theta}_C(\mathbf{y})$ will output either 0 or α. The error probability of a protocol Σ is then

$$\begin{aligned} &P(\hat{\theta}_C(\mathbf{y}) \neq \theta_C; \Sigma) \\ &= \frac{1}{2}\left[P_{\hat{\theta}_C(\mathbf{y})|\theta_C}(\alpha|0; \Sigma) + P_{\hat{\theta}_C(\mathbf{y})|\theta_C}(0|\alpha; \Sigma)\right], \end{aligned} \tag{3.1}$$

where we have used the fact that the prior probabilities satisfy $P(\theta_C = 0) = P(\theta_C = \alpha) = 1/2$. The estimator is designed such that $\hat{\theta}_C(0) = 0$ and $\hat{\theta}_C(1) = \alpha$. Therefore, the error is

[2] We note that sequential feedback protocols with N qubit entangled input state are referred to as sequential protocols in some literature, and are strictly stronger than parallel protocols [Yua16; BMQ21].

[3] When d is fixed, we do not expect adaptivity to change the asymptotic scaling of N with α as observed in the case of conventional QCD [CMW16; SHW22]. When d is allowed to adaptively change, a higher scaling may be achieved. However, this is already captured by the sequential protocol.

given by $\frac{1}{2}(P_{y|\theta_C}(0|\alpha) + P_{y|\theta_C}(1|0))$. We now give an overview of the different estimators that can be used in conjunction with the protocols.

3.2.3 Estimators

In this section, we describe the different estimators that can be used for solving HBCD with different protocols. Let the measurement outcomes from applying any protocol be given by $y^k \in \{0,1\}$, indexed by k. We collectively denote the vector of m binary outcomes as $\mathbf{y} = (y^1, \ldots, y^m) \in \{0,1\}^m$. Suppose we set the phases Φ corresponding to the protocol such that $y = 1$ with a high probability for $\theta_C = \alpha$ and $y = 0$ with a high probability for $\theta_C = 0$. Some estimators that can then be used are as follows.

Majority Vote. A simple (albeit suboptimal) estimator for $\hat{\theta}_C$ uses the majority vote (denoted by Maj) of the measurement outcomes:

$$\hat{\theta}_C = \alpha \cdot \mathrm{Maj}(\mathbf{y}). \tag{3.2}$$

Likelihood Ratio Test. The likelihood ratio test (LRT) or the maximum likelihood estimator is the optimal estimator for binary hypothesis testing.

Consider the log-likelihood function:

$$\begin{aligned} L(\mathbf{y};\theta_C) &= \sum_{k=1}^{m} \log p(y^k|\theta_C) && (3.3)\\ &= Y_1 \log p(1|\theta_C) + Y_0 \log p(0|\theta_C), && (3.4) \end{aligned}$$

where $Y_1 = \sum_k y^k$ and $Y_0 = m - Y_1$

From maximum likelihood, we then have that

$$\hat{\theta}_C = \begin{cases} \alpha & , L(\mathbf{y};\alpha) > L(\mathbf{y};0) \\ 0 & , L(\mathbf{y};\alpha) \leq L(\mathbf{y};0) \end{cases} \tag{3.5}$$

3.3 Performance guarantees for protocols

Another metric of performance of the protocols is the scaling of the minimal number of queries N required to solve $\mathsf{HBCD}(\alpha, \epsilon, \rho_h)$. As α becomes smaller, solving HBCD becomes more difficult and hence N should grow. We can then define two scaling limits of N with α.

Definition 7 (Standard quantum limit and Heisenberg scaling in HBCD.)**.** *Suppose that $\epsilon \in [0, 1/2)$, ρ_h, S, and d are given. For $0 < \alpha \ll 1$, let $N(\alpha)$ be the number of queries needed to solve $\mathsf{HBCD}(\alpha, \epsilon, \rho_h)$ by $\Sigma(N, d, Z, S)$. We say that a depth-d S protocol achieves the standard quantum limit (SQL) if $N(\alpha) = \Theta(\alpha^{-2})$ and Heisenberg scaling if $N = \Theta(\alpha^{-1})$.*

The SQL and the Heisenberg limit are defined in quantum metrology for parameter estimation in terms of the number of access to an unknown physical system of interest. This corresponds to the number of interactions N between $\mathcal{H}$ and $\mathcal{M}$. In parameter estimation,

a protocol is said to achieve SQL when the number of queries N required to achieve an estimation error α_{PE} scales as $N \sim \alpha_{\mathrm{PE}}^{-2}$ and the Heisenberg limit when $N \sim \alpha_{\mathrm{PE}}^{-1}$ [GLM11b; Bra+18; Pez+18a]. Similarly, we can model the problem of discriminating the value of θ_C from $\{0, \alpha\}$ in HBCD as estimating the value of θ_C. In this case, we succeed if the estimation error is smaller than half of the angle difference $(\alpha/2)$. Definition 7 is then evident.

Perfect discrimination. We now discuss advantages of using sequential protocols in HBCD.

Theorem 16 (Perfect discrimination in HBCD)**.** *Let $\alpha \in (0, 2\pi)$. Suppose there is a constant j which takes the value of 1 for $\alpha \in [0, \frac{\pi}{4}] \cup [\frac{3\pi}{4}, \frac{5\pi}{4}] \cup [\frac{7\pi}{4}, 2\pi)$, 2 for $[\frac{3\pi}{8}, \frac{5\pi}{8}] \cup [\frac{11\pi}{8}, \frac{13\pi}{8}]$ and 3 for the rest. There exists a sequential protocol $\Sigma(N, d = N, Z, S = \text{sequential})$ that solves HBCD$(\alpha, \epsilon = 0, \rho_h)$ with at most $N = j\lceil \frac{2\pi}{\beta} \rceil$ queries.*

Here, β is an effective rotation angle on the measurement qubit (to be shown soon). To prove the theorem, we first make a diagonal unitary matrix with four query iterations. The controlled rotation then becomes a single-qubit R_Z gate on $\mathcal{M}$ with its rotation angle being either $-\beta$ for $\theta_C = 0$ or β for $\theta_C = \alpha$. Using this rotation on $\mathcal{M}$, we accumulate the phase $\pm\beta$ in the measurement qubit so that measurement qubit is $|0\rangle$ for $\theta_C = 0$ and $|1\rangle$ for $\theta_C = \alpha$ [RC21; Mar+21a].

Proof of Theorem 16. Let us firstly consider $\alpha \in \mathcal{D}_1 = [0, \frac{\pi}{4}] \cup [\frac{3\pi}{4}, \frac{5\pi}{4}] \cup [\frac{7\pi}{4}, 2\pi)$ and prove that the perfect discrimination is possible in this case. Later, we will extend the results to all α. Consider the following unitary matrix which involves four applications of queries

$$\check{Q}(\theta_C, \psi) = Q_4 Q_3 Q_2 Q_1, \tag{3.6}$$

with $\phi_n = 0$ and $\psi_n = \psi$ for $\forall n$.

In this section we use computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ where the first qubit is in $\mathcal{H}$ and the second qubit is in $\mathcal{M}$. Then the matrix elements of $\check{Q}$ is given as follows:

$$\check{Q} = \begin{bmatrix} P_1(x,a) & 0 & iR(x,a) & 0 \\ 0 & P_2(x,a) & 0 & i\frac{1}{a^k}R(x,a) \\ iR(x,a) & 0 & P_3(x,a) & 0 \\ 0 & i\frac{1}{a^k}R(x,a) & 0 & P_4(x,a), \end{bmatrix}, \tag{3.7}$$

where we use parametrizations $x \equiv \cos\theta_C$ and $a \equiv \exp(i\psi)$, $P_i(x,a)$ is a 4-degree polynomial of x and a, which is even in θ_C,

$$P_1(x,a) = a^2 - a(1+a)(3+a)x^2 + (1+a)^3x^4 \tag{3.8}$$

$$P_2(x,a) = \frac{a - (1+a)(1+3a)x^2 + (1+a)^3x^4}{a^3} \tag{3.9}$$

$$P_3(x,a) = a(a - (1+a)(1+3a)x^2 + (1+a)^3x^4) \tag{3.10}$$

$$P_4(x,a) = \frac{a^2 - a(1+a)(3+a)x^2 + (1+a)^3x^4}{a^4}, \tag{3.11}$$

and $R(x,a)$ is a function that has a following form

$$R(x,a) = x\sqrt{1-x^2}a(1-a)(-2a + (1+a)^2x^2). \tag{3.12}$$

An important observation for $\check{Q}$ is that all the off-diagonal elements share $R(x,a)$. For a given x there exists $\tilde{a}(x)$ such that

$$R(x,\tilde{a}(x))=0, \tag{3.13}$$

if and only if $\theta_C \in \mathcal{D}_1$.

Now, we know that θ_C is a Bernoulli random variable in our HBCD, taking a value of 0 or α. Since we assume that $\alpha \in \mathcal{D}_1$, we choose $a=\tilde{a}(\alpha)$. Then we can readily show that $\check{Q}$ becomes diagonal both for $\theta_C=0$ and for $\theta_C=\alpha$:

$$\check{Q}\left(0,-i\log(\tilde{a})\right)=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&e^{2i\beta}&0\\0&0&0&e^{-2i\beta}\end{bmatrix}, \tag{3.14}$$

and

$$\check{Q}\left(\alpha,-i\log(\tilde{a})\right)=\begin{bmatrix}e^{i\beta}&0&0&0\\0&e^{-i\beta}&0&0\\0&0&e^{i\beta}&0\\0&0&0&e^{-i\beta}.\end{bmatrix}, \tag{3.15}$$

where the rotation angle β is given by $\beta=-i\log(P_1(\cos\alpha,\tilde{a}))$.

Now we can obtain an analytical solution that achieves the perfect discrimination between $\theta_C=0$ and $\theta_C=\alpha$ when $\alpha\in\mathcal{D}$. First, when β has the form of $\beta=\frac{\pi}{2n}$ ($n\in\mathbb{N}$), we make a n-th power of $\check{Q}^n$.

$$\check{Q}^n\left(\theta_C=0,\psi=-i\log(a_{\mathrm{sol}})\right)=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix}, \tag{3.16}$$

and

$$\check{Q}^n\left(\theta_C=\alpha,\psi=-i\log(a_{\mathrm{sol}})\right)=\begin{bmatrix}i&0&0&0\\0&-i&0&0\\0&0&i&0\\0&0&0&-i.\end{bmatrix}, \tag{3.17}$$

In this case, perfect discrimination is done with $N=4n$ queries with $\rho_m=\frac{1}{2}(|0\rangle+i|1\rangle)(\langle 0|-i\langle 1|)$ and $\vec{\phi}_{\mathrm{sol}}\equiv\{\phi_1,\cdots\phi_{4n}\}=\{0,0,\cdots,0,\frac{\pi}{4}\}$. The overall unitary matrix $U=Q_{4N}\cdots Q_1$ reads:

$$U\left(\theta_C=0\right)=\frac{1}{\sqrt{2}}\begin{bmatrix}1&i&0&0\\i&1&0&0\\0&0&-1&-i\\0&0&-i&-1\end{bmatrix}, \tag{3.18}$$

and

$$U\left(\theta_C = \alpha\right) = \frac{1}{\sqrt{2}} \begin{bmatrix} i & 1 & 0 & 0 \\ -1 & -i & 0 & 0 \\ 0 & 0 & i & 1 \\ 0 & 0 & -1 & -i \end{bmatrix}. \tag{3.19}$$

We can readily check that regardless the state in $\mathcal{H}$ the final state in $\mathcal{M}$ is $|1\rangle$ for $\theta_C = 0$ and $|0\rangle$ for $\theta_C = \alpha$. Therefore this unitary achieves perfect discrimination. When $\frac{\pi}{2(n-1)} < \beta < \frac{\pi}{2n}$, we can also construct the function that satisfies Eqs. (3.18) and (3.19) by choosing the angles ϕ_n [Ros+22]. Therefore proof is done for $\alpha \in \mathcal{D}_1$.

When $\alpha \notin \mathcal{D}_1$, we consider a process

$$S_2 = Q_2 Q_1 \tag{3.20}$$

for $\alpha \in \mathcal{D}_2 = [\frac{3\pi}{8}, \frac{5\pi}{8}] \cup [\frac{11\pi}{8}, \frac{13\pi}{8}]$ and

$$S_3 = Q_3 Q_2 Q_1 \tag{3.21}$$

for the rest, i.e. $\alpha \in \mathcal{D}_3 = [\frac{\pi}{4}, \frac{3\pi}{8}] \cup [\frac{5\pi}{8}, \frac{3\pi}{4}] \cup [\frac{5\pi}{4}, \frac{11\pi}{8}] \cup [\frac{13\pi}{8}, \frac{7\pi}{4}]$. In both cases, we choose $\psi_n = \phi_n = 0$ for all n. Then we show that S has a form

$$S_n = \begin{bmatrix} \cos(n\theta_C) & i\sin(n\theta_C) & 0 & 0 \\ i\sin(n\theta_C) & \cos(n\theta_C) & 0 & 0 \\ 0 & 0 & \cos(n\theta_C) & i\sin(n\theta_C) \\ 0 & 0 & i\sin(n\theta_C) & \cos(n\theta_C) \end{bmatrix}, \tag{3.22}$$

which is nothing but $R_x(n\theta_C)$ rotation gate on $\mathcal{H}$. We can readily show that $2\alpha \in \mathcal{D}_1$ for $\alpha \in \mathcal{D}_2$, and $3\alpha \in \mathcal{D}_1$ for $\alpha \in \mathcal{D}_3$. Therefore both cases reduce to the first case, $\alpha \in \mathcal{D}_1$.

An important consequence of the fact that the polynomials P_i are even polynomials of the unknown angle θ_C is that the discrimination protocol does not depend on the initial state of the hidden system. First consider the situation that the initial state of the hidden system is pure. Then it can be expressed in the eigenbasis of the channel Hamiltonian (i.e., σ_x)

$$|\psi_{\mathcal{H}}\rangle = \sqrt{\alpha}|+\rangle + \sqrt{1-\alpha}|-\rangle. \tag{3.23}$$

Applying the proposed composite protocol $U(\theta_C = 0)$ to this initial state then $\mathcal{M}$ does not flip independently of $|\psi_{\mathcal{H}}\rangle$. Similarly, applying $U(\theta_C = 0)$ to the composite systems, we see $\mathcal{M}$ is flipped independently of $|\phi_{\mathcal{H}}\rangle$. As a result, when $\alpha \in \mathcal{D}$, then the discrimination protocol is independent of the initial state.

Lastly, we briefly mention why this solution only achieves the SQL. The number of query needs for this solution is characterized by β. Let us assume that α is small. Then $x = \cos(\alpha) = 1 - \frac{\alpha}{2} + O)(\alpha^2)$. By solving $R(x, a) = 0$ and $\beta = -i\log(P_1(\cos\alpha, \tilde{a}))$ for small α, we obtain $\beta = 2\alpha^2$. This means that the solution in Theorem 16 obeys the SQL. □

Although the above sequential protocol only achieves the SQL, our numerical results and information-theoretic bound show that sequential protocols can be designed to attain the Heisenberg limit.

Weakness of the non-sequential protocol. Since any entanglement improves conventional QCD [PW09a; BCP19], one may expect parallel protocols to be better than sequential protocols. Indeed, this expectation is valid when the channel is noisy and error correction is unavailable [Zho+18]. However, when the process is noiseless but the state is noisy, the opposite is true; the sequential protocol outperforms the parallel protocol. When the query depth $d = 1$ and the initial state in $\mathcal{H}$ is maximally mixed, we show that discrimination is impossible for non-sequential protocols.

Theorem 17 (Impossible case for depth-1 non-sequential protocols)**.** *Suppose that $\rho_h = \frac{\mathbb{I}}{2}$, and S is the multi-shot or parallel protocol. For any $\epsilon \in [0, 1/2)$ and $\alpha \in (0, 2\pi)$, the protocols $\Sigma(N, d = 1, Z, S)$ cannot solve $\mathit{HBCD}(\alpha, \epsilon, \rho_h)$. That is, Σ does not obtain any information on θ_C through M.*

Proof. We show this by keeping track of the quantum state with each application of a query in the circuit. After applying C, the state becomes $\rho_h \otimes \rho_m$. Notice that the rotation on the hidden qubit does not change ρ_h because it is the maximally mixed state. Then we apply the controlled rotation gate and $e^{i\phi_1\sigma_z}$, and obtain

$$\frac{1}{2}(|0\rangle\langle 0|R_x(\phi_1)R_z(\psi_1)\rho_m R_z(-\psi_1)R_x(-\phi_1) + |1\rangle\langle 1|R_x(\phi_1)\rho_m R_x(-\phi_1)). \tag{3.24}$$

This is independent of θ_C, and thus, the measurement outcome of the measurement qubit does not determine θ_C at all. The situation does not change even when the parallel protocol is used. Since $R_x(\theta_C)^{\otimes N} 1^{\otimes N} R_x^\dagger(\theta_C)^{\otimes N} = 1^{\otimes N}$, the rotations on the hidden qubits do not change the state of the hidden qubits through the controlled rotations. Thus, the measurement qubit does not depend on whether $\theta_C = 0$ ore α. □

The key idea to the above proof is that the maximally mixed state remains invariant under single qubit rotations. Therefore, the state (ρ_M) of $\mathcal{M}$ before measurement does not depend on the value of θ_C. However, if $d \geq 2$ queries are used, ρ_M correlates with θ_C through the controlled interaction. Thus, protocols with $d \geq 2$ queries are strictly better than non-sequential protocols of $d = 1$. Next, we prove the asymptotic number of queries required to solve the HBCD for $d = 2$. The multi-shot protocol with a fixed depth cannot achieve the Heisenberg limit (Theorem 18), which is illustrated numerically later.

Theorem 18 (Standard quantum limit in HBCD by multi-shot protocol)**.** *For all $\epsilon \in [0, 1/2)$ and ρ_h, depth-2 multi-shot protocol achieves the SQL.*

Proof. Here, we compute the asymptotic scaling of a multi-shot protocol with constant query depth i.e., $d = \text{const}$. Since the asymptotic scaling does not change for adaptive protocols [CMW16], we consider non-adaptive protocols. Suppose the minimum error probability of a single shot is p_s. Assume that the minimization is done for ρ_i, ϕ_i,ψ_n, and M, we estimate the lower bound for p_s.

The minimum error probability to distinguish two pure states is given by the Helstrom bound. In distinguishing two quantum channels, the minimum error probability of single-shot measurement P_s is obtained by minimizing the Helstrom bound over possible input states.

Suppose that we have a quantum circuit U to discriminate two channels. The action of U is different for $\theta_C = 0$ and for $\theta_C = \alpha$. Let U_i $(i = 1, 2)$ be a unitary operator in

the estimation protocol before measurement for $\theta_C = 0$ and α, respectively. Then the error probability of a one-shot measurement is given as follows.

$$P_{\mathrm{s}}(\hat{\theta}_C \neq \theta_C) = \min_{|\psi\rangle} \frac{1 - \sqrt{1 - |\langle\psi|U_1^\dagger U_2|\psi\rangle|^2}}{2}, \tag{3.25}$$

where $|\psi\rangle$ is the initial state. Since $|\langle\psi|U_1^\dagger U_2|\psi\rangle| \leq 1$, this minimization is equivalent to:

$$P_{\mathrm{s}}(\hat{\theta}_C \neq \theta_C) = \frac{1 - \sqrt{1 - \min_{|\psi\rangle}|\langle\psi|U_1^\dagger U_2|\psi\rangle|^2}}{2}. \tag{3.26}$$

Here, $\min_{|\psi\rangle}|\langle\psi|U_1^\dagger U_2|\psi\rangle|$ is nothing but the operator norm $\|U_1^\dagger U_2\|$. Therefore, we look for the bound of $\|U_1^\dagger U_2\|$.

We use a known bound for the operator norm. Consider a $K \times L$ matrix

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_L \end{pmatrix}, \tag{3.27}$$

where $\mathbf{a}_m$ is K dimensional vector. Then the operator norm is bounded as follows. For all l,

$$\|A\| \geq \|\mathbf{a}_l\| \tag{3.28}$$

We apply (3.28) to $\|U_1^\dagger U_2\|$ for $d = 2$. Then we obtain

$$P_{\mathrm{s}} \geq \frac{1 - 2\alpha}{2}. \tag{3.29}$$

In the multi-shot protocol, the measurement is performed m times. An estimate $\hat{\theta}_C$ is determined through LRT (Eq. 3.5) on these measurement outcomes. The probability of error P_e is given by

$$P_e = \frac{1}{2}\left(P_{\hat{\theta}_C|\theta_C}(0|\alpha) + P_{\hat{\theta}_C|\theta_C}(\alpha|0)\right). \tag{3.30}$$

This can be bounded as [Cov05]

$$2^{-mC(p_0,p_\alpha)} \geq P_e \geq \frac{1}{4}\exp\left(-mD(p_0||p_\alpha)\right), \tag{3.31}$$

where m is the number of measurements, p_0 (or p_α) is the probability distribution over measurements outcomes y when the truth is $\theta_C = 0$ (or $\theta_C = \alpha$), $C(\cdot)$ is the Chernoff information bound and $D(\cdot)$ is the KL divergence.

To achieve an error probability of at most ϵ, we can then show that

$$m = O\left(\frac{\log(1/4\epsilon)}{4\alpha^2}\right) \tag{3.32}$$

This completes the proof. □

The theorem implies that HBCD becomes challenging with decreasing α, and the minimum distinguishable value of α scales as $\alpha \sim N^{-1/2}$ with increasing N.

The advantages of sequential protocols over non-sequential protocols in HBCD are evident from Theorems 16-18. The sequential protocol alone enables perfect discrimination. Additionally, non-sequential protocols with $d = 1$ cannot determine θ_C regardless of the number of queries, while the sequential protocol can.

3.4 Heisenberg-limited learning by sequential protocol

The possibility of achieving the Heisenberg limit (Def. 7) is still unanswered. We first derive a lower bound on N required to solve HBCD.

3.4.1 Lower bound

Theorem 19 (Fundamental limit of HBCD)**.** *Any protocol* $\Sigma(N, d, Z, S)$ *with* $N < \frac{1}{\sqrt{2(1-\cos\alpha)}}$ *cannot solve* $\mathsf{HBCD}(\alpha, \epsilon = 0, \rho_h)$.

Proof. Let U_i $(i = 1, 2)$ be a unitary operator in the estimation protocol for $\theta_C = 0$ and α, respectively. For U_1 and U_2 to be perfectly discriminated, the following necessary conditions must be satisfied.

$$\mathcal{D}(U_1, U_2) = 0, \tag{3.33}$$

where the distance is defined by

$$\mathcal{D}(U_1, U_2) = \min_{\eta} |\langle \eta | U_1^\dagger U_2 | \eta \rangle| \tag{3.34}$$

and the minimization is done over all pure states [Aci01]. We evaluate (3.33) using the spectral norm of a matrix with the help of its subadditivity:

$$\|A_1 A_2 - B_1 B_2\| \leq \|A_1 - B_1\| + \|A_2 - B_2\|. \tag{3.35}$$

Then, the upper bound of the distance between U_1 and U_2 is obtained:

$$\|U_1 - U_2\| \leq \sum_{i=1}^{N} \|C(\theta_C = 0) - C(\theta_C = \alpha)\| = N\sqrt{2(1 - \cos\alpha)}. \tag{3.36}$$

It quantifies the distance between two unitary operators. Then we translate the distance of operators to the distinguishability of them. Substituting the upper bound of the distance into Eq. (3.33), we obtain

$$\mathcal{D}(U_1, U_2) \geq 1 - N\sqrt{2(1 - \cos\alpha)}. \tag{3.37}$$

Therefore, the necessary condition for (3.33) is

$$N \geq \frac{1}{\sqrt{2(1 - \cos\alpha)}}, \tag{3.38}$$

i.e., this gives a lower bound of the query complexity for perfect discrimination shown in Fig. 3.3. Here we can explicitly see that the bound for N is the Heisenberg limit for small α,

$$N \sim \frac{1}{\alpha} \tag{3.39}$$

□

Expanding the bound on N in Theorem 19 around $\alpha \ll 1$ gives us that the Heisenberg limit is indeed the optimal scaling, i.e., $N = \Omega(\alpha^{-1})$.

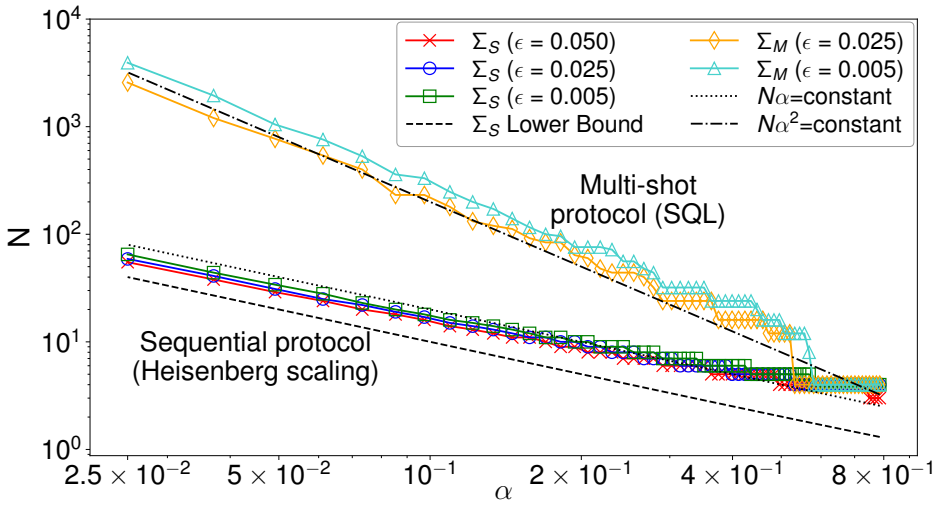


Figure 3.3: Number of queries N sufficient for solving $\mathsf{HBCD}(\alpha, \epsilon, \rho_h = \mathbb{I}/2)$. For the sequential protocol (Σ_S), we measure only once and use a phase sequence Φ of length N. For the multi-shot protocol (Σ_M), we use queries of depth $d = 4$ and measure multiple times. Trends for different values of $\epsilon \in \{0.005, 0.025, 0.05\}$ are shown for Σ_S and $\epsilon \in \{0.025, 0.005\}$ for Σ_M.

3.4.2 Numerical experiments

We now present numerical evidence that the Heisenberg limit is indeed achieved by the sequential protocol while the multi-shot protocol using constant depth queries only achieves the SQL. We solve the HBCD problem through measurements on $\mathcal{M}$ shown in Figure 3.2(b,c) using a maximally mixed state ($\rho_h = \frac{\mathbb{I}}{2}$) on $\mathcal{H}$. The state on $\mathcal{M}$ depends on the specified phase sequence Φ. If some Φ of length N sets the state of $\mathcal{M}$ to be $|1\rangle$ for $\theta_C = \alpha$ and $|0\rangle$ for $\theta_C = 0$, then $\mathsf{HBCD}(\alpha, \epsilon = 0, \rho_h)$ is solved with one shot.

For the sequential protocol, we attempt to solve HBCD by measuring once and with error probability $\epsilon \in [0, 1/2)$. The goal is to set the outcome y of measuring $\mathcal{M}$ in the computational basis such that

$$P_{y|\theta_C}(1|\alpha) - P_{y|\theta_C}(1|0) \geq 1 - 2\epsilon, \tag{3.40}$$

where we have used Eq. 3.1. To determine Φ that satisfies Eq. 3.40, we solve the following optimization problem

$$\arg\min_{\Phi} \left(1 - P_{y|\theta_C}(1|\alpha; \Phi) + P_{y|\theta_C}(1|0; \Phi)\right)^2, \tag{3.41}$$

with the additional constraint $\psi_n = \psi, \forall n \in [N]$. Details of the optimization is given in A.3. We claim that Φ succeeds in $\mathsf{HBCD}(\alpha, \epsilon, \rho_h)$ if the solution to Eq. (3.41) satisfies Eq. (3.40), i.e., $R(\Phi) \leq 4\epsilon^2$ where $R(\cdot)$ corresponds to the loss function defined inside Eq. 3.41. Given α, we determine the minimal number of queries required by starting with $N = 1$ and incrementing the value of N by one until the solution to Eq. 3.41 satisfies Eq. 3.40.

For the multi-shot protocol with constant depth-d queries, we use a phase sequence Φ of length d but may measure $\mathcal{M}$ $m \geq 1$ times to solve HBCD with error probability $\epsilon \in [0, 1/2)$.

For a given value of α, we determine Φ of length d by solving the optimization problem of Eq. 3.41. We determine $m^\star$ or the smallest number of shots required to achieve an error ϵ by evaluating Eq. 3.1, considering the estimator based on the likelihood-ratio test over the measurement outcomes A.3. The total number of queries required is then $N = d \cdot m^\star$.

In Fig. 3.3, we show the numerically determined trends of N required by the sequential and multi-shot protocols to solve $\mathsf{HBCD}(\alpha, \epsilon, \rho_h = \mathbb{I}/2)$. As expected, N increases as α decreases and approaches zero for both protocols. In particular, we observe a scaling of SQL for the multi-shot protocol but crucially a Heisenberg limited scaling $N \sim O(\alpha^{-1})$ for the sequential protocol.

3.5 Practical operation

We have commented so far on the performance of different discrimination protocols (Def. 6) in terms of the number of queries N required to solve an HBCD problem $\mathsf{HBCD}(\alpha, \epsilon, \rho_h)$ (Def. 5). In this section, we will comment on the performance of discrimination protocols in terms of detection probability under constraints on the total number of queries allowed.

In classical binary hypothesis testing [Cov05], the performance of any estimator (or decision rule) can be specified fully in terms of the detection probability P_D and the false-alarm probability P_F, defined as follows

$$P_D = \mathbb{P}\left(\hat{\theta}_C(\mathbf{y}) = \alpha \mid \theta_C = \alpha\right), \tag{3.42}$$

$$P_F = \mathbb{P}\left(\hat{\theta}_C(\mathbf{y}) = \alpha \mid \theta_C = 0\right). \tag{3.43}$$

For an estimator, it is desired to have a high value of P_D with a lower value of P_F. There may, of course, be additional criteria. For the sequential protocol, we want to achieve P_D higher than a certain value (say a_D) while keeping P_F below a certain threshold (say a_F) for the minimal length of the phase sequence Φ. In the multishot protocol with constant query depth d, we have the same goal but for the minimal number of shots m.

In Section 3.2.3, we noted that the estimator $\hat{\theta}_C(\cdot)$ of choice for both the sequential and multishot protocols is the likelihood ratio test (LRT, Eq. 3.5). We can rewrite this in the following form for a discrimination protocol Σ (Def. 6)

$$\hat{\theta}_C = \begin{cases} \alpha & , \frac{p_{\mathbf{y}|\theta_C}(\mathbf{y}|\alpha;\Sigma)}{p_{\mathbf{y}|\theta_C}(\mathbf{y}|0;\Sigma)} \geq \eta, \\ 0 & , \frac{p_{\mathbf{y}|\theta_C}(\mathbf{y}|\alpha;\Sigma)}{p_{\mathbf{y}|\theta_C}(\mathbf{y}|0;\Sigma)} < \eta, \end{cases} \tag{3.44}$$

where $\eta \in [0, \infty)$ is some threshold determined by the choice of prior probabilities and cost criterion. Formerly in Eq. 3.5, this threshold had the value of 1 as we considered the prior probabilities of θ_C to be uniform and the cost of making any decision to have equal risk. We note that specifying the value of η specifies the decision rule in Eq. 3.44. In fact, we can describe the decision regions in terms of the measurement outcomes $\mathbf{y} \in \{0,1\}^m$ as follows:

$$\mathcal{D}(\eta) = \{\mathbf{y} : p_{\mathbf{y}|\theta_C}(\mathbf{y}|\alpha;\Sigma) / p_{\mathbf{y}|\theta_C}(\mathbf{y}|0;\Sigma) \geq \eta\}, \tag{3.45}$$

$$\bar{\mathcal{D}}(\eta) = \{\mathbf{y} : p_{\mathbf{y}|\theta_C}(\mathbf{y}|\alpha;\Sigma) / p_{\mathbf{y}|\theta_C}(\mathbf{y}|0;\Sigma) < \eta\}, \tag{3.46}$$

where $\mathcal{D}$ denotes the set of measurement outcomes $\mathbf{y}$ for which the LRT makes the decision of $\hat{\theta}_C = \alpha$ and $\bar{\mathcal{D}}$ denotes the complement of $\mathcal{D}$ or the set of measurement outcomes for which the LRT makes the decision of $\hat{\theta}_C = 0$.

Moreover, it is known that the LRT maximizes the detection probability for a given upper bound on the false-alarm probability i.e., LRT is optimal under the Neyman-Pearson criterion [Hel94]. Each value of η can thus be associated with a particular (P_D, P_F) operating point. Varying the value of η varies the decision regions (Eq. 3.46) and thus allows us to analyze the trade-off between detection probability and false-alarm probability. The resulting curve of (P_D, P_F) points from varying $\eta \in [0, \infty)$ is called the operating characteristic.

We now adapt the operating characteristics to our quantum setting of HBCD in a similar fashion to that of quantum detector operating characteristics (QDOC) in [Med+19] and will also call them by the same name. For a given discrimination protocol $\Sigma(N, d, Z, S)$ (see Def. 6 for details on inputs) which includes specifying the type of protocol S and phase sequence Φ, we generate QDOC by varying the decision regions (Eq. 3.46) i.e., the value of η in LRT. Note that the phase sequence Φ and total number of queries N used are then fixed a priori. Let us now analyze QDOCs for the sequential and multi-shot protocols in solving HBCD for $\alpha = 0.1$ in different scenarios.

Sequential protocol. In Figure 3.4, we plot QDOCs for the sequential protocol in solving HBCD for $\alpha = 0.1$ for increasing values of N. We observe the QDOCs are monotonically increasing with P_F as expected. There is only one intermediate point of (P_D, P_F) in between $(0, 0)$ and $(1, 1)$ as we only measure once in the sequential protocol. This operating point corresponds to the value of $\eta = 1$. Values of operating points along the piece-wise linear segments can be obtained through randomization [Cov05]. In Figure 3.4, as the number of concatenated queries N is increased, the detection probability increases reaching $P_D > 0.95$ for $N = 16$. This is expected but somewhat surprisingly, we obtain this higher detection probability at a negligible increase in false-alarm probability.

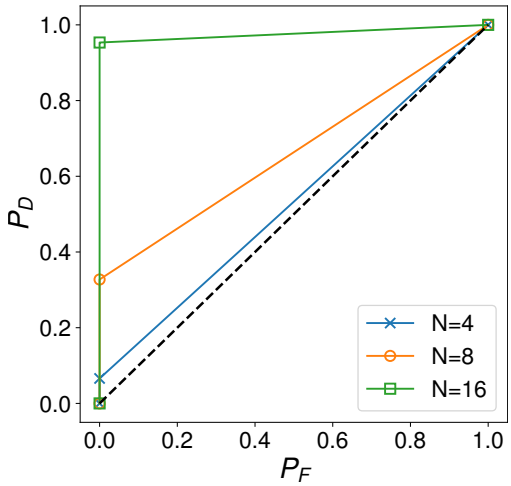


Figure 3.4: Quantum detector operating characteristics (QDOC) of the sequential protocol in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$ with increasing length N of phase sequence Φ.

Multi-shot protocol. In Figure 3.5, we plot QDOCs for the mutli-shot protocol with queries of constant depth $d = 8$ and increasing number of shots m. For higher values of m, there are more number of intermediate operating points corresponding to values of $\eta = \left(\frac{p_{y|\theta_C}(1|\alpha)}{p_{y|\theta_C}(1|0)}\right)^{Y_1} \left(\frac{p_{y|\theta_C}(0|\alpha)}{p_{y|\theta_C}(0|0)}\right)^{m-Y_1}$ with $Y_1 \in \{0, 1, \ldots, m\}$ denoting the number of measurement outcomes being one. Increasing the total number of queries N by increasing m allows higher detection probabilities to be achieved and yet again at negligible increase in false-alarm probability.

Fixed resource budget. We now consider the scenario where the total number of queries allowed to be used by any protocol is fixed to $N = 16$. In Figure 3.6, we compare the QDOC of the sequential protocol against various multi-shot protocols. We find that even under constraints of experimental resources, it is advantageous to use a sequential protocol to obtain a higher detection probability than any corresponding multi-shot protocol with same resource constraints.

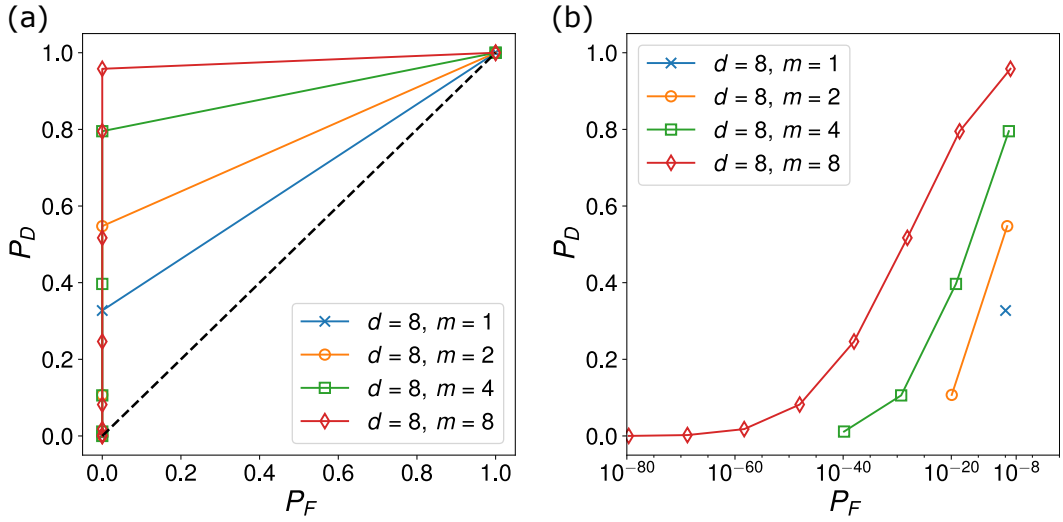


Figure 3.5: Quantum detector operating characteristics (QDOC) of the multi-shot protocol with a fixed query depth of $d = 8$ in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$ with increasing number of shots m. (a) Trend of P_D with P_F considering linear scales on both x-axis and y-axis. (b) Trend of P_D with P_F for intermediate operating points obtained by each protocol considering a log-scale on the x-axis to illustrate that P_F remains orders of magnitude below 1 for all values of P_D.

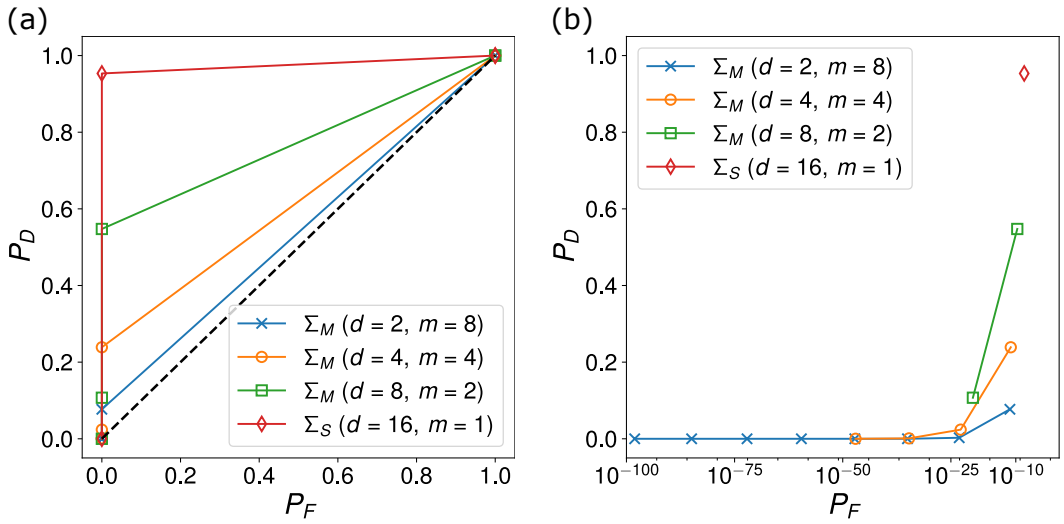


Figure 3.6: Quantum detector operating characteristics (QDOC) of a sequential protocol (Σ_S) and multi-shot protocol (Σ_M) with a fixed budget of $N = 16$ in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$. For the multi-shot protocol, we show OCs with increasing query depth d and decreasing number of shots m such that $N = d \cdot m = 16$. (a) Trend of P_D with P_F considering linear scales on both x-axis and y-axis. (b) Trend of P_D with P_F for intermediate operating points obtained by each protocol considering a log-scale on the x-axis to illustrate that P_F remains orders of magnitude below 1 for all values of P_D.

3.6 Conclusion

In this chapter, we proposed the HQCD problem and analyzed the performance of different protocols on HBCD. We showed sequential protocols outperform multi-shot and parallel protocols in HBCD. Notably, our work shows Heisenberg limited scaling (HLS) can be achieved with a sequential protocol in HQCD for a single-qubit channel. We expect these results to have interesting implications for other learning tasks on hidden quantum systems. Theoretically, one question is whether our results can be extended to problems of discriminating multi-qubit channels on hidden systems or learning multi-qubit channels with continuous parameters. If the hidden quantum channels correspond to unitary Hamiltonian evolution, could we learn the Hamiltonian [Hua+23; Dut+23a] at HLS? Our results suggest the guiding principle for measuring the properties of the hidden system requires transmitting information through interactions between the hidden and measurement systems when they are separated. The sequential protocol conveys information through repeated interactions and could achieve HLS. Conversely, the parallel protocol fails to estimate the channel because the initial state does not have entanglement across the two systems. We hope this interpretation could be proved for other learning problems on hidden channels.

Code and data availability: Code for the different discrimination protocols in solving HBCD numerically and data are available on GitHub [4].

[4] https://github.com/arkopaldutt/HiddenBCD

Chapter 4

Active learning of quantum system Hamiltonians

4.1 Introduction

Hamiltonian learning constitutes the problem of learning the Hamiltonian governing the dynamics of a quantum system given finite classical and quantum resources. This is a fundamental problem encountered in identification of quantum systems [Col+05; Col+06; BY12], operation of quantum information devices [Val+19; Wan+17], validation of theoretical physical models, and has implications for computational bounds on quantum algorithms [FG98; DM15; Liu+16]. In the calibration of quantum computers alone, it is a significant step in each of the following tasks: system characterization, learning device parameters, different sources of noise, gate design [Inn+20] and control strategies for implementing robust quantum gates with high fidelity. Moreover, a quantum computer typically requires frequent recalibrations to account for drift in parameters over time requiring multiple iterations of some Hamiltonian learning routine.

The resource requirements for learning a generic many-body Hamiltonian rise exponentially with the system size [MRL08]. Even for a fixed system size, however, the achievable learning error ϵ is fundamentally limited by the number of queries N made to the quantum system. In particular, through repeated system queries, N in general scales as ϵ^{-2} as a consequence of the central limit theorem. This is commonly referred to as the standard quantum limit (SQL) or shot noise limited scaling. Using quantum resources, however, a number of approaches have shown a much better *Heisenberg limited* scaling of $N \sim \epsilon^{-1}$. The Heisenberg limit is known to be fundamental [GLM04a; ZPK10; GLM11a; TA14; Pez+18b], under a wide range of assumptions [IF07; Ber+15; SBD16; Gór+20], but it is typically only saturated with the help of quantum resources.

This has been achieved using entanglement [GLM06] such as NOON states [Bol+96; LKD02], but can also be accomplished without entanglement under certain circumstances. For example in the problem of phase estimation [NC10], one has to estimate the phase ϕ in an unitary operator of form $U = \exp(-i\phi H)$ where H is a Hermitian operator and ϕ can also be interpreted as the strength of coupling in a Hamiltonian. It has been shown that Heisenberg limited scaling can be achieved using multi-round protocols using both adaptive

measurements [Hig+07b; Wis+09] and predetermined non-adaptive measurement sequences [Hig+09; KLY15]. In contrast, there has not been such a detailed study in the case of learning a general many-body or a multi-parametric Hamiltonian. This motivates understanding how might just classical resources be employed to solve the Hamiltonian learning problem with a scaling which surpasses the SQL, and ideally achieves the Heisenberg limit.

Even if scalings higher than SQL cannot be achieved, it is still desirable to minimize resource requirements. This may be accomplished by changing the estimation procedure used for Hamiltonian learning in combination with engineered experiments. Fast Fourier transform (FFT) and linear regression are some of the traditional estimation methods which still form the powerhouse of modern Hamiltonian learning strategies [Aru+20a]. It has been shown that adopting alternate estimation methods such as Bayesian estimation [EHF19], stochastic estimation [Kra+19], and neural-network based Hamiltonian reconstruction [Val+19] can reduce resource requirements and improve scalability. We will call the reduction in resource requirements achieved by replacing one Hamiltonian learning strategy with another as *query advantage*. This will obviously depend on the two strategies being compared. We will call the strategy being replaced as the baseline. The baseline and our proposed replacements will be discussed in detail later. This brings us to the primary question we tackle in this work: what is a common framework for Hamiltonian learning that can achieve query advantage even if scalings higher than SQL are not achieved, just using classical resources?

One effective framework for surpassing central limit theorem bounds when possible and/or achieving query advantage is *active learning*, e.g. using optimal experiment design. In [KWR04], this idea was explored for quantum state tomography, process tomography and Hamiltonian learning given a model, but in an offline manner. This has also been used to reduce experiment budget and propose different control schemes [GR02]. Active learning of a Hamiltonian is a more challenging problem than for quantum state tomography [Nun+10] where one optimizes over different measurements and process tomography [GNS19] where one additionally optimizes over initial states, due to the additional control parameter of system time evolution. Active learning thus provides a general framework for making adaptive queries to the quantum system, comprised of initial state, measurements and system time evolution during Hamiltonian learning. In fact, [YF15] has shown that with adaptive feedback control, Heisenberg limited scaling can be reached in principle, but a recipe for this is only given for estimation of a single Hamiltonian parameter and the procedure requires prior information of the parameter. This was later extended to multi-parameter Hamiltonians by [KU18]. A common ingredient of these works and the earlier mentioned multi-round protocols for phase estimation, is trading the cost of using physical quantum resources for cost in time resources [GLM06].

A sequential active learner for Hamiltonian learning based on the criteria of Bayes risk was proposed as part of Robust Online Hamiltonian learning (ROHL) in [Gra+12] and later Qinfer [Gra+17]. However, this active learner operates in a sequential manner, proposing a query to be made one shot at a time. This limits its applicability to current hardware where batching queries (or quantum circuits) is essential to overcome costs of typically expensive computation of risk functions for different queries being considered, access over the cloud, compilation, system queues, and latencies between classical electronics and the quantum hardware. It is then necessary to ensure any active learner we introduce for Hamiltonian learning operates in batch-mode.

In this work, we introduce a Hamiltonian active learning algorithm (HAL) operating in batch-mode based on the criteria of Fisher information, which is a way of measuring the information content of different queries and naturally appears in the bound on the errors achieved by a Hamiltonian learner. This resulting variant of HAL is called HAL-FI. We also introduce another variant of HAL for the task of predicting queries to the Hamiltonian, which uses the criteria of Fisher information ratio (FIR). We call the resulting active learning algorithm HAL-FIR. We demonstrate the performance of HAL-FI experimentally on IBM Quantum devices which are based on the superconducting cross-resonance (CR) gate. Compared with passive learning which scales as SQL, we show that HAL-FI with a fixed space of queries also has an asymptotic scaling of SQL but is able to achieve a constant reduction of more than 96.9% in number of queries required for a desired learning error in learning the two-qubit CR Hamiltonian on a 20-qubit IBM Quantum device. HAL-FI can achieve more than 33% reduction in number of queries required when compared to current standard methods used for Hamiltonian learning such as the sequential active learner of Qinfer. We finally show that queries involving exponentially growing system evolution time to the quantum devices suffices during learning to achieve Heisenberg limited scaling with HAL-FI when prior information is available. This is another example of trading physical quantum resources with time resources as highlighted before.

The chapter is organized as follows. In Sec 4.2, we formally describe the problem of Hamiltonian learning, and the concept of an active learner. In Sec. 4.3, we present the HAL algorithms of HAL-FI and HAL-FIR. To illustrate the performance of HAL-FI, we consider the example of calibrating CR gates on IBM Quantum devices. In Sec. 4.4, we describe our experimental setup, provide a theoretical description of the Hamiltonian model of the CR gate and physical models of the different noise sources affecting the quantum devices. Further, we provide details of our experiments on evaluating the performance of HAL-FI. Finally in Sec. 4.5, we compare the computational cost of different learners, describe the amount of query advantage that can be obtained using HAL-FI and specify the conditions under which Heisenberg limited rate of convergence or even super-Heisenberg limited rate of convergence can be achieved. Specifically for CR gates, we show that HAL-FI can be used to learn an accurate Hamiltonian using only a fraction of the queries required by currently used methodologies, resulting in reduction of queries of around two or three orders of magnitude over currently used methods for particular learning tasks.

4.2 Hamiltonian Learning

In this section, we present a description of the problem of Hamiltonian learning for a general quantum system (Section 4.2.1), and in the presence of different noise sources (Section 4.2.2). We will introduce the concept of an active learner in this context (Section 4.2.3). Notation used for this work will be defined as introduced.

4.2.1 Problem Statement

This section of the chapter describes the unknown Hamiltonian of interest, specifies our query setting, formally describes the different Hamiltonian learning tasks, and the estimators that

are used for Hamiltonian learning.

Unknown Hamiltonian

Let H be a partially or fully unknown system Hamiltonian over n qubits. We can represent H in terms of the n-qubit Pauli operators as $H = \sum_{P \in \{\sigma_I, \sigma_X, \sigma_Y, \sigma_Z\}^{\otimes n}} c_P P$ with $\{\sigma_I, \sigma_X, \sigma_Y, \sigma_Z\}$ representing the single-qubit Pauli operators and where $c_P \in \mathbb{R}$ are the corresponding coefficients. We could also represent H as a Hermitian matrix in $\mathbb{C}^{2^n \times 2^n}$ under the usual computational basis.

We assume that we have a model for the unknown Hamiltonian denoted by $H(\boldsymbol{\theta})$, parametrized by a real vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^m$ of length m. We use Θ to denote the set of all possible values over the Hamiltonian parameters. The model may be derived from first principles based on the understanding of the physics of the quantum system or proposed through empirical observations. We further assume that the system Hamiltonian is time-independent.

Let the unknown Hamiltonian of the quantum system be $H^\star$ and the true Hamiltonian parameters be $\boldsymbol{\theta}^\star$. We refer to the quantum system of whose Hamiltonian we wish to learn as an *oracle* that we can query and which returns measurement outcomes upon querying. We denote the random variable of query to the oracle by x and the resulting output by the oracle as the measurement outcome random variable y corresponding to a single shot of the qubits being readout in the standard computational basis. The pair of a query and its corresponding output is called an example and is denoted by (x, y). The alphabet of query x is referred to as the query space and we denote it by $\mathcal{Q}$. The distribution from which queries are sampled from $\mathcal{Q}$ is referred to as the *query distribution* and denoted by q. Commonly, $y \in \mathcal{Y} = \{0, 1\}^{n_r}$ where n_r is the number of qubits being readout. Our goal is to then learn the parameters $\boldsymbol{\theta}$ of the Hamiltonian with error ϵ from examples of the form $\{(x^{(i)}, y^{(i)})\}_{i=1}^N$ while minimizing the number of queries made to the oracle or the query complexity N.

Specifying a Query

We first describe what we mean by a query before specifying the different learning objectives. The query comprises three parts: measurement operators M, initial state preparation operators U, and control parameters t. A schematic of how the query is used in a quantum circuit to evolve a quantum system is shown in Figure 4.1. In Figure 4.2, we show the oracle of a quantum device receiving the input of queries of form $x = (M, U, t)$ and returning the corresponding outputs of a single shot of the qubits on the device being readout denoted by a binary string of length n_r: $y \in \{0, 1\}^{n_r}$.

Measurement Operators The measurement operators $M \in \mathcal{M}$ specify the admissible set of unitary operators that can be applied after evolving the system Hamiltonian and before measuring each qubit in the Z basis (as typical of current hardware). This in turn specifies the basis set $\{|\psi_m\rangle\}$ used to generate the final measurement observation results. The permissible set of M is typically constrained to be single-qubit operators or shallow circuits, because it is experimentally realistic to implement these on current hardware with

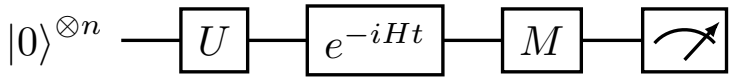


Figure 4.1: Quantum circuit picture of how the query $x = (M, U, t)$ is used. The input to the quantum circuit is the zero state $|0\rangle^{\otimes n}$. Application of the preparation operator U on $|0\rangle^{\otimes n}$ is used to create an initial state before interacting with the system Hamiltonian H for time t. Finally, a measurement operator M is applied to the evolved state before measuring in the usual computational basis denoted by the meter. The output is a single shot of qubits being readout.

high fidelity. For our applications, we will consider M to correspond to single-qubit operators corresponding to measuring under a Pauli basis.

Initial State Preparation Operators Together with the measurement operator, an initial state $|\psi_0\rangle$ must also be specified, in order to determine a measurement result from H. This initial state can in principle be any unit vector in the Hilbert space, but much like for measurement operators, we only allow a subset of possible states to be specified. We consider $|\psi_0\rangle = |0\rangle^{\otimes n}$ and allow for realistic single qubit unitary operators. A common initial state specification is to provide a unitary operator $U = \bigotimes_{i=1}^{n} U_i$ as a tensor product of single qubit operators such as single-qubit Pauli operators and Hadamard gates acting on each of the n qubits in the system. This then determines the U element of x. We denote the set of all considered preparation operators as $\mathcal{U}$

Control Parameters The control parameters are the last element of x, and are typically a set of classical numbers. For example, one canonical control parameter is the time t for which H should be applied. Other control parameters may modulate interactions between qubits. In this work, we consider only the Hamiltonian evolution time t as a control parameter and denote the set of all possible evolution times as $\mathcal{T}$.

After defining the set of measurement operators $\mathcal{M}$, the set of preparation operators $\mathcal{U}$, and the set of Hamiltonian evolution times $\mathcal{T}$, the resulting query space is given by $\mathcal{Q} = \mathcal{M} \times \mathcal{U} \times \mathcal{T}$. We require $\mathcal{Q}$ to be *complete*, i.e., there exist queries in $\mathcal{Q}$ that are informative about each of the Hamiltonian parameters $\boldsymbol{\theta}_i$ so that Hamiltonian learning can succeed. As we will see later in our discussion on active learning in Section 4.2.3, completeness of $\mathcal{Q}$ is equivalent to the condition of there existing a set of queries or query distribution for which the resulting Fisher information matrix is full rank and invertible.

Learning Framework

Having fully described a query, we are now in a position to formalize the problem of Hamiltonian learning. While doing so, we draw parallels to and introduce language from machine learning and statistical learning theory.

We are given a query space $\mathcal{Q} = \mathcal{M} \times \mathcal{U} \times \mathcal{T}$ constructed as discussed above and a space over the measurement outcomes $\mathcal{Y} = \{0, 1\}^{n_r}$. We consider the class of Hamiltonian models $\mathcal{H}$. A Hamiltonian $H \in \mathcal{H}$ defines a map from $\mathcal{Q}$ to the set of measurement outcomes taking

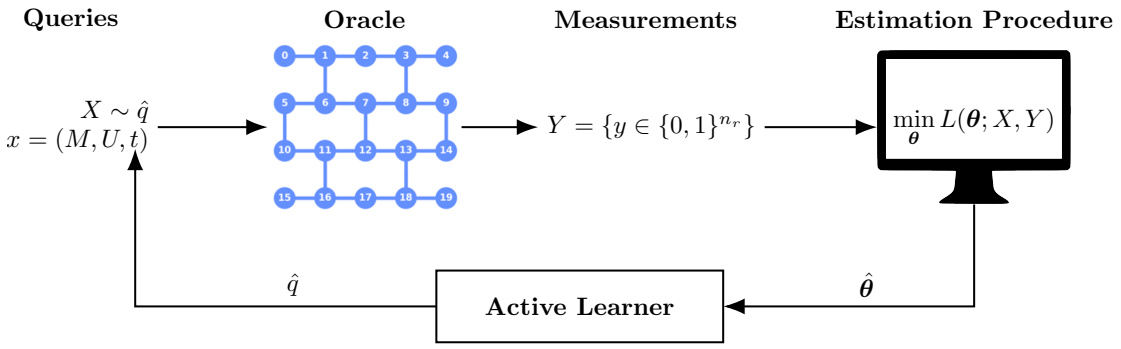


Figure 4.2: Schematic of Hamiltonian learning with an active learner: The *oracle* constitutes the unknown Hamiltonian and noise sources, with the true parameter vector of $\boldsymbol{\theta}^\star$ which is unknown and needs to be learned. Here, we show the device coupling map of the 20-qubit IBM Quantum device *ibmq_ boeblingen*. Learning is carried out on training examples of the form (x, y) where x are queries inputted to the oracle specifying the measurement observable M, preparation operator U and system time evolution t, and y are the corresponding measurement outcomes outputted by the oracle. An illustrative quantum circuit picture of the oracle in the noiseless case is shown in Figure 4.1. The set of queries inputted to the oracle is denoted as X and the corresponding set of measurement outcomes outputted by the oracle as Y. An estimation procedure is run on the training examples of (X, Y) to learn a model parameter estimate $\hat{\boldsymbol{\theta}}$. The complete top row corresponds to how a passive learner operates and is also called open-loop Hamiltonian learning. We add a feedback loop to introduce an active learner which uses the current estimate of model parameters $\hat{\boldsymbol{\theta}}$ to prescribe the distribution $\hat{q}$ from which queries to the oracle should be sampled from next. This process is then repeated during learning until an accurate estimate $\hat{\boldsymbol{\theta}}$ is obtained. The Hamiltonian Active Learning (HAL) algorithm introduced in this work comprises the estimation procedure and active learner shown in this schematic. The active learner is described in Section 4.2.3 and the HAL algorithm in Section 4.3.

values in $\mathcal{Y}$ by Born's rule. In particular we assume the availability of a model description $\mathcal{H} = \{H(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \subset \mathbb{R}^m$ parametrized by the vector $\boldsymbol{\theta}$ and where Θ is considered to be the space over the parameters. When no prior information is available to constrain the parameter space, we can consider $\Theta = \mathbb{R}^m$. We denote the true Hamiltonian as $H^\star$ and assume that $H^\star \in \mathcal{H}$ i.e., it is realizable. We denote the parameter vector of the true Hamiltonian as $\boldsymbol{\theta}^\star$. Our goal is to learn an estimate of $\boldsymbol{\theta}^\star$ which we denote by $\hat{\boldsymbol{\theta}}$. With regards to notation, we use $\boldsymbol{\theta}$ whenever we make statements that are true for any parameter in Θ.

For the parameter vector $\boldsymbol{\theta}$, the label (or measurement outcome) y given query $x = (M, U, t)$ is produced with the conditional probability

$$p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta}) = \sum_z \left| \langle yz | M e^{-iH(\boldsymbol{\theta})t} U | 0^{\otimes n} \rangle \right|^2 \tag{4.1}$$

as per Born's rule and assuming the absence of noise. The summation is over the hidden measurement outcomes of the $(n - n_r)$ qubits that are not read out which we denote by z and $|\cdot|$ is used to denote the absolute value.

Given the dataset of N training examples $D = \{(x^{(i)}, y^{(i)})\}_{i \in [N]}$, the goal is to learn the

Hamiltonian parameter vector $\boldsymbol{\theta}$ which characterizes the conditional probability (Eq. 4.1) over measurement outcomes in $\mathcal{Y}$ given queries in $\mathcal{Q}$. This is a supervised learning task for which a suitable loss function is

$$\text{Loss}(\boldsymbol{\theta}; D) = \mathcal{L}\left(D; \boldsymbol{\theta}\right) + R(\boldsymbol{\theta}) \tag{4.2}$$

where $\mathcal{L}$ is a suitable error function for assessing the chosen model on the training dataset, and R is a regularization function penalizing model complexity. The latter is chosen to incorporate prior information, enforce conditions such as sparsity through ℓ_1-norm and generalize to unseen data. The parameter estimate is determined by minimizing the loss function $\hat{\boldsymbol{\theta}} = \text{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta}; D)$. After obtaining an estimate $\hat{\boldsymbol{\theta}}$, it is common to evaluate the performance against a testing dataset which includes examples not seen during training. This testing dataset which we denote as $D_{\text{test}} = \{(x^{(i)}, y^{(i)})\}_{i\in[N_{\text{test}}]}$ contains N_{test} queries and their corresponding outputs. The N_{test} queries are sampled from the query space $\mathcal{Q}$ according to a testing distribution which we denote by p_{test}. This allows us to further distinguish between the problems of *model inference* and *prediction* against p_{test}.

Model inference The goal is to learn a parameter estimate $\hat{\boldsymbol{\theta}}$ within a prescribed learning error ϵ with the fewest queries N, without making special considerations for the testing distribution p_{test} which may change or may be unknown. We use root mean squared error (RMSE) as $\mathcal{L}$ in Eq. 4.2

$$\text{RMSE}(\hat{\boldsymbol{\theta}}; \boldsymbol{\xi}) = \left(\sum_{i=1}^{m} \mathbb{E}\left[\left(\frac{\hat{\theta}_i}{\xi_i} - \frac{\theta_i^\star}{\xi_i}\right)^2\right]\right)^{1/2} \tag{4.3}$$

where $\boldsymbol{\theta}^\star$ is the parameter vector corresponding to the unknown Hamiltonian $H^\star$, and $\boldsymbol{\xi}$ is the vector to non-dimensionalize and normalize $\boldsymbol{\theta}$. This is done to account for different relative magnitudes of each parameter component $\theta_i^\star$ and to ensure that the RMSE does not explode due to contributions of few parameter components. These normalization factors $\boldsymbol{\xi}$ can be obtained during estimation or be available through prior information. When an unbiased estimator is used, $\mathcal{L}$ may be reduced to being the square root of the variance, $\left[\sum_j \text{Var}(\theta_j/\xi_j)\right]^{1/2}$. It is typically hard to ensure that an estimator is unbiased, even though this is desired. To find $\hat{\boldsymbol{\theta}}$ that minimizes the RMSE given D, we use estimation methods based on the asymptotically unbiased maximum likelihood estimator (MLE) and Bayesian estimation, all of which we will discuss in Section 4.2.1.

Prediction against a testing distribution p_{test} The goal is to learn an estimate $\hat{\boldsymbol{\theta}}$ that will allow us to perform well on predicting the likelihood function of measurement outcomes given queries sampled from p_{test}. In such a scenario, $\mathcal{L}$ is still the RMSE but the performance of the estimate is assessed through the testing error:

$$\text{Testing Error} = \mathbb{E}_{x\sim p_{\text{test}}, y\sim p_{\text{y|x}}(\cdot|x;\boldsymbol{\theta}^\star)}[L(y|x;\hat{\boldsymbol{\theta}}) - L(y|x;\boldsymbol{\theta}^\star)], \tag{4.4}$$

where we have denoted the negative log-likelihood $-\log p_{\text{y|x}}\left(y|x;\boldsymbol{\theta}\right)$ in shorthand by $L(y|x;\boldsymbol{\theta})$ for a general measurement outcome y and query x given model parameters $\boldsymbol{\theta}$. We have chosen

the testing error as the expectation of the difference in negative log-likelihood when using the estimate $\boldsymbol{\theta}$ and truth $\boldsymbol{\theta}^\star$ with respect to the testing distribution p_{test}. This difference in negative log-likelihood is also often called the log-likelihood ratio [Sou+17]. It is then desired to learn $\hat{\boldsymbol{\theta}}$ such that the testing error is lower than a given error ϵ. The error ϵ here may need to be specified relative to how the negative log-likelihood scales which will depend on the Hamiltonian of interest and the queries themselves.

Given N training examples, we then say that we have succeeded at model inference (prediction) if we are able to produce an estimate $\hat{\boldsymbol{\theta}}$ such that the RMSE (testing error) is bounded by some error parameter ϵ. We want to accomplish these learning tasks by minimizing the resource requirements or number of queries N.

We distinguish between the two problems of model inference and prediction against p_{test} as the ideal training datasets for these two learning tasks may come from two different distributions. This will influence how the training data is chosen when we introduce an active learner into Hamiltonian learning, as we will discuss in Section 4.3. The performance of the Hamiltonian learning algorithms in tackling the learning problem 4.2.1 will also be useful for cross-validation and testing the robustness of these methods.

Estimators

We now discuss the different types of estimators that may be adopted for learning an estimate $\boldsymbol{\theta}$ of the Hamiltonian parameters from the N measurement outcomes $Y = \{y^{(i)}\}_{i\in[N]}$ obtained from the quantum system upon N queries $X = \{x^{(i)}\}_{i\in[N]}$. Collectively, we refer to them as the training data $D = (X, Y) = \{(x^{(i)}, y^{(i)})\}_{i\in[N]}$. The active learner for Hamiltonian learning we introduce in this work will be designed for use with the following maximum likelihood estimator (MLE) but is also compatible with the other estimators mentioned here.

Maximum Likelihood Estimation. The maximum likelihood estimator (MLE) determines an estimate of the parameters $\hat{\boldsymbol{\theta}}$ from the data D through

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} -\log p_{\text{y|x}}\left(y^{(i)}|x^{(i)};\boldsymbol{\theta}\right), \tag{4.5}$$

where $-\log p_{\text{y|x}}\left(y^{(i)}|x^{(i)};\boldsymbol{\theta}\right)$ is the negative log-likelihood of the measurement outcome $y^{(i)}$ given the query $x^{(i)}$ and when considering the model parameters $\boldsymbol{\theta}$. The optimization problem of Eq. 4.5 is typically non-convex and thus specific procedures may be required for avoiding local minima and converging to the global minimum.

Regression. Estimation based on regression is used as a proxy for the MLE or sometimes even as part of the first step in a procedure to solve the optimization of MLE (Eq. 4.5). Let the empirical likelihood computed from measurement outcomes corresponding to query x be given by $\hat{p}_{y|x}$. Suppose we have a model $l(x;\boldsymbol{\theta})$ (e.g, Gaussian or sinusoidal) of the likelihood function for query x given parameters $\boldsymbol{\theta}$. We can then obtain an estimate by performing least squares:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\hat{p}_{Y|X} - l(X;\boldsymbol{\theta})||_2^2, \tag{4.6}$$

where we have denoted $\hat{p}_{Y|X}$ as the empirical likelihood computed for different queries in queries X using corresponding measurement outcomes in Y from D. Note that Eq. 4.6 could be further improved by weighting the empirical likelihoods by their variances due to shot noise.

Bayesian Estimation. Instead of the frequentist approach above, we can also adopt a Bayesian approach during estimation. The Hamiltonian parameters $\boldsymbol{\theta}$ are treated as a random vector whose distribution is updated according to Baye's rule:

$$p(\boldsymbol{\theta}|D) = \frac{p(Y|\boldsymbol{\theta}, X)p(\boldsymbol{\theta})}{p(Y|X)}, \tag{4.7}$$

where $p(Y|\boldsymbol{\theta}, X)$ is the likelihood function over measurement outcomes Y given queries X and parameter vector $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$ and $p(\boldsymbol{\theta}|D)$ is the posterior distribution after incorporation of training data $D = (X, Y)$. Often, we will use the Baye's rule iteratively and in that case the posterior from one step becomes the prior for the next step. When a single point parameter estimate is desired, we will use the posterior mean $\hat{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}|D}[\boldsymbol{\theta}]$ which is optimal for minimizing RMSE [KLY15].

Computing the denominator in Eq. 4.7 and the mean over the posterior distribution is computationally expensive and usually intractable in practice. This computational burden is reduced if structural information about the distributions in Eq. 4.7 are known (e.g., Gaussian) or can be reliably enforced. This is done in Bayesian estimation methods such as the Kalman filter [Jaz07] and Gaussian mixture model based filter [AS72]. For our applications, we will consider sequential Monte Carlo methods (also called particle filtering) [DJ+09].

We note that Eq. 4.5 is an optimization problem that may be solved in several ways, and similarly there may be different strategies to evaluate Eq. 4.7 in Bayesian estimation. The different programs or algorithms to solve the optimization problem (or Baye's rule) are commonly called *optimization procedures* or *estimation procedures*. Different procedures have different properties such as rate of convergence and optimality such as guarantee of a local or global minimum. In order to improve convergence or search direction, it is common for procedures to require first or second gradient information. The estimation procedure will need to be specified along with the formulated optimization problem at hand for a complete specification of the learning algorithm or learner.

4.2.2 Learning in the Presence of Noise

Previously, in Section 4.2.1, we gave formal statements of the Hamiltonian learning problem in the absence of noise assuming an ideal oracle. However, the oracle is usually noisy due to the presence of different noise sources affecting the quantum system. In this section, we describe the problem of learning in the presence of these noise sources.

Common noise sources include readout noise, decoherence, and imperfect control of the quantum system. State preparation and measurement (SPAM) errors are already accounted for by considering these noise sources. Classical SPAM errors encountered are included in the readout noise and the errors in implementing state preparation or measurement operators fall under imperfect control.

Readout Noise The readout line of the qubit measurement outcomes y is also a classical communication channel and hence suffers from bit-flip errors. The readout noise can then be modeled as a classical bit flip channel where the true measurement outcome $y_i \in \{0, 1\}$ of the ith qubit may be flipped. We denote the observed measurement outcomes' random variable by $\tilde{\text{y}}$ which can be interpreted as a noisy observation of y. We denote the conditional probability of observing $\tilde{y}$ given measurement outcome y which is hidden from us as $p_{\tilde{\text{y}}|\text{y}}(\tilde{y}|y)$. Note that in general we expect the readout channel to be asymmetric i.e., $p_{\tilde{\text{y}}|\text{y}}(1|0) \neq p_{\tilde{\text{y}}|\text{y}}(0|1)$, and this is accounted for in the readout noise model. The probability of observing a noisy measurement outcome $\tilde{y}_i$ of the ith qubit given query x is then given by

$$p_{\tilde{\text{y}}_i|\text{x}}(\tilde{y}_i|x;\boldsymbol{\theta}) = p_{\tilde{\text{y}}_i|\text{y}}(\tilde{y}_i|\tilde{y}_i)p_{\text{y}_i|\text{x}}(\tilde{y}_i|x;\boldsymbol{\theta}) + p_{\tilde{\text{y}}_i|\text{y}}(\tilde{y}_i|1-\tilde{y}_i)p_{\text{y}_i|\text{x}}(1-\tilde{y}_i|x;\boldsymbol{\theta}) \tag{4.8}$$

where $p_{\text{y}|\text{x}}(y|x;\boldsymbol{\theta})$ is given by Eq. 4.1.

Imperfect Control Strategies In general, a Hamiltonian can be decomposed as

$$H(t) = H_d + \sum_{k=1}^{K} H_k(t, u_k(t)) \tag{4.9}$$

where H_d is the drift/free part of the Hamiltonian that is internal to the system and cannot be controlled externally. The terms H_k corresponds to the parts of the Hamiltonian that can be controlled using the control function $u_k(t)$. Depending on the quantum architecture, the control function can be realized as microwave pulses for superconducting qubits, optical pulses for ion traps, etc.

Target operators U_f are then obtained starting from the identity matrix I by implementing controls $u_k(t)$ over time duration $[0, T]$ as

$$U_f = \mathbb{T}\left[\exp\left(-i\int_0^T \left[H_d + \sum_{k=1}^{K} H_k(t, u_k(t))\right] dt\right)\right] \tag{4.10}$$

where $\mathbb{T}$ is the time-ordering operator. These target operators U_f can be preparation operators, measurement operators or different gates e.g., the Clifford gates. Imperfections in these pulses will lead to errors in U_f. For example, the strength of the qubit-qubit interactions is sensitive to variations in the pulse amplitudes. Strong driving can lead to leakage of states outside of the computational subspace. Moreover, bandwidth effects or dispersion can cause leading or trailing edge distortions in the pulse shapes that can lead to errors in the unitary operator implemented.

Decoherence Let us consider the unitary operator of $U(t) = e^{-iHt}$. The application of this unitary operator is accompanied by decoherence on a real quantum system due to interactions with its environment. We model this as a depolarizing channel $\mathcal{E}$

$$\mathcal{E}(\rho(t)) = (1 - p_d(t))\rho(t) + p_d(t)\frac{I}{2^n} \tag{4.11}$$

where $\rho(t) = U(t)\rho(0)U(t)^\dagger$ is the state obtained on application of the unitary $U(t)$ to the initial state $\rho(0)$, $p_d(t)$ is the probability of the state being depolarized and $I/2^n$ is the

maximally mixed state. We have assumed the case of complete depolarization here. To obtain a functional form of $p_d(t)$, we note that up to first order, depolarization events can be assumed to arrive under a Poisson process with rate μ. By a depolarization event, we refer to the occurrence of an error that completely randomizes the quantum state. As the time between Poisson events follows an exponential distribution, we can write

$$p_d(t) = 1 - \exp\left(-\frac{t - t_0}{\mu}\right) \tag{4.12}$$

where t_0 denotes the starting time of the experiment. Denoting the probability of measurement outcomes y given a query x under depolarization errors as $p_{\mathcal{E}}(y|x;\boldsymbol{\theta})$, we have

$$p_{\mathcal{E}}(y|x;\boldsymbol{\theta}) = \exp\left(-\frac{t - t_0}{\mu}\right) p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta}) + \frac{1}{2^n}\left(1 - \exp\left(-\frac{t - t_0}{\mu}\right)\right) \tag{4.13}$$

where $p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})$ is the probability of measurement outcome y assuming no depolarization given by Eq. 4.1. The rate μ can be related to the amplitude relaxation time T_1 and dephasing time T_2 of the qubits on the quantum system. We will describe this in Section 4.4.3 for the IBM Quantum devices considered for application of Hamiltonian learning.

Let the collective set of parameters associated with the different noise models so far described be denoted by $\boldsymbol{\zeta}$. The measurement outcomes y from the oracle will then be a function of the queries x and the true parameters $(\boldsymbol{\theta}^\star, \boldsymbol{\zeta}^\star)$. In order to obtain estimates for $\hat{\zeta}$ in addition to $\hat{\boldsymbol{\theta}}$, we solve the following modified optimization problem over the training examples

$$\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\zeta}}\right) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta},\boldsymbol{\zeta}\in\mathrm{Z}}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\left(x^{(i)}, y^{(i)}; \boldsymbol{\theta}, \boldsymbol{\zeta}\right) + R(\boldsymbol{\theta}, \boldsymbol{\zeta}) \tag{4.14}$$

However, this increases the computational cost of the estimation procedure due to the increase in the number of parameters and the corresponding search space. In practice, prior calibration data can be used to obtain estimates of $\hat{\boldsymbol{\zeta}}$ which are then used in Eq. 4.2 to obtain $\hat{\boldsymbol{\theta}}$.

Hence, the Hamiltonian learning Problems 4.2.1 and 4.2.1 can be restated assuming additional information of the noise model parameters $\hat{\boldsymbol{\zeta}}$ when we have access to a noisy oracle. When describing our application of Hamiltonian learning, we will describe both the Hamiltonian of interest in Section 4.4.1 and the specific noise sources in Section 4.4.3.

4.2.3 Active Learning of Hamiltonians

This subsection introduces the concept of active learning and how an active learner can be used in the context of Hamiltonian learning. We begin by giving a quick overview of different types of learners, and describe how active learning differs from passive learning and online learning. This is followed by an overview of the different active learning (AL) strategies that have been proposed in the literature. We describe which AL strategy seems best suited for Hamiltonian learning and particularly when using the maximum likelihood estimator (MLE) (Eq. 4.2.1). AL strategies based on Fisher information (FI) for Problem 4.2.1 and Fisher information ratio (FIR) for Problem 4.2.1 seem to be notably appropriate. In doing so, we

establish criteria for evaluating the performance of an AL algorithm, and observe that AL algorithms are known to solve some learning problems faster than passive learners. Moreover, under some circumstances, AL algorithms can exceed the central limit theorem bounds.

Types of Learners

An overview of different learners often considered for learning tasks is given below. We will use the passive learner as a baseline and the active learner is the focus of our work. A description of an *offline learner* and *online learner* is given for completeness and to distinguish an active learner from them.

Offline Learner In offline learning, all of the training data is given to the learner at once and a model is learned. The training dataset may be made available to the learner or collected by sampling queries from an arbitrary distribution and then obtaining outputs from an oracle using these as inputs. This is the learning paradigm under which most machine learning tasks operate in.

Online Learner In online learning, the training data is made available in a sequence, typically one at a time by a referee. A query x_t is made available to the learner at the t'th round in a sequence after which the learner constructs an estimate of the output $\hat{y}_t$ to this query. The learner provides $\hat{y}_t$ to the referee, and suffers a loss that depends on $\hat{y}_t$ and the actual output. The learner is then provided with feedback by the referee which the learner can then use to update the model. In this case, the queries that are provided to the learner by the referee may be adversarial or adaptive to the learner's behavior. The learner has no control over the query distribution from which these queries arrive.

Passive Learner In the learning problems described in Sec. 4.2.1, we did not specify the distribution from which the queries x are sampled from. In open-loop Hamiltonian learning, the query distribution remains fixed during learning and all the queries to the oracle are sampled from this distribution. When no prior information is available, it is common to set the query distribution to the uniform distribution over the query space $\mathcal{Q}$. We will refer to this setting as *passive learning* through out this chapter. Combined with a specification of an estimation procedure, the passive learner will serve as a baseline to the active learner which we introduce in the next section.

Active Learner In active learning (AL), the learner has access to the query space $\mathcal{Q}$ and the ability to select queries or decide the query distribution during training using the current estimates of the model parameters $\hat{\boldsymbol{\theta}}$. This is accomplished by introducing a feedback into the open loop Hamiltonian learning approach shown earlier in Figure 4.2. Based on the current estimate $\hat{\boldsymbol{\theta}}$ and the queries made so far combined with their respective outcomes, the active learner proposes a query distribution from which queries should be selected from to send to the oracle. These queries may be sent to the oracle in a sequential manner one at a time or in batches. In Section 4.2.3, we discuss different criteria used for query selection or proposing query distributions.

Here, we distinguish a passive learner from an offline learner only in the number of rounds of training data collection. An offline learner is given access to a complete training dataset or is allowed to collect it through queries to an oracle in one round. On the other hand, a passive learner has continued access to the oracle and is allowed to collect training data until the learning task has been accomplished. The primary difference between the active learner and the online learner is that the former has control over the query distribution from which it will select queries to input to the oracle.

Query Criteria for Active Learner

There have been multiple criteria proposed for query selection but are usually subdivided into the two categories of informativeness and representativeness. Criteria based on informativeness aim to select queries that will reduce the uncertainty of the statistical model and include uncertainty sampling [LG94; Set09], query-by-committee [SOS92; BRK07], and margin [BBZ07]. On the other hand, the goal of representativeness [YBT06; FAJ06] is to ensure selection of queries that exploit the structure of the underlying distribution and are diverse. There has also been exploration into combining the criteria of informativeness and representativeness [Du+15; HJZ14].

Multiple query criteria used in active learning in practice are based on heuristics and empirical evidence [Set12]. Here, we choose the informativeness criteria of Fisher information (FI) and Fisher information ratio (FIR) as they have direct relationships with the different learning problems we introduced in Sec. 4.2.1. This allows us to provide guarantees on the performance of our AL strategy. Later, we discuss how we can ensure that representative queries are also selected.

We introduce some notation before discussing the query criteria for our AL strategy. Let us denote the Fisher Information matrix of a particular query $x \in \mathcal{Q}$ as $\mathcal{I}_x(\boldsymbol{\theta})$ where the (i,j)th element of the matrix is given by

$$\mathcal{I}_x(\boldsymbol{\theta})[i,j] = \mathbb{E}\left[\frac{\partial \log p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \log p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})}{\partial \theta_j}\right] \tag{4.15}$$

and where the expectation is taken with respect to $p(y|x;\boldsymbol{\theta})$. The Fisher Information matrix is equivalently written as $\mathcal{I}_x(\boldsymbol{\theta}) = \mathbb{E}[SS^T]$ where $S = \partial \log p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is commonly called the score vector. Instead of selecting one query at a time, we often require the active learner to select a distribution over the query space $\mathcal{Q}$ which we will call the query distribution. The Fisher Information matrix corresponding to q will then be given by

$$\mathcal{I}_q(\boldsymbol{\theta}) = \mathbb{E}_q[\mathcal{I}_x(\boldsymbol{\theta})] = \sum_{x\in\mathcal{Q}} q(x)\mathcal{I}_x(\boldsymbol{\theta}) \tag{4.16}$$

where the summation can be replaced by an integral in the case of a continuous query space. If the parameters describing the Hamiltonian model are $\boldsymbol{\theta} \in \mathbb{R}^m$, then $\mathcal{I}_q(\boldsymbol{\theta}) \in \mathbb{R}^{m\times m}$. Let us now describe the different query criteria in the context of the different learning tasks.

RMSE of Parameters If the learning objective is to learn the parameters with small RMSE (Problem 4.2.1), a natural query optimization strategy is obtained by noting the

Cramer-Rao bound for unbiased estimators: [CT06]:

$$\mathrm{Cov}(\boldsymbol{\theta}) \geq \frac{1}{N}\mathcal{I}_q^{-1}(\boldsymbol{\theta}) \tag{4.17}$$

$$\sum_i \mathrm{Var}(\theta_i) \geq \frac{1}{N}\mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta})) \tag{4.18}$$

Combining the Cramer-Rao bound with the fact that the parameter estimates converges in probability $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}^\star$ for an unbiased estimator such as MLE and the distribution converges in law as $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) \to \mathcal{N}\left(0, \mathcal{I}_q^{-1}(\boldsymbol{\theta}^\star)\right)$ [LC06; Was13], we get that MLE is an asymptotically efficient estimator with the efficiency equal to the Fisher information over the training examples [Sou+17]. An optimal query distribution can then be obtained through the following query optimization:

$$q^\star = \arg\min_{q \in \mathcal{P}} \mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta}^\star)) \tag{4.19}$$

where $\mathcal{P}$ is the family of all valid probability distributions over the query space $\mathcal{Q}$. We also note that for $q^\star$, $\mathcal{I}_{q^\star}$ must necessarily be invertible and hence full rank. This ensures that the queries inform us about all the Hamiltonian parameters of interest. Also for an unbiased and consistent estimator, the Cramer-Rao bound is likely to be saturated in the limit of large number of queries.

Testing accuracy against testing distribution When the learning objective is to minimize the expected log-likelihood error against a testing distribution p_{test} (learning problem 4.2.1), we use an active learning strategy based on Fisher Information ratio (FIR) $\mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta})\mathcal{I}_{p_{\text{test}}}(\boldsymbol{\theta}))$. The name FIR comes from the scalar case where it can be viewed a ratio of the Fisher information corresponding to the query distribution and that corresponding to the testing distribution. The use of FIR for this learning task can be motivated by noting the following inequality [Sou+17]

$$\mathbb{E}_{x\sim p_{\text{test}}, y\sim p_{\text{y}|\text{x}}(\cdot|x;\boldsymbol{\theta})}[\mathrm{Var}_{\text{D}\sim q(x)p_{\text{y}|\text{x}}(y|x;\boldsymbol{\theta})}[\mathcal{L}(\hat{\boldsymbol{\theta}}_{\text{D}}; x, y) - \mathcal{L}(\boldsymbol{\theta}; x, y)]] \leq \frac{1}{N}\mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta})\mathcal{I}_{p_{\text{test}}}(\boldsymbol{\theta})), \tag{4.20}$$

where the left hand side is the expected variance of the asymptotic distribution of the log-likelihood ratio which can be viewed as a testing error and the right hand side involves the FIR. Minimizing the upper bound would then allow us to control the testing error and hence this suggests using the following query distribution:

$$q^\star = \arg\min_{q \in \mathcal{P}} \mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta}^\star)\mathcal{I}_{p_{\text{test}}}(\boldsymbol{\theta}^\star)). \tag{4.21}$$

We note the Fisher information ratio is related to the Fisher information matrices of the query distribution $\mathcal{I}_q$ and testing distribution $\mathcal{I}_{p_{\text{test}}}$ through the following inequality:

$$\mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta})\mathcal{I}_p(\boldsymbol{\theta})) \leq \mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta})) \cdot \mathrm{Tr}(\mathcal{I}_{p_{\text{test}}}(\boldsymbol{\theta})). \tag{4.22}$$

Thus, we recover the query optimization of Eq. 4.19 when the testing distribution is unknown.

In quantum tomography, such query criteria have been applied in optimal experiment design (OED) or adaptive quantum tomography. Fisher information has been used as a query

criteria for offline OED in quantum state tomography [Nun+10]. Even earlier, an active learner based on Shannon entropy aka maximum uncertainty sampling was considered in [FKF00]. An AL strategy based on Shannon information combined with Bayesian estimation was proposed in [HH12] for the selection of measurement operators during quantum state tomography. Fisher information was again used in [GNS19] where OEDs were analyzed for a family of qubit channels over different design problems. In Hamiltonian learning, Fisher information has been used to comment on heuristic strategies for OED [FGC13] and has been combined with Bayesian estimation to produce a sequential active learner [Gra+12].

Active Learning Strategy

Implicit in the above descriptions of the query criteria is the idea of proposing a query distribution rather than selecting one query at a time during active learning. This demarcates sequential active learning where one query is chosen at a time from batch mode active learning where a batch of queries sampled from a query distribution are selected to be inputted to the oracle. Moreover, combining FI/FIR query criteria with batch mode active learning [Hoi+06] ensures that representative queries are chosen as well.

In what follows, we describe the batch-mode active learning scheme that forms the basis of the AL algorithm for Hamiltonian learning that we discuss in Section 4.3. Given a budget of N queries, the training is divided into multiple rounds. We index each round of the training process as i and denote the batch size as N_b. The number of queries made till the ith round (inclusive) is denoted as $N_{tot}^{(i)}$. In each round, a batch of queries is sampled from the optimal query distribution based on the current parameters' estimate $\hat{\boldsymbol{\theta}}$ and then this estimate is updated using the measurement outcomes of the queries. This is then repeated until all of the budget has been expended. We denote the estimate in the ith round by $\hat{\boldsymbol{\theta}}^{(i)}$ and the optimal query distribution in the ith round based on $\hat{\boldsymbol{\theta}}^{(i)}$ by $q^{(i)}$. The query distribution at the very beginning of the training process $q^{(0)}$ is determined from any available prior information of the parameters or else set to be the uniform distribution over the query space.

What should be the size of the initial set of queries $N_{tot}^{(0)}$? Some suggestions are given in [Cha+15] based on a finite-sample analysis for logistic regression but these do not suffice for the application considered in this chapter. Qualitatively, one hopes that $N_{tot}^{(0)}$ is high enough such that the parameter estimate $\hat{\boldsymbol{\theta}}^{(0)}$ lies close to the true parameter value $\boldsymbol{\theta}^{\star}$ and in a convex basin of the asymptotic negative log-likelihood loss function. However, setting $N_{tot}^{(0)}$ to a very high value may not allow us take advantage of the presence of an active learner and the savings it can provide.

Additionally, it may be advantageous to adaptively change the query space for exploration from one batch to the next. It is not necessary for the query space to remain static or unchanged [Set12] during active learning. In fact, there is an element of adaptively changing the *search* space in many prominent algorithms. In sparse fast Fourier transform [Has+12a; Has+12b], the bins of frequencies are randomly chosen with each iteration in the algorithm. In the related machine learning tool of reinforcement learning, action spaces are changed to eliminate actions [Zah+18], and to generalize over time by parameterizing them [MRK16] or embedding them in a continuous action space [Dul+15].

Adaptively changing the query space is particularly compelling for the application of

active learning to the Hamiltonian learning problem, because evidence suggests that it may result in so-called *Heisenberg*-limited scaling of the number of queries as we noted in Section 4.1. We will also make a case for why we expect the active learner which we will introduce for Hamiltonian learning in Section 4.3 to achieve Heisenberg-limited scaling when possible.

The computational cost of the batch-mode AL scheme is determined by the number of rounds of batches issued and the computational cost of solving the query optimization problems of Equations 4.19 and 4.21. Solving these directly can be challenging but fortunately, the query optimizations can be reformulated as semidefinite programs (SDP) [Sou+17] under the assumptions of differentiability of the log-likelihood function, and invertibility of the Fisher information matrix for query distributions in a compact space around the optimal query distribution and around the uniform distribution. When using the query criteria of Fisher Information ratio (FIR) (Eq. 4.21), the optimization problem [Cha+15; Sou+17] is

$$\arg\min_{\alpha_1,\ldots,\alpha_d} \sum_{i=1}^{m} \alpha_i \quad \text{such that } \sum_{x\in\mathcal{Q}} q(x) = 1, \text{ and } \begin{bmatrix} \mathcal{I}_q(\boldsymbol{\theta}) & \mathbf{e}_j \\ \mathbf{e}_j^T & \alpha_j \end{bmatrix} \succcurlyeq 0,\ j \in [m] \tag{4.23}$$

where we have introduced m auxiliary variables $\alpha_1, ..., \alpha_m$, and $\mathbf{e}_j$ are the eigenvectors of $\mathcal{I}_p(\hat{\boldsymbol{\theta}})$. Recall that the parameter vector $\boldsymbol{\theta}$ has m components. To obtain the SDP program for the query optimization when using the query criteria of Fisher information (FI) (Eq. 4.19) in our AL strategy, we replace $\mathbf{e}_j$ by the eigenvectors of the identity matrix. In this case, $\mathbf{e}_j$ are m-dimensional canonical vectors with 1 in the jth component and zero elsewhere. The computational cost of solving the above SDP programs with a barrier interior-point method is $\mathcal{O}(n_{\mathcal{Q}}^2 m^3 + n_{\mathcal{Q}} m^4 + m^5)$ where we have denoted $n_{\mathcal{Q}} = |\mathcal{Q}|$ as the size of the query space of interest.

Query Advantage

To compare resource requirements of different Hamiltonian learning (HL) methods for accomplishing a learning task, we introduce the concept of *query advantage*. The query advantage (QA) of an HL method in achieving a learning error of ϵ is:

$$\text{QA} = 1 - \frac{\text{number of queries required by method}}{\text{number of queries required by baseline}} \tag{4.24}$$

As discussed before, QA measures the amount of query reduction obtained by selecting a HL method over a baseline strategy. In this work, we consider the passive learner equipped with an appropriately chosen estimation procedure as the baseline. We will specify the estimation procedure when discussing a QA result or it will be clear from context.

The benefits of quantifying QA for an HL method are twofold. Firstly, it allows us to comment on the performance boost obtained by using one HL method over another for accomplishing a learning task. Moreover, we can comment on learning tasks that can be achieved using one HL method but is unattainable by another. Secondly, it gives us a direct way to select a particular HL method based on minimal resources required, from a set of methods for a particular learning task by choosing the method with the highest QA.

As we will see in the next Section 4.3, an active learning strategy is a framework for achieving query advantage and higher learning rates of convergence when possible.

Related Work

Numerous methods exist for quantum Hamiltonian learning with different estimations and admissible experiments (aka query space). This work is not intended to replace these Hamiltonian learning methods but propose a general framework for active learning (or adaptive experiment design) that can replace the step of proposing experiments within these methods.

Active learning for Hamiltonian learning has not received as much attention as has quantum state tomography and quantum process tomography. The only example we could find in literature is that of robust online Hamiltonian learning (ROHL) [Gra+12] which has since been used for scalable Bayesian estimation for Hamiltonian learning [EHF19] and has been extended to Qinfer [Gra+17]. ROHL (or Qinfer) is a sequential active learner that proposes experiments (or queries) one shot at a time by minimizing a *risk* that is a function of the current estimate $\hat{\boldsymbol{\theta}}$ and examples seen so far. The proposed query for the $(k+1)$th shot then typically takes the form

$$x^{(k+1)} = \operatorname*{argmin}_{x \in \mathcal{Q}} R\left(X^{(k)}, Y^{(k)}; \hat{\boldsymbol{\theta}}^{(k)}\right), \tag{4.25}$$

where $X^{(k)}$ is the set of all queries taken so far, $Y^{(k)}$ are the associated measurement outcomes, and $\hat{\boldsymbol{\theta}}^{(k)}$ is the current estimate of the Hamiltonian parameters. In the case of ROHL and Qinfer, this is chosen to be Bayes risk given by

$$R\left(X^{(k)}, Y^{(k)}; \hat{\boldsymbol{\theta}}^{(k)}\right) = \mathbb{E}_{y|\hat{\boldsymbol{\theta}};Q}\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(X^{(k)}, Y^{(k)}))^T Q(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(X^{(k)}, Y^{(k)}))\right] \tag{4.26}$$

where the right right hand side can be interpreted as the expected posterior covariance matrix weighted against Q which is a positive semi-definite matrix indicating the relative scale between the unknown parameters in $\boldsymbol{\theta}$. In practice, minimizing the above risk function is expensive and one may resort to different experimental design heuristics [Hin+18].

The main limitation of the above method is that it is very expensive to decide queries one at a time during learning. This will be true irrespective of our progress in quantum computers as this processing is done on a classical computer, and is especially an obstacle on current hardware given the usual limited access, and latencies within electronics. It is then desirable to batch up queries (or experiments), which has not been considered by earlier work. This work which proposes an active learner operating in batch-mode thus fills this gap.

4.3 Hamiltonian active learning algorithms

We are now in a position to describe how to adapt probabilistic pool-based batch-mode active learning with query criteria of FI and FIR for Hamiltonian learning. The resulting algorithms are collectively called Hamiltonian Active Learning algorithms. We call HAL combined with the query criterion of Fisher Information as HAL-FI and that with Fisher Information Ratio as HAL-FIR. We discuss how the resulting algorithms are expected to achieve query advantage over a specified baseline.

4.3.1 Algorithm

The HAL algorithm is summarized in Algorithm 3. We assume that the unknown Hamiltonian is time-independent and a model parameterized by $\boldsymbol{\theta}^\star$ for the oracle (i.e., Hamiltonian with noise sources) is available to us. Inputs to the HAL algorithm include the initial query distribution $q^{(0)}$ to be used for sampling the initial set of $N_{tot}^{(0)}$ queries and the query optimization algorithm (QOA). When the query criterion is Fisher information (FI), the corresponding QOA is given by Algorithm 4 which solves Eq. 4.19. Similarly when the query criterion is Fisher information ratio (FIR), the corresponding QOA is given by Algorithm 5 which solves Eq. 4.21. Recall that the choice of query criteria and hence QOA depends on the learning task at hand: we consider FI for learning Hamiltonian parameters with low RMSE (Eq. 4.19) and FIR for minimizing the testing error (Eq. 4.21). Additional inputs include the maximum number of batches i_{max} that will be issued during learning which we will denote by N_b and the total experimental budget available.

Using the notation for batch mode active learning as discussed in Sec. 4.2.3, we denote the batch of queries that are sampled at the ith round with respect to the query distribution $q^{(i)}$ as $X_q^{(i)}$ and the corresponding measurement outcomes from inputting these queries to the *oracle* as $Y_q^{(i)}$. The set of all queries made so far at any round is denoted by $X^{(i)}$ and their corresponding measurement outcomes as $Y^{(i)}$. We note that $|X_q^{(0)}|= |Y_q^{(0)}|= N_{tot}^{(0)}$ which is typically larger than $|X_q^{(i)}|= |Y_q^{(i)}|= N_b$, when prior information about the parameters is not available. The initial set of training examples is used to determine an initial value of the parameters which is used for determining an informative albeit suboptimal query distribution $q^{(1)}$ through the given QOA. Note that in the QOA of Algorithms 4,5, we modify the query distribution $q^{(i)}$ obtained through solving Eq. 4.19 or Eq. 4.21, by mixing it with the uniform distribution over the query space p_U i.e., $\mu q^{(i)} + (1-\mu)p_U$ where $0 \leq \mu \leq 1$ is the mixing coefficient. This is done to encourage exploration and is analogous to epsilon-greedy policies in reinforcement learning [SB18]. The value of μ typically depends on the number of queries made so far, and we set it to $\mu = 1 - 1/|X^{(i)}|^{1/6}$ as often used for such active learning algorithms [Sou+17; Cha+15].

Deviating from a vanilla batch model AL scheme, we require an additional input of the query space $\{\mathcal{Q}^{(i)}\}_{i\in[i_{\max}]}$ during training. We allow the query space to adaptively change from one batch to the next. How do we decide how the query space changes from one batch to the next? We remind ourselves that the query distribution in $\mathcal{I}_q$ is a joint probability distribution over the measurement operators, preparation operators and evolution times. Conditioned on a particular evolution time, we would not expect changing $\mathcal{M}$ or $\mathcal{U}$ to help in reducing the query complexity. The query space is then adaptively grown by growing $\mathcal{T}$ linearly or exponentially with each batch. We note that $\mathcal{Q}^{(1)} \subset \mathcal{Q}^{(2)} \subset ... \subset \mathcal{Q}^{(i_{max})}$.

The output of the algorithm is an estimate $\hat{\boldsymbol{\theta}}$ of the true Hamiltonian model parameters $\boldsymbol{\theta}^\star$ learned through active learning using N queries. In the HAL algorithm as presented in Algorithm 3, we use the maximum likelihood estimator in the HAL algorithm but this may be replaced by other estimation methods such as regression or Bayesian estimation as we show later in Section 4.5. HAL-FI produces an estimate $\hat{\boldsymbol{\theta}}$ that has an RMSE lower than what would be obtained without any active learning using a budget of N queries, and batches of size N_b. Similarly, HAL-FIR produces an estimate $\hat{\boldsymbol{\theta}}$ that performs well in prediction of queries to the Hamiltonian against the testing distribution p_{test}.

Algorithm 3 Hamiltonian Active Learning (HAL)

Input: Initial number of queries $N_{tot}^{(0)}$, Batch size N_b, Initial query distribution $q^{(0)}$, Maximum number of batches $i_{\max}$, Adaptively growing query space $\{\mathcal{Q}^{(i)}\}_{i\in[i_{\max}]}$, oracle, query optimization algorithm (QOA)

Output: $\hat{\boldsymbol{\theta}}$

Sample $N_{tot}^{(0)}$ queries $X_q^{(0)}$ from $\mathcal{Q}^{(0)}$ according to $q^{(0)}$
Obtain measurement outcomes $Y_q^{(0)}$ by sending queries $X^{(0)}$ to oracle
Set $X^{(0)} = X_q^{(0)}$ and $Y^{(0)} = Y_q^{(0)}$
Compute MLE estimate: $\hat{\boldsymbol{\theta}}^{(i-1)} = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; X^{(0)}, Y^{(0)})$
for $i = 1 : i_{\max}$ **do**
    Solve $q^{(i)}$ through QOA (Algorithm 4 or 5)
    Sample N_b queries $X_q^{(i)}$ from $\mathcal{Q}^{(i)}$ w.p. $q^{(i)}$
    Update number of queries: $N_{tot}^{(i)} = N_{tot}^{(i-1)} + N_b$
    Obtain measurement outcomes $Y^{(i)}$ by issuing queries $X^{(i)}$ to oracle
    Set $X^{(i)} = X^{(i-1)} \bigcup X_q^{(i)}$ and $Y^{(i)} = Y^{(i-1)} \bigcup Y_q^{(i)}$
    Compute MLE estimate: $\hat{\boldsymbol{\theta}}^{(i)} = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; X^{(i)}, Y^{(i)})$
end for
return $\hat{\boldsymbol{\theta}}^{(i_{\max})}$

Algorithm 4 Query Optimization based on Fisher Information (FI)

Input: Number of queries made so far $N_{tot}^{(i-1)}$, Batch size N_b, Query space $\mathcal{Q}^{(i)}$, Current parameter estimates $\hat{\boldsymbol{\theta}}^{(i)}$

Output: $q^{(i)}$

Set $N_{tot}^{(i)} = N_{tot}^{(i-1)} + N_b$
Solve $q^{(i)} = \arg\min_q \mathrm{Tr}(\mathcal{I}_q^{-1}(\hat{\boldsymbol{\theta}}^{(i)}))$ subject to $\sum_{x\in\mathcal{Q}^{(i)}} q(x) = 1$, and $0 \le q(x) \le 1, \forall x \in \mathcal{Q}^{(i)}$
Obtain uniform distribution over query space: $p_U = 1/|\mathcal{Q}^{(i)}|$
Set mixing coefficient: $\mu = 1 - 1/|N_{tot}^{(i-1)}|^{1/6}$
Modify query distribution: $q^{(i)} = \mu q^{(i)} + (1-\mu)p_U$
return $q^{(i)}$

Algorithm 5 Query Optimization based on Fisher Information Ratio (FIR)

Input: Number of queries made so far $N_{tot}^{(i-1)}$, Batch size N_b, Query space $\mathcal{Q}^{(i)}$, Current parameter estimates $\hat{\boldsymbol{\theta}}^{(i)}$, Testing Distribution p_{test}
Output: $q^{(i)}$

1: Set $N_{tot}^{(i)} = N_{tot}^{(i-1)} + N_b$
2: Compute model fisher information corresponding to p_{test}: $\mathcal{I}_{p_{\text{test}}}(\hat{\boldsymbol{\theta}})$
3: Solve $q^{(i)} = \arg\min_q \text{Tr}\left(\mathcal{I}_q^{-1}(\hat{\boldsymbol{\theta}}^{(i)})\mathcal{I}_{p_{\text{test}}}(\hat{\boldsymbol{\theta}}^{(i)})\right)$ subject to $\sum_{x\in\mathcal{Q}^{(i)}} q(x) = 1$, and $0 \leq q(x) \leq 1, \forall x \in \mathcal{Q}^{(i)}$
4: Obtain uniform distribution over query space: $p_U = 1/|\mathcal{Q}^{(i)}|$
5: Set mixing coefficient: $\mu = 1 - 1/|N_{tot}^{(i-1)}|^{1/6}$
6: Modify query distribution: $q^{(i)} = \mu q^{(i)} + (1-\mu)p_U$
7: **return** $q^{(i)}$

4.3.2 Comment on Query Advantage and Heisenberg Limited Scaling

In this section, we provide intuition as to why we expect HAL-FI to provide query advantage and when Heisenberg limited scaling may be expected in experiment. The intuition and claims made here are backed by empirical evidence in Section 4.5.

We start by comparing the performance of a passive learner L_{PL} and an active learner L_{HAL} which uses HAL-FI. We assume that both of these learners use the same estimation method to estimate $\hat{\boldsymbol{\theta}}$ from the training examples generated. Consider the second round of learning (i=2) with a fixed query space $\mathcal{Q}$ and where both learners have seen the same data $D^{(1)}$ so far. L_{PL} will uses queries uniformly sampled from $\mathcal{Q}$ and L_{HAL} will use HAL-FI to decide the query distribution to use for sampling from $\mathcal{Q}$. The corresponding Fisher information is $\mathcal{I}_{p_U}(\hat{\boldsymbol{\theta}}^{(1)})$ (where p_U is the uniform distribution) and $\mathcal{I}_q(\hat{\boldsymbol{\theta}}^{(1)})$. From Eq. 4.19, we know that $\text{Tr}(\mathcal{I}_q^{-1}(\hat{\boldsymbol{\theta}}^{(1)})) \leq \text{Tr}(\mathcal{I}_{p_U}^{-1}(\hat{\boldsymbol{\theta}}^{(1)}))$. We thus expect L_{HAL} to require the fraction of $r^{(1)} \approx \text{Tr}(\mathcal{I}_q^{-1}(\hat{\boldsymbol{\theta}}^{(1)}))/\text{Tr}(\mathcal{I}_{p_U}^{-1}(\hat{\boldsymbol{\theta}}^{(1)})) \leq 1$ number of queries compared to L_{PL} to achieve the same variance reduction. This is compounded through the following learning rounds, leading to an overall reduction in queries required (or gain in query advantage) by L_{HAL} to reach the same RMSE compared to L_{PL}. Moreover, in latter rounds ($i > 2$), the fraction of queries required by L_{HAL} may further decrease due to more informative training examples available to obtain $\hat{\boldsymbol{\theta}}^{(i)}$ which is then used to obtain the query distribution.

We claim that HAL-FI with an appropriately chosen adaptively growing query space by adaptively growing $\mathcal{T}$ can achieve Heisenberg limited scaling in Hamiltonian parameters where possible. If the query space is not *rich* enough, it will not be possible to determine a sequence of queries to achieve Heisenberg limited scaling. Without an adaptively growing query space (i.e., with fixed query space), the scaling of the number of queries N with error parameter ϵ is $O(1/\epsilon^2)$ as is dictated by the Cramer-Rao bound. The query complexity by using an active learner only improves by a constant factor in such a case. If one chose to adaptively grow the query space during training without an active learning strategy and chose for example an uniform distribution over the new query space (i.e., carry out passive learning), Heisenberg limited scaling would not be expected as it would become

exponentially more unlikely to sample an informative query. Moreover, Heisenberg limited scaling is expected as long as the evolution time is below the decoherence time. Beyond the decoherence time, the oracle will start losing its quantum behavior.

If the so chosen adaptively growing query space cannot be used to achieve Heisenberg limited scaling, it should still be possible to achieve Heisenberg limited scaling for a subset of the Hamiltonian parameters provided that the goal is to now learn these parameters and we are provided information about the other parameters. This learning task often occurs in practice during recalibrations of quantum devices when it is required to learn a subset of parameters that are known to fluctuate significantly with time but information about the other parameters can be used from previous calibrations during estimation. We provide empirical evidence for this claim in Section 4.5 where we consider the application of the HAL-FI to cross-resonance Hamiltonians. The results of HAL-FI for this particular application are summarized in Table 4.1.

What distinguishes HAL-FI from other methods such as Floquet calibration [Aru+20a] which have been shown to achieve Heisenberg limited scaling is that it does not require prior specification of experiments and their order of implementation. This is decided by the HAL-FI during learning. Moreover, HAL-FI utilizes single-shot outimes from queries, instead of requiring expectation values. In practice, this can result in multiple orders of magnitude reduction in queries required. Finally, HAL-FI can achieve query advantage over passive learners for complete query spaces even when Heisenberg limited scaling cannot be achieved.

4.3.3 Computational Cost and Extensions

A consequence of adaptively growing the query space over rounds during learning is that the SDP programs (Eq. 4.23) corresponding to the query optimizations of Eqs. 4.19 and 4.21 (with computational cost $\mathcal{O}(n_{\mathcal{Q}}^2 m^3 + n_{\mathcal{Q}} m^4 + m^5)$) become increasingly more computationally expensive to solve over rounds. If $n_{\mathcal{Q}}$ grows exponentially, each iteration of the query optimization problem becomes more exponentially expensive to solve. This can be circumvented by reducing the number of queries to optimize over using uncertainty filtering [WIB15], thereby effectively reducing the size of the query space $n_{\mathcal{Q}}$ over which the query optimization is carried out. Uncertainty filtering for our application of Hamiltonian learning to the cross-resonance Hamiltonian is discussed in Appendix B.2.2.

The HAL-FI and HAL-FIR algorithms can be generalized to different experimental setups or requirements. The HAL algorithm presented in Algorithm 3 uses the stopping criterion of maximum number of batches of queries issued during learning but other stopping criteria such as the ℓ_2 norm of the differences in consecutive parameter values $||\hat{\boldsymbol{\theta}}^{(i)} - \hat{\boldsymbol{\theta}}^{(i-1)}||_2$ could also be used.

4.4 Hamiltonian learning for a two-qubit superconducting cross-resonance gate: Model and setup

To assess the performance of the HAL algorithms described in Section 4.3 over a passive learner and empirically verify our claims, we consider the application of learning cross-resonance (CR) Hamiltonians on superconducting IBM Quantum devices. In this section,

Adaptivity in Query Space	Scaling of N	Scaling of N (Recalibration)
None	$O(1/\epsilon^2)$	$O(1/\epsilon^2)$
Linearly growing $\mathcal{T}$	$O(1/\epsilon^2)$	$O(1/\epsilon^{2/3})$
Exponentially growing $\mathcal{T}$	$O(1/\epsilon^2)$	$> O(1/\epsilon)$

Table 4.1: Query complexity of HAL-FI for Hamiltonian learning (Problem 4.2.1) under different conditions as observed on a real quantum device. The case of fixed query space corresponds to adaptivity in query space being *none*.

we describe the model of the two-qubit cross-resonance Hamiltonian, and the set of queries that we can make to it. This is followed by a description of the different noise sources that affect the quantum system and the resulting likelihood of measurement outcomes given queries. The evaluation of the likelihood is required by all of our estimators (Section 4.2.1). In turn, we then describe different estimators used for Hamiltonian learning including MLE and a Bayesian estimator. Finally, we describe our implementation of the HAL algorithms in the context of numerical experiments for learning the CR Hamiltonians. A description of the IBM Quantum devices employed for assessing the performance of the HAL algorithms can be found in Appendix B.1.

4.4.1 Cross-Resonance Hamiltonian

The cross resonance (CR) gate is a two-qubit entangling gate for superconducting qubits requiring only microwave control which allows for the use of fixed-frequency transmon qubits [RD10; Cho+11]. Using appropriate pulse sequences such as multi-pulse echos and cancellation tones [She+16], the CR gate can be transformed to a locally equivalent CNOT gate [Zha+03]. Combined with arbitrary single qubit gates, this then forms a complete set of gates for universal quantum computation.

The Hamiltonian of the cross-resonance (CR) gate has the following structure

$$H_{CR} = \frac{\sigma_Z \otimes A}{2} + \frac{\sigma_I \otimes B}{2} \tag{4.27}$$

$$A = c_{ZI}\sigma_I + c_{ZX}\sigma_X + c_{ZY}\sigma_Y + c_{ZZ}\sigma_Z \tag{4.28}$$

$$B = c_{IX}\sigma_X + c_{IY}\sigma_Y + c_{IZ}\sigma_Z \tag{4.29}$$

where $\{\sigma_I, \sigma_X, \sigma_Y, \sigma_Z\}$ are the single-qubit Paulis and $c_{ab} \in \mathbb{R}$ are real coefficients of the Pauli product terms $\sigma_a \otimes \sigma_b$. The above time-independent Hamiltonian description of the cross-resonance gate can be obtained through theoretical models based on effective block-diagonal Hamiltonian techniques [MG20]. In our experiments which we describe in Section 4.4.2, the CR gate is implemented without using an echo [She+16] to refocus the $\sigma_I\sigma_X$, $\sigma_Z\sigma_Z$ and $\sigma_Z\sigma_I$ terms. However, we measure only the target qubit through our queries and thus effectively neglect the $\sigma_Z\sigma_I$ term. The target qubit is typically chosen to be qubit 1 in a $(0, 1)$ qubit pair and is specified for the different quantum devices we consider in Appendix B.1.1. The effective removal of the $\sigma_Z\sigma_I$ term from Eq. 4.29 results in the following simplified CR

Hamiltonian which we consider throughout the rest of our work:

$$H = \sum_{\substack{a\in\{I,Z\}\\ b\in\{X,Y,Z\}}} J_{ab}\sigma_a \otimes \sigma_b \tag{4.30}$$

where we have used the parameter vector $\mathbf{J} = [J_{IX}, J_{IY}, J_{IZ}, J_{ZX}, J_{ZY}, J_{ZZ}]^T$ to denote the non-zero coefficients of the corresponding Pauli product terms and have omitted the subscript CR. Learning the unknown Hamiltonian of the two-qubit CR gate is then reduced to estimating the unknown parameter vector $\mathbf{J}$.

Noting the block-diagonal structure of H, we can express it in the usual computational basis of $(|00\rangle, |01\rangle, |10\rangle, |11\rangle)$ as

$$H = \begin{bmatrix} a_0 & \beta_0^\star & 0 & 0 \\ \beta_0 & -a_0 & 0 & 0 \\ 0 & 0 & a_1 & \beta_1^\star \\ 0 & 0 & \beta_1 & -a_1 \end{bmatrix}, \text{ where } \begin{matrix} a_j = J_{IZ} + (-1)^j J_{ZZ} \\ \beta_j = (J_{IX} + (-1)^j J_{ZX}) + i(J_{IY} + (-1)^j J_{ZY}) \end{matrix} \tag{4.31}$$

where the subscript $j \in \{0, 1\}$ is used to refer to the two different blocks. The two blocks have similar structure with their elements differing in the sign of J_{ZX}, J_{ZY}, and J_{ZZ}. Using a_j, and β_j for $j \in \{0, 1\}$ from Eq. 4.31, we define the following parameters

$$\omega_j = \sqrt{a_j^2 + |\beta_j|^2}, \quad \delta_j = \sin^{-1}\frac{a_j}{\omega_j}, \quad \phi_j = \arg(\beta_j) \tag{4.32}$$

where the subscript $j = 0$ is used to denote the preparation operator $U = \sigma_I\sigma_I$ and $j = 1$ to denote $U = \sigma_X\sigma_I$. We then arrive at an alternate parameterization (related to the spectral decomposition of H) which turns out to be useful for simplifying expressions for probability, likelihood and Fisher information: $\mathbf{\Lambda} = (\omega_0, \delta_0, \phi_0, \omega_1, \delta_1, \phi_1)^T$. The unitary operator $U(t) = e^{-iHt}$ in the usual computational basis is then given by

$$U(t) = \begin{bmatrix} \cos(\omega_0 t) - i\sin(\delta_0)\sin(\omega_0 t) & -ie^{-i\phi_0}\cos(\delta_0)\sin(\omega_0 t) & 0 & 0 \\ -ie^{i\phi_0}\cos(\delta_0)\sin(\omega_0 t) & \cos(\omega_0 t) + i\sin(\delta_0)\sin(\omega_0 t) & 0 & 0 \\ 0 & 0 & \cos(\omega_1 t) - i\sin(\delta_1)\sin(\omega_1 t) & -ie^{-i\phi_1}\cos(\delta_1)\sin(\omega_1 t) \\ 0 & 0 & -ie^{i\phi_1}\cos(\delta_1)\sin(\omega_1 t) & \cos(\omega_1 t) + i\sin(\delta_1)\sin(\omega_1 t) \end{bmatrix} \tag{4.33}$$

where we have used the parameter vector of $\mathbf{\Lambda}$ defined in Eq. 4.32. Moreover, the different components of $\mathbf{\Lambda}$ can be bounded based on their definitions and these bounds will help us later while solving the MLE problems. By construction, we have $-\frac{\pi}{2} \leq \delta_{0,1} \leq \frac{\pi}{2}$ and $\phi_{0,1}$ can be bounded within any interval of size 2π e.g., $-\pi \leq \phi_{0,1} \leq \pi$. Assuming δt is the average time increment between distinct ordered values of evolution times $t \in \mathcal{T}$, $\omega_{0,1}$ can be bounded based on the Nyquist sampling theorem as $0 \leq \omega_{0,1} \leq \frac{\pi}{\delta t}$.

To obtain the physical meaning of the parameter vector $\mathbf{\Lambda}$, we consider Rabi oscillations. For different measurement operators $M \in \mathcal{M}$ and preparation operators $U \in \mathcal{U}$, the corresponding Rabi oscillation is the difference in probability densities of the ground state and excited state of the target qubit with time $t \in \mathcal{T}$. A typical example of the Rabi oscillations from querying the CR Hamiltonian on a noisy quantum system is shown in Figure 4.3. We remark that $\omega_{0,1}$ defines the frequency of the Rabi oscillations for the two different preparation operators we consider. The parameters $\delta_{0,1}$ and $\phi_{0,1}$ define the offsets, amplitudes and phase shifts of the Rabi oscillations. In Figure 4.3, we can see the effects of different noise sources such as readout and depolarization which will be discussed in Section 4.4.3.

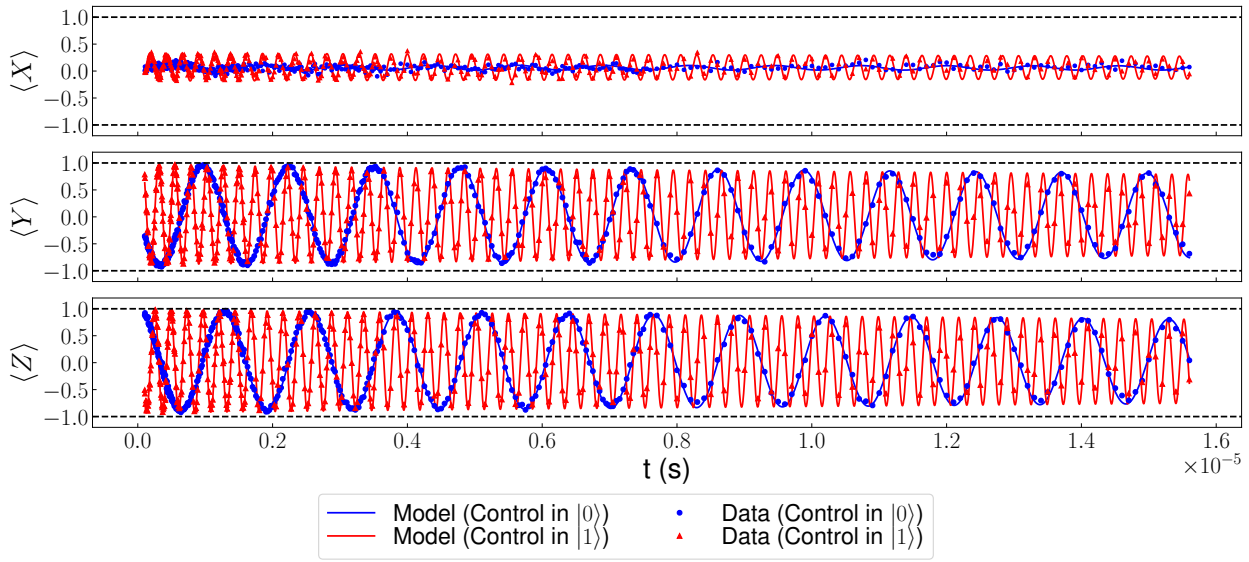


Figure 4.3: Rabi oscillations obtained experimentally through queries to a CR Hamiltonian on the IBM Quantum device (*ibmq_ boeblingen*) for different measurement operators (rows), preparation operators (markers) and evolution times t (x-axis). The set of measurement and preparation operators will be described in Section 4.4.2. The experimental data is indicated by markers and the model fit to the data by lines. The model fit was generated by learning the CR Hamiltonian (Eq. 4.31) from experimental data using MLE and predicting values of Rabi oscillations for the query space using the learned Hamiltonian.

4.4.2 Experimental Setup

In this section, we give a quick overview of the different IBM Quantum devices that we employ for our application of Hamiltonian learning. This is followed by a description of the query space considered for our application.

Quantum Devices

We will present data and results for four different IBM Quantum devices which we call A, B, C, and D. All of these devices are based on superconducting architectures requiring only microwave control [RD10; Cho+11] and consist of fixed-frequency transmon qubits with shared quantum buses [Cho+14]. Device A is a two-qubit device. Device B is a four-qubit device on which we will query only one of the CR gates that can be implemented between two qubits. Device C is a five-qubit device with a bow-tie layout. Device D is the 20-qubit *ibmq_ boeblingen* which was accessed via the IBM Quantum cloud computing service. See Figure B.1 in Appendix B.1.1 for the connectivity maps of all the devices. We give a summary of the properties of the pairs of qubits involved in the cross-resonance Hamiltonians we considered on these devices in Table B.1 in Appendix B.1.1.

Query Space

To prepare the initial state, we consider the set of preparation operators $\mathcal{U} = \{\sigma_I \otimes \sigma_I, \sigma_X \otimes \sigma_I\}$ applied to the pure state $|00\rangle$. Assuming the first (left) qubit is the control and the second (right) qubit is the target, the effect of the preparation operators is to place the control in $|0\rangle$ and $|1\rangle$ respectively. We evolve the initial state $|\psi(0)\rangle$ for time $t \in \mathcal{T}$ which we will specify when discussing the results of our application of Hamiltonian learning in Section 4.5. Finally after obtaining the final state $|\psi(t)\rangle$, we apply the measurement operators in $\mathcal{M} = \{\sigma_I \otimes \exp\left(i\frac{\pi}{4}\sigma_Y\right), \sigma_I \otimes \exp\left(-i\frac{\pi}{4}\sigma_X\right), \sigma_I \otimes \sigma_I\}$ and measure only the second qubit which we have chosen as the target qubit. The query space is then $\mathcal{Q} = \mathcal{M} \times \mathcal{U} \times \mathcal{T}$.

Queries to the CR Hamiltonians between different qubit pairs on the IBM Quantum devices are made through appropriate pulse sequences. These pulse sequences are constructed and executed on the hardware using `Qiskit-Pulse` [Ale+20], which is a pulse programming module within `Qiskit` [Abr+19a]. Description of how a query to the CR gate on a IBM Quantum device is specified is given in Appendix B.1.2.

In our experimental setup, we obtain measurements of the single-shot signal (integrated cavity amplitude) $\mathbf{c}$ from the IBM Quantum devices which is a function of the measurement outcomes $\mathbf{y}$ which we have described earlier. In the following Section 4.4.3, we will discuss how to model the different noise sources that affect our system and how they are determined through experiments.

4.4.3 Estimates of Noise and Nonidealities for Experimental System

Using the noise models presented in Section 4.2.2, we give the specific relevant models for readout noise, imperfect pulse shaping, and decoherence for the different IBM Quantum devices which we study.

Let us introduce some notation that we will use in the following discussions. We use the index k for the queries. The kth query is given by $x^{(k)} = (M^{(k)}, U^{(k)}, t^{(k)})$, and the corresponding measurement outcome of the target qubit as $y^{(k)}$. The measurement outcome $y^{(k)}$ is not directly observed but inferred from the corresponding signal $\mathbf{c}^{(k)}$. We denote the inferred value as $\hat{y}^{(k)}$.

Readout Noise

As discussed in Section 4.2.2, we assume the measurement noise model to be a bit-flip channel with the input of unobserved measurement outcomes y and the output of readout $\tilde{y}$ observed through signal $\mathbf{c}$. For this inference task, we use calibration data of single qubits initially prepared in states $|0\rangle$ or $|1\rangle$, and subsequently measured in the usual computational basis. In Figure 4.4(a), we plot different realizations $\{y^{(k)}, \mathbf{c}^{(k)}\}_{k=1}^{100}$ from IBM Quantum device D *ibmq_boeblingen* that were used for training a binary classifier. The classifier provides us with the ability to predict y given $\mathbf{c}$. Moreover, the misclassification errors $p_{\hat{y}|y}(1|0)$ and $p_{\hat{y}|y}(0|1)$ can be used to approximate the properties of the bit-flip channel, in particular the conditional probabilities of a bit-flip $p_{\tilde{y}|y}(1|0)$ and $p_{\tilde{y}|y}(0|1)$ respectively. As these are independent of values of $\mathbf{c}$ here, we denote the conditional probabilities of a bit-flip as $r_0 = p_{\tilde{y}|y}(1|0)$ and $r_1 = p_{\tilde{y}|y}(0|1)$. The MLE of the parameters incorporating this noise model

is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -\frac{1}{N}\sum_{k=1}^{N} \log\left[p_{\tilde{\mathrm{y}}|\mathrm{x}}(\tilde{y}^{(k)}|x^{(k)};\boldsymbol{\theta})\right] \tag{4.34}$$

$$= \arg\min_{\boldsymbol{\theta}} -\frac{1}{N}\sum_{k=1}^{N} \log\Big[(1-\hat{y}^{(k)})\left((1-r_0)p_{\mathrm{y}|\mathrm{x}}(y^{(k)}=0|x^{(k)};\boldsymbol{\theta}) + r_1(1-p_{\mathrm{y}|\mathrm{x}}(y^{(k)}=1|x^{(k)};\boldsymbol{\theta}))\right) + \tag{4.35}$$

$$\hat{y}^{(k)}\left((1-r_1)(1-p_{\mathrm{y}|\mathrm{x}}(y^{(k)}=1|x^{(k)};\boldsymbol{\theta})) + r_0 p_{\mathrm{y}|\mathrm{x}}(y^{(k)}=0|x^{(k)};\boldsymbol{\theta})\right)\Big] \tag{4.36}$$

where $p_{\mathrm{y}|\mathrm{x}}(y^{(k)}|x^{(k)};\boldsymbol{\theta})$ is given by Eq. 4.1. Alternately, instead of assigning a deterministic result $\hat{y}^{(k)}$ for each $c^{(k)}$, we can incorporate $p_{\mathrm{c}|y}$ directly into the log-likelihood function which could yield a more accurate model. This could be done through our choice of binary classifier or by fitting a distribution to the training data. Noting that single qubit measurement outcomes correspond to their energy levels, we fit Gaussian distributions to the training data and hence obtain a parameteric form of $p_{\mathrm{c}|y}$ in Figure 4.4(b). The MLE is now

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -\frac{1}{N}\sum_{k=1}^{N} \log\left[p_{\mathrm{c}|\mathrm{y}}(c^{(k)}|0)p_{\mathrm{y}|\mathrm{x}}(y^{(k)}=0|x^{(k)};\boldsymbol{\theta}) + p_{\mathrm{c}|\mathrm{y}}(c^{(k)}|1)(1-p_{\mathrm{y}|\mathrm{x}}(y^{(k)}=0|x^{(k)};\boldsymbol{\theta}))\right]. \tag{4.37}$$

A useful tool for calibration and diagnostics is Rabi oscillations which we denote by $p_{\mathrm{rabi}}(x)$. Rabi oscillations are obtained from evaluating the difference in the population densities of the ground state and excited state of the target qubit or $p_{\mathrm{rabi}}(x) = p_{\mathrm{y}|\mathrm{x}}(0|x) - p_{\mathrm{y}|\mathrm{x}}(1|x)$. We can compute Rabi oscillations in two different ways, either through binary classification or through fitting Gaussians. Using the misclassification errors from the binary classifier, we can then write

$$\hat{p}_{\mathrm{rabi}}(x) = p_{\hat{\mathrm{y}}|\mathrm{x}}(0|x)\left(\frac{1-r_0+r_1}{1-r_0-r_1}\right) - p_{\hat{\mathrm{y}}|\mathrm{x}}(1|x)\left(\frac{1+r_0-r_1}{1-r_0-r_1}\right), \tag{4.38}$$

where we have denoted the computed Rabi oscillations as $\hat{p}_{\mathrm{rabi}}(x)$. While $p_{\mathrm{y}|\mathrm{x}}(0|x)$ and $p_{\mathrm{y}|\mathrm{x}}(1|x)$ are guaranteed to be valid probability distributions, the above estimation does not ensure that $\hat{p}_{\mathrm{rabi}}(x)$ is bounded by -1 and 1. To obtain more accurate estimates of the Rabi oscillations, we can solve the following MLE problem

$$\hat{p}_{\mathrm{rabi}}(x) = \arg\min_{q\in[-1,1]}\left(-\sum_{k}\mathbb{1}\{x^{(k)}=x\}\log\left[(1+q)p_{\mathrm{c}|\mathrm{y}}(c^{(k)}|0) + (1-q)p_{\mathrm{c}|\mathrm{y}}(c^{(k)}|1)\right]\right), \tag{4.39}$$

where we use the estimated conditional distributions $p_{\mathrm{c}|y}$ from the Gaussian fits. The indicator function $\mathbb{1}\{x^{(k)}=x\}$ is used to ensure that the summation is only over measurement outcomes of given query x. We can write down the analytical expressions for $p_{\mathrm{rabi}}(x)$ for each measurement operator M and preparation operator U noting that $x = (M, U, t)$ as

$$M_{\langle X\rangle} : p_{\mathrm{rabi}}(x) = \sin\delta_j\cos\delta_j\cos\phi_j + \cos\delta_j\sqrt{1-\cos^2\delta_j\cos^2\phi_j}\cos(2\omega_j t+\alpha_j), \tag{4.40}$$

$$M_{\langle Y\rangle} : p_{\mathrm{rabi}}(x) = \sin\delta_j\cos\delta_j\sin\phi_j + \cos\delta_j\sqrt{1-\cos^2\delta_j\sin^2\phi_j}\cos(2\omega_j t+\gamma_j), \tag{4.41}$$

$$M_{\langle Z\rangle} : p_{\mathrm{rabi}}(x) = 1-2\cos^2\delta_j\sin^2(\omega_j t) = \sin^2\delta_j + \cos^2\delta_j\cos(2\omega_j t), \tag{4.42}$$

where we denote $\alpha_j = \arg((-\sin\delta_j \cos\phi_j) + i(-\sin\phi_j))$ and $\gamma_j = \arg((-\sin\delta_j \sin\phi_j) + i\cos\phi_j)$ in the above expressions. The subscript $j = 0$ is used to denote the preparation operator $U = \sigma_I\sigma_I$ and $j = 1$ to denote $U = \sigma_X\sigma_I$.

In Figure 4.5, we plot the Rabi oscillations for the different $M \in \mathcal{M}$ and $U \in \mathcal{U}$ over the time range of $t \in \mathcal{T}$ using the above two methods. We observe that the Rabi oscillations are not bounded between -1 and $+1$ for the case of $M_{\langle Y \rangle}$ in Figure 4.5(a) when using misclassification error to compensate for the readout noise. The estimated Rabi oscillations in Figure 4.5(b) do not suffer from the same issue.

As we will see later in Section 4.4.4, Rabi oscillations will also be used in our estimation procedure for obtaining an initial guess for the parameter estimate $\hat{\boldsymbol{\theta}}$ that is used as an input to the optimizer for solving the MLE problem. It can also be used as a quantitative tool for ascertaining how well the model fits the data. This will become apparent in the next few sections where we discuss other noise models.

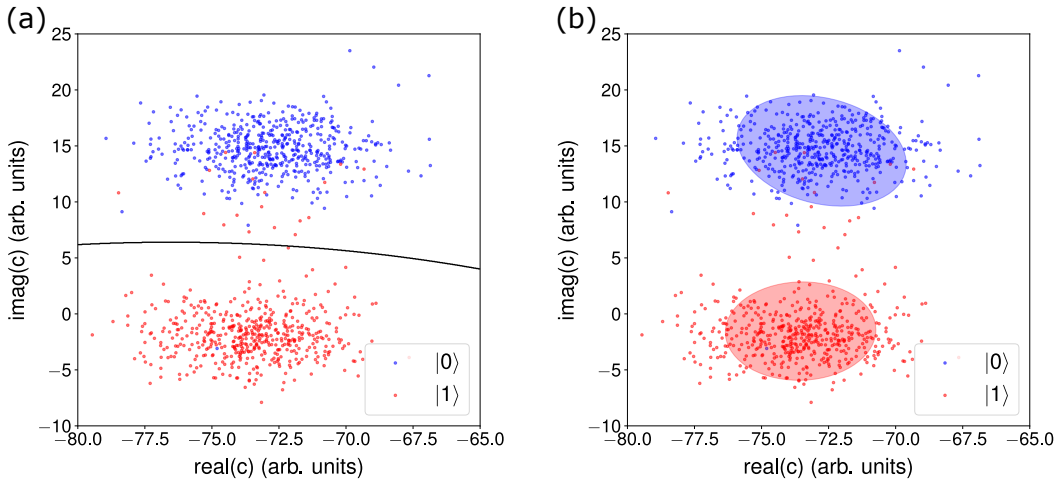


Figure 4.4: Characterization of readout noise from calibration data of single qubit readouts considering a bit-flip channel model using (a) a trained Bayesian naive classifier, and (b) fitting Gaussian distributions. In (a)-(b), the experimental data points of the complex readout signal $\mathbf{c}$ are shown as markers. In (a), the decision boundary is shown as a line. In (b), the contours indicate the single standard deviation.

Imperfect Pulse Shaping

Another nonideality is introduced through the pulses used to control the Hamiltonian and implement different unitary operators. It is convenient to think of cross-resonance control pulses as rectangular pulses that modulate the sinusoidal control pulse. The modulated signal results in unitary operators of the form $U(t_1) = \exp(-iHt_1)$ which we would want to implement in a quantum circuit. However, in practice, rectangular pulses cause significant amounts of signal energy to be distributed above and below the frequency of the control sinusoid. This distribution of energy can potentially excite higher energy states of the superconducting transmon being used as a qubit (i.e., $|2\rangle$ and above) as well as excite neighboring spectator transmon qubits. To minimize such effects, pulse shaping is employed to reduce

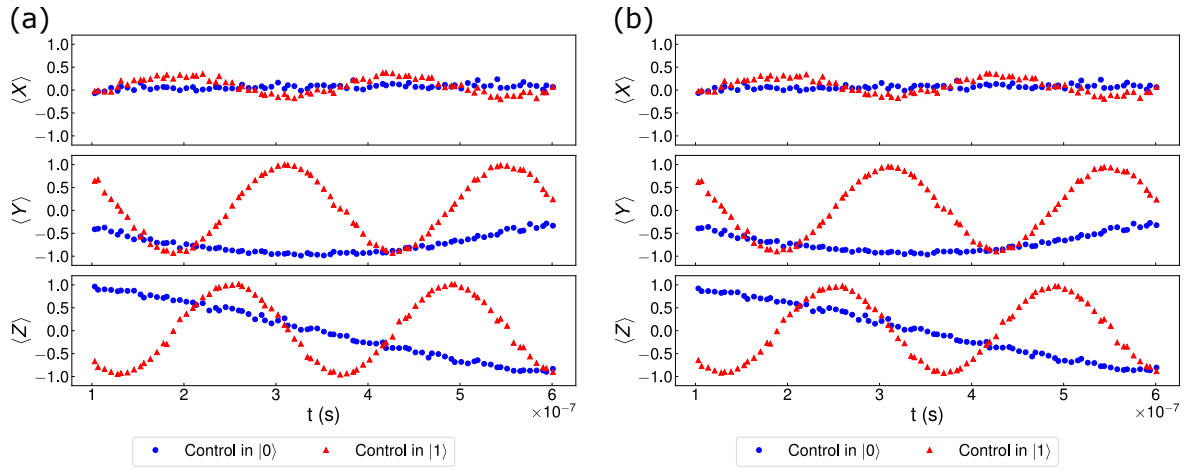


Figure 4.5: Examples of Rabi oscillations computed from experimental data collected from IBM Quantum device D *ibmq_boeblingen* for different measurement operators M (rows), preparation operators U (markers) and evolution times t (x-axis). These were computed (a) using a binary classifier and (b) using Gaussian distribution fits. The difference between the two approaches is negligible when the misclassification error of the binary classifier is low ($r_0, r_1 < 0.01$) as is the case here.

this energy spread by smoothing the rising and falling edges of the pulse, which has the effect of reducing the magnitudes of the frequency artifacts above and below the frequency of the control sinusoid.

Pulse shaping is accomplished by taking a Gaussian-shaped pulse, splitting it in half, and then inserting a rectangular pulse between the halves. This results in the `GaussianSquare` pulse, described in further detail in Section B.1.2. Thus, we actually implement operators of the form $\tilde{U}(t_1) = \mathbb{T}\exp\left(-i\int_0^{t_1}\tilde{H}(t;\boldsymbol{\theta})dt\right)$ where $\mathbb{T}$ is the time ordering operator and $\tilde{H}(t;\boldsymbol{\theta})$ is the cross-resonance Hamiltonian due to imperfect control at any particular time t with parameters $\boldsymbol{\theta}$. Up to a first-order approximation in t, we can model $\tilde{H}(t;\boldsymbol{\theta})$ as (see Appendix B.1.3 for details)

$$\tilde{H}(t;\boldsymbol{\theta}) = v(t)H(\boldsymbol{\theta}), \tag{4.43}$$

where $H(\boldsymbol{\theta})$ is the time-independent cross-resonance Hamiltonian and $v(t)$ is a function of the cross-resonance pulse amplitude. Using this, we can now derive an expression for the unitary of Hamiltonian evolution. We denote Δt_r and Δt_f as the time durations of the rising and falling edges of the shaped pulse. The central portion of $v(t')$ is a rectangular function; i.e., for $\Delta t_r \leq t' \leq t - \Delta t_f$, $v(t') = \mathbb{1}_{t'\in[\Delta t_r, t-\Delta t_f]}$ where t is the duration of the pulse. We then have

$$\tilde{U}(t) = \exp\left(-iH(\boldsymbol{\theta})\int_0^t v(t')dt'\right) = \exp\left(-iH(\boldsymbol{\theta})\left[(t-\Delta t_f-\Delta t_r) + \underbrace{\int_0^{\Delta t_r} v(t')dt' + \int_{t-\Delta t_f}^{\Delta t_f} v(t')dt'}_{\Delta t_{\text{eff}}}\right]\right) \tag{4.44}$$

$$= \exp\left(-iH(\boldsymbol{\theta})(t_{\text{expt}} + \Delta t_{\text{eff}})\right), \tag{4.45}$$

where in the last step, we set $t_{\text{expt}} = t - \Delta t_f - \Delta t_r$ which is the evolution time that is reported in our experiments. It should be noted that this can change from one experimental setup to

another. The value of Δt_{eff} can be interpreted as an effective total edge duration that takes the shapes of the rising and falling edges into account. This first-order pulse-shaping model introduces another model parameter, namely Δt_{eff}, that we need to estimate. We can do this directly in our MLE:

$$[\hat{\boldsymbol{\theta}}, \Delta\hat{t}_{\text{eff}}] = \arg\min_{\boldsymbol{\theta}, \Delta t_{\text{eff}}} -\frac{1}{N}\sum_{k=1}^{N} \log\left[\sum_{y\in\{0,1\}}\left(p_{\text{c|y}}(c^{(k)}|y)\sum_{z\in\{0,1\}}\left|\left\langle yz|M^{(k)}e^{-iH(\boldsymbol{\theta})(t^{(k)}+\Delta t_{\text{eff}})}U^{(k)}|00\right\rangle\right|^2\right)\right] \tag{4.46}$$

While this can be certainly done while learning the Hamiltonian, one can also determine the dependence of Δt_{eff} on the different Hamiltonian parameters using prior calibration data. We determined that Δt_{eff} depends only on the parameters of $\omega_{0,1}$ as $\Delta\hat{t}_{\text{eff}}(\boldsymbol{\theta}) = a/(\omega + b\omega^2)$. In Figure 4.6, we plot the dependence of Δt (short for Δt_{eff}) on ω for the IBM Quantum device D *ibmq_ boeblingen*. The results for the other IBM Quantum devices A, B and C, are shown in Figure B.7 (Appendix B.4). For the IBM Quantum device D (*ibmq_ boeblingen*), the values are $a = 6.2774 \pm 0.01502$ and $b = 1.5086 \times 10^{-9} \pm 0.6104 \times 10^{-9}$s. How we arrived at this dependence is discussed later in Section 4.5. Using this data-driven model allows us to use reduce the number of parameters in the estimation and hence reduce the associated computational cost of the optimization. In the following sections, we will denote this model by $\Delta\hat{t}_{\text{eff}}(\boldsymbol{\theta})$.

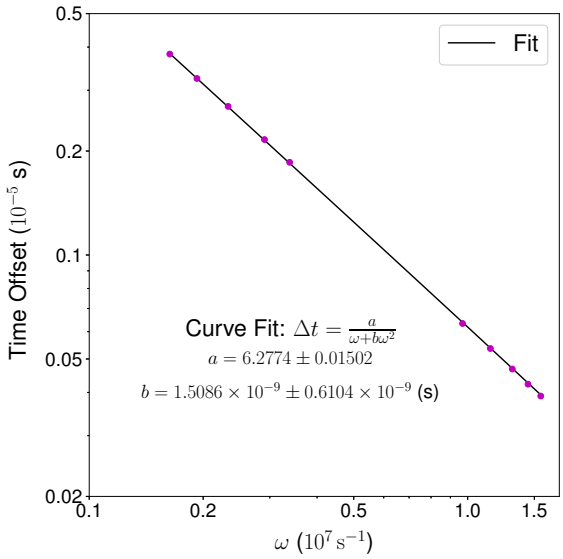


Figure 4.6: Dependence of the time offset Δt on parameters ω for IBM Quantum device D *ibmq_ boeblingen*. The plotted data points correspond to driving the device under different conditions and hence different cross-resonance Hamiltonians. The imperfect pulse shaping model extracted from these experimental data points is shown by a fit and this is later used in the MLE.

Decoherence

Recalling our discussion in Sec.4.2.2, we model decoherence as a depolarization channel acting on the quantum state $\rho(t) = \exp(-iHt)\rho(0)$ produced as a result of Hamiltonian

evolution for a time duration t. The resulting state (from Eq. 4.11) was given by

$$\mathcal{E}(\rho(t)) = (1 - p_d(t))\rho(t) + p_d(t)\frac{I}{2^n} \tag{4.47}$$

One approach to obtain a description of $p_d(t)$ is to assume the functional form $1-\exp(-(t-t_0)/\mu)$, resulting from a Poisson process with rate μ as described in Sec. 4.2.2 and then estimate μ from the training examples after incorporating this into the MLE. Here, we describe a model for $p_d(t)$ using the measured device properties of T_1 and T_2 times of each qubit.

We consider an independent noise model on each of the two qubits used for implementing a cross-resonance gate. Let us denote the amplitude damping and phase damping of kth qubit as $\mathcal{E}_{a,k}$ and $\mathcal{E}_{p,k}$ respectively. They have the following Kraus operators

$$\mathcal{E}_{a,k} : \left\{ \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma_{a,k}} \end{bmatrix}, \begin{bmatrix} 0 & \sqrt{\gamma_{a,k}} \\ 0 & 0 \end{bmatrix} \right\}, \quad \mathcal{E}_{p,k} : \left\{ \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma_{p,k}} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\gamma_{p,k}} \end{bmatrix} \right\} \tag{4.48}$$

where

$$\gamma_{a,k} = 1 - \exp\left(-\Gamma_{a,k} t\right), \quad \gamma_{p,k} = 1 - \exp\left(-\Gamma_{p,k} t\right) \tag{4.49}$$

with

$$\Gamma_{a,k} := \frac{1}{2\mathrm{T}_{1,k}}, \quad \Gamma_{p,k} := \frac{1}{\mathrm{T}_{\phi,k}} \tag{4.50}$$

where $\mathrm{T}_{1,k}$ is the T_1 time of the kth qubit and $\mathrm{T}_{\phi,k}$ is the pure dephasing rate of the kth qubit related to the T_2 time of the kth qubit as

$$\frac{1}{\mathrm{T}_{\phi,k}} = \frac{1}{\mathrm{T}_{2,k}} - \frac{1}{2\mathrm{T}_{1,k}} \tag{4.51}$$

The overall noise operator acting on the two qubits is then given by [Sun+20a]

$$\mathcal{E} = \otimes_{k=1}^{2} (\mathcal{E}_{a,k} \circ \mathcal{E}_{p,k}) \tag{4.52}$$

where $\circ$ indicates taking a composition of the two noise operators. The probability $p_d(t)$ can be based on the unitarity [Wal+15] of the noise operator $\mathcal{E}$ which is a completely positive linear map quantifying the coherence of the noise operator. The probability $p_d(t)$ is then given by

$$\begin{aligned} 1 - p_d(t) = 1 + \frac{1}{15}(&\gamma_{a,1}^2(3\gamma_{a,2}(\gamma_{p,2}-2) - 4\gamma_{p,2} + 7) \\ &+ 4\gamma_{a,1}(\gamma_{p,1}-2)(\gamma_{a,2}^2 + \gamma_{a,2}(\gamma_{p,2}-2) - \gamma_{p,2} + 2) \\ &+ \gamma_{a,2}^2(7 - 4\gamma_{p,1}) - 4\gamma_{a,2}(\gamma_{p,1}-2)(\gamma_{p,2}-2) + 4\gamma_{p,1}\gamma_{p,2} - 8\gamma_{p,1} - 8\gamma_{p,2}) \end{aligned} \tag{4.53}$$

This can be further generalized to n-qubit system (see Appendix G of [Sun+20a] for details). Let the measurement of Hamiltonian evolution for time t followed by this noisy depolarization channel be $\tilde{y}$ and the corresponding Rabi oscillation $\tilde{p}_{\text{rabi}}$. Note that the Rabi oscillations in this case are related to the noiseless case (from Eq. 4.42) as

$$\tilde{p}_{\text{rabi}} = (1 - p_d(t))p_{\text{rabi}} \tag{4.54}$$

In Table 4.2, we give the RMSE betweeen Rabi oscillations obtained using different decoherence models assuming we know the true Hamiltonian parameters and Rabi oscillations inferred from data. For the *single parameter model*, we consider a prefactor of $(1-p_d(t)) = \exp(-(t-t_0)/\mu)$ with the single parameter μ on the Rabi oscillations obtained using no decoherence model. This parameter μ is then estimated from the data and is found to be $(7.75 \pm 0.91) \times 10^{-5}$s. In the *two parameter model*, this is allowed to vary with the state preparation operator U being applied. For $U = \sigma_I \otimes \sigma_I$, we have $\mu = (5.52 \pm 0.82) \times 10^{-5}$s and for $U = \sigma_X \otimes \sigma_I$, $\mu = (2.51 \pm 0.59) \times 10^{-5}$s. It is advantageous to use such models due to the low number of parameters present and when the T_1 or T_2 times of the different qubits are not available. However, if these times are available, one can use the two-qubit decoherence model or the two parameter model as we note from the values of RMSE in Table 4.2.

Decoherence Model	**Root Mean Squared Error**	**KL Divergence**
No decoherence	0.109321	0.012370
Single-qubit decoherence model	0.095903	0.007408
Single Parameter Model	0.091644	0.006515
Two-qubit decoherence model	0.090953	0.006392
Two Parameter Model	0.084289	0.005376

Table 4.2: Comparison of different decoherence models in fitting Rabi oscillations inferred from experimental data collected from IBM Quantum Device D. Kullback-Leibler divergence (KL divergence) is computed as $D_{KL}(p_{\text{data}}||p_{\text{model}})$ where p_{data} is the probability inferred from data and p_{model} is that predicted from the model.

4.4.4 Estimation Procedures for Learning CR Hamiltonians

We have so far described the likelihood of different measurement outcomes given different queries (Sec. 4.4.2) to the two-qubit cross-resonance Hamiltonian (Eq. 4.30), and the different noise sources or non-idealities on the IBM Quantum devices. We now describe our estimation procedures for estimators described earlier in Section 4.2.1, which utilizes this likelihood function for estimating the CR Hamiltonian parameters from the training data $D = (X, Y)$ generated during learning. While all the estimators described here can be adopted with HAL algorithms (Section 4.3), it should be noted that the best performance is expected with the maximum-likelihood estimator (MLE) due to the query criteria used in HAL-FI/HAL-FIR.

Regression

We perform regression on Rabi oscillations (Eq. 4.42) (which is just the difference in likelihood for $y = 0$ and $y = 1$) inferred from the measurement outcomes. Let us consider the parameterization of $\mathbf{\Lambda}$. This is more useful to work with as the frequencies of oscillation in e^{-iHt} are determined by ω_0 and ω_1 unlike $\mathbf{J}$ where it is determined by all the parameters. Instead of directly performing regression as described in Section 4.2.1, we divide our estimation procedure into multiple steps. We create initial estimates of $\omega_{0,1}$ independent of the other parameters through Fast Fourier Transform (FFT) before performing a full regression.

We compute the Rabi oscillations $\hat{p}_{\text{rabi}}$ for each query x through the corresponding measurement outcomes by solving the optimization problem of Eq. 4.39. Initial estimates of the parameters $\omega_{0,1}$ are then obtained by applying a Discrete (Fast) Fourier Transform to the Rabi oscillations. These initial estimates are then refined by fitting regression equations of the form $A\cos(\omega t) + B\sin(\omega t) + C$ to the Rabi oscillations, where the fit minimizes the total L_2 error, the coefficients A, B, and C for each Rabi oscillation are estimated using linear least-squares regression, and a bracketed gradient-based search is performed to refine the estimates of $\omega_{0,1}$. The corresponding optimization problem can be framed as minimizing the following residual error

$$\min E(\mathbf{A}, \omega_{0,1}) = \sum_{t \in \mathcal{T}} \left(p_{\text{rabi}}(t) - \mathbf{A}\boldsymbol{\Omega}(\omega t)\right)^2, \tag{4.55}$$

where the coefficients $\mathbf{A} = (A, B, C)$ are known functions of $\delta_{0,1}, \phi_{0,1}$ through the analytical forms of the Rabi oscillations for the query space considered (see Eq. 4.42) and $\Omega(\omega t)$ is a vector of cosines and sines (fully described in Appendix B.3). Thus, we can then obtain estimates for $\delta_{0,1}, \phi_{0,1}$ from the A, B, and C coefficients of the regression equations for each of the Rabi oscillations. Finally, we obtain an estimate $\hat{\mathbf{\Lambda}}$ by fixing the values of $\omega_{0,1}$ and carrying out a gradient descent procedure using the same cost function. If desired, we output the Hamiltonian parameter estimate $\hat{\mathbf{J}}$ by transforming $\hat{\mathbf{\Lambda}}$ appropriately (Eq. 4.32).

Maximum-Likelihood Estimation

The MLE parameter estimate $\hat{\boldsymbol{\theta}}$ obtained through solving Eq. 4.5 for the CR Hamiltonian incorporating the Hamiltonian model description and presence of different noise sources is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -\frac{1}{N}\sum_{k=1}^{N} \log\left[\sum_{y\in\{0,1\}}\left(p_{\tilde{y}|y}(\tilde{y}^{(k)}|y)\sum_{z\in\{0,1\}}\left[\left(1-p_d\left(t^{(k)}\right)\right)\left|\left\langle yz|M^{(k)}e^{-iH(\boldsymbol{\theta})(t^{(k)}+\Delta\hat{t}_{\text{eff}}(\boldsymbol{\theta}))}U^{(k)}|00\right\rangle\right|^2 + \frac{1}{4}p_d\left(t^{(k)}\right)\right]\right)\right], \tag{4.56}$$

or

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -\frac{1}{N}\sum_{k=1}^{N} \log\left[\sum_{y\in\{0,1\}}\left(p_{c|y}(c^{(k)}|y)\sum_{z\in\{0,1\}}\left[\left(1-p_d\left(t^{(k)}\right)\right)\left|\left\langle yz|M^{(k)}e^{-iH(\boldsymbol{\theta})(t^{(k)}+\Delta\hat{t}_{\text{eff}}(\boldsymbol{\theta}))}U^{(k)}|00\right\rangle\right|^2 + \frac{1}{4}p_d\left(t^{(k)}\right)\right]\right)\right], \tag{4.57}$$

depending on how the readout noise is modeled. The imperfect pulse-shaping model $\Delta\hat{t}_{\text{eff}}(\boldsymbol{\theta})$ was discussed in Section 4.4.3 and $p_d(t)$ is the probability of depolarization associated with the two-qubit decoherence model (Section 4.4.3). The choice of the MLE (Eq. 4.56 or Eq. 4.57) for each IBM Quantum device will be specified in Section 4.5.

The MLE problem (Eq. 4.56 and Eq. 4.57) in either of the two different parameterizations of $\mathbf{J}$ and $\mathbf{\Lambda}$ is nonlinear and non-convex. An example of the energy landscape of the log-likelihood loss function for the IBM Quantum device *ibmq_ boeblingen* is shown in Figure B.6 of Appendix B.3. The presence of multiple local minima makes the MLE problem in general challenging to solve. To ensure that we converge to the global minimum and do not get

stuck in a local minimum during estimation, we divide the estimation into multiple stages. In the first stage, we obtain an initial estimate which is refined over subsequent stages. The estimation procedure is summarized in Algorithm 6.

During learning, this initial estimate may be available from the previous round. When such an initial estimate is not available, we use the estimation based on regression as discussed earlier in Section 4.4.4, and summarized in lines 1-6 of Algorithm 6. This initial estimate is then used as an initial condition to the MLE solve (Eq. 4.56). In Algorithm 6, we have denoted the negative log-likelihood loss function which appears in the optimization problem of Eq. 4.56 as L. We solve the MLE problem using the optimizer of stochastic gradient method of ADAM [KB14] which encourages getting out of any local minima. The parameter estimate produced by ADAM is then refined using the second-order quasi-Newton method of L-BFGS-B [Zhu+97]. The computational complexity of our MLE estimation procedure is dominated by ADAM. This motivates us to directly use L-BFGS-B for estimation for latter batches during learning. The full description, computational details and extensions of the estimation procedure is given in Appendix B.3.

Algorithm 6 MLE Estimation Procedure for Hamiltonian Learning

Input: Training examples of size m: $D = \{(X, Y)\} = \{(x_i, y_i)\}_{i\in[m]}$, initial condition $\hat{\mathbf{J}}_0$ (optional)
Output: $\hat{\mathbf{J}}$

1: **if** no input of $\hat{\mathbf{J}}_0$ **then** ▷ get an initial estimate through regression
2: Obtain Rabi oscillations p_{rabi} from D by solving Eq. 4.39
3: $\omega_{0,1} \leftarrow \text{FFT}(p_{\text{rabi}})$ ▷ FFT of Rabi oscilations p_{rabi} over $\mathcal{T}$, see App. B.3 for details
4: $(\delta_{0,1}, \phi_{0,1}) \leftarrow p_{\text{rabi}}(\Omega(\omega_{0,1}t))^{-1}$ ▷ Regression on Rabi oscillations given $\omega_{0,1}$
5: Refine estimate of $\hat{\mathbf{\Lambda}}_0$ through gradient descent on $E(\mathbf{A}, \omega_{0,1})$ in Eq. 4.55
6: Get $\hat{\mathbf{J}}_0$ by transforming $\hat{\mathbf{\Lambda}}_0$ (Eq. 4.32)
7: **end if**
8: $\hat{\mathbf{J}} \leftarrow \arg\min_{\mathbf{J}} L(\mathbf{J}; X, Y)$ ▷ using $\hat{\mathbf{J}}_0$ as a guess
9: **return** $\hat{\mathbf{J}}$

Bayesian Estimation

For Bayesian estimation, we use a particle filtering algorithm [DJ+09], also called sequential Monte Carlo (SMC) algorithm. In SMC, a distribution (say $p(r)$) is discretely approximated with a distribution that has support only over a finite number of points (say n_p) called particles:

$$p(r) = \sum_{k=1}^{n_p} w_k \delta(r - r_k), \tag{4.58}$$

where w_k is the weight of the kth particle, and r_k is the location of the kth particle. Note that the vector of weights (w_k) can also be thought of as a vector of probabilities, satisfying $0 \leq w_k \leq 1$ and $\sum_{k=1}^{n_p} w_k = 1$. An particle filter is then specified by weights and locations $\{(w_k, r_k)\}_{k\in[n_p]}$ over the set of n_p particles. This can be used to compute the expectation value of a function g as $\sum_{k=1}^{n_p} w_k g(r_k)$.

For learning the CR Hamiltonian, we will be interested in tracking the distribution $p(\mathbf{J}|D)$ of the Hamiltonian parameters given the training data D being collected. It is then beneficial

to think of the particle filter as $\{(w_k, \mathbf{J}_k)\}_{k\in[n_p]}$. Considering the prior distribution as $p(\mathbf{J})$, the initial weights of the particles can be set as $w_k = 1/n_p$ and the locations as random samples $\mathbf{J}_k$, sampled from this prior. With this prior distribution, we can then carry out Bayes rule (Eq. 4.7) to compute the posterior distribution $p(\mathbf{J}|D)$ over the Hamiltonian parameters $\mathbf{J}$ described by particles $\{(w'_k, \mathbf{J}'_k)\}_{k\in[n_p]}$ by setting $\mathbf{J}'_k = \mathbf{J}_k$ and

$$w'_k = \frac{w_k p(Y|X, \mathbf{J}_k)}{\sum_k p(Y|X, \mathbf{J}_k)}. \tag{4.59}$$

In practice, iteratively carrying out Baye's rule as above leads to numerical instabilities due to limited resolution and weights shrinking to zero. Increasing the number of particles n_p obviously delays this while improving the accuracy of the evaluation of expectations of functions using the particles and posterior distributions in the SMC algorithm. Stability is ensured by resampling techniques which adaptively changes locations $\mathbf{J}_k$ of particles to higher weight regions. In our implementation, we use the Liu-West algorithm [LW01].

Finally, we can output a point estimate from the posterior particles

$$\hat{\mathbf{J}} = \sum_{k=1}^{n_p} w'_k \mathbf{J}'_k \tag{4.60}$$

which is the mean of the posterior distribution (Section 4.2.1) and minimizes the RMSE.

We use the implementation of SMC in the Qinfer package [Gra+17] for our purposes. We thus refer the reader to [Gra+12; Gra+17] for details on the SMC algorithm used for Hamiltonian learning. We specify the prior distribution $p(\mathbf{J})$ and the number of particles n_p chosen for different IBM Quantum devices and learners in Section 4.5.

4.4.5 Implementation of HAL Algorithm for Learning CR Hamiltonians

We now discuss the details of the HAL-FI algorithm implemented for learning the CR Hamiltonian on the IBM Quantum devices. As the implementation of HAL-FIR for this application is very similar, it is omitted. Moreover, we will consider the MLE estimator here but the other estimators (Section 4.4.4) can also be adopted.

In our implementation of the HAL-FI algorithm in experiments on learning the CR Hamiltonian, we consider the following inputs. The initial query space which may be changed during the course of the HAl-FI algorithm during training is $\mathcal{Q}^{(0)} = \mathcal{M} \times \mathcal{U} \times \mathcal{T}^{(0)}$ with $\mathcal{M}$ and $\mathcal{U}$ as described in Section 4.4.2. Here, we explicitly denote the superscript on $\mathcal{T}$ which is initially set to $\mathcal{T}^{(0)}$ but may change during training if an adaptive query space strategy is employed. We set $\mathcal{T}^{(0)}$ to be the 81 equispaced times in the interval $[10^{-7}, 6 \times 10^{-7}]s$. We consider the initial number of queries as $N^{(0)}_{\text{tot}} = 2430$ (or five times the number of different queries in the initial query space $\mathcal{Q}^{(0)}$), a constant batch size $N_b = 486$ (or the number of queries in the initial query space $\mathcal{Q}^{(0)}$), and the initial query distribution $q^{(0)}$ as the uniform random distribution over $\mathcal{Q}^{(0)}$.

The initial set of training examples $(X^{(0)}, Y^{(0)})$ are obtained by sampling $X^{(0)}$ from $\mathcal{Q}^{(0)}$ with respect to $q^{(0)}$ in Line 1 of Algorithm 3 and collecting the corresponding set of

measurement outcomes $Y^{(0)}$ through queries to the CR Hamiltonian on the IBM Quantum device in Line 2. The set of training examples $(X^{(i)}, Y^{(i)})$ is progressively increased (Line 10) during learning by adding N_b queries $X_q^{(i)}$ sampled from query distribution $q^{(i)}$ chosen by HAL-FI in Line 6 and collecting the corresponding measurement outcomes $Y_q^{(i)}$ in Line 9. The learning is continued until our query budget is expended or the desired learning error is achieved.

To determine the initial parameter estimate $\hat{\boldsymbol{\theta}}^{(0)}$ from the initial set of training examples (Line 4) and subsequent $\hat{\boldsymbol{\theta}}^{(i)}$ from $(X^{(i)}, Y^{(i)})$ (Line 11), we solve the MLE (Eq. 4.5) for the CR gate incorporating the Hamiltonian model description and presence of different noise sources.

The parameter estimates $\hat{\boldsymbol{\theta}}$ obtained after solving the MLE problem are used by HAL-FI to construct the Fisher information matrices based on the model and obtain the query distribution $q^{(i)}$ by solving the SDP program of Eq. 4.23 in Line 6. The expressions for the Fisher information matrices for different queries (Section 4.4.2) considering the CR Hamiltonian and noise models are given in Appendix B.1. The query space $\mathcal{Q}^{(i)}$ used in Lines 6 and 7 depend on the querying strategy and hence the learning scenario we consider.

As described in Section 4.3.1, we can define four different learning scenarios based on the presence of an active learner and how the query space is adaptively changed during learning. In passive learning, an active learner is not present and we set the query distribution to the uniformly random distribution over $\mathcal{Q}$. In the case of active learning with fixed query space, the query space remains fixed during training i.e., $\mathcal{Q}^{(i)} = \mathcal{Q} \forall i$, and the query distribution is determined by solving Eq. 4.23 using the current estimate of the parameters $\hat{\boldsymbol{\theta}}$ over this fixed query space. When considering active learning with an adaptively growing query space, we consider two different situations on the basis of how the query space is changed between batches during learning. We consider two cases: (i) $\mathcal{Q}^{(i)}$ grows linearly by linearly increasing the $\mathcal{T}^{(i)}$ between batches and (ii) $\mathcal{Q}^{(i)}$ grows exponentially by doubling the allowed set of system interaction time $\mathcal{T}^{(i)}$. The query distribution is then determined by solving the corresponding SDP problem of Eq. 4.23 over the query space $\mathcal{Q}^{(i)}$ corresponding to the ith batch. These different learning scenarios for HAL-FI are summarized in Table 4.3.

Learning Scenario	**Query Space**	**Query Distribution**
Passive Learning	Fixed	uniformly random
Active Learning with Fixed Query Space	Fixed	q through Eq. 4.19
Active Learning with Adaptive Query Space I	Linearly growing $\mathcal{T}$	q through Eq. 4.19
Active Learning with Adaptive Query Space II	Exponentially growing $\mathcal{T}$	q through Eq. 4.19

Table 4.3: Summary of different learning scenarios

4.5 Results

We now present the results of the experiments described in Section 4.4 and show that they support the claims made in Section 4.2.1. In this section, we compare the active learner HAL introduced in Section 4.3 against the passive learner (Section 4.2.3) equipped with different estimators (Section 4.4.4) and the sequential active learner of Qinfer (Section 4.2.3)

in learning the CR Hamiltonian through experiments on different IBM Quantum devices. We finally analyze the results of these experiments to comment on the scaling behavior, classical computational cost and query advantage of the different learners.

In Section 4.5.1, we first describe the datasets used for Hamiltonian learning, hyperparameters of the estimators (Section 4.4.4) used in experiments, and the experimental protocol for evaluating the performance of the learners. In Section 4.5.2, we describe the implementation of the sequential active learner of Qinfer which is also used as a baseline for assessing the performance of HAL-FI. In Section 4.5.3, we show results of the learners on these different datasets. In Section 4.5.4, we describe how HAL-FI can be used to achieve Heisenberg (or super-Heisenberg) limited scaling and evaluate the query advantage of HAL-FI over the baseline considering different learning scenarios.

4.5.1 Data and Experiment Protocol

In this section, we describe the different kinds of datasets that were used in assessing the performance of the HAL-FI algorithm and how they were collected. We then summarize the parameters of the cross-resonance Hamiltonian and the noise sources discussed in Section 4.4.3 for the different IBM Quantum devices (Section 4.4.2).

Datasets from IBM Quantum Devices

The different datasets that we use for Hamiltonian learning are a combination of experimental data collected from the IBM Quantum devices described in Section 4.4.2 and that collected from a *simulator* which we will describe in Section 4.5.1.

Experimental data is collected from the different IBM Quantum devices according to the query space described in Section 4.4.2. The set of evolution times $\mathcal{T}$ is set to 81 equispaced times in the interval $\mathcal{T} = [10^{-7}, 6 \times 10^{-7}]s$. For IBM Quantum devices A, B, and C, there are 200 measurement outcomes (or shots) for each query $\mathbf{x} \in \mathcal{Q}$. For IBM Quantum device D *ibmq_boeblingen*, there are 512 measurement outcomes for each query. Recall from our discussion of the Hamiltonian learning framework in 4.2.1 and HAL algorithm in 4.3.2, the outputs of our queries are not expectation values but rather single shot readouts of the target qubit.

The experimental data is then collected and made available as an offline dataset that the active learner can query. Unlike deploying an active learner in real-time where a particular query can be made to the system unlimited number of times, using experimental datasets imposes the additional constraint of the number of times a query can be made by the active learner due to the limitation on the number of measurement outcomes available for each query. In Appendix B.2, we discuss how we handle this constraint during query optimization.

Parameters of the CR Hamiltonian and Noise Sources for IBM Quantum Devices

Considering the entire collected experimental datasets for each IBM Quantum device (Section 4.4.2) as training data, we compute the Hamiltonian parameters $\mathbf{J}/\mathbf{\Lambda}$, and that of the different noise sources using the estimation procedure specified in Section 4.4.4 for the MLE. We solve the MLE of Eq. 4.56 for IBM Quantum device D which has very low readout

noise and use MLE of Eq. 4.57 for the other IBM Quantum devices. A summary of the estimated parameters of the CR Hamiltonian and noise sources on IBM Quantum device D *ibmq_boeblingen* is shown in Table 4.4 under different drive configurations. We summarize the estimated parameters for the other devices in Appendix B.4.

Drive	**CR Amp.**	**Hamiltonian Parameters [$\times 10^6 s^{-1}$]**		**Noise: Readout (L) & Time Offset (R)**	
Config.	**(arb. units)**	$\mathbf{J} = (J_{IX}, J_{IY}, J_{IZ}, J_{ZX}, J_{ZY}, J_{ZZ})$	(ω_0, ω_1)	(r_0, r_1)	$(\Delta t_{\mathbf{eff},0}, \Delta t_{\mathbf{eff},1})$ **[ns]**
1	0.24	(-3.88, -1.08, -0.24, 5.44, 1.07, 0.21)	(1.57, 9.58)	$(0.012, 0.025)$	$(1965, 289)$
2	0.30	(-4.57, -1.47, -0.29, 6.50, 1.39, 0.41)	(1.94, 11.45)	$(0.0078, 0.033)$	$(1581, 226)$
3	0.36	(-5.12, -1.65, -0.23, 7.52, 1.66, 0.33)	(2.40, 13.07)	$(0.0078, 0.035)$	$(1267, 203)$
4	0.42	(-5.42, -1.95, 0.37, 8.38, 1.90, 0.07)	(2.97, 14.33)	$(0.0078, 0.039)$	$(1016, 182)$
5	0.48	(-5.72, -2.13, 0.03, 9.20, 2.15, 0.11)	(3.48, 15.51)	$(0.0078, 0.023)$	$(862, 166)$

Table 4.4: Summary of estimated CR Hamiltonian parameters for the IBM Quantum device D *ibmq_boeblingen* with different drive configurations (Config.) corresponding to amplitude (Amp.) of CR pulse. We give the Hamiltonian parameters in the parameterization $\mathbf{J}$ and the physically relevant frequency components in $\mathbf{\Lambda}$. The readout noise is defined by the parameters of r_0 and r_1 which are the conditional probabilities of bit flip given the measurement outcomes are $y = 0$ and $y = 1$ respectively (see Section 4.4.3).

Datasets from Simulation

The experimental datasets in Section 4.5.1 contain noise sources other than those modeled even if negligible and may not span a long enough time range over $\mathcal{T}$ for testing different learning scenarios. In order to understand the behavior of the HAL-FI and HAL-FIR algorithms considering all the noise sources are known, we set up a simulator. The advantage of using a simulator over experimental data is that it allows us to assess the limits of the performance of the HAL-FI algorithm in the presence or absence of different noise sources such as decoherence. Studies carried out on the simulator also allow us to test the robustness of the active learner in the presence or absence of different noise sources.

The simulator imitates the different quantum devices but where all the different noise sources are known and perfectly modeled, and which we can query. The Hamiltonian of the simulator is set to that learned from the full set of training examples contained in an experimental dataset collected from a particular quantum device. Thus, we can have simulators for each of the IBM Quantum devices A, B, C, and D, under different drive configurations. All the modeled noise sources of readout noise, imperfect pulse-shaping, and decoherence as described in Section 4.4.3 are included. We apply HAL-FI in real-time on the simulator as there is no limitation on the number of times we select a particular query $\mathbf{x} \in \mathcal{Q}$.

In Figure 4.7, we compare Rabi oscillations computed from a subset of the experimental data collected from IBM Quantum device D *ibmq_boeblingen* under drive configuration 2, and training examples generated on the corresponding simulator. A set of 46800 training examples are generated from both the experimental data and simulator assuming a uniform query distribution over the query space. These training examples are then used to learn Hamiltonian parameters in each case. The small difference in the predicted model Rabi oscillations, computed using the analytical expressions of Eq. 4.42 and the Hamiltonian

parameters learned from these two different datasets, further indicates that the simulator faithfully represents the collected experimental data.

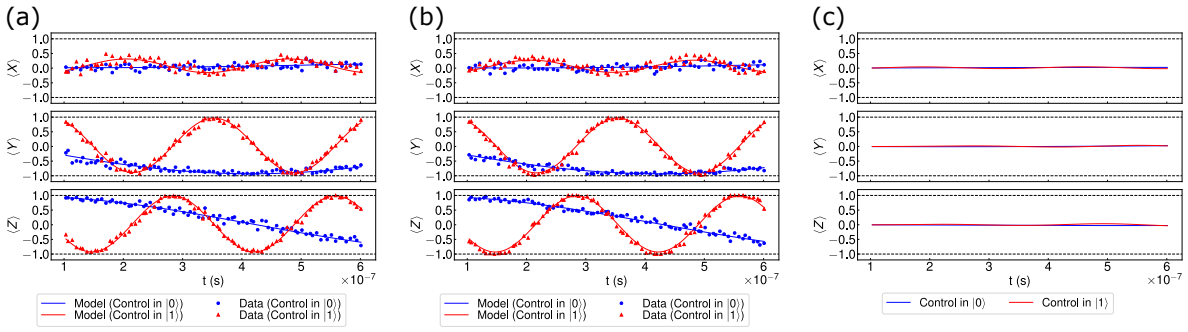


Figure 4.7: Comparison of Rabi oscillations computed from data (46800 queries) of IBM Quantum device D *ibmq_boeblingen* from the (a) collected experimental data, and (b) simulator. In (c), we plot the difference in the model Rabi oscillations predicted from the Hamiltonians learned in (a) and (b). In the subplots of (a)-(c), we plot Rabi oscillations (or difference) for different measurement operators M (rows), preparation operators U (colors), and evolution times t (x-axis), corresponding to the query space described in Section 4.4.2.

Hyperparameters of Estimators for Learning CR Hamiltonians

Among the different estimators (Section 4.4.4) we use for Hamiltonian learning, some of them require the specification of some hyperparameters. In our estimation procedure for MLE (Section 4.4.4), we set the learning rate for the ADAM solve as 10^{-3} and use only the quasi-Newton method after 50 batches. In the Bayesian SMC method (Section 4.2.1), we use n_p particles to track the distribution over $p(\tilde{\mathbf{J}}|D)$ where $\tilde{\mathbf{J}}$ is the normalized version of the Hamiltonian parameter vector $\mathbf{J}$ by 10^6. We choose this normalization factor as this is the order of magnitude expected for these parameters under these drive conditions [MG20]. We then set the prior distribution over the parameters as $p(\tilde{\mathbf{J}})$ as an uniform distribution over $[-10, 10]^6$. On the simulator, we set $n_p = 10^4$ and on the experimental data, we set $n_p = 10^5$. These values were obtained by increasing the number of particles on each oracle until we saw the expected convergence behavior with a passive learner equipped with the SMC method as an estimator.

Protocol for Comparing Performance of Hamiltonian Learning Methods

Query complexity is used for comparing the performance of different learners on the simulator or oracle with access to an experimental dataset. In particular, our main goal is to extract the scaling of the query complexity with respect to the root mean square error (RMSE) of Hamiltonian parameters $\mathbf{J}$. On the simulator as we have access to the truth $\mathbf{J}^\star$, we can compute the RMSE directly. On the experimental data, we use the empirical RMSE,

computed at each round of learning as

$$\text{RMSE} = \sum_{i=1}^{m} \mathbb{E}\left[\left(\frac{\hat{\theta}_i}{\xi_i} - \frac{\theta_i^\star}{\xi_i}\right)^2\right]^{1/2} \approx \frac{1}{N_{\text{runs}}} \sum_{i=1}^{m} \sum_{k=1}^{N_{\text{runs}}} \left[\left(\frac{\hat{\theta}_i^{(k)}}{\xi_i} - \frac{\theta_i^\star}{\xi_i}\right)^2\right]^{1/2}, \tag{4.61}$$

where we approximate the expectation by an average over parameter estimates from N_{runs} runs and the true parameter values $\boldsymbol{\theta}^\star$ by the mean of these runs. The normalization factors ξ_i are selected to be $10^6\,\text{s}^{-1}$ for all i as the components of $\mathbf{J}$ with highest magnitude are expected to the order of $10^{6\pm1}$ [MG20] for these IBM Quantum devices under these drive configurations. We implement the following experimental protocol. For each of the quantum devices described in Section 4.4.2, we compute the empirical RMSE for the learners from 200 independent runs of the simulator and 500 independent runs on the experimental dataset. These number of runs on each oracle were required to obtain accurate scalings of trends in RMSE with number of queries and ensure the uncertainty (or two standard deviations) of each scaling was at most 10%. In each run, we carry out the Hamiltonian learning algorithm for the different learning scenarios as detailed in Section 4.4.5 and summarized in Table 4.3. Additionally, we track the testing error of the learner with number of queries N. The so obtained trends are used to comment on the robustness of the estimation procedure used for MLE (Section 4.4.4) and the benefits of using the active learner HAL-FIR for making predictions of queries to the Hamiltonian (Problem 4.2.1) over a baseline. The testing error is computed empirically as well on a testing dataset collected from the simulator or experimental dataset using p_{test}.

4.5.2 Sequential Active Learner: Qinfer

In this section, we describe the sequential active learner of Qinfer [Gra+12; Gra+17] that was earlier introduced in Section 4.2.3 and how it is used in our numerical experiments on learning the CR Hamiltonian.

The sequential active learner in [Gra+12; Gra+17] which we refer to simply as Qinfer [1] uses Bayes risk (Eq. 4.26) as a query criteria. In our implementation, the Bayes risk is computed by running a *hypothetical update* over a set of *risk* particles in SMC (Section 4.4.4) for each query in the query space $\mathcal{Q}$. These risk particles are a separate set of particles than those used to track the distribution $p(\boldsymbol{\theta}|D)$ in SMC. The weights and locations of these risk particles are set to the same values as those of the particles partaking in estimation however, before computation of the Bayes risk over all queries and after every estimation step. Further details of Bayes risk computation can be found in [Gra+17].

For a fair and systematic comparison of Qinfer against HAL-FI on the simulator and experimental data, we run Qinfer sequentially over batches of queries. Running Qinfer sequentially over $\sim 10^5$ shots/queries would be computationally very expensive, as we would need to update the particles in SMC after collecting just one measurement outcome and then compute Bayes risk over all the queries in $\mathcal{Q}$. Instead, we issue a batch of N_b queries, all of which are the same query and the one with the lowest Bayes risk. The corresponding

[1] In [Gra+12; Gra+17], the package for statistical inference altogether is called Qinfer and the sequential active learner is but just a part of it.

measurement outcomes are then used to update the particles in SMC before computing Bayes risk using the updated risk particles and issuing another batch of queries.

On experimental data, we also need to ensure that query constraints are respected due to limited number of measurement outcomes available for each query. We do this by breaking down a batch of queries into mini-batches if enough number of measurement outcomes is not available for the request query with the lowest Bayes risk. Further details are given in Appendix B.2.

Finally in Qinfer, we use the same prior distribution and same number of particles as in the case of the passive learner equipped with the Bayesian SMC estimator (Section 4.4.4). On the simulator, we set $n_p = 10^4$ and on the experimental data we set $n_p = 10^5$.

4.5.3 Performance of Hamiltonian Learning Methods

In this section, we assess the performance of the different Hamiltonian learning methods introduced so far using the protocol described in Section 4.5.1 in tackling the learning problems posed in Section 4.2.1. We report results of the different Hamiltonian learning scenarios (Table 4.3): (i) passive learner with estimation based on FFT and linear regression, (ii) passive learner with an estimation procedure to solve the MLE problem, (iii) active learner in fixed query space, and (iv) active learner in an adaptively growing (linearly) query space. Results of the exponentially growing query space are postponed to Section 4.5.4 where we comment on the achievability of Heisenberg limited scaling. We use the HAL-FI algorithm which was discussed in Section 4.2.3 in the latter active learning approaches. We also show results of learners with different estimations based on FFT and linear regression (Lin. Reg.), maximum-likelihood (MLE), and Bayesian estimation using the sequential Monte Carlo (SMC) method. We additionally compare HAL-FI which proposes batches of queries during learning against the sequential active learner Qinfer (Section 4.5.2) which proposes queries one at a time using Bayes risk as a query criteria.

We present the convergence behavior of each algorithm during learning under different regimes. For each scenario, we show trends of learning error (RMSE) with number of queries. These trends indicate the performance of each learning algorithm, culminating in evidence of query advantage. For brevity of presentation, we focus on the results obtained from IBM Quantum Device D (*ibmq_boeblingen*) under the drive configuration 2.

We focus on the Hamiltonian learning task of model inference (4.2.1) and touch upon prediction against a testing distribution (4.2.1) to show that HAL is a general framework for tackling both problems when equipped with the appropriate query optimization. We show results on the simulator for two different cases of time range $\mathcal{T}$ of the query space $\mathcal{Q}$. The time range $\mathcal{T}$ should be chosen such that you can estimate the Hamiltonian parameters of interest $\hat{\boldsymbol{\theta}}$. For a given $\mathcal{T}$, the non-zero frequencies that can be successfully detected using FFT [Coh89; PBV15] range from $1/\Delta T$ to $F_s/2 - 1/\Delta T$ at increments of $1/\Delta T$ where ΔT is the length of the time interval $\mathcal{T}$ and F_s is the sampling rate. F_s is decided according to the Nyquist criterion to ensure that aliasing doesn't occur. Accordingly, ΔT must be sufficiently long to see a single cycle of the sinsusoid corresponding to the lowest frequency. We call this as the minimum-frequency criteria (MFC). Note that when using other estimation methods based on MLE or Bayesian estimation methods, $\mathcal{T}$ is not required to satisfy these properties.

We thus show results on the simulator where $\mathcal{T}$ is such that the MFC for frequency

estimation (using FFT) is satisfied and when MFC is not necessarily satisfied. Moreover, in Appendix B.3, we describe how standard FFT can be modified to detect lower frequencies than MFC allows. This is a buildup to our comparison on the experimental dataset where $\mathcal{T}$ is such that MFC is not necessarily satisfied. In the following results, we plot the learning error (RMSE/testing error) versus number of queries made on a log-log scale so that the slope s of the plotted lines can be directly interpreted as the scaling of learning error with complexity $\epsilon \sim N^s$.

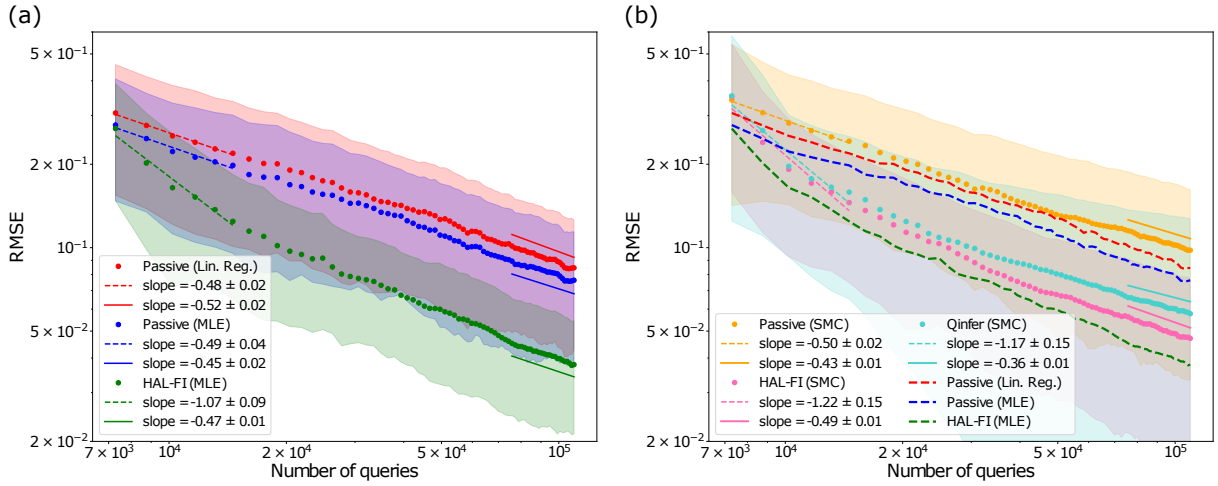


Figure 4.8: Scaling of RMSE with number of queries for different learners on a simulator when MFC is satisfied. In (a), we compare the trends in RMSE of HAL-FI against the passive learner with estimations based on MLE and linear regression (Lin. Reg.). In (b), we compare the trends in RMSE of HAL-FI using the Bayesian estimator of sequential Monte Carlo (SMC) against HAL-FI using MLE, Qinfer using Bayes risk as the query criteria and passive learners equipped with different estimators. In (a)-(b), estimators of each learner are indicated in brackets. Slopes indicate the scaling of RMSE with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner. In (b), we do not show slopes and errors for learners shown earlier in (a) on the left, and their trends are shown for visual reference.

Simulator and Minimum-Frequency criteria is satisfied We consider the query space as defined in Section 4.4.2 with $\mathcal{T}$ set to be the 243 equispaced times in the interval of $[10^{-7}, 18 \times 10^{-7}]$s. A comparison of different learners for Hamiltonian learning considering this query space is shown in Figure 4.8. For the passive learners (with any estimator), we observe a scaling of $\epsilon \sim 1/\sqrt{N}$ or $N \sim \epsilon^{-2}$ in RMSE with number of queries. This is in agreement with the SQL. The approximately constant gap between the passive learner when using an estimator based on MLE and the passive learner when using an estimator based on linear regression corresponds to a constant query reduction. This illustrates how a query advantage can be obtained by changing estimation for Hamiltonian learning.

In Figure 4.8(a), we observe two different scalings for the HAL-FI algorithm with MLE, an initial scaling which is higher than SQL and similar to Heisenberg limited scaling, and a scaling of SQL in the asymptotic query regime. This shows that depending on the desired

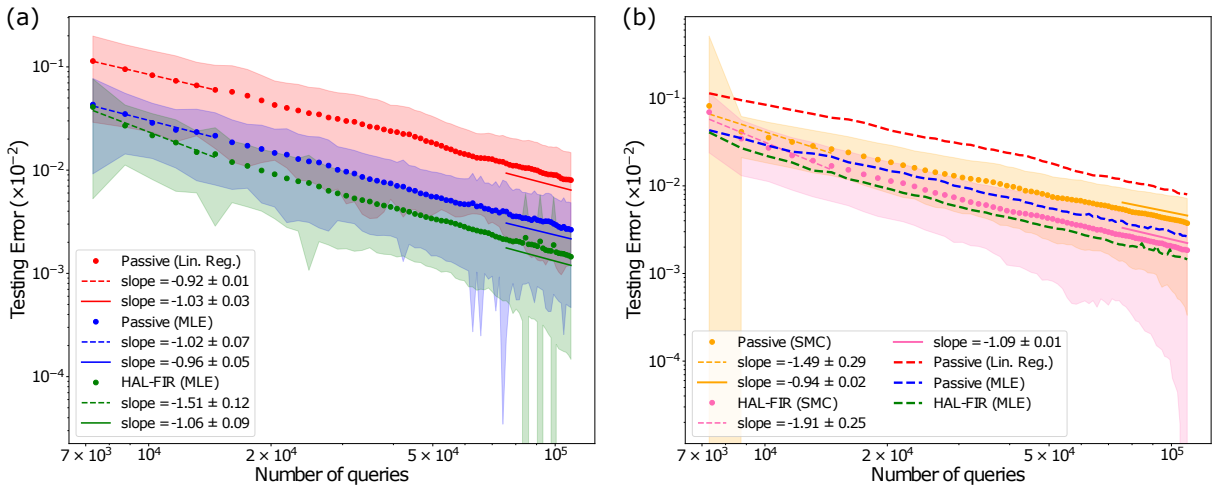


Figure 4.9: Scaling of testing error with number of queries for different learners on a simulator when MFC is satisfied. In (a), we compare the trends in testing error of HAL-FIR against the passive learner with estimations based on MLE and linear regression (Lin. Reg.). In (b), we compare the trends in testing error of HAL-FIR using the Bayesian estimator of sequential Monte Carlo (SMC) against HAL-FIR using MLE, and passive learners equipped with different estimators. In (a)-(b), estimators of each learner are indicated in brackets. Slopes indicate the scaling of testing error with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner. In (b), we do not show slopes and errors for learners shown earlier in (a) on the left, and their trends are shown for visual reference.

learning error, we can expect to see a higher rate of convergence. However, asymptotically the number of queries N required by the active learner HAL-FI is a constant fraction of that required by the passive learner.

We can also consider HAL-FI using Bayesian estimation for Hamiltonian learning. In Figure 4.8(b), we show results comparing learners with the Bayesian estimator SMC, against learners using alternate estimation. Among active learners, we show HAL-FI alongside Qinfer using Bayes risk as a query criteria. Firstly, we note that HAL-FI with SMC outperforms all other learners equipped with the Bayesian estimator including Qinfer. This indicates that HAL-FI can be adopted with different types of estimators in practice, and a query advantage can be obtained. Secondly, similar scaling behavior is seen for the different learners as earlier.

Additionally, we compare the passive learners and active learner HAL-FIR for tackling Problem 4.2.1 in Figure 4.9. The testing error is computed using Eq. 4.21 on a testing dataset of 10^5 i.i.d. samples. The testing distribution p_{test} is considered to be known and set to be the uniform distribution over the query space $\mathcal{Q}$. It is assumed that HAL-FIR is given access to this testing distribution. We observe a scaling of $\epsilon \sim 1/N$ for the the passive learners which is expected as the log-likelihood loss function associated with $\hat{\boldsymbol{\theta}}$ is divergence-free in the asymptotic case [Sou+17]. As in the case of HAL-FI for Hamiltonian learning, we observe a higher initial scaling for HAL-FIR and a a scaling of $\epsilon \sim 1/N$ in the asymptotic regime, consistent with that observed for the passive learner.

As discussed in Section 4.2.1, it is not typical for the testing distribution in Problem 4.2.1

to be known and the result here can viewed as a validation of the Hamiltonian model learned and hence the learners on a set of queries sampled using the testing distribution.

We note sudden peaks in uncertainty associated with the passive learner and HAL-FIR using the MLE estimator for higher values of samples. This might be indicative of traveling between multiple local minima in a larger convex hull when solving the MLE (Eq. 4.56). The trends in testing error are not severely impacted by this in expectation indicating that these are rare events and our learners (with their estimation) are robust.

Having made a case for the robustness of the learners and estimation used in this work on the simulator considering a query case $\mathcal{Q}$ which satisfies MFC, we are now in a position to compare the performance of the different learners when the $\mathcal{Q}$ is not guaranteed or known to satisfy such a criteria. In the results that follow, the observations made here are used as a basis for the behaviour to expect among the learners.

Simulator and Minimum-Frequency criteria is not necessarily satisfied In practice, the range of system evolution times $\mathcal{T}$ corresponding to the query space $\mathcal{Q}$ cannot be known apriori to satisfy MFC for frequency estimation using FFT on Rabi oscillation data. It is then crucial for learners equipped with estimation procedures to either: (i) succeed at Hamiltonian learning given this query space or (ii) alert the user that a longer $\mathcal{T}$ is required upon failure to learn a Hamiltonian model. Here, we ensure the former by modifying the standard FFT routine as discussed earlier in Section 4.4.4. Further details are provided in Appendix B.3. Note that this solve is also carried out as the first step in our estimation procedure for obtaining initial conditions to the MLE solve.

A comparison of the different learners considering a $\mathcal{T}$ that does not satisfy Minimum-Frequency criteria on the simulator is shown in Figure 4.10. We set $\mathcal{T}$ to be the set of 81 equispaced times in the interval of $10^{-7}, 6\times10^{-7}]$s. As obtained earlier on the simulator with a longer $\mathcal{T}$, we observe a SQL scaling of $\epsilon \sim 1/\sqrt{N}$ or ($N \sim \epsilon^{-2}$) in RMSE with number of queries for the passive learners combined with estimation based on linear regression, MLE or the Bayesian estimator of SMC. However, there is now a noticeably wider gap in between the trends corresponding to around 80.6% reduction in queries when using the passive learner with the MLE estimator over the passive learner with estimation based on linear regression.

For the HAL-FI algorithm with MLE, we consider cases of when the query space is fixed and when it is adaptively grown by linearly growing the $\mathcal{T}$. For both cases, we see an initial scaling which is higher than SQL and similar to Heisenberg limited scaling, and a scaling of SQL in the asymptotic query regime. This behaviour is expected from our observations on the simulator earlier. We note that using HAL-FI combined with a linearly growing query space does not show significant improvement over HAL-FI in the fixed query space. This is due to the low rate at which we adaptively grow the query query space during learning and the fact that growing the query space is only advantageous for learning a subset of the Hamiltonian parameters. We discuss this further in Section 4.5.4.

As before, we can also consider HAL-FI using Bayesian estimation for Hamiltonian learning. In Figure 4.10(b), we show results comparing learners with the Bayesian estimator SMC, against learners using alternate estimation. Firstly, we note that in a reversal of fortunes from Figure 4.8(b), the passive learner with the Bayesian estimator SMC performs similarly to the passive learner with MLE, and much better than the passive learner with estimation

based on linear regression. Secondly, we note that while HAL-FI with SMC outperforms all other learners equipped with Bayesian estimators including the sequential active learner of Qinfer, this advantage has reduced from the case where the query space satisfied the MFC criteria. Lastly, similar scaling behavior is seen for the different learners as earlier.

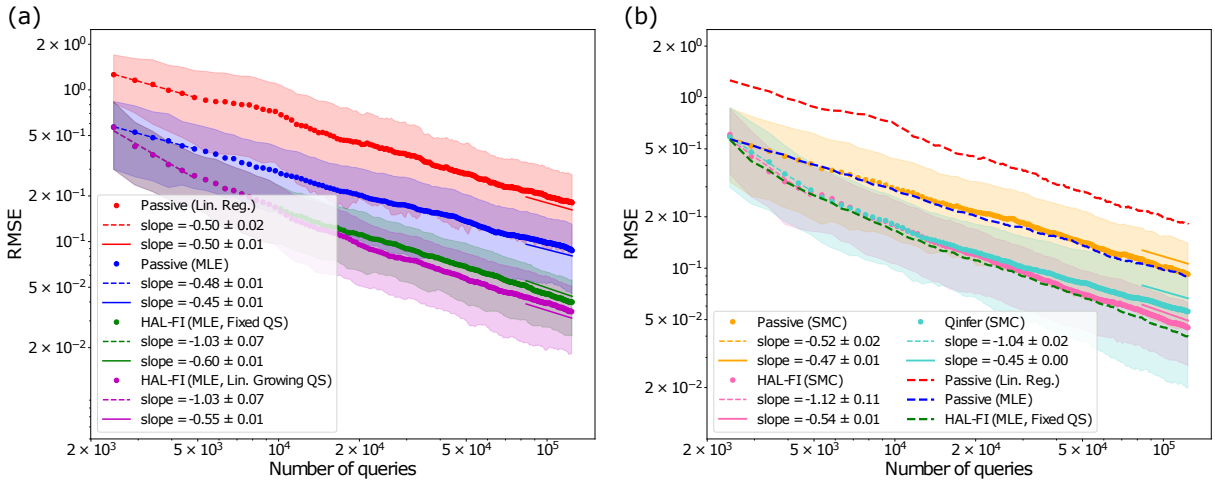


Figure 4.10: Scaling of RMSE with number of queries for different learners on a simulator when MFC is not satisfied. In (a), we compare the trends in RMSE of HAL-FI against the passive learner with estimations based on MLE and linear regression (Lin. Reg.). In (b), we compare the trends in RMSE of HAL-FI using the Bayesian estimator of sequential Monte Carlo (SMC) against HAL-FI using MLE, Qinfer using Bayes risk as the query criteria and passive learners equipped with different estimators. In (a)-(b), estimators of each learner are indicated in brackets. Slopes indicate the scaling of RMSE with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner. In (b), we do not show slopes and errors for learners shown earlier in (a) on the left, and their trends are shown for visual reference.

Experimental Dataset We show a comparison of the performance of the different learners on the *oracle* with access to experimental data in Figure 4.11. As expected from our observations on the simulator, we observe SQL like scalings in RMSE with N for the passive learners. We observe around 82.2% query reduction when using the passive learner with the MLE estimator over estimation based on linear regression, similar to that previously observed on the simulator. This query reduction was computed by fixing the RMSE value at 0.2, and comparing the number of queries required by the passive learner to achieve this RMSE value versus the baseline. The trends themselves are remarkably similar to those obtained on the simulator, supporting the fact that the main noise sources affecting the quantum device were identified and the simulator is a good representation of the real quantum hardware.

For HAL-FI, we only show results for fixed query space as the results for the linearly growing query space are very similar, as was also observed on the simulator. We also only show results for HAL-FI with the MLE estimator as HAL-FI produces best results combined with this estimation and we have already shown that HAL-FI is compatible with Bayesian estimation earlier. We see an initial scaling of $\epsilon \sim 1/N^{3/2}$ (or $N \sim \epsilon^{-2/3}$) in RMSE with

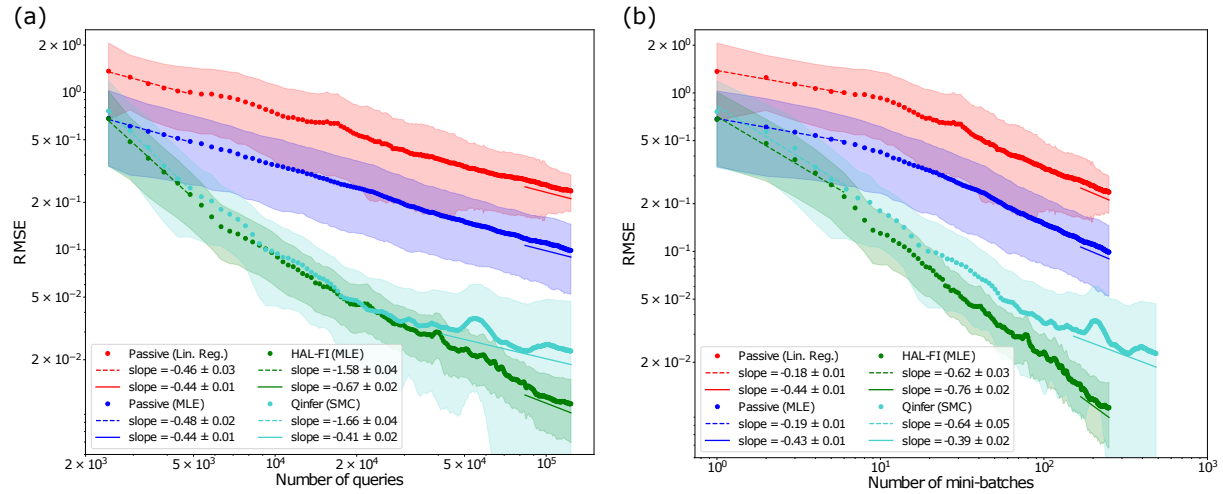


Figure 4.11: Scaling of RMSE for different learners on experimental data with (a) number of queries made and (b) number of mini-batches requested. The time range $\mathcal{T}$ of the query space is such that Minimum-Frequency criteria is not necessarily satisfied for frequency estimation using FFT. We compare the trends in RMSE of HAL-FI using MLE, sequential active learner of Qinfer using Bayes risk as the query criteria and passive learners equipped with different estimators. Estimators of each learner are indicated in brackets. Slopes indicate the scaling of RMSE with number of queries in the finite sample and asymptotic sample regimes. Filled in areas indicate the respective errors on trends for each learner.

number of queries which is higher than SQL and Heisenberg limited scaling, and a scaling of SQL in the asymptotic query regime. The performance of HAL-FI on the experimental data is surprisingly better than that on the simulator. This behaviour is consistent across different drive configurations on *ibmq_ boeblingen*. The initially accelerated learning where we observe super-Heisenberg limited scaling in HAL-FI and lower values of RMSE achieved for a smaller number of N than on the simulator might be due to: (i) noise sources not included in our model that do not contribute significantly to the noise affecting the quantum device but encourage exploration by HAL-FI, and (ii) correlations between the samples collected in the experimental data (e.g., due to thermal fluctuations).

For the sequential active learner Qinfer, we observe similar behavior in scalings as HAL-FI with an initially accelerated scaling higher than Heisenberg limited scaling and a scaling of SQL in the asymptotic query regime. The RMSE achieved by Qinfer is comparable to HAL-FI for lower values of queries but HAL-FI outperforms Qinfer significantly for higher values of queries (obtaining lower RMSE and with lower uncertainty). This is similar to what was observed on the simulator earlier. Moreover, as Qinfer is a sequential active learner, it requests more mini-batches of queries to reach a particular value of RMSE. This is illustrated in Figure 4.11(b) where we compare the number of mini-batches of queries made by HAL-FI against Qinfer. Note that the passive learners and HAL-FI request only one mini-batch for every round during learning. A mini-batch in our numerical experiments directly corresponds to a job containing multiple quantum circuits for execution on cloud based IBM Quantum devices. Each mini-batch has an associated cost of compilation on the device, and latency from classical electronics interfacing with the quantum hardware. Results of Figure 4.11(b)

thus indicate that HAL-FI would be preferred over Qinfer on current hardware with a fixed experimental budget.

We have already seen benefits of using HAL-FI over a passive learner and a sequential active learner such as Qinfer through the lower values of RMSE that can be achieved for a given number of queries. This is analyzed in terms of query advantage in Section 4.5.4.

4.5.4 Analysis

In this section, we analyze the results of the performance of the different learners from Section 4.5.3. We firstly comment on the achievability of the Heisenberg limited scaling by HAL-FI with an adaptively growing query space as claimed in Section 4.2.2. In the process, we consider a different learning scenario motivated by recalibrations of quantum devices. We then describe the query advantage of the active learner under different conditions over the baseline strategy.

Heisenberg limited scaling

In Section 4.5.3, we did not observe Heisenberg limited scaling for HAL-FI (even with an adaptively growing query space). This is due to the fact that the query space is not rich enough to achieve Heisenberg limited scaling i..e., there is no sequence of queries even in the adaptively growing query space to achieve Heisenberg limited scaling. We discuss when Heisenberg limited scaling is achievable for Hamiltonians based on the CR Hamiltonian in Appendix B.5. We show that the behavior of the learners observed so far is expected through another set of experiments in Appendix B.4.

It should however be possible to achieve Heisenberg limited scaling for a subset of Hamiltonian parameters given the query space (see Section 4.4.2) when the task is to learn this subset of Hamiltonian parameters and we are given access to information about the other Hamiltonian parameters. This is exactly the setting of a recalibration where prior information about the Hamiltonian parameters is available from previous calibrations and the goal is to learn the subset of parameters which drift significantly with time while refining estimates of those parameter that do not. Motivated by this, we consider the following learning scenario on *ibm_boeblingen* under drive configuration 3 (see Table 4.4).

We have access to an estimate of the Hamiltonian parameters $\hat{\boldsymbol{\theta}}$ from a previous calibration during which Hamiltonian learning was run on a uniformly sampled set of queries from $\mathcal{Q}$ (of size $N = 2430$). The goal is to then learn the parameters $\omega_{0,1}$ using the different learners at our disposal. We plot comparisons of different learners (using $N_b = 972$) on this recalibration task in Figure 4.12 considering the oracles of the simulator and experimental data. We consider all the learners with estimation based on MLE or linear regression but could also use a Bayesian estimator as we have already noted in Section 4.5.3. We will call the passive learner with estimation based on linear regression as the baseline for these studies.

We observe the SQL scaling in the baseline, passive learner and HAL-FI in the fixed query space. There is nearly a constant gap between the baseline strategy and HAL-FI in the fixed query space, indicating a constant query reduction in achieving a desired learning error. We observe a super-Heisenberg limited scaling of RMSE $\epsilon \sim N^{-3/2}$ in number of

queries in HAL-FI with a linearly growing query space. This is also observed for HAL-FI with an exponentially growing query space in the low sample regime. The deterioration in the scaling of HAL-FI with an exponentially growing query space to the SQL is due to the maximum evolution time corresponding to the growing query space eventually exceeding T_1 and T_2. This is shown in Figure 4.12(b). In fact, the bend in the trend of RMSE versus N for HAL-FI with an exponentially growing query space occurs immediately after the maximum evolution time in $\mathcal{T}$ exceeds T_1. As $\mathcal{T}$ for HAL-FI with a linearly growing query space is grown much more slowly, effects of decoherence are not yet felt and super-Heisenberg limited scaling convergence rate in learning error is achieved.

Qualitatively, it is clear that much lower values of learning error can be achieved with a given budget of queries using HAL-FI with an adaptively growing query space over the baseline for recalibration. This is quantified in terms of query advantage in Section 4.5.4.

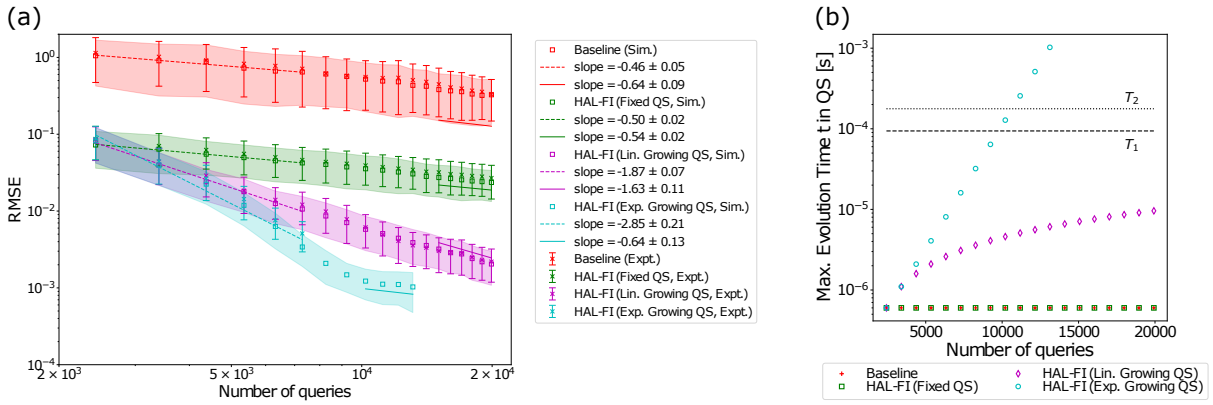


Figure 4.12: (a) Trends of RMSE with number of queries for different learners in fixed or growing query spaces (QS) on experimental data (Expt.) compared against simulator (Sim.). Slopes indicate the scaling of RMSE ϵ with number of queries N in the low query/sample and high query/sample regimes. Filled in areas indicate the respective errors on trends for each learner on the simulator. Passive learner which has similar behavior to HAL-FI in fixed query space is not shown for brevity. (b) Trends of the maximum evolution time associated with query spaces of different learners.

Computational Cost

So far, we have assessed the performance of different learners in achieving a value of RMSE with respect to the number of queries consumed. However, another relevant resource is the classical computational time required for learning the Hamiltonian parameters during estimation or that required in computing query criteria for active learners.

The main contribution to this classical computational time is the evaluation of the likelihood function for different queries, required during estimation and for evaluation of query criteria for active learning. If an analytical expression for the likelihood function is not available, then one needs to carry out expensive quantum simulations. The computational cost of using the SMC estimator in a round during learning scales as $O(|\mathcal{Q}|n_p)$ with the number of particles. On the other hand, the cost (or number of likelihood evaluations) of using MLE

in a round scales as $O(|\mathcal{Q}|n_{\text{iters}})$ with the number of iterations n_{iters} within the estimation procedure (e.g., number of gradient steps, etc.). In HAL-FI, we also account for the number of likelihood evaluations in setting up the Fisher information matrices of different queries for solving the SDP of Eq. 4.19. In Qinfer, we account for the number of likelihood calls in hypothetical updates of the risk particles for evaluating Bayes risk (see Section 4.5.2). In Figure 4.13, we compare the RMSE achieved for increasing computational cost for different learners equipped with the MLE estimator or the Bayesian SMC estimator. Overall, we find that learners equipped with MLE are less computationally expensive compared to those equipped with the Bayesian estimator.

In our experiments, we observe that the number of iterations (e.g., of the quasi-Newton solve) in the estimation procedure of MLE (Section 4.4.4) is lower than the number of particles n_p required in SMC. This translates into about two orders of magnitudes of savings in computational cost when comparing HAL-FI against the sequential active learner of Qinfer on both the simulator and experimental data.

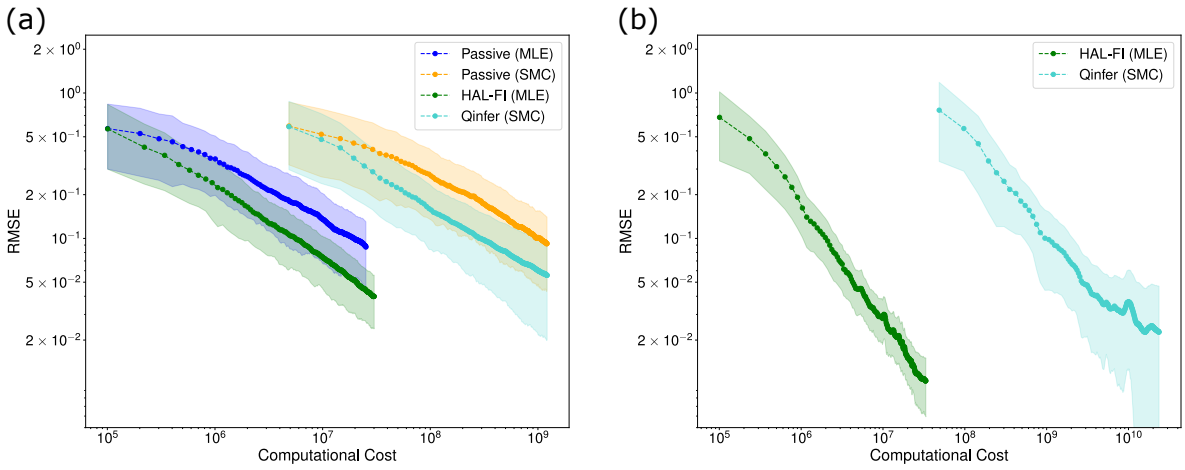


Figure 4.13: Trends of RMSE with computational cost for different learners on (a) simulator where MFC is not necessarily satisfied, and (b) experimental data. The computational cost is measured in terms of the number of likelihood evaluations during learning.

Query Advantage

We have so far compared the trends and obtained the scalings of RMSE ϵ with number of queries N for different learners for Hamiltonian learning (Section 4.5.3) and with prior information (Section 4.5.4). To quantify the benefits of using HAL-FI over other learners, we now evaluate its query advantage (QA), defined in Section 4.2.3 as a performance metric that summarizes the reduction in resources required to achieve a desired learning error. We restrict our attention to HAL-FI with the MLE estimator to simplify the discussion.

We plot the trend of QA as a function of RMSE for active learners HAL-FI and Qinfer over baselines of passive learners equipped with different estimators as achieved on the simulator in Figure 4.14 and as achieved on experimental data in Figure 4.15(a). Values of QA for certain values of RMSE are tabulated in Tables 4.5 and 4.6 for different combinations of

learners and baselines. We observe a QA of around 80% in HAL-FI on the simulator and experimental data over the baseline of the passive learner with estimation based on linear regression, for high values of RMSE. The initial accelerated learning observed for HAL-FI on the simulator in Figure 4.10 and experimental data in Figure 4.11 translates to an accelerated QA for high values of RMSE. The trend in QA for HAL-FI over the passive learner with estimation based on linear regression on the simulator flattens to an asymptotic value of at least 95.1% when the query space is fixed during learning and at least 96.3% when the query space is grown linearly. The corresponding value for HAL-FI on the experimental data in the fixed query space is 99.8% for low values of RMSE. Similarly, the QA of HAL-FI over a passive learner with the MLE estimator is around 82.1% on the simulator and around 99.0% on experimental data for low values of RMSE. In comparison, the sequential active learner of Qinfer is only able to achieve a QA of around 62.7% on the simulator but performs better on the experimental data achieving a QA of around 96.9% over the baseline of passive learner with the MLE estimator.

For QA of other combinations of learners and baselines, we turn our attention to Tables 4.5 and 4.6. We note that HAL-FI can achieve up to a QA of 53.2% and 70.9% in expectation over the baseline of Qinfer on the simulator and experimental data respectively for the lowest value of RMSE achieved by HAL-FI in our numerical experiments in Figure 4.11. There is however a large uncertainty on the estimate of QA on the experimental due to the fluctuations of the trend of Qinfer on the experimental data (Figure 4.11). In practice, we expect that HAL-FI would achieve a QA upwards of 33% (computed by considering one standard deviation below the expected value of QA).

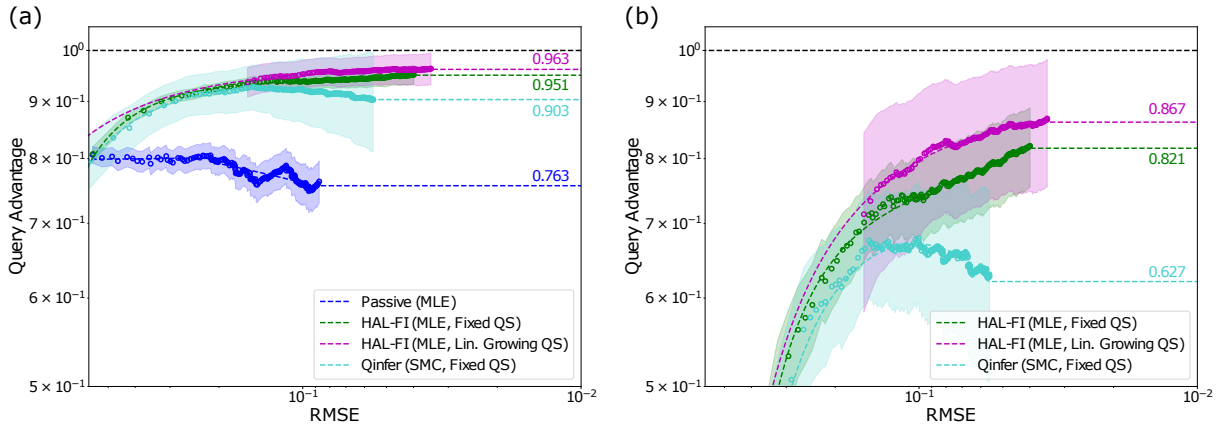


Figure 4.14: Query advantage (QA) of different learners on the simulator when MFC (minimum-frequency criteria) is not satisfied over the following baselines: (a) passive learner with estimation based on linear regression and (b) passive learner with the MLE estimator. In (a)-(b), the data points correspond to values of QA computed from the data points of Figure 4.10 and the lines are the fits to these values of QA. The annotated text for the different trends indicate the QA obtained for the corresponding learner over the baseline for the lowest observed value of RMSE in our numerical experiments. Filled in areas indicate the 95% confidence interval on the data points shown.

Similarly for the learning scenario of recalibration, we plot the QA of HAL-FI over

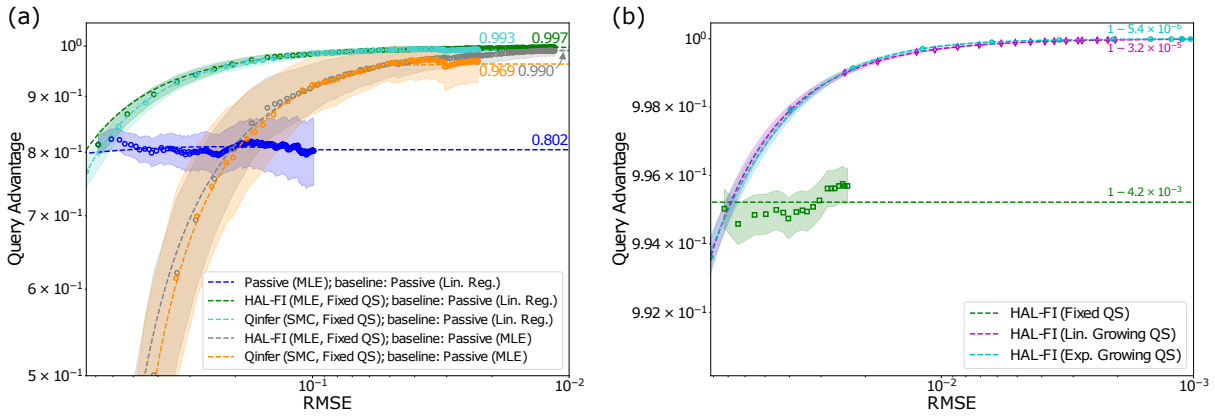


Figure 4.15: Query advantage (QA) of different learners over specified baselines considering the problem of (a) Hamiltonian learning without prior information and (b) Hamiltonian learning with prior information on subset of Hamiltonian parameters from previous calibrations. In (a), we plot the QA of passive learners, HAL-FI, and Qinfer on experimental data over two different baselines. In (b), we plot the QA of HAL-FI over the baseline of the passive learner on the simulator for fixed and adaptively growing query spaces (QS). In (b), the QAs obtained on the simulator closely match those obtained on the experimental data and hence the latter is not shown. In (a)-(b), the data points correspond to values of QA computed from the data points of Figures 4.11,4.12 and the lines are the fits to these values of QA. The annotated text for the different trends indicate the QA obtained for the corresponding learner over the baseline for the lowest observed value of RMSE in our numerical experiments. Filled in areas indicate the 95% confidence interval on the data points shown.

the baseline of passive learner (with estimation based on linear regression or MLE) for Hamiltonian learning when prior information is available in Figure 4.15(b). The trends shown here correspond to that of the simulator as similar behavior is observed for experimental data. The QA of HAL-FI in a fixed query space over the baseline is constant with RMSE and around 99.6% i.e., HAL-FI requires only 0.4% of the queries required by the baseline. For HAL-FI with a linearly growing query space, the query advantage for low values of reported RMSE is around $(100 - 3.2 \times 10^{-3})\%$. Further, we note that the scaling of QA of HAL-FI in a linearly growing query space with RMSE ϵ scales as QA $\sim (1 - O(\epsilon^{-1/6}))$ until decoherence starts effecting the scaling. For HAL-FI with an exponentially growing query space, the QA flattens to around $(100 - 5.4 \times 10^{-4})\%$ for low values of RMSE. Under this learning setting at low values of RMSE around 10^{-3}, HAL-FI has a query space with evolution times far exceeding T_1 or T_2 and thus this reported QA is expected for even lower values of RMSE.

From the analysis of query advantage above, we observe that two orders of magnitude reduction in queries can be obtained over a baseline of passive learner with estimation based on linear regression by adopting HAL-FI. Moreover, during recalibrations, we observe three orders of magnitude reduction when HAL-FI is used with a linearly growing query space and five orders of magnitude reduction in queries when HAL-FI is used with an exponentially growing query space.

Adopting HAL-FI not only allows us to achieve query reduction but it also allows us

to save wall clock time taken to calibrate a quantum device. We again consider the query space of 4.4.2 where a query on average takes $2400ns$ to run (accounting for time duration of implementing measurement pulses). The repetition rate of current IBM Quantum devices for executing circuits is 10kHz. We can then reduce the duration of Hamiltonian learning of all CR Hamiltonians of directly connected qubit pairs on a 20-qubit IBM Quantum device of *ibmq_boeblingen* to reach a RMSE of 5×10^{-2} from around 10 minutes to 5 seconds by using HAL-FI instead of a passive learner with estimation based on regression [She+16]. These timings only take into account the time on the quantum hardware and not additional latencies in classical electronics interfacing with the hardware.

Baseline \ **Learner**	Passive (MLE)	Qinfer (SMC)	HAL-FI (MLE, Fixed QS)	HAL-FI (MLE, Lin. Grow. QS)
Passive (Lin. Reg.)	0.763 ± 0.034 (0.788 ± 0.024)	0.903 ± 0.092 (0.926 ± 0.043)	0.951 ± 0.018 (0.929 ± 0.011)	0.963 ± 0.031 (0.935 ± 0.029)
Passive (MLE)	–	0.627 ± 0.122 (0.668 ± 0.090)	0.821 ± 0.067 (0.750 ± 0.062)	0.867 ± 0.113 (0.811 ± 0.114)
Qinfer (SMC)	–	–	0.532 ± 0.174 (0.416 ± 0.182)	0.658 ± 0.289 (0.572 ± 0.302)
HAL-FI (MLE, Fixed QS)	–	–	–	0.184 ± 0.091 (0.215 ± 0.081)

Table 4.5: Query advantage of various learners considering different baselines on a simulator when MFC is not satisfied. Estimators of each learner are indicated in brackets. For each combination of learner and baseline, we report query advantage at lowest value of RMSE achieved by the learner (compared to extrapolated fit of baseline) and the query advantage at lowest value of RMSE achieved by the baseline in brackets below. The errors represent one standard deviation on the value of expected query advantage.

Baseline \ **Learner**	Passive (MLE)	Qinfer (SMC)	HAL-FI (MLE, Fixed QS)
Passive (Lin. Reg.)	0.802 ± 0.057 (0.809 ± 0.034)	0.993 ± 0.006 (0.953 ± 0.012)	0.997 ± 0.001 (0.957 ± 0.009)
Passive (MLE)	–	$0.969 \pm \underline{0.046}$ (0.917 ± 0.056)	$0.990 \pm \underline{0.011}$ (0.916 ± 0.046)
Qinfer (SMC)	–	–	$0.709 \pm \underline{0.373}$ (0.510 ± 0.169)

Table 4.6: Query advantage (QA) of various learners considering different baselines on experimental data. Estimators of each learner are indicated in brackets. For each combination of learner and baseline, we report query advantage at lowest value of RMSE achieved by the learner (compared to extrapolated fit of baseline) and the query advantage at lowest value of RMSE achieved by the baseline in brackets below. The errors represent one standard deviation on the value of expected query advantage and are underlined in cases where the range of query advantage should be read as [QA-error,1) instead of [QA-error, QA+error].

4.6 Conclusion

In this chapter, we proposed the active learning algorithms of HAL-FI for Hamiltonian learning and HAL-FIR for predictions of queries to a Hamiltonian, sampled from a testing distribution. They both perform in batch-mode making them ideal for use on near-term quantum hardware. The performance of HAL-FI/HAL-FIR was compared against different learners for learning a CR Hamiltonian on the 20-qubit IBM Quantum device *ibmq_boeblingen* on a simulator and experimental data. We showed that HAL-FI can achieve a query advantage of around 99.7% over a passive learner with estimation based on linear regression, 99.0% over a passive learner with MLE estimator, and upwards of 33% over the sequential active learner of Qinfer (equipped with the query criteria of Bayes risk) for low values of learning error on experimental data. Moreover, HAL-FI is able to achieve the same learning error at lower computational cost than passive learners and Qinfer. During recalibration when learning the Hamiltonian with access to information from previous calibrations, we observed that HAL-FI can achieve query advantages of 99.5% over passive learners. Further, we showed that we achieve Heisenberg limited rate of convergence where possible when an active learner is used in conjunction with an adaptive query space during learning before the evolution time of queries exceed qubit T_1 or T_2 and decoherence starts deteriorating information content available in queries.

Overall, using the active learner HAL-FI operating in batch-mode can yield a reduction in resources of up to two orders of magnitude during calibration and five orders of magnitude during recalibration for low values of learning error. This improvement in query complexity has multiple practical consequences besides accelerating Hamiltonian learning during calibration and recalibrations of quantum computers.

Another important calibration step is determining controls to implement desired single and multi-qubit quantum gates. This often relies on building a Hamiltonian model of the true environment i.e., the quantum computer. This can be accomplished by HAL to ensure minimal queries are used. Gates, once implemented, are characterized through quantum process tomography which is not query efficient but could be accelerated with an active learner. For the specific application of learning the CR Hamiltonian, more noise sources can be included as they become relevant e.g., leakage errors which become pronounced under strong driving. It would also be interesting to see if one can achieve asymptotic Heisenberg limited scaling even in the presence of noise such as decoherence by using appropriate quantum error correction protocols. It should be noted that Hamiltonian learning and for that matter quantum process tomography both suffer from an exponential scaling in the size of the quantum device n and using an active learner only ensures a better scaling in ϵ or query advantage. One could get a better scaling in n if additional information such as the structure of the Hamiltonian was known to be a k-local Hamiltonian which can be learned with a query complexity that scales as $O(\text{poly}(n))$ [Ans+21].

There are other calibration steps where Hamiltonian learning may not be required but the concept of active learner can be introduced. This would particularly be advantageous where a variety of experiments can be carried out but it is not clear which of them are more informative for the learning task. For example, a query efficient method is desired for learning cross-talk on a superconducting quantum device [Abr+19b; Dai+21].

Additionally, there is room for improvement and extension in the algorithm itself. Currently, expert knowledge is required to specify a complete query space to HAL to ensure all the Hamiltonian parameters can be learned and possibly with Heisenberg limited scaling. It is desirable to remove this expert and replace them with a method to synthesize queries. The current active learning strategy is also based on the query criterion of Fisher information but this could be modified to incorporate cost of different queries and incentivize exploration. Moreover, a more general query criteria could possibly be learned through reinforcement learning as illustrated in recent work on classical applications [Pan+18].

Code and data availability

Code for the passive learner and the active learner HAL using estimators based on regression and MLE is available at [Dut23a]. Code for the learners including Qinfer for the cross-resonance Hamiltonian, all equipped with a Bayesian estimator is available at [Dut23b]. Data sets generated from the simulator are included in the above repositories. Please contact the thesis author for access to the experimental data sets collected from the IBM quantum devices and used in this chapter.

Chapter 5

Benchmarking randomized measurements of quantum chemsitry Hamiltonians

5.1 Introduction

The electronic structure problem, and more specifically the problem of approximating ground states, is one of the outstanding challenges in computational chemistry. Over nearly the past century, an enormous amount of scholarship has gone into developing classical methods for this task (Hartree-Fock [HJO00], MP2 [HJO00], configuration interaction [HJO00], coupled cluster [HJO00], density functional theory [HK64; KS65], quantum Monte Carlo [HLR94], and DMRG [CH02], to name some notable examples), and in recent decades a large proportion of scientific HPC resources have been dedicated to solving it (see for example [Aus+]). The computational intensiveness of the electronic structure problem has contributed to the motivation for developing benchmarks for ranking algorithms across many scientific areas including computational chemistry, including for example QM7 [BR09; Rup+12], QM9 [Rud+12; Ram+14], and W4-11 [KDM11]. Several of these listed benchmarks in computational chemistry involve datasets for tasks pertaining to molecular property prediction, but there are also benchmarks that focus on algorithmic or methodological aspects without relying on specific datasets, for example, basis set benchmarking [KTM07; MBS11] for describing electronic structure of molecules, conformal benchmarking [Fri+17] to assess algorithms for exploring the low-energy conformational space of molecules, and reaction path benchmarking [MH19] to compare optimization methods for finding chemical reaction pathways.

The classical computational chemistry benchmarks mentioned above share the idea of testing algorithms against a common computation or prediction task on a set of common objects. This idea is prevalent throughout benchmarking for not only comparing different algorithms but also devices with respect to certain measures of performance. For example, the LINPACK benchmark [Don+79; Don87], which was used to rank the top supercomputers in the world, involves the common task of solving linear systems of equations $Ax = b$ where the input matrix A is a pseudo-random dense matrix. The common computation is solving linear systems of equations in the case of LINPACK and molecular prediction tasks

for the computational chemistry benchmarks. Additionally, as part of the benchmark, the performance of different algorithms is assessed on the chosen task by testing it on a set of objects. In the case of LINPACK, these objects are pseudo-random dense matrices. In the QM databases, the objects are small organic molecules [BR09; Rup+12; Rud+12]. The hardness and generality of the set of objects in the benchmark determine how well it will predict the performance of algorithms in practice. This has been particularly successful in the context of machine learning for image classification (e.g., MNIST [Den12], CIFAR [KH09], ImageNet [KSH12]) and object detection (e.g., MS COCO [Lin+14]). In MS COCO, for example, algorithms for object recognition are assessed in the broader context of scene understanding and are tested against images of complex everyday scenes containing common objects in their natural context. By including typically occurring objects in practice as part of the testing suite for benchmarking, there has been an improvement in the development of state-of-art algorithms for object recognition [Kha+22; Min+22].

Returning our attention to the electronic structure problem, on a quantum computer the primary challenge in the classical methods, that of representing highly entangled and correlated wavefunctions, is removed in principle. This, together with the classical hardness of the ground state problem and the enormous amount of resources dedicated to it, has motivated the development of quantum algorithms for the task [McA+20; Lee+23], although new challenges arise. Numerous methods have been designed, including near-term quantum algorithms such as variational quantum eigensolvers (VQE) [Per+14; McC+16; Kan+17; Gri+19], quantum approximate optimization algorithm (QAOA) [FGG14; Mol+18; Far+22], and quantum subspace expansion (QSE) methods [McC+17; Col+18a; PM19; Mot+20]. Fault-tolerant algorithms for ground state estimation, which are aimed at future high-accuracy quantum computers, include quantum phase estimation (QPE) [Kit95], its variants [AL99a; PW09b], algorithm in [GTC19] which uses linear combination of unitaries (LCU) [CW12], and those using quantum signal processing [Gil+19; LT20; LT22; DL23]. On currently existing and near-term quantum computers, without error correction, near-term algorithms are preferred for use instead of fault-tolerant algorithms whose circuit depths and qubit counts will require error correction. These near-term algorithms including VQE are typically hybrid quantum-classical algorithms involving sequential rounds of measurements of parametrized quantum circuits or short time Hamiltonian simulation and classical post-processing along with classical optimization.

A common subroutine across many of these algorithms is that of observable estimation or estimating $\mathrm{Tr}(\rho H)$ (e.g., [Du+10; Lan+10; Wan+15; OMa+16; She+17; Pae+17; Hem+18; San+18; Col+18b; Dum+18; Kok+19; Kan+19; Gan+19; Sag+19; McC+19; SM19; Nam+20; Aru+20b; Kre+21; Löt+21; Kis+22]) for a given n-qubit quantum state ρ resulting from a short depth quantum circuit and an n-qubit Hamiltonian H (or in general any observable). Physical Hamiltonians H can be decomposed into a linear combination of L n-qubit Pauli operators: they form a basis for the Hermitian operators, and local observables have polynomial-sized decompositions in the Pauli basis. Other decompositions of H include LCU [CW12; Kir+22] or one-sparse matrices [AT03; Ber+07; CK11]) but these are impractical in the near-term because of the relatively complex quantum circuits required to estimate expectation values of the terms. In contrast for Pauli decompositions, we could estimate $\mathrm{Tr}(\rho H)$ simply by estimating $\mathrm{Tr}(\rho Q)$ independently [Per+14] for each of the Pauli terms Q in the Pauli decomposition of H.

However, this procedure is typically inefficient as (in general) subsets of Pauli terms will commute and thus be co-measurable. Generally, however, a commuting set of Paulis can only be simultaneously measured by applying a depth-$\Theta(n)$ Clifford circuit to map them to their common eigenbasis. In the near-term, when circuit depth is at a premium due to the lack of error correction, it is desirable to use all of the circuit depth for the state preparation instead of measurements. *Locally* commuting Pauli operators, i.e., operators that have all non-identity single-qubit Pauli matrices in common are used instead. They can be measured simultaneously in the same local basis by applying one layer of single-qubit gates followed by measurement in the computational basis and these are the type of measurements we consider access to in this chapter.

In recent years, two main approaches for the observable estimation problem using local Pauli measurements have emerged: (i) randomized measurements [HKP20; Had+22; Hil+21; LLB21; KG22; Elb+23] in which local Pauli measurement bases are drawn from a distribution over the n-qubit Pauli operators [HKP20; Had+22; Hil+21] or generated via a sampling procedure that does not require explicit access to the distribution [Had21], and (ii) grouping methods [Gok+20; YVI20; VYI20; Cra+21; Wu+23; YGI23; Shl+23], which combine Pauli terms into locally compatible subsets for simultaneous measurement either systematically [Cra+21; Wu+23] or using ad-hoc heuristics [Kan+17; Hem+18; VYI20]. There also other approaches: [ASS21] uses a set of informationally complete positive operator-valued measurements to solve the observable estimation problem and [HKP21] obtains deterministic sequences of Pauli measurements to be made by derandomizing randomized measurements. Notably, among the listed methods are those based on classical shadows [HKP20; Had+22; Hil+21; Wu+23] which are asymptotically optimal [HKP20] requiring only $O(3^w \log L)$ measurements for a Hamiltonian with L Paulis in its Pauli decomposition and maximum number of non-identity Paulis in any Pauli term being w. Despite the potential of these methods in ideal scenarios, little is known about their behavior on quantum devices in presence of noise. Experimental studies have only been carried out so far on small molecular Hamiltonians [Str+21] or quantum states over few qubits [Zha+21].

Given the large suite of options, a natural question at this point is: How do we systematically select measurement methods for the common quantum computation of estimating $\mathrm{Tr}(\rho H)$ in hybrid quantum-classical algorithms? One way to tackle this is to follow the classical approach and propose a benchmark. This is not without precedent on the quantum side: recently a "quantum LINPACK" benchmark [DL21] was proposed for ranking computational power of quantum computers; in direct analogy to LINPACK, it involves solving the quantum linear system problem. A challenge in designing benchmarks is to have predictive power regarding the performance of the candidate algorithms beyond the dataset tested against.

In this work, we take a similar approach to classical benchmarks by considering the common quantum computation task of observable estimation $\mathrm{Tr}(\rho H)$ on a set of common chemistry Hamiltonians and quantum states. In analogy to classical computational chemistry benchmarks [BR09; Rud+12], we consider sets of quantum states particular to the problem of ground state estimation as well as those states that naturally occur during the runtime of a hybrid quantum-classical algorithm. This culminates in the proposed benchmark in this work called CSHOREBench: Common States and Hamiltonians for ObseRvable Estimation.

In addition to commenting on the objects considered as part of the data set of CSHOREBench,

performance metrics used to rank measurement methods on these objects need to be defined. An important aspect in designing or selecting the measurement method and estimator is the amount of resources required. Considering the performance metric of accuracy, one selection criterion is to minimize the number of measurements required on the quantum device in achieving a given accuracy. However, this only takes into account the quantum resources used and may come at a prohibitive computational cost on classical computers in setting up the measurement methods or running the estimator on the data acquired from the experiment step. To capture a representative performance metric for all of the costs associated to a measurement procedure, classical and quantum, we propose a heuristic that incorporates different resources' utilization in the observable estimation problem.

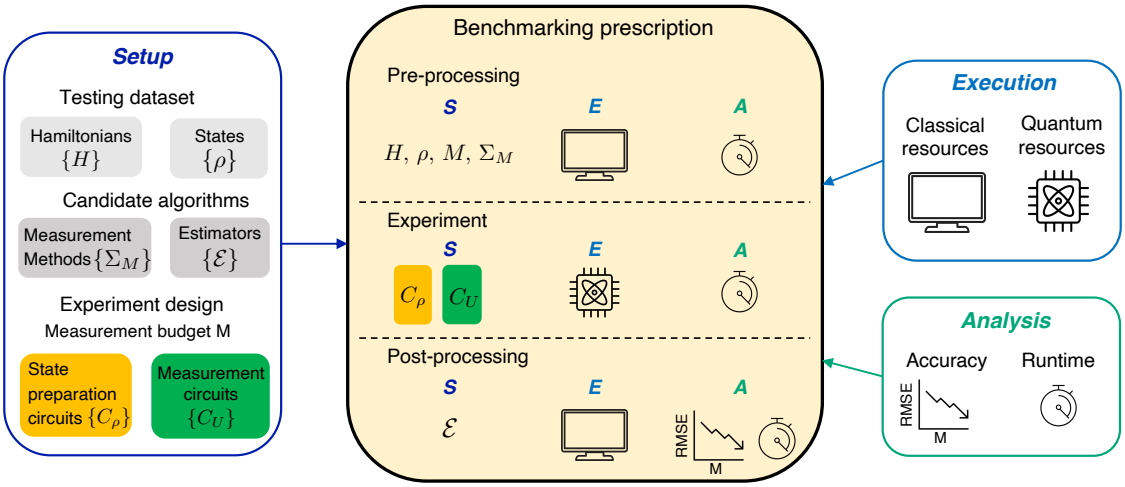


Figure 5.1: Prescription for benchmarking a candidate measurement protocol (measurement method Σ_M and estimator $\mathcal{E}$) on a Hamiltonian H and state ρ against measurement budget M. The benchmarking stages of setup (S), execution (E) and analysis (A) are shown for each step of a general measurement protocol. General measurement protocols, to be discussed in Section 5.2.4, can be divided into steps of classical pre-processing where measurement samples are generated, experiments on a quantum device and classical post-processing on data acquired. Each step has a benchmarking stage associated with it. As shown (left), setup defines the testing dataset, candidate algorithms and experiment design. Execution (right) defines the computational resources available for any step. Analysis (right) defines the metrics associated with each step.

As part of benchmarking different measurement protocols on a given H and ρ, it will be convenient to break up each measurement protocol into the steps of pre-processing where the Pauli bases are generated using a measurement method, experiments where a state is measured in these bases on a quantum device and post-processing where an estimate is obtained from the data acquired using an estimator. In the rest of this chapter, we will complete the benchmarking process for each of these steps in CSHOREBench as commonly done in machine learning and computational chemistry benchmarks, consisting of the following stages: (i) *setup*, (ii) *execution*, and (ii) *analysis*. An illustration of this benchmarking pro-

cess (or prescription) is depicted in Figure 5.1. (i) *Setup* involves defining the test dataset of Hamiltonians and states under consideration, candidate measurement methods, estimators and the benchmarking experiment design. (ii) *Execution* defines the computational resources available for executing the step such as classical computing (e.g., CPU, distributed computing, parallelization, simulator, etc.) and quantum computing (e.g., QPU, modular quantum devices etc.). (iii) *Analysis* defines the performance metrics and the methods (e.g., empirical, inferential, etc.) to evaluate the performance metrics. The main distinction of CSHOREBench from classical computational chemistry benchmarks is the availability of quantum devices for execution and as highlighted in the paragraph before, utilization of this resource will be important to account for. Finally, it is desirable that a benchmark is reproducible and reflects reality (or the performance obtained in experiment on quantum hardware). We demonstrate this by including data and analysis of CSHOREBench from experiments on IBM quantum devices.

This chapter is organized as follows. In Section 5.2, we first formalize the problem of estimating $\mathrm{Tr}(\rho H)$ for a given n-qubit Hamiltonian and access to an unknown n-qubit quantum state ρ under the constraints of measuring in the Pauli basis. We then describe the benchmarking strategy followed in CSHOREBench in Section 5.2.3 and then describe its setup, execution and analysis in the context of a general measurement protocol for estimating $\mathrm{Tr}(\rho H)$ in Section 5.2.4. This allows to kick off our discussion of different estimators that may be employed with various measurement methods in Section 5.3. In Section 5.4, we describe randomized and derandomized measurement methods using the framework of decision diagrams. In Section 5.5, we formally describe the data set of molecular Hamiltonians and states considered as part of CSHOREBench before presenting the experimental protocol we follow. In Section 5.6, we report our results from CSHOREBench on the convergence behavior and resource utilization of various measurement methods. Finally in Section 5.7, we comment on our benchmarking results and possible extensions of this work.

5.2 Background

In this section, we introduce the problem of observable estimation, i.e., estimating $\mathrm{Tr}(\rho H)$ via Pauli measurements, given an n-qubit quantum Hamiltonian (or observable) H and an n-qubit quantum state ρ. This is followed by the description of our benchmarking strategy for the observable estimation problem in Section 5.2.3. We then describe the different steps of setup, execution and analysis of CSHOREBench through a presentation of the general measurement protocol for estimating $\mathrm{Tr}(\rho H)$.

A formal description of CSHOREBench is presented in Section 5.5 after describing the measurement methods in Section 5.4 and estimators in Section 5.3. We now begin by introducing relevant notation.

5.2.1 Notation

We will denote the set of n-qubit Pauli operators as $\mathcal{P}_n = \{I, X, Y, Z\}^{\otimes n}$, the set of n-fold tensor products of the single-qubit Pauli matrices $\{I, X, Y, Z\}$. At times, it will be convenient to consider the set of n-fold tensor products of non-identity single-qubit Pauli

matrices, which we denote by $\Omega_n = \{X, Y, Z\}^n$. For any n-qubit Pauli operator Q, we refer to its action on the jth qubit as Q_j and hence have $Q = \bigotimes_{j=1}^n Q_j$. We denote the support of a Pauli operator as $\text{supp}(Q) = \{j | Q_j \neq I\}$ and its weight as $\text{wt}(Q) = |\text{supp}(Q)|$.

We say that the n-qubit Pauli operator B covers n-qubit Pauli operator Q (or Q is covered by B) if Q can be obtained from B by replacing some of the local Pauli matrices on single-qubits with identity. We then write $Q \triangleright B$. We extend the same notation to sets of Pauli operators on the left hand side, e.g., $S \triangleright B$ if and only if all Pauli operators in S are covered by B. For example, $\{XXI, IXX, XIX\} \triangleright XXX$ but $\{ZZI, IZZ, ZIZ\} \not\triangleright XXX$.

5.2.2 Observable estimation: Learning problem of measuring quantum Hamiltonians

Consider an n-qubit Hamiltonian decomposed as a linear combination of L Pauli terms

$$H = \sum_{j=1}^{L} \alpha_j Q^{(j)} \tag{5.1}$$

where $Q^{(j)} \in \mathcal{P}_n$ are n-qubit Pauli operators and $\alpha_j \in \mathbb{R}$ are the corresponding coefficients. We call the set $\mathbf{Q} \coloneqq \{Q^{(j)}\}_{j \in [L]}$ the *target* observables where we used the notation $[L] = \{1, 2, ..., L\}$.

The *observable expectation problem* is then as follows. Given an n-qubit quantum state ρ (prepared by some quantum circuit), the goal is to estimate $E \coloneqq \text{Tr}(\rho H)$ within error $\epsilon \in (0, 1/2)$ using as few prepare-and-measure repetitions as possible. Note that H can represent any physical observable; an typical example would be the qubit representation of a molecular Hamiltonian, as studied in the earliest papers to consider grouping of commuting Pauli measurements [Kan+17; Hem+18]. In the process of obtaining an estimate $\hat{E}$ of E, we will obtain estimates of the $\text{Tr}(\rho Q^{(j)})$, which will be denoted by $\hat{\omega}^{(j)}$. The true value of $\text{Tr}(\rho Q^{(j)})$ will be denoted by $\omega^{(j)}$.

The main constraint that we will impose on our learning problem is that once ρ has been prepared on a quantum device, we are only allowed to use measurements corresponding to n-qubit Pauli operators to learn values of $\text{Tr}(\rho Q^{(j)})$ and hence $\text{Tr}(\rho H)$. This ensures that we do not have any access to quantum resources such as entanglement for learning, and our measurement circuits are composed of single-qubit operators. As discussed above, this is a reasonable constraint to impose on existing and near-term noisy quantum hardware where one would want to prioritize depth in the state preparation circuit over depth in the measurement circuit.

5.2.3 Strategy for CSHOREBench

The goal of CSHOREBench is to assess the performance of and rank different measurement methods in estimating $\text{Tr}(\rho H)$ through local Pauli measurements on a quantum computer, for n-qubit Hamiltonians H and n-qubit quantum states ρ prepared on a quantum computer. To start the description of the benchmarking setup, we need to define the set of objects, i.e., types of Hamiltonians H and quantum states ρ considered as part of the test suite for our candidate measurement methods.

Analogous to classical benchmarks of MS COCO [Lin+14] and on QM datasets [BR09; Rud+12] (described in Section 5.1), we consider the broader learning context of these measurement methods when used in near-term hybrid quantum-classical algorithms, which is typically ground state estimation. We thus consider objects from this natural context. The set of Hamiltonians considered here include small molecular Hamiltonians of varying sizes and with varying Pauli weight distributions. We consider different types of states that would be expected during the runtime of a hybrid quantum-classical algorithm such as VQE. For example, we benchmark against the Hartree-Fock (HF) state, which is classically simulatable and a possible initial state for many ground estimation algorithms. During the course of a (successful) VQE run, one could also expect to see an approximate ground state at the very end, but for the purpose of benchmarking such states are not desirable since they may be difficult to prepare. Instead, we benchmark against quasi-random states prepared by a typical low-depth ansatz with random parameter settings. These random states serve as a proxy for typical intermediate states obtained in the middle of a VQE optimization, since although in that case the parameter setting would not be random, it would not in general bear any particular relation to the Pauli decomposition of the target observable. The overall code base of CSHOREBench is designed such that any new Hamiltonian can be easily added to the existing dataset and measurement methods benchmarked against it.

The most popular metric used so far is that of accuracy, i.e., root mean square error (RMSE) in the estimate of $\mathrm{Tr}(\rho H)$ for a given budget of measurements. However, a highly accurate measurement method may not be useful in practice as the classical computational runtime required for set up or optimization may be prohibitive and the quantum resources required too demanding. It is thus imperative to analyze the resources utilized in obtaining an accurate estimate through a measurement method. We further stress that we need to account for both classical and quantum resources as the subroutine of obtaining expectation values with respect to different quantum observables is inherently hybrid quantum-classical in nature, requiring different experiments to be executed on the quantum device and classical post-processing of the measurements in addition to pre-processing to decide the experiments themselves.

In the next section, we discuss a general measurement protocol and comment on resource utilization in the different steps of the protocol.

5.2.4 General measurement protocol

In this section, we describe the general procedure along with resource utilization for the problem of estimating $\mathrm{Tr}(\rho H)$ on n qubits, given measurement budget M. The measurement budget is equivalent to the number of total shots we are allowed gather from the quantum device or the number of times the device is queried.

The general procedure is schematically depicted in Figure 5.2 and involves three steps: (i) pre-processing on a classical computer (CPU), (ii) experiments on a quantum device (or QPU for quantum processing unit), and (iii) further post-processing of data acquired from the quantum device on a classical computer. We now describe each of these steps in detail. We will also explicitly state the benchmarking setup, execution, and analysis associated with each step. First, we describe the experiments executed on the quantum device as this decides the formulation of the pre-processing and post-processing steps.

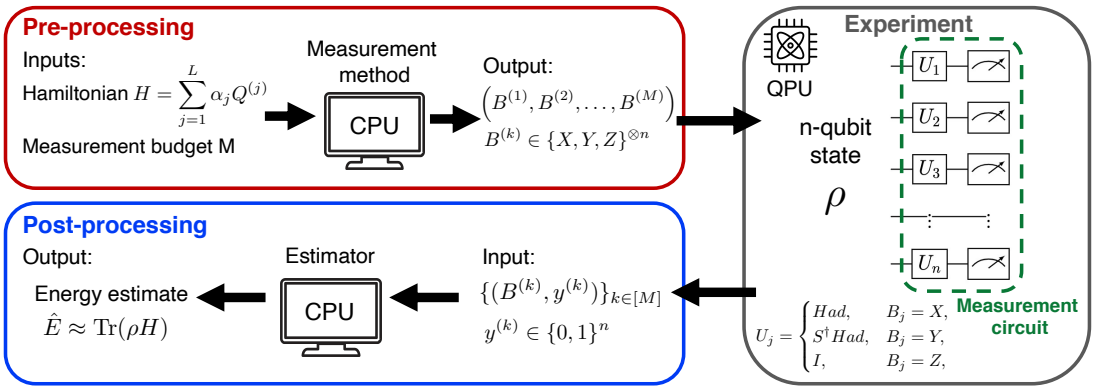


Figure 5.2: Schematic of estimation of $\mathrm{Tr}(\rho H)$: The procedure is divided into three steps of (i) pre-processing on a classical computer (CPU), (ii) experiments on a quantum device or quantum processing unit (QPU), and (iii) post-processing on the CPU. In (i), the measurement method plays the central role of deciding which measurement bases (denoted by B here) to execute on the QPU given inputs of an n-qubit Hamiltonian and measurement budget M. In (ii), experiments are executed on the QPU using the inputs of the Pauli measurement bases output by step (i). A measurement circuit corresponding to an arbitrary basis B is shown inset (Had denotes the Hadamard gate, and $S^\dagger$ denotes the inverse phase gate, which are used to transform the local measurement basis). Finally, in (iii) estimation is carried out on measurement results of the form (B, y) where B are the Pauli measurement bases and y are the corresponding measurement outcomes from the QPU.

Experiment. Each experiment on the quantum device involves the preparation of the n-qubit quantum state of interest followed by a measurement circuit. In an arbitrary step of VQE, this state would correspond to a parametrized quantum circuit or an ansatz with a certain set of assigned parameters. In quantum Krylov methods, this state may correspond to a certain time-evolved state. After the state is prepared, a measurement circuit is applied which involves application of single-qubit unitaries $\otimes_{i=1}^{n} U_i$ followed by a measurement in the computational basis. We denote the outcome of a measurement which is an n-bit binary string as $y \in \{0,1\}^n$. For any arbitrary qubit j, the single-qubit unitary U_j corresponds to measuring qubit j in a non-trivial Pauli basis

$$U_j = \begin{cases} Had, & B_j = X, \\ S^\dagger Had, & B_j = Y, \\ I, & B_j = Z, \end{cases} \tag{5.2}$$

where we have denoted the n-qubit Pauli basis as B, the subscript j denotes the Pauli matrix on qubit j, Had is the Hadamard gate, and $S = \mathrm{diag}(1, i)$ is the phase gate. As discussed earlier in Section 5.1, we restrict measurement circuits to involve measurements in the Pauli basis due to depth limitations on currently available noisy quantum hardware and hence the preference for shallow measurement circuits. In summary, the input to the experiment step for a measurement budget M is a set of M Pauli measurement bases $\{B^{(k)}\}_{k \in [M]}$ corresponding to the measurement circuits executed on the quantum device and the output from this

step is a set of measurement outcomes with each one corresponding to a measurement basis $\{(B^{(k)}, y^{(k)})\}_{k\in[M]}$. This output is then later used in the post-processing step for obtaining an estimate of $\mathrm{Tr}(\rho H)$. This completes the description of the experiment step.

The benchmarking setup includes defining the state preparation circuit (either after compiling a classical description of a digital quantum circuit to be implemented or setting the parameters in a parameterized quantum circuit), and the measurement circuits to be used for the chosen measurement bases. As part of the benchmarking execution, we need to define the type of quantum computing resource being utilized (e.g., QPU, modular, etc.) as well as any classical computing resource (e.g., CPU, distributed, parallel, etc.) utilized for compilation and error mitigation. Finally, as part of the benchmarking analysis, this step involves the quantum wall-clock time over the experiments executed and the metric of computational runtime associated with compilation or error mitigation.

Pre-processing. In this step, a measurement method is used to propose a set of Pauli measurement bases $\{B^{(k)}\}_{k\in[M]}$ that is inputted to the experiment step. The inputs to the measurement method are the n-qubit Hamiltonian H, measurement budget M, and any available prior information of the quantum state ρ. Taking the measurement basis operator on any particular qubit to be the identity I corresponds to not measuring that qubit, and hence does not reveal any information about a target observable Q. Therefore, we consider the alphabet of measurement bases as $\mathcal{Q} = \{X, Y, Z\}^{\otimes n}$ and call it the query space. We denote a distribution over $\mathcal{Q}$ as β and call it the query distribution. The probability mass associated with a Pauli operator $P \in \mathcal{Q}$ is given by $\beta(P)$.

The benchmarking setup includes defining the Hamiltonian H under consideration, quantum state ρ to be measured (either through a classical description of the state preparation circuit required to be implemented or parameters associated with a parameterized quantum circuit), and measurement method Σ_S to be used for generating the samples. As part of the benchmarking execution, we need to define the type of computational resource being utilized (e.g., CPU, distributed computing, parallelization, etc.). Finally, as part of the benchmarking analysis, this step only involves the metric of computational runtime associated with the measurement method generating M samples.

Post-processing. After measurement outcomes $\{(B^{(k)}, y^{(k)})\}_{k\in[M]}$ against M Pauli measurement bases are acquired in the experiment step, they are passed on to an estimator $\mathcal{E}$ in the post-processing step. Suppose there is a Pauli measurement basis B which is queried m_B times and the corresponding measurement outcomes are $\{y_B^{(k)}\}_{k\in[m_B]}$. We can then compute an estimate, which we denote by $\hat{\omega}(B)$, of $\mathrm{Tr}(\rho B)$ as follows:

$$\hat{\omega}(B) = \frac{1}{m_B}\sum_{k=1}^{m_B}\mu^{(k)}(B) = \frac{1}{m_B}\sum_{k=1}^{m_B}\prod_{j=1}^{n}(-1)^{y_j^{(k)}}, \tag{5.3}$$

where $\mu^{(k)}(B) = \prod_{j=1}^{n}(-1)^{y_j^{(k)}}$ is eigenvalue measurement of ρ in the basis B corresponding to the outcome $y^{(k)}$. But, it turns out we can do even more with these measurements. Suppose there is a target Pauli term Q in the decomposition of the Hamiltonian H (Eq. 5.1) which is covered by B, i.e., all the non-trivial single-qubit Pauli matrices in the tensor product

of Q coincide with those in B. The eigenvalue measurement of ρ in basis Q could then be obtained from the measurement outcomes of measuring ρ in basis B.

Let us make this more concrete by introducing some relevant notation. For a full weight Pauli operator B, we let $\mu(B, j) = (-1)^{y_j}$ denote the eigenvalue measurement when qubit j is measured in the basis B_j and as corresponding to measurement outcome y. For for a subset $A \subset [n]$, we define $\mu(B, A) := \prod_{j \in A} \mu(B, i) = \prod_{j \in A} (-1)^{y_j}$ with the convention that $\mu(B, \emptyset) = 1$. The eigenvalue measurement of ρ in basis Q corresponding to measurement outcome $y^{(k)}$ of measuring ρ in basis B is then given by $\mu^{(k)}(B, \text{supp}(Q))$. We thus note that eigenvalue measurements of ρ in basis B can then be used to obtain estimates of not only $\text{Tr}(\rho B)$ but also of $\text{Tr}(\rho Q)$ for all Q that are covered by B. These estimates can then be combined to give an estimate $\hat{E}(\rho)$ of $\text{Tr}(\rho H)$. The goal of the estimator $\mathcal{E}$ is to do this in a computationally efficient fashion while using the available measurements to come up with an accurate estimate. Here, we have hinted at the Monte-Carlo estimator [RC99]. In Section 5.3, we will give an overview of different estimators that can be used in the post-processing step.

The benchmarking setup includes defining the estimator $\mathcal{E}$ used. As part of the benchmarking execution, we need to define the type of computational resource being utilized (e.g., CPU, distributed computing, parallelization, etc.). Finally, the benchmarking analysis involves the accuracy metric of the output learning error (RMSE) and performance metric of computational runtime associated with running the estimator on the acquired data.

A sequential algorithmic viewpoint of the general protocol for observable estimation discussed thus so far is presented in Algorithm 7. While benchmarking different measurement methods, it will also be convenient to define the PEP-SEA matrix, used as an abbreviation of (P)re-processing (E)xperiment (P)ost-processing - (S)etup (E)xecution (A)nalysis matrix, that summarizes the different stages of benchmarking for each step of the general measurement protocol. The PEP-SEA matrix is summarized in Table 5.1.

We have noted that the design of estimators and measurement methods can be interdependent. For example, a trivial estimator could be designed to estimate $\text{Tr}(\rho B)$ for a measurement basis B that also occurs as a target Pauli term in the Hamiltonian H using the corresponding measurement outcomes but not use the same measurement outcomes to estimate $\text{Tr}(\rho Q)$ for a Pauli Q that is covered by B. This would put a severe restriction on sets of useful measurement bases to the experiment and limit the flexibility of the measurement methods. In this work, we only discuss estimators which are designed with compatibility of different Pauli operators in mind, and discuss which of those estimators may be equipped with the various measurement methods presented here.

In Section 5.3, we describe different estimators that can be used in post-processing. In Section 5.4, we discuss various measurement methods that either construct a query distribution β in order to sample measurement bases B from it or use a routine to sample measurement bases B without direct access to the underlying distribution. We will also comment on the compatibility of different measurement methods with different estimators in Section 5.4.

Algorithm 7 Estimation of $\mathrm{Tr}(\rho H)$ through different measurement methods

Input: Measurement budget M, Hamiltonian H, Measurement Method Σ_S, Estimator $\mathcal{E}$

1: **for** sample $m = 1 : M$ **do**
2: Generate measurement basis $B \in \{X, Y, Z\}^{\otimes n}$ using Σ_S
3: **for** qubit $k = 1 : n$ **do**
4: Measure qubit k in basis B_k
5: Save eigenvalue measurement $\mu(B, k) \in \{-1, +1\}$
6: **end for**
7: Update observable expectation $\hat{E}$ using estimator $\mathcal{E}$ on acquired eigenvalue measurements
8: **end for**

Output: $\hat{E}$

	Setup (S)	**Execution (E)**	**Analysis (A)**
Pre-processing (P)	Meas. budget: M Objects: H, ρ Meas. Method: Σ_S	Classical computational resource utilized (CPU, distributed, parallel)	Classical runtime
Experiment (E)	State preparation circuit for ρ Meas. circuit: $\{U_j\}_{j\in[n]}$ from $\{B_j\}_{j\in[n]}$	Quantum computing resource utilized (QPU, modular)	Quantum wall-clock time Classical wall-clock time (error mitigation, compilation) Quantum coherence time
Post-processing (P)	Estimator: $\mathcal{E}$	Classical computational resource utilized (CPU, distributed, parallel)	Classical runtime Output accuracy (RMSE)

Table 5.1: Definition of the PEP-SEA matrix.

5.2.5 Summary of performance metrics in CSHOREBench

To account for computational runtime in addition to the convergence of the algorithm at hand, one common approach used across different classical computational chemistry packages is to measure the wall-clock time required to achieve some threshold accuracy. As the learning task of estimating $\mathrm{Tr}(\rho H)$ is a subroutine in many hybrid quantum-classical algorithms, we need to account for the computational time spent on the quantum device, computational time spent on the classical device, and any latencies in between the quantum and classical hardware, to measure the overall wall-clock time. We refer to the total time spent on a quantum device as quantum wall-clock time and this includes time duration associated with experiment execution, measurements and resets. We refer to the total time spent on a classical device as classical wall-clock time, which includes setting up different measurement methods (e.g., optimization), post-processing of measurements (e.g., estimation), and compilation of quantum circuits to the native set of gates of the quantum hardware.

A performance metric for resource utilization could thus simply be the sum of the quantum wall-clock time and classical wall-clock time for a measurement method in reaching a cutoff of accuracy. However, quantum computers are not yet a mature technology compared to classical computers. This suggests that a stronger approach to benchmarking would be

to allow some flexibility in weighting the quantum and classical costs, since wall-clock time may not all be equivalent across the classical and quantum phases of the experiment.

Towards this end, we propose a heuristic in Section 5.6 to rank different measurement methods based on a weighted-sum of the quantum wall-clock time and classical wall-clock time to reach an specified cutoff of chemical accuracy in RMSE of the resulting energy estimates. We expect quantum devices to progress rapidly over the next few year, and so for the benchmarks to be robust to this progress, the weights should vary with time and be revised with new advances. A functional form of how these weights may change with time may be useful but is outside the scope of the current chapter.

5.3 Estimators

In this section we discuss different estimators $\mathcal{E}$, which take center stage in the post-processing step of the general protocol for estimating $\mathrm{Tr}(\rho H)$ (Figure 5.2) and that can be used alongside measurement methods as shown in Algorithm 7. Suppose we generate a set of M Pauli measurement bases through a specified measurement method. We will denote this set of measurements by $\mathbf{B} = \{B^{(s)}\}_{s\in[M]}$ where $B^{(s)} \in \Omega_n = \{X, Y, Z\}^{\otimes n}\, \forall s \in [M]$. Recall that we denote the corresponding measurement outcomes as $\mu(B^{(s)}) \in \{-1, +1\}^{\otimes n}$ with $\mu(B^{(s)}, j)$ the measurement outcome on qubit j.

The goal is to use the M examples of $\{(B^{(s)}, \mu(B^{(s)}))\}_{s\in[M]}$ to estimate $\mathrm{Tr}(\rho H)$. We do this by obtaining estimates of $\mathrm{Tr}(\rho Q^{(j)})$ which we denote by $\hat{\omega}^{(j)}$, using one of three possible estimators: a Monte Carlo (MC) estimator, a weighted Monte Carlo (WMC) estimator, and a Bayesian estimator, which we introduce in the next subsection. Given the $\hat{\omega}^{(j)}$, an estimate of $\mathrm{Tr}(\rho H)$ is obtained as

$$\hat{E}_G = \sum_{j=1}^{L} \alpha_j \hat{\omega}^{(j)}. \tag{5.4}$$

5.3.1 Monte-Carlo Estimator

Let us first define the *hit* function

$$h(Q^{(j)}; \mathbf{B}) = \sum_{s=1}^{M} \mathbb{1}\{Q^{(j)} \triangleright B^{(s)}\} \tag{5.5}$$

which counts the number of Pauli measurement bases that cover (hit) $Q^{(j)}$. The MC estimate of $\hat{\omega}^{(j)}$ is then simply given by

$$\hat{\omega}^{(j)} = \begin{cases} \frac{1}{h(Q^{(j)};\mathbf{B})} \sum_{s=1}^{M} \mathbb{1}\{Q^{(j)} \triangleright B^{(s)}\} \mu\left(B^{(s)}, \mathrm{supp}(Q^{(j)})\right), & h(Q^{(j)}; \mathbf{B}) \geq 1 \\ 0, \quad h(Q^{(j)}; \mathbf{B}) = 0. \end{cases} \tag{5.6}$$

This is merely the empirical average of the measurements corresponding to $Q^{(j)}$. As the MC estimator does not require access to a query distribution β, it can be used with any measurement methods, including those that generate measurement bases without giving us direct access to an underlying query distribution. Finally, it can be readily verified that the MC estimator is an unbiased estimator.

Laplace smoothing. In any collected dataset $D = \{(B^{(s)}, \mu(B^{(s)}))\}_{s\in[M]}$, there is a possibility that there are no measurements covering an observable $Q^{(\ell)}$ for some $\ell \in [L]$ and thus the hit $h(Q^{(\ell)}; \mathbf{B}) = 0$. This happens when the query distribution β assigns very low probability to measurement bases covering $Q^{(\ell)}$. For example, this occurs when the measurement method is based on L_1 sampling and the corresponding coefficient α_ℓ in the decomposition of H has a low magnitude relative to the 1-norm of the coefficients $\sum_{j=1}^{L}|\alpha_j|$.

Now suppose that in this situation where the probability of a measurement basis covering $Q^{(\ell)}$ is very low, we do get lucky and obtain a single measurement of it. When this occurs the estimate $\hat{\omega}^{(\ell)}$ will jump from the value 0 (prior to that shot) to the measurement outcome, either $+1$ or -1. This may result in large contribution to the uncertainty in the overall estimate of $\mathrm{Tr}(\rho H)$ from $\alpha_\ell\hat{\omega}^{(\ell)}$ even when α_ℓ is small, if the true value of $Q^{(\ell)}$ far from the obtained measurement outcome (as for example is guaranteed if the true value is close to 0). To avoid this, we artificially adjust the empirical probabilities via Laplace smoothing (also called additive smoothing) [MRS08] as follows

$$P(\lambda^{(j)} = 1) = \frac{m_0^{(j)} + \gamma}{m_0^{(j)} + m_1^{(j)} + 2\gamma}, \quad P(\lambda^{(j)} = -1) = \frac{m_1^{(j)} + \gamma}{m_0^{(j)} + m_1^{(j)} + 2\gamma}, \tag{5.7}$$

where γ is the smoothing parameter and $m_0^{(j)}$ (or $m_1^{(j)}$) are the number of measurements in D which cover $Q^{(j)}$ and correspond to an eigenvalue measurement of $+1$ (or -1). Formally, we have

$$m_k^{(j)} = \frac{1}{2}\sum_{s=1}^{M} \mathbb{1}\{Q^{(j)} \triangleright B^{(s)}\} \left(1 + (-1)^k \mu\left(B^{(s)}, \mathrm{supp}(Q^{(j)})\right)\right), \; k \in \{0, 1\}. \tag{5.8}$$

The resulting MC estimate of $\hat{\omega}^{(j)}$ with Laplace smoothing is then

$$\hat{\omega}^{(j)} = P(\lambda^{(j)} = 1) - P(\lambda^{(j)} = -1) = \frac{1}{h_\gamma(Q^{(j)}; \mathbf{B})} \sum_{s=1}^{M} \mathbb{1}\{Q^{(j)} \triangleright B^{(s)}\} \mu\left(B^{(s)}, \mathrm{supp}(Q^{(j)})\right), \tag{5.9}$$

where $h_\gamma(Q^{(j)}; \mathbf{B})$ is the smoothed hit function related to the original hit function as $h_\gamma(Q^{(j)}; \mathbf{B}) = h(Q^{(j)}; \mathbf{B}) + 2\gamma$. Note that for $\gamma = 0$, we have no smoothing and re-obtain the original MC estimator (Eq. 5.6). The value of $\gamma = 1$ corresponds to the case when we assume the prior probability of $P(\lambda^{(j)} = 1) = P(\lambda^{(j)} = -1) = 1/2$ are uniform. The value of $\gamma = 0.5$ corresponds to the case when the prior probabilities are the Jeffrey's prior. Typically, γ is set to a value in $(0, 1)$ or is treated as a hyperparameter to be fine-tuned later.

5.3.2 Weighted Monte-Carlo Estimator

In the MC estimator, all the eigenvalue measurements are weighted uniformly in computing $\hat{\omega}^{(j)}$. This can be modified by weighting the different samples non-uniformly as follows:

$$\hat{\omega}^{(j)} = \frac{1}{M}\sum_{s=1}^{M} w^{(s)} \mu\left(B^{(s)}, \mathrm{supp}(Q^{(j)})\right), \tag{5.10}$$

where we have introduced weights $\{w^{(s)}\}_{s\in[M]}$ that satisfy $\sum_s w^{(s)} = 1$. This is often desirable to ensure stability of estimation [Cas+96], for incorporating prior information or for the purpose of importance sampling when a proposal distribution is used instead of the query distribution β for generating samples of measurement bases B.

Here we consider weights based on the query distribution, which can be interpreted as a self-normalization. Let the probability of a Pauli operator Q being covered by a query distribution β be denoted by $\xi(Q,\beta)$. This is equal to the probability with respect to β of generating a sample measurement basis B that covers Q:

$$\xi(Q,\beta) = \sum_{B\in\Omega_n} \mathbb{1}\{Q \triangleright B\}\beta(B), \tag{5.11}$$

where we have assumed that the alphabet of β is Ω_n. The resulting WMC estimator from setting the weights as $w^{(s)} = \mathbb{1}\{Q^{(j)}\triangleright B^{(s)}\}/\xi(Q^{(j)},\beta)$ (with $w^{(s)}$ set to 0 if $\mathbb{1}\{Q^{(j)}\triangleright B^{(s)}\} = 0$) is then given by

$$\hat{\omega}^{(j)} = \frac{1}{M}\sum_{s=1}^{M} \frac{\mathbb{1}\{Q^{(j)} \triangleright B^{(s)}\}}{\xi(Q^{(j)},\beta)} \mu\left(B^{(s)}, \mathrm{supp}(Q^{(j)})\right). \tag{5.12}$$

We note that the expectation of the hitting function $\mathbb{E}_\beta[h(Q^{(j)};\mathbf{B}] = M\xi(Q^{(j)},\beta)$. We can thus interpret the weighting in the WMC estimator as assigning a value to $\mathbb{1}\{Q^{(j)} \triangleright B^{(s)}\}/h(Q^{(j)};\mathbf{B})$ according to β and not through the samples actually obtained, in contrast to the MC estimator of Eq. 5.6.

The WMC estimator can be used when the measurement method has access to the query distribution β and we can evaluate $\beta(P)$ for an n-qubit Pauli operator P. In literature, WMC estimators have been used where β is a product distribution [HKP20; Had+22] as well as for more general cases [Hil+21]. It is not immediately clear whether the MC or WMC estimator would be preferred in practice. To answer this question, we compare the behavior of these estimators in Section 5.6.1 in combination with various applicable measurement methods.

Remark. The MC estimator (Eq. 5.6) and WMC estimator (Eq. 5.12) may display different behaviors for low number of samples. However, it can be shown they are asymptotically equivalent.

5.3.3 Bayesian Estimator

We now take a Bayesian approach in obtaining estimates $\hat{\omega}^{(j)}$ of $\mathrm{Tr}(\rho Q^{(j)})$. The estimates $\omega^{(j)}$ are obtained through single-shot measurements of ρ in a Pauli basis B that covers $Q^{(j)}$ (i.e., $Q^{(j)} \triangleright B$), which we denote by the random variable $\lambda^{(j)}$ which takes values in $\{+1,-1\}$. Let the underlying probability of observing $\lambda^{(j)}$ to be $+1$ (-1) be $\theta_0^{(j)}$ $(\theta_1^{(j)})$. We then model the random vector $\theta^{(j)} = (\theta_0^{(j)}, \theta_1^{(j)})$ as a Dirichlet random variable of order 2 and with hyperparameters $a = (a_1, a_2)$, i.e., $P(\theta^{(j)}) = \mathrm{Dir}(2, a)$. We set $a = (1,1)$ which corresponds to a uniformly distributed prior.

Given a data set of M measurements $D = \{(B^{(s)}, \mu(B^{(s)}))\}_{s\in[M]}$ as follows, we can then update the probability distribution over $\theta^{(j)}$ as

$$P(\theta^{(j)}|D) = \frac{P(D|\theta^{(j)})P(\theta^{(j)})}{P(D)} \tag{5.13}$$

where we have denoted $P(\theta^{(j)})$ as the prior probability of $\theta^{(j)}$, $P(D|\theta^{(j)})$ as likelihood or conditional probability of obtaining the measurements in D given $\theta^{(j)}$, $P(D)$ as the evidence, and $P(\theta^{(j)}|D)$ as the posterior probability of $\theta^{(j)}$ given D.

Typically, computing the posterior distribution through Bayes law (Eq. 5.13) is expensive and one needs to resort to alternate approximate methods such as Markov Chain Monte Carlo sampling and particle filter methods [Sär13]. However, in this case, the prior $P(\theta^{(j)})$ is a conjugate prior to the likelihood distribution, which is given by

$$P(D|\theta^{(j)}) = \left(\theta_0^{(j)}\right)^{m_0} \left(\theta_1^{(j)}\right)^{m_1}, \quad m_k = \frac{1}{2} \sum_{s\in[M]} \mathbb{1}\{B^{(s)} \triangleright Q^{(j)}\} \left(1 + (-1)^k \mu\left(B^{(s)}, \mathrm{supp}(Q^{(j)})\right)\right), k \in \{0,1\}. \tag{5.14}$$

This means that our prior $P(\theta^{(j)})$ and posterior $P(\theta^{(j)}|D)$ both belong to the same family of distributions, which in this case are Dirichlet distributions. The posterior is simply given by $P(\theta^{(j)}|D) = \mathrm{Dir}(2, a+m)$ where $m = (m_0, m_1)$. Ultimately, the posterior distribution is given by

$$P(\theta^{(j)}|D) = \frac{\left(\theta_0^{(j)}\right)^{m_0} \left(\theta_1^{(j)}\right)^{m_1}}{B(m_0 + a_0, m_1 + a_1)} \tag{5.15}$$

where $B(k_1, k_2) = \Gamma(k_1)\Gamma(k_2)/\Gamma(k_1 + k_2)$ and $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$ is the gamma function. In practice, we may receive datasets sequentially and the Bayes law (Eq. 5.13) is then computed sequentially with the posterior becoming the prior for the next dataset. Finally, an estimate of the expected value of $\omega^{(j)}$ or its variance can be obtained as

$$\mathbb{E}[\omega^{(j)}] = 2\mathbb{E}[\theta_0^{(j)}] - 1 = \frac{m_0 - m_1}{m_0 + m_1 + 2} \tag{5.16}$$

$$\mathrm{Var}[\omega^{(j)}] = 4\mathbb{E}[\theta_0^{(j)}(1 - \theta_0^{(j)})] = 4\frac{(m_0+1)(m_1+1)}{(m_0+m_1+2)(m_0+m_1+3)} \tag{5.17}$$

Thus, for carrying out Bayesian estimation of $\omega^{(j)}$, it is enough to keep track of the cumulative number of shots corresponding to measurements of $Q^{(j)}$ yielding values of $+1$ and -1. We also observe that the expected value of the estimate $\omega^{(j)}$ coincides with our MC estimate (Eq.) with Laplace error correction of $\gamma = 1$. This is not pure coincidence as Laplace error correction can be motivated through Bayesian estimation.

So far, we have shown how to carry out Bayesian estimation of $\omega^{(j)}$, but in many instances other quantities such as the covariance of estimates of two different target Pauli operators may be of interest. For a discussion on Bayesian estimation in this latter case, we refer the reader to [Shl+23, Appendix B].

5.4 Measurement Methods

In this section, we give an overview of different measurement methods Σ_S that can be used in the pre-processing step of the general protocol (Section 5.2.4). We will primarily focus on randomized measurements where, as the name suggests, there is randomization involved in producing the measurement bases. We also consider how these methods may be derandomized. Among randomized measurements, there are some methods that involve the explicit construction of a query distribution β (with respect to the alphabet $\Omega_n = \{X, Y, Z\}^{\otimes n}$) from

which we can sample to produce our measurement bases, and other methods that involve a routine that allows us to directly produce samples of measurement bases.

To guide our discussion on randomized measurements, we introduce decision diagrams as an efficient data structure for representing query distributions and describe multiple measurement methods. As we introduce these different measurement methods, we will also mention their compatibility with the different estimators presented in Section 5.3.

5.4.1 Randomized Measurements

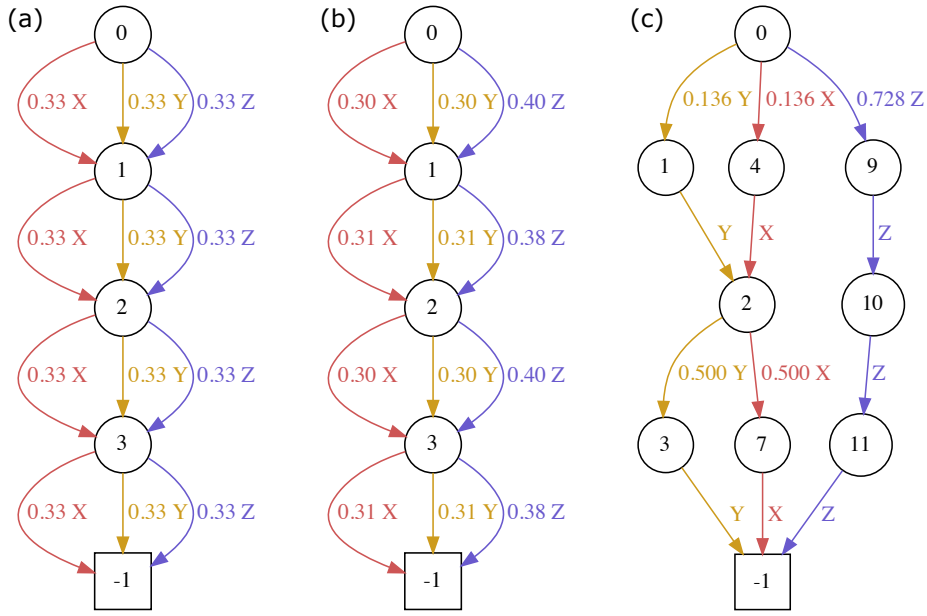


Figure 5.3: Instances of different decision diagrams for H_2 (4 qubits, sto3g basis, JW encoding). We show (a) uniform classical shadows, (b) locally biased classical shadows (LBCS) and (c) an optimized compact decision diagram.

We will start off by considering randomized measurement methods where the central object is a query distribution β (with respect to the alphabet $\Omega = \{X, Y, Z\}^{\otimes n}$) from which Pauli measurement bases are sampled. We could describe a general β through the probability assignments over $3^n - 1$ Pauli matrices in Ω. [1] However, this is exponentially expensive. Moreover, molecular Hamiltonians have $O(n^4)$ Pauli terms [McA+20], and these are the focus of this chapter. We could thus give probability assignments for β over Pauli operators in Ω that cover the different Pauli terms in H. In fact, we can do better as many of the Pauli terms in H can be grouped together, i.e., covered by a single Pauli measurement basis B.

Decision diagrams [Hil+21] give us a compact way to represent a query distribution β that takes advantage of this highlighted structure in H and represent probability assignments for different B in Ω_n efficiently. For our purposes, decision diagrams are acyclic directed graphs. In Figure 5.3, we show different decision diagrams for an H_2 Hamiltonian, corresponding

[1] The unassigned probability can be determined through the normalization $\sum_{x \in \Omega} \beta(x) = 1$.

to different measurement methods (to be discussed later). Common to all these decision diagrams are that the parent node is set to be 0 and the last child node is denoted by -1. Going from node 0 to -1, we have to take a path that takes n directed edges.

To generate a measurement basis from a decision diagram, we follow a path from node 0 to node -1, at each step choosing the next edge from the current node with probability given by its weight. Each edge is also labeled X, Y, or Z, so each choice of an edge in the path represents a choice of local Pauli basis for a corresponding qubit. Each qubit corresponds to a layer in the decision diagram, so the path from node 0 to node -1 defines a choice of local basis for each layer and thus each qubit. Thus decision diagrams can represent a very flexible family of query distributions β in $O(\text{poly}(n))$ memory and allow us to efficiently generate samples from β as well. We now describe how different randomized measurements that have been proposed can be seen as instances of decision diagrams.

Classical shadows. The uniform classical shadows (CS) of [HKP20] considers the query space $\mathcal{Q} = \Omega_n = \{X, Y, Z\}^{\otimes n}$ and the query distribution $\beta(B) = 3^{-n}, \forall B \in \mathcal{Q}$. This corresponds to uniformly randomly picking a measurement basis from $\{X, Y, Z\}$ for each qubit. Despite its simplicity, it was shown in [HKP20] that for a set of L Pauli observables $Q^{(j)}$, using $M = O(\log L/\epsilon^2)$ (factors depending on weight of $Q^{(j)}$ hidden) samples generated from this query distribution suffice to obtain estimates of the expectation values $\text{Tr}(\rho Q^{(j)})$ of all $Q^{(j)}$ up to additive error ϵ. It was also shown that this is asymptotically optimal. A challenge with using CS in practice is the potentially many different circuits required to be compiled and run, while current hardware favors repetitions of the same circuit. Moreover, CS does not incorporate any available prior information of the Hamiltonian H or state ρ. This led to the extension of CS to other randomized methods which we will discuss shortly. We note that all the estimators (MC estimator, importance MC estimator and Bayesian estimator) discussed in Section 5.3 are compatible with this measurement method.

Locally Biased Classical Shadows (LBCS). LBCS [Had+22] is a randomized measurement method that incorporates prior information about the state ρ. In LBCS, the query distribution β over $\mathcal{Q} = \{X, Y, Z\}^{\otimes n}$ is always a product distribution

$$\beta = \prod_{i=1}^{n} \beta_i \tag{5.18}$$

where β_i is the marginal probability of the ith qubit Pauli matrix B_i, but unlike CS, the β_i are not required to be uniform. Instead, the query distribution β is optimized by minimizing the one-shot variance of the estimate $\hat{E}$ with respect to a reference state, subject to the constraint that it has the form (5.18). Ideally, the reference state would be the target state, e.g., a ground state, but $\hat{E}$ is unknown *a priori* for that state. Next best would be to choose a heuristic reference state in the same way one would choose an initial state for VQE or other ground state preparation algorithms, e.g., the Hartree-Fock state, but this leads to a non-convex optimization. LBCS overcomes this by instead minimizing the variance of the estimate against the maximally mixed state, which results in a convex optimization [Had+22]. It has been numerically shown that LBCS can yield more accurate estimates of $\text{Tr}(\rho H)$ than CS on various chemistry Hamiltonians for the same measurement budget [Had+22]. Like

CS, all the estimators (MC estimator, importance MC estimator, and Bayesian estimator) presented in Section 5.3 are compatible with this measurement method.

Compact Decision Diagrams. Both of the above measurement methods are restricted to β being a product distribution, i.e., the marginal distributions over each qubit are independent of each other. However, in general, there will be correlations between measurements of $\mathrm{Tr}(\rho P_j)$ and $\mathrm{Tr}(\rho P_k)$, so is desirable to allow the query distribution account for such correlations. This is where general decision diagrams fit in by allowing us to represent general non-product joint distributions efficiently for molecular Hamiltonians. We refer the reader to [Hil+21; Mat+24] on how to construct compact decision diagrams given the Pauli decomposition of Hamiltonian H (Eq. 5.1) and a qubit ordering.

Once an initial decision diagram is obtained, the edge weights of the DD can be further optimized by minimizing the one-shot variance of the estimate $\hat{E}$ if we have some prior knowledge on the quantum state ρ. In practice, we find that even optimizing against the maximally mixed state is beneficial, like in LBCS. Unlike in LBCS, however, this does not yield a convex optimization problem. A solution at the local minima is still advantageous, yielding higher accuracy for the same measurement budget compared to LBCS as was shown in [Hil+21].

We have so far seen how decision diagrams are useful in representing query distributions β and how one can efficiently generate samples from β on them. In the following sections, we discuss Adaptive Pauli Shadows (APS) and how derandomization of decision diagrams may be performed.

5.4.2 Adaptive Pauli Measurements

We have so far focused on randomized measurement methods that involve construction of a query distribution β that we can directly sample from to generate the samples. Moreover, β remains unchanged during the sampling process. In contrast, Adaptive Pauli Shadows (APS) method [Had21] allows us to sample from β that changes adaptively during the sampling process.

As in Algorithm 7, we index samples by m. Each time we generate a measurement basis $B^{(m)}$, we iteratively sample each single-qubit basis $B_j^{(m)} \in \{X, Y, Z\}$ in some qubit ordering, which is chosen randomly. On the jth qubit in this ordering, we sample a measurement Pauli matrix according to the probability distribution that is the solution to

$$\begin{aligned} \text{minimize:} \quad & \sum_{Q^{(k)} \in \Omega} \frac{\alpha_k^2}{\beta(Q_j^{(k)})} \\ \text{subject to:} \quad & 0 \leq \beta(B_j) \leq 1, \quad \text{for } B_j \in \{X, Y, Z\} \\ & \sum_{B_j \in \{X,Y,Z\}} \beta(B_j) = 1 \end{aligned} \tag{5.19}$$

where the set Ω is defined as

$$\Omega = \{Q \in \mathbf{Q} | Q_j \in \{X, Y, Z\} \text{ and } Q_{j'} \in \{I, B_{j'}^{(s)}\} \ \forall j' < j\} \tag{5.20}$$

and we have used β in place of β_j. An analytical solution based on Lagrange multipliers can be developed for this convex optimization problem. We point the reader to [Had21] for the solution. Iterating over the qubits allows us to generate the sample $B^{(s)}$. The additional cost of this sampling process is only $O(n)$.

The idea of using a distribution solving (5.19) is to appropriately include the information about choices of basis on the previous qubits (in the ordering) in the choice of basis on the jth qubit. The key point to notice is the definition of the set Ω in (5.20): this contains all Pauli operators that are (i) in the target observable (i.e., in the set $\mathbf{Q}$), (ii) X, Y, or Z on the qubit currently being decided, and (iii) compatible with the part of the basis chosen so far.

Constraints (ii) and (iii) are the important ones, and what really separate APS from other measurement methods. Constraint (ii) is used so that the choice on qubit j is independent of operators that are I on qubit j, since these are compatible with any choice on qubit j. Constraint (iii) is used so that the choice on qubit j is independent of operators that are already incompatible with the basis given the choices on the previous qubits. Thus the choice of basis for qubit j only depends on operators whose inclusion in the covered set actually depends on that choice.

As we do not have access to the joint query distribution β for each sample, the Weighted-MC estimator (Eq 5.12) cannot be used with this method. The MC and Bayesian estimators can be used.

5.4.3 Derandomization

We have so far discussed randomized measurement methods that involve random generation of measurement bases. Often, however, it is desirable to have a deterministic sequence of measurement bases to make and that can be repeated across different experiments while retaining the performance of randomized algorithms.

In this section, we introduce the idea of derandomizing the randomized measurements that are obtained by sampling different query distributions β that may correspond to CS, LBCS or in general a decision diagram. Derandomization of CS was proposed in [HKP21] which we extend in a straight-forward fashion to derandomization of a general query distribution β represented on a decision diagram. Notably, we show how relevant computations carried out during derandomizing a general query distribution β can be implemented efficiently when one has access to the corresponding decision diagram.

Let us now introduce some notation that will be relevant to our discussion on derandomization. Recall that the target observables that we are interested in are $\mathbf{Q} = \{Q^{(j)}\}_{j\in[L]}$. Departing slightly from the goal we have considered so far, consider the goal of estimating $\mathrm{Tr}(\rho Q^j)$ within error ϵ for any $j \in [L]$. We denote these estimates as $\hat{\omega}^{(j)}$ and denote the true value of $\mathrm{Tr}(\rho Q^{(j)})$ by $\omega^{(j)}$. This is once again achieved through M single shot measurements of the following Pauli operators: $\mathbf{B} = \{B^{(s)}\}_{s\in[M]}$.

Through derandomization of a randomized algorithm, it is possible to come up with a partially or fully fixed sequence of measurements to be carried out. The key idea is to understand how estimates typically deviate from the truth in a randomized algorithm and how this changes when conditioned upon previous measurements. One way to measure this

deviation is through the *confidence bound* introduced in [HKP21]:

$$\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B}) := \sum_{j=1}^{L} \exp\left(-\frac{\epsilon^2}{2} h(Q^{(j)};\mathbf{B})\right), \tag{5.21}$$

where $h(\cdot)$ is the hit function (Eq. 5.5). It was shown in [HKP21, Lemma 1] that if the confidence bound is upper bounded by $\delta/2$ for some $\delta \in (0,1)$ then each of the empirical estimates $\hat{\omega}^{(j)}$ are within ϵ-distance of the truth $\omega^{(j)}$, with probability at least $1-\delta$.

Derandomization of DD is then completed through the following steps: (i) obtain a confidence bound on estimates for DD, (ii) analyze the confidence bound when conditioned on prior measurements, and (iii) use this conditional expectation bound to design a cost function that will be used for the derandomizing procedure. This procedure is outlined in Appendix C.1. Here, we state the cost function in derandomization of DD.

To motivate the cost function for derandomization of DD, consider the following scenario. Suppose we are given a measurement budget of M and that $\mathbf{B}^\#$ contains the assignments of measurement bases for the first $(m-1)$ samples and first k qubits of the mth measurement basis. That is, we have already generated the first $(m-1)$ samples and Paulis of the first k qubits of the mth measurement basis. We then have the following conditional expectation of the confidence bound (see Appendix C.1)

$$\begin{aligned}\mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})|\mathbf{B}^\#\right] = &\sum_{j\in[L]}\prod_{m'=1}^{m-1}\left(1-\eta\mathbb{1}\{Q^{(j)} \triangleright B^{\#(m')}\}\right)\\ &\times\left(1-\eta\prod_{k'=1}^{k}\mathbb{1}\{Q^{(j)} \triangleright B^{\#(m')}\}\Pr\left[Q^{(j)}_{k+1:n} \text{ covered by DD}|B^{\#(m)}_{1:k}\right]\right)\\ &\times\left(1-\eta\Pr\left[Q^{(j)} \text{ covered by DD}\right]\right)^{M-m},\end{aligned}$$

where $\eta = 1-\exp(-\epsilon^2/2)$. To choose the assignment of the kth qubit of the mth measurement basis, we consider the following cost function

$$B_k^{\#(m)} = \underset{W\in\{X,Y,Z\}}{\mathrm{argmin}}\ C(W) = \underset{W\in\{X,Y,Z\}}{\mathrm{argmin}}\ \mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})|\mathbf{B}^\#, B_k^{(m)} = W\right] \tag{5.22}$$

where $\mathbf{B}^\#$ now corresponds to the assignments of measurement bases over the first $(m-1)$ samples and $(k-1)$ qubits of the mth measurement basis. Note that the above cost function requires the input of the experimental budget M.

An algorithm for derandomization of DD is given in Algorithm 8. Note that all the steps in the algorithm can be computed efficiently on a decision diagram for molecular Hamiltonians. Details are given in Appendix C.1.

Algorithm 8 Derandomization of Decision Diagrams

Input: Measurement budget M, accuracy ϵ, target observables $\mathbf{Q} = \{Q^{(j)}\}_{j\in[L]}$, decision diagram Σ_{DD}
Output: Set of M measurement bases $\mathbf{B}^{\#}$

for $m = 1 : M$ **do**
  for $k = 1 : n$ **do**
    for $W \in \{X, Y, Z\}$ **do**
      Compute $C(W) = \mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})|\mathbf{B}^{\#}, B_k^{(m)} = W\right]$
    end for
    Set $B_k^{\#(m)} \leftarrow \mathrm{argmin}_{W\in\{X,Y,Z\}} C(W)$
  end for
end for
return $\mathbf{B}^{\#}$

The derandomized measurement procedure is compatible only with the MC estimator (Section 5.3.1) and Bayesian estimator (Section 5.3.3) that we have described so far.

Remark. Both measurement methods of APS and derandomization attempt to bring in adaptivity into the sampling process. However, there are qualitative differences besides how these methods themselves are motivated and set up. Derandomization fixes the Pauli operator for a qubit by solving an optimization problem while APS obtains an optimized marginal distribution over Paulis on a qubit and then allows for a randomized Pauli sample. Measurement history is taken into account of derandomization but not in APS.

5.5 CSHOREBench: Common States and Hamiltonians for ObseRvable Estimation Benchmark

Having discussed different randomized and derandomized measurement methods at our disposal for the learning problem of estimating $\mathrm{Tr}(\rho H)$ (Section 5.2) in Section 5.4 and estimators that can be used in conjunction with these methods in Section 5.3, we are now in a position to compare the performance of these different measurements in practice.

In Section 5.2.3, we laid the inspiration and strategy for CSHOREBench. We now formally describe the setup, execution and following analysis, discussed in the context of general measurement protocols in Section 5.2.4, to be carried out as part of this benchmark. CSHOREBench along with the implementation of all measurement methods (Section 5.4) and estimators (Section 5.3) can be found in a Github repository [2].

[2] https://github.com/arkopaldutt/RandMeas

5.5.1 Common quantum computation task and common objects

In Section 5.2.3, we mentioned the types of Hamiltonians and states that we would benchmark against to assess the performance of different measurement methods. We now explicitly state these objects. For CSHOREBench, we consider a set of small molecular electronic Hamiltonians that have been encoded into qubit systems. The molecules that we specifically consider as part of the benchmark here, are listed in Table 5.2. The qubit Hamiltonians are obtained by first representing each molecule by a fermionic Hamiltonian in a particular molecular orbital basis which is then mapped to a qubit Hamiltonian using the Jordan-Wigner (JW) [JW93] encoding. In our experiments we found similar results against the Bravyi-Kitaev [BK02b], and parity [BK02b; SRL12] encodings. We thus only report our results for the JW mapping.

As mentioned earlier in Section 5.2.3, given a Hamiltonian H, the set of states considered are those which we expect during the runtime of a hybrid quantum-classical circuit for ground state estimation such as VQE. These include a properly chosen initialization (e.g., Hartree-Fock (HF) state), target state (e.g., ground state) and appropriately chosen parametrized quantum circuits (also called quantum ansatz). For the purpose of the benchmarking experiments in this chapter, we consider the HF state, ground state and a random quantum ansatz. Another motivation for including random quantum ansatz as one of the states to test on is simply because in most cases, the ground state is difficult to prepare on a near-term quantum device and it is the state that the hybrid quantum-classical algorithm is trying to optimize for.

An overview of the different molecules and the experiments carried out on them are given in Table 5.2. These include tapered Hamiltonians [Bra+17] where qubit degrees of freedom corresponding to exact Z_2 symmetries of a Hamiltonian are removed. At least two qubits can always be removed, corresponding to spin and electron number parity, and often more Z_2 symmetries are available, which can typically be attributed to the point group of the molecule [Bra+17; Set+20].

We will describe the experimental protocol followed in the execution of CSHOREBench in Section 5.5.3. It should be noted that while we consider a small set of Hamiltonians in this work, CSHOREBench can be performed on any set of molecular Hamiltonians such as those included in the QM databases [BR09; Rud+12].

Molecule	**Number of qubits**	**Basis**
H_2 (tapered)	5	3-21g
HeH^+ (tapered)	6	3-21g
HeH^+	8	6-31g
LiH	12	sto6g
N_2	16	sto6g

Table 5.2: Molecules considered for benchmarking of measurement methods on the simulator and quantum device. The Jordan Wigner encoding is considered. For each Hamiltonian, the state is set to be the ground state on the simulator and a random ansatz on quantum devices.

5.5.2 Performance metrics in CSHOREBench

In this section, we describe the different metrics considered as part of the analysis stage of benchmarking different measurement methods in CSHOREBench. In Figure 5.1, we showcased two metrics of relevance, accuracy and runtime, which were later contextualized in Section 5.2.4.

Accuracy. In assessing measurement methods, the main performance metric is the accuracy or learning error achieved given a measurement budget M. For various measurement protocols, we track the root mean squared error (RMSE) with increasing values of measurement budget. This also allows us to investigate the convergence behaviour of different measurement methods and numerically determine their sample complexities.

Runtimes. As we identified in Section 5.2.4 and illustrated in Figure 5.1, each step of the measurement protocol (pre-processing, experiment, and post-processing) has a computational runtime associated with it. For the pre-processing and post-processing steps, this is the classical computational runtimes as these are executed on classical computing. For experiments, this is a classical computational runtime when executed on a simulator and the quantum wall-clock time when executed on a quantum device. Here, we keep track of classical computational runtime as wall-clock time instead of number of FLOPS (floating point operations). It should be noted that these computational runtimes are then specific to our implementation of the measurement methods and estimators.

Classical latencies. To execute experiments on a quantum computer according to the measurement bases generated on a classical computer, there are multiple latencies involved such as compilation of circuits to the native gate set of the quantum computer, loading of circuits through the control electronics interfacing with the quantum computer, and classical post-processing associated with measurement error mitigation. On current control electronics, it is more expensive to run any number of shots over many different circuits than the same number of shots against the same circuit. Thus, we also track the number of unique measurement circuits (or measurement bases) requested by different measurement methods to reach a given value of accuracy and the distribution of shots across different circuits.

Summary of resource utilization. As suggested in Section 5.2.5, we attempt to summarize resource requirements by answering the following question: *How much classical and quantum resources are utilized by a given measurement method Σ_M to reach a cutoff of accuracy 5 milli Hartree in estimating $Tr(\rho H)$ for a given Hamiltonian H and ρ?*

Ideally, we would have chosen the cutoff to be chemical accuracy or 1.593 milli Hartree. However, many of the candidate measurement methods considered here would require far too high measurement budget to reach this cutoff and would not be possible to verify this reasonably in experiment on quantum devices. In the above question, we have chosen 5 milli Hartree as a cutoff as this is achieved by various measurement methods on a simulator and in experiments on a quantum device. The resource utilization at this cutoff should be representative of that at chemical accuracy. In answering the above question, we will

primarily account for classical and quantum runtimes. We will ignore classical latencies as control electronics hardware is fast evolving.

As noted earlier, we cannot simply add the classical and quantum wall-clock times to obtain a value associated with overall resource utilization due to the differing maturities of classical and quantum computing technologies. Rather, we introduce a heuristic for resource utilization that takes the form of a weighted sum of the wall-clock times:

$$\mathrm{R} = w_c \cdot R_{\text{classical}} + w_q \cdot R_{\text{quantum}}, \tag{5.23}$$

where w_c is the weight corresponding to classical computers, w_q is the weight corresponding to quantum computers and $R_{\text{classical}}$ (or R_{quantum}) is the resource utilization by the corresponding type of computing and measured in wall-clock time here. The weights have units of s^{-1} to make R dimensionless and can be interpreted as resources used per second.

The question then arises: How do we design these weights? We could consider a common task for both classical and quantum computers, and then compare their performances. For example, in comparing CPU-centric classical computers and those with access to GPUs, a common task is solving pseudorandom dense linear systems. Towards this, LINPACK has been extended for these computers [DLP03; Jo+15; Pet+18]. Here, as well, we could consider the common task of solving random linear systems of equations on both classical and quantum computers, albeit with different input and output models. However, there have not been any experiments on quantum devices solving large-scale linear systems using the HHL algorithm.

Another way to design these weights is to consider the rate of logical instructions. A natural metric for classical computers is that of FLOPS (floating point operations per second) and for quantum computers is that of CLOPS (circuit layer operations per second) [Wac+21]. A general purpose classical computer has a speed of around 6 GFLOPS and the IBM quantum computer (`ibmq_mumbai`) which we primarily used for our experiments had a speed of around 2.4 KCLOPS during our experiments. Comparing these speeds gives us weights of $(w_c, w_q) = (1, 2.5 \times 10^6)$. However, designing weights in such a manner has flaws of neglecting the finite coherence time of quantum computers, and not accounting for different energy costs of these computing resources. To get around this, we instead suggest different regimes for weights in Table 5.3 and which allows the user of the benchmark to incorporate their own preferences.

Regime	**Weights** (w_c, w_q)
A: Fast QPU	$(1, 1)$
B: Medium Fast QPU	$(1, 1.5 \times 10^2)$
C: Medium Slow QPU	$(1, 2 \times 10^4)$
D: Slow QPU	$(1, 2.5 \times 10^6)$

Table 5.3: Different regimes of weights for classical and quantum computers in our heuristic.

5.5.3 Experimental protocol for CSHOREBench

We have so far described the objects in CSHOREBench and the different performance metrics considered as part of the analysis in CSHOREBench. Let us cross the rest of our SEAs. We

will benchmark against the following candidate measurement methods, previously described in Section 5.4, and included as part of the pre-processing setup: uniform classical shadows (CS), locally biased classical shadows (LBCS), decision diagrams (DD), and derandomization of these methods (Derand. CS, Derand. LBCS and Derand. DD) respectively. We do not include APS (see Appendix C.2.3). In addition to these measurement methods, we will carry out experiments with different estimators (Section 5.3).

We will now describe the experimental protocol followed for each combination of measurement procedure and estimator on a given Hamiltonian. Experiments are executed on an ideal classical simulator and IBM quantum devices. On the simulator, we assume that copies of an unknown quantum state ρ are given to us. On the quantum device, we assume access to a quantum circuit preparing the unknown quantum state ρ. This may be in the form of an ansatz as one would have during a step of the VQE algorithm. In each experiment, the goal is to then estimate the energy $E = \mathrm{Tr}(\rho H)$ of the encoded n-qubit Hamiltonian H.

To compare the performance of the different measurement procedures, we use the metric of root mean square error (RMSE) of the ground state energy against a budget of M Pauli measurements. The measurement budget M will be specified for different cases later. We compute RMSE by independently repeating each experiment with measurement budget M, N times so that we have access to N independent estimates of the energy $\{\hat{E}_G^{(i)}\}_{i\in[N]}$. RMSE is then computed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{E}_G^{(i)} - E_G\right)^2}, \tag{5.24}$$

where E_G is the true energy. We have access to the true energy on the simulator. On the devices, we replace this with the empirical mean computed over N runs.

Classical simulator The ground state and ground energy of each n-qubit molecular Hamiltonian is determined through the Lanczos method [MT20]. The simulator is then initialized with this ground state and this quantum circuit is provided to the measurement procedure in lieu of ρ. As described in Section 5.2.4, depending on the Pauli measurement basis, this quantum circuit is then appended with a measurement circuit and the qubits are measured in computational basis to produce a single shot (of measurement outcomes). All the simulator runs are executed on a cluster which has `Intel Xeon Platinum 8260 (2.4 GHz)` nodes. The classical computational runtimes reported here correspond to these runs and are specific to this CPU. A constraint from the cluster is that all simulations must be completed within four days of wall-clock time and this is imposed on our runs as well. This sets a constraint for the measurement budget on some of the measurement methods being considered.

Quantum device We also benchmark the different measurement methods on IBM quantum devices which include the 16-qubit `ibmq_guadalupe` and 27-qubit `ibmq_mumbai`. Rather than considering access to the ground state as on the simulator, we consider the state preparation circuit to be an excitation preserving ansatz [Bar+18] of depth 25 for our experiments. We then compare the performance of different measurement procedures on the ansatz with a set of (fixed) randomly assigned parameters. A relevant constraint from the hardware side

is that each job on the IBM quantum devices are limited to have 300 different circuits and 10^5 shots each. Further, a sequence of jobs is given a maximum time allocation of 8 hours. As we will see later, this constrains different measurement methods (e.g., CS and LBCS) which generate many unique circuits, that can be benchmarked on the quantum device on large molecules.

5.6 Results

In this section, we carry out a systematic comparison of the performance of different measurement methods (Section 5.4) in CSHOREBench (Section 5.5). Previous comparisons have largely focused on analytical single shot variances against different molecules [Had+22; Hil+21] or small fixed measurement budgets [HKP21; Had21; Shl+23]. Moreover, measurement methods are often equipped with different estimators, making it less clear how much gain in performance one obtains by switching measurement methods.

We first pay special attention to the convergence of accuracy of these measurements. We then evaluate different measures of classical and quantum resource utilization as discussed in Section 5.5.2 for each measurement method. We use the experimental protocols as discussed in Section 5.5.3. In Section 5.6.1, we evaluate the RMSE of a specified measurement method when equipped with different estimators. This is to highlight the advantage one can gain in terms of the number of shots required to achieve a desired accuracy in estimation of $\mathrm{Tr}(\rho H)$ by simply changing the estimator. In Section 5.6.2, we report results from CSHOREBench on a simulator and in Section 5.6.3 on quantum devices for different molecular Hamiltonians and states 5.2. Finally, we comment on the utilization of quantum and classical resources with experiments on IBM quantum devices in Section 5.6.3.

5.6.1 Comparison of estimators

In Section 5.3, we noted that there are multiple different estimators — Monte Carlo (MC), weighted MC (WMC) and Bayesian — that could be used with the various measurement methods. We also noted that asymptotically with proper choice of parameters, all these estimators give the same performance in terms of the expected value of $\mathrm{Tr}(\rho H)$ for a given Hamiltonian H. This has motivated the use of different estimators combined with different measurement methods in previous comparison studies [Had+22; HKP21; Wu+23]. However, this becomes problematic when the measurement budget (or the total number of shots) considered is very low as has been the case in these studies and we are not in the asymptotic regime where the estimators are equivalent. The difference in performance of two combinations of measurements methods with estimators cannot be then properly attributed to the difference in estimators or the difference in measurement methods.

To highlight the advantage one can gain in terms of RMSE achieved in estimating $\mathrm{Tr}(\rho H)$ for a low measurement budget and how this apparent advantage disappears by increasing the number of shots, we consider the problem of estimating the value of $\mathrm{Tr}(\rho H)$ for a fixed measurement method and different estimators. As an illustration, we set the Hamiltonian H to be that of HeH^+ cation (6-31g basis, JW encoding, 8 qubits) and ρ as its ground state. In Figure 5.4(a), we plot the trends of RMSE achieved by the CS method with different

estimators for HeH^+. We observe that at a measurement budget of 10^3 shots (or samples), the Bayesian estimator has a lower RMSE than the MC estimator (although this difference can be removed by increasing the smoothing factor γ to 1), which in turn has a lower RMSE than the WMC estimator. These differences however disappear after 10^5 shots. Similarly, in Figure 5.4(b) for the LBCS measurement method, any advantage offered by one estimator disappears after around 10^4 shots. This is also observed in Figure 5.4(c) for the case of optimized decision diagrams (DD).

The implications of these results are twofold. Firstly, we need to be systematic in our choice of estimator when comparing different measurement methods for low measurement budgets so that we can properly attribute differences in performance to the measurement method at hand. Secondly, for low measurement budgets, a Bayesian estimator or MC estimator with Laplace smoothing is preferred. In the rest of the chapter, we will either fix the estimator across all the measurement methods or state when different estimators are chosen for different measurement methods. We can afford to do the latter as there is negligible difference between any of the estimators for a given measurement method at the high measurement budgets ($> 10^6$ shots) that we consider.

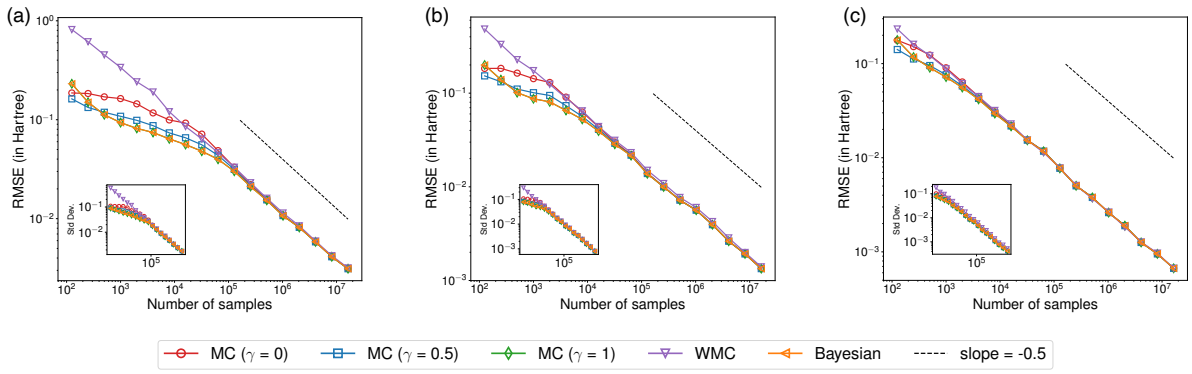


Figure 5.4: Comparison of RMSE achieved by different estimators in numerical simulations on HeH^+ cation (8 qubits, 6-31g basis, JW encoding) with measurement methods of (a) classical shadows (CS), (b) locally biased classical shadows (LBCS), and decision diagrams (DD). The state is set to be the ground state of the Hamiltonian. A common legend is shown for the subfigures. Trends of RMSE achieved by Monte Carlo (MC) with Laplace smoothing of $\gamma \in \{0, 0.5, 1\}$, weighted Monte Carlo (WMC) and Bayesian estimators on the same sets of measurements collected for each measurement method are shown. Averaged values over 200 independent runs are shown. Inset shows the standard deviation or uncertainty on these expectation values for each estimator.

5.6.2 Experiments on classical simulator

We now turn our attention to comparing the performance of different measurement methods in estimating $\mathrm{Tr}(\rho H)$ on a classical simulator for molecular Hamiltonians (those given in Table 5.2) and their ground states ρ. In all the following experiments, the estimator will be set to be the Bayesian estimator with the exception of N_2 where it is set to be the MC estimator. The latter choice is made to reduce the classical post-processing runtime for N_2.

Figures 5.5 and 5.6 show convergence behaviour of RMSE achieved by different measurement methods against measurement budget up to around 1.6×10^7 samples. In Table 5.4, we summarize resource utilization by different measurement methods against different molecular Hamiltonians considering metrics discussed in Section 5.5. We now describe our observations for the set of molecular Hamiltonians in Table 5.2.

Tapered Hamiltonians. We show convergence results for the tapered Hamiltonian H_2 (3-21g, JW, 5 qubits) in Figure 5.5(a) and tapered Hamiltonian HeH^+ (3-21g, JW, 6 qubits) in Figure 5.5(b). We observe that CS performs very similar to LBCS for these smaller sized Hamiltonians. Derandomization methods also perform very similarly to decision diagrams with Derand. LBCS narrowly improving on H_2 and decision diagrams on HeH^+.

8 to 16 qubit Hamiltonians. We show results for HeH^+ (6-31g, JW, 8 qubits) in Figure 5.6(a), LiH (sto6g, JW, 12 qubits) in Figure 5.6(b), and N_2 in Figure 5.6(c). Across all these three molecular Hamiltonians, we find that optimized decision diagrams (Dec. Diag.) and derandomization (Derand.) are best, achieving the same RMSE compared to CS or LBCS with fewer shots. In particular for HeH^+, decision diagrams are able to achieve chemical accuracy of 1.6 milli-Hartree, and require around 22% fewer shots than LBCS. By using LBCS itself, we obtain a constant query reduction (as observed after 2×10^4 shots) of 40% compared to CS.

Similarly, for LiH we notice a large gap in performance in between LBCS and CS versus decision diagrams and any of the derandomized methods. Overall, in this case, Derand. CS is the preferred measurement method, shaving off nearly two orders of magnitude of shots required to reach high accuracy compared to CS, although it is extremely close to the other derandomized methods and decision diagrams. On N_2, we do not benchmark CS given its poor performance on previous molecules and the high number of unique measurement circuits needed to be evaluated on the simulator. On N_2, we observe that Derand. LBCS is the most accurate method followed by decision diagrams.

Resource utilization. In Table 5.4, we report the resource utilization for estimating $\mathrm{Tr}(\rho H)$ within an accuracy cutoff of 5 milli Hartree. While we have so far commented on accuracy, let us consider the other metrics. We notice that across the board, Derand. DD requires the least number of unique circuits to be executed. This has implications for reduced time for quantum compilation of measurement circuits on quantum hardware. Regarding classical pre-processing runtime, generating samples from product distributions is fast which makes CS and LBCS attractive in this respect. For decision diagrams (DD), the largest contribution to the classical pre-processing runtime is the runtime taken to construct and optimize the decision diagrams with time taken to generate samples for 5 milli Hartree accuracy being 10s of seconds up to LiH and taking roughly 490s for N_2. The runtimes for constructing DDs are reported in Appendix C.2.2. For Derand DD, the largest contribution to classical pre-processing runtime is the derandomization process itself. In terms of total runtime on the simulator, decision diagrams and derandomization start becoming preferred over CS and LBCS as we move to the larger molecules of LiH and N_2.

In Table 5.5, we report estimated resource utilization for estimating $\mathrm{Tr}(\rho H)$ within an

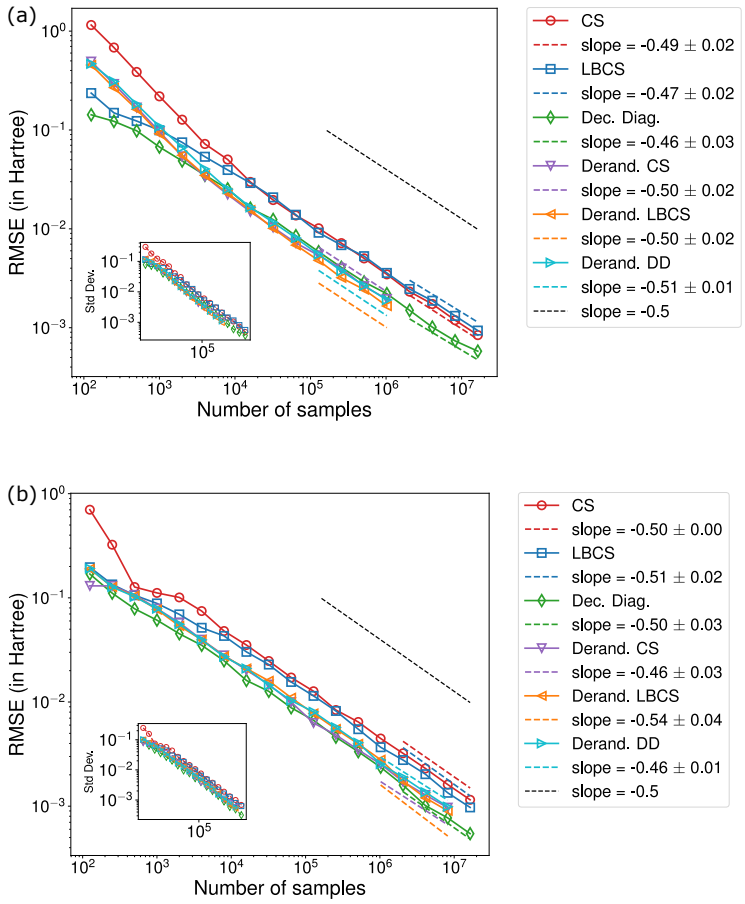


Figure 5.5: Comparison of RMSE achieved in numerical simulations by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as the ground state and H is the Hamiltonian of (a) tapered H_2 (5 qubits, 3-21g basis, JW encoding), and (b) tapered HeH^+ (6 qubits, 3-21g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator.

accuracy cutoff of 5 milli Hartree. The quantum runtimes are predicted assuming that most of the quantum wall-clock time is due to delay between executions of circuits which is around 500μs on the quantum devices which we ran our experiments on. This is an example of how CSHOREBench could be used as a tool for selecting measurement methods before running any experiment on the quantum device.

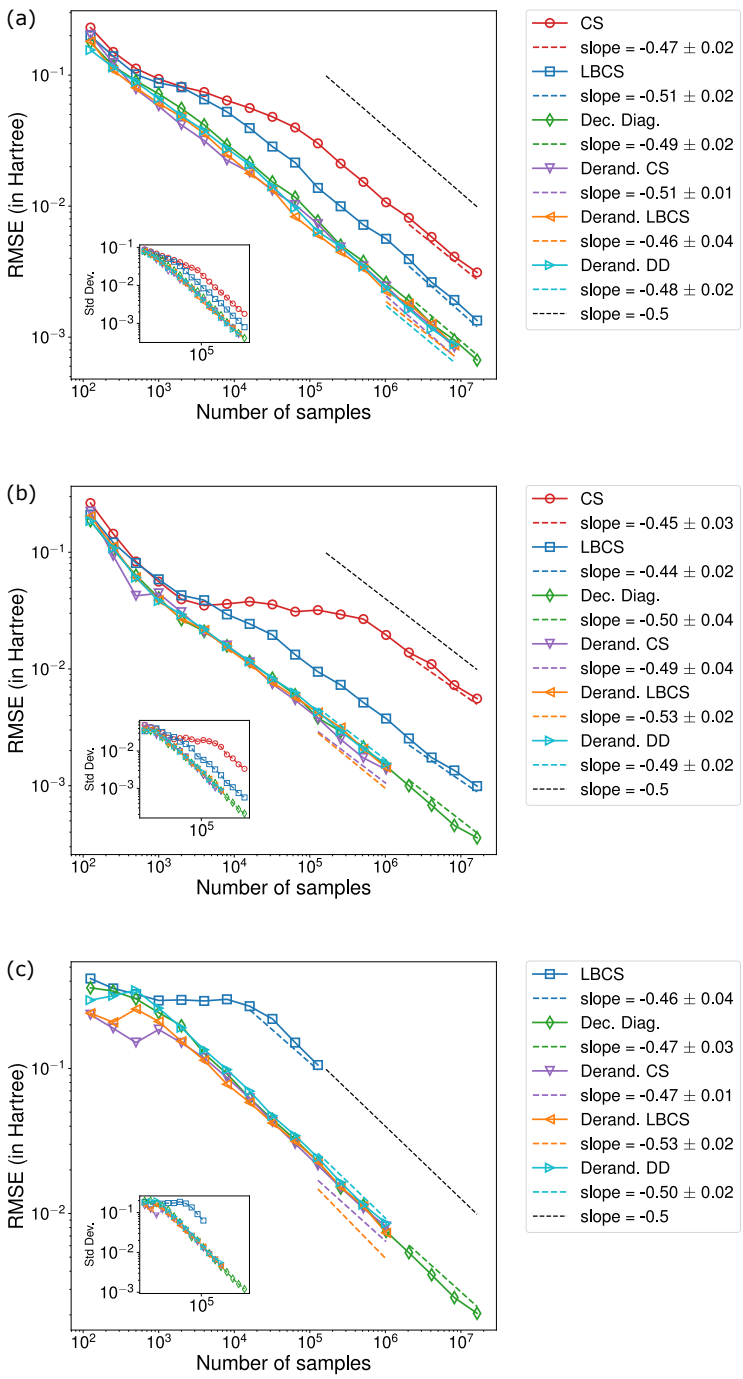


Figure 5.6: Comparison of RMSE achieved in numerical simulations by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as the ground state and H is the Hamiltonian of (a) HeH^+ (8 qubits, 6-31g basis, JW encoding), (b) LiH (12 qubits, sto-6g basis, JW encoding), and (c) N_2 (16 qubits, sto-6g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator in (a)-(b) and the MC estimator in (c).

Measurement Method	# shots required	# unique circuits	Median # of shots per circuit (all, top 5%, bottom 5%)	Classical pre-process. runtime [s]	Classical post-process. runtime [s]	Classical simulator runtime [s]	Total classical runtime [s]
H_2, 5 qubits (3-21g, JW)							
CS	4.75×10^5	243	(1954, 1954, 1954)	1.21×10^2	4.94	4.27	1.30×10^2
LBCS	5.11×10^5	243	(944, 12350, 94)	9.59×10^1	7.13	3.73	1.07×10^2
Dec. Diag.	1.79×10^5	44	(1335, 15122, 36)	$\mathbf{4.39 \times 10^1}$	1.59	0.91	$\mathbf{4.64 \times 10^1}$
Derand. CS	1.46×10^5	49	(4921, 5628, 1)	1.72×10^3	**1.52**	0.88	1.72×10^3
Derand. LBCS	$\mathbf{1.16 \times 10^5}$	55	(18, 4462, 1)	2.31×10^3	1.62	0.92	2.31×10^3
Derand. DD	1.53×10^5	**32**	(5453, 5461, 1)	1.86×10^3	1.67	**0.71**	1.86×10^3
HeH^+, 6 qubits (6-31g, JW)							
CS	8.25×10^5	729	(1132, 1132, 1132)	1.52×10^2	30.63	13.50	1.96×10^2
LBCS	6.35×10^5	729	(372, 4078, 37)	$\mathbf{1.16 \times 10^2}$	18.19	11.40	$\mathbf{1.46 \times 10^2}$
Dec. Diag.	$\mathbf{2.13 \times 10^5}$	234	(253, 5718, 7)	5.63×10^2	**5.60**	3.41	5.72×10^2
Derand. CS	2.37×10^5	188	(665, 2790, 1)	5.24×10^3	8.80	2.59	5.25×10^3
Derand. LBCS	3.11×10^5	184	(950, 3616, 1)	7.13×10^3	8.91	2.64	7.14×10^3
Derand. DD	3.06×10^5	**123**	(3589,3610, 1)	6.08×10^3	8.04	**2.10**	6.09×10^3
HeH^+, 8 qubits (6-31g, JW)							
CS	5.58×10^6	6561	(850, 850, 850)	1.14×10^3	575.43	2.06×10^2	1.92×10^3
LBCS	9.96×10^5	6561	(81, 1001, 10)	$\mathbf{2.47 \times 10^2}$	105.92	1.44×10^2	$\mathbf{4.97 \times 10^2}$
Dec. Diag.	2.92×10^5	781	(99, 2907, 3)	3.57×10^3	25.33	1.96×10^1	3.61×10^3
Derand. CS	2.56×10^5	545	(43, 2345, 1)	9.16×10^3	26.40	1.30×10^1	9.20×10^3
Derand. LBCS	$\mathbf{2.05 \times 10^5}$	517	(24, 1869, 1)	7.52×10^3	21.80	1.18×10^1	7.55×10^3
Derand. DD	2.34×10^5	**209**	(1407, 1960, 1)	9.94×10^3	**20.30**	**7.13**	9.97×10^3
LiH, 12 qubits (sto6g, JW)							
CS	2.32×10^7	531441	(50, 58, 43)	8.81×10^3	2910	2.24×10^5	2.36×10^5
LBCS	5.62×10^5	128002	(2, 27, 1)	$\mathbf{2.15 \times 10^2}$	204	1.71×10^4	1.75×10^4
Dec. Diag.	8.82×10^4	802	(32, 693, 2)	2.56×10^4	6.99	3.22×10^2	2.59×10^4
Derand. CS	$\mathbf{7.13 \times 10^4}$	1814	(4, 433, 1)	7.13×10^3	4.89	1.09×10^2	$\mathbf{7.25 \times 10^3}$
Derand. LBCS	8.78×10^4	976	(11, 514, 1)	8.76×10^3	4.55	7.86×10^1	8.84×10^3
Derand. DD	9.32×10^4	**243**	(462, 468, 1)	3.17×10^4	**4.51**	$\mathbf{6.45 \times 10^1}$	3.18×10^4
N_2, 16 qubits (sto6g, JW)							
Dec. Diag.	2.36×10^6	1134	(536, 11084, 9)	$\mathbf{1.39 \times 10^5}$	**123**	1.81×10^4	$\mathbf{1.57 \times 10^5}$
Derand. CS	3.88×10^6	10015	(778, 35225, 1)	7.18×10^5	310	7.84×10^4	7.97×10^5
Derand. LBCS	$\mathbf{2.07 \times 10^6}$	5628	(73, 10274, 1)	3.91×10^5	173	2.83×10^4	4.20×10^5
Derand. DD	3.48×10^6	**488**	(8714, 9754, 2)	5.31×10^5	140	$\mathbf{5.95 \times 10^3}$	5.37×10^5

Table 5.4: Comparison of resource utilization with experiments on a classical simulator by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ to achieve an accuracy of 5×10^{-3} Hartree, with H set to be different different molecular Hamiltonians (Table 5.2) and ρ as the ground state. Values shown are averages over 192 independent runs against different metrics. Metrics shown are described in Section 5.5.2. Note that the contribution to classical post-processing time is from the estimator which is set to be the Bayesian estimator for all molecules except for N_2 where it is set to be the MC estimator. Lowest values obtained for each metric against a Hamiltonian is boldfaced.

Measurement Method	Predicted quantum runtime [s]	log(R) A	B	C	D
H_2, 5 qubits (3-21g, JW)					
CS	2.4×10^2	5.9	10.5	15.4	20.2
LBCS	2.6×10^2	5.9	10.6	15.4	20.3
Dec. Diag.	9.0×10^1	**4.9**	9.5	14.4	19.2
Derand. CS	7.3×10^1	7.5	9.4	14.2	19.0
Derand. LBCS	$\mathbf{5.8 \times 10^1}$	7.8	**9.3**	**14.0**	**18.8**
Derand. DD	7.6×10^1	7.6	9.5	14.2	19.1
HeH^+, 6 qubits (6-31g, JW)					
CS	4.1×10^2	6.4	11.0	15.9	20.8
LBCS	3.2×10^2	**6.1**	10.8	15.7	20.5
Dec. Diag.	$\mathbf{1.1 \times 10^2}$	6.5	**9.7**	**14.6**	**19.4**
Derand. CS	1.2×10^2	8.6	10.0	14.7	19.5
Derand. LBCS	1.6×10^2	8.9	10.3	15.0	19.8
Derand. DD	1.5×10^2	8.7	10.3	14.9	19.8
HeH^+, 8 qubits (6-31g, JW)					
CS	2.8×10^3	8.3	12.9	17.8	22.7
LBCS	5.0×10^2	**6.6**	11.2	16.1	20.9
Dec. Diag.	1.5×10^2	8.2	10.1	14.9	19.7
Derand. CS	1.3×10^2	9.1	10.3	14.8	19.6
Derand. LBCS	$\mathbf{1.0 \times 10^2}$	8.9	**10.0**	**14.5**	**19.4**
Derand. DD	1.2×10^2	9.2	10.2	14.7	19.5
LiH, 12 qubits (sto6g, JW)					
CS	1.2×10^4	10.1	14.4	19.3	24.1
LBCS	2.8×10^2	**6.6**	10.7	15.5	20.4
Dec. Diag.	4.4×10^1	10.2	10.4	13.7	18.5
Derand. CS	$\mathbf{3.6 \times 10^1}$	8.9	**9.4**	**13.5**	**18.3**
Derand. LBCS	4.4×10^1	9.1	9.6	13.7	18.5
Derand. DD	4.7×10^1	10.4	10.6	13.8	18.6
N_2, 16 qubits (sto6g, JW)					
Dec. Diag.	1.2×10^3	**11.9**	**12.7**	17.0	21.8
Derand. CS	1.9×10^3	13.5	13.8	17.5	22.3
Derand. LBCS	1.0×10^3	12.9	13.2	**16.9**	**21.7**
Derand. DD	1.7×10^3	13.2	13.6	17.4	22.2

Table 5.5: Comparison of predicted resource utilization on IBM Quantum devices by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ to achieve an accuracy of 5×10^{-3} Hartree, with H set to be different molecular Hamiltonians (Table 5.2) and ρ is the ground state. Predicted quantum runtime or the wall-clock that would be spent by the quantum device in executing experiments is shown for the different measurement methods based on the number of shots required to achieve 5×10^{-3} Hartree (Table 5.4) and assuming the quantum runtime would primarily be from delay between circuit executions. Estimates of resource utilization, indicated by R, here account for classical and predicted quantum computing resources via the heuristic in Eq. 5.23 and using the regimes of Table 5.3. Lowest values obtained for each metric against a Hamiltonian is boldfaced.

5.6.3 Experiments on quantum device

As in the case of our experiments on the classical simulator, we consider the molecular Hamiltonians listed in Table 5.2. We set the quantum state ρ to be a random ansatz as discussed in Section 5.5.3. This avoids the significant overhead of conducting a large VQE experiment to simulate the ground state of the Hamiltonian, and instead is representative of a typical state that might appear midway through a VQE optimization.

We carry out a comparison of all the candidate measurement methods for the smaller tapered Hamiltonians (5.2), but consider a subset of the measurement methods for the larger sized molecules. As the number of qubits n of the Hamiltonian H increases, the number of unique measurement bases queried by uniform CS or LBCS can become as high as our measurement budget which become too expensive to run in experiments on IBM quantum devices given current classical latencies (Section 5.5.2) and experimental constraints (Section 5.5.3). We thus limit the number of circuits by considering the methods of decision diagrams and derandomization on the larger sized molecules.

Tapered Hamiltonians. We show convergence results for the tapered Hamiltonian H_2 (3-21g, JW, 5 qubits) in Figure 5.7(a) and tapered Hamiltonian HeH^+ (3-21g, JW, 6 qubits) in Figure 5.7(b). In these experimental results, we find larger differences in performance of the measurement methods than we observed earlier in the simulated results in Figure 5.5. The notable change from our observations on the simulator is that derandomization methods perform poorly compared to decision diagrams. This could be due to effects of measurement noise and choosing hyperparameters for derandomization via tuning on the simulator instead of the quantum device. Overall in the case of tapered H_2, we can reduce the number of samples required to achieve an accuracy of 5 milli Hartree by 57% by switching from Derand. CS to DD, by 57% as well by switching from LBCS to DD, and by around 85% by opting for DD instead of CS. In the case of tapered HeH^+, the resource reduction in the number of samples achieved by DD is more than 83% compared to CS, 57% compared to LBCS, and more than 30% compared to any derandomization method.

8 to 16 qubit Hamiltonians. We show results for HeH^+ (6-31g, JW, 8 qubits) in Figure 5.8(a), LiH (sto6g, JW, 12 qubits) in Figure 5.8(b), and N_2 in Figure 5.8(c). Note that as mentioned earlier, we benchmark the more frugal methods against these molecular Hamiltonians to keep the number of quantum compilations required during experiments reasonable and latencies incurred from circuit loading low. Across the three molecular Hamiltonians, we again find that optimized decision diagrams (Dec. Diag.) outperform derandomization (Derand.) methods. In particular for LiH, decision diagrams are able to achieve an accuracy of 5 milli-Hartree with around 40% fewer shots than Derand. LBCS. On N_2, the difference between DD and derandomization reduces from before with DD achieving a resource reduction of only 25% at 5 milli Hartree. Overall, DD consistently requires fewer measurements compared to derandomization methods.

Resource utilization. In Table 5.6, we report the resource utilization for estimating $\mathrm{Tr}(\rho H)$ within an accuracy cutoff of 5 milli Hartree. We notice that across the board, Derand. DD requires the least number of unique circuits to be executed, as expected from

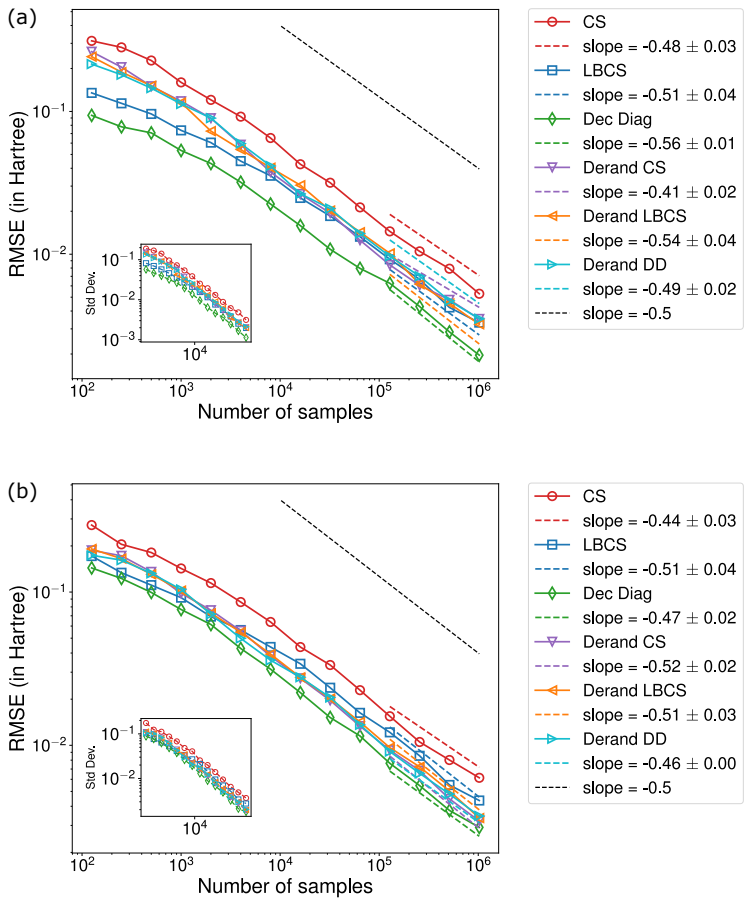


Figure 5.7: Comparison of empirical RMSE achieved in experiments on quantum device by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as a randomly fixed ansatz and H is the Hamiltonian of (a) tapered H_2 (5 qubits, 3-21g basis, JW encoding), and (b) tapered HeH^+ (6 qubits, 3-21g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator.

our results on the simulator. Taking into account both quantum and classical computational resource utilization, decision diagrams are preferred over derandomization. It requires fewer shots to reach the same accuracy in experiments on the device and less classical computational runtime in generating samples. This is also supported by the weighted resource utilization reported in Table 5.6 based on the heuristic and different regimes introduced in Section 5.5.2.

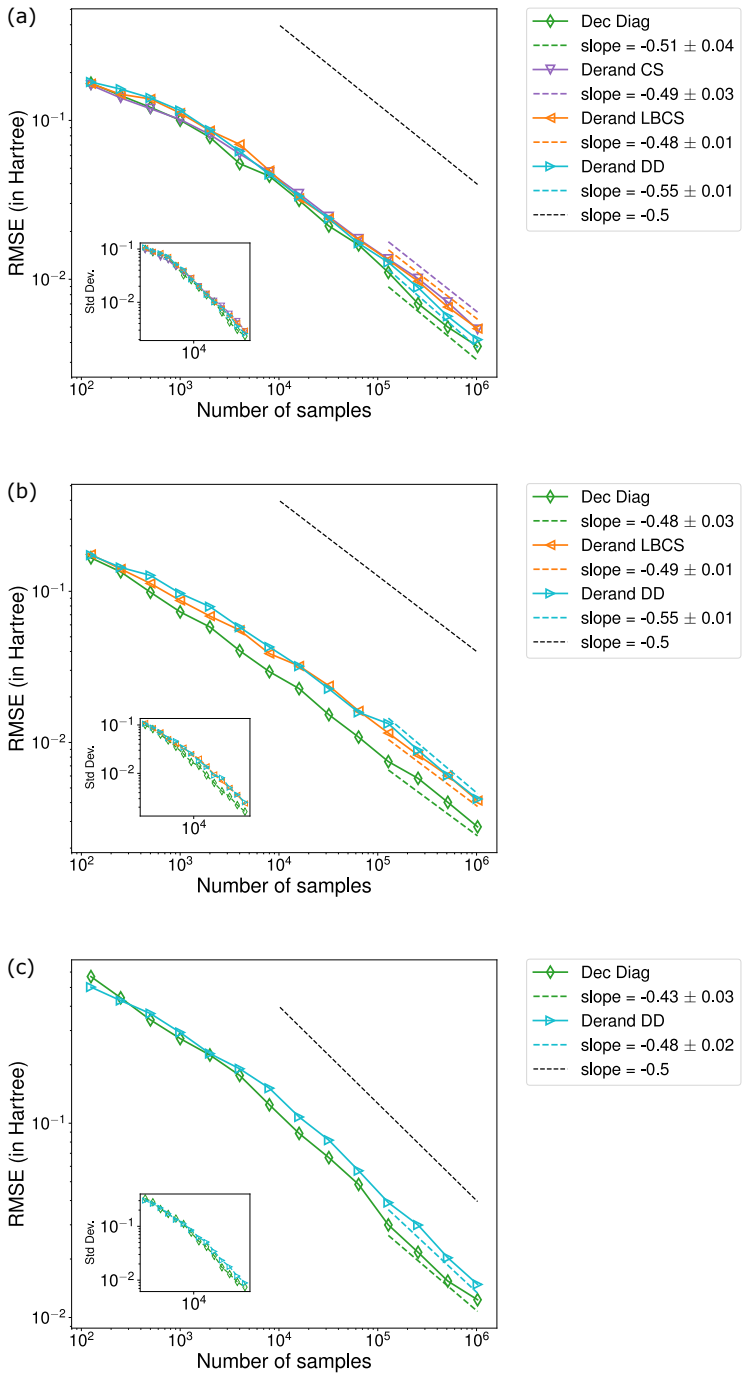


Figure 5.8: Comparison of empirical RMSE achieved in experiments on a quantum device by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ with ρ set as a randomly fixed ansatz and H is the Hamiltonian of (a) HeH^+ (8 qubits, 6-31g basis, JW encoding), (b) LiH (12 qubits, sto-6g basis, JW encoding), and (c) N_2 (16 qubits, sto-6g basis, JW encoding). RMSE is shown with the number of samples (or shots) made. The estimator for each measurement method is set to be the Bayesian estimator.

Measurement Method	# shots required	# unique circuits	Median # shots per circ. (all, top 5%, bottom 5%)	Classical runtime [s] pre-process.	post-process.	latencies	Quantum runtime [s]	log(R) A	B	C	D
H_2, 5 qubits (3-21g, JW)											
CS	1.2×10^6	243	(4917, 5056, 4788)	2.3×10^2	6.5	8.7×10^4	6.1×10^2	6.7	11.4	16.3	21.1
LBCS	4.2×10^5	243	(769, 10081, 76)	8.0×10^1	3.3	3.0×10^4	2.1×10^2	5.7	10.4	15.3	20.1
Dec. Diag.	$\mathbf{1.8 \times 10^5}$	43	(1371, 15526, 37)	$\mathbf{4.3 \times 10^1}$	**1.1**	$\mathbf{1.3 \times 10^4}$	$\mathbf{9.3 \times 10^1}$	**4.9**	**9.5**	**14.4**	**19.3**
Derand. CS	4.3×10^5	49	(15740, 16442, 1)	9.7×10^3	10.1	3.1×10^4	2.2×10^2	9.2	10.6	15.3	20.1
Derand. LBCS	4.1×10^5	55	(28, 15887, 1)	9.9×10^3	9.3	3.0×10^4	2.1×10^2	9.2	10.6	15.3	20.1
Derand. DD	4.9×10^5	**32**	(17429, 17441, 1)	8.5×10^3	8.7	3.5×10^4	2.5×10^2	9.1	10.7	15.4	20.2
HeH^+, 6 qubits (6-31g, JW)											
CS	1.8×10^6	729	(2525, 2731, 2347)	3.5×10^2	6.4	1.4×10^5	9.3×10^2	7.2	11.8	16.7	21.6
LBCS	7.1×10^5	728	(418, 4566, 42)	$\mathbf{1.4 \times 10^2}$	3.9	5.3×10^4	3.6×10^2	**6.2**	10.9	15.8	20.6
Dec. Diag.	$\mathbf{3.0 \times 10^5}$	234	(355, 8071, 10)	5.7×10^2	**1.4**	$\mathbf{2.2 \times 10^4}$	$\mathbf{1.5 \times 10^2}$	6.6	**10.0**	**14.9**	**19.7**
Derand. CS	4.3×10^5	189	(1009, 5071, 1)	9.8×10^3	10.1	3.2×10^4	2.2×10^2	9.2	10.7	15.3	20.1
Derand. LBCS	5.0×10^5	185	(1796, 5824, 1)	1.2×10^4	9.8	3.8×10^4	2.6×10^2	9.4	10.8	15.5	20.3
Derand. DD	4.4×10^5	**124**	(5157, 5179, 1)	8.5×10^3	8.7	3.3×10^4	2.2×10^2	9.1	10.6	15.3	20.1
HeH^+, 8 qubits (6-31g, JW)											
Dec. Diag.	$\mathbf{5.5 \times 10^5}$	787	(181, 5392, 4)	$\mathbf{3.6 \times 10^3}$	**5.3**	$\mathbf{4.1 \times 10^4}$	$\mathbf{2.8 \times 10^2}$	**8.3**	**10.7**	**15.5**	**20.4**
Derand. CS	1.0×10^6	548	(59, 9254, 2)	3.5×10^4	41.5	7.5×10^4	5.1×10^2	10.5	11.6	16.1	21.0
Derand. LBCS	1.0×10^6	518	(25, 9125, 2)	3.5×10^4	34.1	7.5×10^4	5.1×10^2	10.5	11.6	16.1	21.0
Derand. DD	7.2×10^5	**210**	(3971, 6042, 1)	2.3×10^4	28.7	5.4×10^4	3.7×10^2	10.1	11.3	15.8	20.6
LiH, 12 qubits (sto6g, JW)											
Dec. Diag.	$\mathbf{3.2 \times 10^5}$	811	(112, 2504, 7)	$\mathbf{2.6 \times 10^4}$	**12.8**	$\mathbf{2.4 \times 10^4}$	$\mathbf{1.6 \times 10^2}$	**10.2**	**10.8**	**15.0**	**19.8**
Derand. LBCS	7.0×10^5	979	(23, 4090, 2)	6.9×10^4	13.0	5.2×10^4	3.5×10^2	11.1	11.7	15.8	20.6
Derand. DD	7.5×10^5	**284**	(3715, 3727, 1)	7.4×10^4	13.0	5.6×10^4	3.8×10^2	11.2	11.8	15.9	20.7
N_2, 16 qubits (sto6g, JW)											
Dec. Diag.	$\mathbf{8.9 \times 10^6}$	1137	(15298, 40070, 480)	$\mathbf{1.4 \times 10^5}$	516	$\mathbf{6.5 \times 10^5}$	$\mathbf{4.5 \times 10^3}$	**11.9**	**13.6**	**18.3**	**23.1**
Derand. DD	1.2×10^7	**488**	(33477, 94988, 2)	1.6×10^6	369	8.8×10^5	6.2×10^3	14.3	14.7	18.6	23.5

Table 5.6: Comparison of resource utilization with experiments on IBM Quantum devices by different measurement methods in estimating $\mathrm{Tr}(\rho H)$ to achieve an accuracy of 5×10^{-3} Hartree, with H set to be different molecular Hamiltonians (Table 5.2) and ρ is a random ansatz. Values shown are averages over 192 bootstrapped runs against different metrics. Metrics shown are described in Section 5.5.2. Number of unique circuits run contribute to classical latencies such as compilation. Number of shots per circuit (all, top 5%, bottom 5%) summarizes shot distribution across circuits and is an indication of latencies due to circuit loading on control electronics. Classical pre-processing runtime is the wall-clock time spent by measurement methods in constructing a query distribution and generating samples. Classical post-processing runtime is the wall-clock time spent by the Bayesian estimator in computing an estimate from acquired measurement outcomes. Quantum runtime is the wall-clock spent by the quantum device in executing experiments. Estimates of resource utilization, indicated by R, account for both classical and quantum computing resources via the heuristic in Eq. 5.23 and using the regimes of Table 5.3. Lowest values obtained for each metric against a Hamiltonian is boldfaced.

5.7 Conclusion

In this chapter, we propose CSHOREBench(Common States and Hamiltonians for ObseRvable Estimation Benchmark) for benchmarking measurement methods designed to solve the problem of estimating the expectation value $\mathrm{Tr}(\rho H)$ of a quantum Hamiltonian H with respect to a quantum state ρ. We presented the various evaluation criteria and explained the importance of considering utilization of both quantum and classical resources for these hybrid measurement methods in addition to the performance metric of accuracy achieved in estimating $\mathrm{Tr}(\rho H)$. In practice, CSHOREBench is empirically efficient to run on states of

random quantum ansatz with fixed depth. This makes it an attractive tool to select measurement methods for a given Hamiltonian before experiments are executed on a quantum device. In numerical simulations on molecular Hamiltonians, we found that compact decision diagrams (DD) along with derandomization are the most competitive methods. In our experiments on IBM quantum devices, we observed that decision diagrams achieved a given accuracy with fewer quantum measurements than derandomization. However, derandomization of decision diagrams may be preferred if a low number of unique circuits is the most important consideration.

The improvement in query complexity by using compact decision diagrams or derandomized decision diagrams for estimating $\mathrm{Tr}(\rho H)$ has multiple practical implications besides accelerating steps of variational quantum eigensolvers (VQE). For example, in the hybrid quantum-classical algorithm of Krylov subspace methods [PM19; Mot+20], the measurement problem appears when estimating matrix elements for the generalized eigenvalue problem and DDs may be advantageous in that setting as well, albeit optimized differently.

There is still room for improvement in the design of measurement methods themselves. Randomized measurement methods could be improved by incorporating adaptive strategies during learning by building information about the quantum state ρ under consideration in real-time in addition or instead of prior information. One tool to introduce adaptivity would be active learning which has been shown to be useful in practice for the learning tasks of quantum state tomography [Nun+10] and Hamiltonian learning [Dut+23a]. The query complexity of randomized measurement methods could also be improved by allowing low-depth and locally-entangling measurements [Ipp23]. It is known that the query complexity is significantly improved [HKP20] when global Clifford measurements are allowed but these require deeper circuits than one would typically want to expend on the measurement part of a near-term quantum algorithm.

Aside from improving the measurement methods themselves, there is still potential for improving the benchmarking introduced here. We have proposed a heuristic comprising different regimes to summarize utilization of quantum and classical resources. Recommendations of heuristics for other platforms would aid in providing further guidance on which measurement methods to prefer for these platforms. Moreover, the list of measurement methods being benchmarked here could be extended to include grouping methods and other approaches [GK23]. As for the instances considered as examples of common computations, future benchmarks would benefit from extending beyond minimal bases. We used minimal bases in this work in order to reduce the computational cost on the quantum side as far as possible, but these are not physically representative, so using correlation consistent bases would be preferable as we advance toward practically relevant quantum algorithms for chemistry.

Chapter 6

Quantum Bootstrap Embedding: A distributed quantum algorithm for ground state energy estimation

6.1 Introduction

Determining the ground state of large-scale interacting fermionic systems is an important challenge in quantum chemistry, materials science, and condensed matter physics. Just as electronic properties of molecules underpin their chemical reactivity [FYS52; PY84; GNM02], phase diagrams of solid state materials are also determined to a large degree by their ground state electronic structure [LeB+15; ZW15; Kot+06]. However, exact solution to the time-independent Schrodinger equation of a practical many-electron system remains a daunting task because the dimension of the underlying Hilbert space grows exponentially with the number of orbitals, and the computational resources required to perform calculations over such a large space can quickly exceed the capacity of current classical or quantum hardware.

One promising approach to fit a large electronic structure problem into a limited amount of computational resources is to break the original system into smaller fragments, where each fragment can be solved individually from which a solution to the whole is then obtained [Gor+12; Jon+20; SC16]. Efforts along this direction have successfully led to various embedding schemes that significantly expand the complexity of the systems solvable using classical computational resources, such as density-based embedding theories [WSZ15; LHC14], density-matrix embedding theories (DMET) [KC12; KC13; Wou+16; WAK17; Fau+22], various Green's function embedding theories [Het+00; Ma+21; LZ17; Rus+18; BAG03; Kot+06] and the bootstrap embedding theory [WTV16; Ye+19; YTV21]. The essence of such embedding-based methods is to add an additional external potential to each fragment Hamiltonian and then iteratively update the potential until some conditions on certain observables of the system are matched. Nevertheless, due to the significant cost in solving the fragment Hamiltonian itself as the fragment size increases, the applicability of such methods are limited to relatively small fragments, which may lead to incorrect predictions in systems with long-range correlations [Zhe+17]. While approximate fragment solvers such as the coupled-cluster theory or many-body perturbation theory have greatly enhanced

the applicability of such embedding methods at a reduced cost [Zhu+19; SZ19; LKB21], these approximations tend to fail for strongly correlated systems due to limited treatment of electron correlation. In addition, because of limitations on computing k-electron reduced density matrices (k-RDMs for $k > 2$), embedding and observable calculations beyond 2-RDM are difficult in general.

Quantum computers are believed to be promising in tackling electronic structure problems more efficiently [Bau+20], despite evidence for an exponential advantage across chemical space has yet to be found [Lee+22]. One natural idea to circumvent the problems of classical eigensolvers is to use a quantum computer to treat the fragments. By mapping each orbital to a constant (small) number of qubits, the exponentially large (in the number of orbitals) Hilbert space of an interacting fermionic system can be encoded in only a polynomial number of qubits and terms. Indeed, quantum eigensolvers such as the quantum phase estimation (QPE) [AL99b] algorithm has been proposed to achieve an exponential advantage given a properly prepared input state [Asp+05] with non-exponentially small overlap with the exact ground state. More recently, various variants of the variational quantum eigensolver (VQE) [Per+14; Til+22; WHB19; Gri+19; Gri+22] have been demonstrated experimentally on NISQ devices to achieve impressive performance as compared to classical methods. Moreover, k-RDMs (for any k) can be measured through quantum eigensolvers [Cra+10; ZRM21] that may circumvent the difficulty encountered on classical computers. Current quantum computers are small, but quantum embedding provides a way to tailor fragment sizes to fit into such small quantum machines to achieve a solution to the entire problem.

To take the full advantage of these quantum eigensolvers within the embedding framework [Ott+22; Vor+22; MM22; Li+22; MGG20; Ma+21], two open questions immediately arise as a result of the intrinsic nature of quantum systems. Firstly, the wave function of a quantum system collapses when measured. This means any measurement of the fragment wave function is but a statistical sample (akin to Monte Carlo methods), and many measurements are needed to obtain statistical averages with sufficiently low uncertainty in order to achieve a good matching condition for the embedding. Secondly, the best way to perform matching between fragments using results from quantum eigensolvers is not clear, and most likely a new approach needs to be formulated to match fragments. Admittedly, it would be straightforward to first estimate the density matrices by collecting a number of quantum samples and then use the estimated density matrices to minimize the cost function as in classical embedding theories [WTV16; KC12]. But this approach would be very costly especially given the increasing number of elements in qubit reduced density matrices (RDMs) that need to be estimated [Qua+20]. Could there be a quantum method for matching, as opposed to a statistical sampling-based classical approach?

We address the two challenges by providing a quantum coherent matching algorithm and an adaptive sampling schedule, leading to a quantum bootstrap embedding (QBE) method based on classical bootstrap embedding [WTV16]. Instead of matching the RDM element-by-element, the quantum matching algorithm employs a `SWAP` test [Bar+97; Buh+01] to match the full RDM between overlapping regions of the fragments in parallel. Moreover, the quantum amplitude estimation algorithm [Bra+02; Mar+21b] allows an extra quadratic speedup to reach a target accuracy on estimating the fragment overlap. In addition, the adaptive sampling changes the number of samples as the optimization proceeds in order to achieve an increasingly better matching conditions.

It is important to compare the cost of quantum algorithms to classical algorithms carefully to claim any quantum advantage [Lee+22]. Toward this end, to highlight the advantage of the new QBE algorithms, we systematically compare their cost against classical BE algorithms with a biased stochastic eigensolver (the variational Monte Carlo, VMC) and the exact solver (full configuration interaction, FCI) as baselines. Different versions of the QBE algorithms using either QPE or VQE as eigensolvers with classical or quantum (coherent) matching algorithms are also compared among themselves for clarity.

The present work invites a viewpoint of treating quantum computers as *coherent sampling* machines which have three major advantages, as compared to their classical counterparts. First, the exponentially large Hilbert space provided by a quantum computer allows more efficient exact ground state solver (QPE) than their classical counterpart (exact diagonalization). Second, in the case of truncation for seeking approximate solutions, the abundant Hilbert space of quantum computers enable more flexible and expressive variational ansatz than classical computers, leading to more accurate solutions. Third, the coherent nature of quantum computers allows sampling to be performed at a later stage, e.g. after quantum amplitude amplification of matching conditions to extract just the feedback desired, instead of having to read out full state of a system.

The rest of the chapter is organized as follows. Sec. 6.2 overviews bootstrap embedding method at a high level and analyzes its scaling on classical computers, in order to motivate the need for bootstrap embedding on quantum computers. This section serves to set the notation and baseline of comparison for the rest of the chapter. Sec. 6.3 presents the theoretical framework of quantum bootstrap embedding in detail as constraint optimization problems. In Sec. 6.4, we give details of the QBE algorithm to solve the optimization problem. In Sec. 6.5, we apply our methods to hydrogen chains under minimal basis where both classical and quantum simulation results are shown to demonstrate the convergence and sampling advantage of our QBE method. We conclude the chapter in Sec. 6.6 with a summary of comparisons between classical and quantum BE discussed in the chapter, as well as prospects and future directions.

6.2 Ideas of Bootstrap Embedding

The idea of Bootstrap Embedding (BE) for quantum chemistry has recently led to a promising path to tackle large-scale electronic structure problems [WTV16; Ye+19; YTV20]. In this section, we establish the terminology and framework that will be used in the rest of the chapter. We first briefly review BE and outline the main framework of BE for computation on a classical computer in Sec. 6.2.1 and 6.2.2 for non-chemistry readers, to set up the notation. We then begin presenting new material by discussing typical behavior and computational resource requirements for BE on classical computers in Sec. 6.2.3, which leads to the quest for performing BE on a quantum computer in Sec. 6.2.4.

6.2.1 Fragmentation and Embedding Hamiltonians

To provide a foundation for a more concrete exposition of the bootstrap embedding method, we first establish some rigorous notation for discussing molecular Hamiltonians and their

associated Hilbert spaces. We will work with the molecular Hamiltonian under the second quantization formalism. Specifically, given a particular molecule of interest, define $O = \{\phi_\mu \mid \mu = 1, \ldots, N\}$ to be an orthonormal set of single-particle local orbitals (LOs), where N is the total number of orbitals; in this work, these LOs are generated through Löwdin's symmetric orthogonalization method [Löw50]. The full Hilbert space $\mathcal{H}$ for the entire molecular system is thus given by $\mathcal{H} = \mathcal{F}(O)$, where $\mathcal{F}(O)$ denotes the Fock space determined by the LOs in the set O. Further define the creation (annihilation) operator $c_\mu^\dagger$ (c_μ) which creates (annihilates) an electron in the LO ϕ_μ, the molecular Hamiltonian is written in the second-quantized notation

$$\hat{H} = \sum_{\mu\nu=1}^{N} h_{\mu\nu} c_\mu^\dagger c_\nu + \frac{1}{2} \sum_{\mu\nu\lambda\sigma=1}^{N} V_{\mu\nu\lambda\sigma} c_\mu^\dagger c_\nu^\dagger c_\sigma c_\lambda \tag{6.1}$$

where $h_{\mu\nu}$ and $V_{\mu\nu\lambda\sigma}$ are the standard one- and two-electron integrals.

Note that the number of terms in the full molecular Hamiltonian $\hat{H}$ scales polynomially with the total number of orbitals N, but the dimension of $\mathcal{H}$ scales exponentially with N. Clearly, for large N, it will become prohibitively expensive to directly compute the exact full ground state. To circumvent this issue, we divide the full molecule into multiple smaller fragments, each equipped with its own "embedding Hamiltonian" which contains a number of terms that only scales polynomially with the number of orbitals *in the fragment*. Given that there are potentially far fewer orbitals in each fragment than in the whole molecular system, computing the ground state of each fragment's embedding Hamiltonian can be significantly less expensive than computing the ground state of the full system. Furthermore, using the bootstrap embedding procedure to be described later, the ground states of individual fragments can, to a high degree of accuracy, be algorithmically combined to recover the desired electron densities prescribed by the exact ground state of the full system. Thus, this combination of fragmentation and bootstrap embedding can be used to reconstruct the full molecular ground state more efficiently than by direct computation alone.

We now briefly review the construction of embedding Hamiltonians for each fragment. Consider a single fragment associated with a label A, without loss of generality, define $O^{(A)} = \{\phi_\mu \mid \mu = 1, \ldots, N_A\}$ with $N_A \leq N$ to be the set of LOs contained in fragment A; we will refer to $O^{(A)}$ as the set of fragment orbitals. Note that $O^{(A)} \subseteq O$, the set of LOs for the entire molecular system. The construction of the embedding Hamiltonian $\hat{H}^{(A)}$ for fragment A begins with any solution of the ground state of the full system $\hat{H}$. For simplicity, the Hartree-Fock (HF) solution $|\Phi_{\text{HF}}\rangle$ is often used because it is easy to obtain on a classical computer. By invoking a Schmidt decomposition, we can write $|\Phi_{\text{HF}}\rangle$ with the following tensor product structure for $\forall\ A$

$$|\Phi_{\text{HF}}\rangle = \left(\sum_{i=1}^{N_A} \lambda_i^{(A)} |f_i^{(A)}\rangle \otimes |b_i^{(A)}\rangle \right) \otimes |\Psi_{\text{env}}^{(A)}\rangle. \tag{6.2}$$

In the above decomposition, $|f_i^{(A)}\rangle$ represent single-particle fragment states contained in the Fock space $\mathcal{F}(O^{(A)})$ of fragment orbitals. On the other hand, $|b_i^{(A)}\rangle$ and $|\Psi_{\text{env}}^{(A)}\rangle$ represent Slater determinants contained in the "environment" Fock space $\mathcal{F}(O \setminus O^{(A)})$ of the $N - N_A$

orbitals not included in the fragment. The key difference between the single environment state $|\Psi_{\text{env}}^{(A)}\rangle$ and the various "bath" states $|b_i^{(A)}\rangle$ is that the bath states $|b_i^{(A)}\rangle$ are entangled with the fragment states $|f_i^{(A)}\rangle$ while $|\Psi_{\text{env}}^{(A)}\rangle$ is not; this entanglement is quantified by the Schmidt coefficients $\lambda_i^{(A)}$. Crucially, since the HF solution is used, the sum in Eq. (6.2) only has N_A terms (as opposed to 2^{N_A} for a general many-body wave function). Denote the collection of the N_A entangled bath orbitals as $O_{\text{bath}}^{(A)} = \{\beta_\mu \ | \mu = 1, \ldots, N_A\}$, where each of the LOs β_μ are linear combinations of the original LOs not included in the fragment, $\beta_\mu \in \text{Span}\{O \setminus O^{(A)}\}$. Furthermore, we denote the Fock space that corresponds to this set of entangled bath orbitals as $\mathcal{F}(O_{\text{bath}}^{(A)})$.

This tensor product structure of $|\Phi_{\text{HF}}\rangle$ allows us to naturally decompose the Hilbert space $\mathcal{H}$ for the full molecular system into the direct product of two smaller Hilbert spaces, namely

$$\mathcal{H} = \mathcal{H}^{(A)} \otimes \mathcal{H}_{\text{env}}^{(A)}, \tag{6.3}$$

where

$$\mathcal{H}^{(A)} = \mathcal{F}(O^{(A)}) \otimes \mathcal{F}(O_{\text{bath}}^{(A)}) \tag{6.4}$$

is the active fragment embedding space and $\mathcal{H}_{\text{env}}^{(A)}$ contains the remaining states, including $|\Psi_{\text{env}}^{(A)}\rangle$. Note that since both sets $O^{(A)}$ and $O_{\text{bath}}^{(A)}$ have size N_A, the fragment Hilbert space $\mathcal{H}^{(A)}$ is a Fock space spanned of just $2N_A$ single-particle orbitals. The core intuition motivating this decomposition is that, in the exact ground state of the full system, states in $\mathcal{H}_{\text{env}}^{(A)}$ are unlikely to be strongly entangled with the many-body fragment states (consider the approximate HF ground state in Eq. (6.2), where they are perfectly disentangled); therefore, in a mean-field approximation, it is reasonable to entirely disregard the states in $\mathcal{H}_{\text{env}}^{(A)}$ when calculating the ground state electron densities on fragment A. Following this logic, we can define an embedding Hamiltonian $\hat{H}^{(A)}$ for fragment A *only* on the $2N_A$ LOs in $\mathcal{H}^{(A)}$, which will have the form

$$\hat{H}^{(A)} = \sum_{pq}^{2N_A} h_{pq}^{(A)} a_p^{(A)\dagger} a_q^{(A)} + \frac{1}{2} \sum_{pqrs}^{2N_A} V_{pqrs}^{(A)} a_p^{(A)\dagger} a_q^{(A)\dagger} a_s^{(A)} a_r^{(A)}, \tag{6.5}$$

given some creation and annihilation operators $a_p^{(A)\dagger}$ and $a_p^{(A)}$, which respectively create and annihilate electrons in orbitals from the combined set $O^{(A)} \cup O_{\text{bath}}^{(A)}$ for $\mathcal{H}^{(A)}$. The new one- and two- electron integrals $h_{pq}^{(A)}$ and $V_{pqrs}^{(A)}$ can be computed by projecting $\hat{H}$ into the smaller Hilbert space $\mathcal{H}^{(A)}$ (consult the Supporting Information (SI) Sec. D.1 for details on constructing $h_{pq}^{(A)}$ and $V_{qprs}^{(A)}$). Note that since we can choose $2N_A \ll N$, the ground state of this embedding Hamiltonian can be solved at a significantly reduced cost when compared to that of the full system Hamiltonian.

We are hence prepared to generate an embedding Hamiltonian for any arbitrary fragment of the original molecular system. However, the ground state electron densities of the fragment embedding Hamiltonian are unlikely to exactly match those of the full system Hamiltonian because, as mentioned above, the embedding process may neglect some small (but nonzero) entanglement of the fragment orbitals with the environment. Because we can expect interactions in the molecular Hamiltonian to be reasonably local, we anticipate that

the electron densities on orbitals near the edge of the fragment (those closest to the "environment") will deviate most significantly from their true values, while electron densities on orbitals toward the center of the fragment will be most accurate. Note that the uneven distribution of entanglement in molecular systems may likely lead to the potential sensitivity of the BE results to particular choices and partitions of fragments [CM19; WWG21; YV19; Kni13; LKB21; NB22], while how quantum computers may help to reduce such dependence is an open problem.

To improve the accuracy of the fragment ground state wave function near the fragment edge, we employ the technique of bootstrap embedding. Broadly speaking, we first divide the full molecule into overlapping fragments such that the edge of each fragment overlaps with the center of another. Fig. 6.1i illustrates this fragmentation strategy: for example, we see that the edge of fragment A (labeled as orbital 3) coincides with the center of fragment B. We then apply additional local potentials to the edge sites of each fragment to match their electron densities to those on overlapping center sites of adjacent fragments. Because we expect the electron densities computed on the center sites to be closer to their true values, these added local potentials should improve the accuracy of each fragment wave function near the edges. In the next section, we will formalize this edge-to-center matching process rigorously and discuss its implementation on a classical computer.

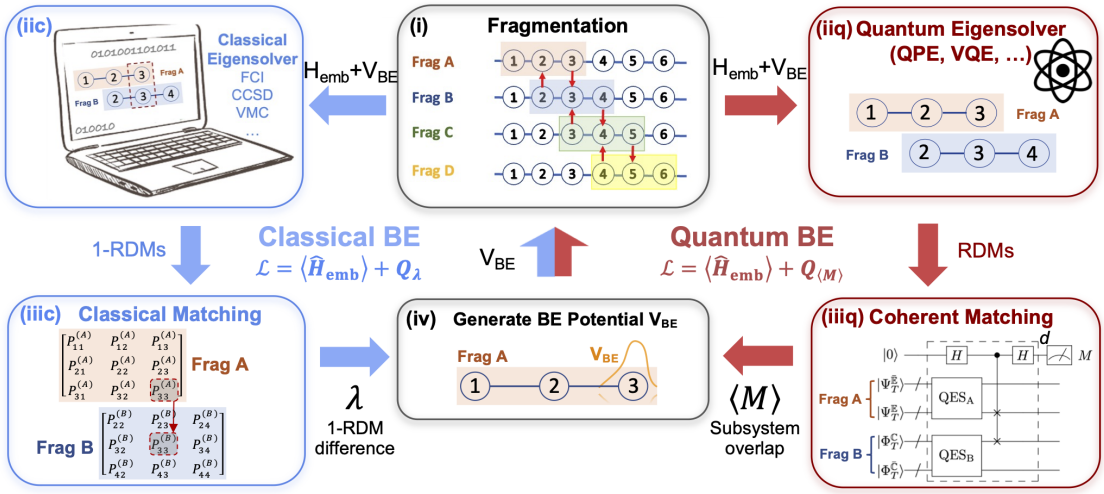


Figure 6.1: Schematic of bootstrap embedding on classical (left, blue arrows) and quantum (right, red arrows) computers. The arrows indicate BE iterative loops that are used to optimize the corresponding objective functions. Starting from panel (i) (upper center), the original system is first broken into overlapping fragments (Fragmentation), where each fragment is solved using a classical (iic) (upper left) or quantum eigensolver (iiq) (upper right). In classical matching, the 1-electron reduced density matrices (1-RDM) on the overlapping sites of adjacent fragments are used to obtain the matching condition (iiic) (lower left), while in the quantum case a coherent matching protocol based on SWAP tests of overlapping sites combined with a single qubit measurement (iiiq) (lower right). The matching results are then used by classical computers to generate the bootstrap embedding potential V_{BE} (iv) (lower center) and the updated fragment embedding Hamiltonian $H_{\mathrm{emb}} + V_{\mathrm{BE}}$ (back to panel (i) in order to minimize a target objective function $\mathcal{L}$ in both classical and quantum case.

6.2.2 Matching Electron Densities: an Optimization Problem

As mentioned in the previous section, we intend to correct the electron density error near a fragment's edge by applying a local potential to the edge; this local potential serves to match the edge electron density of the fragment to the center electron density of an adjacent overlapping fragment, which we expect to be more accurate. In principle, to achieve an exact density matching, all k-electron reduced density matrices (k-RDM, for any k) on the overlapping region have to be matched. However, in practice, such matching beyond the 2-RDM is difficult on a classical computer due to the mathematical challenge that the number of terms in k-RDM in general increases exponentially as k. In addition, almost all electronic structure codes available on classical computers are programmed to deal with only 1- and 2-RDMs, despite the importance of k-RDMs ($k > 2$) for computing observables such as entropy and other multi-point correlation functions [PA18]. Due to this reason, the discussion of density matching process in classical BE in this section will be based on 1-RDMs. We note that the matching process applies similarly if k-RDMs are matched.

We begin by introducing some rigorous notation. Recall that a fragment A is defined by a set of local orbitals $O^{(A)}$ which constitute the fragment. We partition this set of LOs into a subset of edge sites (or orbitals), denoted $\mathbb{E}^{(A)}$, and a subset of center sites, denoted $\mathbb{C}^{(A)}$, such that $\mathbb{E}^{(A)} \cup \mathbb{C}^{(A)} = O^{(A)}$ and $\mathbb{E}^{(A)} \cap \mathbb{C}^{(A)} = \emptyset$. Given the ground state wave function $|\Psi^{(A)}\rangle$ of the embedding Hamiltonian, we further define the 1-electron reduced density matrix (1-RDM) $\mathbf{P}^{(A)}$ according to

$$P^{(A)}_{pq} = \langle \Psi^{(A)} | a^{(A)\dagger}_p a^{(A)}_q | \Psi^{(A)} \rangle \tag{6.6}$$

where $p, q = 1, \ldots, 2N_A$ and the operators $a^{(A)\dagger}_p$ and $a^{(A)}_q$ are defined in the previous section.

Suppose, for example, that the edge of fragment A overlaps with the center of another fragment B so that $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)} \neq \emptyset$. On a high level, the goal of bootstrap embedding is to find a ground state wave function $|\Psi^{(A)}\rangle$, perturbed by local potentials on the edge sites of A, such that $|P^{(A)}_{pq} - P^{(B)}_{pq}| \to 0$ for indices p and q that correspond to orbitals in the set of overlapping sites $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)}$. More generally, and more rigorously, the goal is to find a wave function which minimizes the fragment Hamiltonian energy

$$|\Psi^{(A)}\rangle = \arg\min_{\Psi^{(A)}} \langle \hat{H}^{(A)} \rangle_A \tag{6.7}$$

subject to the constraints

$$\langle a^{(A)\dagger}_p a^{(A)}_q \rangle_A - P^{(B)}_{pq} = 0 \tag{6.8}$$

for *all* other fragments B with $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)} \neq \emptyset$ and for all p, q corresponding to orbitals in $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)}$. Here, we explicitly write the expectation $\langle \cdot \rangle_A = \langle \Psi^{(A)} | \cdot | \Psi^{(A)} \rangle$ in terms of $|\Psi^{(A)}\rangle$ to indicate that the optimization is over the wave function of A.

We can formulate this constrained optimization problem as finding the stationary solution to a Lagrangian by associating a scalar Lagrange multiplier $(\lambda^{(A)}_B)_{pq}$ to Eq. (6.8). Since Eq. (6.8) has to be satisfied for any p, q and B that overlaps with A, these constraint can be rewritten in a more compact vector form $\boldsymbol{\lambda}^{(\boldsymbol{A})}_{\boldsymbol{B}} \cdot \boldsymbol{\mathcal{Q}}_{\textbf{1-RDM}}(\Psi^{(A)}; \mathbf{P}^{(B)})$ where the dot product conceals the implicit sum over p, q, and each component of the vector $\mathcal{Q}_{1-\mathrm{RDM}}(\Psi^{(A)}; \mathbf{P}^{(B)})_{pq}$

represents the constraint associated with Lagrange multiplier $(\lambda_B^{(A)})_{pq}$, given by the left hand side of Eq. (6.8). With this notation, we arrive at the following Lagrangian with the constraint added as an additional term

$$\begin{aligned}\mathcal{L}^{(A)} =& \langle \hat{H}^{(A)} \rangle_A + \mathcal{E}^{(A)} \left(\langle \Psi^{(A)} | \Psi^{(A)} \rangle - 1 \right) \\ &+ \sum_B \boldsymbol{\lambda}_{\boldsymbol{B}}^{(\boldsymbol{A})} \cdot \boldsymbol{\mathcal{Q}}_{\textbf{1-RDM}}(\Psi^{(A)}; \mathbf{P}^{(B)}),\end{aligned} \tag{6.9}$$

where once again the B are fragments adjacent to A with $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)} \neq \emptyset$ and p, q are indices of orbitals contained in the overlapping set $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)}$. Here, the additional constraint with Lagrange multiplier $\mathcal{E}^{(A)}$ is also included to ensure normalization of the ground state wave function $|\Psi^{(A)}\rangle$. Solving for the stationary solution of the Lagrangian in Eq. (6.9) will only result in a ground state wave function for fragment A whose 1-RDM elements at the edge sites match those at the center sites of adjacent overlapping fragments. However, we would instead like to solve for such a ground state for *all* fragments in the molecule simultaneously. Toward this regard, we can combine all individual fragment Lagrangians (of the form of Eq. (6.9)) into a single composite Lagrangian for the whole molecule, given by

$$\mathcal{L} = \sum_{A=1}^{N_{\text{frag}}} \mathcal{L}^{(A)} + \mu \mathcal{P} \tag{6.10}$$

where N_{frag} is the number of fragments in the molecule. Observe that we have added one additional constraint

$$\mathcal{P} = \left(\sum_{A=1}^{N_{\text{frag}}} \sum_{p' \in \mathbb{C}^{(A)}} \langle a_{p'}^{(A)\dagger} a_{p'}^{(A)} \rangle_A \right) - N_e \tag{6.11}$$

with Lagrange multiplier μ to restore the desired total number of electrons in the molecule, N_e. Note in Eq. (6.11) that p' is summed over indices corresponding to orbitals only in $\mathbb{C}^{(A)}$; this is to ensure that there is no double-counting of electrons in the whole molecule. By self-consistently finding ground states $|\Psi^{(A)}\rangle$ for $A = 1, \ldots, N_{\text{frag}}$ which make the composite Lagrangian in Eq. (6.10) stationary, we will have completed the density matching procedure for all fragments, and the process of bootstrap embedding will be complete.

We can gain insight into which wave functions $|\Psi^{(A)}\rangle$ will make the composite Lagrangian $\mathcal{L}$ stationary by differentiating $\mathcal{L}$ with respect to $|\Psi^{(A)}\rangle$ for some fixed fragment A and setting the resulting expression equal to zero. Upon some algebraic manipulation, we can recover the eigenvalue equation

$$(\hat{H}^{(A)} + V_{\text{BE}}) |\Psi^{(A)}\rangle = -\mathcal{E}^{(A)} |\Psi^{(A)}\rangle, \tag{6.12}$$

where V_{BE}, the local bootstrap embedding potential, is given by

$$V_{\text{BE}} = \sum_B \sum_{p,q} (\lambda_B^{(A)})_{pq} a_p^{(A)\dagger} a_q^{(A)} + \mu \sum_{p'} a_{p'}^{(A)\dagger} a_{p'}^{(A)} \tag{6.13}$$

where the p, q are indices of orbitals in the overlapping set $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)}$, and the p' are indices of orbitals in the fragment center $\mathbb{C}^{(A)}$. We see that, when the composite Lagrangian is made stationary with respect to the fragment wave functions, the bare fragment embedding Hamiltonians become dressed with a potential V_{BE} that contains a component local to the edge sites of each fragment (see the left term of Eq. (6.13)). This observation confirms our intuition that adding a local potential to the edge of one fragment will allow the edge site electron density to be matched to that of a center site on an overlapping neighbor. Note that V_{BE} also contains an additional potential on the center sites of each fragment (see the right term of Eq. (6.13)); this is simply to conserve the total electron number in the molecule. Moreover, V_{BE} as in Eq. (6.13) only contains one-body terms because only 1-RDM is used for density matching. In general, V_{BE} will contain up to k-body terms if k-RDMs are used for matching.

On a classical computer, the composite Lagrangian in Eq. (6.10) is made stationary through an iterative optimization algorithm [WTV16] until the edge-to-center matching condition for all fragments is satisfied by some criterion. One possible criterion is to terminate the algorithm when the root-mean-squared 1-RDM mismatch, given by

$$\epsilon = \left[\frac{1}{N_{\mathrm{sites}}} \sum_{A}^{N_{\mathrm{frag}}} \sum_{B} \sum_{p,q} (P_{pq}^{(A)} - P_{pq}^{(B)})^2 \right]^{\frac{1}{2}}, \tag{6.14}$$

drops below some predetermined threshold. Note again that p, q are indices corresponding to orbitals in the overlapping set $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)}$; also, N_{sites} denotes the total number of overlapping sites in the whole molecule, equal to $N_{\mathrm{sites}} = \sum_{A}^{N_{\mathrm{frag}}} \sum_{B} \sum_{p,q} 1$. The final set of density-matched fragment wave functions $\{|\Psi^{(A)}\rangle\}$ for $A = 1, \ldots, N_{\mathrm{frag}}$ which solve the composite Lagrangian can then be used to reconstruct the electron densities and other observables for the full molecular system, as desired.

6.2.3 Resource Requirement and Typical Behavior of BE on Classical Computers

Given the notation established for classical BE, we now begin presenting new material. We discuss the computational resource requirement and typical behaviors of performing BE on classical computers to set the stage for a quantum BE theory. The details of the classical BE algorithms are omitted for succinctness, and we refer the reader to Ref. [WTV16; Ye+19; YTV20; YTV21] for details.

The space and time resource requirement to perform the classical BE can be broken down into two parts: a) the number of iteration steps to reach a fixed accuracy for ϵ (Eq. (6.14)); b) the runtime of the fragment eigensolver. For a), numerical evidence suggests an exponentially fast convergence on total system energy as the number of bootstrap iteration increases (black trace in Fig. 6.2 for FCI), while a proof of the convergence rate has yet to be established.

We focus on resource requirement in b) in the following. Admittedly, an exact classical eigensolver such as full configuration interaction (FCI) can be used to solve the embedding Hamiltonian in Eq. (6.5). However, both the storage space and time requirement scales

exponentially as the the number of orbitals (see blue symbols and dashed line in Fig. 6.3 for the runtime scaling of FCI). Even with the state-of-the-art classical computational resources, exact solutions using FCI are only tractable for systems up to 20 electrons in 20 orbitals [Vog+17].

As a result, classical computation of BE resorts to approximate eigensolvers with only polynomial cost in practice, by properly truncating or sampling from the fragment Hilbert space. One example for truncation is the coupled-cluster singles and doubles (CCSD) [BM07], which scales with N^6 with N being the number of orbitals. Alternately, different flavors of stochastic electronic structure solvers can be employed as fragment solvers in BE. Depending on implementation, these stochastic solvers can be biased or unbiased (if unbiased, with a cost of introducing the phase problem in general) [Mor+21; LPR22; She+19; LCR18]. Collecting each sample on a classical computer usually has similar cost as a mean field theory (roughly $O(N^3)$), while the overall target accuracy ϵ on observable estimation can be achieved with a sampling overhead of roughly $O(\frac{1}{\epsilon^2})$ with a constant prefactor depending on the severity of the sign problem.

Importantly, the sampling feature of these stochastic electronic structure methods on classical computers are strikingly similar to the nature of quantum computers where measurement necessarily collapses the wave function. As a result, only a classical sample (in terms of measurement results) can be obtained from a quantum computer. This similarity suggests a general strategy that many sampling techniques in stochastic classical algorithms can be deployed to design better quantum algorithms. For example, sophisticated importance sampling techniques [NU98; Fou+01] can be employed to greatly improve the sampling efficiency in both classical [LCR18] and quantum cases [HKP20].

Due their shared feature on sampling between classical stochastic algorithm and quantum eigensolvers, we shall use one approximate sign-problem-free flavor of stochastic electronic structure method, the variational Monte Carlo (VMC), to serve as an additional baseline scenario for comparison with quantum BE in later sections. In addition to BE convergence behavior with a FCI solver, Fig. 6.2 also shows, for a VMC eigensolver with single Slater-Jastrow type wave function with two-body Jastrow factors [WM07; Whe+21], the density mismatch converges exponentially fast initially as iteration number increases with varying number of samples. However, partially due to the statistical noise on estimating the 1-RDM (thus the gradient for the optimization), the final density mismatch plateaus to a finite biased value. Comparing the VMC results across different numbers of samples, we can see that the bias improves as the number of samples increases (dashed horizontal lines). It is also evident that the orange trace (640k samples) has smaller fluctuation as compared to blue (160k samples) and grey (40k samples). Strictly speaking, an increase of sample size by a factor of 16 would decrease the statistical fluctuations by a factor of 4. However, our numerical data in Fig. 6.2 only shows *qualitative* but not quantitative agreement with this statement. We attribute part of the bias in the plateau to the intrinsic truncation of the VMC ansatz in addition to statistical fluctuations.

The increasing accuracy of density mismatch with respect to BE iteration also suggests an increasing number of samples are needed. Thus, an optimal number of samples at each BE iteration must be determined to achieve the desired accuracy in the matching conditions. A careful design of such a sampling schedule can potentially save a large amount of computational resources. We defer a thorough discussion of this point to later sections on quantum

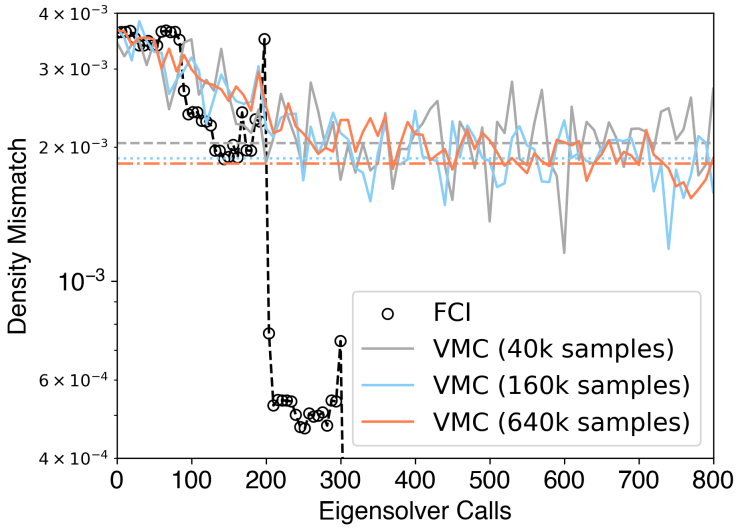


Figure 6.2: Typical convergence of density mismatch with respect to the number of eigensolver calls in classical bootstrap embedding with a deterministic eigensolver (FCI, black circle) and a stochastic eigensolver (VMC) with different number of samples (grey, blue, and orange solid lines). The horizontal dashed lines shows the final plateaued value of the density mismatch for VMC, while the FCI data converges to 10^{-6} after 700 eigensolver calls (not shown on the figure). The discrete jumps around 200 and 300 eigensolver calls are due to switching to the next BE iteration. The data is obtained for an H_8 linear chain under STO-3G basis. See SI Sec. D.9.2 for computational details.

BE.

6.2.4 The Quest for BE on Quantum Computers

By employing the coherent superposition and entanglement of quantum states, the limitation of an exact classical solver can be overcome by substituting it with an exact quantum eigensolver such as the quantum phase estimation (QPE) algorithm [AL99b]. This section directly compares the cost between the two exact eigensovlers on quantum and classical computers, the QPE and the FCI solvers, using hydrogen chains where the initial trial state with a non-vanishing overlap with the exact eigenstate for QPE can be efficiently prepared on classical computers.

Fig. 6.3 compares the runtime (gate depth) of FCI and QPE for finding the ground state of linear hydrogen chain H_n for different system size n. Clearly, the QPE runtime scales only polynomially as the system size increases as expected [Asp+05; Lee+22], while its classical counterpart (FCI) has an exponentially increasing runtime. Note the runtime is normalized to the case of $n = 1$ for each solver separately (see SI Sec. D.9). The dramatic advantage in the runtime scaling of quantum over classical eigensolvers demonstrated above suggests formulating BE on a quantum computer can bring significant benefits.

One might think that the eigensolver at the heart of the classical BE algorithm could simply be replaced with a quantum one. However, as mentioned before, there are two outstanding challenges for such a quantum bootstrap embedding (QBE) method. First, just

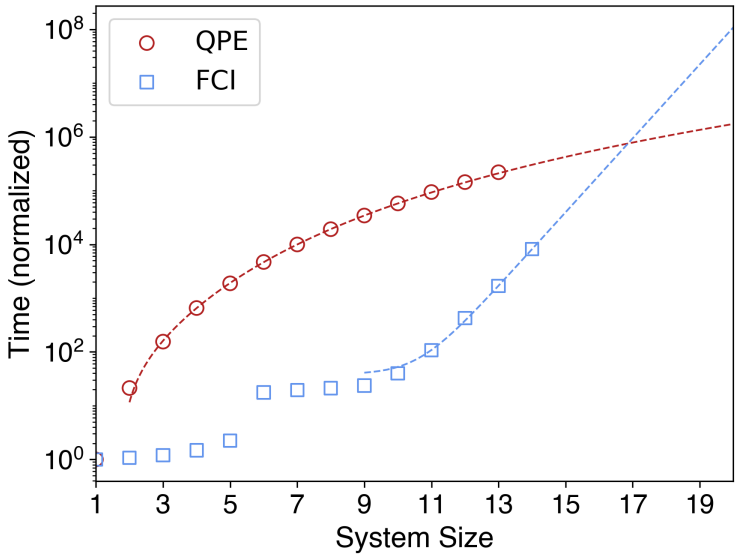


Figure 6.3: Runtime (normalized) as a function of system size n for finding the ground state of a linear hydrogen chain H_n at STO-3G basis, comparing an exact classical solver (FCI, blue square) and an exact quantum solver (QPE, red circle) on real classical and quantum devices. Red (blue) dashed line shows a polynomial (exponential) fit to the QPE (FCI) runtime. Note the crossover at large system size.

as in classical stochastic methods, the results of a quantum eigensolver need to be measured for later use, but quantum wave functions collapse after measurement. Therefore, sampling from the quantum eigensolver is required, and the optimal sampling strategy is unclear. Secondly, with quantum wave function from quantum eigensolvers, it is not wise to achieve matching between fragments in the same way as classical BE, as many incoherent samples are needed to obtain a good estimation of the 1-RDM elements. Clearly, performing matching in a quantum way is desired.

In the next two sections (Secs. 6.3 and 6.4), we present how we address these two challenges by an adaptive quantum sampling scheduling algorithm and a quantum coherent matching algorithm in detail.

6.3 Quantum Bootstrap Embedding Methods

In previous sections, we have seen potential advantages of performing bootstrap embedding on quantum computers, and discussed two major challenges of doing so. In this section, we present the theoretical formulation of our bootstrap embedding method on a quantum computer that addresses these challenges.

Sec. 6.3.1 first set up notations and discuss a few aspects of locality and global symmetry on performing embedding of fermions on quantum computers. Sec. 6.3.2 discuss a naive extension of the classical BE algorithm on quantum computers by matching individual elements of the RDMs directly, and highlight the disadvantage of doing so. Sec. 6.3.3 introduces the `SWAP` test circuit and show that it achieves the matching between two RDMs

coherently. In 6.3.4, we discuss some subtleties on why it is impossible to incorporate this coherent matching condition into the Lagrange multiplier optimization method, and present an alternative quadratic penalty method to perform the optimization.

6.3.1 Fermion-Qubit Mapping - Global Symmetry vs. Locality

When mapping electronic structure problem to qubits on quantum computers, it is well-known that the global anti-symmetric property of fermionic wave functions necessarily leads to an overhead in operator lengths or qubit counts [Tra+18]. On the other hand, chemical information is usually local if represented using localized single-particle orbitals [ER63; Wan62]. In the case of performing bootstrap embedding, this tension between locality of chemical information and global fermionic anti-symmetry is more subtle. Because bootstrap embedding intrinsically uses the fermionic occupation number in the local orbitals (LOs) to perform matching, it is therefore convenient to preserve such locality when constructing the mapping. Throughout the discussion, without loss of generality, we assume a mapping that preserves fermionic local occupation number, such as the Jordan-Wigner mapping where each spin-orbital is mapped to one qubit. Our discussion equally applies to cases where a non-local mapping is used (such as parity mapping). In that case, a unitary transformation from the non-local mapping to a local mapping will be required before actually computing the matching conditions.

It is possible to formulate QBE using matching conditions on either qubit reduced density matrices (RDMs) [NC10] or k-electron RDMs [Maz12] for all k, both with an exponential number of matrix elements. For simplicity, in the present work we use qubit RDMs in our QBE and leave an efficient formulation in terms of fermionic k-electron RDMs for future work. The full density matrix of fragment A is thus provided by $\rho^{(A)} = |\Psi_A\rangle\langle\Psi_A|$. Given an orbital set $R \subset O^{(A)}$ for $O^{(A)}$ being set of orbitals in fragment A. Let $\rho_R^{(A)}$ signify the RDM obtained from $\rho^{(A)}$ by tracing out the set of qubits not in R. Specially, if R only contains orbitals on the edge (center) of fragment A, then $\rho_R^{(A)}$ represents information about the density information (for example the occupation number) on the edge (center) of A.

These RDMs can be expanded under an arbitrary set of orthonormal basis $\{\Sigma_\alpha\}$ as follows

$$\rho_R^{(A)} = \frac{I + \sum_{\alpha=1}^{4^m-1} \langle\Sigma_\alpha\rangle_A \, \Sigma_\alpha}{2^m} \tag{6.15}$$

where $\langle\Sigma_\alpha\rangle_A = \langle\Psi_A|\Sigma_\alpha|\Psi_A\rangle = \mathrm{Tr}[\rho^{(A)} \, \Sigma_\alpha]$, $\forall\alpha \in [1, 4^m - 1]$, and $m = |R|$ is the number of orbitals in the set R. One convenient orthonormal basis set is the generalized Gell-Mann basis [BK08]. In the special case of a 1-qubit RDM, $\{\Sigma_\alpha\}$ ($\alpha = x, y, z$) is the familiar Pauli matrices.

6.3.2 Naive RDM Linear Matching and its Disadvantage

A naive implementation of BE on a quantum computer is to simply replace 1-RDM in Eq. (6.6) with the qubit RDM in Eq. (6.15) on the fragment overlapping regions. Such an extension imposes matching constraints on each elements of the RDMs, resulting the

following constraint vector in analogous to Eq. (6.8)

$$\mathcal{Q}_{lin}(\rho_R^{(A)};\rho_R^{(B)}) = \begin{bmatrix} \langle\Sigma_1\rangle_A - \langle\Sigma_1\rangle_B \\ \vdots \\ \langle\Sigma_{4^m-1}\rangle_A - \langle\Sigma_{4^m-1}\rangle_B \end{bmatrix} = \mathbf{0}. \tag{6.16}$$

It is obvious that $\rho_R^{(A)} - \rho_R^{(B)} = 0$, if and only if all the $(4^m - 1)$ components in the above constraint are satisfied.

Similarly, we can associate a scalar Lagrange multiplier to each constraint in Eq. (6.16) and use this linear RDM constraint in place of the 1-RDM constraint $\mathcal{Q}_{\textbf{1-RDM}}(\Psi^{(A)};\mathbf{P}^{(B)})$ in Eq. (6.9). Finding the stationary point of this new Lagrangian gives the same eigenvalue equation as Eq. (6.12) with a new BE potential given by

$$V_{\text{BE}} = \sum_{B\neq A, \mathbb{C}_B\cap\mathbb{E}_A\neq\emptyset} \boldsymbol{\lambda}_B^{(A)} \cdot [I \otimes \boldsymbol{\Sigma}_{\mathbf{r}} \otimes I] \tag{6.17}$$

where $\boldsymbol{\Sigma}_{\mathbf{r}} = \left[\Sigma_1, \cdots, \Sigma_\alpha, \cdots, \Sigma_{4^m-1}\right]$ is a $(4^m - 1)$-dimensional vector of the orthonormal basis in Eq. (6.15), and $\boldsymbol{\lambda}_B^{(A)}$ is the Lagrange multipliers now modulating the local potentials on each qubit basis, and n is the number of overlapping sites between A and B.

To perform the optimization, the eigenvalue equation Eq. (6.12) with the above new BE potential in (6.17) can be solved on a quantum computer to obtain an updated wave function for fragment A. By iteratively solving the eigenvalue equation and updating the Lagrange multipliers $\{\boldsymbol{\lambda}, \mu\}$ using either gradient-based or gradient-free methods [CSV09], an algorithm can be formulated to solve the optimization problem. For completeness, we document the algorithm from the naive linear matching of RDMs in Sec. D.8 of the SI.

The above is a convenient way to impose the constraint on quantum computers, but it is computationally costly as the number of constraints in (6.16) increases exponentially as the number of overlapping sites n on neighboring fragments. For each constraint equation, the expectation values $\langle\Sigma_\alpha\rangle$ has to be measured on the quantum computer, which therefore introduces an exponential overhead on the sampling complexity.

In the next section, we introduce a simple alternative to evaluate the mismatch between two RDMs on a quantum computer much faster based on a `SWAP` test.

6.3.3 Coherent Quantum Matching from `SWAP` Test

The wave functions of two overlapping fragments are stored coherently as many amplitudes that suppose with each other. The beauty of quantum computers and algorithms lies at the ability to coherently manipulating such amplitudes simultaneously. We may naturally ask: are there quantum algorithms or circuits that can coherently achieve matching between an exponentially large number of amplitudes, without explicitly measuring each amplitude?

In quantum information, there is a class of quantum protocols to perform the task of estimating the overlap between two wave functions or RDMs under various assumptions [Fan+20]. Among these protocols, the `SWAP` test is widely used [HM13; Buh+01]. Such a `SWAP` test on a quantum computer can also be naturally implemented by simple controlled-`SWAP` operations as in Fig. 6.4, showing a `SWAP` test between two qubits. The essence of a

SWAP test is to entangle the symmetric and anti-symmetric subspaces of the two quantum states ($|\phi\rangle$ and $|\psi\rangle$) to a single ancillary qubit, such that the quantum state of the system before the final measurement is

$$|\Psi\rangle = \frac{1}{2}\left[|0\rangle\left(|\phi\rangle|\psi\rangle + |\psi\rangle|\phi\rangle\right) + |1\rangle\left(|\phi\rangle|\psi\rangle - |\psi\rangle|\phi\rangle\right)\right]. \tag{6.18}$$

By measuring the top *single* ancillary qubit in the usual computational Z-basis (collapsing it to either the $|0\rangle$ or $|1\rangle$ state), the overlap of the two qubit wave function, $|\langle\phi|\psi\rangle|$, can be directly obtained from the measurement outcome probability:

$$\text{Prob}[M=0] = \frac{1+|\langle\phi|\psi\rangle|^2}{2}, \tag{6.19}$$

without requiring explicit estimation of the density matrix elements of each individual qubit.

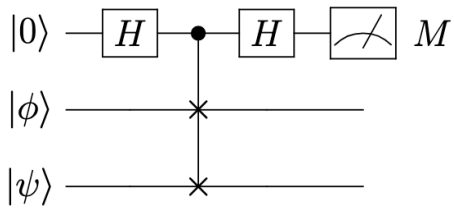


Figure 6.4: Quantum circuit of a SWAP test between two qubits (lower, with state $|\phi\rangle$ and $|\psi\rangle$). The circuit is composed of two Hadamard gate (H), a controlled-SWAP operation in between, and a final Z-basis measurement M on an additional ancilla qubit (top), where $M = 0, 1$.

Can we recast the linear matching conditions as linear combination of several SWAP tests? Observe that an equivalent condition alternative to Eq. (6.16) is the following quadratic matching condition

$$\mathcal{Q}_{quad}(\rho_R^{(A)}; \rho_R^{(B)}) = \text{Tr}[\left(\rho_R^{(A)} - \rho_R^{(B)}\right)^2] = 0. \tag{6.20}$$

Interestingly, the above quadratic constraint can be rewritten as a linear combination of three different multi-qubit generalization of the SWAP tests (with each repeated multiple times), regardless of the number of overlapping sites (Fig. 6.1iiiq). Two of the SWAP tests are to estimate the purity of $\rho_R^{(A)}$ and $\rho_R^{(B)}$ each, while the third one is to estimate the overlap between $\rho_R^{(A)}$ and $\rho_R^{(B)}$. See SI Sec. D.3 for a proof of the equivalence between the two quantum matching conditions) and Sec. D.4 on how to generalize the SWAP test on two qubits to a multi-qubit setting and how to relate the SWAP test results to the quadratic constraint.

The reformulation of the quadratic constraint allows us to estimate the mismatch between two fragments by measuring only a single ancilla qubit (estimating three different amplitudes). As compared to the linear constraint case where an exponentially large number of constraints have to be estimated individually ($4^m - 1$ where $m = |R|$ is the number of

overlapping sites again), the quadratic matching based on SWAP tests achieves an exponential saving in the types of measurements required.

Furthermore, the reduction of the mismatch to the estimation of only a few (three) amplitudes in SWAP tests allows an additional quadratic speedup by amplifying the amplitude of the ancilla qubit before measure it. We will discuss more details on how to achieve the quadratic speedup in Sec. 6.4.3. Admittedly, such amplitude amplification algorithm may be applied even to the naive linear RDM matching by boosting individual RDM amplitude, but the resulting quantum circuit will be much more complicated.

6.3.4 Optimization Using the Quadratic Penalty Method

With an efficient way to estimate the quadratic penalty constraint established in Eq. (6.20), it now appears feasible to use this new constraint in Eq. (6.9) as in the case of linear constraint. However, the nature of the quadratic matching in Eq. (6.20) makes the same Lagrange multiplier optimization method used in the linear case invalid. We first discuss in more detail why this approach fails, in Sec. 6.3.4; we then describe an alternative way of treating the quadratic constraint as a penalty term to optimize the resulting objective function in Sec. 6.3.4.

Violation of the Constraint Qualification

A necessary condition to use the Lagrange multiplier method for constraint optimization is that the gradient of the constraint itself with respect to system variables has to be non-zero at the solution point (this guarantees a non-zero effective potential to be added to the original Hamiltonian), a.k.a., constraint qualification [MF67; Ber16]. Specifically, we require $\nabla \mathcal{Q}_{quad}(\rho_R^{(A)}; \rho_R^{(B)}) \neq 0$ when $\rho_R^{(A)} = \rho_R^{(B)}$.

Unfortunately, in the quadratic case, we have

$$\nabla \mathcal{Q}_{quad}(\rho_R^{(A)}; \rho_R^{(B)}) \propto \rho_R^{(A)} - \rho_R^{(B)} = 0 \tag{6.21}$$

when $\rho_R^{(A)}$ and $\rho_R^{(B)}$ matches, which violates the above condition. Note that any high-order constraint other than linear order will violate the constraint qualification. The existence of such constraint qualification makes sense also from a physical point of view. Because the gradient $\nabla \mathcal{Q}_{quad}(\rho_R^{(A)}; \rho_R^{(B)})$ enters the eigenvalue equation (6.13) as the BE potential V_{BE} modulated by the Lagrange multipliers. The vanishing of this potential near the solution point means there is no way to modulate V_{BE} by adjusting the Lagrange multipliers, and therefore will lead to failure of convergence of the Lagrange multiplier.

Alternatively, the quadratic constraint can be treated as a penalty by using $\lambda_B^{(A)} \mathcal{Q}_{quad}(\rho_R^{(A)}; \rho_R^{(B)})$ to substitute the constraint $\boldsymbol{\lambda_B^{(A)}} \cdot \boldsymbol{\mathcal{Q}_{\text{1-RDM}}}(\Psi^{(A)}; \mathbf{P}^{(B)})$ in Eq. (6.9). We can then employ the quadratic penalty method [Sta+10] to minimize this cost function. To highlight the distinction of quadratic penalty method from the Lagrange multiplier method, we use "cost function" instead of "Lagrangian" to refer to the objective function in the quadratic penalty case.

Details of the Quadratic Penalty Method

The idea of the penalty method is to use the constraint as a penalty where the magnitude of $\lambda_B^{(A)}$ serves as a weight to the penalty. Initially, $\lambda_B^{(A)}$ is set to a small constant, and then we treat the resulting cost function as an unconstrained minimization where its minimum is found by varying the wave functions. The next step is to increase $\lambda_B^{(A)}$ to a larger value leading to a new Lagrangian, which is then minimize again by varying the wave function parameters. This procedure is repeated until the penalty parameter $\lambda_B^{(A)}$ is large enough to guarantee a small mismatch $\mathcal{Q}_{quad}(\rho_r^{(A)}; \rho_r^{(B)})$. In our case, we choose all $\lambda_B^{(A)} = \lambda$ for all pairs of adjacent fragments.

It is helpful to note that optimization of the wave function is done again using the eigenvalue equation as in Eq. (6.12) by tuning the BE potential V_{BE}. In other words, for a fixed penalty parameter λ, the fragment Lagrangian $\mathcal{L}_A(\{V_{\mathrm{BE}}\})$ is minimized with respect to V_{BE}. For a particular parametrization in terms of local potentials $\{v_\alpha\}$ on the edge sites of fragment A

$$V_{\mathrm{BE}}(\{v_\alpha\}) = \sum_{\alpha=0}^{M} v_\alpha \; I \otimes \Sigma_\alpha \otimes I, \tag{6.22}$$

where $\{\Sigma_\alpha\}$ is a set of Hermitian generator basis of size M on the edge sites of fragment A (can be Pauli operators for a single edge site), and $\{v_\alpha\}$ is the corresponding local potential (real numbers). Note that M in Eq. (6.22) can be much smaller than the total number of generators (4^m) on the edge sites, because in each bootstrap embedding iteration, only a small local potential is added to the Hamiltonian. This perturbative nature of the bootstrap embedding iteration allows us to expand the BE potential V_{BE} in each iteration under the Hermitian generator basis from the previous iteration, such that the BE potential in each iteration is diagonal dominant, i.e., $M \ll 4^m$ where n is the number of edge sites on any fragment A.

To update $\{v_\alpha\}$, we derive the following gradient (SI Sec. D.5)

$$\frac{d\mathcal{L}^{(A)}}{dv_\alpha} = \sum_{n' \neq 0} \left[\mathbf{C}^\dagger (\mathbf{I} \otimes \mathbf{W}_{\mathbf{ff}}^{(\mathbf{n}')} \otimes \mathbf{I}) \mathbf{C}^{(\mathbf{n}')}\right] \times \left[\mathbf{C}^{(\mathbf{n}')\dagger} \left(\mathbf{H}^{(\mathbf{A})} + \mathcal{E}_{\mathbf{0}}^{(\mathbf{A})} + 2\lambda \left(\mathbf{I} \otimes (\mathbf{æ}_{\mathbb{E}_\mathbf{A}} - \mathbf{æ}_{\mathbb{C}_\mathbf{B}}) \otimes \mathbf{I}\right)\right) \mathbf{C}\right] \tag{6.23}$$

$\forall \alpha \in [0, M]$, that can, in principle, be used to perform the updating of V_{BE} to minimize $\mathcal{L}^{(A)}$. In the above, $\mathbf{C}^{(\mathbf{n})}$ is the eigenvector of the n-th eigenstate ($n \geq 1$) while $\mathbf{C}$ is the eigenvector of the ground state, $\mathbf{W}_{\mathbf{ff}}^{(\mathbf{n}')}$ is a perturbation matrix between ground state and the n'-th eigenstate for the α-th Pauli basis at the edge site of fragment A, whereas $\rho_{\mathbb{E}_A}$ and $\rho_{\mathbb{C}_B}$ are the RDM at the edge and center sites of fragment A and B, respectively.

The above gradient in Eq. (6.23) is only formally useful, but computing it exactly requires all the eigenstates to be known (not only the ground state) which is clearly very costly if possible. Nevertheless, it serves as a good starting point to develop *approximated* updating scheme or to perform bootstrap embedding for excited states. We leave such topics for future investigation. In the present work, instead of using Eq. (6.23) to update V_{BE}, we employ gradient-free schemes to update $\{v_\alpha\}$ and measure the required expectation values using `SWAP` test to obtain the mismatch to evaluate the cost function $\mathcal{L}^{(A)}$.

We note that one additional advantage of this quadratic penalty method is that it can be easily integrated with variational eigensolvers [Til+22] by treating the quadratic penalty

as an additional term in the VQE cost function [KN21]. The drawback is that the optimized wave function only *exactly* equals to the true wave function when the penalty goes to infinity $\lambda \to \infty$. Practically, we find that choosing the penalty parameter large enough is sufficient to obtain satisfactory results.

6.4 Quantum Bootstrap Embedding Algorithms

Given the theoretical formulation of QBE method in Sec. 6.3, we present a general hybrid quantum-classical algorithm in this section that can be practically used to solve the BE problem on quantum computers to find the BE potentials V_{BE} that satisfies the matching condition.

In our quantum bootstrap embedding algorithm, the electronic structure problem of the total system is formulated as a minimization of a composite objective function with a penalty term constructed from the matching conditions on the full qubit RDMs on overlapping regions of adjacent fragments. We then design an iterative hybrid quantum-classical algorithm to solve the optimization problem, where a quantum subroutine as an eigensolver is employed to prepare the ground state of fragment Hamiltonian. The quantum matching algorithm employs a `SWAP` test [Bar+97; Buh+01] between wave functions of two fragments to evaluate the matching conditions, which is a dramatic improvement as compared to the straightforward method of measuring an exponential number (with respect to the number of qubits on the fragment edge) of RDM elements. Additionally, the quantum bootstrap embedding framework is internally self-consistent without the need to match fragment density matrices to external more accurate solutions. The adaptive sampling changes the number of samples as the optimization proceeds in order to achieve an increasingly better matching conditions. We note that the `SWAP` test adds only little computational cost to quantum eigensolvers which can be readily performed on current NISQ devices. The amplitude amplified coherent quantum matching requires iterative application of eigensolvers multiple times which are more suitable for small fault-tolerant quantum computers.

The rest of this section is organized as follows. Sec. 6.4.1 gives an outline of the QBE algorithm with the quadratic penalty method. Sec. 6.4.2 discusses possible choices of quantum eigensolvers with an analysis on sampling complexities. We then present a way to achieve an additional quadratic speedup by using coherent amplitude estimating algorithm in Sec. 6.4.3.

6.4.1 The Algorithm

We present a high-level framework of the main algorithm in this section. As a comparison, the QBE algorithm with naive linear matching can be found in SI Sec. D.8. Code for the algorithms and data for generating the plots are available as open source on github [Liu+22].

To quantify the mismatch across all fragments, we define $\Delta\rho$ to be the root mean square density matrix mismatch averaged over all the overlapping sites of all the fragments according to

$$\Delta\rho = \sqrt{\frac{1}{N_{sites}} \sum_{A,B} \sum_{r \in \mathbb{E}^{(A)} \cap \mathbb{C}^{(B)}} \mathrm{Tr}[\left(\rho_r^{(B)} - \rho_r^{(A)}\right)^2]} \tag{6.24}$$

where $\mathrm{Tr}[\left(\rho_r^{(B)} - \rho_r^{(A)}\right)^2] = \mathcal{Q}_{quad}(\rho_r^{(A)}; \rho_r^{(B)})$ as in Eq. (6.20), which may also be recognized as the Frobenius norm of $(\rho_r^{(B)} - \rho_r^{(A)})$. N_{sites} is the total number of terms in the double sum in Eq. (6.24), $N_{sites} = \sum_{A \neq B} |\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)}|$, with $|\mathcal{S}|$ denoting the number of elements in set $\mathcal{S}$.

The cost function $\mathcal{L}^{(A)}(\lambda)$ being optimized is discussed in Sec. 6.3.4. For clarity, we write it explicitly here

$$\mathcal{L}^{(A)}(\lambda) = \langle \hat{H}^{(A)} \rangle_A + \sum_B \lambda \mathcal{Q}_{quad}(\rho_R^{(A)}; \rho_R^{(B)}), \tag{6.25}$$

with $\mathcal{Q}_{quad}$ given by Eq. (6.20). We have omitted the term $\mathcal{E}^{(A)}$ for simplicity since the normalization of the wave function is guaranteed for a fault-tolerant quantum computer. However, this term can be important on a noisy quantum computer where the purity of the wave function can be contaminated. Note the expectation value in Eq. (6.25) has to be estimated by collecting samples on a quantum computer.

The quantum bootstrap embedding algorithm with quadratic penalty method is presented below in Alg. 9. The algorithm takes as its input the total Hamiltonian of the original system, and then perform the fragmentation and parameter initialization, followed by the main optimization loop to achieve the matching. Finally, it returns the optimized BE potential $V_{\mathrm{BE}}^{(A)}$ for any fragment A and the final mismatch $\Delta\rho$. Inside the main loop (line 9 of Alg. 9), the cost function $\mathcal{L}^{(A)}(\lambda)$ for each fragment A is minimized for a fixed penalty parameter λ (line 10 and 11). The penalty λ is then increased geometrically (line 12) until the mismatch criteria is met, i.e., $\Delta\rho \leq \varepsilon$.

Algorithm 9 Quantum bootstrap embedding algorithm: quadratic penalty method

Input: Geometry of the total molecular system and the associated *ab initio* Hamiltonian.

Output: $\left(H^{(A)} + V_{\mathrm{BE}}^{(A)}\right)$ for all A

1: **for** A = 1 to N_{frag} **do** ▷ Initial fragmentation: Divide the full molecular system into N_{frag} overlapping fragments
2: Generate $H^{(A)}$ using Eq. (D.1) of Appendix D.1
3: Set $V_{\mathrm{BE}}^{(A)} = 0$
4: **end for**
5: Set initial penalty factor $\lambda = 1$; set initial mismatch $\Delta\rho > \epsilon$. ▷ Parameter initialization
6: **while** $\Delta\rho > \varepsilon$ **do** ▷ Main loop
7: **for** A = 1 to N_{frag} **do**
8: Minimize $\mathcal{L}^{(A)}(\lambda)$ as in Eq.(6.25): Repeatedly generate $V_{\mathrm{BE}}^{(A)}$ and estimate the penalty loss function $\mathcal{L}^{(A)}(\lambda)$ using `SWAP` test.
9: **end for**
10: $\lambda \leftarrow \gamma\lambda$, for some fixed $\gamma > 1$ ▷ Increase penalty parameter
11: **for** A = 1, N_{frag} **do** ▷ Update mismatch
12: Estimate $\mathcal{Q}_{quad}(\rho_r^{(A)}; \rho_r^{(B)})$ using $N_{samp}^{\texttt{SWAP}}$ (Eq. (6.27)) samples for each `SWAP` test.
13: **end for**
14: Classically compute the mismatch $\Delta\rho$ using Eq. (6.24).
15: **end while**
16: **return** $\left(H^{(A)} + V_{\mathrm{BE}}^{(A)}\right)$ for all A, $\Delta\rho$.

A key step of the algorithm is the minimization of $\mathcal{L}^{(A)}(\lambda)$ at line 11, which consists of

repeatedly generating the BE potential $V_{\mathrm{BE}}^{(A)}$ and estimate the mismatch using SWAP test. BE potentials $V_{\mathrm{BE}}^{(A)}$ are generated differently for different optimization algorithms. In our implementation, a quasi-Newton method, the L-BFGS-B [Byr+95] algorithm, is used at line 11 for minimizing $\mathcal{L}^{(A)}(\lambda)$, where $V_{\mathrm{BE}}^{(A)}$ is proposed by the optimizer in order to estimate the inverse Hessian matrix to steer the optimization properly. Alternatively, if derivative-free methods such as Nelder-Mead [NM65] is used, $V_{\mathrm{BE}}^{(A)}$ will be generated in a high-dimensional simplex defined by the coefficients $\{v_\alpha\}$ in Eq. (6.22), which is repeatedly refined.

Once $V_{\mathrm{BE}}^{(A)}$ is generated, the first term in the cost function in Eq. (6.25) is estimated by invoking the quantum eigensolver for the Hamiltonian $\left(H^{(A)} + V_{\mathrm{BE}}^{(A)}\right)$. The second term, the mismatch in Eq. (6.25) can be estimated by measurement outcomes of the ancilla qubit in the SWAP test (Sec. 6.3). The mismatch estimation at line 13 is performed in the same way as those in line 11. Note that the number of samples $N_{samp}^{\texttt{SWAP}}$ (Eq. (6.27)) for the SWAP test estimation can be changed adaptively in different BE iterations for different accuracy, which we discuss in detail in the next section.

6.4.2 Eigensolver Subroutines and Sampling Complexity

Two major quantum eigensolvers, QPE [SHF13] and VQE [Til+22] can be used in line 11 and 14 of Alg. 9 to estimate the cost function. QPE is an exact eigensolver, where the system wave function collapses to the exact ground state regardless of the number of evaluation qubits used. In contrast to QPE, VQE is an approximate eigensolver and the results depends on the choice of ansatz and the optimization algorithm used.

A crucial feature of a quantum eigensolver is its probabilistic nature, in a sense that any measurement collapses the entire quantum state. This perspective allows us to treat a quantum eigensolver as a sign-problem-free sampling oracle for correlated electronic structure problems where Ref. [Hug+22] provides a concrete example.

The stochastic nature also means a more careful treatment on the number of samples is required to fully quantify any potential quantum speedup. In general, for typical iterative mixed quantum-classical algorithms, some parameters are usually passed from one iteration to the next, where the parameters are estimated by repeatedly sampling from a quantum eigensolver oracle through proper measurement. This means the uncertainty on these parameters estimated from one iteration has to be small enough to avoid a divergence of the algorithm as iteration continues.

In particular in the bootstrap embedding case, the sampling accuracy on the fragment overlap of each iteration has to be good enough such that the uncertainty of the mismatch passed to the next iteration will not spoil the iteration and lead to diverging results as iterations continue. In the following, sampling complexities of classical matching and SWAP-test-based quantum matching are compared.

When estimating the overlap S to an accuracy ϵ naively by density matrix tomography (TMG) of individual RDM elements, it is shown under mild assumptions that the total number of samples required (Sec. D.6 of SI)

$$N_{samp}^{\mathrm{TMG}}(S, \epsilon, n) = \mathcal{O}(e^n)\left(\frac{D}{\epsilon^2}\right), \tag{6.26}$$

where n is the number of qubits on the overlapping region, and D is a system-dependent constant as a function of the two RDMs. In contrast, the quantum matching based on SWAP test costs

$$N_{samp}^{\texttt{SWAP}}(S,\epsilon) = \left(\frac{1-S^2}{8}\right)\frac{1}{\epsilon^2}, \tag{6.27}$$

which is independent of the size n of the overlapping region of two fragments. This demonstrates that our quadratic quantum matching achieves an exponential speedup compared to naive tomography of density matrices. This dramatic speedup is perhaps not that surprising because we only care about one particular observable (the overlap) instead of the full subsystem RDMs. Therefore, if the observable can be mapped to measurement outcome of few qubits by some quantum operations (SWAP test in this case), advantages are expected in general.

Moreover, the dependence of $N_{samp}^{\texttt{SWAP}}(S,\epsilon)$ on the overlap S and estimation accuracy ϵ allows an adaptive sampling schedule to be implemented for line 11 and 14 of Alg. 9. For example, we may use the overlap S estimated from the previous BE iteration to compute the required $N_{samp}^{\texttt{SWAP}}$ in the current BE iteration. The accuracy ϵ can also be dynamically tuned according to the error of the first term in Eq. (6.25), as well as the value of the penalty parameter λ. For example, at the beginning BE iterations, the mismatch ($\Delta\rho$ or more precisely $\mathcal{Q}_{quad}(\rho_r^{(A)};\rho_r^{(B)})$) is large so that a moderate ϵ suffices. As the BE iteration proceeds, the overlap converges exponentially, therefore an exponentially decreasing ϵ has to be used as well. A numerical value of ϵ needs be determined from case to case.

In addition, Eq. (6.27) suggests an interesting behavior. As the QBE algorithm proceeds and the overlap S increases, fewer samples are needed to achieve a target accuracy. If S approaches 1 exponentially fast as $S \sim 1 - e^{-\gamma\cdot n_{\text{iter}}}$ for some constant γ, then the required number of samples for SWAP will decrease exponentially as BE iteration n_{iter} goes $N_{samp}^{\texttt{SWAP}} \sim e^{-\gamma\cdot n_{\text{iter}}}/\epsilon^2$. In practice, the overlap of two subsystem can never approach 1 but saturates to a constant $0 < c < 1$ when matching is achieved, and therefore $N_{samp}^{\texttt{SWAP}} \sim (1-c)/\epsilon^2$ still obeys the $1/\epsilon^2$ scaling generally. This, on the other hand, suggests that a larger overlapping region is advantageous to reduce $N_{samp}^{\texttt{SWAP}}$ because the RDM of a larger subsystem of a pure state will have greater purity (hence larger c) in general.

6.4.3 Additional Quadratic Speedup

The core of many quantum speedups over classical algorithms lie at the ability of quantum computers to directly manipulate the *probability amplitude* instead of probability itself, while classical computers only have access to probability. With this idea, the above perspective of treating a quantum eigensolver as an oracle where some amplitude is estimated through proper measurements allows us to achieve an additional quadratic speedup in our quantum bootstrap embedding algorithm. This section compares two different versions of quantum matching algorithms in QBE, the SWAP and the SWAP+AE algorithms. However, the same argument of quadratic speedup applies to classical sampling based eigensolvers such as VMC as discussed in detail at the end of this section.

The intuition is that instead of directly measuring a small quantum amplitude to accumulate enough counts to reduce the error bar, we may use quantum algorithms to first

amplify the amplitude before the measurement. One way of understanding this is that Eq. (6.27) contains an overlap-dependent prefactor $(1 - S^2)$ as discussed above. If the overlap S (as a probability amplitude) can be manipulated on the quantum computer *easily* such that $(1 - S^2)$ is on the order of ϵ, then $N_{samp}^{\texttt{SWAP}}$ will be proportional to only $1/\epsilon$ instead of $1/\epsilon^2$. There are well-established ways of performing such amplitude amplification task via coherent quantum algorithms [Bra+02]. See SI Sec. D.7 for the construction of the amplitude amplification and binary search quantum algorithm.

In particular, in each iteration of the algorithm, it can be shown that by combining oblivious amplitude amplification and a binary search protocol, estimating the overlap up to precision ϵ between adjacent fragments takes $N_{samp}^{\texttt{SWAP}+AE}$ samples (state preparation and `SWAP` tests)

$$N_{samp}^{\texttt{SWAP}+\text{AE}} = \frac{\sqrt{2}}{2\ln(2)\epsilon} \ln^2(\frac{1}{\epsilon}), \tag{6.28}$$

regardless of the overlap S.

Comparing (6.28) with (6.27), the above analysis suggests that our coherent quantum matching algorithm achieves a quadratic speed up (up to a factor of polylog$(\frac{1}{\epsilon})$) as compared to the `SWAP` test based quantum matching algorithm, which is consistent with typical behavior of a Grover-type of search algorithm. Moreover, in contrast to (6.26), an exponential advantage is present with respect to the size of the overlapping region, indicating the benefit of using our quadratic QBE algorithm for fragment matching in the presence of large overlapping region.

In the above, we leverage amplitude amplification to achieve a quadratic speedup of a quantum subroutine based on `SWAP` test. More generally, such amplitude amplification technique can be utilized to achieve a general quadratic speedup in the required number of samples for any Monte Carlo classical algorithms [WA08; YA12; Mon15]. This can be understood by realizing that classical probability distributions may be encoded in the amplitudes of the quantum state of a quantum computer, where measurements performed after some unitary quantum computation is similar to *sample* from the quantum computer to extract the probability distributions. When treating the unitary quantum computation part as a quantum blackbox, it is then easier to understand the quadratic speedup in the number of samples as compared to classical Monte Carlo methods. In our case, the quantum blackbox is the quantum eigensolver used to find the ground state for each fragment, while the classical blackbox is the stochastic classical eigensolvers such as VMC.

6.5 Results and Discussions

With the theoretical foundation and algorithms discussed in previous sections, we present numerical results in this section using a typical benchmark system in quantum chemistry, hydrogen chains under minimal basis. In Sec. 6.5.1, we demonstrate the convergence of the QBE algorithm with an exact solver (at infinite sampling limit) using an H_8 molecule with STO-3G basis. In Sec. 6.5.2, we present numerical evidence for the sampling advantages of the QBE algorithm in terms of overlapping fragment size (non-interacting H_4 molecule

with STO-3G basis) and target precision over incoherent estimation and classical VMC sampling (H_8 molecule under STO-3G basis). Numerical results using *approximate* variational quantum eigensolvers (VQE) on a random spin model and a perturbed H_4 molecule are documented in Sec. D.9.5 of SI for interest readers, where a similar BE convergence is established at the beginning iterations but later plateaus, likely due to intrinsic VQE ansatz truncation errors. A detailed discussion of BE+VQE goes beyond the scope of this work which we leave for future investigation.

6.5.1 Convergence of QBE in Infinite Sampling Limit

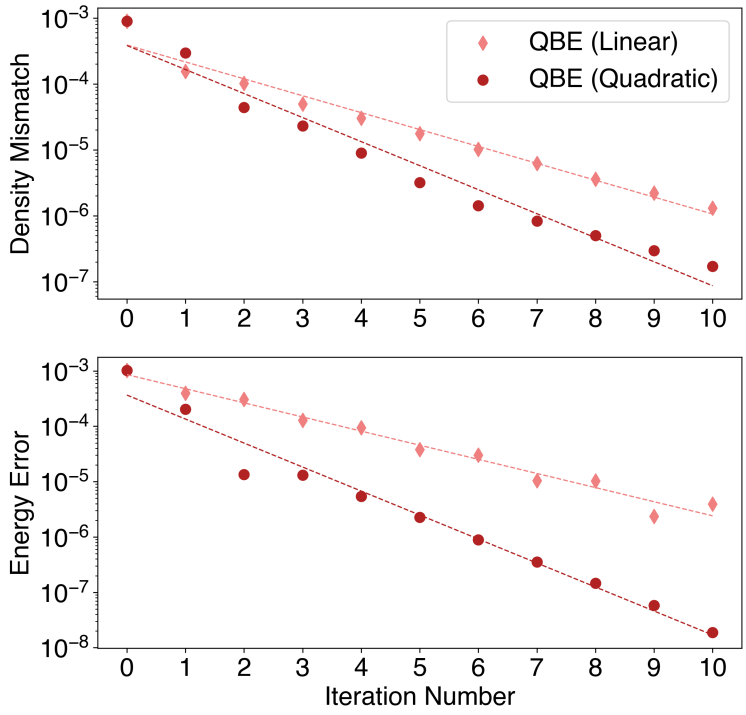


Figure 6.5: Convergence of the quantum bootstrap embedding algorithms on (a) density mismatch and (b) energy error for the linear constraint (pink) and quadratic penalty method (red) in the infinite sample limit for an H_8 molecule. The dashed trend lines in both panels indicate an exponential fit.

We focus on demonstrating the convergence of QBE in the infinite sampling limit by using exact deterministic solver with the quadratic constraint in Eq. (6.20) and linear constraint in Eq. (6.16). As a standard benchmark system for electronic structure, we perform QBE on a H_8 chain under a minimal STO-3G basis, which is fragmented into six overlapping fragments each with six embedding orbitals. Fig. 6.5a shows the exponential convergence of the density mismatch for an H_8 molecule in both linear and quadratic constraint cases. This convergence behavior of QBE matches the convergence of classical BE in Fig. 6.2 with exact classical solver (FCI), demonstrating the correctness of the new constraints. The agreement on the convergence with classical BE in Fig. 6.2 is expected since at infinite sampling limit,

the outer iterations in both classical and quantum BE are the same classical optimization routine.

To quantify how much energy error the final converged result has, Fig. 6.5b shows the absolute value of the error in energy using the energy in the last (11^{th}) iteration as a reference. We can see that the energy errors from both the linear and quadratic constraint algorithm exhibit similar exponential convergence as the density mismatch. Moreover, the energy in both cases converge to the same value within 10^{-6} in the last iteration (not shown in the figure). We note that the linear constraint case shows a slightly oscillatory convergence, while the quadratic case is free of such oscillatory behavior. The fact that quadratic appears to converge slightly faster than linear may be coincidence for the system investigated, and the convergence rate in general depends on the optimization algorithm chosen. See Sec. D.9.4 of the SI for a detailed description on definition of the energy.

6.5.2 Sampling Advantage of Coherent Quantum Matching

In the previous section, we have seen that our quantum bootstrap embedding algorithm convergence as expected in the infinite sampling limit. It is also seen (in the SI) that the approximate VQE leads to biased behavior on the density matching. In practice, only a finite number of samples can be collected on a quantum computer, and we will focus on theis scenario in this section. In particular, we present numerical data demonstrating the sampling advantage of our coherent quantum matching algorithm. Sec. 6.5.2 discusses the sampling advantage of the quantum matching algorithm for an overlapping region of increasing size, echoing the analytical sampling complexity derived in Sec. 6.4.2. In Sec. 6.5.2, the additional quadratic speedup in estimating the overlap via amplitude amplification and binary search (AE) is presented, which agrees with the theoretical sampling complexity in Sec. 6.4.3.

Advantage in Fragment Overlap Size

To perform bootstrap embedding, it is usually advantageous to partition the system into fragments with large overlapping region to increase the convergence rate, because a large overlapping region necessarily means more information is provided to update the local potential for the following BE iteration. However, as is seen in Eq. (6.26) of Sec. 6.4.2, a larger overlapping size also lead to a potentially exponentially higher sampling complexity versus the number of qubits in the overlapping region if estimating the overlap naively from density matrix tomography (TMG). The quantum matching algorithm implemented by a `SWAP` test (Fig. 6.1iiiq) bypass the need for density matrix tomography, and therefore leads to a sample complexity as in Eq. (6.27) independent of the size of the overlapping region.

To validate our theoretical sample complexity, a simulation of the quantum matching algorithm with QPE as an eigensolver for two identical H_4 chain is performed using a noiseless Qiskit AerSimulator (see SI Sec. D.9.3 for more details) for an increasing overlap region ranging from 2 to 4, 6, and 8 qubits (schematic in Fig. 6.6). In the simulation, we first use QPE to prepare the ground state for two non-interacting H_4 molecules separately. A `SWAP` test is then performed on relevant qubits in the overlapping region between the two H_4 molecules. The evaluation qubits for QPE and the ancilla qubit for `SWAP` test are all measured afterwards. Post-selection on the QPE evaluation qubits are performed in order to

select the ground states of H_4 molecules. The SWAP test results are processed and converted to the estimation on the overlap S.

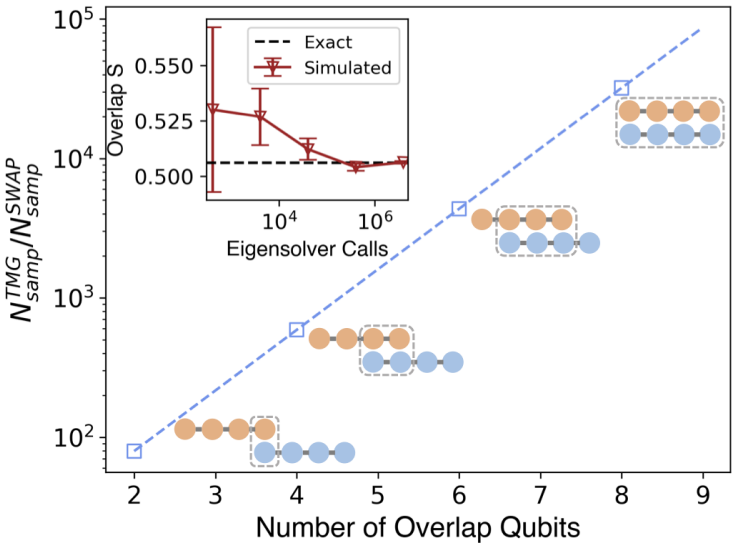


Figure 6.6: Sampling complexity ratio of naive density matrix tomography (TMG) and SWAP test versus number of qubits in the overlapping region for a target precision $\epsilon = 0.001$ on overlap S. The inset shows a simulated convergence of overlap (S) estimation using quantum matching (SWAP) for the case of two overlapping qubits. Data are obtained from a non-interacting chain of H_4 (see SI Sec. D.9.3 for details).

The inset of Fig. 6.6 shows the estimated overlap S as a function of sample size (number of eigensolver calls) in the case of two overlap qubits. The estimated overlap converges to the exact value (black dashed horizontal line) for roughly four million samples within 5×10^{-4} (error bar invisible for the last data point). This demonstrates the correctness of our quantum matching algorithm.

By repeating similar estimation as described above for increasingly larger overlapping regions, the exponential sampling advantage of the quantum matching algorithm over naive density matrix tomography is evident in Fig. 6.6. As we can see, to achieve a constant target precision of $\epsilon = 0.001$ on the overlap S, the ratio between the SWAP test estimation and the naive tomography estimation for the required number of eigensolver calls increases exponentially as the number of qubits.

We note that in general, overlaps between density matrices are not low-rank observables, so the sampling complexity of estimating it is likely to be high. However, more efficient sampling schemes may exist than the naive density matrix tomography as presented in Eq. (6.26). For example, by sampling the differences in the RDMs between the current and the previous BE iterations, the sampling complexity could be much better than exponential. We leave this for future investigation.

Additional Quadratic Speedup in Accuracy

We have seen in the previous section that the quantum matching implemented by a SWAP test shows a potentially exponential sampling advantage in terms of the size of the overlapping region as compared to naive density matrix tomography (compare Eqs. (6.27) to (6.26)). However, the sample complexity in the estimation accuracy ϵ follows the same scaling of $1/\epsilon^2$ as classical sampling based eigensolvers such as VMC. As is derived in Sec. 6.4.3, we see that the sample complexity can be reduced to roughly $1/\epsilon$ with a coherent quantum matching algorithm (SWAP+AE), by combining amplitude estimation and a binary search protocol, thus achieving a quadratic speedup. In this section, we present concrete numerical data demonstrating this quadratic speedup.

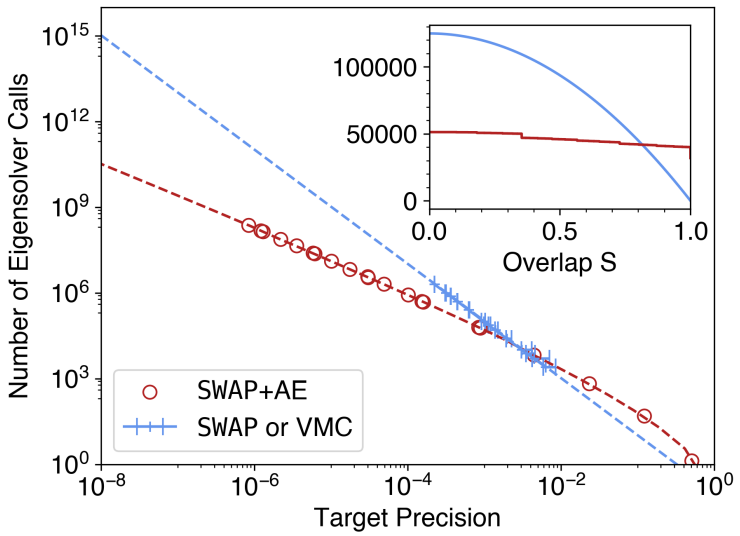


Figure 6.7: Number of eigensolver calls required as a function of target precision at overlap $S = 0.4$, comparing SWAP or VMC (blue) and SWAP+AE (red) estimation for the H_8 chain with STO-3G basis. The blue dashed line shows the number of samples (eigensolver calls) needed in SWAP test as derived in Eq. (6.27), while the red dashed line plots a more accurate version of Eq. (6.28) (Sec. D.7.3 of SI) with red circles highlighting a few data points spanning low to high target precisions. The blue scatter points are the number of VMC eigensolver calls required to achieve the corresponding target precision on the 1-RDM overlap estimation for the same H_8 molecule. The inset plots the number of eigensolver calls as a function of the overlap S for a fixed target precision $\epsilon = 0.001$. Note the crossover in both plots.

Fig. 6.7 shows that for a single BE iteration, the required number of samples (eigensolver calls) on estimating the RDM overlap S between two adjacent fragments as a function of the required precision on the overlap, comparing the SWAP test based quantum matching (blue) and the coherent overlap estimation combining the SWAP test and amplitude estimation (SWAP+AE) (red). We can see that the required number of samples increases quadratically as the accuracy ϵ increases for the SWAP test based estimation. In contrast, the slope of the SWAP+AE sample complexity is reduced to roughly half of the SWAP test, demonstrating the quadratic speedup.

To compare the classical VMC sampling convergence with the quantum overlap estimation method, we also overlay the number of VMC eigensolver calls (blue marks) versus target precision on estimating the overlap on top of the SWAP test sampling complexity for the same H_8 molecule. The general agreement between the VMC eigensolver calls and the derived SWAP test eigensolver calls highlights the similarity of a classical stochastic electronic structure methods and a quantum incoherent matching algorithm in terms of blackbox sampling complexity, echoing the idea of treating quantum computers as coherent sampling machines. It is worthwhile noting that this quadratic speedup is only advantageous in the high precision (small ϵ) limit, as is evident from the existence of a crossing point in Fig. 6.7 (between 10^{-4} and 10^{-2}), which defines a critical ϵ^*. For $\epsilon < \epsilon^*$, SWAP+AE is favored whereas the SWAP test wins when $\epsilon > \epsilon^*$.

Moreover, in addition to the dependence on estimation accuracy ϵ, the sampling complexity also depends on the value of the overlap S. The inset of Fig. 6.7 compares the number of eigensolver calls using SWAP (blue) and the SWAP+AE estimation (red) for estimating the overlap during quantum matching. In more detail, the sample complexity for the SWAP test decreases quadratically as the overlap S approaches 1 (Eq. (6.27)). As a comparison, the SWAP+AE stays roughly a constant for the coherent quantum matching ((6.28)), because the amplitude amplification process used in the present work is agnostic to the value of the amplitude (overlap S), i.e., oblivious amplitude amplification [YLC14; Ber+14]. The slight drop in sample complexity in the SWAP+AE approach (red line, inset of Fig. 6.7) is due to the discrete bit representation of S (Sec. D.7.2 of SI). The different scaling on S between these two algorithms leads to a crossover of the sampling complexity at roughly $S = 0.8$ for a target precision of $\epsilon = 0.001$. This crossover suggests that the plain SWAP test is advantageous for large overlaps, while amplitude estimation works better for small overlaps S.

In addition, as mentioned in the previous section, as the bootstrap embedding iteration proceeds, the exponential convergence of the density mismatch (overlap S) suggests the need for an exponentially increasing accuracy ϵ on the overlap estimation. This further means the number of samples per iteration in the SWAP test should increases exponentially as the the number of iterations. Similarly, SWAP+AE achieves a square-root speedup in the total sample numbers (remains exponential). We note that there may exist ways of sampling the overlap in the current BE iteration *normalized* by the previous BE iteration to accelerate this requirement on a large number of samples, which we leave for future investigation.

6.6 Conclusion

In conclusion, we have developed a general quantum bootstrap embedding method to find the ground state of large electronic structure problems on a quantum computer by taking advantage of quantum algorithms. We formulated the original electronic structure problem as a optimization problem using a quadratic penalty to impose matching condition of adjacent fragments. A coherent quantum matching algorithm based on the SWAP test achieves efficient matching with an exponential sampling advantage compared to naive RDM tomography. By estimating the amplitude that encodes the overlap information combing an amplitude amplification and binary search protocol, an additional quadratic speedup is achieved. In

addition, an adaptive sampling scheme is used based on previous overlap information and the desired target accuracy to improve the sampling efficiency.

We demonstrate the performance of the QBE algorithm using a linear hydrogen molecule under minimal basis. Our QBE algorithm is shown to achieve exponential convergence in density mismatch and energy error similar to classical bootstrap embedding. However, instead of the exponential cost of an exact classical solver (full configuration interaction), quantum eigensolvers such as quantum phase estimation can solve the fragment electronic structure exactly without incurring the exponential cost. Approximate quantum eigensolvers (QES) are likely to achieve exponential speedup compared to FCI. However, such exponential speedup depends on detailed implementation and the easy of input state preparation.

We have also compared sampling advantage of different versions of quantum matching algorithms over classical BE+VMC 1-RDM matching for achieving the same accuracy, where QBE+TMG (full RDM matching) is potentially exponentially slower than classical BE+VMC (1-RDM matching) because the exponentially large number of full RDM elements to estimate (Sec. 6.4.2 and 6.5.2). QBE+`SWAP`+AE achieves quadratic speedup as compared to classical BE+VMC and QBE+`SWAP` (Sec. 6.4.3 and 6.5.2). Different choices of quantum eigensolvers and matching algorithms are summarized in the flow chart in Fig. 6.8, where accuracy and speedups are labeled for each method.

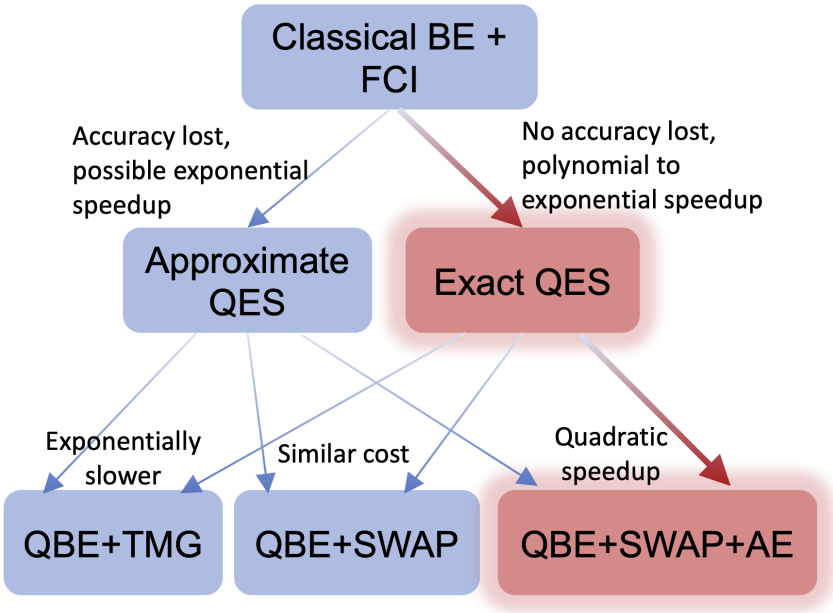


Figure 6.8: Summary of different choices of quantum eigensolvers (QES) and matching algorithms discussed in the present work, with speedup and cost labeled on each arrow accordingly. Overall, the best algorithm (QBE+SWAP+AE with exact QES) is highlighted in red. Note that approximate QES are likely to achieve exponential speedup as compared to classical FCI solver. It is however not guaranteed and depends on specific implementation and the ease of input state preparation. We therefore use "possible exponential speedup" for it.

While we have made progress toward solving electronic structure problems employing quantum resources in bootstrap embedding, there are several open questions to explore in the future. One immediate task is to perform more thorough benchmark comparing different versions of QBE and classical BE algorithms in terms of both speedup and accuracy quantitatively. Beyond benchmark, at the algorithmic level, it is important to reconstruct [QR21; NIB22] the total system density matrices from subsystem ones in order to compute observables other than the energy. Ideally, quantum algorithms that can perform the reconstruction process would be desired. Moreover, we have established how the bootstrap embedding potential can affect the system energy including the excited states in Eq. (6.23). Future works on developing a QBE algorithm targeting excited states [Mit+21] or finite temperature electronic structures [Zha99; LCR18; Sun+20b] would be of great interest. Alternative constraint optimization methods such as the augmented Lagrangian method can also be explored to achieve potentially better convergence [Fau+22].

In addition, the idea of quantum matching proposed in the present work could also be exploited further in other embedding theories to harness quantum computers and resources, including but not limited to embedding schemes based on wave functions, density matrices, and Green's functions [SC16]. In these contexts, it is likely that more sophisticated quantum primitives and algorithms could accomplish quantum matching more efficiently than the simple `SWAP` test we employ. For example, it is possible that higher order matching, or matching of derivatives, could be accomplished quantum-mechanically, thus side-stepping sampling noise.

More broadly, these quantum embedding theories and algorithms enabled by quantum computation resources open new possibilities in chemistry, physics, and quantum information. In the near term, molecules with more complex valence electronic structures such as polyacetylene or polyacene chains beyond minimal basis can be treated with QBE on current noisy quantum computers with a few hundred qubits. In the longer term, large molecular systems in catalysis [FKB19; ZLY12] and protein-ligand binding complexes [War14; Pro+20] likely can be simulated at a much higher accuracy by combining state-of-the-art quantum and classical computational resources in embedding properly. In condensed matter and material science, quantum bootstrap embedding may be adapted to periodic systems [PHG19; Rus+18; Chi+16] for quantum material design [Hea+20] and probing phase diagrams of various lattice models [Qin+22] close to the thermodynamic limit.

Finally, from a viewpoint of quantum information, the concept of embedding is closely related to entanglement. Understanding the connection between the performance of quantum embedding algorithms and fragment-bath entanglement entropy may provide a general way to describe and understand the complexity of chemical and physical problems from a quantum information perspective [Din+20; DS20; Wil13]. Current quantum computers are small – we believe our quantum bootstrap embedding method provides a general strategy to use multiple small quantum machines to solve large problems in chemistry and beyond [Har20; Son+21]. We look forward to future development in these directions.

Chapter 7

Conclusion and Outlook

In this thesis, we set out to tackle one of the biggest challenges of the scalibility of quantum computers, learning of quantum systems during device characterization and execution of quantum algorithms. We showcased how resource requirements of learning tasks across the life cycle of quantum computing could be reduced by exploiting available prior information. This prior information was presented to us in different ways: prior knowledge on the structure or class of quantum states (Chap. 2) or channels (Chap. 3), locality of Hamiltonians, prior knowledge of system parameters (Chap. 4), and structure in addition to sparsity of target observables in a given basis (Chap. 5,6). Of course, the question remains: How much more can we do with prior information in reducing resource requirements of learning tasks across the life cycle of quantum computing?

I would like to conclude with an outlook for some of the learning tasks considered in this thesis. I state questions for different learning tasks considering a particular type of prior information being available i.e., structure of the quantum system being learned, and how we might accelerate learning of different quantum systems to achieve Heisenberg limited scaling. Furthermore, as the size of quantum devices increase, we can expect new architecture designs to be proposed. It then becomes relevant to consider how current learning methods for device characterization may be adapted to such architectures and the possibility of new approaches. I thus state some questions that arise in the context of characterizating modular quantum devices.

Role of structure in learning and testing. In the context of learning quantum states, stabilizer states are notable as they are known to be efficiently learnable in $O(n)$ samples and $O(n^3)$ computational time complexity [Mon17a]. Additionally, in the PAC learning model introduced by Aaranson [Aar07], stabilizer states can be efficiently learned in $O(n^4)$ computational complexity [Roc18]. In Chap. 2 of this thesis, we identified quantum phase states as an another class of states that can be learned efficiently in the exact learning model but notably include states that are hard to simulate classically. This raises the question:

Can we learn quantum phase states in the PAC learning model in $O(poly(n))$ computational complexity?

Further, what can we say regarding the learnability of other structured quantum states

e.g., low-rank stabilizer decompositions [Bra+19] in the exact and PAC learning models?

In many situations, prior information is not always available but we have now at hand a set of efficient learning algorithms given prior information. We thus can choose to *test* if our learning algorithm is applicable. For example, in the context of quantum state tomography, given access to copies of an unknown quantum state but with no prior information, we can test if these unknown quantum states satisfy a property (e.g., being stabilizer states) which we can then exploit during learning (e.g., by utilizing learning algorithms applicable for unknown stabilizer states). This constitutes the question of property testing. The classes of stabilizer states [GNW21] and matrix product states [SW22] are known classes of quantum states that can be tested efficiently in sample complexity growing at most polynomially in n. This motivates the following question.

Can we test if a given unknown quantum state $|\psi\rangle$ is a degree-d quantum phase state or ϵ-far away? What is the corresponding sample complexity?

It should be noted that there exists a separation in the sample complexity of learning and testing stabilizer states. While, learning can be accomplished in $O(n)$ samples, the testing algorithm only requires 6 copies [GNW21]. Could there then be classes of quantum states that are efficiently testable but not efficiently learnable?

Achieving Heisenberg limited scaling. The last few years have been exciting for Hamiltonian learners with new approaches [Fra+22] and results on efficient learning algorithms for k-local Hamiltonians [Yu+23; Hua+23; Bak+23; HKT22]. It was shown that one can learn k-local Hamiltonians at the Heisenberg limited scaling [Hua+23] when single-qubit control is allowed. In general, however, it was shown that Heisenberg limited learning cannot be achieved if control is not available [DOS23].

Another physically relevant class of quantum Hamiltonians is that of sparse Hamiltonians. While we have learning algorithms [Yu+23] that can learn s-sparse Hamiltonians in $O(sn)$ sample complexity, the total interaction time with the quantum system over these queries, scales as the standard quantum limit or worse. This raises the question:

Can we learn s-sparse Hamiltonians at the Heisenberg limited scaling while retaining a query complexity with a linear dependence on n and s?

Of particular interest is then when the Hamiltonian dynamics $U = \exp(-iHt)$ acts on a hidden quantum system as studied in Chap. 3 for the problem of channel discrimination. In this case, not much is known about the classes of Hamiltonians that are efficiently learnable.

Can we learn s-sparse Hamiltonians and k-local Hamiltonians on hidden quntum systems efficiently with query complexity $O(\text{poly}(n))$ and at the Heisenberg limited scaling?

Answering the above question would have interesting implications for learning dynamics of systems undergoing chemical reactions or involving collisions without disrupting the system dynamics severely.

Characterization of modular quantum devices Recently, there have been many proposals to scale the size of quantum devices through modular or distributed designs. In such architectures, multiple small quantum devices (or nodes) are linked together to create one large multinode quantum computer (MNQC) [Ang+22]. The interlinks between different nodes may be classical in the very beginning but are eventually be expected to be quantum interlinks. It then becomes relevant to consider how these MNQCs may be characterized and calibrated. One approach is to apply current learning methods for device characterization and perhaps exploit prior information on connectivity and sparsity in the MNQC. The quantum subsystems or individual quantum nodes forming the MNQC would then be calibrated individually and then the internode quantum gates calibrated.

However, there is potentially another alternative approach that may require less resources and be easier to scale. If we can dare to imagine simulating quantum systems with quantum devices, we might as well dare to consider learning quantum systems with other quantum devices.

Could we calibrate smaller quantum subsystems which constitute the MNQC using a fully calibrated quantum computer which is also a part of the MNQC?

Just as we had introduced an ancillary measurement system to learn hidden quantum channels on a physical system with limited control in Chap. 3, we could now consider using a fully calibrated quantum node to calibrate quantum nodes through weak internode interactions. This is distinct from the setting of [Wie+14a; Wie+14b] which would rely on using noisy `SWAP` gates between the calibrated and uncalibrated quantum nodes. Instead, as in Chap. 3, we could use protocols inspired from quantum signal processing for learning the Hamiltonian on the uncalibrated device or even learn how to drive the Hamiltonian on the uncalibrated quantum node to implement a desired multi-qubit gate and in a more principled manner. Quantum signal processing has been highly successful in algorithm design [Mar+21b; Gil+19] and there seems to be hints that it may be useful for even learning of quantum systems [Sug+23; DGN22].

Overall, the future seems exciting for new larger quantum computers as well as new principled approaches for learning such systems.

Appendix A

Appendix to Chapter 3

In Section A.1, we describe prior work on protocols designed for accomplishing conventional quantum channel discrimination. In Section A.2, we give motivation for the query used in our discrimination protocols for solving the HBCD problem. In Section A.3, we give details on the numerical experiments on assessing the performance of the sequential and multi-shot protocols on different HBCD problems.

A.1 Perspective on measurement protocols for conventional QCD

The field of QCD has witnessed a plethora of studies aimed at exploring various aspects of this important area. We refer to Ref.[BMQ21] and references therein for the overview of progress in QCD. Here, we summarize perspectives provided by prior work that study the performance of different measurement protocols employed in conventional QCD. Comparison between different strategies has also been studied in quantum channel estimation (QCE), which is also referred to as quantum metrology. Therefore, we connect to studies in both fields in this section of the appendix.

Among many measurement protocols, the most relevant protocols for our study are entanglement-free protocols [DFY07; Bra+18; Hig+07a; Reh+18; RS21; KK22; DeB+23]. These protocols have garnered interest due to their simplicity in preparing probe states, which offers practical advantages in experimental implementations [Hig+07a; DeB+23]. In these studies, entanglement-free sequential protocols were shown to perform as well as entanglement-assisted protocols in QCD [DFY07; Reh+18; DeB+23] and in QCE [Hig+07a; KK22; RS21]. However, there is no study that has shown entanglement-free protocols can outperform entanglement-assisted ones.

More generally, hierarchies among protocols have also been studied under the assumption that the use of any operation and any state is allowed for QCD [CDP08; Har+10; BMQ21] and for QCE [GLM11b; KD13; DM14]. In QCE, it is shown that the maximum performance of entanglement-free protocols is equivalent to that of entanglement-assisted protocols for unitary channels, and is conjectured to be no better than those for general channels [DM14]. In QCD, a hierarchy among entanglement-assisted protocols has been shown [BMQ21]. However, no study showing an advantage given by entanglement-free protocols has been reported

in these studies either.

Prior studies suggest that despite the ease of probe state preparation and the potential merits of entanglement-free protocols offer in terms of experimental simplicity, it is unknown whether they exhibit an advantage over entangled protocols in the context of quantum channel discrimination and channel estimation. This observation raises interesting questions regarding the limitations and practical implications of entanglement-free protocols, which warrant further investigation to gain deeper insights into their potential optimization and applicability in quantum information processing tasks.

A.2 Relation to Quantum Signal Processing

The query (Def. 4) as used in our discrimination protocols to solve HBCD is inspired from Quantum Signal Processing (QSP). The qubit in the measurement system $\mathcal{M}$ (see Fig. 3.2) can be viewed as the ancilla qubit typically used in QSP [LC17] with the rotations of $\exp(i\phi_n\sigma_x)$ for any $n \in [N]$ as the processing operators. The signal operators then involve the application C on $\mathcal{H}$, and the controlled interaction on the composed system of $\mathcal{H}$ and $\mathcal{M}$. Note that the signal operator in this case looks different from that typically used in QSP.

Moreover in connection to QSP, tracing out the hidden system $\mathcal{H}$ of the query results in a Completely Positive Trace-Preserving map with non-Markovian process, which can be interpreted as a noisy channel. In the sequential protocol, the query involves the application of a tunable X-rotation gate and a noisy channel alternately, which has a structure of QSP. However, it is not expected that the number of the query used to solve the HBCD saturates the Heisenberg limit with a noisy channel.

The result changes when partial information in $\mathcal{H}$ is known. In particular, it is noted that although the initial state and rotation angle of C are still unknown, it is known that C is the X-rotation gate and that the interaction between $\mathcal{H}$ and $\mathcal{M}$ is a controlled rotation. In this scenario, the result shows that the Heisenberg limit and perfect discrimination can be achieved using Quantum Signal Processing (QSP). This is an interesting finding as it suggests that even with partial information, QSP can be utilized to achieve highly precise measurements and accurate discrimination.

A.3 Numerical Experiments

In our numerical experiments for solving the problem of HBCD using the sequential protocol or multi-shot protocol, we consider the circuit of Figure A.1, shown with a phase sequence of length K. Let us describe how this compares to the corresponding circuits described in Figure 3.2. Here, $\rho_h = I/2$ and ρ_m are prepared through the action of the single qubit gate of $R_x(\phi_0)$ on the zero state $|0\rangle$. In a simplification from Figure 3.2(a), we consider the phase of the controlled gate to be the same across multiple applications i.e., $\psi_1 = \ldots = \psi_K = \psi$. We then denote the overall phase sequence as $\Phi = \{\phi_0, \phi_1, \ldots, \phi_K, \psi\}$, which now includes ϕ_0. Note that in the main text, we typically denote the length of the phase sequence K as N for the sequential protocol as we measure only once, and as d for the multi-shot protocol

with constant depth queries.

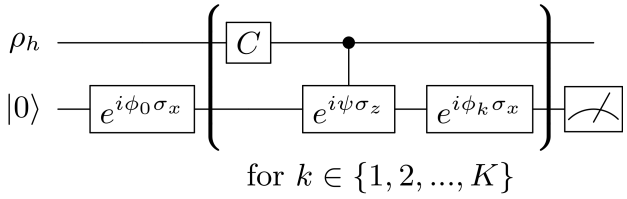


Figure A.1: Quantum circuit used in numerical experiments for HBCD using the sequential or multi-shot protocols

Having described the circuit to be used in our protocols, we are now in a position to describe how a phase sequence Φ is optimized for the problem of HBCD and further details of our experimental procedures for the sequential and multi-shot protocols.

Optimization of phase sequences. Given length K, we obtain a numerically optimized phase sequence $\hat{\Phi}$ by solving the optimization of Eq. 3.41 using the quasi-Newton method of L-BFGS with a particular choice of initial conditions. Denoting the initial condition as Φ^0, we set its first $K+1$ components as

$$\Phi^0_{1:K+1} = \left\{\phi^0_0, \phi^0_1, \ldots, \phi^0_{K-1}, \phi^0_K\right\} = \left\{\frac{\pi}{4}, 0, \ldots, 0, \frac{\pi}{4}\right\}. \tag{A.1}$$

This choice of initial conditions for the phases was inspired by work on optimization of phases in quantum signal processing [Don+21; WDL22], using gradient-based methods. The last component $\Phi^0_{K+2} = \psi^0$ is set randomly by sampling from a normal distribution with zero mean and unit variance. We prepare n_{reps} such initial conditions and then run the L-BFGS algorithm to solve Eq. 3.41. This results in n_{reps} different phase sequence solutions from which we select the one with the lowest loss. In our numerical experiments on HBCD with the sequential and multi-shot protocols, we found that choosing $n_{reps} = 10$, this choice of initial conditions and optimization method yielded solutions at the global minima of the loss function in Eq. 3.41.

Sequential protocol. In our numerical experiments with the sequential protocol, we assume that we only measure once. The goal is to then determine the length N of the phase sequence Φ at which we are able to discriminate $\theta_C = \alpha$ from $\theta_C = 0$. As described in the main text, we do this by starting at $N = 1$ and increasing the value of N by one until we determine a numerically optimized phase sequence Φ (through the optimization procedure described above) that yields a probability of error less than equal to a given error parameter $\epsilon \in [0, 1/2)$ or probability of success greater than equal to $1 - \epsilon$. In Figure A.2, we show how the probability of success varies with N for $\alpha = 0.1$. As illustrated in the figure, the probability of success increases with N, reaching a value of > 0.95 for $N = 16$.

Multi-shot protocol. In our numerical experiments with the multi-shot protocol with constant depth d queries, we fix the length of phase sequence to d and measure m times. The goal is to then determine the minimal number of shots $m^\star$ required to solve HBCD of

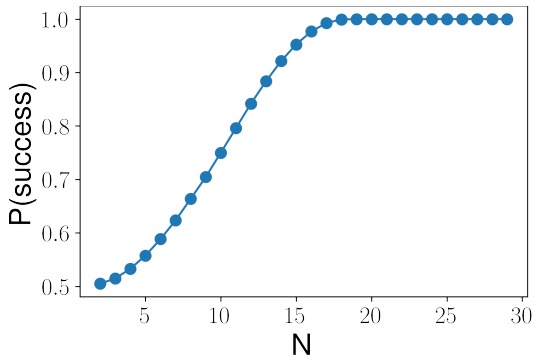


Figure A.2: Probability of success of the sequential protocol in HBCD of $\theta_C = \alpha = 0.1$ from $\theta_C = 0$ with increasing length N of phase sequence Φ.

$\theta_C = \alpha$ from $\theta_C = 0$ with an error probability below ϵ with a numerically optimized phase sequence Φ of length d. This needs to be done empirically as an analytical expression of the error probability over m shots is not available to us. We now describe our experimental procedure for determining the value of $m^\star$ for a given value of α and d. Such experimental procedures are common in the statistical learning community [Lok+18].

For a given value of α, we first determine a numerically optimized phase sequence Φ of length d to solve HBCD using the optimization procedure described earlier. We then generate $L \geq 1$ independent sets (indexed by t) of m measurement outcomes by sampling θ_C^t uniformly from $\{0, \alpha\}$ for each set and using Φ. The likelihood ratio test (LRT) (Eq. 3.5) is then used to determine the estimate $\hat{\theta}_C$ on each of the L sets of measurement outcomes yielding L estimates: $\{\hat{\theta}_C^t\}_{t\in[L]}$. If all the estimates $\hat{\theta}_C^t$ correctly match the corresponding truth θ_C^t for all $t \in \{1, 2, \ldots, L\}$, we say that the multi-shot protocol succeeded in HBCD within a given error probability ϵ.

We now describe how to obtain the value of L given the error probability ϵ, required in our numerical experiments to guarantee that the multi-shot protocol succeeds with a probability above $1 - \epsilon$ with confidence at least 95%. Let the probability of success on any of the L sets i.e., $P(\hat{\theta}_C^t = \theta_C^t)$ be equal to p. Note that the outcome of $\hat{\theta}_C^t$ being equal to θ_C^t through the course of our numerical experiment is then equivalent to generating flips of an unfair coin with the probability of success equal to p. Assuming a uniform initial prior on p, let us denote $P_{post}(p|L)$ as the posterior probability over p after a series of L successful estimations, which is given by the Beta distribution for this Bernoulli process. We then have the probability of confidence p_{conf} in the value of p given successive L_{succ} successful estimations as

$$p_{conf} = \int_{1-\epsilon}^{1} P_{post}(p|L_{succ} = L)dp. \tag{A.2}$$

We require that $p_{conf} \geq 0.95$, which is obtained first for $L = 59$ for $\epsilon = 0.05$, $L = 119$ for $\epsilon = 0.025$, and $L = 598$ for $\epsilon = 0.005$. These values of L were used in all our numerical experiments with the multi-shot protocol in this work.

Appendix B

Appendix to Chapter 4

B.1 Details of Cross-Resonance Hamiltonian

In this section of the Appendix, we firstly give details of the different IBM Quantum devices employed for assessing the performance of the HAL algorithms (HAL-FI and HAL-FIR) proposed in Section 4.3 for learning cross-resonance (CR) Hamiltonians (Eq. 4.31). In Section B.1.2, we discuss how the queries from Section 4.4.2 are implemented in practice on the devices. In Section 4.4.3, we stated a model for the noise source of imperfect pulse-shaping. Here in Section B.1.3, we describe how this model was obtained. This is then followed by relevant analytical expressions of likelihood and Fisher information in Section B.1.4 considering the query space in Section 4.4.2.

B.1.1 Description of IBM Quantum Devices

We consider CR Hamiltonains on the four different IBM Quantum devices described in Section 4.4.2. The connectivity maps of these devices are shown in Figure B.1. We consider CR gates on particular qubit pairs on each device which are summarized in Table B.1. In Table B.1, we describe the properties of each qubit involved in the CR gate including their T_1 or T_2 times and the average infidelity of single-qubit gates.

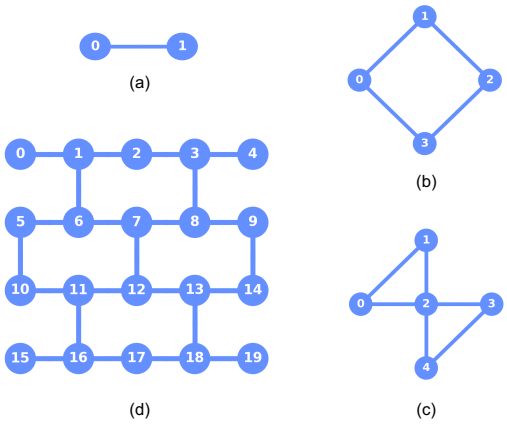


Figure B.1: Connectivity maps for (a) IBM Quantum device A, (b) IBM Quantum device B, (c) IBM Quantum device C, and (d) IBM Quantum device D *ibmq_ boeblingen*. Each node represents a physical qubit on the chip and the presence of an edge between two nodes in the connectivity map indicates that a CR gate can be applied between these two nodes.

Device	**Qubit**		**Qubit Freq. (GHz)**		$\mathbf{T_1}$ (μs)		$\mathbf{T_2}$ (μs)		**Error per gate** ($\times 10^{-4}$)	
	Ctrl	**Targ**	**Ctrl**	**Targ**	**Ctrl**	**Targ**	**Ctrl**	**Targ**	**Ctrl**	**Targ**
A	1	0	5.0593	4.8441	65.4 ± 10.8	32.9 ± 9.5	71.1 ± 5.9	57.7 ± 9.2	5.65 ± 0.13	8.85 ± 0.46
B	3	0	5.1482	4.9273	63.2 ± 11.7	78.1 ± 24.9	73.9 ± 9.9	124.3 ± 21.5	5.92 ± 0.30	6.29 ± 0.23
C	0	1	5.3613	5.2910	34.2 ± 2.5	45.0 ± 19.1	39.2 ± 3.3	63.1 ± 7.6	18.3 ± 1.0	20.6 ± 1.3
	0	2	5.3613	5.2543	34.2 ± 2.5	45.0 ± 19.1	35.8 ± 0.7	52.4 ± 4.2	18.3 ± 1.0	9.23 ± 0.26
	1	2	5.2910	5.2543	39.2 ± 3.3	63.1 ± 37.6	35.8 ± 0.7	52.4 ± 4.2	20.6 ± 1.3	9.23 ± 0.26
D	0	1	5.0466	4.8468	94.0 ± 6.0	75.7 ± 17.0	177.2 ± 44.8	128.1 ± 29.7	2.39 ± 0.12	3.12 ± 0.11

Table B.1: Relevant parameters of IBM Quantum devices. The qubit used as the control or target qubit is indicated by its number in the device connectivity map shown in Fig B.1. Error per gate refers to the average infidelity of single-qubit gates implemented on that qubit.

B.1.2 Experimental Implementation of Query Space

Queries to the CR Hamiltonians between different qubit pairs on the IBM Quantum devices are made through appropriate pulse sequences. These pulse sequences are constructed and executed on the hardware using `Qiskit-Pulse` [Ale+20], which is a pulse programming module within `Qiskit` [Abr+19a] and serves as a front-end implementation of the `OpenPulse` interface [McK+18]. Each `Qiskit-Pulse` [Ale+20] program consists of pulses, channels and instructions. Here, we describe a `Qiskit-Pulse` program and describe how a query to a CR gate on a IBM Quantum device is specified.

A pulse is a time-series of complex-valued amplitudes with maximum unit norm and which we denote as a_k where $k \in [n-1]$ corresponds to the time stamp. The difference between these time-stamps is considered to be dt which is typically the sample rate of the waveform generator. The output signal thus has an amplitude of

$$A_k = \text{Re}\left[e^{i2\pi f k dt + \gamma} a_k\right] \tag{B.1}$$

at time kdt where f and γ are a modulation frequency and phase. A pulse is specified in `Qiskit-Pulse` by specifying the individual amplitudes a_k and the phase ϕ. Alternatively, one can use parametric pulse shapes that are offered by the library such as `Gaussian`, `GaussianSquare`, etc. These pulses are then implemented on the hardware via channels which label signal lines used for transmitting and receiving signals between the control electronics, and the hardware. In particular, these are implemented on the `PulseChannel` and are used to control the system Hamiltonian to implement different gates.

To implement a query to a quantum device, we need an equivalent description of the quantum circuit shown in Fig. 4.1 in the form of a pulse schedule. We discuss this here in parallel with a description of the query space $\mathcal{Q}$ considered for learning CR Hamiltonians. To prepare the initial state, we consider the set of preparation operators $\mathcal{U} = \{\sigma_I \otimes \sigma_I, \sigma_X \otimes \sigma_I\}$ applied to the pure state $|00\rangle$. The single-qubit gates are implemented as a sequence of Gaussian pulses of the appropriate amplitude and duration [Ale+20]. Assuming the first (left) qubit is the control and the second (right) qubit is the target, the effect of the preparation operators is to place the control in $|0\rangle$ and $|1\rangle$ respectively. We evolve the initial state $|\psi(0)\rangle$ for time $t \in \mathcal{T}$ which we will specify when discussing the results of our application of Hamiltonian learning in Section 4.5. This is done by switching the CR interaction on for time duration t which is done in practice by implementing a `GaussianSquare` pulse of duration t. A `GaussianSquare` pulse is a square pulse with truncated Gaussian-shaped rising and falling edges. We discuss later in Section 4.4.3 how using this pulse may introduce non-idealities in the system evolution. Finally after obtaining the final state $|\psi(t)\rangle$, we apply the measurement operators in $\mathcal{M} = \{\sigma_I \otimes \exp\left(i\frac{\pi}{4}\sigma_Y\right), \sigma_I \otimes \exp\left(-i\frac{\pi}{4}\sigma_X\right), \sigma_I \otimes \sigma_I\}$ and measure only the second qubit which we have chosen as the target qubit. The query space is then $\mathcal{Q} = \mathcal{M} \times \mathcal{U} \times \mathcal{T}$. An example of a pulse schedule is shown in Fig. B.2 highlighting the different parts of the query. Moreover, in our experimental setup, we obtain measurements of the single-shot signal (integrated cavity amplitude) $\mathbf{c}$ which is a function of the measurement outcomes $\mathbf{y}$ which we have described earlier.

B.1.3 Modeling Pulse Shapes

We now describe how the imperfect pulse-shaping model stated in Section 4.4.3 was obtained. We consider cross-resonance control pulses (also called `GaussianSquare`) whose time-varying amplitudes are rectangular-shaped envelopes with tapered rising and falling edges, where the tapering is designed to minimize the signal energy that falls above and below the frequency of the sinsoid that is being modulated by the pulse envelope. The resulting unitary operators thus have the form $\tilde{U}(t) = \mathbb{T}\exp\left(-i\int_0^t \tilde{H}(t')dt'\right)$, where $\mathbb{T}$ is the time ordering operator, $\tilde{H}$ is the Hamiltonian at any particular time given by $\tilde{H}(t') = H(t', v(t'))$, H is the cross-resonance Hamiltonian (with the dependence on parameters $\boldsymbol{\theta}$ not shown), and $v(t')$ is the time-varying pulse envelope.

Let Δt_r and Δt_f be, respectively, the durations of the rising and falling edges of the shaped pulse envelope. The central portion of $v(t')$ is then a rectangular function such that, for $t' \in [\Delta t_r, t - \Delta t_f]$, $v(t') = V_{\max}\mathbb{1}_{t' \in [\Delta t_r, t-\Delta t_f]}$ where t is the total duration of the pulse,

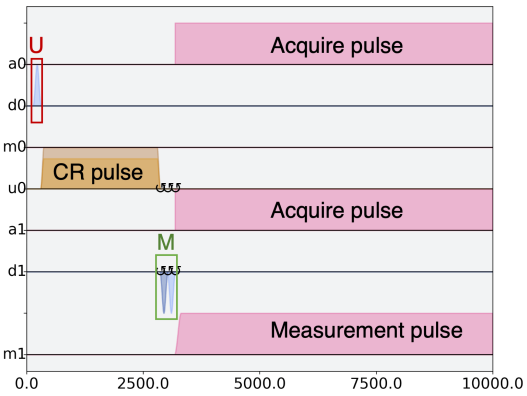


Figure B.2: An example of a CR pulse schedule on the IBM Quantum device *ibmq_ boeblingen* considering to the query of $x = (M, U, t)$ where $M = \sigma_I \otimes \exp\left(-i\frac{\pi}{4}\sigma_X\right)$, $U = \sigma_X \otimes \sigma_I$, and time duration $t = 6 \times 10^{-7}$s. The x-axis corresponds to time normalized by $dt = 2.22 \times 10^{-10}$ (Eq. B.1). The different channels corresponding to each qubit (y-axis) are written as the type of channel (see plot legend) followed by qubit number. Qubit 0 is set to be the control qubit and 1 to be the target qubit. The envelope of the different pulses are shown in each channel. The rotations on the drive or control channels indicate virtual Z gates. An equivalent representation of the quantum circuit is shown in Fig. 4.1.

and $V_{\max}$ is the amplitude of this central rectangular portion. We thus have

$$\tilde{U}(t) = \exp\left(-i\mathbb{T}\int_0^t \tilde{H}(t')dt'\right) \tag{B.2}$$

$$= \exp\left(-i\mathbb{T}\int_{t-\Delta t_f}^{t} H(t', v(t'))dt'\right) \exp\left(-i\mathbb{T}\int_{\Delta t_r}^{t-\Delta t_f} H(t', V_{\max})dt'\right) \exp\left(-i\mathbb{T}\int_0^{\Delta t_r} H(t', v(t'))dt'\right) \tag{B.3}$$

$$= \exp\left(-i\mathbb{T}\int_{t-\Delta t_f}^{t} H(t', v(t'))dt'\right) \exp\left(-iH_{\max}\int_{\Delta t_r}^{t-\Delta t_f} dt'\right) \exp\left(-i\mathbb{T}\int_0^{\Delta t_r} H(t', v(t'))dt'\right) \tag{B.4}$$

$$= \exp\left(-i\mathbb{T}\int_{t-\Delta t_f}^{t} H(t', v(t'))dt'\right) \exp\left(-iH_{\max}t_{\text{expt}}\right) \exp\left(-i\mathbb{T}\int_0^{\Delta t_r} H(t', v(t'))dt'\right) \tag{B.5}$$

where $t_{\text{expt}} = t - \Delta t_f - \Delta t_r$ and $H_{\max} = H(t', V_{\max})$ which is constant assuming that any signal distortions that are introduced by the control electronics and/or along the signal path to the quantum device are negligible.

The above equation decomposes $\tilde{U}(t)$ into the time evolution of the Hamiltonian that corresponds to the central rectangular-pulse portion of the control pulse with pre- and post-rotations that are determined by the tapered rising and falling edges of the pulse. In general, there is a nonlinear relationship between the shapes of these edges and the resulting pre- and post-rotations. However, based on the results reported in [She+16] and shown in Figure B.3, the cross-resonanace Hamiltonian parameters tend to vary fairly linearly with respect to the overall pulse amplitude. The pre- and post-rotations can thus be approximated by assuming

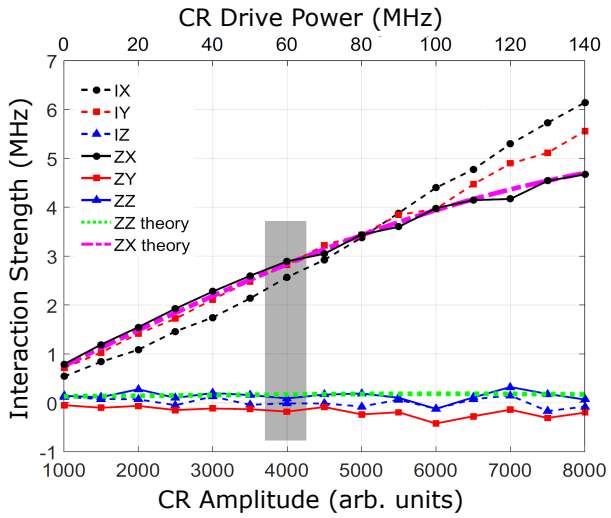


Figure B.3: Hamiltonian parameters $\mathbf{J}$ as a function of the amplitude of the control pulse or drive as reported in [She+16].

a first-order model for the time-varing Hamiltonian parameters given by

$$\mathbf{J}(t') \approx \frac{v(t')}{V_{\max}} \mathbf{J}_{\max} \tag{B.6}$$

where $\mathbf{J}_{\max}$ is the vector of parameters for the Hamiltonian $H_{\max}$ of the central recetangular portion of the pulse envelope. The time-varying Hamiltonian is then approximated by

$$H(t', v(t')) \approx \frac{v(t')}{V_{\max}} H_{\max} \,. \tag{B.7}$$

The overall unitary operator $\tilde{U}(t)$ is then approximated by

$$\tilde{U}(t) \approx \exp\left(-iH_{\max}\frac{1}{V_{\max}}\int_{t-\Delta t_f}^{t} v(t')dt'\right) \exp\left(-iH_{\max}t_{\text{expt}}\right) \exp\left(-iH_{\max}\frac{1}{V_{\max}}\int_{0}^{\Delta t_r} v(t')dt'\right) \tag{B.8}$$

$$= \exp\left(-iH_{\max}(t_{\text{expt}} + \Delta t_{\text{eff}})\right) \tag{B.9}$$

where

$$\Delta t_{\text{eff}} = \frac{1}{V_{\max}}\left(\int_{0}^{\Delta t_r} v(t')dt' + \int_{t-\Delta t_f}^{t} v(t')dt'\right) \,. \tag{B.10}$$

B.1.4 Likelihood Function and Fisher Information Matrix for the CR Hamiltonian

In this section, we give the expressions for the likelihood function of $p_{\text{y|x}}(y|x;\boldsymbol{\theta})$ and the Fisher information (FI) matrix. We consider the experimental setup as described in Section 4.4 and query space $\mathcal{Q}$ as described in Section 4.4.2.

Recall from Section 4.2.3, the FI matrix of a query x is given by

$$\mathcal{I}_x(\boldsymbol{\theta})[i,j] = \mathbb{E}\left[\frac{\partial \log p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \log p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})}{\partial \theta_j}\right] \tag{B.11}$$

where $\log p(y|x;\boldsymbol{\theta})$ is the log-likelihood of the measurement outcome y given the query x. In most cases in practice, the Fisher information matrix must be computed empirically in a Monte Carlo fashion. However, here we have a model of the CR Hamiltonian and models of the different noise sources affecting the quantum system available to us. We can thus evaluate the FI matrix analytically for different queries and in the presence or absence of noise.

In Absence of Noise

Likelihood In the noiseless case, the likelihood function of different measurement outcomes $y \in \{0,1\}$ given query $x = (M, U, t) \in \mathcal{Q}$ is

$$p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta}) = \sum_{z\in\{0,1\}} \left|\langle yz|Me^{-iH(\boldsymbol{\theta})t}U|00\rangle\right|^2 \tag{B.12}$$

Evaluating this for the different queries in the query space as described in Section 4.4.2, we obtain

$$p_{\mathrm{y|x}}(0|x;\boldsymbol{\theta}) = \begin{cases} \frac{1}{2}((\cos(\omega_j t) + \sin(\phi_j)\cos(\delta_j)\sin(\omega_j t))^2 + (\sin(\delta_j)\sin(\omega_j t) + \cos(\phi_j)\cos(\delta_j)\sin(\omega_j t))^2) & , x = (M_{\langle X\rangle}, U_j, t) \\ \frac{1}{2}((\cos(\omega_j t) - \cos(\phi_j)\cos(\delta_j)\sin(\omega_j t))^2 + (\sin(\delta_j)\sin(\omega_j t) + \sin(\phi_j)\cos(\delta_j)\sin(\omega_j t))^2) & , x = (M_{\langle Y\rangle}, U_j, t) \\ 1 - (\cos(\delta_j)\sin(\omega_j t))^2 & , x = (M_{\langle Z\rangle}, U_j, t) \end{cases} \tag{B.13}$$

where we have used the index $j \in \{0,1\}$ to refer to the different preparation operators $U_0 = \sigma_I\sigma_I$ and $U_1 = \sigma_X\sigma_I$. The measurement operators are: $M_{\langle X\rangle} = \sigma_I \otimes \exp\left(i\frac{\pi}{4}\sigma_Y\right)$, $M_{\langle Y\rangle} = \sigma_I \otimes \exp\left(-i\frac{\pi}{4}\sigma_X\right)$, and $M_{\langle Z\rangle} = \sigma_I \otimes \sigma_I$.

Fisher Information Matrix Noting that the measurement outcome $y \in \{0,1\}$, we have

$$\mathcal{I}_x(\boldsymbol{\theta})[i,j] = \sum_{y\in\{0,1\}} \frac{1}{p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})}\frac{\partial p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial p_{\mathrm{y|x}}(y|x;\boldsymbol{\theta})}{\partial \theta_j} \tag{B.14}$$

$$= \frac{1}{p_{\mathrm{y|x}}(0|x;\boldsymbol{\theta})(1 - p_{\mathrm{y|x}}(0|x;\boldsymbol{\theta}))}\frac{\partial p_{\mathrm{y|x}}(0|x;\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial p_{\mathrm{y|x}}(0|x;\boldsymbol{\theta})}{\partial \theta_j} \tag{B.15}$$

where in the second step, we have used the fact that $p_{\mathrm{y|x}}(1|x;\boldsymbol{\theta}) = 1 - p_{\mathrm{y|x}}(0|x;\boldsymbol{\theta})$ and $\frac{\partial p_{\mathrm{y|x}}(1|x;\boldsymbol{\theta})}{\partial \theta_i} = -\frac{\partial p_{\mathrm{y|x}}(0|x;\boldsymbol{\theta})}{\partial \theta_i}$.

The FI matrix elements can also be expressed using the Rabi oscillation of a query $p_{\mathrm{rabi}}(x;\boldsymbol{\theta})$ as follows

$$\mathcal{I}_x(\boldsymbol{\theta})[i,j] = \frac{1}{1 - p^2_{\mathrm{rabi}}(x;\boldsymbol{\theta})}\frac{\partial p_{\mathrm{rabi}}(x;\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial p_{\mathrm{rabi}}(x;\boldsymbol{\theta})}{\partial \theta_j} \tag{B.16}$$

The FI matrices $\mathcal{I}_x(\boldsymbol{\theta})$ for the different queries $x = (M, U, t)$ (Section 4.4.2) depend on the parameterization of choice. We denote the FI matrix considering the parameterization

of $\mathbf{\Lambda}$ as $\mathcal{I}_x(\mathbf{\Lambda})$. Note that the Fisher information matrices $\mathcal{I}_x(\mathbf{J})$ for the parameterization of $\mathbf{J}$ is related to the former through the jacobian of $\mathbf{\Lambda}$ with respect to $\mathbf{J}$.

$$D_{\mathbf{\Lambda},\mathbf{J}} = \begin{bmatrix} \frac{\mathrm{Re}(\beta_0)}{\omega_0} & -\frac{\mathrm{Re}(\beta_0)a_0}{|\beta_0|a_0} & -\frac{\mathrm{Im}(\beta_0)}{|\beta_0|^2} & \frac{\mathrm{Re}(\beta_1)}{\omega_1} & -\frac{\mathrm{Re}(\beta_1)a_1}{|\beta_1|a_1} & -\frac{\mathrm{Im}(\beta_1)}{|\beta_1|^2} \\ \frac{\mathrm{Im}(\beta_0)}{\omega_0} & -\frac{\mathrm{Im}(\beta_0)a_0}{|\beta_0|a_0} & \frac{\mathrm{Re}(\beta_0)}{|\beta_0|^2} & \frac{\mathrm{Im}(\beta_1)}{\omega_1} & -\frac{\mathrm{Im}(\beta_1)a_1}{|\beta_1|a_1} & \frac{\mathrm{Re}(\beta_1)}{|\beta_1|^2} \\ \frac{a_0}{\omega_0} & \frac{|\beta_0|}{\omega_0^2} & 0 & \frac{a_1}{\omega_1} & \frac{|\beta_1|}{\omega_1^2} & \\ \frac{\mathrm{Re}(\beta_0)}{\omega_0} & -\frac{\mathrm{Re}(\beta_0)a_0}{|\beta_0|a_0} & -\frac{\mathrm{Im}(\beta_0)}{|\beta_0|^2} & -\frac{\mathrm{Re}(\beta_1)}{\omega_1} & \frac{\mathrm{Re}(\beta_1)a_1}{|\beta_1|a_1} & \frac{\mathrm{Im}(\beta_1)}{|\beta_1|^2} \\ \frac{\mathrm{Im}(\beta_0)}{\omega_0} & -\frac{\mathrm{Im}(\beta_0)a_0}{|\beta_0|a_0} & \frac{\mathrm{Re}(\beta_0)}{|\beta_0|^2} & -\frac{\mathrm{Im}(\beta_1)}{\omega_1} & \frac{\mathrm{Im}(\beta_1)a_1}{|\beta_1|a_1} & -\frac{\mathrm{Re}(\beta_1)}{|\beta_1|^2} \\ \frac{a_0}{\omega_0} & \frac{|\beta_0|}{\omega_0^2} & 0 & -\frac{a_1}{\omega_1} & -\frac{|\beta_1|}{\omega_1^2} & \end{bmatrix} \tag{B.17}$$

where the (i,j)th element is given by $\partial\Lambda_j/\partial J_i$ with $a_{0,1}$ and $\beta_{0,1}$ as defined as in Eq. 4.31.

Note that $\mathcal{I}_x(\mathbf{\Lambda})$ is of rank-1 for each query and takes a block form of $\begin{bmatrix} \mathcal{I}_0 & 0 \\ 0 & 0 \end{bmatrix}$ for $U = \sigma_I \otimes \sigma_I$ and $\begin{bmatrix} 0 & 0 \\ 0 & \mathcal{I}_1 \end{bmatrix}$ for $U = \sigma_X \otimes \sigma_I$. This indicates that queries involving $U = \sigma_I \otimes \sigma_I$ are informative about the Hamiltonian parameters $(\omega_0, \delta_0, \phi_0)$ and those involving $U = \sigma_X \otimes \sigma_I$ are informative about $(\omega_1, \delta_1, \phi_1)$.

In Presence of Noise Sources and Nonidealities

In Section 4.4.3, we modeled the effect of different noise sources and nonidealities on the quantum system. In particular, we discussed the effect of imperfections in control in Section 4.4.3, effect of decoherence in Section 4.4.3 and how the observed measured outcome is subject to readout noise in Section 4.4.3. We consider the two-qubit decoherence model from Section 4.4.3, and the readout noise model of a bit-flip channel based on binary classification from Section 4.4.3. We denote the noisy observed measurement outcome as $\tilde{y}$ and the hidden measurement outcome before the effect of the bit-flip channel as y.

Likelihood The likelihood function in the presence of the noise sources of readout noise, imperfect-pulse shaping, and decoherence, is then given by

$$p_{\tilde{\mathrm{y}}|\mathrm{x}}(\tilde{y}|x;\boldsymbol{\theta}) = p_{\tilde{\mathrm{y}}|\mathrm{y}}(\tilde{y}|y) \sum_{z\in\{0,1\}} \left[(1-p_d(t))\, p_{\mathrm{yz}|\mathrm{x}}\left(yz|(M,U,t+\Delta t_{\mathrm{eff}}(\boldsymbol{\theta}))\,;\boldsymbol{\theta}\right) + \frac{1}{4}p_d(t) \right] \tag{B.18}$$

$$\begin{aligned} &= (1-p_d(t)) \left[(1-r_{1-\tilde{y}}) p_{\mathrm{y}|\mathrm{x}}(\tilde{y}|(M,U,t+\Delta t_{\mathrm{eff}}(\boldsymbol{\theta}));\boldsymbol{\theta}) + r_{\tilde{y}} p_{\mathrm{y}|\mathrm{x}}(1-\tilde{y}|(M,U,t+\Delta t_{\mathrm{eff}}(\boldsymbol{\theta}));\boldsymbol{\theta}) \right] \\ &+ \frac{1}{2}p_d(t)\left(1-r_{1-\tilde{y}}+r_{\tilde{y}}\right) \end{aligned} \tag{B.19}$$

with the probability of the two-qubit string yz given by

$$p_{\mathrm{yz}|\mathrm{x}}\left(yz|(M,U,t)\,;\boldsymbol{\theta}\right) = \left|\left\langle yz|Me^{-iH(\boldsymbol{\theta})t)}U|00\right\rangle\right|^2 \tag{B.20}$$

and the probability $p_{\mathrm{y}|\mathrm{x}}(y|x)$ given by Eq. 4.1 (the noiseless case). In the expressions above, we have used the tuple representation of the query x, $p_d(t)$ which is the depolarization probability associated with the two-qubit decoherence model as discussed in Section 4.4.3 and (r_0, r_1) which are the readout noise parameters as introduced in Section 4.4.3.

Fisher Information Matrix The FI matrix for a given query x is given by

$$\mathcal{I}_x(\boldsymbol{\theta})[i,j] = \frac{1}{p_{\tilde{\mathrm{y}}|\mathrm{x}}(0|x;\boldsymbol{\theta})(1-p_{\tilde{\mathrm{y}}|\mathrm{x}}(0|x;\boldsymbol{\theta}))} \frac{\partial p_{\tilde{\mathrm{y}}|\mathrm{x}}(0|x;\boldsymbol{\theta})}{\partial \theta_i} \frac{p_{\tilde{\mathrm{y}}|\mathrm{x}}(0|x;\boldsymbol{\theta})}{\partial \theta_j} \tag{B.21}$$

where

$$\frac{\partial p_{\tilde{\mathrm{y}}|\mathrm{x}}(0|x;\boldsymbol{\theta})}{\partial \theta_i} = (1-r_0-r_1)\,(1-p_d(t))\,\frac{\partial p_{\mathrm{y}|\mathrm{x}}(0|x;\boldsymbol{\theta})}{\partial \theta_i} \tag{B.22}$$

and

$$\frac{\partial p_{\mathrm{y}|\mathrm{x}}(0|x;\boldsymbol{\theta})}{\partial \theta_i} = \begin{cases} \frac{\partial p(y=0|(M,U,t+\Delta t_{\mathrm{eff}}(\boldsymbol{\theta}));\boldsymbol{\theta})}{\partial \theta_i}, & \theta_i \notin \{\omega_0,\omega_1\} \\ \frac{\partial p(y=0|(M,U,t+\Delta t_{\mathrm{eff}}(\boldsymbol{\theta}));\boldsymbol{\theta})}{\partial (\theta_i \Delta t_{\mathrm{eff}}(\boldsymbol{\theta}))} \frac{\partial (\theta_i \Delta t_{\mathrm{eff}}(\boldsymbol{\theta}))}{\partial \theta_i}, & \theta_i \in \{\omega_0,\omega_1\} \end{cases} \tag{B.23}$$

Note that special attention has to be given when taking the derivative with respect to $\omega_{0,1}$ as $\Delta t_{\mathrm{eff}}(\boldsymbol{\theta})$ (Section 4.4.3) actually only has a dependence on these two components in $\boldsymbol{\Lambda}$ and appears in the effective evolution time t with a prefactor of $\omega_{0,1}$ in the likelihood (see Eq. B.13).

B.2 Computational Details of Query Optimization

In Section 4.5.1, we pointed out that the number of shots available for each query in the experimental datasets collected from the IBM Quantum devices are limited. In this section of the Appendix, we describe how the query optimizations for HAL-FI (Eq. 4.19) and HAL-FIR (Eq. 4.21) are solved under shot constraints for each query. We then describe how the computational cost of query optimization can be reduced through uncertainty filtering of the query space.

B.2.1 Different Query Optimizations and Strategies for Handling Query Constraints

We describe how constraints can be handled for the query optimization (Eq. 4.19) in HAL-FI (Algorithm 4) but the same approach can also be used for HAL-FIR. Let us consider the ith round in active learning. The Hamiltonian parameter estimate in this round is $\hat{\boldsymbol{\theta}}^{(i)}$. Let us denote the number of shots available for each query $x \in \mathcal{Q}$ in the ith round as $N^{(i)}_{\mathrm{shots}}(x)$. The number of shots available before learning has started is then $N^{(0)}_{\mathrm{shots}}(x)$. Let the total number of shots that have already been made against query x by the ith round (inclusive) be denoted as $N^{(i)}_{\mathrm{tot}}(x)$. We will denote the total number of shots made over all queries inputted to the oracle by the ith round (inclusive) as $N^{(i)}_{\mathrm{tot}}$.

We can frame the query optimization problem under shot constraints in two different ways, motivated by the asymptotic optimal query distribution associated with HAL-FI:

$$q^\star = \arg\min_{q \in \mathcal{P}(\mathcal{Q})} \mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta}^\star)) \tag{B.24}$$

where $\mathcal{P}(\mathcal{Q})$ is the family of all probability distributions over the specified query space $\mathcal{Q}$. As discussed in Section 4.2.3, we cannot solve for this distribution in practice as this requires

us to have access to the true parameters $\boldsymbol{\theta}^\star$, and hence we solve for a sub-optimal query distribution $q^{(i)}(\hat{\boldsymbol{\theta}}^{(i)})$ given the current parameter estimates $\hat{\boldsymbol{\theta}}$. During active learning, we can solve for $q^{(i)}(\hat{\boldsymbol{\theta}}^{(i)})$ by viewing it as the query distribution over all the queries that have inputted to the oracle (i.e., including from previous rounds) or as the query distribution associated with the current ith batch of queries that will be issued. These two viewpoints lead us to two different approaches of how the shot constraints are handled and how the queries to be inputted to the oracle are sampled from the query distribution.

Firstly, consider the following query optimization problem

$$q^{(i)} = \arg\min_q \mathrm{Tr}(\mathcal{I}_q^{-1}(\hat{\boldsymbol{\theta}}^{(i)})) \tag{B.25}$$

$$\text{subject to} \sum_{x\in\mathcal{Q}^{(i)}} q(x) = 1, \text{ and} \frac{N_{\text{tot}}^{(i-1)}(x)}{N_{\text{tot}}^{(i)}} \leq q(x) \leq \min\left\{1, \frac{N_{\text{shots}}^{(i)}(x)}{N_{\text{tot}}^{(i)}}\right\} \forall x \in \mathcal{Q}^{(i)} \tag{B.26}$$

with corresponding sampling of queries in that batch as:

$$\text{Sample } N_b \text{ queries } X_q^{(i)} \text{ from } \mathcal{Q}^{(i)} \text{ w.p. } q_b^{(i)} \tag{B.27}$$

where $q_b^{(i)}(x) = q^{(i)}(x) - \frac{N_{\text{tot}}^{(i-1)}(x)}{N_{\text{tot}}^{(i)}}$. The query optimization in this case considers all the queries that have been sampled from earlier batches during active learning. The query distribution $q^{(i)}$ is then the distribution over all the queries made so far and $q_b^{(i)}$ is the query distribution for the batch to be issued. Further, the Fisher information matrix associated with $q^{(i)}$ is guaranteed to be invertible but that associated with $q_b^{(i)}$ may be non-invertible. This suggests that the outcomes of queries made in this round of the active learning procedure may not be informative about all the Hamiltonian parameters.

In order to ensure that the query distributions for each batch are informative about all the Hamiltonian parameters, one may alternately solve the following query optimization problem

$$q^{(i)} = \arg\min_q \mathrm{Tr}(\mathcal{I}_q^{-1}(\hat{\boldsymbol{\theta}}^{(i)})) \tag{B.28}$$

$$\text{subject to} \sum_{x\in\mathcal{Q}^{(i)}} q(x) = 1, \text{ and } 0 \leq q(x) \leq \min\left\{\frac{N_{\text{shots}}^{(i)}(x)}{N_b^{(i)}}, 1\right\}, \forall x \in \mathcal{Q}^{(i)} \tag{B.29}$$

where $N_b^{(i)}$ is the size of the batch of queries being issued to the oracle in the ith round. Queries of the issued batch are then sampled as

$$\text{Sample } N_b \text{ queries } X_q^{(i)} \text{ from } \mathcal{Q}^{(i)} \text{ w.p. } q^{(i)} \tag{B.30}$$

where $q^{(i)}$ is now the query distribution for each batch. This query optimization can be viewed as a greedy approach of query selection. The Fisher information matrix associated with $q^{(i)}$ is guaranteed to be invertible and hence informative about all the Hamiltonian parameters. We thus solve the query optimization considering shot constraints on each query according to Eq. B.29 for our application.

As we sample queries randomly according to $q^{(i)}$ and not proportional to $q^{(i)}$, the resulting X_q might not satisfy query constraints exactly and an additional pruning step is required.

Moreover, to reduce the computational cost of the query optimization solve, we consider a subset of queries $\mathcal{Q}^{(i)}_{\text{filtered}} \subset \mathcal{Q}^{(i)}$ for which shots are still available.

In Algorithm 10, we summarize the steps taken to handle query constraints during query optimization for HAL-FI. It takes the inputs of the query space $\mathcal{Q}^{(i)}$ of the current ith round of active learning, number of shots available for each query and the size of the batch of queries to be issued. The inputted query space is first filtered by retaining only those queries for which shots are available. The resulting filtered query space is denoted by $\mathcal{Q}^{(i)}_{\text{filtered}}$, and we then compute the query distribution $q^{(i)}$ according to Eq. B.29. We then sample queries for the batch X_q according to this distribution after mixing and with incorporation of a pruning step. Note that lines 4 and 5 were also used in the original query optimization algorithms of HAL-FI (Algorithm 4) and HAL-FIR (Algorithm 5) to encourage exploration. An illustration of the operation of the query optimization and handling of query constraints is shown in Figure B.4 for a particular run of the HAL-FI learner.

Algorithm 10 Handling query constraints during query optimization in HAL-FI

Input: Current query space $\mathcal{Q}^{(i)}$, number of shots available for each query $N^{(i)}_{\text{shots}}(x)$, size of batch of queries requested $N^{(i)}_b$, total number of queries made so far $N^{(i-1)}_{tot}$

Output: Query set X_q of size $N^{(i)}_b$

1: Filter $\mathcal{Q}^{(i)}$ by retaining queries for which shots are available: $\mathcal{Q}^{(i)}_{\text{filtered}} = \{x \in \mathcal{Q}^{(i)} | N^{(i)}_{\text{shots}}(x) > 0\}$
2: Obtain HAL-FI query distribution $q^{(i)}$ by solving the query optimization of Eq. B.29 with input of $\mathcal{Q}^{(i)}_{\text{filtered}}$
3: Obtain uniform distribution over filtered query space: $p_U = 1/|\mathcal{Q}^{(i)}_{\text{filtered}}|$
4: Set mixing coefficient: $\mu = 1 - 1/|N^{(i-1)}_{tot}|^{1/6}$
5: Modify query distribution: $q^{(i)} = \mu q^{(i)} + (1-\mu)p_U$
6: Sample X_q from $\mathcal{Q}^{(i)}_{\text{filtered}}$ according to $q^{(i)}$
7: Prune queries from X_q that cannot be made and randomly assign valid queries
8: **return** X_q

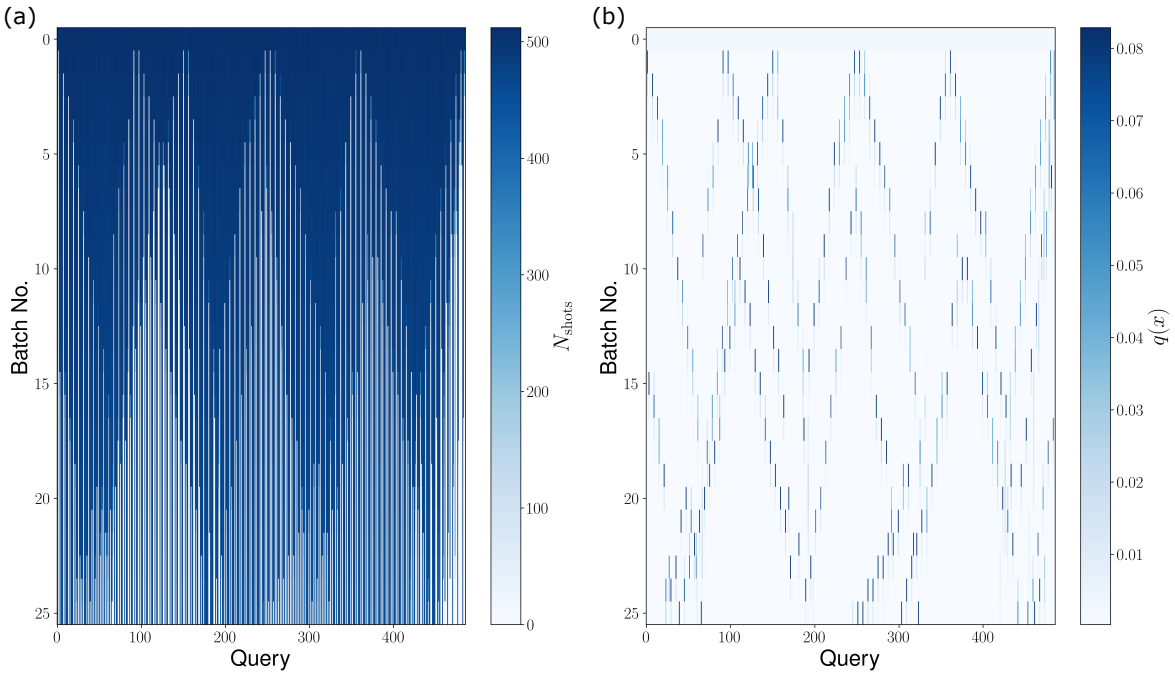


Figure B.4: Visualization of queries being selected for each batch during a particular run of HAL-FI for $N_0 = 2000$ and 25 batches of size $N_b = 5000$. In (a), we plot the available number of shots for each query after batches of queries are made during active learning. In (b), we plot the query distribution for each batch during active learning. The number of shots available for each query before learning starts is $N^{(0)}_{\text{shots}}(x) = 512\,\forall x$. Half the total number of shots available in this dataset is exhausted by the end of learning. Note that the different parameter values considered for HAL-FI are for stress testing the query constraints' handling procedure and are not tuned for the HAL-FI algorithm.

B.2.2 Uncertainty Filtering of Query Space

Uncertainty filtering becomes a crucial step during query optimization when we consider HAL-FI with an adaptively growing query space. We noted in Section 4.2.3 that the computational cost of query optimization (Eq. 4.19) scales as $\mathcal{O}(n_{\mathcal{Q}}^2 m^3 + n_{\mathcal{Q}} m^4 + m^5)$ where $n_{\mathcal{Q}} = |\mathcal{Q}|$ is the number of queries in the query space, and m is the length of the parameter vector $\boldsymbol{\theta}$. This computational cost is alleviated through filtering of the query space based on entropy $S(x)$. The entropy of the different queries can be computed from the model probability expressions available to us (see Appendix B.1) given the current Hamiltonian parameter estimate $\hat{\boldsymbol{\theta}}$. The filtered query space based on entropy is determined as follows:

$$\mathcal{Q}_S = \{x | x \in \mathcal{Q}, S(x) > \tau \times \max_{x' \in \mathcal{Q}} S(x')\} \tag{B.31}$$

where we set the threshold $\tau = 0.95$ i.e., we only consider queries with entropy that is at least 0.95 times the highest entropy. This value of τ was chosen to ensure that at most only half of the queries are retained after uncertainty filtering of the query space. An illustration of uncertainty filtering of a query space is given in Figure B.5.

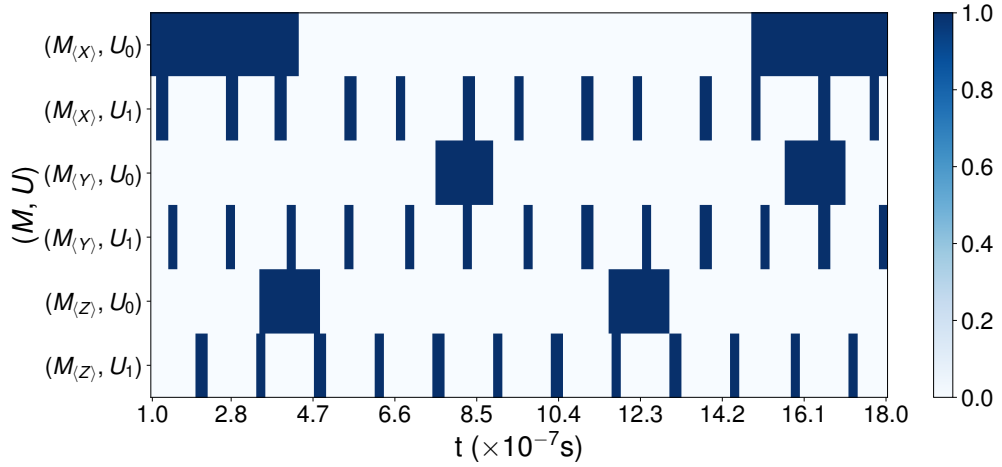


Figure B.5: Uncertainty filtering of $\mathcal{Q}$ as defined in Section 4.4.2 with $\mathcal{T}$ set to be a sequence of 243 linearly equispaced evolution times in $[10^{-7}, 18 \times 10^{-7}]s$. The The x-axis corresponds to the system evolution times $t \in \mathcal{T}$. The y-axis indicates the different combinations of measurement operators and preparation operators available for each query in $\mathcal{Q}$. The different preparation operators are denoted as $U_0 = \sigma_I\sigma_I$ and $U_1 = \sigma_X\sigma_I$. We consider the $\boldsymbol{\theta}^\star$ corresponding to IBM Quantum device D *ibmq_ boeblingen* under drive configuration 2, with the different noise sources of readout noise, imperfect pulse-shaping, and decoherence being accounted for. Queries colored as dark blue are retained in the filtered query space and rest is filtered out.

B.3 Estimation Procedure for Learning Cross-Resonance Hamiltonians

In this section of the Appendix, we discuss in detail the estimation procedures of Section 4.4.4 for the estimator based on regression (Section 4.4.4) and the MLE estimator (Section 4.4.4), used for learning cross-resonance Hamiltonians (Eq. 4.31). These estimators can be used by both the passive learner and the active learners of HAL-FI/HAL-FIR (Section 4.3). Additionally, we can also use Bayesian estimators such as the sequential Monte Carlo method discussed in Section 4.2.1. In the following discussion, we consider the experimental setup from Section 4.4. We will also discuss how these estimation procedures can be improved or extended to other Hamiltonians.

Let us recall the notation introduced in Section 4.3 and Section 4.4.4. The different rounds of active learning are indexed by $i \in [i_{\max}]$. In each round of active learning, we use an estimation procedure divided into multiple steps. We index each of these fractional steps by k. The Hamiltonian parameter estimate at the kth fractional step in the ith round of learning will then be denoted by $\hat{\boldsymbol{\theta}}^{(i,k)}$ with the parameter estimate in the ith round at the end of the estimation procedure denoted simply by $\hat{\boldsymbol{\theta}}^{(i)}$. The training examples available at the ith round is given by $(X^{(i)}, Y^{(i)})$.

We now discuss the estimation procedures for (i) regression and (ii) maximum-likelihood estimation (MLE).

B.3.1 Regression

As mentioned in Section 4.4.4, the first step of our estimation procedure involves frequency estimation followed by a nonlinear regression solve combined with a gradient descent procedure on the Rabi oscillations inferred from data (Eq. 4.39). We denoted these inferred Rabi oscillations by $\hat{p}_{\text{rabi}}(t)$.

Estimates are obtained by fitting nonlinear regression equations of the form $A\cos(\omega t) + B\sin(\omega t) + C$ to the Rabi oscillations, where the fit minimizes the sum of the weighted L_2 errors across the corresponding time series. For a given value of ω, the coefficients A, B, and C are estimated using weighted linear least-squares regression. Fast Fourier Transforms (FFTs) [DR93] are used to perform an efficient grid search over these weighted least-squares fits to obtain initial estimates of ω. Bracketed gradient-based search is then used to refine the estimates. The resulting coefficients A, B, and C for each Rabi oscillation are subsequently used to obtain estimates of the corresponding δ and ϕ Hamiltonian parameters. We note this analysis in general assumes periodic signals. In order to successfully estimate the Rabi frequencies using standard FFT, the set of time points $\mathcal{T}$ should span at least the time duration established by the minimum-frequency criterion defined in Section 4.5.3.

For each Rabi oscillation, the following weighted least-squares *normal* equations are solved to estimate the A, B, C coefficients for a given value of ω, where the summations are performed over time points $t \in \mathcal{T}$, and where the w_t's are weighting factors

$$\begin{bmatrix} \sum_t w_t \cos^2(\omega t) & \sum_t w_t \cos(\omega t)\sin(\omega t) & \sum_t w_t \cos(\omega t) \\ \sum_t w_t \cos(\omega t)\sin(\omega t) & \sum_t w_t \sin^2(\omega t) & \sum_t w_t \sin(\omega t) \\ \sum_t w_t \cos(\omega t) & \sum_t w_t \sin(\omega t)\sin(\omega t) & \sum_t w_t \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \sum_t w_t \hat{p}_{\text{rabi}}(t)\cos(\omega t) \\ \sum_t w_t \hat{p}_{\text{rabi}}(t)\sin(\omega t) \\ \sum_t w_t \hat{p}_{\text{rabi}}(t) \end{bmatrix} \tag{B.32}$$

The weights w_t are equal to one over the estimation variance of $\hat{p}_{\text{rabi}}(t)$ for the corresponding time points t. The purpose of the weights is to account for the heteroskedasticity of the estimation errors in $\hat{p}_{\text{rabi}}(t)$.

If we write Eq. B.32 in the form $R\mathbf{a} = \mathbf{b}$, where $\mathbf{a}$ is the vector of A, B, C coefficients that satisfies Eq. B.32, then the residual weighted squared error is given by

$$\mathcal{E}^2 = \sum_t w_t \hat{p}^2_{\text{rabi}}(t) - \mathbf{a}^T R\mathbf{a} \ . \tag{B.33}$$

For a group of Rabi oscillations that share the same frequency ω, the optimization problem is to find the ω that minimizes the sum of the weighted squared errors $\mathcal{E}^2$ across those time series. Fourier techniques are used to make the initial search computationally efficient by noting that Eq. B.32 can be rewritten as

$$\begin{bmatrix} \frac{1}{2}\sum_t w_t + \frac{1}{2}\sum_t w_t \cos(2\omega t) & \frac{1}{2}\sum_t w_t \sin(2\omega t) & \sum_t w_t \cos(\omega t) \\ \frac{1}{2}\sum_t w_t \sin(2\omega t) & \frac{1}{2}\sum_t w_t - \frac{1}{2}\sum_t w_t \cos(2\omega t) & \sum_t w_t \sin(\omega t) \\ \sum_t w_t \cos(\omega t) & \sum_t w_t \sin(\omega t) & \sum_t w_t \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \sum_t w_t \hat{p}_{\text{rabi}}(t)\cos(\omega t) \\ \sum_t w_t \hat{p}_{\text{rabi}}(t)\sin(\omega t) \\ \sum_t w_t \hat{p}_{\text{rabi}}(t) \end{bmatrix} \tag{B.34}$$

which simplifies to

$$\begin{bmatrix} \frac{1}{2}\mathcal{F}\{w_t\}(0) + \frac{1}{2}\mathrm{Re}(\mathcal{F}\{w_t\}(2\omega)) & -\frac{1}{2}\mathrm{Im}(\mathcal{F}\{w_t\}(2\omega)) & \mathrm{Re}(\mathcal{F}\{w_t\}(\omega)) \\ -\frac{1}{2}\mathrm{Im}(\mathcal{F}\{w_t\}(2\omega)) & \frac{1}{2}\mathcal{F}\{w_t\}(0) - \frac{1}{2}\mathrm{Re}(\mathcal{F}\{w_t\}(2\omega)) & -\mathrm{Im}(\mathcal{F}\{w_t\}(\omega)) \\ \mathrm{Re}(\mathcal{F}\{w_t\}(\omega)) & -\mathrm{Im}(\mathcal{F}\{w_t\}(\omega)) & \mathcal{F}\{w_t\}(0) \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(\mathcal{F}\{w_t\hat{p}_{\mathrm{rabi}}(t)\}(\omega)) \\ -\mathrm{Im}(\mathcal{F}\{w_t\hat{p}_{\mathrm{rabi}}(t)\}(\omega)) \\ \mathcal{F}\{w_t\hat{p}_{\mathrm{rabi}}(t)\}(0) \end{bmatrix}, \quad \text{(B.35)}$$

where

$$\mathcal{F}\{x(t)\}(\omega) = \sum_t x(t)e^{-i\omega t}, \quad \text{(B.36)}$$

is the discrete-time Fourier transform of $x(t)$ over $t \in \mathcal{T}$, $x(t) \in \{w_t, w_t\hat{p}_{\mathrm{rabi}}(t)\}$. Standard FFTs allow us to efficiently calculate the various Fourier coefficients in Eq. B.35 at fixed intervals in the spectrum using

$$F\{x(\frac{n}{F_s} + t_0)\}(k) = \sum_{n=0}^{N-1} x(\frac{n}{F_s} + t_0)e^{-i\frac{2\pi}{N}kn} \quad \text{(B.37)}$$

The Fourier coefficients calculated via the above equation can then be used in Eq. B.35 to perform a computationally efficient grid search over possible values of $\omega \in \{1\frac{2\pi F_s}{N}, 2\frac{2\pi F_s}{N}, 3\frac{2\pi F_s}{N}, ..., \pi F_s\}$, with an initial estimate for ω obtained by minimizing Eq. B.33 summed over the corresponding Rabi oscillations. Bracketed gradient-based search can then be performed using Eq. B.32 directly at a higher computational cost to refine these grid-search estimates.

For grid search purposes, we have also found it useful to sample frequency amplitudes at intermediate half steps to better avoid local minima, particularly when $\mathcal{T}$ does not satisfy the minimum-frequency criteria. Such intermediate sampling is accomplished by multiplying signals of the form $x(\frac{n}{F_s} + t_0)$ by a rotating exponential frequency and then computing a second FFT:

$$G\{x(\frac{n}{F_s} + t_0)\}(k) = F\{x(\frac{n}{F_s} + t_0)\}(k + \frac{1}{2}) = \sum_{n=0}^{N-1} \left[x(\frac{n}{F_s} + t_0)e^{-i\frac{\pi}{N}n}\right] e^{-i\frac{2\pi}{N}kn} \quad \text{(B.38)}$$

For general multi-parameter Hamiltonians, the above initial estimation procedure must be modified to ensure that all the frequency components in a Rabi oscillation corresponding to (M, U) can be faithfully extracted. The model must then assume that each Fourier series has up to K modes. Normal equations for the same can then be setup. Such an approach has been employed in Bayesian spectral analysis [Bre13] and non-stationary time-series estimation [LBK21; She+21].

B.3.2 Maximum Likelihood Estimation

The estimation procedure for MLE (Section 4.4.4) requires an initial estimate $\hat{\boldsymbol{\theta}}^{(i,0)}$ which ideally lives in the same convex basin as the global minimum of the MLE. This allows

for a more localized search to be carried out and a stochastic gradient descent procedure should allow us to jump out of any smaller local minima here if present. When an initial estimate cannot be provided by the learner, the estimates obtained through regression as just discussed is used as an initial guess to the MLE.

We solve the MLE 4.5 through a combination of SGD applied on different parameterizations and the quasi-Newton method for further refinement. Addition of the latter step helps us in saving computationally expensive hyperparameter tuning that is required.

1. SGD solve using ADAM considering the $\mathbf{\Lambda}$ parameterization and learning rate of $\eta^i_{\mathbf{\Lambda}}$ using the input of $\hat{\boldsymbol{\theta}}^{(i,0)}$. This returns the output of $\hat{\boldsymbol{\theta}}^{(i,1)}$.
2. SGD solve using ADAM considering the $\mathbf{J}$ parameterization and learning rate of $\eta^i_{\mathbf{J}}$ using the input of $\hat{\boldsymbol{\theta}}^{(i,1)}$. This returns the output of $\hat{\boldsymbol{\theta}}^{(i,2)}$.
3. MLE solve considering the $\mathbf{J}$ parameterization using LFBGS-B using the input of $\hat{\boldsymbol{\theta}}^{(i,2)}$. This returns the output of $\hat{\boldsymbol{\theta}}^{(i,3)}$.

We set the learning rate η^i for ADAM [KB14] in the ith round of active learning according to the number of queries already made. We consider the learning rate to be $\eta^i \propto \frac{1}{\sqrt{|X^{(i)}|}}$ i.e., the learning rate is reduced inversely to the square root of the number of training examples. This ensures a more localized search as we progress in the learning. We consider $\eta^0 = 10^{-3}$. We found that carrying out step 2 after step 1 gave us more accurate estimates of $\hat{\theta}$ than just carrying out step 1. Moreover, after the first few rounds of HAL, we can skip steps 1 and 2. We can carry out step 3 directly using an initial condition of $\hat{\theta}^{(i,0)}$ from initial estimation or $\hat{\theta}^{(i-1)}$ from the previous round.

B.3.3 Energy Landscapes of Negative Log-Likelihood Loss for Cross-Resonance Hamiltonian

We ascertain the efficacy of the estimation procedure by visualizing the energy landscapes of the negative log-likelihood loss function (Eq. 4.56) and inspecting the location of the $\hat{\boldsymbol{\theta}}$ in the landscape. In Figure B.6, we plot the energy landscape obtained from an experimental dataset, for the two different parameterizations $\mathbf{J}$ and $\mathbf{\Lambda}$. The energy landscapes indicate the nonlinear and non-convex nature of the MLE of Eq. 4.56. These energy landscapes also indicate why solving the MLE in the parameterization of $\mathbf{\Lambda}$ using ADAM is carried out before solving the MLE in the parameterization of $\mathbf{J}$. The slices along specific components of $\mathbf{\Lambda}$ display more convex like nature than those along components of $\mathbf{J}$. It should also be noted that there is a global minimum present in the energy landscapes which we are able to identify using our estimation procedure.

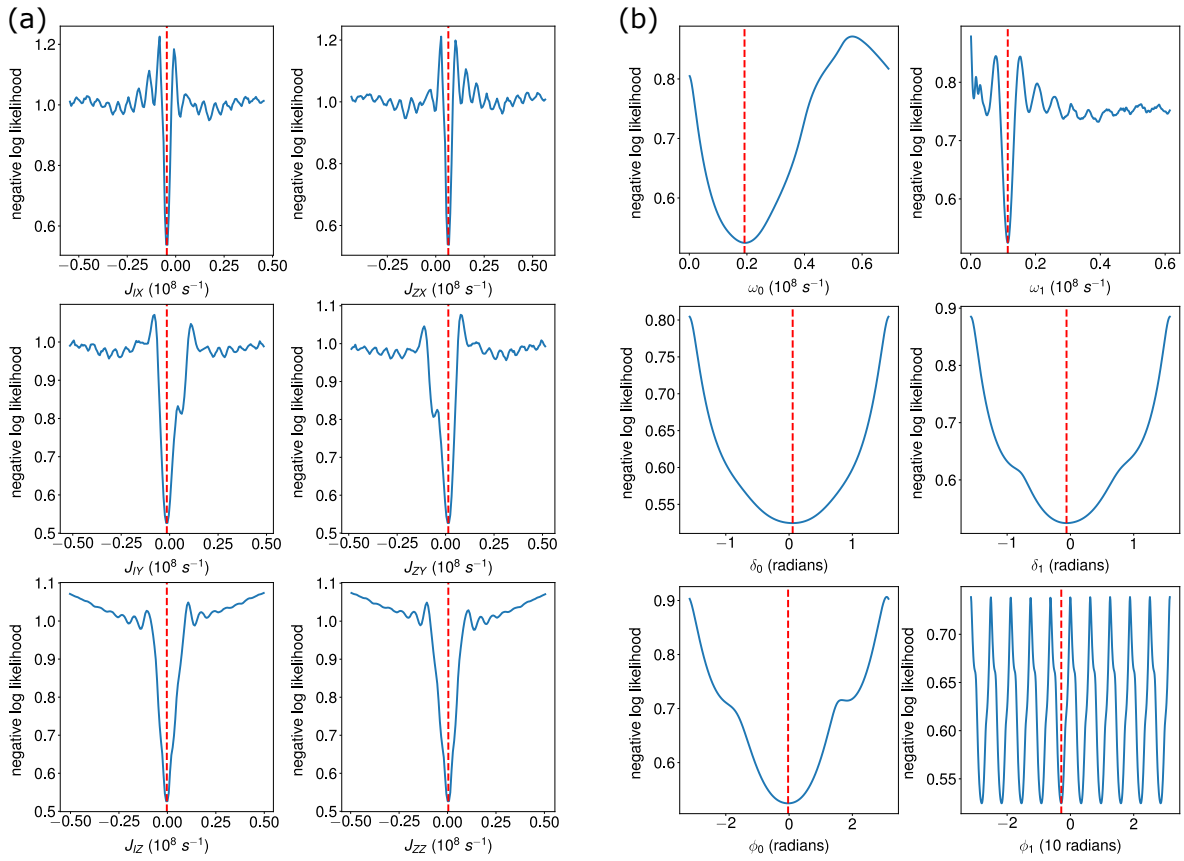


Figure B.6: Slices of energy landscapes of the log-likelihood loss function along the different parameter components considering experimental data collected from IBM Quantum device D *ibmq_boeblingen* under drive configuration 3 (Table 4.4) (a) using parameterization $\mathbf{J}$, and (b) using parameterization $\mathbf{\Lambda}$. In each slice of θ_i (x-axis), we fix the values of the other components as obtained through estimation and evaluate the negative log-likelihood loss (y-axis) by changing the value of θ_i. We indicate the Hamiltonian parameter estimate $\hat{\theta}_i$ as obtained through our estimation procedure by a dashed red line.

B.3.4 Incorporating uncertainty from shot noise

The inferred Rabi oscillations $p_{\text{rabi}}(t)$ used for estimation are sensitive to the number of shots made for a given query $x = (M, U, t)$. This variability in the Rabi oscillations leads to a variability in the estimates of $\hat{\boldsymbol{\theta}}^{(i,0)}$ produced. This variability is particularly high during the initial rounds of HAL when there are only a few shots of each query present in the set of training examples. In order to accurately account for this variability and hence include the uncertainty in our estimates of the Hamiltonian parameters, we consider the following procedure. Let us consider the initial estimation of Algorithm 6 as the procedure applied on a particular realization of the Rabi oscillations. We construct n_{rep} realizations of the Rabi oscillations considering the observed Rabi oscillation data and sampling according to the binomial distribution associated with the number of shots for each query. For each of these realizations, we obtain frequency estimates of $\omega_{0,1}$ through the initial estimation as described above. We then fit the parametric distributions of log-normal distributions to

frequency estimates from each each realization. We then continue with the initial estimation procedure and MLE considering each of these realizations.

Other ways of incorporating uncertainty in the Hamiltonian parameter estimates during the estimation procedure would be to adopt a Bayesian learning framework or stochastic process regression (e.g., Gaussian process regression). This is left for future work.

B.4 Learned Hamiltonian Parameters and Learning Error on IBM Quantum Devices

In this section of the Appendix, we give a summary of the estimated cross-resonance Hamiltonian parameters on different IBM Quantum devices not already discussed in Section 4.5 and lend further support to the performance of different learners on the 20-qubit IBM Quantum device D *ibmq_ boeblingen* under drive configuration 2 (Table 4.5.1).

B.4.1 Summary of Model Parameters on IBM Quantum Devices

In Section 4.5.1, we described the estimated Hamiltonian parameters and the noise model parameters for the IBM Quantum device D *ibmq_ boeblingen*. Here, we give a similar summary for the other IBM Quantum devices (A, B, and, C).

Considering the entire experimental datasets collected for each of these IBM Quantum devices (Section 4.4.2) as training data, we compute the Hamiltonian parameters using our estimation procedure (Section 4.4.4), and that of the different noise sources (Section 4.4.3). We summarize these parameters for the different devices in Table B.2. These Hamiltonian parameters serve as approximations of the true parameters $\mathbf{J}^\star$. Here, the readout noise parameters (r_0, r_1) are given to indicate the amount of readout noise possible in these devices and serve as a proxy for the conditional distributions of the readout given the measurement outcome used in the final MLE (Eq. 4.56). The time offset $\Delta t_{\text{eff},i}$ introduced due to imperfect control is also specified and where the subscript i indicates dependence on the preparation operator $U_0 = \sigma_I \sigma_I$ and $U_1 = \sigma_X \sigma_I$.

Device	Drive Config.	Hamiltonian Parameters [$\times 10^6 \mathrm{s}^{-1}$]		Noise: Readout and Time Offset	
		$\mathbf{J} = (J_{IX}, J_{IY}, J_{IZ}, J_{ZX}, J_{ZY}, J_{ZZ})$	(ω_0, ω_1)	(r_0, r_1)	$(\Delta t_{\mathrm{eff},0}, \Delta t_{\mathrm{eff},1})$ **[ns]**
A	0	$(58.47, 3.68, -5.00, 10.76, 2.29, -0.52)$	$(69.71, 47.94)$	$(0.160, 0.215)$	$(15, 101)$
	1	$(38.93, 2.34, 0.26, 11.15, -3.30, -0.30)$	$(50.09, 28.36)$	$(0.150, 0.210)$	$(97, 187)$
	2	$(19.35, 0.12, 0.69, 10.80, -0.66, 0.24)$	$(30.17, 8.60)$	$(0.220, 0.150)$	$(178, 699)$
	3	$(-0.21, -1.68, 0.20, 10.47, 1.50, -0.86)$	$(10.28, 11.20)$	$(0.145, 0.150)$	$(579, 532)$
	4	$(-20.11, -1.35, 0.73, 10.55, 0.94, -1.13)$	$(9.58, 30.80)$	$(0.190, 0.185)$	$(594, 174)$
B	0	$(30.03, 3.62, 0.49, 1.75, -0.16, -0.31)$	$(31.97, 28.54)$	$(0.110, 0.140)$	$(149, 175)$
	1	$(15.34, 1.85, 0.19, 1.81, -0.75, -0.59)$	$(17.19, 13.80)$	$(0.090, 0.070)$	$(318, 407)$
	2	$(0.89, 0.72, 0.24, 1.82, -0.54, -0.34)$	$(2.72, 1.67)$	$(0.120, 0.160)$	$(2235, 3712)$
	3	$(-13.71, -2.31, -0.54, 1.78, 0.09, 0.09)$	$(12.15, 15.69)$	$(0.130, 0.160)$	$(472, 353)$
	4	$(-28.45, -2.19, -1.20, 1.56, 2.15, 0.27)$	$(26.91, 30.35)$	$(0.110, 0.100)$	$(187, 161)$
$\mathrm{C/CR}_{01}$	0	$(-8.52, -2.15, -0.26, 10.93, 0.85, 0.32)$	$(2.74, 19.69)$	$(0.200, 0.160)$	$(2280, 286)$
	1	$(-3.88, -2.26, -0.35, 10.88, 1.46, 0.43)$	$(7.04, 15.24)$	$(0.120, 0.160)$	$(869, 376)$
	2	$(0.58, -1.81, -0.45, 10.81, 0.83, 1.24)$	$(11.46, 10.70)$	$(0.080, 0.070)$	$(518, 563)$
	3	$(4.86, -1.66, 0.08, 10.85, 0.44, -0.14)$	$(15.75, 6.35)$	$(0.070, 0.110)$	$(369, 966)$
	4	$(9.53, -0.17, 0.29, 10.76, -0.17, -0.32)$	$(20.29, 1.36)$	$(0.070, 0.120)$	$(276, 4835)$
$\mathrm{C/CR}_{02}$	0	$(9.42, -0.71, 0.27, 12.21, -0.71, -0.25)$	$(21.67, 2.84)$	$(0.070, 0.060)$	$(255, 2123)$
	1	$(6.17, -0.46, 0.09, 11.96, -0.59, -0.26)$	$(18.16, 5.81)$	$(0.070, 0.110)$	$(311, 1038)$
	2	$(2.59, 0.05, -0.26, 11.90, -1.66, -0.17)$	$(14.58, 9.46)$	$(0.050, 0.090)$	$(397, 629)$
	3	$(-1.03, -0.10, -0.16, 11.99, -0.63, -0.18)$	$(10.99, 13.03)$	$(0.090, 0.060)$	$(541, 445)$
	4	$(-4.53, 0.10, -0.31, 12.04, -0.18, 0.39)$	$(7.50, 16.59)$	$(0.100, 0.110)$	$(811, 344)$

Table B.2: Summary of estimated CR Hamiltonian parameters for the IBM Quantum devices A, B, and C. We give the Hamiltonian parameters in the parameterization $\mathbf{J}$ and the physically relevant frequency components in $\mathbf{\Lambda}$. The readout noise is defined by the parameters of r_0 and r_1 which are the conditional probabilities of bit flip given the measurement outcomes are $y = 0$ and $y = 1$ respectively (see Section 4.4.3). We show the results for CR Hamiltonians between two different pairs of qubits on Device C, specified as (control qubit, target qubit): $(0, 1)$ (CR_{01}) and $(0, 2)$ (CR_{02}).

As mentioned earlier in Section 4.4.3, we fit the estimated values of Δt_{eff} to the Hamiltonian parameters to obtain a model for the time-offset. This is shown in Figure B.7 and is used in the MLE (Eq. 4.56).

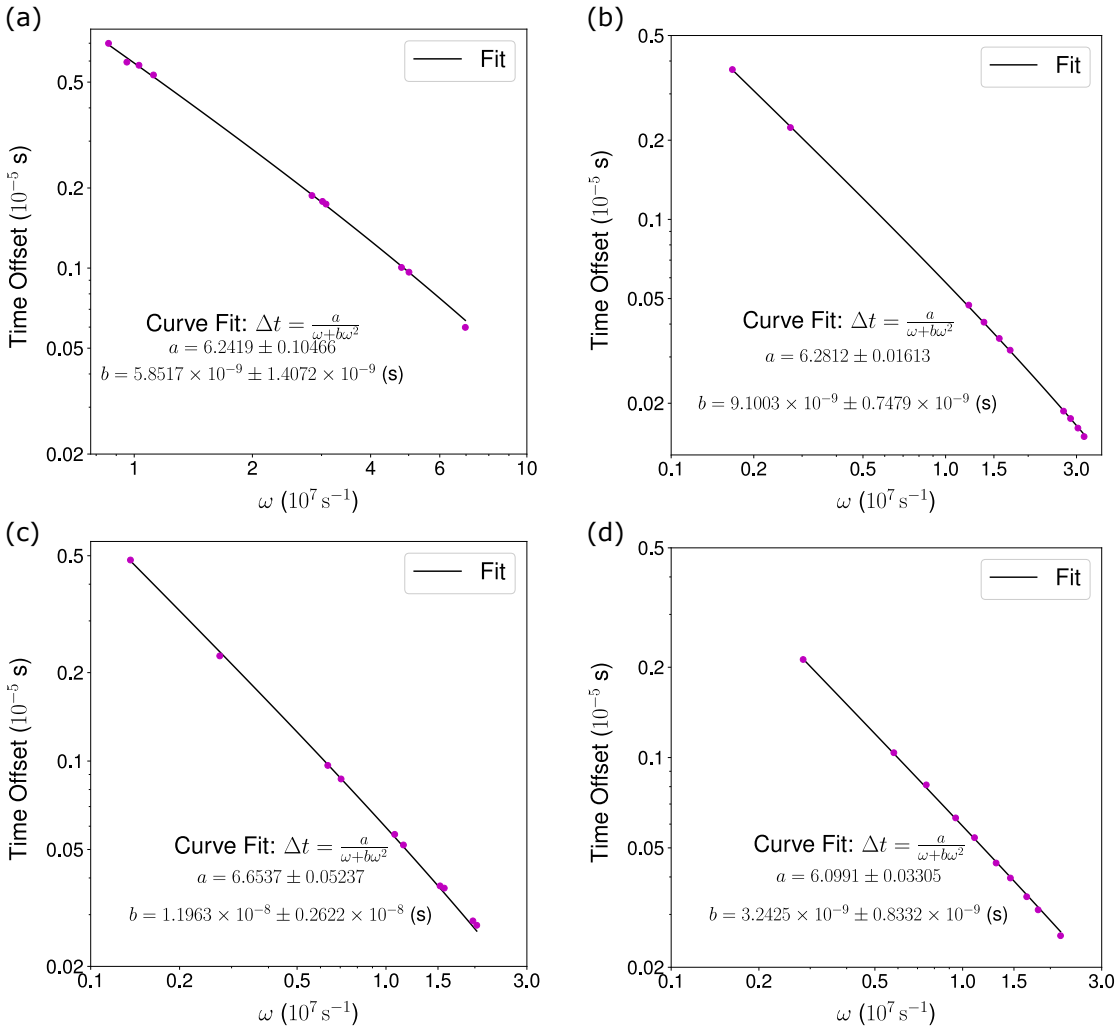


Figure B.7: Dependence of the time offset Δt on parameters ω for IBM Quantum devices (a) A, (b) B, (c) C CR_{01}, and (d) C CR_{02}. The plotted data points correspond to driving the device under different conditions and hence different cross-resonance Hamiltonians. The imperfect pulse shaping model extracted from these experimental data points is shown by a fit and this is later used in the MLE.

B.4.2 Expected trends of learning error

In Section 4.5.3, we assessed the performance of the HAL-FI and HAL-FIR algorithms in different learning scenarios on IBM Quantum device D *ibmq_boeblingen* under drive configuration 2, where the query distribution was learned in real-time. Here, we lend support that the trends observed in Figures 4.10–4.12 are expected.

To determine the behavior of the learners in an idealized setting, we consider the case where we have access to the optimal query distribution during learning. In Figure B.8, we show the trend of RMSE for HAL-FI with a fixed query space assuming access to the query distribution $q(\boldsymbol{\theta}^\star)$ during training and that the Cramer-Rao bound is saturated. We follow the same protocol from Section 4.5.1 as we did for our earlier experiments. For the baseline

strategy and a passive learner, the query distribution corresponds to an uniform distribution over the query space. As expected, using a passive learner does not change the scaling in the finite query nor the asymptotic query regimes. A scaling of $\epsilon \sim 1/\sqrt{N}$ or $N \sim \epsilon^{-2}$ is observed which is in line with the standard quantum limit (SQL). For HAL-FI, we observe an initial scaling in RMSE with number of queries which is higher than SQL but this reduces to SQL in the asymptotic query regime. Thus, our results from Section 4.5.3 is in agreement with what we observe here. Asymptotically, we expect a constant savings in the number of queries or resources required when employing an active learner with a fixed query space compared to a passive learner.

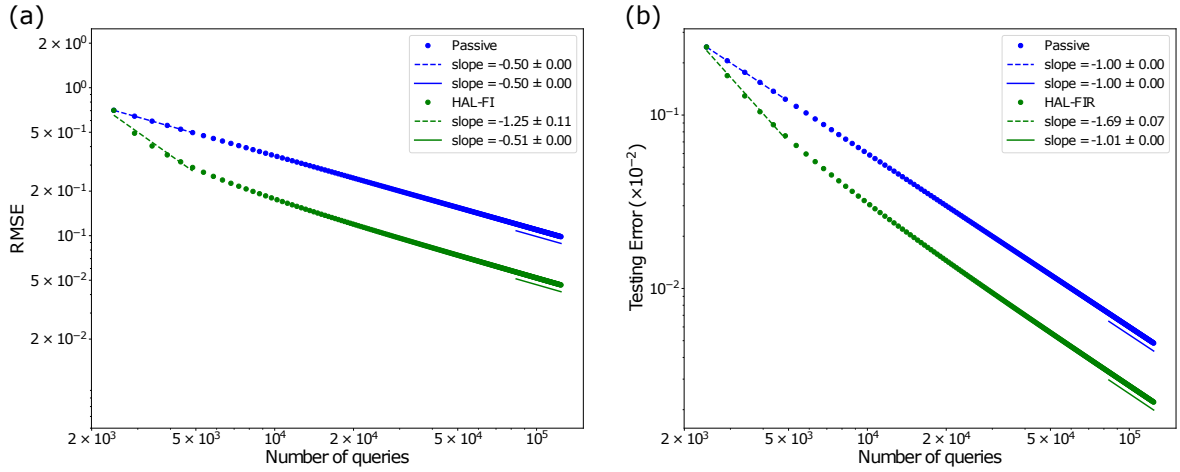


Figure B.8: Scaling of learning testing error with number of queries for Hamiltonian learning considering access to the asymptotic optimal query distribution $q(\boldsymbol{\theta}^\star)$ of HAL-FI or HAL-FIR. We show trends of (a) RMSE and (b) testing error obtained upon analysis of the lower Cramer-Rao bounds. We consider the Hamiltonian parameters of $\boldsymbol{\theta}^\star$ as determined from the experimental dataset of IBM Quantum device D *ibmq_ boeblingen* under drive configuration 2. We plot the trends of learning error with number of queries for HAL-FI/HAL-FIR against the passive learner which uses the uniform distribution over $\mathcal{Q}$.

Likewise for Hamiltonian learning with prior information, we can compute the RMSE with number of queries for HAL-FI with an adaptively growing query space and access to $q^{(i)}(\boldsymbol{\theta}^\star)$ for each ith batch during learning. In Figure B.9, we show the trend of RMSE with number of queries for HAL-FI with a linearly growing query space and an exponentially growing query space. We observe that the passive learner has a scaling of the SQL in the asymptotic query regime. HAL-FI in the linearly growing query space achieves super-Heisenberg scaling. We note that this supports the trend of RMSE achieved during the experiments in Section 4.5.4.

As noted in Section 4.5.4, HAL-FI with an exponentially growing query space also achieves Heisenberg limited scaling until the evolution times being included in the query space reach the magnitude of the decoherence time T_1 and T_2. In this case, HAL-FI avoids selecting higher evolution times as the information gained from these measurement outcomes will tend to zero. Thus, we expect that the Heisenberg limit to be achieved for the finite query setting.

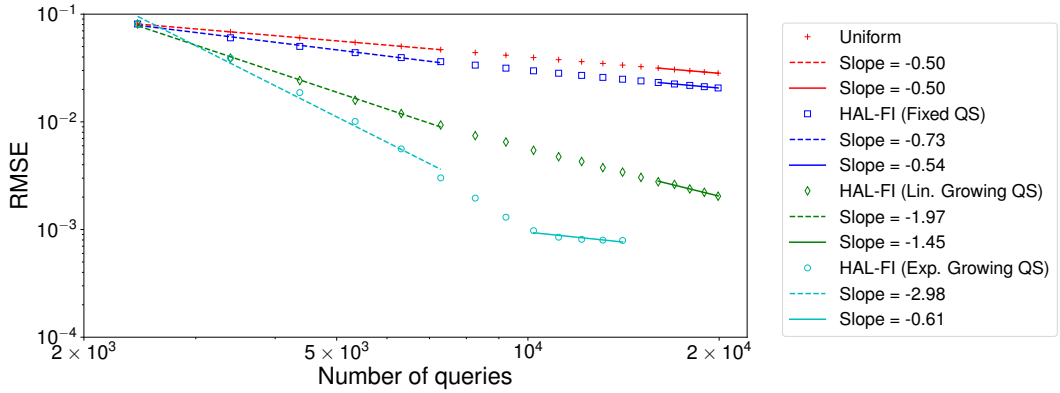


Figure B.9: Hamiltonian learning with acess to prior information of subset of parameters during recalibration: Scaling of RMSE with Number of Queries. We assume access to the asymptotic optimal query distribution $q(\boldsymbol{\theta}^\star)$ of HAL-FI and analysis of the Cramer-Rao Bounds. We consider the Hamiltonian parameters of $\boldsymbol{\theta}^\star$ as determined from the experimental dataset of IBM Quantum device D *ibmq_ boeblingen* under drive configuration 3.

B.4.3 Sparse Query Distributions

Another consequence of using HAL-FI is the sparsity of the asymptotic query distribution during learning. This is confirmed by visualizing the optimal query distribution of HAL-FI with the fixed query space (Section 4.4.2) considering IBM Quantum device *ibmq_ boeblingen* under drive configuration 2 in Figure B.10. It is interesting to note that this was achieved even though sparsity was not incorporated into the learning problem. This can be explained by realizing that the most informative queries are in fact sparse over the query space. It should be noted however that this is typically not the query distribution that HAL-FI has access to during learning as the true parameters $\boldsymbol{\theta}^\star$ are not available and the query distribution obtained through optimization (Eq. 4.19) is modified by mixing with the uniform distribution (see Section 4.3.1).

B.5 Heisenberg Limited Scaling in Cross-Resonance Type Hamiltonians

In this section of the Appendix, we discuss if Heisenberg limited scaling (HLS) can be achieved in different quantum systems with simplified cross-resonance Hamiltonians. The simplified cross-resonance Hamiltonians are obtained by removing particular Pauli product terms from the cross-resonance (CR) Hamiltonian (4.4.1) that we have studied so far. We give examples of such Hamiltonians (equipped with the query space described in Section 4.4) in Appendix B.5.1 where HLS is achieved for all the parameters and examples where HLS is not achieved for all the parameters during Hamiltonian learning in Section B.5.2. For examples of HLS, we describe query distributions obtained by HAL-FI and relate them to query distributions obtained through zero crossings of Rabi oscillations or maximum entropy.

In the following examples, we consider reduction to a single interaction in the two qubit

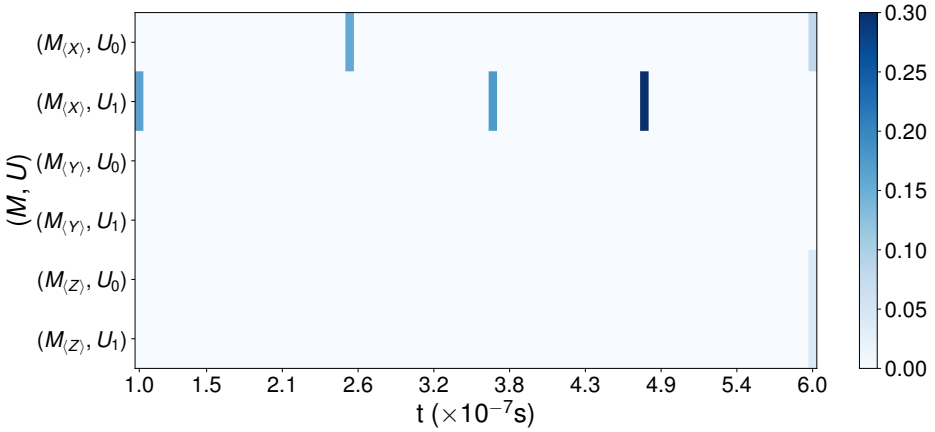


Figure B.10: Asymptotic optimal query distribution $q(\boldsymbol{\theta}^\star)$ for HAL-FI with a fixed query space (Section 4.4.2) on *ibmq_ boeblingen* under drive configuration 2 (Table 4.4). We consider different noise sources of readout noise, imperfect pulse-shaping, and decoherence. The y-axis indicates the different combinations of measurement operators and preparation operators available for each query in $\mathcal{Q}$. The different preparation operators are denoted as $U_0 = \sigma_I\sigma_I$ and $U_1 = \sigma_X\sigma_I$. The x-axis corresponds to $\mathcal{T}$ which is set to 81 equispaced evolution times in $[10^{-7}, 6 \times 10^{-7}]s$. The query distribution is color-coded according to the colormap on the right.

system and a three interaction example which we call the simplified cross-resonance (SCR) Hamiltonian. This reduction from the full CR Hamiltonian is achieved by setting the appropriate Hamiltonian parameters to zero and introducing the matrix R that denotes which parts of the CR Fisher information matrix are involved in the estimation of these respective Hamiltonian parameters. Suppose the reduced set of parameters are collected into the vector $\boldsymbol{\theta}_R$, then the Cramer-Rao bound is now

$$\sum_i \mathrm{Var}((\boldsymbol{\theta}_R)_i) \geq \frac{1}{N}\mathrm{Tr}(\mathcal{I}_q^{-1}(\boldsymbol{\theta}_R)) = \frac{1}{N}\mathrm{Tr}(R^{-T}\mathcal{I}_q^{-1}(\boldsymbol{\theta})R^{-1}) \tag{B.39}$$

where $\mathcal{I}_q(\boldsymbol{\theta}_R)$ is the reduced Fisher information matrix corresponding to the query distribution q. In the last step, we noted the relation of the reduced Fisher information matrix with the full Fisher information matrix as $\mathcal{I}_q(\boldsymbol{\theta}_R) = R\mathcal{I}_q(\boldsymbol{\theta})R^T$.

B.5.1 Examples

Single Interaction Two-Qubit System

The Hamiltonian of interest in this case is

$$H = J_{ZX}\sigma_Z \otimes \sigma_X \tag{B.40}$$

This may be obtained by considering the parameter set of $\mathbf{J} = (0, 0, 0, J_{ZX}, 0, 0)^T$ and $R = [0, 0, 0, 1, 0, 0]$. The values of the parameters in the alternate parameterization of

$\mathbf{\Lambda} = (\omega_0, \delta_0, \phi_0, \omega_1, \delta_1, \phi_1)^T = (J_{ZX}, 0, 0, J_{ZX}, 0, \pi)$ where we have assumed $J_{ZX} > 0$. The Rabi oscillations for different queries in this case are as follows:

$$(M_{\langle X\rangle}, U_j, t) : p_{\text{rabi}}(x) = 0 \tag{B.41}$$
$$(M_{\langle Y\rangle}, U_j, t) : p_{\text{rabi}}(x) = \sin(2\omega_j t) \tag{B.42}$$
$$(M_{\langle Z\rangle}, U_j, t) : p_{\text{rabi}}(x) = \cos(2\omega_j t) \tag{B.43}$$

where the index $j \in \{0, 1\}$ corresponds to different preparation operators $U_0 = \sigma_I\sigma_I$ and $U_1 = \sigma_X\sigma_I$. We note that the measurement operator of $M_{\langle X\rangle}$ is not informative about the frequency $\omega_0 = J_{ZX}$ for this system and can also be noted from considering the corresponding Fisher information. In order to learn the parameter of interest ω_0, it is enough to consider one of the queries in $\{M_{\langle Y\rangle}, M_{\langle Z\rangle}\} \times \{\sigma_I\sigma_I, \sigma_X\sigma_I\}$ and a suitable time range $\mathcal{T}$.

Let us select the query of $(M_{\langle Z\rangle}, \sigma_I\sigma_I, t)$ and suppose our query distribution over the time range is based on the zeros the Rabi oscillations. Given a time range $\mathcal{T}$, we consider values of $t_k = \frac{\pi}{4\omega_0} + \frac{k\pi}{2\omega_0}$ where $k \in \mathbb{N}$. Queries with these system evolution times have the maximum entropy for the considered (M, U).

Note that the Fisher information of a query in this case is given by $\mathcal{I}_x(\omega_0) = 4t^2$. Through the Cramer-Rao bound, we then have

$$\epsilon^2 = \text{Var}(\omega_0) \geq \sum_{k=1}^{N} \frac{1}{4t_k^2} \approx \begin{cases} \frac{1}{2}\frac{\omega^2}{N\pi^2} & , \text{ Fixed space of zero crossings} \\ \frac{3}{4}\frac{\omega^2}{N^3\pi^2} & , \text{ Linearly spaced zero crossings} \end{cases} \tag{B.44}$$

where we have set the learning error to be achieved as ϵ. We can thus expect to achieve a scaling of $N \sim \epsilon^{-3/2}$ when using linearly spaced zero crossings. For exponentially spaced zero crossings of the Rabi oscillations, the variance approaches zero at an increasing rate.

Two Interaction Systems

Consider the following Hamiltonian

$$H = J_{IX}\sigma_I \otimes \sigma_X + J_{ZX}\sigma_Z \otimes \sigma_X \tag{B.45}$$

The reduced set of parameters is then $\mathbf{\Lambda}_R = (\omega_0, \omega_1) = (|J_{IX} + J_{ZX}|, |J_{IX} - J_{ZX}|)$. As in the earlier single interaction example, we can choose the queries that contain the zero crossings of the Rabi oscillations. The Rabi oscillations are given by

$$(M_{\langle X\rangle}, U_j, t) : p_{\text{rabi}}(x) = 0 \tag{B.46}$$
$$(M_{\langle Y\rangle}, U_j, t) : p_{\text{rabi}}(x) = \sin(2\omega_j t) \tag{B.47}$$
$$(M_{\langle Z\rangle}, U_j, t) : p_{\text{rabi}}(x) = \cos(2\omega_j t) \tag{B.48}$$

where $j \in \{0, 1\}$ is used as an index to denote the preparation operators $U_0 = \sigma_I\sigma_I$ and $U_1 = \sigma_X\sigma_I$. A complete set of queries to estimate (ω_0, ω_1) with Heisenberg limited scaling would then be

$$\mathcal{Q} = \left\{(M_{\langle Z\rangle}, \sigma_I\sigma_I, t_k) : t_k = \frac{\pi}{4\omega_0} + \frac{k\pi}{2\omega_0}, k \in \mathbb{N}\right\} \bigcup \left\{(M_{\langle Z\rangle}, \sigma_X\sigma_I, t_k) : t_k = \frac{\pi}{4\omega_1} + \frac{k\pi}{2\omega_1}, k \in \mathbb{N}\right\} \tag{B.49}$$

The set of evolution times chosen here also correspond to those with maximum entropy. The Fisher information of a query made through either set of measurement or preparation operators in the query space is given by $\mathcal{I}_x(\omega_0, \omega_1) = 4t^2$. Thus, through the argument we made in the previous section, we can also learn the parameters of unitary here with Heisenberg limited scaling.

B.5.2 Examples of non-HLS scaling during Hamiltonian learning

So far, we have given examples of Hamiltonians obtained through simplification of the CR Hamiltonian, that can be learned with HLS scaling. We now give examples of Hamiltonians, which cannot be learned with HLS scaling using the query space described in Section 4.4.2.

Two Interaction Systems

Now, let us consider the following alternate Hamiltonian (modified slightly from the previous example discussed)

$$H = J_{IY}\sigma_I \otimes \sigma_Y + J_{ZX}\sigma_Z \otimes \sigma_X \tag{B.50}$$

where a complete reduced set of parameters is $\mathbf{\Lambda}_R = (\omega_0, \phi_0) = \left(\sqrt{J_{ZX}^2 + J_{IY}^2}, \tan^{-1}\left(\frac{J_{IY}}{J_{ZX}}\right)\right)$. This set of parameters contains a frequency in addition to a phase. Fisher information matrices in the $\mathbf{\Lambda}_R$ parameterization considering queries of the form $(M, \sigma_I\sigma_I, t)$ where we select one particular preparation operator is given by

$$M_{\langle X\rangle} : \mathcal{I} = \frac{1}{1 - \sin^2(\phi_0)\sin^2(2\omega_0 t)} \begin{bmatrix} 4t^2\sin^2(\phi_0)\cos^2(2\omega_0 t) & \frac{1}{2}t\sin(2\phi_0)\sin(4\omega_0 t) \\ \frac{1}{2}t\sin(2\phi_0)\sin(4\omega_0 t) & \cos^2(\phi_0)\sin^2(2\omega_0 t) \end{bmatrix} \tag{B.51}$$

$$M_{\langle Y\rangle} : \mathcal{I} = \frac{1}{1 - \cos^2(\phi_0)\sin^2(2\omega_0 t)} \begin{bmatrix} 4t^2\cos^2(\phi_0)\cos^2(2\omega_0 t) & -\frac{1}{2}t\sin(2\phi_0)\sin(4\omega_0 t) \\ -\frac{1}{2}t\sin(2\phi_0)\sin(4\omega_0 t) & \sin^2(\phi_0)\sin^2(2\omega_0 t) \end{bmatrix} \tag{B.52}$$

$$M_{\langle Z\rangle} : \mathcal{I} = \begin{bmatrix} 4t^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{B.53}$$

Fisher information matrices in $\mathbf{J}_R$ is given by

$$\mathcal{I}_J = \begin{bmatrix} \frac{J_{IY}^2}{\omega_0^2}\mathcal{I}_{11} + \frac{2J_{IY}J_{ZX}}{\omega_0^3}\mathcal{I}_{12} + \frac{J_{ZX}^2}{\omega_0^4}\mathcal{I}_{22} & \frac{J_{ZX}}{\omega_0^3}\mathcal{I}_{12} - \frac{J_{IY}}{\omega_0^3}\mathcal{I}_{12} + \frac{J_{IY}J_{ZX}}{\omega_0^2}(\mathcal{I}_{11} - \mathcal{I}_{12}) \\ . & \frac{J_{ZX}^2}{\omega_0^2}\mathcal{I}_{11} - \frac{2J_{IY}J_{ZX}}{\omega_0^3}\mathcal{I}_{12} + \frac{J_{IY}^2}{\omega_0^4}\mathcal{I}_{22} \end{bmatrix} \tag{B.54}$$

where we have related them to elements of $\mathcal{I}$ as given above in Eq. B.53. We observe that in order to obtain HLS in (J_{IY}, J_{ZX}), it is necessary to set $\mathcal{I}_{22} = 0$ and $\mathcal{I}_{12} \neq 0$ to ensure $\mathcal{I}_J$ is full rank and there is an explicit dependence on the variable t that we can take advantage of, for HLS. However, the required conditions cannot be achieved simultaneously here. This suggests that the current set of queries cannot be used to achieve HLS.

Three Interaction Simplified Cross-Resonance Gate

The Hamiltonian of interest in this case is

$$H = J_{IX}\sigma_I \otimes \sigma_X + J_{IY}\sigma_I \otimes \sigma_Y + J_{ZX}\sigma_Z \otimes \sigma_X \tag{B.55}$$

This may be obtained by considering the parameter set of $\mathbf{J} = (J_{IX}, J_{IY}, 0, J_{ZX}, 0, 0)^T$. Defining R is a bit more tricky in this case compared to the previous example. Let us first look at the alternate parameterization of

$$\mathbf{\Lambda} = \begin{bmatrix} \sqrt{(J_{IX}+J_{ZX})^2 + J_{IY}^2} \\ 0 \\ \tan^{-1}\left(\frac{J_{IY}}{J_{IX}+J_{ZX}}\right) \\ \sqrt{(J_{IX}-J_{ZX})^2 + J_{IY}^2} \\ 0 \\ \tan^{-1}\left(\frac{J_{IY}}{J_{IX}-J_{ZX}}\right) \end{bmatrix}. \tag{B.56}$$

Note that the reduced parameterization of Λ_R is an over-parameterization with four non-zero components compared to $\mathbf{J}_R$ which only has three non-zero components. The Rabi oscillations in this case are given by

$$(M_{\langle X\rangle}, U_j, t) : p_{\text{rabi}}(x) = \sin(\phi_j)\sin(2\omega_j t) \tag{B.57}$$
$$(M_{\langle Y\rangle}, U_j, t) : p_{\text{rabi}}(x) = \cos(\phi_j)\sin(2\omega_j t) \tag{B.58}$$
$$(M_{\langle Z\rangle}, U_j, t) : p_{\text{rabi}}(x) = \cos(2\omega_j t) \tag{B.59}$$

The Fisher information matrices can be obtained by looking at the CR Fisher information matrices (Appendix B.1) and simplifying them.

$$M_{\langle X\rangle} : \mathcal{I}_j = \frac{1}{1-\sin^2(\phi_j)\sin^2(2\omega_j t)} \begin{bmatrix} 4t^2\sin^2(\phi_j)\cos^2(2\omega_j t) & \frac{1}{2}t\sin(2\phi_j)\sin(4\omega_j t) \\ \frac{1}{2}t\sin(2\phi_j)\sin(4\omega_j t) & \cos^2(\phi_j)\sin^2(2\omega_j t) \end{bmatrix} \tag{B.60}$$
$$M_{\langle Y\rangle} : \mathcal{I}_j = \frac{1}{1-\cos^2(\phi_j)\sin^2(2\omega_j t)} \begin{bmatrix} 4t^2\cos^2(\phi_j)\cos^2(2\omega_j t) & -\frac{1}{2}t\sin(2\phi_j)\sin(4\omega_j t) \\ -\frac{1}{2}t\sin(2\phi_j)\sin(4\omega_j t) & \sin^2(\phi_j)\sin^2(2\omega_j t) \end{bmatrix} \tag{B.61}$$
$$M_{\langle Z\rangle} : \mathcal{I}_j = \begin{bmatrix} 4t^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{B.62}$$

If we were to consider the zero crossings of the Rabi oscillations as in the previous example, the queries and their corresponding Fisher information matrices are of the following form

$$M_{\langle X\rangle} : t_k(M_{\langle X\rangle}) = \frac{\pi}{2\omega_j} + \frac{k\pi}{2\omega_j},\ \mathcal{I}_j = \begin{bmatrix} 4t_k^2\sin^2(\phi_j) & 0 \\ 0 & 0 \end{bmatrix} \tag{B.63}$$
$$M_{\langle Y\rangle} : t_k(M_{\langle Y\rangle}) = \frac{\pi}{2\omega_j} + \frac{k\pi}{2\omega_j},\ \mathcal{I}_j = \begin{bmatrix} 4t_k^2\cos^2(\phi_j) & 0 \\ 0 & 0 \end{bmatrix} \tag{B.64}$$
$$M_{\langle Z\rangle} : t_k(M_{\langle Z\rangle}) = \frac{\pi}{4\omega_j} + \frac{k\pi}{2\omega_j},\ \mathcal{I}_j = \begin{bmatrix} 4t_k^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{B.65}$$

where $k \in \mathbb{N}$. It should be noted that the evolution times $t_k(M)$ being selected are a function of the measurement operator involved in the query which is made explicit through the argument M. As $\mathbf{\Lambda}_R$ is an over-parameterization, let us look at the query Fisher information matrix $\mathcal{I}_q(\mathbf{J}_R)$ for the above set of queries.

$$\mathcal{I}_q(\mathbf{J}_R) = \sum_k \sum_{M\in\{M_{\langle X\rangle}, M_{\langle Z\rangle}\}} 4t_k^2(M) \begin{bmatrix} \left(\frac{\partial\omega_0}{\partial J_{IX}}\right)^2 + \left(\frac{\partial\omega_1}{\partial J_{IX}}\right)^2 & \left(\frac{\partial\omega_0}{\partial J_{IX}}\frac{\partial\omega_0}{\partial J_{IY}} + \frac{\partial\omega_1}{\partial J_{IX}}\frac{\partial\omega_1}{\partial J_{IY}}\right) & \left(\frac{\partial\omega_0}{\partial J_{IX}}\frac{\partial\omega_0}{\partial J_{ZX}} + \frac{\partial\omega_1}{\partial J_{IX}}\frac{\partial\omega_1}{\partial J_{ZX}}\right) \\ & \left(\frac{\partial\omega_0}{\partial J_{IY}}\right)^2 + \left(\frac{\partial\omega_1}{\partial J_{IY}}\right)^2 & \left(\frac{\partial\omega_0}{\partial J_{IY}}\frac{\partial\omega_0}{\partial J_{ZX}} + \frac{\partial\omega_1}{\partial J_{IY}}\frac{\partial\omega_1}{\partial J_{ZX}}\right) \\ & & \left(\frac{\partial\omega_0}{\partial J_{ZX}}\right)^2 + \left(\frac{\partial\omega_1}{\partial J_{ZX}}\right)^2 \end{bmatrix} \tag{B.66}$$

where we have only given the upper-triangular part of the symmetric matrix. It can be shown that for these queries, $\mathcal{I}_q(\mathbf{J}_R)$ is rank deficient and thus non-invertible. This was foreshadowed by the fact that $\mathcal{I}_q(\mathbf{\Lambda})$ was informative in ω_0 and ω_1 but not one of the phases $\phi_{0,1}$. Hence, it is more appropriate to consider $R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ for these set of queries. It can be verified through an analysis similar to the single interaction system that we can achieve a scaling of $N \sim O(\epsilon^{-3/2})$ and hence make improvements over than SQL.

If we wish to learn J_{IY} as well, it is necessary to introduce other queries such that the Fisher information matrix is non-zero for the corresponding parameter of interest. Let us start by changing our learning task to the simpler challenge of learning the parameters (ω_0, δ_0). In this case, it is enough to consider only queries of the form $(M_{\langle X \rangle}, \sigma_I \otimes \sigma_I, t)$ where the time range $t \in \mathcal{T}$ needs to be determined. We immediately observe that

$$\mathcal{I}_q(\mathbf{\Lambda}_R)^{-1} \propto \sum_k \frac{1}{1 - \sin^2(\phi_0)\sin^2(2\omega_0 t_k)} \begin{bmatrix} \cos^2(\phi_j)\sin^2(2\omega_j t) & -\frac{1}{2}t\sin(2\phi_j)\sin(4\omega_j t) \\ -\frac{1}{2}t\sin(2\phi_j)\sin(4\omega_j t) & 4t_k^2\sin^2(\phi_j)\cos^2(2\omega_j t) \end{bmatrix} \quad \text{(B.67)}$$

and the variance of parameter ϕ_0

$$\mathrm{Var}(\phi_0) \geq \frac{1}{N}\sum_{k=1}^{N} \frac{1 - \sin^2(\phi_0)\sin^2(2\omega_0 t_k)}{\cos^2(\phi_0)\sin^2(2\omega_0 t_k)} \quad \text{(B.68)}$$

where the term inside the sum on the right hand side is fixed for any periodic or equi-spaced set of evolution times t_k and thus HLS cannot be achieved using such a set of queries. One key to ensure achieving Heisenberg limited scaling is to introduce an explicit dependence on the variable of system evolution time t into the corresponding Fisher information. We note that this is not followed by the different set of measurement operators considered here.

Appendix C

Appendix to Chapter 5

In Section C.1, we give details of the measurement method derandomization of decision diagrams which was introduced in Section 5.4.3 of the main paper. Specifically, we show how the associated cost function is obtained and how it may be computed efficiently on decision diagrams corresponding to molecular Hamiltonians. In Section C.2, we give details of the molecular Hamiltonians from Table 5.2 including their Pauli weight distributions and the query distributions obtained for these molecular Hamiltonians via LBCS or decision diagrams. Finally, we also comment on the convergence behavior of Adaptive Pauli Shadows (APS) as observed on the simulator.

C.1 Details of derandomizing decision diagrams

In this section of the Appendix, we give the technical details of derandomizing decision diagrams (Derand DD) which was proposed in Section 5.4.3 of the main paper. We will first show how the cost function for Derand DD is obtained by considering the general query distribution of the decision diagram to be β and then show how the cost function can be computed efficiently on a decision diagram. As discussed in Section 5.4.3, our starting point is the *confidence bound* introduced in [HKP21]. For completeness, we show the proof in [HKP21, Lemma 2] to motivate the confidence bound and then proceed to obtain the cost function for Derand DD.

Confidence bound on estimates. Recall that we denote the estimates $\mathrm{Tr}(\rho Q^{(j)})$ as $\hat{\omega}^{(j)}$ and their true value as $\omega^{(j)}$. We analyze the probability of a large deviation of the estimates

obtained from post-processing on the measurement outcomes against bases $\mathbf{B}$

$$\Pr\left[\max_{j\in[L]}|\hat{\omega}^{(j)}-\omega^{(j)}|\geq\epsilon\right] = \Pr\left[\bigcup_{j\in[L]}|\hat{\omega}^{(j)}-\omega^{(j)}|\geq\epsilon\right] \tag{C.1}$$

$$\leq \sum_{j\in[L]}\Pr\left[|\hat{\omega}^{(j)}-\omega^{(j)}|\geq\epsilon\right] \quad \text{(Union bound)} \tag{C.2}$$

$$\leq 2\sum_{j\in[L]}\exp\left(-\frac{\epsilon^2}{2}h(Q^{(j)};\mathbf{B})\right) \quad \text{(Hoeffding's inequality)} \tag{C.3}$$

$$= 2\sum_{j\in[L]}\prod_{s\in[M]}\exp\left(-\frac{\epsilon^2}{2}\mathbb{1}\{Q^{(j)}\triangleright B^{(s)}\}\right) \tag{C.4}$$

$$= 2\sum_{j\in[L]}\prod_{s\in[M]}\left(1-\eta\mathbb{1}\{Q^{(j)}\triangleright B^{(s)}\}\right), \tag{C.5}$$

where $\eta = 1-\exp(-\epsilon^2/2)$. We see that the probability of the estimate $\hat{\omega}^{(j)}$ deviating from the truth reduces exponentially with the number of measurement bases that hit/cover the target observable $Q^{(j)}$. We will call the upper bound without the constant factor (as derived above) of this probability as the confidence bound:

$$\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B}) = \sum_{j\in[L]}\prod_{s\in[M]}\left(1-\eta\mathbb{1}\{Q^{(j)}\triangleright B^{(s)}\}\right). \tag{C.6}$$

Expectation of the confidence bound. To obtain the cost function for derandomization, it is desirable to compute the expectation of the confidence bound which is given by

$$\mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})\right] = \sum_{j\in[L]}\prod_{s\in[M]}\mathbb{E}\left[\left(1-\eta\mathbb{1}\{Q^{(j)}\triangleright B^{(s)}\}\right)\right] \tag{C.7}$$

$$= \sum_{j\in[L]}\left(1-\eta\mathbb{E}\left[\mathbb{1}\{Q^{(j)}\triangleright B^{(s)}\}\right]\right)^M \tag{C.8}$$

$$= \sum_{j\in[L]}\left(1-\eta\xi(Q^{(j)},\beta)\right)^M, \tag{C.9}$$

where the expectation is with respect to the query distribution β, we have used the fact that each $B^{(s)}$ is sampled independently and identically from the distribution β for all $s\in[M]$, and where $\xi(Q^{(j)},\beta) = \Pr[Q^{(j)}$ covered by $\beta]$ denoting the coverage probability (notation introduced in Section 5.3.2). If the query distribution $\beta = \prod_{k\in[n]}\beta_k$ is a product distribution (with the marginal distribution the kth qubit dented as β_k) as in the case of CS and LBCS, we have the following simplified expression

$$\mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})\right] = \sum_{j\in[L]}\left(1-\eta\prod_{k=1}^{n}\beta_k(Q_k^{(j)})^{\mathbb{1}\{Q_k^{(j)}\neq I\}}\right)^M. \tag{C.10}$$

Cost function for derandomization. Let us now discuss how a cost function for derandomization of decision diagrams may be obtained using the confidence bound. Suppose $\mathbf{B}^{\#}$ contains the assignments of measurement bases for the first $(m-1)$ samples and first k qubits of the mth measurement basis. We then have the following conditional expectation

$$\begin{aligned}\mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})|\mathbf{B}^{\#}\right] = \sum_{j\in[L]}\prod_{m'=1}^{m-1}\left(1-\eta\mathbb{1}\{Q^{(j)}\triangleright B^{\#^{(m')}}\}\right) \\ \times\left(1-\eta\prod_{k'=1}^{k}\mathbb{1}\{Q^{(j)}\triangleright B^{\#^{(m')}}\}\Pr\left[Q^{(j)}_{k+1:n}\text{ covered by }\beta|B^{\#^{(m)}}_{1:k}\right]\right) \\ \times\left(1-\eta\Pr[Q^{(j)}\text{ covered by }\beta]\right)^{M-m},\end{aligned} \tag{C.11}$$

where we have denoted $B^{\#^{(m)}}$ as the mth measurement basis in $B^{\#}$, $B^{\#^{(m)}}_k$ as the kth qubit Pauli of the mth measurement basis, used the sub-scripted Pauli operator $Q^{(j)}_{k+1:n}$ to denote $Q^{(j)}_{k+1:n} = \otimes_{\ell=k+1}^{n} Q^{(j)}_\ell$ and similarly $B^{\#^{(m)}}_{1:k} = \otimes_{\ell=1}^{k} B^{\#^{(m)}}_\ell$. In the above expression, we also have $\Pr\left[Q^{(j)}_{k+1:n}\text{ covered by }\beta|B^{\#^{(m)}}_{1:k}\right] = \xi(\otimes_{\ell=1}^{k}B^{\#^{(m)}}_\ell \otimes_{\ell'=k+1}^{n} Q^{(j)}_{\ell'},\beta)$.

In the special case of the query distribution β being a production distribution (e.g., as in LBCS), we have the following expression for the conditional expectation of the confidence bound

$$\begin{aligned}\mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})|\mathbf{B}^{\#}\right] = \sum_{j\in[L]}\prod_{m'=1}^{m-1}\left(1-\eta\mathbb{1}\{Q^{(j)}\triangleright B^{\#^{(m')}}\}\right) \\ \times\left(1-\eta\prod_{k'=1}^{k}\mathbb{1}\{Q^{(j)}\triangleright B^{\#^{(m')}}\}\prod_{k'=k+1}^{n}\beta_{k'}(Q^{(j)}_{k'})^{\mathbb{1}\{Q^{(j)}_{k'}\neq I\}}\right) \\ \times\left(1-\eta\prod_{k=1}^{n}\beta_k(Q^{(j)}_k)^{\mathbb{1}\{Q^{(j)}_k\neq I\}}\right)^{M-m}.\end{aligned} \tag{C.12}$$

To choose the assignment of the kth qubit of the mth measurement basis, we then consider the following cost function

$$B^{\#^{(m)}}_k = \underset{W\in\{X,Y,Z\}}{\mathrm{argmin}}\ C(W) = \underset{W\in\{X,Y,Z\}}{\mathrm{argmin}}\ \mathbb{E}\left[\mathrm{CONF}_\epsilon(\mathbf{Q};\mathbf{B})|\mathbf{B}^{\#}, B^{(m)}_k = W\right] \tag{C.13}$$

where $\mathbf{B}^{\#}$ now corresponds to the assignments of measurement bases over the first $(m-1)$ samples and $(k-1)$ qubits of the mth measurement basis. Note that the above cost function requires the input of the experimental budget M. As was done in [HKP21] for derandomizing CS, we can remove the dependence on the measurement budget in the cost function by removing the third term in the product of Eq. C.12.

Derandomization of a general query distribution β is then given by Algorithm 8 shown in the main part of the paper.

Fast computation of the cost function on decision diagrams. As part of computing the cost function C.13 quickly, we need to be able to perform quick computation of the

conditional probability $\Pr\left[Q^{(j)}_{k+1:n} \text{ covered by } \beta | B^{\#(m)}_{1:k}\right]$ in Eq. C.12. We will refer to this as the conditional probability of coverage.

To compute the conditional probability of coverage, let us now introduce some relevant notation and useful routines. We will refer to the decision diagram using its graph $G = (V, E)$ where we have denoted the set of nodes as V and edges as E. We will refer to the directed edge from node u to node v as $E(u, v)$. As we mentioned in Section 5.4, each edge has a label associated with it, namely the Pauli decision being taken. This will be denoted by $P(u, v)$.

For each node $v \in V$, we define the coverable set C_v as the non-identity Pauli terms in H that can be covered by measurements starting from the root of the decision diagram (which we denote as r and for example, corresponds to node 0 in the decision diagram of Figure 5.3(c)). For example, in Figure 5.3(c), the target observable $ZIII$ will be a part of the coverable set C_9 as the first qubit Pauli Z is covered by the edge $E(0, 9)$. We compute the coverable sets for each node recursively.

The coverable sets at the root node r (or node 0 in Figure 5.3) and at the terminal node t (or node -1 in Figure 5.3) are set to be all the non-identity Pauli operators in the decomposition of the Hamiltonian H i.e., $C_r = C_t := \mathbf{Q} \setminus I^{\otimes n}$. Suppose node u is the child of node r, then $C_u = \{Q | Q \in C_r, Q_1 \in \{P(r, u), I\}\}$. We can do this for the nodes corresponding to the children of the root node but generally, nodes may have multiple parents. To obtain coverage for a general node v, we move down the decision diagram one layer of edges at a time and update the coverage of a node v in layer ℓ based on its parents in the previous layer $\ell - 1$ as $C_v = \bigcup_{u \in \text{Parents}(v)} \{Q | Q \in C_u, Q_\ell \in \{P(u, v), I\}\}$.

Now, given that the current state of a measurement basis being proposed is at node v, we can compute the conditional probability $\Pr[Q_{k+1:n} \text{ covered by DD } |v]$ (where we have denoted the query distribution β in Eq. C.12 by its DD) of covering any $Q \in C_v$ in a recursive fashion as

$$\begin{aligned}\Pr[Q_{k:n} \text{ covered by DD } |v, Q \in C_v] = &\sum_{w \in \text{Children}(w)} \mathbb{1}\{Q_k \triangleright P(v, w)\} \\ &\times \Pr[P(v, w)] \Pr[Q_{k+1:n} \text{ covered by DD } |w, Q \in C_W],\end{aligned} \tag{C.14}$$

where we have used $\Pr[P(v, w)]$ to denote the probability of taking the decision of Pauli measurement $P(v, w)$ that is available to us from the DD. We should note that there at most three children for any node v corresponding to the three decisions of the single-qubit Paulis $\{X, Y, Z\}$. Finally, to compute the conditional coverage probability $\Pr\left[Q^{(j)}_{k+1:n} \text{ covered by } \beta | B^{\#(m)}_{1:k}\right]$, it is enough to note that making the sequence of measurements in $B^{\#(m)}_{1:k}$ will place us at a node v in the DD. This node v will be unique. We can argue this by noting that if taking the decisions in $B^{\#(m)}_{1:k}$ starting from the root node took us down to two different nodes, then these two nodes would have been merged from the initialization of the DD.

Overall, as part of the fast computation of the cost function in derandomization on DDs, we firstly move from the root node downwards to the terminal node to determine the node coverages C_v, and then secondly move from the terminal node upwards to the root node determine the conditional probabilities of coverage.

C.2 Details of molecules and measurement methods

In this section of the Appendix, we describe the Pauli distributions of the different molecular Hamiltonians considered in Table 5.2 and illustrate query distributions obtained for these molecules via LBCS or DD. The Pauli decompositions of the Hamiltonians considered in this paper are available in our code repository.

C.2.1 Pauli weight distributions

The sample complexity of classical shadows [HKP20] is known to be $O(3^w \log L)$ where w is the maximum weight of any Pauli term in the Pauli decomposition of H. However, in practice, the dependence of the sample complexity on w may be better for some measurement methods depending on mass of the query distribution on higher weight Paulis and the coefficients of higher weight Paulis in H.

Thus, it might be desirable to include Hamiltonians of the same size but with different Pauli weight distributions as part of benchmarking measurement methods. Here, in Figure C.1, we visualize the Pauli weight distributions of the Hamiltonians from Table 5.2 which range from unimodal to skewed to bimodal.

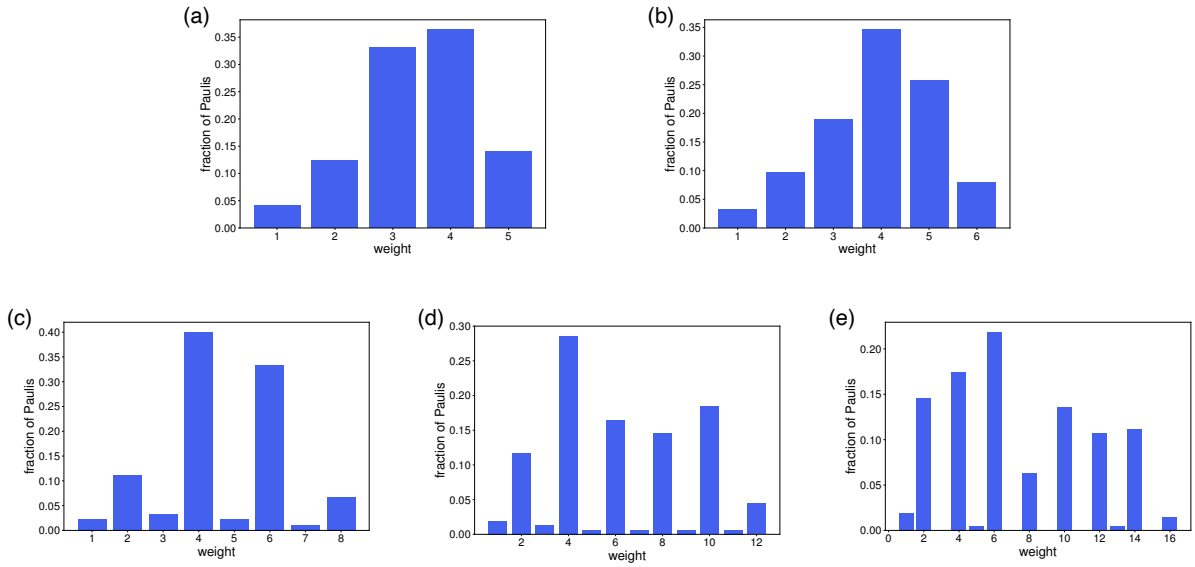


Figure C.1: Pauli weight distributions across non-identity target Paulis in Hamiltonians of different molecules. Molecules correspond to those in Table 5.2 and as shown are (a) H_2(5 qubits, 3-21g, JW), (b) HeH^+ (6 qubits, 3-21g, JW), (c) HeH^+ (8 qubits, 6-31g, JW), (d) LiH (12 qubits, sto6g, JW), and (e) N_2 (16 qubits, sto6g, JW).

C.2.2 Query distributions for different molecules

We now visualize different instances of query distributions as obtained from the measurement methods of LBCS and decision diagrams. In Figure C.2, we show the product query distributions obtained in LBCS by optimizing the variance of the energy estimate considering the state to be a maximally mixed state (Section 5.4).

Instead of showing illustrations of the decision diagrams (DD), we present relevant details of the optimized compact decision diagrams for different molecules in Table C.1. It should be noted that the number of paths for each DD corresponds to the maximum number of unique measurement circuits that are being considered for each molecule. For example, the DD of LiH only considers 810 unique measurement circuits in contrast to CS for LiH which would consider $3^{12} \approx 5.3 \times 10^5$ measurement circuits for high measurement budgets. Reducing the unique number of measurement circuits can help in reducing classical latencies as we mentioned in Section 5.5 and as observed in Section 5.6.3. As in the case of LBCS, the decision diagrams for each molecule are optimized by minimizing the variance of the energy estimate considering the state to be the maximally mixed state. We refer to resulting cost obtained from the optimized DD as the diagonal cost and is shown in Table C.1 with that for LBCS given as a reference. Finally, constructing and optimizing the decision diagram can be expensive. Hence, we report the classical computational runtimes required in obtaining the decision diagrams used in this work and in computing the classical pre-processing times as part of resource utilization (Tables 5.4,5.6).

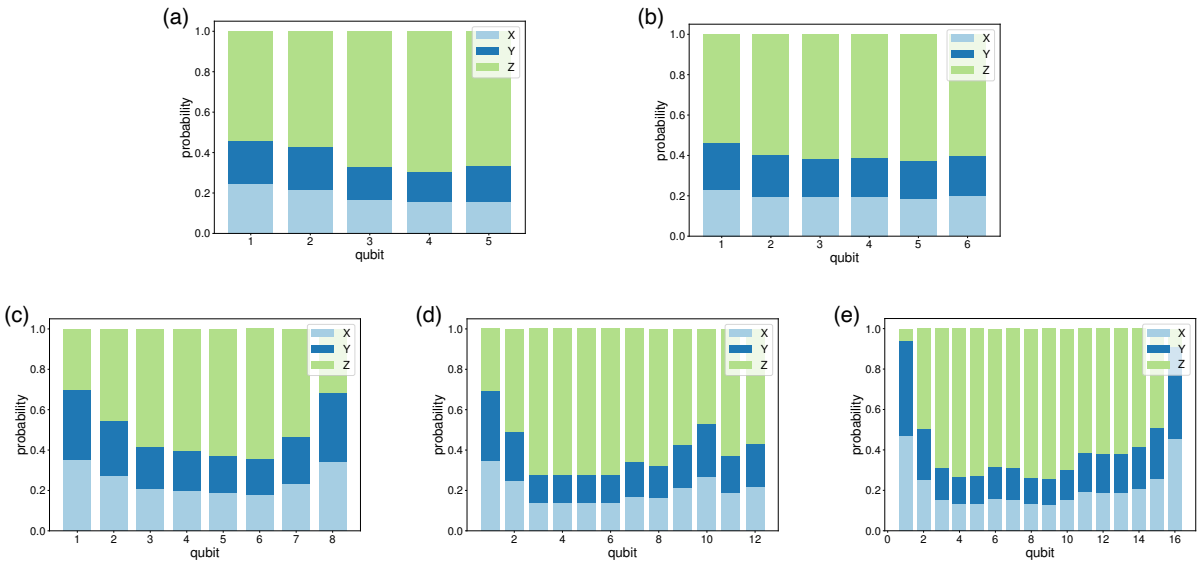


Figure C.2: Query distributions in LBCS for Hamiltonians of different molecules. As the query distribution in LBCS is a product distribution, the marginal probability distribution over single-qubit Paulis are shown for each qubit. Molecules correspond to those in Table 5.2 and as shown are (a) H_2(5 qubits, 3-21g, JW), (b) HeH^+ (6 qubits, 3-21g, JW), (c) HeH^+ (8 qubits, 6-31g, JW), (d) LiH (12 qubits, sto6g, JW), and (e) N_2 (16 qubits, sto6g, JW).

Molecule	| Nodes |	| Edges |	| Paths |	Diagonal cost of Dec. Diag.	Diagonal cost of LBCS (for reference)	Computational runtime of constructing DD [s]
H_2, 5 qubits (3-21g, JW)	40	81	43	14.63	26.59	2.88×10^1
HeH^+, 6 qubits (3-21g, JW)	75	155	174	17.11	35.56	5.45×10^2
HeH^+, 8 qubits (6-31g, JW)	168	337	792	16.69	42.14	3.54×10^3
LiH, 12 qubits (sto6g, JW)	348	600	810	25.45	42.44	2.56×10^4
N_2, 16 qubits (sto6g, JW)	1118	1551	1137	10836.12	12121.62	1.39×10^5

Table C.1: Details of decision diagrams for different molecular Hamiltonians (Table 5.2). Number of paths in a decision diagram correspond to number of unique measurement circuits and contributes to classical latencies such as compilation time and circuit loading. Diagonal cost of a query distribution corresponds to the one-shot variance of the energy estimate considering ρ to be the maximally mixed state. The reported computational runtimes account for both initialization and optimization of the decision diagrams.

C.2.3 Convergence behavior of Adaptive Pauli Shadows

In Figure C.3, we plot the trend of RMSE in estimating ground state energy of the tapered Hamiltonians from Table 5.2 in numerical simulation, considering for the measurement method of Adaptive Pauli Shadows (APS) and other measurement methods. It has been shown earlier in [Had21] that APS outperforms other measurement methods at the low measurement budget of 10^3 shots and this is observed here as well. However, this does not continue for higher measurement budgets and we instead observe a weird convergence behavior for APS. Such behavior has been reported earlier in [Shl+23]. Hence, APS was not included as a candidate measurement method in CSHOREBench.

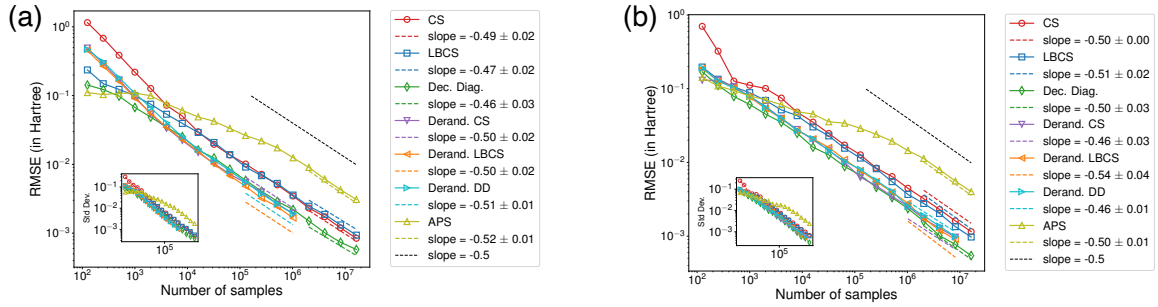


Figure C.3: Comparison of RMSE achieved in numerical simulations by different measurement methods including APS in estimating $\mathrm{Tr}(\rho H)$ with ρ set as the ground state and H is the Hamiltonian of (a) tapered H_2 (5 qubits, 3-21g basis, JW encoding), and (b) tapered HeH^+ (6 qubits, 3-21g basis, JW encoding). RMSE is shown with the number of samples made. The estimator for each measurement method is set to be the Bayesian estimator.

Appendix D

Appendix to Chapter 6

D.1 Construction of the Embedding Hamiltonians

The one- and two-electron integrals $h_{pq}^{(A)}$ and $V_{pqrs}^{(A)}$ of the embedding Hamiltonian for a chosen fragment A in Eq. (6.5) of the main text can be obtained from transforming the original integrals of the total system by a projector $T^{(A)}$

$$\begin{aligned} h_{pq}^{(A)} &= \sum_{\mu\nu}^{N} T_{\mu p}^{(A)} F_{\mu\nu}^{(\mathrm{env},A)} T_{\nu q}^{(A)} \\ V_{pqrs}^{(A)} &= \sum_{\mu\nu\lambda\sigma}^{N} T_{\mu p}^{(A)} T_{\nu q}^{(A)} V_{\mu\nu\lambda\sigma} T_{\lambda r}^{(A)} T_{\sigma s}^{(A)}, \end{aligned} \tag{D.1}$$

where $F^{(\mathrm{env},A)}$ is the Fock matrix of the environment and $T^{(A)}$ is a projection matrix that transform the one- and two-electron integrals in Eq. (6.1) to the EO basis. In practice, $T^{(A)}$ can be obtained by performing an SVD of the off-diagonal Hartree-Fock density matrix $P_{\mu\nu}^{(A)}$,

$$P_{\mu\nu}^{(A)} = U_A \Sigma_A V_A^{\dagger} \tag{D.2}$$

$$T^{(A)} = \left[\begin{array}{c|c} I & \\ \hline & U_A \end{array}\right]. \tag{D.3}$$

The integral transformations in Eq. (D.1) can be performed efficiently on classical computers.

D.2 Basis Transformation Unitary from Localized to Canonical Molecular Orbitals

Bootstrap embedding requires *spatially* local information on overlapping sites of a quantum system to match, whereas this local basis may not be the same as the most convenient computational basis. For example, many state preparation ansatz on a quantum computer is designed to work the best under the canonical molecular orbitals (delocalized, obtained from a Hartree-Fock calculation) as the computational basis, where a basis transformation

to local atomic basis are required afterwards to extract spatially local information. In this section, we give a unitary that can achieve such a basis transformation from MOs to LOs on a quantum computer, from an early result due to Thouless [Tho60] which is also employed in a recent work [Qua+20].

Given two set of single-particle orbitals, $\{\psi_p\}$ and $\{\phi_q\}$ to represent the MOs and the LOs, respectively. Furthermore, denote the basis rotation between the two as

$$\phi_p = \sum_{q=1}^{n} [e^{ih}]_{pq} \ \psi_q \tag{D.4}$$

where h is an $n \times n$ hermitian matrix, $[e^{ih}]_{pq}$ is the (p,q)-th element of the orbital rotation unitary. For notation purpose, we also associate each MO ψ_q with a creation and annhilation operator $a_p^\dagger, a_p$.

This orbital rotation from the MOs to the LOs will induce a unitary transformation on the Slater determinants written under these two set of orbitals. More concretely, suppose $|\Psi\rangle$ and $|\Phi\rangle$ are representation of the same mean-field state using orbitals $\{\psi_p\}$ and $\{\phi_q\}$, respectively, then it can be shown [Tho60] that the transformation between $|\Psi\rangle$ and $|\Phi\rangle$ is essentially a unitary operator generated by a 1-body operator in the MO basis

$$|\Phi\rangle = U_h |\Psi\rangle \tag{D.5}$$

where

$$U_h = e^{i \sum_{pq=1}^{n} h_{pq} a_p^\dagger a_q}. \tag{D.6}$$

The above results on the transformation between two Slater determinants can be easily generalized to two arbitrary many-body quantum states $|\mathbf{\Phi}\rangle$ and $|\mathbf{\Psi}\rangle$ that represent the same underlying quantum state using the two different orbital sets, as we will show in the following. We can always write $|\mathbf{\Phi}\rangle$ as a linear combination of many Slater determinants constructed from the orbibital set $\{\phi_p\}$

$$|\mathbf{\Phi}\rangle = \sum_{\alpha} C_\alpha |\Phi_\alpha\rangle. \tag{D.7}$$

Using the unitary transformation U_h for each Slater determinant in $|\mathbf{\Phi}\rangle$, we have

$$|\mathbf{\Phi}\rangle = \sum_{\alpha} C_\alpha |\Phi_\alpha\rangle = \sum_{\alpha} C_\alpha U_h |\Psi_\alpha\rangle = U_h |\mathbf{\Psi}\rangle, \tag{D.8}$$

where we have defined a new many-body state $|\mathbf{\Psi}\rangle$ using the same many-body coefficients C_α but with the old MO Slater determinants $|\Psi_\alpha\rangle$. Equivalently, this means the transformation of a many-body state under orbital rotations follows the same unitary U_h.

On a quantum computer, in the case of Jordan-Wigner mapping, there is a direct one-to-one correspondence between Slater determinants and qubit states. Therefore the unitary transformation on a quantum computer to transform a state from MO to LO representation is to write U_h in its Jordan-Wigner form using

$$a_p^\dagger = \sigma_p^+ \otimes Z_{p-1}^{\rightarrow}, \quad a_q = \sigma_q^- \otimes Z_{q-1}^{\rightarrow}, \tag{D.9}$$

where

$$Z_j^{\rightarrow} = Z_j \otimes Z_{j-1} \otimes \cdots \otimes Z_1 \otimes Z_0. \tag{D.10}$$

Note that the hermitian matrix h can be obtained from standard quantum chemistry package such as PySCF [Sun+18].

D.3 Proof of Equivalence of the Linear and Quadratic Constraint

In this section, we prove the equivalent of the linear and the quadratic constraints in Eqs. (6.16) and (6.20).

Recall that in the main text (Eq. (6.15)), in a general case of m overlapping qubits, the mixed state reduced density matrices $\rho_R^{(A)}$ can be written as

$$\rho_R^{(A)} = \frac{I + \sum_{\alpha=1}^{4^m-1} \langle \Sigma_\alpha \rangle_A \, \Sigma_\alpha}{2^m}. \tag{D.11}$$

In the special case of $m = 1$, we recover the usual expression for a single-qubit density matrix.

The forward direction of deriving Eq. (6.20) from (6.16) is trivial, because if $\langle \Sigma_\alpha \rangle_A = \langle \Sigma_\alpha \rangle_B$ for all α, this means $\rho_R^{(A)} = \rho_R^{(B)}$ which leads to Eq. (6.20).

Now we focus on showing the reverse is true by deriving Eq. (6.16) from (6.20). Substitute (D.11) into (6.20), we obtain

$$\mathrm{Tr}[(\rho_R^{(A)} - \rho_R^{(B)})^2] = \frac{1}{2^m} \sum_{\alpha,\beta=1}^{4^m-1} (\langle \Sigma_\alpha \rangle_A - \langle \Sigma_\alpha \rangle_B)(\langle \Sigma_\beta \rangle_A - \langle \Sigma_\beta \rangle_B) \mathrm{Tr}[\Sigma_\alpha \Sigma_\beta]. \tag{D.12}$$

Choose a convenient basis for the Hermitian generators Σ_α by express then as tensor product of m SU(2) Paulis $\{\sigma_{\alpha_m}\}$

$$\Sigma_\alpha = \sigma_{\alpha_1} \otimes \sigma_{\alpha_2} \otimes \cdots \otimes \sigma_{\alpha_m}, \tag{D.13}$$

we immediately see that

$$\begin{aligned}
\mathrm{Tr}[\Sigma_\alpha \Sigma_\beta] &= \mathrm{Tr}[(\sigma_{\alpha_1}\sigma_{\beta_1}) \otimes (\sigma_{\alpha_2}\sigma_{\beta_2}) \otimes \cdots \otimes (\sigma_{\alpha_m}\sigma_{\beta_m})] && \text{(D.14)} \\
&= \prod_{s=1}^{m} \mathrm{Tr}[\sigma_{\alpha_s}\sigma_{\beta_s}] && \text{(D.15)} \\
&= \prod_{s=1}^{m} (2\delta_{\alpha_s,\beta_s}) && \text{(D.16)} \\
&= 2^m \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}} && \text{(D.17)}
\end{aligned}$$

where we define a composite index (bold font) $\boldsymbol{\alpha} = (\alpha_1\alpha_2\cdots\alpha_k)$ and $\boldsymbol{\beta} = (\beta_1\beta_2\cdots\beta_k)$. Substitute (D.17) into (D.12) and note that $\delta_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \delta_{\alpha\beta}$, we obtain

$$\mathrm{Tr}[(\rho_R^{(A)} - \rho_R^{(B)})^2] = \sum_{\alpha=1}^{4^m-1} \left(\langle \Sigma_\alpha \rangle_A - \langle \Sigma_\alpha \rangle_B\right)^2. \tag{D.18}$$

Given $\langle \Sigma_\alpha \rangle$'s as real numbers in usual implementation of electronic structure problems, then (D.18) guarantees each individual term in the sum being zero, i.e.,

$$\langle \Sigma_\alpha \rangle_A - \langle \Sigma_\alpha \rangle_B = 0, \quad \forall\, i \in [1, 4^m - 1]. \tag{D.19}$$

This is equivalent to $\rho_R^{(A)} = \rho_B^{(B)}$. This completes the proof. □

D.4 Estimating Quadratic Penalty from Subsystem SWAP Test

In this section, we present details of the subsystem SWAP test and discuss how it can be used to estimate the quadratic penalty mismatch in (6.20) of the main text.

D.4.1 Quantum Circuit of the SWAP Test

The SWAP test as shown in Fig. 6.4 of the main text between two qubits can be directly generalized to a SWAP test between two quantum registers each of which contains multiple qubits. The idea is to use the upper ancilla qubit to perform multiple controlled-SWAP operations between all pairs of qubits in the two registers. For example, Fig. D.5 performs a SWAP test between two pairs of qubits using a single ancilla qubit.

In our case, instead of performing SWAP test on the entire wave function of two fragments, we are interested to apply it to a subsystem of each fragment. In particular, denote qubits corresponding to the entire embedding orbitals as X_1 X_0 for fragment $X = A, B$, where A_0 and B_0 are the subsystem on the overlapping region, whereas A_1 and B_1 are the rest of the embedding orbitals. Then a SWAP test between A_0 and B_0 can be performed in Fig. D.1.

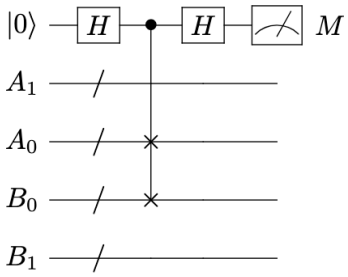


Figure D.1: Subsystem SWAP test between the overlapping regions of fragment A and B, where a controlled SWAP operation is performed on A_0 and B_0. The measurement probability of the top ancilla qubit encodes information of the overlap.

D.4.2 Ancilla Measurement Probability

In the following, we show that the measurement probability of M in the upper ancilla qubit can be directly related to the overlap between the reduced density matrices on region A_0 and B_0. This derivation can be performed in an arbitrary computational basis. For simplicity,

we derive the results in the Schmidt basis of A_0 and A_1 (likewise for B). Interest readers are encouraged to perform the derivation in an arbitrary basis as an exercise.

From Schmidt decomposition, the total wave function for fragment A and B can be written as

$$|\Psi_A\rangle = \sum_j a_j |A_{1,j}\rangle |A_{0,j}\rangle, \tag{D.20}$$

$$|\Psi_B\rangle = \sum_k b_k |B_{0,k}\rangle |B_{1,k}\rangle. \tag{D.21}$$

Denote the measurement outcome of the ancilla in Fig. D.1 as M, it can be shown that

$$\text{Prob}[M=0] = \frac{1}{2}[1 + \sum_{j,k} |a_j|^2 \ |b_k|^2 \ |\langle A_{0,j}|B_{0,k}\rangle|^2], \tag{D.22}$$

regardless of their environment A_1 and B_1.

The above result can be understood as the overlap of the two reduced density matrix of fragments A and B in their overlapping region, as we can explicitly compute this in the following. From Eq. (D.20) and (D.21), the reduced density matrices of each is calculated to be

$$\begin{aligned}
\rho_{A_0} =& \text{Tr}_{A_1}[|\Psi_A\rangle\langle\Psi_A|] \\
=& \text{Tr}_{A_1}[\sum_{jj'} a_j a_{j'}^* |A_{1,j}\rangle |A_{0,j}\rangle\langle A_{1,j'}|\langle A_{0,j'}|] \\
=& \sum_k \sum_{jj'} a_j a_{j'}^* \langle A_{1,k}||A_{1,j}\rangle |A_{0,j}\rangle\langle A_{1,j'}|\langle A_{0,j'}||A_{1,k}\rangle \\
=& \sum_k \sum_{jj'} a_j a_{j'}^* \delta_{jk} |A_{0,j}\rangle\langle A_{0,j'}|\delta_{j'k} \\
=& \sum_j |a_j|^2 |A_{0,j}\rangle\langle A_{0,j}|,
\end{aligned} \tag{D.23}$$

similarly,

$$\rho_{B_0} = \text{Tr}_{B_1}[|\Psi_B\rangle\langle\Psi_B|] = \sum_k |b_k|^2 |B_{0,k}\rangle\langle B_{0,k}|. \tag{D.24}$$

Then the overlap S^2 between the above two reduced density matrices are

$$\begin{aligned}
& S^2(\rho_{A_0}, \rho_{B_0}) = \text{Tr}[\rho_{A_0}\rho_{B_0}] \\
=& \text{Tr}[\sum_{jk} |a_j|^2 |b_k|^2 |B_{0,k}\rangle\langle B_{0,k}||A_{0,j}\rangle\langle A_{0,j}|] \\
=& \sum_l \sum_{jk} |a_j|^2 |b_k|^2 \langle A_{0,l}||B_{0,k}\rangle\langle B_{0,k}||A_{0,j}\rangle\langle A_{0,j}||A_{0,l}\rangle \\
=& \sum_{jk} |a_j|^2 |b_k|^2 |\langle A_{0,j}|B_{0,k}\rangle|^2,
\end{aligned} \tag{D.25}$$

which agrees with the second term in the measurement probability in Eq. (D.22). Therefore, we can reconstruct the overlap in (D.25) using the measurement probability as

$$S^2(\rho_{A_0}, \rho_{B_0}) = 2\ \text{Prob}[M = 0] - 1. \tag{D.26}$$

In the special case of $\rho_{A_0} = \rho_{B_0}$, the overlap as defined above reduces to evaluating the purity of a density matrix

$$\begin{aligned} S^2(\rho_{A_0}, \rho_{A_0}) =& \text{Tr}[\rho_{A_0}^2] \\ =& \sum_{jk} |a_j|^2 |a_k|^2 |\langle A_{0,j}|A_{0,k}\rangle|^2 \\ =& \sum_{jk} |a_j|^2 |a_k|^2 \delta_{jk} \\ =& \sum_{j} |a_j|^4 \end{aligned} \tag{D.27}$$

Since $\sum_j |a_j|^2 = 1$, and therefore $S^2(\rho_{A_0}, \rho_{A_0}) \leq 1$. When ρ_{A_0} corresponds to a pure state, there will be only one non-zero coefficient $a_0 = 1$ and the rest being zero, leading to $S^2(\rho_{A_0}, \rho_{A_0}) = 1$. This agrees with the definition of purity.

D.4.3 Connection to Quadratic Penalty

In our quantum bootstrap embedding algorithm, the quadratic constraint can be rewritten using the definition of overlap in the previous section as

$$\begin{aligned} \text{Tr}[(\rho_{A_0} - \rho_{B_0})^2] &= \text{Tr}[(\rho_{A_0})^2] + \text{Tr}[(\rho_{B_0})^2] - 2\text{Tr}[\rho_{A_0}\rho_{B_0}] \\ &= S^2(\rho_{A_0}, \rho_{A_0}) + S^2(\rho_{B_0}, \rho_{B_0}) - 2S^2(\rho_{A_0}, \rho_{B_0}). \end{aligned} \tag{D.28}$$

It is seen that the RHS of the above equation contains three `SWAP` tests: one for the overlap between ρ_{A_0} and ρ_{B_0} in (D.25), and the other two for the purity of ρ_{A_0} and ρ_{B_0} respectively as in (D.27).

D.5 Eigenvalue Equations for the Quadratic Penalty Method

In Sec. 6.3.4 of the main text, we have seen that the quadratic constraint at the solution point has zero gradient with respect to the wave function parameters, and therefore does not satisfy the constraint qualification condition to use the Lagrange multiplier method for the optimization. Instead, we use a quadratic penalty method to perform the optimization of the loss function where an inherent eigenvalue equation with an effective bootstrap embedding potential V_{BE} is solved using a quantum eigensolver. The goal is to adjust V_{BE} such that two overlapping fragments match. In this section, we derive a rigorous expression for updating V_{BE} by taking the parameters in V_{BE} instead of the wave function as fundamental variable. For clarity, we keep the notation $\lambda_B^{(A)}$ to refer to the penalty for fragment A and B, but all the penalty parameters are kept the same, $\lambda_B^{(A)} = \lambda$, as is also mentioned in the main text.

This can be achieved by taking the functional variation $\delta V_{\rm BE}$ and find the stationary point of $\mathcal{L}_A$. At the stationary point, we have

$$\delta\mathcal{L}_A = \sum_{ai\mu} \frac{\partial \mathcal{L}_A}{\partial C_{ai\mu}} \frac{\partial C_{ai\mu}}{\delta V_{\rm BE}} \delta V_{\rm BE} + \sum_{ai\mu} \frac{\partial \mathcal{L}_A}{\partial C^*_{ai\mu}} \frac{\partial C^*_{ai\mu}}{\delta V_{\rm BE}} \delta V_{\rm BE} = 0 \tag{D.29}$$

for any $\delta V_{\rm BE}$. This implies the equation of motion

$$\sum_{ai\mu} \frac{\partial \mathcal{L}_A}{\partial C_{ai\mu}} \frac{\partial C_{ai\mu}}{\delta V_{\rm BE}} + \sum_{ai\mu} \frac{\partial \mathcal{L}_A}{\partial C^*_{ai\mu}} \frac{\partial C^*_{ai\mu}}{\delta V_{\rm BE}} = 0. \tag{D.30}$$

In the following, we derive separately $\frac{\partial \mathcal{L}_A}{\partial C_{ai\mu}}$ (Sec. D.5.1) and $\frac{\partial C_{ai\mu}}{\delta V_{BE}}$ (Sec. D.5.2), and then combine everything together in Sec. D.5.3 to obtain the overall gradient on how to updating the BE potential in the eigenvalue equation.

D.5.1 Derivative $\frac{\partial \mathcal{L}_A}{\partial C_{ai\mu}}$

Consider the derivative of $\partial\mathcal{L}_A$ with respect to $C_{ai\mu}$:

$$\frac{d\mathcal{L}_A}{dC_{ai\mu}} = \frac{d}{dC_{ai\mu}} \left[\langle H^{(A)} \rangle_A - \mathcal{E}(\langle I \rangle_A - 1) + \sum_{B\neq A}^{N_{frag}} \lambda_B^{(A)} {\rm Tr}[(\rho_{\mathbb{E}_A} - \rho_{\mathbb{C}_B})^2] \right]. \tag{D.31}$$

Let's evaluate this term by term. The linear terms are easy:

$$\frac{d\langle H^{(A)} \rangle}{C^*_{bj\nu}} = \sum_{ai\mu} C_{ai\mu} \langle bj\nu | H^{(A)} | ai\mu \rangle, \tag{D.32}$$

$$\frac{d\mathcal{E}(\langle I \rangle - 1)}{C^*_{bj\nu}} = \mathcal{E} \sum_{ai\mu} C_{ai\mu} (I \otimes I \otimes I)_{ai\mu, bj\nu}, \tag{D.33}$$

$$\frac{d{\rm Tr}[\rho_{\mathbb{E}_A} \rho_{\mathbb{C}_B}]}{dC^*_{bj\nu}} = \sum_i C_{bi\nu} \left(\sum_{a\mu} D_{ja\mu} D^*_{ia\mu} \right) = \sum_i (\rho_{\mathbb{C}_B})_{ji} C_{bi\nu}. \tag{D.34}$$

While the derivative involving quadratic term $\frac{d{\rm Tr}[\rho^2_{\mathbb{E}_A}]}{C^*_{ck\delta}}$ is a little tricky. First, note that

$$(\rho_{\mathbb{E}_A})^2 = \sum_{ii',jj'} \left(\sum_{a\mu} C_{ai\mu} C^*_{ai'\mu} \right) \left(\sum_{b\nu} C_{bj\nu} C^*_{bj'\nu} \delta_{i'j} |i\rangle\langle j'| \right), \tag{D.35}$$

from which we obtain

$$\begin{aligned}
{\rm Tr}[(\rho_{\mathbb{E}_A})^2] &= \sum_{ij} \left(\sum_{a\mu} C_{ai\mu} C^*_{aj\mu} \right) \left(\sum_{b\nu} C_{bj\nu} C^*_{bi\nu} \right) && \text{(D.36)} \\
&= \sum_{ij,ab,\mu\nu} \left(C_{ai\mu} C_{bj\nu} \right) \left(C^*_{aj\mu} C^*_{bi\nu} \right) && \text{(D.37)} \\
&= \sum_{ij,ab,\mu\nu, i\neq j || a\neq b || \mu\neq\nu} \left(C_{ai\mu} C_{bj\nu} \right) \left(C^*_{aj\mu} C^*_{bi\nu} \right) + \sum_{ia\mu} \left(C_{ai\mu} C_{ai\mu} \right) \left(C^*_{ai\mu} C^*_{ai\mu} \right). && \text{(D.38)}
\end{aligned}$$

Note in order to take the derivative of the above with respect to $C^*_{ck\delta}$, we have separated the summation as two different terms because depending on whether the conditon of $a \neq b || j \neq i || \mu \neq \nu$ is met or not, the derivative will be different. Now evaluate the derivative of the above two terms separately, we have

$$\begin{aligned}\frac{d\mathrm{Tr}[\rho^2_{\mathbb{E}_A}]}{dC^*_{ck\delta}} &= \sum_{ij,ab,\mu\nu,i\neq j||a\neq b||\mu\neq\nu} \left(C_{ai\mu}C_{bj\nu}\right)\left(C^*_{aj\mu}\delta_{bc}\delta_{ik}\delta_{\nu\delta} + C^*_{bi\nu}\delta_{ac}\delta_{jk}\delta_{\mu\delta}\right) \\ &\quad + \sum_{ia\mu}\left(C_{ai\mu}C_{ai\mu}\right)\left(2C^*_{ai\mu}\delta_{ac}\delta_{jk}\delta_{\mu\delta}\right) \qquad \text{(D.39)} \\ &= \sum_{ja\mu,k\neq j||a\neq c||\mu\neq\delta}\left(C_{ak\mu}C_{cj\delta}\right)C^*_{aj\mu} + \sum_{ib\nu,i\neq k||b\neq c||\nu\neq\delta}\left(C_{ci\delta}C_{bk\nu}\right)C^*_{bi\nu} + 2C^2_{ck\delta}C^*_{ck\delta}. \qquad \text{(D.40)}\end{aligned}$$

Now in each of the above terms, combine a C and a C^* we can recover some elements of $\rho_{\mathbb{E}_A}$. For example, in the first term we have

$$\begin{aligned}\sum_{ja\mu,k\neq j||a\neq c||\mu\neq\delta}\left(C_{ak\mu}C_{cj\delta}\right)C^*_{aj\mu} &= \sum_{j,k\neq j||a\neq c||\mu\neq\delta}\left(\sum_{a\mu}C_{ak\mu}C^*_{aj\mu}\right)C_{cj\delta} \qquad \text{(D.41)} \\ &= \sum_{j,k\neq j||a\neq c||\mu\neq\delta}\left(\rho^{(a\mu)}_{\mathbb{E}_A}\right)_{kj}C_{cj\delta}, \qquad \text{(D.42)}\end{aligned}$$

where in the last line the superscript $^{(a\mu)}$ on $\rho^{(a\mu)}_{\mathbb{E}_A}$ simply means the implicit summation is over dummy variables $a\mu$.

With this notation, we can collect all terms in $\frac{d\mathrm{Tr}[\rho^2_{\mathbb{E}_A}]}{C^*_{ck\delta}}$ and write it as

$$\frac{d\mathrm{Tr}[\rho^2_{\mathbb{E}_A}]}{dC^*_{bj\nu}} = 2\sum_{i,k\neq i||b\neq c||\nu\neq\delta}\left(\rho^{(b\nu)}_{\mathbb{E}_A}\right)_{ki}C_{ci\delta} + 2\left(\rho_A\right)_{ck\delta,ck\delta}C_{ck\delta}, \qquad \text{(D.43)}$$

which consists of two terms. With the derivative of $\mathrm{Tr}[\rho^2_{\mathbb{E}_A}]$, we can combine this with the derivative of $\mathrm{Tr}[\rho_{\mathbb{E}_A}\rho_{\mathbb{C}_B}]$ to get

$$\begin{aligned}\frac{d\mathrm{Tr}[(\rho_{\mathbb{E}_A}-\rho_{\mathbb{C}_B})^2]}{dC^*_{ck\delta}} &= \frac{d\mathrm{Tr}[(\rho_{\mathbb{E}_A})^2]}{C^*_{ck\delta}} - 2\frac{d\mathrm{Tr}[\rho_{\mathbb{E}_A}\rho_{\mathbb{C}_B}]}{C^*_{ck\delta}} \\ &= 2\sum_{i,k\neq i||b\neq c||\nu\neq\delta}\left(\rho^{(b\nu)}_{\mathbb{E}_A} - \rho^{(b\nu)}_{\mathbb{C}_B}\right)_{ki}C_{ci\delta} + 2\left[\left(\rho_A\right)_{ck\delta,ck\delta} - \left(\rho_B\right)_{kc\delta,kc\delta}\right]C_{ck\delta}. \qquad \text{(D.44)}\end{aligned}$$

Notice the subscripts of ρ_B on the last term of RHS is in different order as compared to ρ_A due to the distinction between center and edge sites.

Combine this with the derivative of the other terms, we obtain the following eigenvalue

equation

$$\frac{d\mathcal{L}_A}{dC^*_{ck\delta}} = \sum_{ai\mu} C_{ai\mu}\langle ck\delta|H^{(A)}|ai\mu\rangle + \mathcal{E}\sum_{ai\mu} C_{ai\mu}(I\otimes I\otimes I)_{ck\delta,ai\mu}$$
$$+ 2\lambda_B^{(A)}\left[\sum_{i,(a,i,\mu)\neq(c,k,\delta)}\left(\rho^{(a\mu)}_{\mathbb{E}_A} - \rho^{(a\mu)}_{\mathbb{C}_B}\right)_{ki} C_{ci\delta} + \left[(\rho_A)_{ck\delta,ck\delta} - (\rho_B)_{kc\delta,kc\delta}\right] C_{ck\delta}\right] = 0,\ \forall c,k,\delta. \tag{D.45}$$

This equation seems to be difficult to rewrite into matrix notation, but actually they are easy if written under the full density matrix of fragment A and B. In terms of full density matrices of the fragments, the first term in the effective potential is

$$\sum_{i,(a,i,\mu)\neq(c,k,\delta)}\left(\rho^{(a\mu)}_{\mathbb{E}_A} - \rho^{(a\mu)}_{\mathbb{C}_B}\right)_{ki} C_{ci\delta} = \sum_{ai\mu,(a,i,\mu)\neq(c,k,\delta)}\left[(\rho_A)_{ak\mu,ai\mu} - (\rho_B)_{ka\mu,ia\mu}\right] C_{ci\delta}. \tag{D.46}$$

It can also be recognized that the second term in the effective potential is essentially

$$\left[(\rho_A)_{ck\delta,ck\delta} - (\rho_B)_{kc\delta,kc\delta}\right] C_{ck\delta} = \sum_{ai\mu,(a,i,\mu)=(c,k,\delta)}\left[(\rho_A)_{ak\mu,ai\mu} - (\rho_B)_{ka\mu,ia\mu}\right] C_{ci\delta}. \tag{D.47}$$

Substitute the above two equations into Eq. (D.45), we have

$$\frac{d\mathcal{L}_A}{dC^*_{ck\delta}} = \sum_{ai\mu} C_{ai\mu}\langle ck\delta|H^{(A)}|ai\mu\rangle + \mathcal{E}\sum_{ai\mu} C_{ai\mu}(I\otimes I\otimes I)_{ck\delta,ai\mu}$$
$$+ 2\lambda_B^{(A)}\left[\left(\sum_{ai\mu,(a,i,\mu)\neq(c,k,\delta)} + \sum_{ai\mu,(a,i,\mu)=(c,k,\delta)}\right)\left[(\rho_A)_{ak\mu,ai\mu} - (\rho_B)_{ka\mu,ia\mu}\right] C_{ci\delta}\right] = 0,\ \forall c,k,\delta$$
$$= \sum_{ai\mu} C_{ai\mu}\langle ck\delta|H^{(A)}|ai\mu\rangle + \mathcal{E}\sum_{ai\mu} C_{ai\mu}(I\otimes I\otimes I)_{ck\delta,ai\mu} + 2\lambda_B^{(A)}\sum_{ai\mu} C_{ci\delta}\left[(\rho_A)_{ak\mu,ai\mu} - (\rho_B)_{ka\mu,ia\mu}\right] \tag{D.48}$$
$$= \sum_{ai\mu}\langle ck\delta|H^{(A)}|ai\mu\rangle C_{ai\mu} + \mathcal{E}\sum_{ai\mu}(I\otimes I\otimes I)_{ck\delta,ai\mu}C_{ai\mu} + 2\lambda_B^{(A)}\sum_i\left(\rho_{\mathbb{E}_A} - \rho_{\mathbb{C}_B}\right)_{ki} C_{ci\delta} = 0,\ \forall c,k,\delta \tag{D.49}$$

D.5.2 Derivative $\frac{\partial C_{ai\mu}}{\delta V_{\mathrm{BE}}}$

In this section, we focus on deriving $\frac{\partial C_{ai\mu}}{\delta V_{BE}}$ or $\frac{\partial C^*_{ai\mu}}{\delta V_{BE}}$.

Use wave function perturbation theory on the following eigenvalue equation

$$(H^{(A)} + V_{\mathrm{BE}})|\psi_{A,n}\rangle = -\mathcal{E}^{(A)}_n|\psi_{A,n}\rangle, \tag{D.50}$$

where n labels different eigenstates. Now given a small variation of V_{BE}, the eigenstates and eigenenergies will change. To 1st-order perturbation, we can write the change of each

eigenstate as

$$\delta|\psi_{A,n}\rangle = \sum_{n'\neq n} \frac{\langle\psi_{A,n'}|(\delta V_{BE})|\psi_{A,n}\rangle}{\mathcal{E}_n^{(A)} - \mathcal{E}_{n'}^{(A)}}|\psi_{A,n'}\rangle. \tag{D.51}$$

This is a change on the eigenstate, and not yet exact what we want (we want change on the coefficients in front of basis vector, $C_{ai\mu}^{(n)}$, note the superscript labels the n-th eigenstate). To do this, let's further write the eigenstates in terms of all the coefficients,

$$|\psi_{A,n}\rangle = \sum_{ai\mu} C_{ai\mu}^{(n)}|ai\mu\rangle, \tag{D.52}$$

then the above equation becomes an array of coupled-system of equations:

$$\sum_{ai\mu} dC_{ai\mu}^{(n)}|ai\mu\rangle = \sum_{n'\neq n}\sum_{a''i''\mu'',a'i'\mu',ai\mu} C_{a'i'\mu'}^{(n')*}C_{ai\mu}^{(n)}\frac{\langle a'i'\mu'|(\delta V_{BE})|ai\mu\rangle}{\mathcal{E}_n^{(A)} - \mathcal{E}_{n'}^{(A)}}C_{a''i''\mu''}^{(n')}|a''i''\mu''\rangle. \tag{D.53}$$

Multiply both sides with $\langle a'''i'''\mu'''|$, we arrives at the following equation on the coefficients after relabeling the $ai\mu$ index

$$dC_{ai\mu}^{(n)} = \sum_{n'\neq n}\sum_{a''i''\mu'',a'i'\mu'} C_{a''i''\mu''}^{(n')*}C_{a'i'\mu'}^{(n)}\frac{\langle a''i''\mu''|(\delta V_{BE})|a'i'\mu'\rangle}{\mathcal{E}_n^{(A)} - \mathcal{E}_{n'}^{(A)}}C_{ai\mu}^{(n')}, \tag{D.54}$$

for any index (a, i, μ).

To further simplify the above equation, we introduce parametrization of $\delta V_{\rm BE}$ as linear combination of local potentials on the edge sites $V_{\rm BE} = \sum_{\alpha=0}^{4^m} v_\alpha\ I\otimes\Sigma_\alpha\otimes I$, where the first and the last identity operators act on the center and the bath sites by definition. Therefore, a functional variation of $V_{\rm BE}$ can be parametrized as a small change in the scalar coefficients v_α

$$\delta V_{\rm BE} = \sum_{\alpha=0}^{4^m} dv_\alpha\ I\otimes\Sigma_\alpha\otimes I, \tag{D.55}$$

and the matrix elements are

$$\langle a''i''\mu''|(\delta V_{\rm BE})|a'i'\mu'\rangle = \delta_{a'a''}\delta_{\mu'\mu''}\sum_{\alpha=0}^{4^m}\langle i''|\Sigma_\alpha|i'\rangle dv_\alpha. \tag{D.56}$$

This leads to the simplifed expression for $dC_{ai\mu}^{(n)}/\delta V_{BE}$

$$\frac{dC_{ai\mu}^{(n)}}{\delta V_{\rm BE}} = \sum_{\alpha=0}^{4^m}\frac{dC_{ai\mu}^{(n)}}{dv_\alpha} = \sum_{n'\neq n}\sum_{i'',a'i'\mu'} C_{a'i''\mu'}^{(n')*}C_{a'i'\mu'}^{(n)}\frac{\sum_{\alpha=0}^{4^m}\langle i''|\Sigma_\alpha|i'\rangle}{\mathcal{E}_n^{(A)} - \mathcal{E}_{n'}^{(A)}}C_{ai\mu}^{(n')}. \tag{D.57}$$

In particular, for the ground eigenstate, we have (omitting the superscript $^{(0)}$)

$$\frac{dC_{ck\delta}^{*}}{\delta V_{\rm BE}} = \sum_{\alpha=0}^{4^m}\frac{dC_{ck\delta}^{*}}{dv_\alpha} = \sum_{n'\neq 0}\sum_{i',ai\mu} C_{ai'\mu}^{(n')}C_{ai\mu}^{*}\frac{\sum_{\alpha=0}^{4^m}\langle i|\Sigma_\alpha|i'\rangle}{\mathcal{E}_0^{(A)} - \mathcal{E}_{n'}^{(A)}}C_{ck\delta}^{(n')*}. \tag{D.58}$$

D.5.3 Gradient of Cost Function versus BE Potential

Now we are ready to put everything together to obtain a final expression for the gradient of the cost function versus the BE potential.

Substitute Eq. (D.58) and Eq. (D.49) into (D.30), we obtain

$$\sum_{ck\delta}\frac{d\mathcal{L}_A}{dC^*_{ck\delta}}\frac{dC^*_{ck\delta}}{\delta V_{BE}}=\sum_{ck\delta}\left[\sum_{ai\mu}\langle ck\delta|H^{(A)}|ai\mu\rangle C_{ai\mu}+\mathcal{E}_0^{(A)}\sum_{ai\mu}(I\otimes I\otimes I)_{ck\delta,ai\mu}C_{ai\mu}+2\lambda_B^{(A)}\sum_i\left(\rho_{\mathbb{E}_A}-\rho_{\mathbb{C}_B}\right)_{ki}C_{ci\delta}\right]$$
$$\times\left[\sum_{n'\neq 0}\sum_{i',ai\mu}C^{(n')}_{ai'\mu}C^*_{ai\mu}\frac{\sum_{\alpha=0}^{4^m}\langle i|\Sigma_\alpha|i'\rangle}{\mathcal{E}_0^{(A)}-\mathcal{E}_{n'}^{(A)}}C^{(n')*}_{ck\delta}\right]=0. \tag{D.59}$$

Writing this in matrix form,

$$\sum_{ck\delta}\left\{\left(\mathbf{H^{(A)}C}\right)_{ck\delta}+\left(\mathcal{E}_{\mathbf{0}}^{\mathbf{(A)}}\mathbf{C}\right)_{ck\delta}+2\lambda_B^{(A)}\left[\left(\mathbf{I}\otimes\left(\mathbf{æ}_{\mathbb{E}_\mathbf{A}}-\mathbf{æ}_{\mathbb{C}_\mathbf{B}}\right)\otimes\mathbf{I}\right)\mathbf{C}\right]_{ck\delta}\right\}$$
$$\times\left[\sum_{n'\neq 0}\left[\mathbf{C}^\dagger\left(\mathbf{I}\otimes\sum_{\mathbf{ff=0}}^{\mathbf{4^m}}\mathbf{W}_{\mathbf{ff}}^{(\mathbf{n'})}\otimes\mathbf{I}\right)\mathbf{C}^{(\mathbf{n'})}\right]\mathbf{C}^{(\mathbf{n'})*}\right]_{ck\delta}=0. \tag{D.60}$$

or

$$\sum_{n'\neq 0}\left[\mathbf{C}^\dagger\left(\mathbf{I}\otimes\sum_{\mathbf{ff=0}}^{\mathbf{4^m}}\mathbf{W}_{\mathbf{ff}}^{(\mathbf{n'})}\otimes\mathbf{I}\right)\mathbf{C}^{(\mathbf{n'})}\right]\times\left[\mathbf{C}^{(\mathbf{n'})\dagger}\left(\mathbf{H^{(A)}}+\mathcal{E}_{\mathbf{0}}^{\mathbf{(A)}}+2\lambda_B^{(A)}\left(\mathbf{I}\otimes\left(\mathbf{æ}_{\mathbb{E}_\mathbf{A}}-\mathbf{æ}_{\mathbb{C}_\mathbf{B}}\right)\otimes\mathbf{I}\right)\right)\mathbf{C}\right]=0. \tag{D.61}$$

where $\mathbf{W}_{\mathbf{ff}}^{(\mathbf{n'})}{}_{ii'}=\frac{\langle i|\Sigma_\alpha|i'\rangle}{\mathcal{E}_0^{(A)}-\mathcal{E}_{n'}^{(A)}}$. Writing this with respect to each parameter v_α in V_{BE}, we have

$$\frac{d\mathcal{L}_A}{dv_\alpha}=\sum_{n'\neq 0}\left[\mathbf{C}^\dagger\left(\mathbf{I}\otimes\mathbf{W}_{\mathbf{ff}}^{(\mathbf{n'})}\otimes\mathbf{I}\right)\mathbf{C}^{(\mathbf{n'})}\right]$$
$$\times\left[\mathbf{C}^{(\mathbf{n'})\dagger}\left(\mathbf{H^{(A)}}+\mathcal{E}_{\mathbf{0}}^{\mathbf{(A)}}+2\lambda_B^{(A)}\left(\mathbf{I}\otimes\left(\mathbf{æ}_{\mathbb{E}_\mathbf{A}}-\mathbf{æ}_{\mathbb{C}_\mathbf{B}}\right)\otimes\mathbf{I}\right)\right)\mathbf{C}\right],\quad\forall\alpha\in[0,4^m]. \tag{D.62}$$

From this gradient, we may update all the fundamental parameters $\{\lambda_B^{(A)},v_\alpha\}$ using gradient descent (or other update scheme) to minimized the Lagrangian, as is typically performed in optimization. However, to compute this gradient exactly in (D.62), it is required that all the eigenstates are known (not only the ground state) which is clearly very costly and not that useful. Nevertheless, it serves as a good starting point to develop *approximated* updating scheme. One possible approximation is to truncate the summation over n' to only a low energy subspace. More efficient approximations are left for future investigation.

D.6 Sample Complexity for Estimating the Overlap from Tomography

We have seen that the linear and quadratic constraints are equivalent in Sec. D.3, and presented that an efficient quantum circuit based on `SWAP` test in Sec. D.4. In this section,

we derive the sample complexity of estimating the RDM mismatch by naively sampling individual RDM element from tomography, to demonstrate the advantage of using the SWAP test in quadratic matching.

Recall that the quadratic mismatch between two RDMs reduces to the sum of element-wise distance (squared) of their individual RDM elements from (D.18)

$$S^2(\rho_A, \rho_B) = \sum_{\alpha=1}^{4^m-1} \left(\langle \Sigma_\alpha \rangle_A - \langle \Sigma_\alpha \rangle_B\right)^2 . \tag{D.63}$$

where m is the number of qubits in the overlapping region.

Let's denote the true value of

$$\lim_{\text{sample size}\to\infty} \langle \Sigma_\alpha \rangle_X = R_\alpha^{(X)} \tag{D.64}$$

for $X = A, B$. Then, in tomography the estimate uncertainty of $\{R_\alpha^{(A)}, R_\alpha^{(B)}\}$ will propagate to the uncertainty of S^2 via

$$\text{var}(S^2) = \boldsymbol{J} \cdot \begin{bmatrix} C_{\{R_\alpha^{(A)}\}} & \\ & C_{\{R_\alpha^{(B)}\}} \end{bmatrix} \cdot \boldsymbol{J}^T, \tag{D.65}$$

where $\boldsymbol{J} = \nabla_{\{R_\alpha^{(A)}, R_\alpha^{(B)}\}} S^2 = [\frac{\partial S^2}{\partial R_1^{(A)}}, \frac{\partial S^2}{\partial R_2^{(A)}}, \cdots, \frac{\partial S^2}{\partial R_{4^m-1}^{(A)}}, \frac{\partial S^2}{\partial R_1^{(B)}}, \frac{\partial S^2}{\partial R_2^{(B)}}, \cdots, \frac{\partial S^2}{\partial R_{4^m-1}^{(B)}}]$ is the Jacobian, $C_{\{R^{(A)}\}}$ and $C_{\{R^{(B)}\}}$ are the co-variance matrix of the RDM elements. $\text{var}(\cdot)$ denotes the variance.

Note that $C_{\{R^{(A)}\}}$ and $C_{\{R^{(B)}\}}$ will be system-dependent, and for now let us assume there is not co-variance between individual elements of ρ_A and ρ_B, and therefore $C_{\{R^{(A)}\}}$ and $C_{\{R^{(B)}\}}$ will be diagonal with diagonal elements being the variance of each RDM element.

Moreover, by substituting (D.63) into $\boldsymbol{J}$, we can explicitly evaluate

$$\begin{aligned} \boldsymbol{J}_{A,\alpha} &= \frac{\partial S^2}{\partial R_\alpha^{(A)}} = 2\left(R_\alpha^{(A)} - R_\alpha^{(B)}\right), \\ \boldsymbol{J}_{B,\alpha} &= \frac{\partial S^2}{\partial R_\alpha^{(B)}} = -2\left(R_\alpha^{(A)} - R_\alpha^{(B)}\right). \end{aligned} \tag{D.66}$$

This gives

$$\text{var}(S^2) = 4\sum_{i=1}^{4^m-1} \left(R_\alpha^{(A)} - R_\alpha^{(B)}\right)^2 \left[\text{var}(R_\alpha^{(A)}) + \text{var}(R_\alpha^{(B)})\right]. \tag{D.67}$$

Now assume that each element of the RDM is estimated by the same amount of samples $N_{samp,0}$, then from binomial distribution, the variance of each $R_\alpha^{(A)}$ and $R_\alpha^{(B)}$ is

$$\text{var}(R_\alpha^{(X)}) = \frac{R_\alpha^{(X)}(1 - R_\alpha^{(X)})}{N_{samp,0}}, \ \forall X = A, B. \tag{D.68}$$

Substitute this into (D.67), we have

$$\mathrm{var}(S^2) = \frac{4DS^2}{N_{samp,0}}, \tag{D.69}$$

where D is a system-dependent constant

$$D = \frac{\sum_{\alpha=1}^{4^m-1} \left(R_\alpha^{(A)} - R_\alpha^{(B)}\right)^2 \left(R_\alpha^{(A)}(1-R_\alpha^{(A)}) + R_\alpha^{(B)}(1-R_\alpha^{(B)})\right)}{\sum_{\alpha=1}^{4^m-1} \left(R_\alpha^{(A)} - R_\alpha^{(B)}\right)^2} \tag{D.70}$$

Given a target accuracy ϵ on S, then $\mathrm{var}(S^2) = (2S\epsilon)^2$, from the above equation we can solve to obtain the required number of samples as for each individual RDM elements

$$N_{samp,0} = \frac{D}{\epsilon^2}. \tag{D.71}$$

There are $4^n - 1$ elements for n-qubit overlapping region, leading to an overall sampling complexity of

$$N_{samp}^{\mathrm{TMG}} = \mathcal{O}(e^n) \cdot N_{samp,0} = \mathcal{O}(e^n)\frac{D}{\epsilon^2}, \tag{D.72}$$

for estimating the overlap to ϵ accuracy from density matrix tomography. The reason why $\mathcal{O}(e^n)$ instead of naive $4^n - 1$ is because there are commuting Pauli operators that can be estimated simultaneously. However, the exponential scaling in terms of the number of overlapping qubits n remains.

As is mentioned in the main test, overlaps between density ma- trices are not low-rank observables, so the sampling com- plexity of estimating it is likely to be high. However, more efficient sampling schemes may exist. For example, by sampling the differences in the RDMs between the current and the previous BE iterations, the sampling complexity could be much better than exponential. One simple way of doing this is to use the diagonal basis of the previous iteration as the measurement basis in the current iteration to perform the RDM sampling. We leave this for future investigation.

D.7 Details of Quantum Amplitude Estimation and Quadratic Speedup

In this section, we describe in detail how a quantum amplitude estimation can be implemented by combining an oblivious amplitude amplification (Sec. D.7.1) with a binary search algorithm (Sec. D.7.2). We then outline how a binary search derive the sample complexity needed to achieve a constant precision ϵ in the overlap in our coherent matching algorithm and compare that with a classical incoherent sampling estimation scheme, demonstrating a quadratic speedup of the former (Sec. D.7.3).

D.7.1 Amplitude Amplification

From Theorem 2 of Ref. [Mar+21b], it is shown that given a state preparation process U from an initial state $|B_0\rangle$, the overlap a of this prepared state with another state $|A_0\rangle$, i.e., $a := \langle A_0|U|B_0\rangle$, can be transformed by a d-degree polynomial $P(a)$ such that $|1 - P(a)| < \delta$ using rotations $A_\phi = e^{i\phi|A_0\rangle\langle A_0|}$ and $B_\phi = e^{i\phi|B_0\rangle\langle B_0|}$ by the following quantum circuit

$$P(a) = \langle A_0| \left[\prod_{j=1}^{d/2} U B_{\phi_{2j-1}} U^\dagger A_{\phi_{2j}} \right] U|B_0\rangle, \tag{D.73}$$

where $d = \mathcal{O}(\frac{1}{a}\log(1/\delta))$, and the rotation angles ϕ_{2j} and ϕ_{2j-1} can be efficiently computed classically. This is a generalization of Grover's search algorithm where the rotation B_ϕ is similar to the diffusion operator (reflection about the average), and A_ϕ is analogous to the Grover's reflection about the target state. The difference is that the rotation angles ϕ's can be fractions of π and is thus more general which combines the optimality of Grover's algorithm for unstructured search and the fixed-point property [YLC14].

In our case, denote $S = |\langle \Psi^{\mathbb{E}}|\Phi^{\mathbb{C}}\rangle|$ as the overlap of the edge of the first fragment and center of the second fragment in their ground states, and choose the following for state $|A_0\rangle$ and $|B_0\rangle$

$$|A_0\rangle = \frac{1}{\sqrt{2(1+S^2)}}|0\rangle \left(|\Psi^{\bar{\mathbb{E}}}\rangle|\Psi^{\mathbb{E}}\rangle|\Phi^{\mathbb{C}}\rangle|\Phi^{\bar{\mathbb{C}}}\rangle + |\Psi^{\bar{\mathbb{E}}}\rangle|\Phi^{\mathbb{C}}\rangle|\Psi^{\mathbb{E}}\rangle|\Phi^{\bar{\mathbb{C}}}\rangle \right) \tag{D.74}$$

$$|B_0\rangle = |0\rangle|\Psi_T^{\bar{\mathbb{E}}}\rangle|\Psi_T^{\mathbb{E}}\rangle|\Phi_T^{\mathbb{C}}\rangle|\Phi_T^{\bar{\mathbb{C}}}\rangle, \tag{D.75}$$

where $|B_0\rangle$ is the input trial states (thus the subscript "T") for the quantum eigensolver, while $|A_0\rangle$ is the symmetric subspace of the `SWAP` test. Let U be the state preparation circuit (in our case the two quantum eigensolver + the `SWAP` test) given by Fig. D.2.

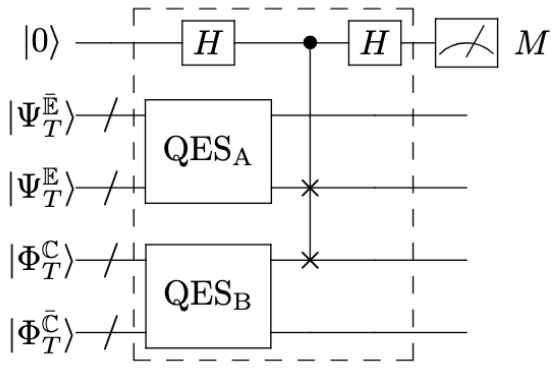


Figure D.2: Quantum circuit to estimate the ground state overlap between subsystems of two fragments, composed of two quantum eigensolver (QES) for two fragments ground state wave function followed by a `SWAP` test. The circuit in the dashed box is U in Eq. (D.73) which will be repeated multiple times during the amplitude amplification process as will be discussed in the following.

Given these choices, it can be verified that the

$$U|B_0\rangle = a|A_0\rangle + \sqrt{1-a^2}|A_\perp\rangle \tag{D.76}$$

where

$$a = \langle A_0|U|B_0\rangle = \sqrt{\frac{1+S^2}{2}}, \tag{D.77}$$

and

$$|A_\perp\rangle = \frac{1}{\sqrt{2(1-S^2)}}|1\rangle\left(|\Psi^{\bar{\mathbb{E}}}\rangle|\Psi^{\mathbb{E}}\rangle|\Phi^{\mathbb{C}}\rangle|\Phi^{\bar{\mathbb{C}}}\rangle - |\Psi^{\bar{\mathbb{E}}}\rangle|\Phi^{\mathbb{C}}\rangle|\Psi^{\mathbb{E}}\rangle|\Phi^{\bar{\mathbb{C}}}\rangle\right). \tag{D.78}$$

Note the choice of $|A_\perp\rangle$ has a lot of degrees of freedom, as long as it is normalized and orthogonal to $|A_0\rangle$. In particular, $|A_\perp\rangle = |1\rangle|\boldsymbol{\Psi}\rangle$ for any $|\boldsymbol{\Psi}\rangle$ (with support on the two system registers) will work as long as the ancilla qubit is at state $|1\rangle$.

Moreover, we choose $|B_\perp\rangle$ such that

$$U|B_\perp\rangle = -a|A_\perp\rangle + \sqrt{1-a^2}|A_0\rangle \tag{D.79}$$

which leads to an explicit expression for $|B_\perp\rangle$

$$\begin{aligned}
|B_\perp\rangle &= -aU^{-1}|A_\perp\rangle + \sqrt{1-a^2}U^{-1}|A_0\rangle \\
&= \frac{1}{2}|0\rangle\otimes \\
&\left[-\sqrt{\frac{1+S^2}{1-S^2}}\left(|\Psi_T^{\bar{\mathbb{E}}}\rangle|\Psi_T^{\mathbb{E}}\rangle|\Phi_T^{\mathbb{C}}\rangle|\Phi_T^{\bar{\mathbb{C}}}\rangle - |\Psi_T^{\bar{\mathbb{E}}}\rangle|\Phi_T^{\mathbb{C}}\rangle|\Psi_T^{\mathbb{E}}\rangle|\Phi_T^{\bar{\mathbb{C}}}\rangle\right)\right. \\
&\left.+\sqrt{\frac{1-S^2}{1+S^2}}\left(|\Psi_T^{\bar{\mathbb{E}}}\rangle|\Psi_T^{\mathbb{E}}\rangle|\Phi_T^{\mathbb{C}}\rangle|\Phi_T^{\bar{\mathbb{C}}}\rangle + |\Psi_T^{\bar{\mathbb{E}}}\rangle|\Phi_T^{\mathbb{C}}\rangle|\Psi_T^{\mathbb{E}}\rangle|\Phi_T^{\bar{\mathbb{C}}}\rangle\right)\right].
\end{aligned} \tag{D.80}$$

It then follows that $\{|A_0\rangle, |A_\perp\rangle\}$ and $\{|B_0\rangle, |B_\perp\rangle\}$, each forms a 2-dimensional subspace where U can be expanded upon

$$\begin{aligned}
U =& a|A_0\rangle\langle B_0| + \sqrt{1-a^2}|A_\perp\rangle\langle B_0| \\
&- a|A_\perp\rangle\langle B_\perp| + \sqrt{1-a^2}|A_0\rangle\langle B_\perp| \qquad \text{(D.81)} \\
=& \begin{bmatrix} a & \sqrt{1-a^2} \\ \sqrt{1-a^2} & -a \end{bmatrix}. \qquad \text{(D.82)}
\end{aligned}$$

Given access to A_ϕ and B_ϕ, the rest of amplitude amplification follows the same as Ref. [Mar+21b] by using Eq. (D.73).

We should note that the rotation B_ϕ is easy to construct from

$$B_\phi = e^{i\phi|B_0\rangle\langle B_0|} = U_T e^{i\phi|0\rangle^{\otimes}\langle 0|^{\otimes}} U_T^\dagger, \tag{D.83}$$

because U_T is a known unitary that can prepare the initial trial state (for example Hartree-Fock state) from the zero state

$$|0\rangle \otimes |\Psi_T^{\bar{\mathbb{E}}}\rangle|\Psi_T^{\mathbb{E}}\rangle|\Phi_T^{\mathbb{C}}\rangle|\Phi_T^{\bar{\mathbb{C}}}\rangle = U_T|0\rangle^{\otimes}. \tag{D.84}$$

At first glance, the rotation A_ϕ seems to be more difficult and it requires an oracle to prepare $|A_0\rangle$ from the zero state

$$|A_0\rangle = V|0\rangle^{\otimes}, \tag{D.85}$$

such that $A_\phi = e^{i\phi|A_0\rangle\langle A_0|} = Ve^{i\phi|0\rangle^{\otimes}\langle 0|^{\otimes}}V^\dagger$. However, because the SWAP test entangles the symmetric and anti-symmetric subspace of the two system register separately with $|0\rangle$ and $|1\rangle$ state of the ancilla, we can simply choose

$$A_\phi = e^{i\phi(|0\rangle\langle 0|-|1\rangle\langle 1|)\otimes I^{\otimes}} = e^{i\phi Z\otimes I^{\otimes}}, \tag{D.86}$$

which is just a single-qubit Z rotation on the ancilla, and the identity operator $I^{\otimes}$ has support on the two system registers. It can be readily verified that when applying A_ϕ to a linear combination of $|A_0\rangle$ and $|A_\perp\rangle$ that

$$\begin{aligned} A_\phi(c_0|A_0\rangle + c_1|A_\perp\rangle) &= e^{i\phi(|0\rangle\langle 0|-|1\rangle\langle 1|)\otimes I^{\otimes}}(c_0|0\rangle\otimes|\Psi_{A_0}\rangle + c_1|1\rangle\otimes|\Psi_{A_\perp}\rangle) \\ &= c_0e^{i\phi}|0\rangle\otimes|\Psi_{A_0}\rangle + c_1e^{-i\phi}|1\rangle\otimes|\Psi_{A_\perp}\rangle \\ &= e^{i\phi}c_0|A_0\rangle + e^{-i\phi}c_1|A_\perp\rangle, \end{aligned} \tag{D.87}$$

which imposes a relative phase of 2ϕ between the target state $|A_0\rangle$ and the unwanted state $|A_\perp\rangle$ as desired.

With these construction, the overall quantum circuit for the amplitude amplification is given in Fig. D.3. with the state prep circuit U given in Fig. D.2, and the rotation $A_{\phi_{2j}}$

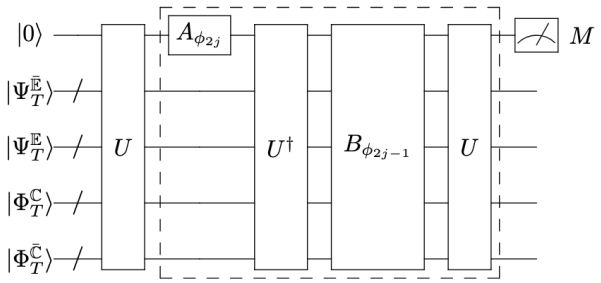


Figure D.3: Quantum circuit for fixed-point oblivious amplitude amplification of the coherent quantum matching. The quantum gates in dashed box corresponds to the gates in bracket of Eq. (D.73) which needs to be repeated by $d/2$ times, where d will be determined by the slope of the amplification polynomial $\frac{1}{\Delta_k}$ (more about this in the next section).

and $B_{\phi_{2j-1}}$ defined in Eq. (D.86) and (D.83) respectively. Note that this circuit requires $U^\dagger$ which means the two quantum eigensolvers in U has to be run in backwards, which is certainly possible for solvers such as QPE and VQE.

D.7.2 Estimate the Amplitude from Binary Search

The above amplitude amplification can be combined with a binary search algorithm to estimate the magnitude of the amplitude $a = \sqrt{\frac{1+S^2}{2}}$ up to precision $\epsilon = \frac{1}{2^n}$ in a bit-by-bit fashion.

Denote a binary representation of the amplitude a as

$$a := [b_{n-1}b_{n-2}\cdots b_1 b_0] = \frac{1}{2^n}\left(b_{n-1}2^{n-1} + b_{n-1}2^{n-1} + \cdots + b_0\right). \tag{D.88}$$

We perform the following two steps repeatedly to determine b_k for $k = n-1, n-2, \cdots, 1, 0$:

1. Perform a fixed-point amplitude amplification using a polynomial in Ref. [Mar+21b] with slope $\frac{1}{\Delta_k}$ (this determines the depth of the circuit d as in Fig. D.3) and precision ϵ, where Δ_k is determined from all previous estimations on $b_{n-1}, \cdots, b_{k+1}$ by

$$\Delta_k = \frac{\sqrt{2}}{2^n}\left(\sum_{l=k+1}^{n-1} b_l 2^l + 2^k\right). \tag{D.89}$$

2. Measure the ancilla of the SWAP test by collecting $N_\epsilon = \lceil \frac{2}{\epsilon} \rceil$ samples and then set b_k to the expectation value of the estimated outcome, $b_k = \langle M \rangle$.

It can be verified that this protocol works because each time after the amplitude amplification, if $\langle M \rangle = 1$, then the amplitude is in the interval $[0.b_{n-1}b_{n-2}\cdots b_{k+1}1, 0.b_{n-1}b_{n-2}\cdots b_{k+1}1 + \frac{1}{2^{n-k}}]$; otherwise if $\langle M \rangle = 0$, then the amplitude is in $[0.b_{n-1}b_{n-2}\cdots b_{k+1}0, 0.b_{n-1}b_{n-2}\cdots b_{k+1}1]$. We choose N_ϵ to be large enough such that it is sufficient to tell if the amplified amplitude is within the range of $[1-\epsilon, 1]$ with high probability. For a Bernoulli distribution with $p \in [(1-\epsilon)^2, 1]$, we require the standard deviation of the estimation for amplitude to be roughly ϵ, i.e.

$$\Delta a = |\frac{da}{dp}|\Delta p = \frac{1}{2}\sqrt{\frac{(1-p)}{N_\epsilon}} = \epsilon, \tag{D.90}$$

which gives

$$N_\epsilon = \frac{(1-p)}{4\epsilon^2} = \frac{1-(1-\epsilon)^2}{4\epsilon^2} = \frac{1}{2\epsilon} - \frac{1}{4}. \tag{D.91}$$

Therefore, a choice of $N_\epsilon = \lceil \frac{1}{2\epsilon} \rceil$ suffices.

D.7.3 Quadratic Speedup

We demonstrate the quantum speedup due to amplitude amplification (AA) by estimate the sample complexity required to achieve a constant precision on estimating a.

We first estimate the total number of samples in the `SWAP` test + amplitude estimation (AE) approach by combining AA and binary search. For each digit b_k, each sample takes a total of $\frac{1}{\Delta_k}\log(\frac{1}{\epsilon})$ queries to the eigensolver in the amplitude amplification for Δ_k in Eq. (D.89), and we need N_ϵ (Eq. (D.91)) samples to estimate whether the amplified amplitude is within the range of $[1-\epsilon, 1]$, which gets us to the number of queries to the eigensolver for estimating b_k to be $N_\epsilon \cdot \frac{1}{\Delta_k}\log(\frac{1}{\epsilon})$. Therefore, the total number of queries to the eigensolver will be the sum of the cost of estimating each bit of a, given by

$$N_{samp}^{\texttt{SWAP}+\text{AE}} = \frac{1}{2\epsilon}\ln(\frac{1}{\epsilon})\sum_{k=0}^{n-1}\frac{1}{\Delta_k} \leq \frac{\sqrt{2}}{2\epsilon}\ln(\frac{1}{\epsilon})\log_2(\frac{1}{\epsilon}) = \frac{\sqrt{2}}{2\ln(2)\epsilon}\ln^2(\frac{1}{\epsilon}), \tag{D.92}$$

where we have used the following inequalities on

$$\Delta_k \geq \Delta_{n-1} = \frac{\sqrt{2}}{2^n} b_{n-1} 2^{n-1} = \frac{1}{\sqrt{2}}, \forall\ k \tag{D.93}$$

derived from Eq. (D.89) and the fact that $b_{n-1} = 1$ from the definition of $a = \sqrt{\frac{1+S^2}{2}}$. This gives us

$$\sum_{k=0}^{n-1} \frac{1}{\Delta_k} \leq n\Delta_{n-1} = \sqrt{2}n = \sqrt{2}\log_2(1/\epsilon). \tag{D.94}$$

Note the above query complexity is independent of the amplitude a (or the overlap S) because our estimation algorithm is constructed using fixed-point oblivious amplitude amplification.

As a comparison, in the case of only using `SWAP` test (no AA), the total number of samples $N_{samp}^{\texttt{SWAP}}$ required to estimate a to precision ϵ has to satisfy

$$\epsilon = \Delta a = |\frac{da}{dp}|\Delta p = \frac{1}{2}\sqrt{\frac{(1-p)}{N_{samp}^{\texttt{SWAP}}}}, \tag{D.95}$$

leading to

$$N_{samp}^{\texttt{SWAP}} = \left(\frac{1-S^2}{8}\right)\frac{1}{\epsilon^2} \tag{D.96}$$

after substituting $p = a^2 = \frac{1+S^2}{2}$. Comparing (D.96) and (D.92), we observe a quadratic speedup up to a polylog factor.

D.8 QBE Algorithm Using Naive RDM Linear Matching

In this section, for completeness, we present a QBE algorithm to perform bootstrap embedding on quantum computers by naively matching all the RDM matrix elements one by one. This scheme is inefficient as discussed in the main text due to the exponential measurement overhead. For concreteness, the algorithm as written here uses gradient descent to perform the optimization of the Lagrangian (with constraint added by Lagrange multipliers), but other gradient-based or gradient-free optimization can be used as well. We ignore the final step of tuning the global chemical potential.

Algorithm 11 Algorithm for QBE with naive linear constraint

Input: Geometry of the total molecular system and the associated *ab initio* Hamiltonian.

Output: $\left(H^{(A)} + V_{\mathrm{BE}}^{(A)}\right)$ for all A

for A = 1 to N_{frag} **do** ▷ Initial fragmentation: Divide the full molecular system into N_{frag} overlapping fragments
  Generate $H^{(A)}$ using Eq. (D.1) in Appendix D.1
  Compute the initial full ground state density matrix: $\rho^{(A)} \leftarrow \texttt{eigensolver}(H^{(A)})$
  Compute the single-qubit reduced density matrices $\rho_r^{(A)}$ for all $r \in O^{(A)}$
  Set $V_{\mathrm{BE}}^{(A)} = 0$
end for
Compute the initial value of the average mismatch $\Delta\rho$;
Set $\boldsymbol{\lambda}_B^{(A)} = 0$ for $\forall A, B$; `iter` $= 0$, `gd_step` $= 0$ ▷ Parameter initialization
Set the Lagrange multiplier convergence thresholds $\epsilon_{\Delta\rho}$ and $\epsilon_{\Delta\mu}$ to their desired initial values.
while $\Delta\rho > \epsilon_{\Delta\rho}$ **do** ▷ Main loop
  for A = 1 to N_{frag} **do**
    Set the current learning rate $\eta \leftarrow \texttt{lr\_schedule}(\texttt{iter})$
    for `gd_step` $= 0$; `gd_step` $++$; `gd_step` $< N_{steps}$ **do**
      for B = 1 to N_{frag} **do**
        if $\mathbb{E}^{(A)} \cap \mathbb{C}^{(B)} \neq \emptyset$ **then** State Estimate the vector $\Delta\overset{\smile}{\mathbf{B}}^{(\mathbf{A})}(\rho^{(A)}, \rho^{(B)})$ as defined in Eq. (6.16) by using $N_{samp,0}$ (Eq. (D.71)) quantum samples to estimate each RDM element in Eq. (6.16).
          Update the Lagrange multiplier vector $\overset{\smile}{\mathbf{B}}^{(\mathbf{A})}$ using gradient descent: $\boldsymbol{\lambda}_B^{(A)} \leftarrow \boldsymbol{\lambda}_B^{(A)} - \eta\Delta\boldsymbol{\lambda}_B^{(A)}$.
          Generate the BE potential $V_{\mathrm{BE}}^{(A)}$ as defined in Eq. (6.17)
          Update the Hamiltonian matrix for fragment A classical: $H^{(A)} \leftarrow H^{(A)} + \eta V_{\mathrm{BE}}^{(A)}$
        end if
      end for
      Reduce the learning rate η according to $\eta \leftarrow \eta \cdot [1 - (\texttt{gd\_step}/N_{steps})]$
    end for
    Update the mismatch error $\Delta\rho$ by estimating it using quantum samples
  end for
  for A = 1 to N_{frag} **do**
    Update the full ground state density matrix: $\rho^{(A)} \leftarrow \texttt{eigensolver}(H^{(A)})$
    Compute the single-qubit reduced density matrices $\rho_r^{(A)}$ for all $r \in O^{(A)}$
    Increment `iter` by one: `iter` $\leftarrow$ `iter` $+ 1$
  end for
end while
return $H^{(A)}$ for all A, $\Delta\rho$.

In the above algorithm, $\texttt{eigensolver}(H^{(A)})$ denotes the quantum eigensolver is called to find the ground state of Hamiltonian $H^{(A)}$. Moreover, `lr schedule(iter)` is a learning rate (step size) schedule to improve convergence in gradient-descent algorithms. It returns a value for the learning rate corresponding to the `iter`-th BE iteration. One typical choice of a learning rate schedule is `lr schedule(iter)` $\propto 1/$`iter`. One crucial step of the above algorithm is to estimate the mismatch $\Delta\rho$ as defined in Eq. (6.24). Here in the naive linear matching algorithm, this is accomplished by perform tomography on each RDM element as described in Sec. D.6, and then classically compute the mismatch (Eq. (6.24)) using the estimated RDM elements.

D.9 Computational Details

D.9.1 FCI and QPE Eigensolver Runtime Benchmark

Full configuration interaction (FCI) is chemists' version of exact diagonalization. In FCI, the full Hamiltonian is expanded under Slater determinant basis (i.e., *configurations*), and then a restricted Hilbert space of interest (for example, with fixed particle number and spin multiplicity) is exactly diagonalized to find the eigenstate state of interest. In the present work, we perform FCI calculation using PySCF [Sun+18] on H_n ($n = 2, 4, 6, 8, 10, 12, 14$) under STO-3G basis with a fixed distance of 1 Å. The real runtime is recorded for different n and then normalized according to the runtime of H_2.

Quantum phase estimation (QPE) is a quantum algorithm to estimate the eigenstate energy and prepare the eigenstate wave function of a given Hamiltonian, whose accuracy can be systematically improved to the exact result. We give a brief overview here and refer the readers to Ref. [SHF13] for more details on QPE including improved versions.

In QPE, the exact ground state $|\Psi_0^A\rangle$ (or an excited state) can be prepared on a quantum computer using the quantum phase estimation algorithm followed by post-selection, given a trial state $|\Psi_{in}^A\rangle$ is taken as the input quantum state as follows

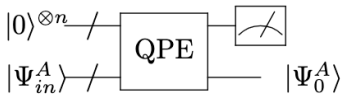


Figure D.4: Schematic for quantum phase estimation.

Analysis shows that the success probability of the post-selection process is

$$\text{Prob}[\textit{success}] = |\langle\Psi_{in}^A|\Psi_0^A\rangle|^2\left(\frac{\sin(\pi\xi)}{\pi\xi}\right)^2 \geq |\langle\Psi_{in}^A|\Psi_0^A\rangle|^2 \left(\frac{4}{\pi^2}\right), \tag{D.97}$$

where $0 < \xi < \frac{1}{2}$ is the distance between the measurement outcome x (in the upper register in Fig. D.4) and the true ground state energy $\theta_0 2^n - x$.

In the present work, for QPE runtime, H_n ($n = 1, 2, 3, 4, 5$) under STO-3G basis with Jordan-Wigner mapping are used as benchmark systems. We use one evaluation qubit due to device constraint. Note the number of evaluation qubits only introduces a constant scaling factor in the absolute gate depth and will not change the scaling behavior of the QPE solver. The quantum phase estimation circuit for different hydrogen molecules is transpiled using a FAKE_MUMBAI backend available in Qiskit [Abr+19a] with a basis gate set composed of R_x, R_y, R_z, and CNOT gates. The resulting total gate depth is recorded as an estimation to the runtime of the QPE circuit. To account for the non-unity success probability of the QPE due to the finite overlap between the initial Hartree-Fock trial state and the exact ground state, the element of the CI vector corresponding to the Hartree-Fock contribution to the FCI ground state is extracted. The QPE gate depth is then rescaled by the square of the overlap amplitude. Due to the stochastic nature of the classical transpilation algorithm, the QPE circuit of each molecule is repeatedly transpiled 5 times and the smallest gate depth is

used in our data. We believe this procedure results in good estimation of the real runtime of the QPE eigensolver, which includes the number of repetition required to account for the failure probability of QPE.

D.9.2 Classical Bootstrap Embedding with VMC and FCI as Eigensolver

We implement a classical BE algorithm using variational Monte Carlo (VMC) as a stochastic eigensolver to generate the VMC data in Fig. 6.2 of the main text. A gradient-descent algorithm is used for the optimization. The real-space formulation of VMC with single determinant two-body Jastrow factor wavefunction as implemented in the PyQMC package [Whe+21] is used to obtain the 1-RDM of each fragment for the H_8 molecule. The analytical form of the two-body Jastrow term is given in Ref. [WM07] which is applied to the single Slater determinants as the VMC trial wave function. The 1-RDMs of the adjacent fragments are then matched as described in Sec. 6.2 of the main text.

In the calculation of H_8 (6 fragments in total), for each BE iteration, the VMC eigensolver is called 10 times for matching all overlapping sites. Each time the VMC is called, optimization of the Jastrow factors is first performed using roughly 10k MC samples, and then additional MC samples are accumulated in the production run to evaluate the 1-RDM. In the production run, the Monte Carlo (MC) sampling process is performed for N_{blocks} blocks. By default of PyQMC, there are 10 steps per block and 1000 configurations. We run the BE with VMC eigensolver with three different choice of $N_{blocks} = 4, 16, 64$, leading to a total number of 40k, 160k, 640k MC samples in each BE iteration. We run a total of 200 BE iterations in each of the three case and plot the first 80 iterations in Fig. 6.2. In the case of FCI as eigensolver, we the only difference is that the 1-RDM is computed using the FCI solver, and the rest of BE iteration follows from Sec. 6.2. A step size of 0.05 is used for both VMC and FCI solver to perform the gradient descent based optimization in the two cases.

The density mismatch from BE with FCI and VMC eigensolver is plotted against each other. Note that since a first-quantized real space formulation of VMC is used, the VMC and FCI (second-quantized) calculation are performed in slightly different Hilbert space. This leads to different values of density mismatch from FCI and VMC in the initial BE iteration. For ease of comparison, the FCI density mismatch is rescaled by a factor of 3.8 to match the initial mismatch of the VMC.

In Fig. 6.2, an initial exponential convergence on the density mismatch is observed for both VMC and FCI solver. The VMC solver later plateau at a mismatch of roughly 2×10^{-3}, due to the statistic fluctuation on the estimated 1-RDM matrix elements from finite number of MC samples. The plateaued value of the density mismatch is calculated by taking an average of all BE iterations after iteration 30. It can be seen that as the number of MC samples increased from 40k to 640k, the fluctuations as well as the plateaued values of the density mismatch are reduced. Note since the energy is roughly quadratic in 1-RDM elements, therefore an error bar of 2×10^{-3} on 1-RDM elements will propagate to a reasonable accuracy (mH) on the energy.

D.9.3 SWAP Test Circuit in Quantum Bootstrap Embedding

Two H_4 molecules each with a bond length of 0.5 Å under STO-3G basis is used to generate Fig. 6.6 of the main text. For ease of classical simulation, the two-body terms are ignored resulting in an non-interacting Hamiltonian. An overall circuit for the SWAP estimation composed of two QPE eigensolver is given in Fig. D.5 which uses 27 qubits overall. Each QPE circuit is further displayed in Fig. D.6 for clarity. The resulting histogram of the QPE estimation is shown in Fig. D.7, which is proportional to the probability given in Eq. (D.97). The highest peak with a measurement outcome of "11011" on the 5 evaluation qubits corresponds to the many-body ground state of H_4. Post-selection is performed with this peak to ensure SWAP test is indeed estimating the overlap between the ground state. In our example, the success probability of this post-selection is roughly 0.7. Therefore, the overall post-selection success probability for two QPE solvers is $0.7^2 \approx 0.5$.

D.9.4 Quantum Bootstrap Embedding Calculation

In this section, we give more computational details of the QBE calculation in infinite sampling limit as in Fig. 6.5 of the main text.

QBE Iterations. For the quadratic penalty optimization, we set the penalty parameter $\lambda = 1$ initially. In each BE iteration, the penalty parameter λ is increased by a factor of 25. For the linear constraint optimization, a gradient descent algorithm as described in Alg. 11 is used, where the initial step size of the gradient descent is set to 1.

Calculation of Total System Energy One important step after BE calculation converges is to reconstruct the total system observables. One common observable is the total system energy. In classical BE (CBE), the total system energy can be reconstructed from 1- and 2-RDM of each fragment projected to the center sites. To be concrete, the classical BE energy is defined as

$$E_{\mathrm{CBE}} = \sum_{A}^{N_{\mathrm{frag}}} \sum_{p \in \mathbb{C}_A} \left[\sum_{q}^{2N_A} \left(h_{pq}^{(A)} - \frac{1}{2} G_{pq}^{(A),env} \right) P_{pq}^{(A)} + \frac{1}{2} \sum_{qrs}^{2N_A} V_{pqrs}^{(A)} \Gamma_{pqrs}^{(A)} \right], \tag{D.98}$$

where $G^{(A),env}$ is the Coulomb-exchange part of the Fock matrix in the embedding basis, $h^{(A)}$ and $V^{(A)}$ are the fragment 1- and 2-electron integrals, $P^{(A)}$ and $\Gamma^{(A)}$ are the fragment one- and two-electron reduced density matrices, respectively.

In quantum bootstrap embedding (QBE), one can certainly perform fermionic 1- and 2-RDM tomography on a quantum computer for the fragments, and then use Eq. (D.98) to compute the total system energy. In our case, for simplicity, we define a similar notion of QBE energy which is used in Fig. 6.5 of the main text to compute the energy

$$E_{\mathrm{QBE}} = \sum_{A=1}^{N_{\mathrm{frag}}} \mathrm{Tr}[\rho^{(A)} H_{\mathbb{C}}^{(A)}], \tag{D.99}$$

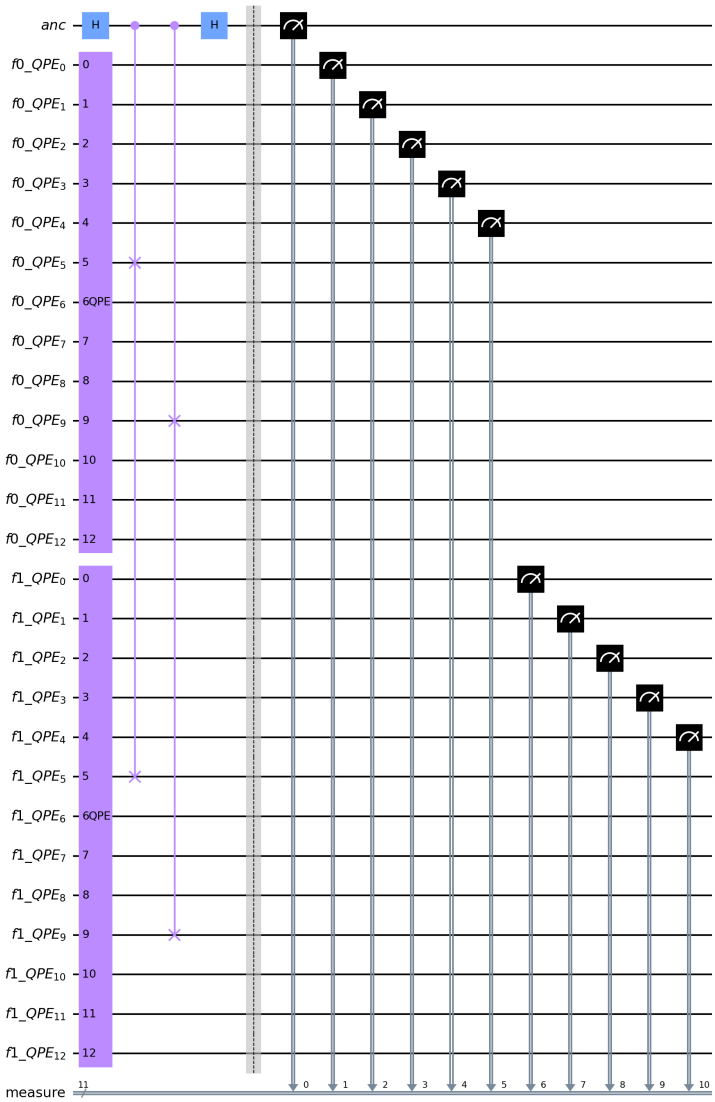


Figure D.5: `SWAP` test circuit between two H_4 molecule with one overlapping site. Note that each QPE uses 8 system qubits and 5 evaluation qubits. The top qubit is the control ancilla for `SWAP` test. The overall circuit is composed of 27 qubits.

where $\rho^{(A)}$ is the full density matrix of fragment A, and $H_{\mathbb{C}}^{(A)}$ is a projected version of $H^{(A)}$ that involves the center site and interaction between center and the rest sites on fragment

Figure D.6: A QPE circuit with for H_4 molecule, where a Hartree-Fock initial trial state preparation circuit is also showed at the beginning. The circuit block labelled as "Hamiltonian" after the four initial Pauli X are a basis transformation unitary from canonical MO basis to localized orbitals (LOs).

A. More concretely, we write $H_{\mathbb{C}}^{(A)}$ as a sum of one- and two-body terms

$$
\begin{aligned}
H_{\mathbb{C}}^{(A)} &= H_{1,\mathbb{C}}^{(A)} + H_{2,\mathbb{C}}^{(A)}, \\
H_{1,\mathbb{C}}^{(A)} &= \frac{1}{2}\left[\sum_{p\in\mathbb{C}_A}\sum_{q}^{2N_A} h_{pq}^{(A)} + \sum_{q\in\mathbb{C}_A}\sum_{p}^{2N_A} h_{pq}^{(A)}\right] a_p^\dagger a_q, \\
H_{2,\mathbb{C}}^{(A)} &= \frac{1}{8}\left[\sum_{p\in\mathbb{C}_A}\sum_{qrs}^{2N_A} V_{pqrs}^{(A)} + \sum_{q\in\mathbb{C}_A}\sum_{prs}^{2N_A} V_{pqrs}^{(A)} + \sum_{r\in\mathbb{C}_A}\sum_{pqs}^{2N_A} V_{pqrs}^{(A)} + \sum_{s\in\mathbb{C}_A}\sum_{pqr}^{2N_A} V_{pqrs}^{(A)}\right] a_p^\dagger a_q^\dagger a_s a_r.
\end{aligned}
\tag{D.100}
$$

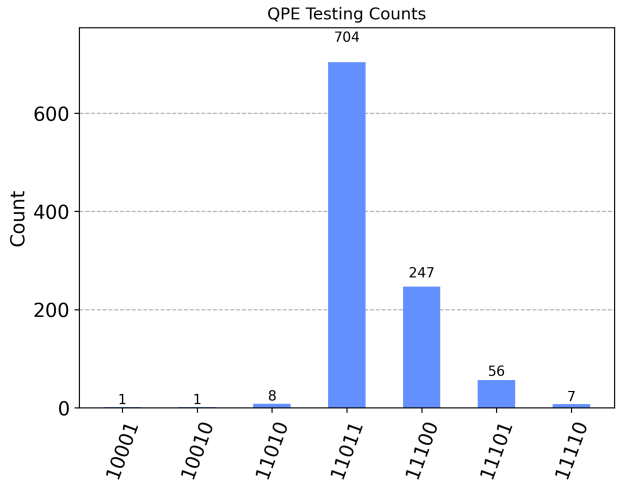


Figure D.7: Histogram of the measurement outcome of the 5 evaluation qubits for estimating the H_4 ground state energy with 1024 shots. Post-selection is performed on the highest peak "11011".

Note that E_{QBE} as defined in Eq. (D.100) does not exactly equal to E_{CBE} as used in classical BE calculation. There may be better ways to reconstruct E_{CBE} on a quantum computer without using fermionic density matrix tomography, and we leave this for future investigation.

D.9.5 Details of VQE Eigensolver

In this section, we discuss how the Variational Quantum Eigensolver (VQE) can be applied as a subroutine in the Quantum Bootstrap Embedding (QBE) method for computing the ground state energy and ground state vector of different fragment Hamiltonians. Our simulations with VQE were carried out on a noiseless simulator provided in Qiskit [Abr+19a].

VQE is a hybrid quantum-classical algorithm for computing the ground state (and excited states) of a Hamiltonian H, using the variation principal $\langle\psi|H|\psi\rangle \geq E_g$ where $|\psi\rangle$ is the normalized quantum state and E_g is the true ground state energy of H. In QBE, we may use VQE to find the ground state energy and ground state vector of the fragment Hamiltonians. Given an n-qubit fragment Hamiltonian $H^{(A)}$ of fragment A and appropriately chosen ansatz circuit $U_A(\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$, the steps of VQE are:

1. Prepare the state $|\psi(\boldsymbol{\theta})\rangle = U_A(\boldsymbol{\theta})|\psi_0\rangle$ on the quantum device where $|\psi_0\rangle$ is an initial state. $|\psi_0\rangle$ is typically chosen such that it can be prepared classically efficiently and has a non-vanishing overlap with the exact ground state of the fragment Hamiltonian (e.g., Hartree-Fock state of $H^{(A)}$).
2. Measure the expectation value $\langle\psi(\boldsymbol{\theta})|H^{(A)}|\psi(\boldsymbol{\theta})\rangle$ given a budget of n_{shots} shots (which we will specify later). We do this using the largest-degree first (LDF) algorithm [WP67] considering the Pauli decomposition of $H^{(A)}$.
3. Update ansatz parameters $\boldsymbol{\theta}$ through a classical optimizer, that will minimize the expectation value. The classical optimizer may involve computation of gradient steps or be a gradient-free method.

4. Repeat the above steps 1-3 until convergence or stopping criteria (e.g., maximum number of iterations, norm of gradient, etc.) is met.
5. Output $|\psi(\boldsymbol{\theta})\rangle$ as the ground state vector and $\langle\psi(\boldsymbol{\theta})|H^{(A)}|\psi(\boldsymbol{\theta})\rangle$ as the ground state energy, using the final values of the ansatz parameters $\boldsymbol{\theta}$.

The example considered using a VQE solver in this work is a 4-qubit random spin model and a perturbed H_4 linear molecule with open boundary condition under STO-3G basis. The Hamiltonian for the H_4 molecule is generated first by using a H-H bond length of 1 Å(with atom labeled as 1, 2,3, 4 from left to right). The H_4 molecule is then fragmented into two fragments, where fragment A has atom 1,2,3 with atom 4 as a bath site. Similarly, fragment B has atom 2,3,4 and uses atom 1 as the bath site. Since the native H_4 is too small to perform any meaningful bootstrap embedding (the two fragment already matches initially), we manually perturb atom 3 on both fragment A (right edge site) and fragment B (center site) by adding a chemical potential of $+1$ and -0.5, respectively.

The initial qubit Hamiltonian of each fragment is then obtained considering the Jordan-Wigner encoding. The fragment Hamiltonians are stored as set of tuples $\{(\alpha_Q, Q)\}$ corresponding to its Pauli decomposition $H_f = \sum_Q \alpha_Q Q$ where $Q \in \{I, X, Y, Z\}^{\otimes n}$ is a distinct Pauli operator and α_Q is the corresponding non-zero coefficient. This allows for a more compact storage rather than holding the entire matrices corresponding to the Hamiltonians in memory as there are only a polynomial in n number of non-zero coefficients α_Q for each fragment Hamiltonian. Moreover, from one iteration to the next in QBE with linear constraints, this set of tuples can be efficiently updated by updating a coefficient or appending a new Pauli operator Q' along with its coefficient $\alpha_{Q'}$ according to Eq. (6.17), where $\alpha_{Q'}$ is one component of $\boldsymbol{\lambda}_B^{(A)}$, and Q' is one-component of $I \otimes \boldsymbol{\Sigma}_\mathbf{r} \otimes I$.

The ansatz of each fragment in the 4-qubit random spin model is considered to be the two-local ansatz and the initial state is chosen to be $|+\rangle^{\otimes 3}$ state. The ansatz of each fragment in the H_4 chain is considered to be the UCCSD ansatz and the initial state is chosen to the Hartree-Fock state. We use the classical optimizer of the quasi-Newton method L-BFGS. The learning rate in QBE for all models is fixed to a constant value of 0.1 across iterations.

Fig. D.8 shows the convergence of the density mismatch for the 4-qubit spin model as the number of eigenvalue calls (note BE iteration number is roughly proportional to the number of eigensolver calls), comparing the VQE results (blue symbols) and exact classical eigensolver (green symbols). As is expected, the mismatch converges exponentially as the number of eigensolver calls increases, and the VQE results closely follow the exact results for large to intermediate density mismatch values. As the mismatch is reduced to roughly 10^{-4} at about 500 eigensolver calls, the VQE results start to deviates from the exact results. The shaded area shows the uncertainty of the VQE results. The inset (red symbols) plots the deviation of the expected density mismatch (averaged over 100 independent runs) obtained from VQE with respect to the exact results, which plateaus around 10^{-5}. We tentatively attribute this deviation from the exact result to the intrinsic ansatz truncation error of VQE.

In Fig. D.9(a), we show the convergence of the density mismatch on the H_4 chain as a number of eigenvalue calls, comparing the VQE results (blue symbols) and exact classical eigensolver (green symbols). As observed earlier for the random spin model, the mismatch converges exponentially as the number of eigensolver calls increases, and the VQE results closely follow the exact results up to around 30 eigensolver calls. As the mismatch is reduced

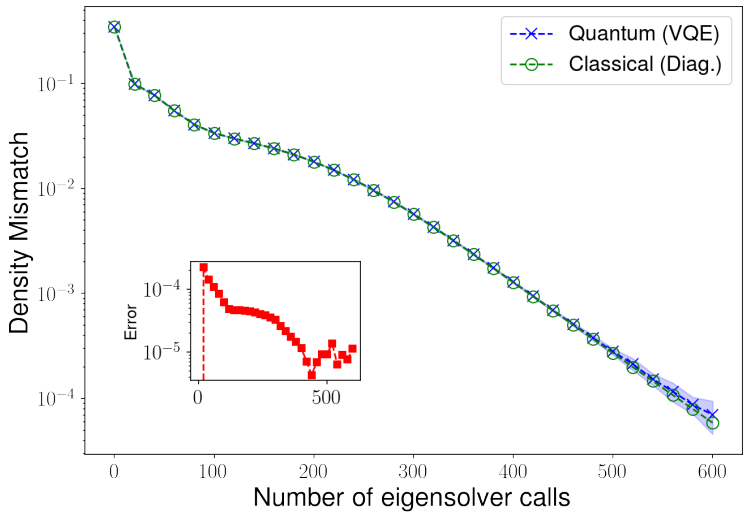
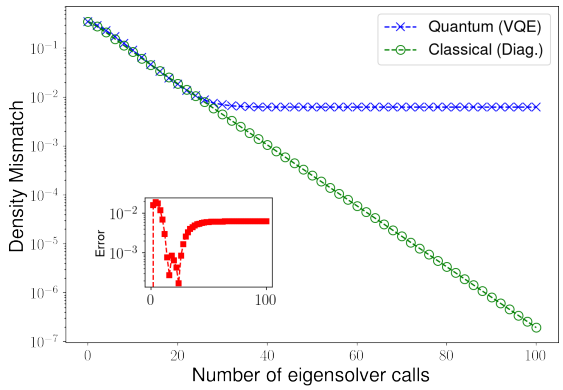
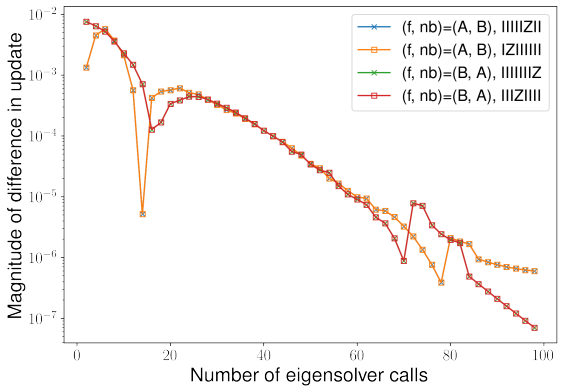


Figure D.8: Quantum bootstrap embedding convergence of the density mismatch for the 4-qubit spin model versus the number of eigenvalue calls, comparing VQE simulation (blue cross) and classical exact diagonalization (green circle). The blue shaded area shows the standard deviation from VQE estimations. The inset shows the absolute error from the VQE estimation of the mismatch to the exact value from exact diagonalization, versus number of eigenvalue calls. The number of shots considered in each step of VQE for measuring the ansatz is fixed at 10^4.

below 10^{-2}, the VQE results start to deviate from the exact results. We tentatively attribute this deviation from the exact result to the intrinsic ansatz truncation error of VQE. To get a finer-grained undertanding, in Fig D.9(b), the difference in the updates between the exact solver and the VQE solver at each QBE iteration is plotted. We can see that this difference is small throughout. This difference decreases as BE iteration goes, because the absolute value of the update also decrease due to increasingly better mismatch.



(a) Convergence behavior of density mismatch



(b) Comparison of updates in fragment Hamiltonians with QBE iterations

Figure D.9: Quantum bootstrap embedding on H_4 chain. (a) We compare the convergence of the density mismatch versus the number of eigenvalue calls, comparing VQE simulation (blue cross) and classical exact diagonalization (green circle). The inset shows the absolute error from the VQE estimation of the mismatch to the exact value from exact diagonalization, versus number of eigenvalue calls. The number of shots considered in each step of VQE for measuring the ansatz is fixed at 10^4. (b) We compare the difference in the magnitude of the updates of the Pauli terms (being updated or appended during QBE) in the fragment Hamiltonians between the exact solver and the VQE solver, for each pair of fragment f and neighbor nb with number of iterations. For each pair of (f, nb), we indicate the Pauli term for each trend in the legend.

References

[Fey82] Richard P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21.6 (June 1982), pp. 467–488. DOI: https://doi.org/10.1007/BF02650179.

[Llo96] Seth Lloyd. "Universal quantum simulators". In: *Science* 273.5278 (1996), pp. 1073–1078.

[Chi+18] Andrew M Childs, Dmitri Maslov, Yunseong Nam, Neil J Ross, and Yuan Su. "Toward the first quantum simulation with quantum speedup". In: *Proceedings of the National Academy of Sciences* 115.38 (2018), pp. 9456–9461.

[HHL09] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. "Quantum Algorithm for Linear Systems of Equations". In: *Phys. Rev. Lett.* 103 (15 Oct. 2009), p. 150502. DOI: 10.1103/PhysRevLett.103.150502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.150502.

[Sho94] P.W. Shor. "Algorithms for quantum computation: discrete logarithms and factoring". In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. 1994, pp. 124–134. DOI: 10.1109/SFCS.1994.365700.

[WBL12] Nathan Wiebe, Daniel Braun, and Seth Lloyd. "Quantum Algorithm for Data Fitting". In: *Phys. Rev. Lett.* 109 (5 Aug. 2012), p. 050505. DOI: 10.1103/PhysRevLett.109.050505. URL: https://link.aps.org/doi/10.1103/PhysRevLett.109.050505.

[BS17] F. L. Brandao and K. M. Svore. "Quantum Speed-Ups for Solving Semidefinite Programs". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2017, pp. 415–426. DOI: 10.1109/FOCS.2017.45. URL: https://doi.ieeecomputersociety.org/10.1109/FOCS.2017.45.

[McA+20] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. "Quantum computational chemistry". In: *Rev. Mod. Phys.* 92 (1 Mar. 2020), p. 015003. DOI: 10.1103/RevModPhys.92.015003. URL: https://link.aps.org/doi/10.1103/RevModPhys.92.015003.

[Kas+11] Ivan Kassal, James D Whitfield, Alejandro Perdomo-Ortiz, Man-Hong Yung, and Alán Aspuru-Guzik. "Simulating chemistry using quantum computers". In: *Annual review of physical chemistry* 62 (2011), pp. 185–207.

[Ber+14] Dominic W Berry, Andrew M Childs, Richard Cleve, Robin Kothari, and Rolando D Somma. "Exponential improvement in precision for simulating sparse Hamiltonians". In: *Proceedings of the forty-sixth annual ACM symposium on Theory*

of computing. 2014, pp. 283–292. URL: https://dl.acm.org/doi/10.1145/2591796.2591854.

[Ber+17] Dominic W Berry, Andrew M Childs, Aaron Ostrander, and Guoming Wang. "Quantum algorithm for linear differential equations with exponentially improved dependence on precision". In: *Communications in Mathematical Physics* 356 (2017), pp. 1057–1081.

[Kro23] Hari Krovi. "Improved quantum algorithms for linear and nonlinear differential equations". In: *Quantum* 7 (2023), p. 913.

[Bog+23] Michael Bogobowicz, Rodney Zemmel, Scarlett Gao, Mateusz Masiowski, Niko Mohr, Henning Soller, and Matija Zesko. "Quantum technology sees record investments, progress on talent gap". In: (Apr. 24, 2023). URL: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/quantum-technology-sees-record-investments-progress-on-talent-gap.

[Cas20] Sara Castellanos. "AI, Quantum R & D Funding to Remain a Priority Under Biden". In: *The Wall Street Journal* (Nov. 9, 2020). URL: https://www.wsj.com/articles/ai-quantum-r-d-funding-to-remain-a-priority-under-biden-11604944800 (visited on 11/29/2023).

[Bru+19] Colin D Bruzewicz, John Chiaverini, Robert McConnell, and Jeremy M Sage. "Trapped-ion quantum computing: Progress and challenges". In: *Applied Physics Reviews* 6.2 (2019).

[Kja+20] Morten Kjaergaard, Mollie E. Schwartz, Jochen Braumüller, Philip Krantz, Joel I.-J. Wang, Simon Gustavsson, and William D. Oliver. "Superconducting Qubits: Current State of Play". In: *Annual Review of Condensed Matter Physics* 11.1 (2020), pp. 369–395. DOI: 10.1146/annurev-conmatphys-031119-050605. eprint: https://doi.org/10.1146/annurev-conmatphys-031119-050605. URL: https://doi.org/10.1146/annurev-conmatphys-031119-050605.

[Che+14] Yu Chen et al. "Qubit Architecture with High Coherence and Fast Tunable Coupling". In: *Phys. Rev. Lett.* 113 (22 Nov. 2014), p. 220502. DOI: 10.1103/PhysRevLett.113.220502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.220502.

[Ngu+19] Long B. Nguyen, Yen-Hsiang Lin, Aaron Somoroff, Raymond Mencia, Nicholas Grabon, and Vladimir E. Manucharyan. "High-Coherence Fluxonium Qubit". In: *Phys. Rev. X* 9 (4 Nov. 2019), p. 041041. DOI: 10.1103/PhysRevX.9.041041. URL: https://link.aps.org/doi/10.1103/PhysRevX.9.041041.

[Bar+14] Rami Barends, Julian Kelly, Anthony Megrant, Andrzej Veitia, Daniel Sank, Evan Jeffrey, Ted C White, Josh Mutus, Austin G Fowler, Brooks Campbell, et al. "Superconducting quantum circuits at the surface code threshold for fault tolerance". In: *Nature* 508.7497 (2014), pp. 500–503.

[Hon+20] Sabrina S. Hong, Alexander T. Papageorge, Prasahnt Sivarajah, Genya Crossman, Nicolas Didier, Anthony M. Polloreno, Eyob A. Sete, Stefan W. Turkowski, Marcus P. da Silva, and Blake R. Johnson. "Demonstration of a parametrically activated entangling gate protected from flux noise". In: *Phys. Rev. A* 101 (1 Jan. 2020), p. 012302. DOI: 10.1103/PhysRevA.101.012302. URL: https://link.aps.org/doi/10.1103/PhysRevA.101.012302.

[GCS17] Jay M Gambetta, Jerry M Chow, and Matthias Steffen. "Building logical qubits in a superconducting quantum computing system". In: *npj quantum information* 3.1 (2017), p. 2.

[Kel+15] Julian Kelly, Rami Barends, Austin G Fowler, Anthony Megrant, Evan Jeffrey, Theodore C White, Daniel Sank, Josh Y Mutus, Brooks Campbell, Yu Chen, et al. "State preservation by repetitive error detection in a superconducting quantum circuit". In: *Nature* 519.7541 (2015), pp. 66–69.

[Aru+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779 (2019), pp. 505–510.

[22a] *Expanding the IBM Quantum roadmap to anticipate the future of quantum-centric supercomputing*. https://research.ibm.com/blog/ibm-quantum-roadmap-2025. Accessed: 2023-11-29. May 10, 2022.

[22b] *Google Quantum AI: roadmap and mission*. https://ai.google/static/documents/approach-quantum-computing.pdf. Accessed: 2023-11-29. Sept. 7, 2022.

[20] *Scaling IonQ's Quantum Computers: The Roadmap*. https://ionq.com/posts/december-09-2020-scaling-quantum-computer-roadmap. Accessed: 2023-11-29. Dec. 9, 2020.

[Pre18] John Preskill. "Quantum computing in the NISQ era and beyond". In: *Quantum* 2 (2018), p. 79.

[Ang+22] James Ang, Gabriella Carini, Yanzhu Chen, Isaac Chuang, Michael Austin DeMarco, Sophia E Economou, Alec Eickbusch, Andrei Faraon, Kai-Mei Fu, Steven M Girvin, et al. "Architectures for multinode superconducting quantum computers". In: *arXiv preprint arXiv:2212.06167* (2022).

[Haa+17] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. "Sample-optimal tomography of quantum states". In: *IEEE Transactions on Information Theory* 63.9 (2017), pp. 5628–5641.

[Leu00] Debbie Wun Chi Leung. "Towards robust quantum computation". PhD thesis. Stanford University, California, Jan. 2000.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge, UK: Cambridge University Press, 2010. DOI: 10.1017/CBO9780511976667.

[She+16] Sarah Sheldon, Easwar Magesan, Jerry M. Chow, and Jay M. Gambetta. "Procedure for systematically tuning up cross-talk in the cross-resonance gate". In: *Phys. Rev. A* 93 (6 June 2016), 060302(R). DOI: 10.1103/PhysRevA.93.060302. URL: https://link.aps.org/doi/10.1103/PhysRevA.93.060302.

[Cro+19] Andrew W. Cross, Lev S. Bishop, Sarah Sheldon, Paul D. Nation, and Jay M. Gambetta. "Validating quantum computers using randomized model circuits". In: *Phys. Rev. A* 100 (3 Sept. 2019), p. 032328. DOI: 10.1103/PhysRevA.100.032328. URL: https://link.aps.org/doi/10.1103/PhysRevA.100.032328.

[Wac+21] Andrew Wack, Hanhee Paik, Ali Javadi-Abhari, Petar Jurcevic, Ismael Faro, Jay M Gambetta, and Blake R Johnson. "Scale, Quality, and Speed: three key attributes to measure the performance of near-term quantum computers". In: *arXiv preprint arXiv:2110.14108* (2021).

[McK+23] David C McKay, Ian Hincks, Emily J Pritchett, Malcolm Carroll, Luke CG Govia, and Seth T Merkel. "Benchmarking Quantum Processor Performance at Scale". In: *arXiv preprint arXiv:2311.05933* (2023).

[Moo98] Gordon E Moore. "Cramming more components onto integrated circuits". In: *Proceedings of the IEEE* 86.1 (1998), pp. 82–85.

[KSL95] Andreas Kuehlmann, Arvind Srinivasan, and David P LaPotin. "Verity—a formal verification program for custom CMOS circuits". In: *IBM Journal of Research and Development* 39.1.2 (1995), pp. 149–165.

[Lin94] Peter Lindskog. *Algorithms and tools for system identification using prior knowledge*. Tech. rep. Linkoeping Univ.(Sweden). Dept. of Electrical Engineering, 1994.

[SM71] A.P. Sage and J.L. Melsa. *System Identification*. Mathematics in science and engineering. Academic Press, 1971. ISBN: 9780126144505. URL: https://books.google.com/books?id=iJXoS9OOb0kC.

[ÅE71] K.J. Åström and P. Eykhoff. "System identification—A survey". In: *Automatica* 7.2 (1971), pp. 123–162. ISSN: 0005-1098. DOI: https://doi.org/10.1016/0005-1098(71)90059-8. URL: https://www.sciencedirect.com/science/article/pii/0005109871900598.

[Joh96] Tor A. Johansen. "Identification of non-linear systems using empirical data and prior knowledge—an optimization approach". In: *Automatica* 32.3 (1996), pp. 337–356. ISSN: 0005-1098. DOI: https://doi.org/10.1016/0005-1098(95)00146-8. URL: https://www.sciencedirect.com/science/article/pii/0005109895001468.

[Wah91] B. Wahlberg. "System identification using Laguerre models". In: *IEEE Transactions on Automatic Control* 36.5 (1991), pp. 551–562. DOI: 10.1109/9.76361.

[Kan+17] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". In: *Nature* 549 (2017), pp. 242–246.

[McC+16] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". In: *New Journal of Physics* 18.2 (2016), p. 023023.

[OW16] Ryan O'Donnell and John Wright. "Efficient quantum tomography". In: *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. 2016, pp. 899–912.

[BBO20] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien. "Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States". In: *Phys. Rev. X* 10 (3 Sept. 2020), p. 031064. DOI: 10.1103/PhysRevX.10.031064. URL: https://link.aps.org/doi/10.1103/PhysRevX.10.031064.

[RBM18] Nicholas C Rubin, Ryan Babbush, and Jarrod McClean. "Application of fermionic marginal constraints to hybrid quantum algorithms". In: *New Journal of Physics* 20.5 (May 2018), p. 053020. DOI: 10.1088/1367-2630/aab919. URL: https://dx.doi.org/10.1088/1367-2630/aab919.

[Cra+10] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai

Liu. "Efficient quantum state tomography". In: *Nature Communications* 1.1 (2010), pp. 1–7.

[Mon17a] Ashley Montanaro. "Learning stabilizer states by Bell sampling". In: *arXiv preprint arXiv:1707.04012* (2017).

[Tót+10] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter. "Permutationally Invariant Quantum Tomography". In: *Phys. Rev. Lett.* 105 (25 Dec. 2010), p. 250403. DOI: 10.1103/PhysRevLett.105.250403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.105.250403.

[Ans+21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. "Sample-efficient learning of interacting quantum systems". In: *Nature Physics* 17.8 (2021), pp. 931–935.

[HKT21] Jeongwan Haah, Robin Kothari, and Ewin Tang. "Optimal learning of quantum Hamiltonians from high-temperature Gibbs states". In: *arXiv preprint arXiv:2108.04842* (2021).

[CN97] Isaac L Chuang and Michael A Nielsen. "Prescription for experimental determination of the dynamics of a quantum black box". In: *Journal of Modern Optics* 44.11-12 (1997), pp. 2455–2467. DOI: 10.1080/09500349708231894. eprint: https://www.tandfonline.com/doi/pdf/10.1080/09500349708231894. URL: https://www.tandfonline.com/doi/abs/10.1080/09500349708231894.

[MRL08] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. "Quantum-process tomography: Resource analysis of different strategies". In: *Phys. Rev. A* 77 (3 Mar. 2008), p. 032322. DOI: 10.1103/PhysRevA.77.032322. URL: https://link.aps.org/doi/10.1103/PhysRevA.77.032322.

[Car22] Matthias C Caro. "Learning quantum processes and Hamiltonians via the Pauli transfer matrix". In: *arXiv preprint arXiv:2212.04471* (2022).

[YRC20] Yuxiang Yang, Renato Renner, and Giulio Chiribella. "Optimal Universal Programming of Unitary Gates". In: *Phys. Rev. Lett.* 125 (21 Nov. 2020), p. 210501. DOI: 10.1103/PhysRevLett.125.210501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.210501.

[Bis+10] Alessandro Bisio, Giulio Chiribella, Giacomo Mauro D'Ariano, Stefano Facchini, and Paolo Perinotti. "Optimal quantum learning of a unitary transformation". In: *Phys. Rev. A* 81 (3 Mar. 2010), p. 032324. DOI: 10.1103/PhysRevA.81.032324. URL: https://link.aps.org/doi/10.1103/PhysRevA.81.032324.

[Haa+23] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. "Query-optimal estimation of unitary channels in diamond distance". In: *arXiv preprint arXiv:2302.14066* (2023).

[Low09] Richard A. Low. "Learning and testing algorithms for the Clifford group". In: *Phys. Rev. A* 80 (5 Nov. 2009), p. 052314. DOI: 10.1103/PhysRevA.80.052314. URL: https://link.aps.org/doi/10.1103/PhysRevA.80.052314.

[LC22] Ching-Yi Lai and Hao-Chung Cheng. "Learning quantum circuits of some T gates". In: *IEEE Transactions on Information Theory* 68.6 (2022), pp. 3951–3964.

[CNY] Thomas Chen, Shivam Nadimpalli, and Henry Yuen. "Testing and Learning Quantum Juntas Nearly Optimally". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1163–1185. DOI: 10.

1137/1.9781611977554.ch43. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611977554.ch43. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611977554.ch43.

[Har+10] Aram W. Harrow, Avinatan Hassidim, Debbie W. Leung, and John Watrous. "Adaptive versus nonadaptive strategies for quantum channel discrimination". In: *Phys. Rev. A* 81 (3 Mar. 2010), p. 032339. DOI: 10.1103/PhysRevA.81.032339. URL: https://link.aps.org/doi/10.1103/PhysRevA.81.032339.

[HW12] Aram W Harrow and Andreas Winter. "How many copies are needed for state discrimination?" In: *IEEE Transactions on Information Theory* 58.1 (2012), pp. 1–2.

[Inn+20] Luca Innocenti, Leonardo Banchi, Alessandro Ferraro, Sougato Bose, and Mauro Paternostro. "Supervised learning of time-independent Hamiltonians for gate design". In: *New Journal of Physics* 22.6 (June 2020), p. 065001. DOI: 10.1088/1367-2630/ab8aaf. URL: https://doi.org/10.1088/1367-2630/ab8aaf.

[GLM04a] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum-Enhanced Measurements: Beating the Standard Quantum Limit". In: *Science* 306.5700 (2004), pp. 1330–1336. DOI: 10.1126/science.1104149. eprint: https://www.science.org/doi/pdf/10.1126/science.1104149. URL: https://www.science.org/doi/abs/10.1126/science.1104149.

[GLM11a] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Advances in quantum metrology". In: *Nature Photonics* 5.4 (2011), p. 222.

[Aru+23] Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. "Optimal Algorithms for Learning Quantum Phase States". In: *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*. Ed. by Omar Fawzi and Michael Walter. Vol. 266. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 3:1–3:24. ISBN: 978-3-95977-283-9. DOI: 10.4230/LIPIcs.TQC.2023.3. URL: https://drops.dagstuhl.de/opus/volltexte/2023/18313.

[HKT22] Jeongwan Haah, Robin Kothari, and Ewin Tang. "Optimal learning of quantum Hamiltonians from high-temperature Gibbs states". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. 2022, pp. 135–146. DOI: 10.1109/FOCS54457.2022.00020.

[Bak+23] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. "Learning quantum Hamiltonians at any temperature in polynomial time". In: *arXiv preprint arXiv:2310.02243* (2023).

[GCC22] Andi Gu, Lukasz Cincio, and Patrick J Coles. "Practical black box hamiltonian learning". In: *arXiv preprint arXiv:2206.15464* (2022).

[Fra+22] Daniel Stilck Franca, Liubov A Markovich, VV Dobrovitski, Albert H Werner, and Johannes Borregaard. "Efficient and robust estimation of many-qubit Hamiltonians". In: *arXiv preprint arXiv:2205.09567* (2022).

[Hua+23] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. "Learning Many-Body Hamiltonians with Heisenberg-Limited Scaling". In: *Phys. Rev. Lett.* 130 (20 May 2023), p. 200403. DOI: 10.1103/PhysRevLett.130.200403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.200403.

[Sha+11] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz. "Estimation of many-body quantum Hamiltonians via compressive sensing". In: *Phys. Rev. A* 84 (1 July 2011), p. 012107. DOI: 10.1103/PhysRevA.84.012107. URL: https://link.aps.org/doi/10.1103/PhysRevA.84.012107.

[Yu+23] Wenjun Yu, Jinzhao Sun, Zeyao Han, and Xiao Yuan. "Robust and efficient Hamiltonian learning". In: *Quantum* 7 (2023), p. 1045.

[DOS23] Alicja Dutkiewicz, Thomas E O'Brien, and Thomas Schuster. "The advantage of quantum control in many-body Hamiltonian learning". In: *arXiv preprint arXiv:2304.07172* (2023).

[FL11] Steven T. Flammia and Yi-Kai Liu. "Direct Fidelity Estimation from Few Pauli Measurements". In: *Phys. Rev. Lett.* 106 (23 June 2011), p. 230501. DOI: 10.1103/PhysRevLett.106.230501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.230501.

[MGE11] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. "Scalable and Robust Randomized Benchmarking of Quantum Processes". In: *Phys. Rev. Lett.* 106 (18 May 2011), p. 180504. DOI: 10.1103/PhysRevLett.106.180504. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.180504.

[Erh+19] Alexander Erhard, Joel J Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban A Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. "Characterizing large-scale quantum computers via cycle benchmarking". In: *Nature communications* 10.1 (2019), p. 5347.

[BY20] Robin Blume-Kohout and Kevin C Young. "A volumetric framework for quantum computer benchmarks". In: *Quantum* 4 (2020), p. 362.

[McC+19] Alexander J McCaskey, Zachary P Parks, Jacek Jakowski, Shirley V Moore, Titus D Morris, Travis S Humble, and Raphael C Pooser. "Quantum chemistry as a benchmark for near-term quantum computers". In: *npj Quantum Information* 5.1 (2019), pp. 1–8.

[Eis+20] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. "Quantum certification and benchmarking". In: *Nature Reviews Physics* 2.7 (2020), pp. 382–390.

[Car+21] Jose Carrasco, Andreas Elben, Christian Kokail, Barbara Kraus, and Peter Zoller. "Theoretical and Experimental Perspectives of Quantum Verification". In: *PRX Quantum* 2 (1 Mar. 2021), p. 010102. DOI: 10.1103/PRXQuantum.2.010102. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.010102.

[GKK19] Alexandru Gheorghiu, Theodoros Kapourniotis, and Elham Kashefi. "Verification of quantum computation: An overview of existing approaches". In: *Theory of computing systems* 63 (2019), pp. 715–808.

[ŠB20] Ivan Šupić and Joseph Bowles. "Self-testing of quantum systems: a review". In: *Quantum* 4 (2020), p. 337.

[KR21] Martin Kliesch and Ingo Roth. "Theory of Quantum System Certification". In: *PRX Quantum* 2 (1 Jan. 2021), p. 010201. DOI: 10.1103/PRXQuantum.2.010201. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.010201.

[LMS22] Noah Linden, Ashley Montanaro, and Changpeng Shao. "Quantum vs. classical algorithms for solving the heat equation". In: *Communications in Mathematical Physics* 395.2 (2022), pp. 601–641.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". In: *Nature Physics* 16.10 (June 2020), pp. 1050–1057. DOI: 10.1038/s41567-020-0932-7. URL: https://doi.org/10.1038%2Fs41567-020-0932-7.

[Elb+23] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoıt Vermersch, and Peter Zoller. "The randomized measurement toolbox". In: *Nature Reviews Physics* 5.1 (Jan. 1, 2023), pp. 9–24. DOI: 10.1038/s42254-022-00535-2. URL: https://doi.org/10.1038/s42254-022-00535-2.

[BJS11] Michael J Bremner, Richard Jozsa, and Dan J Shepherd. "Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 467.2126 (2011), pp. 459–472.

[BMS17] Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. "Achieving quantum supremacy with sparse and noisy commuting quantum computations". In: *Quantum* 1 (2017), p. 8.

[Mon17b] Ashley Montanaro. "Quantum circuits and low-degree polynomials over $\mathbb{F}_2$". In: *Journal of Physics A: Mathematical and Theoretical* 50.8 (Jan. 2017), p. 084002. DOI: 10.1088/1751-8121/aa565f. URL: https://doi.org/10.1088/1751-8121/aa565f.

[JLS18] Zhengfeng Ji, Yi-Kai Liu, and Fang Song. "Pseudorandom quantum states". In: *Annual International Cryptology Conference*. Springer. 2018, pp. 126–152.

[BS19] Zvika Brakerski and Omri Shmueli. "(Pseudo) Random Quantum States with Binary Phase". In: *Theory of Cryptography Conference*. Springer. 2019, pp. 229–250.

[AQY21] Prabhanjan Ananth, Luowen Qian, and Henry Yuen. "Cryptography from Pseudorandom Quantum States". In: *arXiv preprint arXiv:2112.10020* (2021).

[Ira+21] Sandy Irani, Anand Natarajan, Chinmay Nirkhe, Sujit Rao, and Henry Yuen. "Quantum search-to-decision reductions and the state synthesis problem". In: *arXiv preprint arXiv:2111.02999* (2021).

[Ros+13] Matteo Rossi, Marcus Huber, Dagmar Bruß, and Chiara Macchiavello. "Quantum hypergraph states". In: *New Journal of Physics* 15.11 (2013), p. 113022.

[TMH19] Yuki Takeuchi, Tomoyuki Morimae, and Masahito Hayashi. "Quantum computational universality of hypergraph states with Pauli-X and Z basis measurements". In: *Scientific Reports* 9.1 (2019), pp. 1–14.

[Sug+23] Sho Sugiura, Arkopal Dutt, William J Munro, Sina Zeytinoğlu, and Isaac L Chuang. "Power of sequential protocols in hidden quantum channel discrimination". In: *arXiv preprint arXiv:2304.02053* (2023).

[DCS17] Rafał Demkowicz-Dobrza ński, Jan Czajkowski, and Pavel Sekatski. "Adaptive Quantum Metrology under General Markovian Noise". In: *Phys. Rev. X* 7 (4 Oct. 2017), p. 041009. DOI: 10.1103/PhysRevX.7.041009. URL: https://link.aps.org/doi/10.1103/PhysRevX.7.041009.

[Kat+22] Or Katz, Meirav Pinkas, Nitzan Akerman, and Roee Ozeri. "Quantum logic detection of collisions between single atom–ion pairs". In: *Nature Physics* 18.5 (2022), pp. 533–537. ISSN: 1745-2481. DOI: 10.1038/s41567-022-01517-y. URL: https://doi.org/10.1038/s41567-022-01517-y.

[Pec+21] M. Pechal, G. Salis, M. Ganzhorn, D. J. Egger, M. Werninghaus, and S. Filipp. "Characterization and Tomography of a Hidden Qubit". In: *Phys. Rev. X* 11 (4 Nov. 2021), p. 041032. DOI: 10.1103/PhysRevX.11.041032. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.041032.

[Dut+23a] Arkopal Dutt, Edwin Pednault, Chai Wah Wu, Sarah Sheldon, John Smolin, Lev Bishop, and Isaac L. Chuang. "Active learning of quantum system Hamiltonians yields query advantage". In: *Phys. Rev. Res.* 5 (3 July 2023), p. 033060. DOI: 10.1103/PhysRevResearch.5.033060. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.5.033060.

[EHF19] Tim J Evans, Robin Harper, and Steven T Flammia. "Scalable Bayesian Hamiltonian Learning". In: *arXiv preprint arXiv:1912.07636* (2019).

[Aru+20a] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Andreas Bengtsson, Sergio Boixo, Michael Broughton, Bob B Buckley, et al. "Observation of separated dynamics of charge and spin in the Fermi-Hubbard model". In: *arXiv preprint arXiv:2010.07965* (2020).

[Gra+12] Christopher E Granade, Christopher Ferrie, Nathan Wiebe, and David G Cory. "Robust online Hamiltonian learning". In: *New Journal of Physics* 14.10 (Oct. 2012), p. 103013. DOI: 10.1088/1367-2630/14/10/103013. URL: https://doi.org/10.1088/1367-2630/14/10/103013.

[Gra+17] Christopher Granade, Christopher Ferrie, Ian Hincks, Steven Casagrande, Thomas Alexander, Jonathan Gross, Michal Kononenko, and Yuval Sanders. "QInfer: Statistical inference software for quantum applications". In: *Quantum* 1 (2017), p. 5.

[Pat+23] Siddhartha Patra, Saeed S Jahromi, Sukhbinder Singh, and Roman Orus. "Efficient tensor network simulation of IBM's largest quantum processors". In: *arXiv preprint arXiv:2309.15642* (2023).

[Dut+23b] Arkopal Dutt, William Kirby, Rudy Raymond, Charles Hadfield, Sarah Sheldon, Isaac L Chuang, and Antonio Mezzacapo. "Practical Benchmarking of Randomized Measurement Methods for Quantum Chemistry Hamiltonians". In: *arXiv preprint arXiv:2312.07497* (2023).

[Lee+23] Seunghoon Lee, Joonho Lee, Huanchen Zhai, Yu Tong, Alexander M Dalzell, Ashutosh Kumar, Phillip Helms, Johnnie Gray, Zhi-Hao Cui, Wenyuan Liu, et al. "Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry". In: *Nature Communications* 14.1 (2023), p. 1952.

[Per+14] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor". In: *Nature Communications* 5.1 (July 2014), p. 4213. URL: http://dx.doi.org/10.1038/ncomms5213 (visited on 07/23/2014).

[McC+17] Jarrod R. McClean, Mollie E. Kimchi-Schwartz, Jonathan Carter, and Wibe A. de Jong. "Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states". In: *Phys. Rev. A* 95 (4 Apr. 2017), p. 042308. DOI: 10.1103/PhysRevA.95.042308. URL: https://link.aps.org/doi/10.1103/PhysRevA.95.042308.

[PM19] Robert M. Parrish and Peter L. McMahon. "Quantum Filter Diagonalization: Quantum Eigendecomposition without Full Quantum Phase Estimation". In: *arXiv preprint, arXiv:1909.08925* (2019). URL: https://doi.org/10.48550/arXiv.1909.08925.

[Had+22] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. "Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows". In: *Communications in Mathematical Physics* 391.3 (2022), pp. 951–967. DOI: 10.1007/s00220-022-04343-8. URL: https://doi.org/10.1007/s00220-022-04343-8 (visited on 05/01/2022).

[Hil+21] Stefan Hillmich, Charles Hadfield, Rudy Raymond, Antonio Mezzacapo, and Robert Wille. "Decision Diagrams for Quantum Measurements with Shallow Circuits". In: *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2021, pp. 24–34. DOI: 10.1109/QCE52317.2021.00018. URL: https://doi.ieeecomputersociety.org/10.1109/QCE52317.2021.00018.

[KG22] Dax Enshan Koh and Sabee Grewal. "Classical shadows with noise". In: *Quantum* 6 (2022), p. 776.

[YVI20] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F. Izmaylov. "Measuring All Compatible Operators in One Series of Single-Qubit Measurements Using Unitary Transformations". In: *Journal of Chemical Theory and Computation* 16.4 (Apr. 14, 2020), pp. 2400–2409. URL: https://doi.org/10.1021/acs.jctc.0c00008.

[VYI20] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F. Izmaylov. "Measurement optimization in the variational quantum eigensolver using a minimum clique cover". In: *The Journal of Chemical Physics* 152.12 (2020), p. 124114. DOI: 10.1063/1.5141458. eprint: https://doi.org/10.1063/1.5141458. URL: https://doi.org/10.1063/1.5141458.

[Wu+23] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. "Overlapped grouping measurement: A unified framework for measuring quantum states". In: *Quantum* 7 (2023), p. 896.

[Shl+23] Ariel Shlosberg, Andrew J. Jena, Priyanka Mukhopadhyay, Jan F. Haase, Felix Leditzky, and Luca Dellantonio. "Adaptive estimation of quantum observables". In: *Quantum* 7 (Jan. 2023), p. 906. ISSN: 2521-327X. DOI: 10.22331/q-2023-01-26-906. URL: https://doi.org/10.22331/q-2023-01-26-906.

[Liu+23] Yuan Liu, Oinam R. Meitei, Zachary E. Chin, Arkopal Dutt, Max Tao, Isaac L. Chuang, and Troy Van Voorhis. "Bootstrap Embedding on a Quantum Computer". In: *Journal of Chemical Theory and Computation* 19.8 (2023). PMID: 37001026, pp. 2230–2247. DOI: 10.1021/acs.jctc.3c00012. eprint: https://doi.org/10.1021/acs.jctc.3c00012. URL: https://doi.org/10.1021/acs.jctc.3c00012.

[Kit95] Alexei Y. Kitaev. "Quantum measurements and the Abelian Stabilizer Problem". In: *arXiv preprint, arXiv:quant-ph/9511026* (1995). arXiv: quant-ph/9511026 [quant-ph]. (Visited on).

[Gri+19] Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. "An adaptive variational algorithm for exact molecular simulations on a

quantum computer". In: *Nature Communications* 10.1 (2019), pp. 1–9. URL: https://www.nature.com/articles/s41467-019-10988-2.

[Col+18a] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi. "Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm". In: *Phys. Rev. X* 8 (1 Feb. 2018), p. 011021. DOI: 10.1103/PhysRevX.8.011021. URL: https://link.aps.org/doi/10.1103/PhysRevX.8.011021.

[Mot+20] Mario Motta, Chong Sun, Adrian TK Tan, Matthew J O'Rourke, Erika Ye, Austin J Minnich, Fernando GSL Brandao, and Garnet Kin-Lic Chan. "Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution". In: *Nature Physics* 16.2 (2020), pp. 205–210.

[Goi+22] Joshua J Goings, Alec White, Joonho Lee, Christofer S Tautermann, Matthias Degroote, Craig Gidney, Toru Shiozaki, Ryan Babbush, and Nicholas C Rubin. "Reliably assessing the electronic structure of cytochrome p450 on today's classical computers and tomorrow's quantum computers". In: *Proceedings of the National Academy of Sciences* 119.38 (2022), e2203533119.

[DJN15] Victorita Dolean, Pierre Jolivet, and Frédéric Nataf. *An introduction to domain decomposition methods: algorithms, theory, and parallel implementation*. SIAM, 2015.

[Ape+22] Apeldoorn van Joran, Arjan Cornelissen, Andras Gilyén, and Giacomo Nannicini. *Quantum tomography using state-preparation unitaries*. arXiv:2207.08800. 2022.

[Yue22] Henry Yuen. "An Improved Sample Complexity Lower Bound for Quantum State Tomography". In: *arXiv preprint arXiv:2206.11185* (2022).

[BCD05] Dave Bacon, Andrew M Childs, and Wim van Dam. "Optimal measurements for the dihedral hidden subgroup problem". In: *arXiv preprint quant-ph/0501044* (2005).

[HEB04] Marc Hein, Jens Eisert, and Hans J Briegel. "Multiparty entanglement in graph states". In: *Physical Review A* 69.6 (2004), p. 062311.

[Sch02] Dirk Schlingemann. "Stabilizer Codes Can Be Realized as Graph Codes". In: *Quantum Info. Comput.* 2.4 (June 2002), pp. 307–323. ISSN: 1533-7146.

[CGK17] Shawn X Cui, Daniel Gottesman, and Anirudh Krishna. "Diagonal gates in the Clifford hierarchy". In: *Physical Review A* 95.1 (2017), p. 012329.

[GC99] Daniel Gottesman and Isaac L Chuang. "Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations". In: *Nature* 402.6760 (1999), pp. 390–393.

[BG16] Sergey Bravyi and David Gosset. "Improved classical simulation of quantum circuits dominated by Clifford gates". In: *Physical Review Letters* 116.25 (2016), p. 250501.

[Mon12] Ashley Montanaro. "The quantum query complexity of learning multilinear polynomials". In: *Information Processing Letters* 112.11 (2012), pp. 438–442.

[BV97] Ethan Bernstein and Umesh Vazirani. "Quantum complexity theory". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1411–1473.

[Röt09] Martin Rötteler. "Quantum algorithms to solve the hidden shift problem for quadratics and for functions of large Gowers norm". In: *International Sym-*

posium on Mathematical Foundations of Computer Science. Springer. 2009, pp. 663–674.

[BMS16] Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. "Average-case complexity versus approximate simulation of commuting quantum computations". In: *Physical Review Letters* 117.8 (2016), p. 080501.

[NBG21] Leonardo Novo, Juani Bermejo-Vega, and Raúl Garcıa-Patrón. "Quantum advantage from energy measurements of many-body quantum systems". In: *Quantum* 5 (2021), p. 465.

[ASW15] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. "Reed-Muller Codes for Random Erasures and Errors". In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '15. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 297–306. ISBN: 9781450335362. DOI: 10.1145/2746539.2746575. URL: https://doi.org/10.1145/2746539.2746575.

[GNW21] David Gross, Sepehr Nezami, and Michael Walter. "Schur–Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations". In: *Communications in Mathematical Physics* 385.3 (2021), pp. 1325–1393.

[ASY20] Emmanuel Abbe, Amir Shpilka, and Min Ye. "Reed–Muller Codes: Theory and Algorithms". In: *IEEE Transactions on Information Theory* 67.6 (2020), pp. 3251–3277. DOI: 10.1109/TIT.2020.3004749. URL: https://doi.org/10.1109/TIT.2020.3004749.

[DM09] Stark C. Draper and Sheida Malekpour. "Compressed sensing over finite fields". In: *2009 IEEE International Symposium on Information Theory*. IEEE. 2009, pp. 669–673. DOI: 10.1109/ISIT.2009.5205666.

[Roc+19] Andrea Rocchetto, Scott Aaronson, Simone Severini, Gonzalo Carvacho, Davide Poderini, Iris Agresti, Marco Bentivegna, and Fabio Sciarrino. "Experimental learning of quantum states". In: *Science Advances* 5.3 (2019), eaau1946.

[Lyu05] Vadim Lyubashevsky. "The parity problem in the presence of noise, decoding random linear codes, and the subset sum problem". In: *Approximation, randomization and combinatorial optimization. Algorithms and techniques*. Springer, 2005, pp. 378–389.

[NS94] Noam Nisan and Mario Szegedy. "On the degree of Boolean functions as real polynomials". In: *Computational Complexity* 4.4 (1994), pp. 301–313.

[BK02a] Howard Barnum and Emanuel Knill. "Reversing quantum dynamics with near-optimal quantum and classical fidelity". In: *Journal of Mathematical Physics* 43.5 (2002), pp. 2097–2106.

[BC99] A. Bernasconi and B. Codenotti. "Spectral analysis of Boolean functions as a graph eigenvalue problem". In: *IEEE Transactions on Computers* 48.3 (1999), pp. 345–351. DOI: 10.1109/12.755000.

[EF01] Yonina C Eldar and G David Forney. "On quantum detection and the square-root measurement". In: *IEEE Transactions on Information Theory* 47.3 (2001), pp. 858–872.

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error correcting codes*. Vol. 16. Elsevier, 1977.

[Hol73] Alexander Semenovich Holevo. "Bounds for the quantity of information transmitted by a quantum communication channel". In: *Problemy Peredachi Informatsii* 9.3 (1973), pp. 3–11.

[Bra+19] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. "Simulation of quantum circuits by low-rank stabilizer decompositions". In: *Quantum* 3 (2019), p. 181.

[ODo14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. DOI: 10.1017/CBO9781139814782.

[Mon19] Ashley Montanaro. "Pretty simple bounds on quantum state discrimination". In: *arXiv preprint arXiv:1908.08312* (2019).

[DD03] Jeroen Dehaene and Bart De Moor. "Clifford group, stabilizer states, and linear and quadratic operations over GF(2)". In: *Physical Review A* 68.4 (2003), p. 042318.

[Mon07] Ashley Montanaro. "On the distinguishability of random quantum states". In: *Communications in Mathematical Physics* 273.3 (2007), pp. 619–636.

[ZCC08] Bei Zeng, Xie Chen, and Isaac L. Chuang. "Semi-Clifford operations, structure of $\mathcal{C}_k$ hierarchy, and gate complexity for fault-tolerant quantum computation". In: *Phys. Rev. A* 77 (4 Apr. 2008), p. 042313. DOI: 10.1103/PhysRevA.77.042313. URL: https://link.aps.org/doi/10.1103/PhysRevA.77.042313.

[Roc18] Andrea Rocchetto. "Stabiliser states are efficiently PAC-Learnable". In: *Quantum Inf. Comput.* 18.7&8 (2018), pp. 541–552.

[Lia23] Daniel Liang. "Clifford Circuits can be Properly PAC Learned if and only if $\mathsf{RP} = \mathsf{NP}$". In: *Quantum* 7 (2023), p. 1036.

[Ací01] A. Acín. "Statistical distinguishability between unitary operations". In: *Phys. Rev. Lett.* 87 (17 2001), p. 177901. DOI: 10.1103/PhysRevLett.87.177901. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.177901.

[DFY07] Runyao Duan, Yuan Feng, and Mingsheng Ying. "Entanglement is Not Necessary for Perfect Discrimination between Unitary Operations". In: *Phys. Rev. Lett.* 98 (10 Mar. 2007), p. 100503. DOI: 10.1103/PhysRevLett.98.100503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.98.100503.

[Cal+08] J. Calsamiglia, R. Muñoz-Tapia, Ll. Masanes, A. Acin, and E. Bagan. "Quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states". In: *Phys. Rev. A* 77 (3 Mar. 2008), p. 032311. DOI: 10.1103/PhysRevA.77.032311. URL: https://link.aps.org/doi/10.1103/PhysRevA.77.032311.

[ZP20] Q. Zhuang and S. Pirandola. "Ultimate Limits for Multiple Quantum Channel Discrimination". In: *Phys. Rev. Lett.* 125 (8 2020), p. 080505. DOI: 10.1103/PhysRevLett.125.080505. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.080505.

[Pir+19] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira. "Fundamental limits to quantum channel discrimination". In: *npj Quantum Inf.* 5 (2019), p. 50. DOI: https://doi.org/10.1038/s41534-019-0162-y.

[WW19] Xin Wang and Mark M. Wilde. "Resource theory of asymmetric distinguishability for quantum channels". In: *Phys. Rev. Res.* 1 (3 Dec. 2019), p. 033169.

DOI: 10.1103/PhysRevResearch.1.033169. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.1.033169.

[NC11] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th. USA: Cambridge University Press, 2011. ISBN: 1107002176.

[GLM04b] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum-Enhanced Measurements: Beating the Standard Quantum Limit". In: *Science* 306.5700 (2004), pp. 1330–1336. DOI: 10.1126/science.1104149. eprint: https://www.science.org/doi/pdf/10.1126/science.1104149. URL: https://www.science.org/doi/abs/10.1126/science.1104149.

[Bra+18] Daniel Braun, Gerardo Adesso, Fabio Benatti, Roberto Floreanini, Ugo Marzolino, Morgan W. Mitchell, and Stefano Pirandola. "Quantum-enhanced measurements without entanglement". In: *Rev. Mod. Phys.* 90 (3 Sept. 2018), p. 035006. DOI: 10.1103/RevModPhys.90.035006. URL: https://link.aps.org/doi/10.1103/RevModPhys.90.035006.

[DFY09] Runyao Duan, Yuan Feng, and Mingsheng Ying. "Perfect Distinguishability of Quantum Operations". In: *Phys. Rev. Lett.* 103 (21 Nov. 2009), p. 210501. DOI: 10.1103/PhysRevLett.103.210501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.210501.

[Pez+18a] Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. "Quantum metrology with nonclassical states of atomic ensembles". In: *Rev. Mod. Phys.* 90 (3 Sept. 2018), p. 035005. DOI: 10.1103/RevModPhys.90.035005. URL: https://link.aps.org/doi/10.1103/RevModPhys.90.035005.

[IHY85] N. Imoto, H. A. Haus, and Y. Yamamoto. "Quantum nondemolition measurement of the photon number via the optical Kerr effect". In: *Phys. Rev. A* 32 (4 Oct. 1985), pp. 2287–2292. DOI: 10.1103/PhysRevA.32.2287. URL: https://link.aps.org/doi/10.1103/PhysRevA.32.2287.

[GLP98] Philippe Grangier, Juan Ariel Levenson, and Jean-Philippe Poizat. "Quantum non-demolition measurements in optics". In: *Nature* 396.6711 (1998), pp. 537–542. DOI: 10.1038/25059. URL: https://doi.org/10.1038/25059.

[Sch+05] P. O. Schmidt, T. Rosenband, C. Langer, W. M. Itano, J. C. Bergquist, and D. J. Wineland. "Spectroscopy Using Quantum Logic". In: *Science* 309.5735 (2005), pp. 749–752. DOI: 10.1126/science.1114375. eprint: https://www.science.org/doi/pdf/10.1126/science.1114375. URL: https://www.science.org/doi/abs/10.1126/science.1114375.

[Xia+13] Ze-Liang Xiang, Sahel Ashhab, J. Q. You, and Franco Nori. "Hybrid quantum circuits: Superconducting circuits interacting with other quantum systems". In: *Rev. Mod. Phys.* 85 (2 Apr. 2013), pp. 623–653. DOI: 10.1103/RevModPhys.85.623. URL: https://link.aps.org/doi/10.1103/RevModPhys.85.623.

[Hel76] C. W. Helstrom. *Quantum Detection and Estimation Theory*. Mathematics in Science and Engineering: A Series of Monographs and Textbooks. Academic Press, Inc., 1976.

[Hig+07a] Brendon L Higgins, Dominic W Berry, Stephen D Bartlett, Howard M Wiseman, and Geoff J Pryde. "Entanglement-free Heisenberg-limited phase estima-

tion". In: *Nature* 450.7168 (2007), pp. 393–396. ISSN: 1476-4687. DOI: 10.1038/nature06257. URL: https://doi.org/10.1038/nature06257.

[Reh+18] Junaid ur Rehman, Ahmad Farooq, Youngmin Jeong, and Hyundong Shin. "Quantum channel discrimination without entanglement". In: *Quantum Information Processing* 17.10 (2018), p. 271. ISSN: 1573-1332. DOI: 10.1007/s11128-018-2037-0. URL: https://doi.org/10.1007/s11128-018-2037-0.

[RS21] Junaid ur Rehman and Hyundong Shin. "Entanglement-Free Parameter Estimation of Generalized Pauli Channels". In: *Quantum* 5 (July 2021), p. 490. ISSN: 2521-327X. DOI: 10.22331/q-2021-07-01-490. URL: https://doi.org/10.22331/q-2021-07-01-490.

[KK22] Stewart A. Koppell and Mark A. Kasevich. *Optimal Dose-Limited Phase Estimation without Entanglement*. 2022. arXiv: 2203.10137 [quant-ph].

[DeB+23] Kyle DeBry, Jasmine Sinanan-Singh, Colin D. Bruzewicz, David Reens, May E. Kim, Matthew P. Roychowdhury, Robert McConnell, Isaac L. Chuang, and John Chiaverini. "Experimental Quantum Channel Discrimination Using Metastable States of a Trapped Ion". In: *Phys. Rev. Lett.* 131 (17 Oct. 2023), p. 170602. DOI: 10.1103/PhysRevLett.131.170602. URL: https://link.aps.org/doi/10.1103/PhysRevLett.131.170602.

[LC17] Guang Hao Low and Isaac L. Chuang. "Optimal Hamiltonian Simulation by Quantum Signal Processing". In: *Phys. Rev. Lett.* 118 (1 Jan. 2017), p. 010501. DOI: 10.1103/PhysRevLett.118.010501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.010501.

[Yua16] Haidong Yuan. "Sequential Feedback Scheme Outperforms the Parallel Scheme for Hamiltonian Parameter Estimation". In: *Phys. Rev. Lett.* 117 (16 Oct. 2016), p. 160801. DOI: 10.1103/PhysRevLett.117.160801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.117.160801.

[BMQ21] Jessica Bavaresco, Mio Murao, and Marco Túlio Quintino. "Strict Hierarchy between Parallel, Sequential, and Indefinite-Causal-Order Strategies for Channel Discrimination". In: *Phys. Rev. Lett.* 127 (20 Nov. 2021), p. 200504. DOI: 10.1103/PhysRevLett.127.200504. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.200504.

[PW09a] Marco Piani and John Watrous. "All Entangled States are Useful for Channel Discrimination". In: *Phys. Rev. Lett.* 102 (25 June 2009), p. 250501. DOI: 10.1103/PhysRevLett.102.250501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.102.250501.

[BCP19] Joonwoo Bae, Dariusz Chru ści ński, and Marco Piani. "More Entanglement Implies Higher Performance in Channel Discrimination Tasks". In: *Phys. Rev. Lett.* 122 (14 Apr. 2019), p. 140404. DOI: 10.1103/PhysRevLett.122.140404. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.140404.

[CMW16] Tom Cooney, Milán Mosonyi, and Mark M Wilde. "Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication". In: *Communications in Mathematical Physics* 344.3 (2016), pp. 797–829. DOI: 10.1007/s00220-016-2645-4. URL: https://doi.org/10.1007/s00220-016-2645-4.

[SHW22] Farzin Salek, Masahito Hayashi, and Andreas Winter. "Usefulness of adaptive strategies in asymptotic quantum channel discrimination". In: *Phys. Rev. A* 105 (2 Feb. 2022), p. 022419. DOI: 10.1103/PhysRevA.105.022419. URL: https://link.aps.org/doi/10.1103/PhysRevA.105.022419.

[GLM11b] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Advances in quantum metrology". In: *Nature Photonics* 5.4 (Mar. 2011), pp. 222–229. ISSN: 1749-4893. DOI: 10.1038/nphoton.2011.35. URL: http://dx.doi.org/10.1038/nphoton.2011.35.

[RC21] Zane M. Rossi and Isaac L. Chuang. "Quantum hypothesis testing with group structure". In: *Phys. Rev. A* 104 (1 July 2021), p. 012425. DOI: 10.1103/PhysRevA.104.012425. URL: https://link.aps.org/doi/10.1103/PhysRevA.104.012425.

[Mar+21a] John M. Martyn, Zane M. Rossi, Andrew K. Tan, and Isaac L. Chuang. "Grand Unification of Quantum Algorithms". In: *PRX Quantum* 2 (4 Dec. 2021), p. 040203. DOI: 10.1103/PRXQuantum.2.040203. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.040203.

[Ros+22] Zane M. Rossi, Jeffery Yu, Isaac L. Chuang, and Sho Sugiura. "Quantum advantage for noisy channel discrimination". In: *Phys. Rev. A* 105 (3 Mar. 2022), p. 032401. DOI: 10.1103/PhysRevA.105.032401. URL: https://link.aps.org/doi/10.1103/PhysRevA.105.032401.

[Zho+18] Sisi Zhou, Mengzhen Zhang, John Preskill, and Liang Jiang. "Achieving the Heisenberg limit in quantum metrology using quantum error correction". In: *Nature Communications* 9.1 (Jan. 2018). ISSN: 2041-1723. DOI: 10.1038/s41467-017-02510-3. URL: http://dx.doi.org/10.1038/s41467-017-02510-3.

[Cov05] Thomas M Cover. *Elements of Information Theory*. John Wiley & Sons, Ltd, 2005. ISBN: 9780471748823. DOI: https://doi.org/10.1002/047174882X.fmatter. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/047174882X. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.

[Hel94] Carl W. Helstrom. *Elements of Signal Detection and Estimation*. USA: Prentice-Hall, Inc., 1994. ISBN: 013808940X.

[Med+19] Catherine Medlock, Alan Oppenheim, Isaac Chuang, and Qi Ding. "Operating characteristics for binary hypothesis testing in quantum systems". In: *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2019, pp. 1136–1145.

[Col+05] Jared H. Cole, Sonia G. Schirmer, Andrew D. Greentree, Cameron J. Wellard, Daniel K. L. Oi, and Lloyd C. L. Hollenberg. "Identifying an experimental two-state Hamiltonian to arbitrary accuracy". In: *Phys. Rev. A* 71 (6 June 2005), p. 062312. DOI: 10.1103/PhysRevA.71.062312. URL: https://link.aps.org/doi/10.1103/PhysRevA.71.062312.

[Col+06] Jared H. Cole, Andrew D. Greentree, Daniel K. L. Oi, Sonia G. Schirmer, Cameron J. Wellard, and Lloyd C. L. Hollenberg. "Identifying a two-state Hamiltonian in the presence of decoherence". In: *Phys. Rev. A* 73 (6 June 2006), p. 062333. DOI: 10.1103/PhysRevA.73.062333. URL: https://link.aps.org/doi/10.1103/PhysRevA.73.062333.

[BY12] Daniel Burgarth and Kazuya Yuasa. "Quantum System Identification". In: *Phys. Rev. Lett.* 108 (8 Feb. 2012), p. 080502. DOI: 10.1103/PhysRevLett.108.080502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.080502.

[Val+19] Agnes Valenti, Evert van Nieuwenburg, Sebastian Huber, and Eliska Greplova. "Hamiltonian learning for quantum error correction". In: *Phys. Rev. Research* 1 (3 Nov. 2019), p. 033092. DOI: 10.1103/PhysRevResearch.1.033092. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.1.033092.

[Wan+17] Jianwei Wang, Stefano Paesani, Raffaele Santagati, Sebastian Knauer, Antonio A Gentile, Nathan Wiebe, Maurangelo Petruzzella, Jeremy L O'Brien, John G Rarity, Anthony Laing, et al. "Experimental Quantum Hamiltonian Learning". In: *Nature Physics* 13.6 (2017), pp. 551–555.

[FG98] Edward Farhi and Sam Gutmann. "Analog analogue of a digital quantum computation". In: *Phys. Rev. A* 57 (4 Apr. 1998), pp. 2403–2406. DOI: 10.1103/PhysRevA.57.2403. URL: https://link.aps.org/doi/10.1103/PhysRevA.57.2403.

[DM15] Rafał Demkowicz-Dobrza ński and Marcin Markiewicz. "Quantum computation speedup limits from quantum metrological precision bounds". In: *Phys. Rev. A* 91 (6 June 2015), p. 062322. DOI: 10.1103/PhysRevA.91.062322. URL: https://link.aps.org/doi/10.1103/PhysRevA.91.062322.

[Liu+16] Nana Liu, Jayne Thompson, Christian Weedbrook, Seth Lloyd, Vlatko Vedral, Mile Gu, and Kavan Modi. "Power of one qumode for quantum computation". In: *Phys. Rev. A* 93 (5 May 2016), p. 052304. DOI: 10.1103/PhysRevA.93.052304. URL: https://link.aps.org/doi/10.1103/PhysRevA.93.052304.

[ZPK10] Marcin Zwierz, Carlos A. Pérez-Delgado, and Pieter Kok. "General optimality of the Heisenberg limit for quantum metrology". In: *Phys. Rev. Lett.* 105 (18 Oct. 2010), p. 180402. DOI: 10.1103/PhysRevLett.105.180402. URL: https://link.aps.org/doi/10.1103/PhysRevLett.105.180402.

[TA14] Géza Tóth and Iagoba Apellaniz. "Quantum metrology from a quantum information science perspective". In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (Oct. 2014), p. 424006. DOI: 10.1088/1751-8113/47/42/424006. URL: https://doi.org/10.1088/1751-8113/47/42/424006.

[Pez+18b] Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. "Quantum metrology with nonclassical states of atomic ensembles". In: *Rev. Mod. Phys.* 90 (3 Sept. 2018), p. 035005. DOI: 10.1103/RevModPhys.90.035005. URL: https://link.aps.org/doi/10.1103/RevModPhys.90.035005.

[IF07] Hiroshi Imai and Akio Fujiwara. "Geometry of optimal estimation scheme for SU(D) channels". In: *Journal of Physics A: Mathematical and Theoretical* 40.16 (Mar. 2007), pp. 4391–4400. DOI: 10.1088/1751-8113/40/16/009. URL: https://doi.org/10.1088/1751-8113/40/16/009.

[Ber+15] Dominic W. Berry, Mankei Tsang, Michael J. W. Hall, and Howard M. Wiseman. "Quantum Bell-Ziv-Zakai bounds and Heisenberg limits for waveform estimation". In: *Phys. Rev. X* 5 (3 Aug. 2015), p. 031018. DOI: 10.1103/PhysRevX.5.031018. URL: https://link.aps.org/doi/10.1103/PhysRevX.5.031018.

[SBD16] Magdalena Szczykulska, Tillmann Baumgratz, and Animesh Datta. "Multi-parameter quantum metrology". In: *Advances in Physics: X* 1.4 (2016), pp. 621–

639. DOI: 10.1080/23746149.2016.1230476. eprint: https://doi.org/10.1080/23746149.2016.1230476. URL: https://doi.org/10.1080/23746149.2016.1230476.

[Gór+20] Wojciech Górecki, Rafał Demkowicz-Dobrza ński, Howard M. Wiseman, and Dominic W. Berry. "π-Corrected Heisenberg Limit". In: *Phys. Rev. Lett.* 124 (3 Jan. 2020), p. 030501. DOI: 10.1103/PhysRevLett.124.030501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.030501.

[GLM06] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum Metrology". In: *Phys. Rev. Lett.* 96 (1 Jan. 2006), p. 010401. DOI: 10.1103/PhysRevLett.96.010401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.010401.

[Bol+96] J. J. Bollinger, Wayne M. Itano, D. J. Wineland, and D. J. Heinzen. "Optimal frequency measurements with maximally correlated states". In: *Phys. Rev. A* 54 (6 Dec. 1996), R4649–R4652. DOI: 10.1103/PhysRevA.54.R4649. URL: https://link.aps.org/doi/10.1103/PhysRevA.54.R4649.

[LKD02] Hwang Lee, Pieter Kok, and Jonathan P Dowling. "A quantum Rosetta stone for interferometry". In: *Journal of Modern Optics* 49.14-15 (2002), pp. 2325–2338. DOI: 10.1080/09500340210000011536. eprint: https://doi.org/10.1080/09500340210000011536. URL: https://doi.org/10.1080/09500340210000011536.

[Hig+07b] Brendon L Higgins, Dominic W Berry, Stephen D Bartlett, Howard M Wiseman, and Geoff J Pryde. "Entanglement-free Heisenberg-limited phase estimation". In: *Nature* 450.7168 (2007), pp. 393–396.

[Wis+09] Howard M Wiseman, Dominic W Berry, Stephen D Bartlett, Brendon L Higgins, and Geoffrey J Pryde. "Adaptive measurements in the optical quantum information laboratory". In: *IEEE Journal of Selected Topics in Quantum Electronics* 15.6 (2009), pp. 1661–1672.

[Hig+09] B L Higgins, D W Berry, S D Bartlett, M W Mitchell, H M Wiseman, and G J Pryde. "Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements". In: *New Journal of Physics* 11.7 (July 2009), p. 073023. DOI: 10.1088/1367-2630/11/7/073023. URL: https://doi.org/10.1088/1367-2630/11/7/073023.

[KLY15] Shelby Kimmel, Guang Hao Low, and Theodore J. Yoder. "Robust calibration of a universal single-qubit gate set via robust phase estimation". In: *Phys. Rev. A* 92 (6 Dec. 2015), p. 062315. DOI: 10.1103/PhysRevA.92.062315. URL: https://link.aps.org/doi/10.1103/PhysRevA.92.062315.

[Kra+19] Stefan Krastanov, Sisi Zhou, Steven T Flammia, and Liang Jiang. "Stochastic estimation of dynamical variables". In: *Quantum Science and Technology* 4.3 (2019), p. 035003.

[KWR04] Robert Kosut, Ian A Walmsley, and Herschel Rabitz. "Optimal experiment design for quantum state and process tomography and Hamiltonian parameter estimation". In: *arXiv preprint quant-ph/0411093* (2004).

[GR02] J. M. Geremia and Herschel Rabitz. "Optimal identification of Hamiltonian information by closed-loop laser control of quantum systems". In: *Phys. Rev. Lett.* 89 (26 Dec. 2002), p. 263902. DOI: 10.1103/PhysRevLett.89.263902. URL: https://link.aps.org/doi/10.1103/PhysRevLett.89.263902.

[Nun+10] J. Nunn, B. J. Smith, G. Puentes, I. A. Walmsley, and J. S. Lundeen. "Optimal experiment design for quantum state tomography: Fair, precise, and minimal

tomography". In: *Phys. Rev. A* 81 (4 Apr. 2010), p. 042109. DOI: 10.1103/PhysRevA.81.042109. URL: https://link.aps.org/doi/10.1103/PhysRevA.81.042109.

[GNS19] Yonatan Gazit, Hui Khoon Ng, and Jun Suzuki. "Quantum process tomography via optimal design of experiments". In: *Phys. Rev. A* 100 (1 July 2019), p. 012350. DOI: 10.1103/PhysRevA.100.012350. URL: https://link.aps.org/doi/10.1103/PhysRevA.100.012350.

[YF15] Haidong Yuan and Chi-Hang Fred Fung. "Optimal feedback scheme and universal time scaling for Hamiltonian parameter estimation". In: *Phys. Rev. Lett.* 115 (11 Sept. 2015), p. 110401. DOI: 10.1103/PhysRevLett.115.110401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.110401.

[KU18] Naoto Kura and Masahito Ueda. "Finite-error metrological bounds on multiparameter Hamiltonian estimation". In: *Phys. Rev. A* 97 (1 Jan. 2018), p. 012101. DOI: 10.1103/PhysRevA.97.012101. URL: https://link.aps.org/doi/10.1103/PhysRevA.97.012101.

[Sou+17] Jamshid Sourati, Murat Akcakaya, Todd K. Leen, Deniz Erdogmus, and Jennifer G. Dy. "Asymptotic Analysis of Objectives Based on Fisher Information in Active Learning". In: *J. Mach. Learn. Res.* 18.1 (Jan. 2017), pp. 1123–1163. ISSN: 1532-4435.

[Jaz07] Andrew H Jazwinski. *Stochastic processes and filtering theory*. New York, USA: Courier Corporation, 2007.

[AS72] D. Alspach and H. Sorenson. "Nonlinear Bayesian estimation using Gaussian sum approximations". In: *IEEE Transactions on Automatic Control* 17.4 (1972), pp. 439–448. DOI: 10.1109/TAC.1972.1100034.

[DJ+09] Arnaud Doucet, Adam M Johansen, et al. "A tutorial on particle filtering and smoothing: Fifteen years later". In: *Handbook of Nonlinear Filtering* 12.656-704 (2009), p. 3.

[LG94] David D Lewis and William A Gale. "A sequential algorithm for training text classifiers". In: *SIGIR'94*. Springer. 1994, pp. 3–12.

[Set09] Burr Settles. *Active learning literature survey*. Tech. rep. 1648. University of Wisconsin-Madison, Department of Computer Sciences, Jan. 2009.

[SOS92] H. S. Seung, M. Opper, and H. Sompolinsky. "Query by Committee". In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. COLT '92. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 1992, pp. 287–294. ISBN: 089791497X. DOI: 10.1145/130385.130417. URL: https://doi.org/10.1145/130385.130417.

[BRK07] Robert Burbidge, Jem J. Rowland, and Ross D. King. "Active Learning for Regression Based on Query by Committee". In: *Proceedings of the 8th International Conference on Intelligent Data Engineering and Automated Learning*. IDEAL'07. Birmingham, UK: Springer-Verlag, 2007, pp. 209–218. ISBN: 3540772251.

[BBZ07] Maria-Florina Balcan, Andrei Broder, and Tong Zhang. "Margin Based Active Learning". In: *Proceedings of the 20th Annual Conference on Learning Theory*. COLT'07. San Diego, CA, USA: Springer-Verlag, 2007, pp. 35–50. ISBN: 9783540729259.

[YBT06] Kai Yu, Jinbo Bi, and Volker Tresp. "Active Learning via Transductive Experimental Design". In: *Proceedings of the 23rd International Conference on Machine Learning*. ICML '06. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 2006, pp. 1081–1088. ISBN: 1595933832. DOI: 10.1145/1143844.1143980. URL: https://doi.org/10.1145/1143844.1143980.

[FAJ06] Patrick Flaherty, Adam Arkin, and Michael Jordan. "Robust design of biological experiments". In: *Advances in Neural Information Processing Systems*. Ed. by Y. Weiss, B. Schölkopf, and J. Platt. Vol. 18. MIT Press, 2006, pp. 363–370. URL: https://proceedings.neurips.cc/paper/2005/file/3953630da28e5181cffca1278517e3cf-Paper.pdf.

[Du+15] Bo Du, Zengmao Wang, Lefei Zhang, Liangpei Zhang, Wei Liu, Jialie Shen, and Dacheng Tao. "Exploring representativeness and informativeness for active learning". In: *IEEE Transactions on Cybernetics* 47.1 (2015), pp. 14–26.

[HJZ14] Sheng-Jun Huang, Rong Jin, and Zhi-Hua Zhou. "Active learning by querying informative and representative examples". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.10 (2014), pp. 1936–1949.

[Set12] Burr Settles. "Active Learning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6.1 (2012), pp. 1–114.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. USA: Wiley-Interscience, 2006. ISBN: 0471241954.

[LC06] Erich L Lehmann and George Casella. *Theory of point estimation*. New York, USA: Springer Science & Business Media, 2006.

[Was13] Larry Wasserman. *All of statistics: a concise course in statistical inference*. New York, USA: Springer Science & Business Media, 2013.

[FKF00] Dietmar G. Fischer, Stefan H. Kienle, and Matthias Freyberger. "Quantum-state estimation by self-learning measurements". In: *Phys. Rev. A* 61 (3 Feb. 2000), p. 032306. DOI: 10.1103/PhysRevA.61.032306. URL: https://link.aps.org/doi/10.1103/PhysRevA.61.032306.

[HH12] F. Huszár and N. M. T. Houlsby. "Adaptive Bayesian quantum tomography". In: *Phys. Rev. A* 85 (5 May 2012), p. 052120. DOI: 10.1103/PhysRevA.85.052120. URL: https://link.aps.org/doi/10.1103/PhysRevA.85.052120.

[FGC13] Christopher Ferrie, Christopher E Granade, and David G Cory. "How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies". In: *Quantum Information Processing* 12.1 (2013), pp. 611–623.

[Hoi+06] Steven C. H. Hoi, Rong Jin, Jianke Zhu, and Michael R. Lyu. "Batch Mode Active Learning and Its Application to Medical Image Classification". In: *Proceedings of the 23rd International Conference on Machine Learning*. ICML '06. Pittsburgh, Pennsylvania, USA: ACM, 2006, pp. 417–424. ISBN: 1595933832. DOI: 10.1145/1143844.1143897. URL: https://doi.org/10.1145/1143844.1143897.

[Cha+15] Kamalika Chaudhuri, Sham M. Kakade, Praneeth Netrapalli, and Sujay Sanghavi. "Convergence Rates of Active Learning for Maximum Likelihood Estimation". In: *Advances in Neural Information Processing Systems*. NIPS'15. Montreal, Canada: ACM, Dec. 2015, pp. 1090–1098.

[Has+12a] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. "Nearly Optimal Sparse Fourier Transform". In: *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*. STOC '12. New York, USA: Association for Computing Machinery, 2012, pp. 563–578. ISBN: 9781450312455. DOI: 10.1145/2213977.2214029. URL: https://doi.org/10.1145/2213977.2214029.

[Has+12b] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. "Simple and Practical Algorithm for Sparse Fourier Transform". In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '12. Kyoto, Japan: SIAM, 2012, pp. 1183–1194.

[Zah+18] Tom Zahavy, Matan Haroush, Nadav Merlis, Daniel J. Mankowitz, and Shie Mannor. "Learn What Not to Learn: Action Elimination with Deep Reinforcement Learning". In: *Advances in Neural Information Processing Systems*. NIPS'18. Montréal, Canada: Curran Associates Inc., 2018, pp. 3566–3577.

[MRK16] Warwick Masson, Pravesh Ranchod, and George Konidaris. "Reinforcement Learning with Parameterized Actions". In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI'16. Phoenix, Arizona: AAAI Press, 2016, pp. 1934–1940.

[Dul+15] Gabriel Dulac-Arnold, Richard Evans, Hado van Hasselt, Peter Sunehag, Timothy Lillicrap, Jonathan Hunt, Timothy Mann, Theophane Weber, Thomas Degris, and Ben Coppin. "Deep reinforcement learning in large discrete action spaces". In: *arXiv preprint arXiv:1512.07679* (2015).

[Hin+18] Ian Hincks, Thomas Alexander, Michal Kononenko, Benjamin Soloway, and David G Cory. "Hamiltonian learning with online bayesian experiment design in practice". In: *arXiv preprint arXiv:1806.02427* (2018).

[SB18] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*. USA: MIT Press, 2018.

[WIB15] Kai Wei, Rishabh Iyer, and Jeff Bilmes. "Submodularity in Data Subset Selection and Active Learning". In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, July 2015, pp. 1954–1963. URL: https://proceedings.mlr.press/v37/wei15.html.

[RD10] Chad Rigetti and Michel Devoret. "Fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies". In: *Phys. Rev. B* 81 (13 Apr. 2010), p. 134507. DOI: 10.1103/PhysRevB.81.134507. URL: https://link.aps.org/doi/10.1103/PhysRevB.81.134507.

[Cho+11] Jerry M. Chow et al. "Simple All-Microwave Entangling Gate for Fixed-Frequency Superconducting Qubits". In: *Phys. Rev. Lett.* 107 (8 Aug. 2011), p. 080502. DOI: 10.1103/PhysRevLett.107.080502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.107.080502.

[Zha+03] Jun Zhang, Jiri Vala, Shankar Sastry, and K. Birgitta Whaley. "Geometric theory of nonlocal two-qubit operations". In: *Phys. Rev. A* 67 (4 Apr. 2003), p. 042313. DOI: 10.1103/PhysRevA.67.042313. URL: https://link.aps.org/doi/10.1103/PhysRevA.67.042313.

[MG20] Easwar Magesan and Jay M. Gambetta. "Effective Hamiltonian models of the cross-resonance gate". In: *Phys. Rev. A* 101 (5 May 2020), p. 052308. DOI:

10.1103/PhysRevA.101.052308. URL: https://link.aps.org/doi/10.1103/PhysRevA.101.052308.

[Cho+14] Jerry M Chow, Jay M Gambetta, Easwar Magesan, David W Abraham, Andrew W Cross, BR Johnson, Nicholas A Masluk, Colm A Ryan, John A Smolin, Srikanth J Srinivasan, et al. "Implementing a strand of a scalable fault-tolerant quantum computing fabric". In: *Nature Communications* 5.1 (2014), pp. 1–9. URL: https://doi.org/10.1038/ncomms5015.

[Ale+20] Thomas Alexander, Naoki Kanazawa, Daniel J Egger, Lauren Capelluto, Christopher J Wood, Ali Javadi-Abhari, and David C McKay. "Qiskit Pulse: Programming Quantum Computers Through the Cloud with Pulses". In: *Quantum Science and Technology* 5.4 (Aug. 2020), p. 044006. DOI: 10.1088/2058-9565/aba404. URL: https://doi.org/10.1088/2058-9565/aba404.

[Abr+19a] Héctor Abraham, Ismail Yunus Akhalwaya, Gadi Aleksandrowicz, Thomas Alexander, Eli Arbel, Abraham Asfaw, Carlos Azaustre, Panagiotis Barkoutsos, George Barron, Luciano Bello, et al. *Qiskit: An Open-source Framework for Quantum Computing*. Version 0.7.2. Jan. 2019. DOI: 10.5281/zenodo.2562111. URL: https://doi.org/10.5281/zenodo.2562111.

[Sun+20a] Neereja Sundaresan, Isaac Lauer, Emily Pritchett, Easwar Magesan, Petar Jurcevic, and Jay M. Gambetta. "Reducing Unitary and Spectator Errors in Cross Resonance with Optimized Rotary Echoes". In: *PRX Quantum* 1 (2 Dec. 2020), p. 020318. DOI: 10.1103/PRXQuantum.1.020318. URL: https://link.aps.org/doi/10.1103/PRXQuantum.1.020318.

[Wal+15] Joel Wallman, Chris Granade, Robin Harper, and Steven T Flammia. "Estimating the coherence of noise". In: *New Journal of Physics* 17.11 (Nov. 2015), p. 113020. DOI: 10.1088/1367-2630/17/11/113020. URL: https://doi.org/10.1088/1367-2630/17/11/113020.

[KB14] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[Zhu+97] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization". In: *ACM Trans. Math. Softw.* 23.4 (Dec. 1997), pp. 550–560. ISSN: 0098-3500. DOI: 10.1145/279232.279236. URL: https://doi.org/10.1145/279232.279236.

[LW01] Jane Liu and Mike West. "Combined parameter and state estimation in simulation-based filtering". In: *Sequential Monte Carlo methods in practice*. New York, USA: Springer-Verlag, 2001, pp. 197–223.

[Coh89] Leon Cohen. "Time-Frequency Distributions – A review". In: *Proceedings of the IEEE* 77.7 (1989), pp. 941–981.

[PBV15] Reza Parhizkar, Yann Barbotin, and Martin Vetterli. "Sequences with minimal time–frequency uncertainty". In: *Applied and Computational Harmonic Analysis* 38.3 (2015), pp. 452–468.

[Abr+19b] Deanna M. Abrams, Nicolas Didier, Shane A. Caldwell, Blake R. Johnson, and Colm A. Ryan. "Methods for Measuring Magnetic Flux Crosstalk between Tunable Transmons". In: *Phys. Rev. Applied* 12 (6 Dec. 2019), p. 064022. DOI:

10.1103/PhysRevApplied.12.064022. URL: https://link.aps.org/doi/10.1103/PhysRevApplied.12.064022.

[Dai+21] X. Dai et al. "Calibration of Flux Crosstalk in Large-Scale Flux-Tunable Superconducting Quantum Circuits". In: *PRX Quantum* 2 (4 Oct. 2021), p. 040313. DOI: 10.1103/PRXQuantum.2.040313. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.040313.

[Pan+18] Kunkun Pang, Mingzhi Dong, Yang Wu, and Timothy Hospedales. "Meta-learning transferable active learning policies by deep reinforcement learning". In: *AutoML Workshop at ICML* (2018).

[Dut23a] Arkopal Dutt. *HAL*. Version 0.1.0-alpha. Apr. 2023. DOI: 10.5281/zenodo.7859448. URL: https://doi.org/10.5281/zenodo.7859448.

[Dut23b] Arkopal Dutt. *BayesianHL*. Version 0.1.0-alpha. Apr. 2023. DOI: 10.5281/zenodo.7859452. URL: https://doi.org/10.5281/zenodo.7859452.

[HJO00] Trygve Helgaker, Poul Jorgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, Ltd, 2000. ISBN: 9781119019572. DOI: https://doi.org/10.1002/9781119019572.

[HK64] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Phys. Rev.* 136 (3B Nov. 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864. URL: https://link.aps.org/doi/10.1103/PhysRev.136.B864.

[KS65] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Phys. Rev.* 140 (4A Nov. 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133. URL: https://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[HLR94] B. L. Hammond, William A. Lester, and Peter J. Reynolds. *Monte Carlo methods in ab initio quantum chemistry*. Singapore: World Scientific Singapore, 1994. ISBN: 9789814317245. DOI: https://doi.org/10.1142/1170.

[CH02] Garnet Kin-Lic Chan and Martin Head-Gordon. "Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group". In: *The Journal of Chemical Physics* 116.11 (Mar. 2002), pp. 4462–4476. ISSN: 0021-9606. DOI: 10.1063/1.1449459. eprint: https://pubs.aip.org/aip/jcp/article-pdf/116/11/4462/10838240/4462_1_online.pdf. URL: https://doi.org/10.1063/1.1449459.

[Aus+] Brian Austin et al. *NERSC-10 Workload Analysis*. URL: https://portal.nersc.gov/project/m888/nersc10/workload/N10_Workload_Analysis.latest.pdf.

[BR09] Lorenz C. Blum and Jean-Louis Reymond. "970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13". In: *Journal of the American Chemical Society* 131.25 (2009). PMID: 19505099, pp. 8732–8733. DOI: 10.1021/ja902302h. eprint: https://doi.org/10.1021/ja902302h. URL: https://doi.org/10.1021/ja902302h.

[Rup+12] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning". In: *Phys. Rev. Lett.* 108 (5 Jan. 2012), p. 058301. DOI: 10.1103/PhysRevLett.108.058301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.058301.

[Rud+12] Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond. "Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17". In: *Journal of Chemical Information and Modeling* 52.11 (2012). PMID: 23088335, pp. 2864–2875. DOI: 10.1021/ci300415d. eprint: https://doi.org/10.1021/ci300415d. URL: https://doi.org/10.1021/ci300415d.

[Ram+14] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. "Quantum chemistry structures and properties of 134 kilo molecules". In: *Scientific data* 1.1 (2014), pp. 1–7.

[KDM11] Amir Karton, Shauli Daon, and Jan M.L. Martin. "W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data". In: *Chemical Physics Letters* 510.4 (2011), pp. 165–178. ISSN: 0009-2614. DOI: https://doi.org/10.1016/j.cplett.2011.05.007. URL: https://www.sciencedirect.com/science/article/pii/S0009261411005616.

[KTM07] Amir Karton, Peter R. Taylor, and Jan M. L. Martin. "Basis set convergence of post-CCSD contributions to molecular atomization energies". In: *The Journal of Chemical Physics* 127.6 (Aug. 2007), p. 064104. ISSN: 0021-9606. DOI: 10.1063/1.2755751. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.2755751/14757432/064104_1_online.pdf. URL: https://doi.org/10.1063/1.2755751.

[MBS11] Michael S. Marshall, Lori A. Burns, and C. David Sherrill. "Basis set convergence of the coupled-cluster correction, δ MP2 CCSD (T): Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases". In: *The Journal of Chemical Physics* 135.19 (Nov. 2011), p. 194102. ISSN: 0021-9606. DOI: 10.1063/1.3659142. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.3659142/15444624/194102_1_online.pdf. URL: https://doi.org/10.1063/1.3659142.

[Fri+17] Nils-Ole Friedrich, Christina de Bruyn Kops, Florian Flachsenberg, Kai Sommer, Matthias Rarey, and Johannes Kirchmair. "Benchmarking Commercial Conformer Ensemble Generators". In: *Journal of Chemical Information and Modeling* 57.11 (2017). PMID: 28967749, pp. 2719–2728. DOI: 10.1021/acs.jcim.7b00505. eprint: https://doi.org/10.1021/acs.jcim.7b00505. URL: https://doi.org/10.1021/acs.jcim.7b00505.

[MH19] Satoshi Maeda and Yu Harabuchi. "On Benchmarking of Automated Methods for Performing Exhaustive Reaction Path Search". In: *Journal of Chemical Theory and Computation* 15.4 (2019). PMID: 30860828, pp. 2111–2115. DOI: 10.1021/acs.jctc.8b01182. eprint: https://doi.org/10.1021/acs.jctc.8b01182. URL: https://doi.org/10.1021/acs.jctc.8b01182.

[Don+79] Jack J Dongarra, Cleve Barry Moler, James R Bunch, and Gilbert W Stewart. *LINPACK users' guide*. SIAM, 1979.

[Don87] Jack J Dongarra. "The LINPACK benchmark: An explanation". In: *International Conference on Supercomputing*. Springer. 1987, pp. 456–474.

[Den12] Li Deng. "The MNIST Database of Handwritten Digit Images for Machine Learning Research". In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 141–142. DOI: 10.1109/MSP.2012.2211477.

[KH09] A. Krizhevsky and G. Hinton. *Learning multiple layers of features from tiny images*. Apr. 2009. URL: https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[Lin+14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. "Microsoft COCO: Common Objects in Context". In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer. Springer International Publishing, 2014, pp. 740–755. ISBN: 978-3-319-10602-1.

[Kha+22] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. "Transformers in Vision: A Survey". In: *ACM Comput. Surv.* 54.10s (Sept. 2022). ISSN: 0360-0300. DOI: 10.1145/3505244. URL: https://doi.org/10.1145/3505244.

[Min+22] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. "Image Segmentation Using Deep Learning: A Survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7 (2022), pp. 3523–3542. DOI: 10.1109/TPAMI.2021.3059968.

[FGG14] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A quantum approximate optimization algorithm". In: *arXiv preprint arXiv:1411.4028* (2014).

[Mol+18] Nikolaj Moll et al. "Quantum optimization using variational algorithms on near-term quantum devices". In: *Quantum Science and Technology* 3.3 (June 2018), p. 030503. DOI: 10.1088/2058-9565/aab822. URL: https://dx.doi.org/10.1088/2058-9565/aab822.

[Far+22] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou. "The quantum approximate optimization algorithm and the sherrington-kirkpatrick model at infinite size". In: *Quantum* 6 (2022), p. 759.

[AL99a] Daniel S. Abrams and Seth Lloyd. "Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors". In: *Phys. Rev. Lett.* 83 (24 Dec. 1999), pp. 5162–5165. DOI: 10.1103/PhysRevLett.83.5162. URL: https://link.aps.org/doi/10.1103/PhysRevLett.83.5162.

[PW09b] David Poulin and Pawel Wocjan. "Preparing Ground States of Quantum Many-Body Systems on a Quantum Computer". In: *Phys. Rev. Lett.* 102 (13 Apr. 2009), p. 130503. DOI: 10.1103/PhysRevLett.102.130503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.102.130503.

[GTC19] Yimin Ge, Jordi Tura, and J. Ignacio Cirac. "Faster ground state preparation and high-precision ground energy estimation with fewer qubits". In: *Journal of Mathematical Physics* 60.2 (Feb. 2019), p. 022202. ISSN: 0022-2488. DOI: 10.1063/1.5027484. eprint: https://pubs.aip.org/aip/jmp/article-pdf/

doi/10.1063/1.5027484/13434463/022202_1_online.pdf. URL: https://doi.org/10.1063/1.5027484.

[CW12] Andrew M. Childs and Nathan Wiebe. "Hamiltonian Simulation Using Linear Combinations of Unitary Operations". In: *Quantum Info. Comput.* 12.11–12 (Nov. 2012), pp. 901–924. ISSN: 1533-7146.

[Gil+19] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. "Quantum Singular Value Transformation and beyond: Exponential Improvements for Quantum Matrix Arithmetics". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2019. Phoenix, AZ, USA: Association for Computing Machinery, 2019, pp. 193–204. ISBN: 9781450367059. DOI: 10.1145/3313276.3316366. URL: https://doi.org/10.1145/3313276.3316366.

[LT20] Lin Lin and Yu Tong. "Near-optimal ground state preparation". In: *Quantum* 4 (2020), p. 372.

[LT22] Lin Lin and Yu Tong. "Heisenberg-Limited Ground-State Energy Estimation for Early Fault-Tolerant Quantum Computers". In: *PRX Quantum* 3 (1 Feb. 2022), p. 010318. DOI: 10.1103/PRXQuantum.3.010318. URL: https://link.aps.org/doi/10.1103/PRXQuantum.3.010318.

[DL23] Zhiyan Ding and Lin Lin. "Even Shorter Quantum Circuit for Phase Estimation on Early Fault-Tolerant Quantum Computers with Applications to Ground-State Energy Estimation". In: *PRX Quantum* 4 (2 May 2023), p. 020331. DOI: 10.1103/PRXQuantum.4.020331. URL: https://link.aps.org/doi/10.1103/PRXQuantum.4.020331.

[Du+10] Jiangfeng Du, Nanyang Xu, Xinhua Peng, Pengfei Wang, Sanfeng Wu, and Dawei Lu. "NMR Implementation of a Molecular Hydrogen Quantum Simulation with Adiabatic State Preparation". In: *Phys. Rev. Lett.* 104 (3 2010), p. 030502. DOI: 10.1103/PhysRevLett.104.030502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.030502.

[Dum+18] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski. "Cloud Quantum Computing of an Atomic Nucleus". In: *Phys. Rev. Lett.* 120 (2018), p. 210501.

[Kok+19] C. Kokail et al. "Self-verifying variational quantum simulation of lattice models". In: *Nature* 569.7756 (2019), pp. 355–360. DOI: 10.1038/s41586-019-1177-4. URL: https://doi.org/10.1038/s41586-019-1177-4.

[Kan+19] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. "Error mitigation extends the computational reach of a noisy quantum processor". In: *Nature* 567.7749 (Mar. 1, 2019), pp. 491–495. DOI: 10.1038/s41586-019-1040-7. URL: https://doi.org/10.1038/s41586-019-1040-7.

[Gan+19] Marc Ganzhorn, Daniel J Egger, Panagiotis Barkoutsos, Pauline Ollitrault, Gian Salis, Nikolaj Moll, Marco Roth, A Fuhrer, P Mueller, Stefan Woerner, et al. "Gate-efficient simulation of molecular eigenstates on a quantum computer". In: *Physical Review Applied* 11.4 (2019), p. 044092.

[Sag+19] Ramiro Sagastizabal, Xavier Bonet-Monroig, Malay Singh, M Adriaan Rol, CC Bultink, Xiang Fu, CH Price, VP Ostroukh, N Muthusubramanian, A Bruno,

et al. “Experimental error mitigation via symmetry verification in a variational quantum eigensolver”. In: *Physical Review A* 100.1 (2019), p. 010302.

[SM19] Scott E Smart and David A Mazziotti. “Quantum-classical hybrid algorithm using an error-mitigating N-representability condition to compute the Mott metal-insulator transition”. In: *Physical Review A* 100.2 (2019), p. 022517.

[Nam+20] Yunseong Nam et al. “Ground-state energy estimation of the water molecule on a trapped-ion quantum computer”. In: *npj Quantum Information* 6.1 (Apr. 3, 2020), p. 33. DOI: 10.1038/s41534-020-0259-3. URL: https://doi.org/10.1038/s41534-020-0259-3.

[Aru+20b] Frank Arute et al. “Hartree-Fock on a superconducting qubit quantum computer”. In: *Science* 369.6507 (2020), pp. 1084–1089. DOI: 10.1126/science.abb9811. eprint: https://www.science.org/doi/pdf/10.1126/science.abb9811. URL: https://www.science.org/doi/abs/10.1126/science.abb9811.

[Kre+21] Michael Kreshchuk, Shaoyang Jia, William M. Kirby, Gary Goldstein, James P. Vary, and Peter J. Love. “Light-Front Field Theory on Current Quantum Computers”. In: *Entropy* 23.5 (2021). ISSN: 1099-4300. DOI: 10.3390/e23050597. URL: https://www.mdpi.com/1099-4300/23/5/597.

[Lan+10] Benjamin P Lanyon et al. “Towards quantum chemistry on a quantum computer”. In: *Nature Chemistry* 2.2 (2010), pp. 106–111.

[Löt+21] Erik Lötstedt, Kaoru Yamanouchi, Takashi Tsuchiya, and Yutaka Tachikawa. “Calculation of vibrational eigenenergies on a quantum computer: Application to the Fermi resonance in CO2”. In: *Physical Review A* 103.6 (2021), p. 062609.

[Kis+22] Oriel Kiss, Michele Grossi, Pavel Lougovski, Federico Sanchez, Sofia Vallecorsa, and Thomas Papenbrock. “Quantum computing of the ^{6}Li nucleus via ordered unitary coupled clusters”. In: *Phys. Rev. C* 106 (3 Sept. 2022), p. 034325. DOI: 10.1103/PhysRevC.106.034325. URL: https://link.aps.org/doi/10.1103/PhysRevC.106.034325.

[Wan+15] Ya Wang et al. “Quantum simulation of helium hydride cation in a solid-state spin register”. In: *ACS Nano* 9.8 (2015), pp. 7769–7774.

[OMa+16] P. J. J. O’Malley et al. “Scalable Quantum Simulation of Molecular Energies”. In: *Phys. Rev. X* 6 (3 July 2016), p. 031007. DOI: 10.1103/PhysRevX.6.031007. URL: https://link.aps.org/doi/10.1103/PhysRevX.6.031007.

[She+17] Yangchao Shen, Xiang Zhang, Shuaining Zhang, Jing-Ning Zhang, Man-Hong Yung, and Kihwan Kim. “Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure”. In: *Physical Review A* 95.2 (2017), p. 020501.

[Pae+17] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O’Brien, and M. G. Thompson. “Experimental Bayesian Quantum Phase Estimation on a Silicon Photonic Chip”. In: *Phys. Rev. Lett.* 118 (10 2017), p. 100503. DOI: 10.1103/PhysRevLett.118.100503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.100503.

[Hem+18] Cornelius Hempel et al. “Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator”. In: *Phys. Rev. X* 8 (3 July 2018), p. 031022. DOI: 10.1103/PhysRevX.8.031022. URL: https://link.aps.org/doi/10.1103/PhysRevX.8.031022.

[San+18] Raffaele Santagati et al. "Witnessing eigenstates for quantum simulation of Hamiltonian spectra". In: *Science Advances* 4.1 (2018). DOI: 10.1126/sciadv.aap9646. eprint: https://advances.sciencemag.org/content/4/1/eaap9646.full.pdf. URL: https://advances.sciencemag.org/content/4/1/eaap9646.

[Col+18b] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi. "Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm". In: *Phys. Rev. X* 8 (1 Feb. 2018), p. 011021. DOI: 10.1103/PhysRevX.8.011021. URL: https://link.aps.org/doi/10.1103/PhysRevX.8.011021.

[Kir+22] William Kirby, Bryce Fuller, Charles Hadfield, and Antonio Mezzacapo. "Second-Quantized Fermionic Operators with Polylogarithmic Qubit and Gate Complexity". In: *PRX Quantum* 3 (2 June 2022), p. 020351. DOI: 10.1103/PRXQuantum.3.020351. URL: https://link.aps.org/doi/10.1103/PRXQuantum.3.020351.

[AT03] Dorit Aharonov and Amnon Ta-Shma. "Adiabatic Quantum State Generation and Statistical Zero Knowledge". In: *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*. STOC '03. San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 20–29. ISBN: 1581136749. URL: https://doi.org/10.1145/780542.780546.

[Ber+07] Dominic W. Berry, Graeme Ahokas, Richard Cleve, and Barry C. Sanders. "Efficient Quantum Algorithms for Simulating Sparse Hamiltonians". In: *Communications in Mathematical Physics* 270.2 (Mar. 1, 2007), pp. 359–371. URL: https://doi.org/10.1007/s00220-006-0150-x.

[CK11] Andrew M. Childs and Robin Kothari. "Simulating Sparse Hamiltonians with Star Decompositions". In: *Theory of Quantum Computation, Communication, and Cryptography*. Ed. by Wim van Dam, Vivien M. Kendon, and Simone Severini. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 94–103. ISBN: 978-3-642-18073-6. URL: https://doi.org/10.1007/978-3-642-18073-6_8.

[LLB21] Joseph M Lukens, Kody JH Law, and Ryan S Bennink. "A Bayesian analysis of classical shadows". In: *npj Quantum Information* 7.1 (2021), pp. 1–10.

[Had21] Charles Hadfield. "Adaptive Pauli Shadows for Energy Estimation". In: *arXiv preprint arXiv:2105.12207* (2021).

[Gok+20] Pranav Gokhale, Olivia Angiuli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, Margaret Martonosi, and Frederic T. Chong. "$O(N^3)$ Measurement Cost for Variational Quantum Eigensolver on Molecular Hamiltonians". In: *IEEE Transactions on Quantum Engineering* 1 (2020), pp. 1–24. DOI: 10.1109/TQE.2020.3035814.

[Cra+21] Ophelia Crawford, Barnaby van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley. "Efficient quantum measurement of Pauli operators in the presence of finite sampling error". In: *Quantum* 5 (2021), p. 385.

[YGI23] Tzu-Ching Yen, Aadithya Ganeshram, and Artur F. Izmaylov. "Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates". In: *npj Quantum Information* 9.1 (Feb. 22, 2023), p. 14. DOI: 10.1038/s41534-023-00683-y. URL: https://doi.org/10.1038/s41534-023-00683-y.

[ASS21] Atithi Acharya, Siddhartha Saha, and Anirvan M. Sengupta. "Shadow tomography based on informationally complete positive operator-valued measure". In: *Phys. Rev. A* 104 (5 Nov. 2021), p. 052418. DOI: 10.1103/PhysRevA.104.052418. URL: https://link.aps.org/doi/10.1103/PhysRevA.104.052418.

[HKP21] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Efficient Estimation of Pauli Observables by Derandomization". In: *Phys. Rev. Lett.* 127 (3 July 2021), p. 030503. DOI: 10.1103/PhysRevLett.127.030503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.030503.

[Str+21] GI Struchalin, Ya A Zagorovskii, EV Kovlakov, SS Straupe, and SP Kulik. "Experimental estimation of quantum state properties from classical shadows". In: *PRX Quantum* 2.1 (2021), p. 010307.

[Zha+21] Ting Zhang, Jinzhao Sun, Xiao-Xu Fang, Xiaoming Zhang, Xiao Yuan, and He Lu. "Experimental quantum state measurement with classical shadows". In: *arXiv preprint arXiv:2106.10190* (2021).

[DL21] Yulong Dong and Lin Lin. "Random circuit block-encoded matrix and a proposal of quantum LINPACK benchmark". In: *Phys. Rev. A* 103 (6 June 2021), p. 062412. DOI: 10.1103/PhysRevA.103.062412. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.062412.

[RC99] Christian P Robert and George Casella. *Monte Carlo statistical methods*. Vol. 2. Springer, 1999.

[MRS08] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*. Cambridge University Press, 2008. ISBN: 978-0-521-86571-5. DOI: 10.1017/CBO9780511809071. URL: https://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf.

[Cas+96] George Casella, Juan Ferrándiz, Daniel Peña, David Rios Insua, José M Bernardo, PA Garcıa-López, A González, J Berger, AP Dawid, Thomas J Diciccio, et al. "Statistical inference and Monte Carlo algorithms". In: *Test* 5 (1996), pp. 249–344.

[Sär13] Simo Särkkä. *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139344203.

[Mat+24] Ryosuke Matsuo, Shigeru Yamashita, Shin-ichi Minato, and Rudy Raymond. "Optimizing Decision Diagrams for Measurements of Quantum Circuits". In: *Proceedings of the 29th Asia and South Pacific Design Automation Conference. To appear*. ASPDAC '24. Incheon, Korea, 2024.

[JW93] Pascual Jordan and Eugene Paul Wigner. "über das paulische äquivalenzverbot". In: *The Collected Works of Eugene Paul Wigner*. Springer, 1993, pp. 109–129.

[BK02b] Sergey B Bravyi and Alexei Yu Kitaev. "Fermionic quantum computation". In: *Annals of Physics* 298.1 (2002), pp. 210–226.

[SRL12] Jacob T Seeley, Martin J Richard, and Peter J Love. "The Bravyi-Kitaev transformation for quantum computation of electronic structure". In: *The Journal of chemical physics* 137.22 (2012), p. 224109.

[Bra+17] Sergey Bravyi, Jay M. Gambetta, Antonio Mezzacapo, and Kristan Temme. "Tapering off qubits to simulate fermionic Hamiltonians". In: *arXiv preprint,*

arXiv:1701.08213 (2017). arXiv: 1701.08213 [quant-ph]. URL: https://arxiv.org/abs/1701.08213 (visited on).

[Set+20] Kanav Setia, Richard Chen, Julia E. Rice, Antonio Mezzacapo, Marco Pistoia, and James D. Whitfield. "Reducing Qubit Requirements for Quantum Simulations Using Molecular Point Group Symmetries". In: *Journal of Chemical Theory and Computation* 16.10 (Oct. 13, 2020), pp. 6091–6097. DOI: 10.1021/acs.jctc.0c00113. URL: https://doi.org/10.1021/acs.jctc.0c00113.

[DLP03] Jack J. Dongarra, Piotr Luszczek, and Antoine Petitet. "The LINPACK Benchmark: past, present and future". In: *Concurrency and Computation: Practice and Experience* 15.9 (2003), pp. 803–820. DOI: https://doi.org/10.1002/cpe.728. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.728. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.728.

[Jo+15] Gangwon Jo, Jeongho Nah, Jun Lee, Jungwon Kim, and Jaejin Lee. "Accelerating LINPACK with MPI-OpenCL on Clusters of Multi-GPU Nodes". In: *IEEE Transactions on Parallel and Distributed Systems* 26.7 (2015), pp. 1814–1825. DOI: 10.1109/TPDS.2014.2321742.

[Pet+18] A. Petitet, R. C. Whaley, J. Dongarra, and A. Cleary. *HPL - A Portable Implementation of the High-Performance LINPACK Benchmark for Distributed-Memory Computers*. https://www.netlib.org/benchmark/hpl/. 2018.

[MT20] Per-Gunnar Martinsson and Joel A Tropp. "Randomized numerical linear algebra: Foundations and algorithms". In: *Acta Numerica* 29 (2020), pp. 403–572.

[Bar+18] Panagiotis Kl Barkoutsos, Jerome F Gonthier, Igor Sokolov, Nikolaj Moll, Gian Salis, Andreas Fuhrer, Marc Ganzhorn, Daniel J Egger, Matthias Troyer, Antonio Mezzacapo, et al. "Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions". In: *Physical Review A* 98.2 (2018), p. 022322.

[Ipp23] Matteo Ippoliti. "Classical shadows based on locally-entangled measurements". In: *arXiv preprint arXiv:2305.10723* (2023).

[GK23] Alexander Gresch and Martin Kliesch. "Guaranteed efficient energy estimation of quantum many-body Hamiltonians using ShadowGrouping". In: *arXiv preprint arXiv:2301.03385* (2023).

[FYS52] Kenichi Fukui, Teijiro Yonezawa, and Haruo Shingu. "A molecular orbital theory of reactivity in aromatic hydrocarbons". In: *The Journal of Chemical Physics* 20.4 (1952), pp. 722–725. URL: https://aip.scitation.org/doi/abs/10.1063/1.1700523.

[PY84] Robert G Parr and Weitao Yang. "Density functional approach to the frontier-electron theory of chemical reactivity". In: *Journal of the American Chemical Society* 106.14 (1984), pp. 4049–4050. URL: https://pubs.acs.org/doi/pdf/10.1021/ja00326a036.

[GNM02] Jeff Greeley, Jens K Nørskov, and Manos Mavrikakis. "Electronic structure and catalysis on metal surfaces". In: *Annual Review of Physical Chemistry* 53.1 (2002), pp. 319–348. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.53.100301.131630.

[LeB+15] J. P. F. LeBlanc et al. "Solutions of the two-dimensional Hubbard model: benchmarks and results from a wide range of numerical algorithms". In: *Physical Review X* 5.4 (2015), p. 041041. URL: https://journals.aps.org/prx/abstract/10.1103/PhysRevX.5.041041.

[ZW15] Huihuo Zheng and Lucas K Wagner. "Computation of the correlated metal-insulator transition in vanadium dioxide from first principles". In: *Physical Review Letters* 114.17 (2015), p. 176401. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.114.176401.

[Kot+06] Gabriel Kotliar, Sergej Y Savrasov, Kristjan Haule, Viktor S Oudovenko, O Parcollet, and CA Marianetti. "Electronic structure calculations with dynamical mean-field theory". In: *Reviews of Modern Physics* 78.3 (2006), p. 865. URL: https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.78.865.

[Gor+12] Mark S Gordon, Dmitri G Fedorov, Spencer R Pruitt, and Lyudmila V Slipchenko. "Fragmentation methods: A route to accurate calculations on large systems". In: *Chemical Reviews* 112.1 (2012), pp. 632–672. URL: https://pubs.acs.org/doi/full/10.1021/cr200093j.

[Jon+20] Leighton O Jones, Martın A Mosquera, George C Schatz, and Mark A Ratner. "Embedding methods for quantum chemistry: applications from materials to life sciences". In: *Journal of the American Chemical Society* 142.7 (2020), pp. 3281–3295. URL: https://pubs.acs.org/doi/abs/10.1021/jacs.9b10780.

[SC16] Qiming Sun and Garnet Kin-Lic Chan. "Quantum embedding theories". In: *Accounts of Chemical Research* 49.12 (2016), pp. 2705–2712. URL: https://doi.org/10.1021/acs.accounts.6b00356.

[WSZ15] Tomasz A Wesolowski, Sapana Shedge, and Xiuwen Zhou. "Frozen-density embedding strategy for multilevel simulations of electronic structure". In: *Chemical Reviews* 115.12 (2015), pp. 5891–5928. URL: https://doi.org/10.1021/cr500502v.

[LHC14] Florian Libisch, Chen Huang, and Emily A Carter. "Embedded correlated wavefunction schemes: Theory and applications". In: *Accounts of Chemical Research* 47.9 (2014), pp. 2768–2775. URL: https://pubs.acs.org/doi/abs/10.1021/ar500086h.

[KC12] Gerald Knizia and Garnet Kin-Lic Chan. "Density matrix embedding: A simple alternative to dynamical mean-field theory". In: *Physical Review Letters* 109.18 (2012), p. 186404. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.109.186404.

[KC13] Gerald Knizia and Garnet Kin-Lic Chan. "Density matrix embedding: A strong-coupling quantum embedding theory". In: *Journal of Chemical Theory and Computation* 9.3 (2013), pp. 1428–1432. URL: https://pubs.acs.org/doi/10.1021/ct301044e.

[Wou+16] Sebastian Wouters, Carlos A Jiménez-Hoyos, Qiming Sun, and Garnet K-L Chan. "A practical guide to density matrix embedding theory in quantum chemistry". In: *Journal of Chemical Theory and Computation* 12.6 (2016), pp. 2706–2719. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.6b00316.

[WAK17] Sebastian Wouters, Carlos A. Jiménez-Hoyos, and Garnet KL Chan. "Five years of density matrix embedding theory". In: *Fragmentation: toward accurate cal-*

culations on complex molecular systems (2017), pp. 227–243. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119129271.ch8.

[Fau+22] Fabian M Faulstich, Raehyun Kim, Zhi-Hao Cui, Zaiwen Wen, Garnet Kin-Lic Chan, and Lin Lin. "Pure State v-Representability of Density Matrix Embedding Theory". In: *Journal of Chemical Theory and Computation* 18.2 (2022), pp. 851–864. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.1c01061.

[Het+00] MH Hettler, M Mukherjee, M Jarrell, and HR Krishnamurthy. "Dynamical cluster approximation: Nonlocal dynamics of correlated electron systems". In: *Physical Review B* 61.19 (2000), p. 12739. URL: https://journals.aps.org/prb/abstract/10.1103/PhysRevB.61.12739.

[Ma+21] He Ma, Nan Sheng, Marco Govoni, and Giulia Galli. "Quantum embedding theory for strongly correlated states in materials". In: *Journal of Chemical Theory and Computation* 17.4 (2021), pp. 2116–2125. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.0c01258.

[LZ17] Tran Nguyen Lan and Dominika Zgid. "Generalized self-energy embedding theory". In: *The Journal of Physical Chemistry Letters* 8.10 (2017), pp. 2200–2205. URL: https://pubs.acs.org/doi/abs/10.1021/acs.jpclett.7b00689.

[Rus+18] Alexander A Rusakov, Sergei Iskakov, Lan Nguyen Tran, and Dominika Zgid. "Self-energy embedding theory (SEET) for periodic systems". In: *Journal of Chemical Theory and Computation* 15.1 (2018), pp. 229–240. URL: https://pubs.acs.org/doi/abs/10.1021/acs.jctc.8b00927.

[BAG03] S Biermann, F Aryasetiawan, and A Georges. "First-Principles Approach to the Electronic Structure of Strongly Correlated Systems: Combining the G W Approximation and Dynamical Mean-Field Theory". In: *Physical Review Letters* 90.8 (2003), p. 086402. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.90.086402.

[WTV16] Matthew Welborn, Takashi Tsuchimochi, and Troy Van Voorhis. "Bootstrap embedding: An internally consistent fragment-based method". In: *The Journal of Chemical Physics* 145.7 (2016), p. 074102. URL: https://aip.scitation.org/doi/10.1063/1.4960986.

[Ye+19] Hong-Zhou Ye, Nathan D Ricke, Henry K Tran, and Troy Van Voorhis. "Bootstrap embedding for molecules". In: *Journal of Chemical Theory and Computation* 15.8 (2019), pp. 4497–4506. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.9b00529.

[YTV21] Hong-Zhou Ye, Henry K Tran, and Troy Van Voorhis. "Accurate Electronic Excitation Energies in Full-Valence Active Space via Bootstrap Embedding". In: *Journal of Chemical Theory and Computation* 17.6 (2021), pp. 3335–3347. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.0c01221.

[Zhe+17] Bo-Xiao Zheng, Chia-Min Chung, Philippe Corboz, Georg Ehlers, Ming-Pu Qin, Reinhard M Noack, Hao Shi, Steven R White, Shiwei Zhang, and Garnet Kin-Lic Chan. "Stripe order in the underdoped region of the two-dimensional Hubbard model". In: *Science* 358.6367 (2017), pp. 1155–1160. URL: https://www.science.org/doi/10.1126/science.aam7127.

[Zhu+19] Tianyu Zhu, Carlos A Jiménez-Hoyos, James McClain, Timothy C Berkelbach, and Garnet Kin-Lic Chan. "Coupled-cluster impurity solvers for dynamical

mean-field theory". In: *Physical Review B* 100.11 (2019), p. 115154. URL: https://journals.aps.org/prb/abstract/10.1103/PhysRevB.100.115154.

[SZ19] Avijit Shee and Dominika Zgid. "Coupled cluster as an impurity solver for Green's function embedding methods". In: *Journal of Chemical Theory and Computation* 15.11 (2019), pp. 6010–6024. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.9b00603.

[LKB21] Bryan TG Lau, Gerald Knizia, and Timothy C Berkelbach. "Regional embedding enables high-level quantum chemistry for surface science". In: *The Journal of Physical Chemistry Letters* 12.3 (2021), pp. 1104–1109. URL: https://pubs.acs.org/doi/abs/10.1021/acs.jpclett.0c03274.

[Bau+20] Bela Bauer, Sergey Bravyi, Mario Motta, and Garnet Kin-Lic Chan. "Quantum algorithms for quantum chemistry and quantum materials science". In: *Chemical Reviews* 120.22 (2020), pp. 12685–12717. URL: https://pubs.acs.org/doi/10.1021/acs.chemrev.9b00829.

[Lee+22] Seunghoon Lee et al. "Is there evidence for exponential quantum advantage in quantum chemistry?" In: *arXiv preprint arXiv:2208.02199* (2022). URL: https://arxiv.org/abs/2208.02199.

[AL99b] Daniel S Abrams and Seth Lloyd. "Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors". In: *Physical Review Letters* 83.24 (1999), p. 5162. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.83.5162.

[Asp+05] Alán Aspuru-Guzik, Anthony D Dutoi, Peter J Love, and Martin Head-Gordon. "Simulated quantum computation of molecular energies". In: *Science* 309.5741 (2005), pp. 1704–1707. URL: https://www.science.org/doi/10.1126/science.1113479.

[Til+22] Jules Tilly et al. "The variational quantum eigensolver: a review of methods and best practices". In: *Physics Reports* 986 (2022), pp. 1–128. URL: https://www.sciencedirect.com/science/article/pii/S0370157322003118.

[WHB19] Daochen Wang, Oscar Higgott, and Stephen Brierley. "Accelerated variational quantum eigensolver". In: *Physical Review Letters* 122.14 (2019), p. 140504. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.122.140504.

[Gri+22] Harper R Grimsley, George S Barron, Edwin Barnes, Sophia E Economou, and Nicholas J Mayhall. "ADAPT-VQE is insensitive to rough parameter landscapes and barren plateaus". In: *arXiv preprint arXiv:2204.07179* (2022). URL: https://arxiv.org/abs/2204.07179.

[ZRM21] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. "Fermionic Partial Tomography via Classical Shadows". In: *Phys. Rev. Lett.* 127 (11 Sept. 2021), p. 110504. DOI: 10.1103/PhysRevLett.127.110504. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.110504.

[Ott+22] Matthew Otten, Matthew R. Hermes, Riddhish Pandharkar, Yuri Alexeev, Stephen K. Gray, and Laura Gagliardi. "Localized Quantum Chemistry on Quantum Computers". In: *Journal of Chemical Theory and Computation* 18.12 (Dec. 2022), pp. 7205–7217. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.2c00388. URL: https://doi.org/10.1021/acs.jctc.2c00388.

[Vor+22] Christian Vorwerk, Nan Sheng, Marco Govoni, Benchen Huang, and Giulia Galli. "Quantum embedding theories to simulate condensed systems on quantum computers". In: *Nature Computational Science* 2.7 (2022), pp. 424–432. URL: https://www.nature.com/articles/s43588-022-00279-0.

[MM22] Lana Mineh and Ashley Montanaro. "Solving the Hubbard model using density matrix embedding theory and the variational quantum eigensolver". In: *Phys. Rev. B* 105 (12 Mar. 2022), p. 125117. DOI: 10.1103/PhysRevB.105.125117. URL: https://link.aps.org/doi/10.1103/PhysRevB.105.125117.

[Li+22] Weitang Li, Zigeng Huang, Changsu Cao, Yifei Huang, Zhigang Shuai, Xiaoming Sun, Jinzhao Sun, Xiao Yuan, and Dingshun Lv. "Toward practical quantum embedding simulation of realistic chemical systems on near-term quantum computers". In: *Chemical Science* 13.31 (2022), pp. 8953–8962. URL: https://doi.org/10.1039/D2SC01492K.

[MGG20] He Ma, Marco Govoni, and Giulia Galli. "Quantum simulations of materials on near-term quantum computers". In: *npj Computational Materials* 6.1 (2020), pp. 1–8. URL: https://www.nature.com/articles/s41524-020-00353-z.

[Qua+20] Google AI Quantum et al. "Hartree-Fock on a superconducting qubit quantum computer". In: *Science* 369.6507 (2020), pp. 1084–1089.

[Bar+97] Adriano Barenco, André Berthiaume, David Deutsch, Artur Ekert, Richard Jozsa, and Chiara Macchiavello. "Stabilization of Quantum Computations by Symmetrization". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1541–1557. DOI: 10.1137/S0097539796302452. eprint: https://doi.org/10.1137/S0097539796302452. URL: https://doi.org/10.1137/S0097539796302452.

[Buh+01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. "Quantum Fingerprinting". In: *Phys. Rev. Lett.* 87 (16 Sept. 2001), p. 167902. DOI: 10.1103/PhysRevLett.87.167902. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.167902.

[Bra+02] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. "Quantum amplitude amplification and estimation". In: *Contemporary Mathematics* 305 (2002), pp. 53–74. URL: https://doi.org/10.1090/conm/305/05215.

[Mar+21b] John M Martyn, Zane M Rossi, Andrew K Tan, and Isaac L Chuang. "Grand unification of quantum algorithms". In: *PRX Quantum* 2.4 (2021), p. 040203. URL: https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.2.040203.

[YTV20] Hong-Zhou Ye, Henry K Tran, and Troy Van Voorhis. "Bootstrap embedding for large molecular systems". In: *Journal of Chemical Theory and Computation* 16.8 (2020), pp. 5035–5046. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.0c00438.

[Löw50] Per-Olov Löwdin. "On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals". In: *The Journal of Chemical Physics* 18.3 (1950), pp. 365–375. URL: https://aip.scitation.org/doi/10.1063/1.1747632.

[CM19] Daniel Claudino and Nicholas J Mayhall. "Automatic partition of orbital spaces based on singular value decomposition in the context of embedding theories".

In: *Journal of Chemical Theory and Computation* 15.2 (2019), pp. 1053–1064. URL: https://pubs.acs.org/doi/abs/10.1021/acs.jctc.8b01112.

[WWG21] Jonathan M Waldrop, Theresa L Windus, and Niranjan Govind. "Projector-based quantum embedding for molecular systems: An investigation of three partitioning approaches". In: *The Journal of Physical Chemistry A* 125.29 (2021), pp. 6384–6393. URL: https://doi.org/10.1021/acs.jpca.1c03821.

[YV19] Hong-Zhou Ye and Troy Van Voorhis. "Atom-based bootstrap embedding for molecules". In: *The journal of physical chemistry letters* 10.20 (2019), pp. 6368–6374. URL: https://pubs.acs.org/doi/abs/10.1021/acs.jpclett.9b02479.

[Kni13] Gerald Knizia. "Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts". In: *Journal of chemical theory and computation* 9.11 (2013), pp. 4834–4843. URL: https://pubs.acs.org/doi/abs/10.1021/ct400687b.

[NB22] Max Nusspickel and George H Booth. "Systematic improvability in quantum embedding for real materials". In: *Physical Review X* 12.1 (2022), p. 011046. URL: https://journals.aps.org/prx/abstract/10.1103/PhysRevX.12.011046.

[PA18] Francesco Parisen Toldin and Fakher F. Assaad. "Entanglement Hamiltonian of Interacting Fermionic Models". In: *Phys. Rev. Lett.* 121 (20 Nov. 2018), p. 200602. DOI: 10.1103/PhysRevLett.121.200602. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.200602.

[Vog+17] Konstantinos D Vogiatzis, Dongxia Ma, Jeppe Olsen, Laura Gagliardi, and Wibe A De Jong. "Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations". In: *The Journal of Chemical Physics* 147.18 (2017), p. 184111. URL: https://aip.scitation.org/doi/10.1063/1.4989858.

[BM07] Rodney J. Bartlett and Monika Musiał. "Coupled-cluster theory in quantum chemistry". In: *Review of Modern Physics* 79 (1 Feb. 2007), pp. 291–352. DOI: 10.1103/RevModPhys.79.291. URL: https://link.aps.org/doi/10.1103/RevModPhys.79.291.

[Mor+21] Miguel A Morales-Silva, Kenneth D Jordan, Luke Shulenburger, and Lucas K Wagner. "Frontiers of stochastic electronic structure calculations". In: *The Journal of Chemical Physics* 154.17 (2021), p. 170401. URL: https://aip.scitation.org/doi/abs/10.1063/5.0053674.

[LPR22] Joonho Lee, Hung Q. Pham, and David R. Reichman. "Twenty Years of Auxiliary-Field Quantum Monte Carlo in Quantum Chemistry: An Overview and Assessment on Main Group Chemistry and Bond-Breaking". In: *Journal of Chemical Theory and Computation* (Oct. 2022). Publisher: American Chemical Society. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.2c00802. URL: https://doi.org/10.1021/acs.jctc.2c00802.

[She+19] James Shee, Benjamin Rudshteyn, Evan J Arthur, Shiwei Zhang, David R Reichman, and Richard A Friesner. "On achieving high accuracy in quantum chemical calculations of 3 d transition metal-containing systems: a comparison of auxiliary-field quantum monte carlo with coupled cluster, density functional theory, and experiment for diatomic molecules". In: *Journal of Chemical Theory*

and Computation 15.4 (2019), pp. 2346–2358. URL: https://doi.org/10.1021/acs.jctc.9b00083.

[LCR18] Yuan Liu, Minsik Cho, and Brenda Rubenstein. "Ab initio finite temperature auxiliary field quantum Monte Carlo". In: *Journal of Chemical Theory and Computation* 14.9 (2018), pp. 4722–4732. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.8b00569.

[NU98] M Peter Nightingale and Cyrus J Umrigar. *Quantum Monte Carlo Methods in Physics and Chemistry*. 525. Springer Science & Business Media, 1998. URL: https://link.springer.com/book/9780792355519.

[Fou+01] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. "Quantum Monte Carlo simulations of solids". In: *Review of Modern Physics* 73 (1 Jan. 2001), pp. 33–83. DOI: 10.1103/RevModPhys.73.33. URL: https://link.aps.org/doi/10.1103/RevModPhys.73.33.

[WM07] Lucas K Wagner and Lubos Mitas. "Energetics and dipole moment of transition metal monoxides by quantum Monte Carlo". In: *The Journal of chemical physics* 126.3 (2007), p. 034105. URL: https://aip.scitation.org/doi/abs/10.1063/1.2428294.

[Whe+21] William A. Wheeler et al. *PyQMC: an all-Python real-space quantum Monte Carlo code, v0.5.1*. 2021. URL: https://github.com/WagnerGroup/pyqmc.

[Tra+18] Andrew Tranter, Peter J Love, Florian Mintert, and Peter V Coveney. "A comparison of the bravyi–kitaev and jordan–wigner transformations for the quantum simulation of quantum chemistry". In: *Journal of Chemical Theory and Computation* 14.11 (2018), pp. 5617–5630. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.8b00450.

[ER63] Clyde Edmiston and Klaus Ruedenberg. "Localized Atomic and Molecular Orbitals". In: *Review of Modern Physics* 35 (3 July 1963), pp. 457–464. DOI: 10.1103/RevModPhys.35.457. URL: https://link.aps.org/doi/10.1103/RevModPhys.35.457.

[Wan62] Gregory H. Wannier. "Dynamics of Band Electrons in Electric and Magnetic Fields". In: *Review of Modern Physics* 34 (4 Oct. 1962), pp. 645–655. DOI: 10.1103/RevModPhys.34.645. URL: https://link.aps.org/doi/10.1103/RevModPhys.34.645.

[Maz12] David A Mazziotti. "Two-electron reduced density matrix as the basic variable in many-electron quantum chemistry and physics". In: *Chemical Reviews* 112.1 (2012), pp. 244–262. URL: https://pubs.acs.org/doi/10.1021/cr2000493.

[BK08] Reinhold A Bertlmann and Philipp Krammer. "Bloch vectors for qudits". In: *Journal of Physics A: Mathematical and Theoretical* 41.23 (May 2008), p. 235303. DOI: 10.1088/1751-8113/41/23/235303. URL: https://dx.doi.org/10.1088/1751-8113/41/23/235303.

[CSV09] Andrew R Conn, Katya Scheinberg, and Luis N Vicente. *Introduction to Derivative-Free Optimization*. SIAM, 2009. URL: https://epubs.siam.org/doi/book/10.1137/1.9780898718768.

[Fan+20] M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti. "Beyond the Swap Test: Optimal Estimation of Quantum State Overlap". In:

Phys. Rev. Lett. 124 (6 Feb. 2020), p. 060503. DOI: 10.1103/PhysRevLett.124.060503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.060503.

[HM13] Aram W Harrow and Ashley Montanaro. "Testing product states, quantum Merlin-Arthur games and tensor optimization". In: *Journal of the ACM (JACM)* 60.1 (2013), pp. 1–43. URL: https://dl.acm.org/doi/abs/10.1145/2432622.2432625.

[MF67] Olvi L Mangasarian and Stan Fromovitz. "The Fritz John necessary optimality conditions in the presence of equality and inequality constraints". In: *Journal of Mathematical Analysis and Applications* 17.1 (1967), pp. 37–47. URL: https://www.sciencedirect.com/science/article/pii/0022247X67901631.

[Ber16] D. Bertsekas. *Nonlinear Programming*. Athena scientific optimization and computation series. Athena Scientific, 2016, pp. 317–330. ISBN: 9781886529052. URL: https://books.google.com/books?id=TwOujgEACAAJ.

[Sta+10] A Staszczak, Mario Stoitsov, A Baran, and Witold Nazarewicz. "Augmented Lagrangian method for constrained nuclear density functional theory". In: *The European Physical Journal A* 46.1 (2010), pp. 85–90. URL: https://link.springer.com/article/10.1140/epja/i2010-11018-9.

[KN21] Kohdai Kuroiwa and Yuya O Nakagawa. "Penalty methods for a variational quantum eigensolver". In: *Physical Review Research* 3.1 (2021), p. 013197. URL: https://journals.aps.org/prresearch/abs/10.1103/PhysRevResearch.3.013197.

[Liu+22] Yuan Liu, Arkopal Dutt, Max Tao, and Zachary E. Chin. *Quantum Bootstrap Embedding*. https://github.com/yuanliu1/QBootstrapEmbedding. 2022.

[Byr+95] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. "A Limited Memory Algorithm for Bound Constrained Optimization". In: *SIAM Journal on Scientific Computing* 16.5 (1995), pp. 1190–1208. DOI: 10.1137/0916069. eprint: https://doi.org/10.1137/0916069. URL: https://doi.org/10.1137/0916069.

[NM65] J. A. Nelder and R. Mead. "A Simplex Method for Function Minimization". In: *The Computer Journal* 7.4 (Jan. 1965), pp. 308–313. ISSN: 0010-4620. DOI: 10.1093/comjnl/7.4.308. eprint: https://academic.oup.com/comjnl/article-pdf/7/4/308/1013182/7-4-308.pdf. URL: https://doi.org/10.1093/comjnl/7.4.308.

[SHF13] Krysta M. Svore, Matthew B. Hastings, and Michael Freedman. "Faster Phase Estimation". In: *Quantum Information and Computation* (2013). URL: https://arxiv.org/abs/1304.0741.

[Hug+22] William J Huggins, Bryan A O'Gorman, Nicholas C Rubin, David R Reichman, Ryan Babbush, and Joonho Lee. "Unbiasing fermionic quantum Monte Carlo with a quantum computer". In: *Nature* 603.7901 (2022), pp. 416–420. URL: https://www.nature.com/articles/s41586-021-04351-z.

[WA08] Pawel Wocjan and Anura Abeyesinghe. "Speedup via quantum sampling". In: *Phys. Rev. A* 78 (4 Oct. 2008), p. 042336. DOI: 10.1103/PhysRevA.78.042336. URL: https://link.aps.org/doi/10.1103/PhysRevA.78.042336.

[YA12] Man-Hong Yung and Alán Aspuru-Guzik. "A quantum–quantum Metropolis algorithm". In: *Proceedings of the National Academy of Sciences* 109.3 (2012), pp. 754–759. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1111758109.

[Mon15] Ashley Montanaro. "Quantum speedup of Monte Carlo methods". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471.2181 (2015), p. 20150301. URL: https://doi.org/10.1098/rspa.2015.0301.

[YLC14] Theodore J. Yoder, Guang Hao Low, and Isaac L. Chuang. "Fixed-Point Quantum Search with an Optimal Number of Queries". In: *Phys. Rev. Lett.* 113 (21 Nov. 2014), p. 210501. DOI: 10.1103/PhysRevLett.113.210501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.210501.

[QR21] Xiao-Liang Qi and Daniel Ranard. "Emergent classicality in general multipartite states and channels". In: *Quantum* 5 (2021), p. 555. URL: https://quantum-journal.org/papers/q-2021-09-28-555/.

[NIB22] Max Nusspickel, Basil Ibrahim, and George H Booth. "On the effective reconstruction of expectation values from ab initio quantum embedding". In: *arXiv preprint arXiv:2210.14561* (2022). URL: https://arxiv.org/abs/2210.14561.

[Mit+21] Abhishek Mitra, Hung Q Pham, Riddhish Pandharkar, Matthew R Hermes, and Laura Gagliardi. "Excited states of crystalline point defects with multireference density matrix embedding theory". In: *The Journal of Physical Chemistry Letters* 12.48 (2021), pp. 11688–11694. URL: https://pubs.acs.org/doi/10.1021/acs.jpclett.1c03229.

[Zha99] Shiwei Zhang. "Finite-temperature Monte Carlo calculations for systems with fermions". In: *Physical Review Letters* 83.14 (1999), p. 2777. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.83.2777.

[Sun+20b] Chong Sun, Ushnish Ray, Zhi-Hao Cui, Miles Stoudenmire, Michel Ferrero, and Garnet Kin-Lic Chan. "Finite-temperature density matrix embedding theory". In: *Physical Review B* 101.7 (2020), p. 075131. URL: https://journals.aps.org/prb/abstract/10.1103/PhysRevB.101.075131.

[FKB19] Jessica G Freeze, H Ray Kelly, and Victor S Batista. "Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists". In: *Chemical Reviews* 119.11 (2019), pp. 6595–6612. URL: https://pubs.acs.org/doi/10.1021/acs.chemrev.8b00759.

[ZLY12] Hong-Cai Zhou, Jeffrey R Long, and Omar M Yaghi. *Introduction to Metal–Organic Frameworks*. 2012. URL: https://pubs.acs.org/doi/10.1021/cr300014x.

[War14] Arieh Warshel. "Multiscale modeling of biological functions: from enzymes to molecular machines (Nobel Lecture)". In: *Angewandte Chemie International Edition* 53.38 (2014), pp. 10020–10031. URL: https://onlinelibrary.wiley.com/doi/10.1002/anie.201403689.

[Pro+20] Andrew H. Proppe et al. "Bioinspiration in light harvesting and catalysis". In: *Nature Reviews Materials* 5.11 (2020), pp. 828–846. URL: https://www.nature.com/articles/s41578-020-0222-0.

[PHG19] Hung Q Pham, Matthew R Hermes, and Laura Gagliardi. "Periodic electronic structure calculations with the density matrix embedding theory". In: *Journal of Chemical Theory and Computation* 16.1 (2019), pp. 130–140. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.9b00939.

[Chi+16] Wael Chibani, Xinguo Ren, Matthias Scheffler, and Patrick Rinke. "Self-consistent Green's function embedding for advanced electronic structure methods based on

a dynamical mean-field concept". In: *Physical Review B* 93.16 (2016), p. 165106. URL: https://journals.aps.org/prb/abstract/10.1103/PhysRevB.93.165106.

[Hea+20] Kade Head-Marsden, Johannes Flick, Christopher J Ciccarino, and Prineha Narang. "Quantum information and algorithms for correlated quantum matter". In: *Chemical Reviews* 121.5 (2020), pp. 3061–3120. URL: https://pubs.acs.org/doi/10.1021/acs.chemrev.0c00620.

[Qin+22] Mingpu Qin, Thomas Schäfer, Sabine Andergassen, Philippe Corboz, and Emanuel Gull. "The Hubbard model: A computational perspective". In: *Annual Review of Condensed Matter Physics* 13 (2022), pp. 275–302. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev-conmatphys-090921-033948.

[Din+20] Lexin Ding, Sam Mardazad, Sreetama Das, Szilárd Szalay, Ulrich Schollwöck, Zoltán Zimborás, and Christian Schilling. "Concept of orbital entanglement and correlation in quantum chemistry". In: *Journal of Chemical Theory and Computation* 17.1 (2020), pp. 79–95. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.0c00559.

[DS20] Lexin Ding and Christian Schilling. "Correlation paradox of the dissociation limit: A quantum information perspective". In: *Journal of Chemical Theory and Computation* 16.7 (2020), pp. 4159–4175. URL: https://pubs.acs.org/doi/pdf/10.1021/acs.jctc.0c00054.

[Wil13] Mark M Wilde. *Quantum Information Theory*. Cambridge University Press, 2013. URL: https://www.cambridge.org/core/books/quantum-information-theory/9DC2CA59F45636D4F0F30D971B677623.

[Har20] Aram W Harrow. "Small quantum computers and large classical data sets". In: *arXiv preprint arXiv:2004.00026* (2020). URL: https://arxiv.org/abs/2004.00026.

[Son+21] Wooyeong Song, Marcin Wieśniak, Nana Liu, Marcin Pawłowski, Jinhyoung Lee, Jaewan Kim, and Jeongho Bang. "Tangible reduction in learning sample complexity with large classical samples and small quantum system". In: *Quantum Information Processing* 20.8 (2021), p. 275. URL: https://link.springer.com/article/10.1007/s11128-021-03217-7.

[Aar07] Scott Aaronson. "The learnability of quantum states". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 463.2088 (2007), pp. 3089–3114.

[SW22] Mehdi Soleimanifar and John Wright. "Testing matrix product states". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2022, pp. 1679–1701. DOI: 10.1137/1.9781611977073.68. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611977073.68. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611977073.68.

[Wie+14a] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David Cory. "Quantum Hamiltonian learning using imperfect quantum resources". In: *Phys. Rev. A* 89 (4 Apr. 2014), p. 042314. DOI: 10.1103/PhysRevA.89.042314. URL: https://link.aps.org/doi/10.1103/PhysRevA.89.042314.

[Wie+14b] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and D. G. Cory. "Hamiltonian Learning and Certification Using Quantum Resources". In: *Phys.*

Rev. Lett. 112 (19 May 2014), p. 190501. DOI: 10.1103/PhysRevLett.112.190501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.112.190501.

[DGN22] Yulong Dong, Jonathan Gross, and Murphy Yuezhen Niu. "Beyond heisenberg limit quantum metrology through quantum signal processing". In: *arXiv preprint arXiv:2209.11207* (2022).

[CDP08] Giulio Chiribella, Giacomo M. D'Ariano, and Paolo Perinotti. "Memory Effects in Quantum Channel Discrimination". In: *Phys. Rev. Lett.* 101 (18 Oct. 2008), p. 180501. DOI: 10.1103/PhysRevLett.101.180501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.101.180501.

[KD13] Jan Kołodyński and Rafał Demkowicz-Dobrzański. "Efficient tools for quantum metrology with uncorrelated noise". In: *New Journal of Physics* 15.7 (July 2013), p. 073043. DOI: 10.1088/1367-2630/15/7/073043. URL: https://dx.doi.org/10.1088/1367-2630/15/7/073043.

[DM14] Rafal Demkowicz-Dobrza ński and Lorenzo Maccone. "Using Entanglement Against Noise in Quantum Metrology". In: *Phys. Rev. Lett.* 113 (25 Dec. 2014), p. 250801. DOI: 10.1103/PhysRevLett.113.250801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.250801.

[Don+21] Yulong Dong, Xiang Meng, K. Birgitta Whaley, and Lin Lin. "Efficient phase-factor evaluation in quantum signal processing". In: *Phys. Rev. A* 103 (4 Apr. 2021), p. 042419. DOI: 10.1103/PhysRevA.103.042419. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.042419.

[WDL22] Jiasu Wang, Yulong Dong, and Lin Lin. "On the energy landscape of symmetric quantum signal processing". In: *Quantum* 6 (Nov. 2022), p. 850. ISSN: 2521-327X. DOI: 10.22331/q-2022-11-03-850. URL: https://doi.org/10.22331/q-2022-11-03-850.

[Lok+18] Andrey Y. Lokhov, Marc Vuffray, Sidhant Misra, and Michael Chertkov. "Optimal structure and parameter learning of Ising models". In: *Science Advances* 4.3 (2018), e1700791. DOI: 10.1126/sciadv.1700791. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.1700791. URL: https://www.science.org/doi/abs/10.1126/sciadv.1700791.

[McK+18] David C McKay, Thomas Alexander, Luciano Bello, Michael J Biercuk, Lev Bishop, Jiayin Chen, Jerry M Chow, Antonio D Córcoles, Daniel Egger, Stefan Filipp, et al. "Qiskit backend specifications for OpenQASM and OpenPulse experiments". In: *arXiv preprint arXiv:1809.03452* (2018).

[DR93] Alok Dutt and Vladimir Rokhlin. "Fast Fourier transforms for nonequispaced data". In: *SIAM Journal on Scientific computing* 14.6 (1993), pp. 1368–1393.

[Bre13] G Larry Bretthorst. *Bayesian spectrum analysis and parameter estimation*. Vol. 48. New York, USA: Springer Science & Business Media, 2013.

[LBK21] Henning Lange, Steven L. Brunton, and J. Nathan Kutz. "From Fourier to Koopman: Spectral Methods for Long-Term Time Series Prediction". In: *J. Mach. Learn. Res.* 22.1 (Jan. 2021). ISSN: 1532-4435.

[She+21] Daniel E. Shea, Rajiv Giridharagopal, David S. Ginger, Steven L. Brunton, and J. Nathan Kutz. "Extraction of Instantaneous Frequencies and Amplitudes in Nonstationary Time-Series Data". In: *IEEE Access* 9 (2021), pp. 83453–83466. DOI: 10.1109/ACCESS.2021.3087595.

[Tho60] David J Thouless. "Stability conditions and nuclear rotations in the Hartree-Fock theory". In: *Nuclear Physics* 21 (1960), pp. 225–232. URL: https://www.sciencedirect.com/science/article/pii/0029558260900481.

[Sun+18] Qiming Sun et al. "PySCF: the Python-based simulations of chemistry framework". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 8.1 (2018), e1340. URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340.

[WP67] Dominic JA Welsh and Martin B Powell. "An upper bound for the chromatic number of a graph and its application to timetabling problems". In: *The Computer Journal* 10.1 (1967), pp. 85–86. URL: https://academic.oup.com/comjnl/article/10/1/85/376064.