Classical and quantum simulation of chemical and physical systems

Thesis by
Jiace Sun

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended Sep, 16, 2024

Jiace Sun
ORCID: 0000-0002-7329-0585

ACKNOWLEDGEMENTS

I want to thank my advisors, Professor Thomas F. Miller III and Austin J. Minnich, for their help and mentoring during my graduate studies. I would also like to thank my thesis committee, Professor Garnet Chan, Professor William Andrew Goddard III, and Professor Xie Chen, for their suggestions for my research and future career and their patience and help in scheduling all the meetings.

I feel lucky to be a graduate student at Caltech, where I did great research in various areas, and met so many talented young scientists who always provided valuable suggestions and supported me throughout all the difficult periods. I appreciate all the suggestions and efforts from my collaborators: Professor Nawaf Bou-Rabee at Rutgers University and Dr. Jorge L. Rosa-Raíces for research on molecular quantum dynamics; Dr. Lixue Cheng (my wife), Dr. Tamara Husch, Dr. J. Emiliano Deustua, Dr. Zhuoran Qiao, and Dr. Vignesh C. Bhethanabotla for research on machine learning of quantum chemistry; Dr. Iretomiwa Esho, Dr. Peishi Cheng, and Dr. Shi-Ning Sun for research on electron-phonon dynamics; Dr. Hirsh Kamakari, Dr. Mario Motta, and Dr. Yaodong Li for research about entanglement dynamics. Without your help and collaboration, I could not make such progress in these areas.

Finally, I want to thank my parents, who are always supporting my dream and helping me to realize it. Special thanks to my wife Lixue for accompanying me and collaborating on many of the research projects.

ABSTRACT

Various quantum mechanics effects have been found and widely studied in different microscopic systems, such as quantum nuclear effects and electron correlation in molecular systems, electron-phonon coupling in crystal systems, and quantum Zeno effects in open quantum systems. However, exact numerical simulations require exponentially scaled classical resources. In this thesis, we study these quantum systems by a series of classical or quantum methods, which include semiclassical, ab initio, machine learning, and quantum computing approaches.

In Chapter 2, we develop the molecular-orbital-based machine learning (MOB-ML) method as a general-purpose method to learn molecular electronic structure properties. By preserving physical constraints, including invariance conditions and size consistency, MOB-ML is shown to be able to capture both weak and strong interactions. Furthermore, the Gaussian Process framework is extended for learning both scalar properties such as energies, and linear-response properties like dipole moments with the rotationally equivariant derivative kernel. With these improvements, MOB-ML shows not only significantly higher learning rates for organic molecules, non-covalent interactions, and transition states but also excellent transferability from small systems to large systems.

In Chapter 3, we develop a generalized class of integrators in the thermostatted ring-polymer molecular dynamics (T-RPMD) method, which is a semi-classical quantum dynamics method to capture various types of molecular nuclear quantum effects, including zero point energy, quantum tunneling, and kinetic isotopic effects. Such generalized integrators are carefully designed to be strong stable and dimension-free, which are essential for robust numerical computations. In particular, a so-called "BCOCB" integrator is proved to be superior in terms of accuracy and efficiency in the harmonic limit. Such superiority is further verified in strongly anharmonic systems featured by liquid water.

In Chapter 4, we develop an ab initio-based semi-analytical model of electron-phonon scattering to describe the transport and noise behavior in GaAs, which is a widely-used semiconductor. Such a semi-analytical model lifts a few approximations in the standard ab initio calculation of intervalley scatterings, which were believed to be the origin of the failure to capture the nonmonotonic noise phenomena. We find qualitatively unchanged transport and noise properties and agreements

on the scattering rates between the photoluminescence experiments. These results indicate the most probable origin of the nonmonotonic noise behavior is the formation of space-charge domains rather than the intervalley scattering.

In Chapter 5, we simulate the challenging measurement-induced phase transitions (MIPT) behavior in quantum many-body systems on a superconducting quantum processor. Due to the intrinsic exponential scaling of the quantum state tomography and post-selection process, traditional simulations of MIPT were limited to a few qubits. With the recently introduced linear cross-entropy benchmarking, such exponential overhead is eliminated, and the correct critical behavior of MIPT is observed on a 22-qubit system. Our work paves the way for the studies of open quantum systems on large-scale near-term quantum devices.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. Kamakari, H., Sun, J., Li, Y., Thio, J. J., Gujarati, T. P., Fisher, M., Motta, M. & Minnich, A. J. Experimental demonstration of scalable cross-entropy benchmarking to detect measurement-induced phase transitions on a superconducting quantum processor. *arXiv preprint arXiv:2403.00938*. https://arxiv.org/abs/2403.00938 (2024).
   Contribution: J.S. designed and implemented the circuit compression algorithm, and contributed to writing the article.

2. Sun, J. & Minnich, A. J. Transport and noise of hot electrons in GaAs using a semianalytical model of two-phonon polar optical phonon scattering. *Phys. Rev. B* **107,** 205201. https://link.aps.org/doi/10.1103/PhysRevB.107.205201 (20 May 2023).
   Contribution: J.S. participated in the conception of the project, designed and implemented the algorithm, performed the numerical simulation, analyzed the results, and wrote the article.

3. Cheng, L., Sun, J., Deustua, J. E., Bhethanabotla, V. C. & Miller Thomas F., I. Molecular-orbital-based machine learning for open-shell and multi-reference systems with kernel addition Gaussian process regression. *The Journal of Chemical Physics* **157,** 154105. ISSN: 0021-9606. https://doi.org/10.1063/5.0110886 (Oct. 2022).
   Contribution: J.S. participated in the conception of the project, designed and implemented the algorithm, and contributed to writing the article.

4. Cheng, P. S., Sun, J., Sun, S.-N., Choi, A. Y. & Minnich, A. J. High-field transport and hot-electron noise in GaAs from first-principles calculations: Role of two-phonon scattering. *Phys. Rev. B* **106,** 245201. https://link.aps.org/doi/10.1103/PhysRevB.106.245201 (24 Dec. 2022).
   Contribution: J.S. participated in implementing the algorithm.

5. Sun, J., Cheng, L. & Miller Thomas F., I. Molecular dipole moment learning via rotationally equivariant derivative kernels in molecular-orbital-based machine learning. *The Journal of Chemical Physics* **157,** 104109. ISSN: 0021-9606. https://doi.org/10.1063/5.0101280 (Sept. 2022).
   Contribution: J.S. participated in the conception of the project, designed and implemented the algorithm, trained the ML models, and wrote the article.

6. Husch, T., Sun, J., Cheng, L., Lee, S. J. R. & Miller Thomas F., I. Improved accuracy and transferability of molecular-orbital-based machine learning: Organics, transition-metal complexes, non-covalent interactions, and transition states. *The Journal of Chemical Physics* **154,** 064108. ISSN: 0021-9606. https://doi.org/10.1063/5.0032362 (Feb. 2021).

Contribution: J.S. participated in the algorithm design and implementation of the improved features, and contributed to writing the article.

7. Rosa-Raíces, J. L., Sun, J., Bou-Rabee, N. & Miller Thomas F., I. A generalized class of strongly stable and dimension-free T-RPMD integrators. *The Journal of Chemical Physics* **154,** 024106. ISSN: 0021-9606. https://doi.org/10.1063/5.0036954 (Jan. 2021).
Contribution: J.S. conceptualized the project, formulated the theory, and contributed to writing the article.

8. Sun, J., Cheng, L. & Miller III, T. F. Molecular energy learning using alternative blackbox matrix-matrix multiplication algorithm for exact Gaussian process. *arXiv preprint arXiv:2109.09817.* https://arxiv.org/abs/2109.09817 (2021).
Contribution: J.S. conceptualized the project, designed and implemented the algorithm, trained the ML models, and wrote the article.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

LIST OF TABLES

Number *Page*

Chapter 1

INTRODUCTION

Understanding the behavior of physical or chemical systems is vital to various real-world and scientific applications, such as discovering materials [1, 2], drugs [3, 4], and new physics mechanisms [5]. At the atomic length scale, such systems are governed by quantum mechanics, which exhibits unique characteristics of superposition, entanglement, and wave-particle duality that are absent in classical mechanics. Such quantum systems can be simulated by diagonalizing the time-independent Hamiltonian for static behavior or solving the time-dependent Schrodinger equation for dynamical behavior [6]. However, exact solutions are not available except for systems involving one or two degrees of freedom, and numerical solutions generally suffer from the exponential growth of the many-body Hilbert space with respect to the system size [7, 8]. Despite these difficulties, various methods have been developed in the last a few decades to simulate different types of quantum systems. In this thesis, we consider three types of quantum systems, i.e. molecular systems, periodic crystal systems, and non-equilibrium many-body quantum systems. We review the previous efforts to simulate these systems by semi-classical, ab initio, machine learning, and quantum computing methods, and introduce our contribution to the simulation methods.

Molecular systems: Molecular systems are decomposed of nuclei and electrons, thus the Hamiltonian includes the nuclear kinetic energy, electronic kinetic energy, nuclear-nuclear repulsion, electron-electron repulsion, and nuclear-electron attractions. Since nuclei have much larger mass than electrons ($m_H \approx 1836 m_e$), electrons are relaxing much faster than nuclei. The so-called adiabatic or Born-Oppenheimer approximation [9]

$$\psi(\{\boldsymbol{R}_\alpha\},\{\boldsymbol{r}_\beta\}) \approx \psi_n(\{\boldsymbol{R}_\alpha\})\psi_e(\{\boldsymbol{r}_\beta\};\{\boldsymbol{R}_\alpha\}) \tag{1.1}$$

is then typically applied to decouple electrons, where ψ_n and ψ_e are nuclear and electron wavefuction respectively, and $\{\boldsymbol{R}_\alpha\}$, $\{\boldsymbol{r}_\beta\}$ are nuclear and electron coordinates. The electronic Hamiltonian $\hat{H}_e$ is solved for each given nuclear positions $\{\boldsymbol{R}_\alpha\}$, and the nuclear Hamiltonian $\hat{H}_n$ is solved with the electronic state energy as a function of $\{\boldsymbol{R}_\alpha\}$ (also known as the potential energy surface (PES)) obtained in the former task. In certain cases, such as conical intersection [10, 11], Born-Oppenheimer approximation fails, and non-adiabatic effects could be important [12, 13].

The electronic structure problem aims to obtain the eigenvalue and eigenstates of $\hat{H}_e$ as a function of the nuclear positions. Most of the electronic structure methods

employ a single-electron basis set (also referred as atomic orbitals) $\{\phi_\mu\}$ and decompose the wavefunction to the superposition of atomic orbitals [14, 15]. With such atomic orbitals, $\hat{H}_e$ can be written in the second-quantized form [16]

$$\hat{H}_e = \sum_{\mu\nu} h_{\mu\nu} c_\mu^\dagger c_\nu + \sum_{\mu\nu\lambda\sigma} V_{\mu\nu\lambda\sigma} c_\mu^\dagger c_\nu^\dagger c_\sigma c_\lambda, \tag{1.2}$$

where $c_\mu^\dagger$ and c_μ are the creation and annihilation operators of atomic orbitals. The two-electron term $\sum_{\mu\nu\lambda\sigma} V_{\mu\nu\lambda\sigma} c_\mu^\dagger c_\nu^\dagger c_\sigma c_\lambda$ results in the correlation between electrons, whose exact description requires diagonalization in exponentially-scaled Hilbert space, which is computationally infeasible. The mean-field-level Hartree–Fock (HF) theory [17] is a well-known approximation that neglects the effect from electron correlations by representing the wavefunction in the form of a Slater determinant [18]:

$$|\psi\rangle = \prod_{i \in \text{occ}} c_i^\dagger |0\rangle, \tag{1.3}$$

where $c_i^\dagger = \sum_\mu C_{i\mu} c_\mu^\dagger$ is the creation operator of the corresponding molecular orbital $\phi_i = \sum_\mu C_{i\mu} \phi_\mu$, and Eq. 1.3 loops over all the occupied molecular orbitals. Such Slater determinant wavefunction counts for the exchange effects originated from the fermionic anti-commutation requirement of the electron wavefunction. Some of the advanced electron structure methods such as Møller-Plesset perturbation [19], configuration interaction [20], and coupled cluster [21] are based on the HF method, and are known as the Post-Hartree-Fock methods. [22]

Aside from classical ab initio calculations, machine learning and quantum computing algorithms have also emerged to solve electronic structure problems. Most machine learning methods learn electronic energy and other molecular properties as a function of molecular structures [23–25], while other approaches have also been suggested by combining with density functional theory (DFT) or Quantum Monte Carlo[26, 27]. Quantum computing algorithms featured by quantum phase estimation (QPE) [28] and variational quantum eigensolver (VQE) [29] have also been developed to perform the electronic structure calculations on quantum computers. A molecular-orbital-based machine learning (MOB-ML) method is introduced in Chapter 2 of this thesis to learn molecular properties calculated by advanced electronic structure methods with the Hartree–Fock level input.

With the electron structure problem solved, the obtained PES enters into the nuclei Hamiltonian $\hat{H}_n$. The dynamical behavior of nuclei can be solved by the time-dependent Schrodinger equation. Unlike electrons, nuclei behave much more classically due to the heavy mass, and thus in many cases, the nuclei movements can

be well approximated by the classical Newton's laws of motion on the PES. Nevertheless, the quantum nuclear effects appear for light nuclei such as hydrogen and deuterium, and result in various phenomena including kinetic isotope effects, zero point energy, and quantum tunneling. [30, 31] Semi-classical treatments like path integral molecular dynamics (PIMD) [32] and full quantum treatments like multi-configuration time-dependent Hartree (MCTDH) [33] have been used to describe the nuclear quantum effects. Chapter 3 of this thesis develops the PIMD method by generalizing a set of strongly stable and dimension-free integrators for robust numerical computations.

Periodic crystal systems: In periodic crystal systems, atoms are defined on periodic unit cells, and the structure is translational invariant. This implies that the Hamiltonian and the translation operators in the three directions can be simultaneously diagonalized, and thus the eigenstates can be labeled by a real wave vector $\boldsymbol{k}$. A Fourier transformation is applied to the localized atomic orbitals ϕ_μ, which generates a set of Bloch functions $\phi_{\boldsymbol{k}\mu}$ as the atomic orbitals in the crystal electronic structure calculations. Within the mean-field framework via either HF or DFT, the mean-field Hamiltonian becomes

$$\hat{H}_e = \sum_{\boldsymbol{k}\mu} \varepsilon_{\boldsymbol{k}\mu} c^\dagger_{\boldsymbol{k}\mu} c_{\boldsymbol{k}\mu}, \tag{1.4}$$

while the electron correlation can be interpreted as collisions between electrons. Thus the energy levels are a set of energy bands labeled by the index μ. Each of these energy levels evolves smoothly with changes in $\boldsymbol{k}$, forming a smooth band of states.

Unlike the molecular situation, the electronic Hamiltonian is solved only on the equilibrium structure instead of as a function of the nuclear positions, while the atomic oscillations are treated perturbatively. Up to second-order perturbation, the oscillation normal modes (also known as phonons) and frequencies can be obtained by diagonalizing the dynamical matrix in the Fourier basis. Let the phonon creation and annihilation operators with wave vector $\boldsymbol{q}$ and branch i be $b^\dagger_{\boldsymbol{q}i}$ and $b_{\boldsymbol{q}i}$, and the frequency be $\omega_{\boldsymbol{q}i}$, the phonon Hamiltonian is given by

$$\hat{H}_{\text{ph}} = \sum_i \hbar\omega_{\boldsymbol{q}i}(b^\dagger_{\boldsymbol{q}i} b_{\boldsymbol{q}i} + \frac{1}{2}). \tag{1.5}$$

Compared with the molecular case, non-adiabatic effects in crystal, which is effectively the electron-phonon coupling, are more important due to its relation with

the transport [34]. The electron-phonon Hamiltonian can be obtained by Taylor expansion of the nuclear-electron attractions. The leading order electron-phonon Hamiltonian [35] is given by

$$\hat{H}_{\text{e-ph}} = \sum_{\mu\nu i\boldsymbol{k}\boldsymbol{q}} g_{\mu\nu i}(\boldsymbol{k},\boldsymbol{q}) c^{\dagger}_{\boldsymbol{k}+\boldsymbol{q},\mu} c_{\boldsymbol{k},\nu} (b_{\boldsymbol{q}j} + b^{\dagger}_{-\boldsymbol{q}j}), \tag{1.6}$$

which leads to the scattering of a single electron with a phonon created or annihilated. Higher orders of scattering are also available via higher order effects of the leading order Hamiltonian or higher orders of the Hamiltonian. Combining Eq. 1.4, 1.5, and 1.6 gives the Fröhlich Hamiltonian $\hat{H} = \hat{H}_e + \hat{H}_{\text{ph}} + \hat{H}_{\text{e-ph}}$.

Electron-phonon coupling is crucial for understanding many properties of materials. The scattering between electrons and phonons strongly affects the electric and thermal conductivity [36, 37]. Especially, in the Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity [38], electron-phonon interactions lead to the formation of Cooper pairs, which are pairs of electrons that can move through a lattice without resistance. The transport effects are typically described by the semi-classical Boltzmann transport equation [39]. An ab initio-based semi-analytical model of electron-phonon scattering is introduced in Chapter 4 of this thesis to describe the transport and noise behavior in GaAs.

Non-equilibrium many-body quantum systems: Unlike systems at equilibrium, where static properties and steady states dominate the analysis, non-equilibrium systems are characterized by time-dependent processes, driven by external forces, quenches, or interactions with their environment [40–42]. These systems can exhibit a rich variety of phenomena, such as relaxation, thermalization, and the emergence of non-thermal steady states, all of which are dictated by the intricate interplay between coherent quantum dynamics and dissipative effects [43–45].

One particularly intriguing aspect of non-equilibrium many-body dynamics is the study of quantum systems under continuous observation or measurement. In classical systems, measurements simply reveal the state of the system without altering its evolution significantly. However, in quantum mechanics, the act of measurement plays a fundamentally different role due to the collapse of the wavefunction, which can significantly influence the system's dynamics [46]. This interplay between unitary evolution and measurement has led to the discovery of novel phases of matter, termed measurement-induced phase transitions (MIPT) [47, 48].

Measurement-induced phase transitions occur in systems where frequent measurements compete with the unitary evolution driven by a many-body Hamiltonian. As

measurements are introduced, they can either stabilize certain quantum states or disrupt the coherence needed for complex quantum entanglement. When the rate of measurement is varied, the system can undergo a phase transition between different dynamical regimes.

Non-equilibrium many-body quantum systems can be classically simulated by strongly-correlated methods including tensor network [49], Quantum Monte Carlo [50], and dynamical mean-field theory [51]. Besides, quantum computers have also been used to simulate many-body dynamics via Trotterized time evolution, especially for highly entangled systems [52, 53]. Chapter 5 of this thesis introduces a demonstration of the MIPT behavior via the linear cross-entropy benchmarking on a superconducting quantum processor.

The thesis is organized as follows. In Chapter 2 we develop a machine learning solution named orbital-based machine learning (MOB-ML), which is originally introduced by our collaborators, for the electronic structure problem. Compared with the commonly used strategy that learns the electronic properties with the molecular structure as input, MOB-ML uses the Δ-learning strategy that learns the correlation energy (i.e. the difference with the Hartree–Fock energy) with the Hartree–Fock level information as the input. Furthermore, the Nesbet theorem is employed in MOB-ML to provide a rigorous decomposition of the correlation energy to local pair energies. Our developments of MOB-ML include three aspects. First, we carefully design the input features and the machine learning algorithm to ensure important physics constraints including invariance conditions and size consistency, which have been proven to be vital for traditional electronic structure methods. With such physical constraints, better learnability and transferability from small systems to large systems are observed in various types of benchmarking datasets. Second, we develop the alternative black-box matrix-matrix multiplication (AltBBMM) method, which is improved from the original BBMM method, and apply it to MOB-ML. Such AltBBMM method reduces the $O(N^3)$ training cost of the Gaussian Process Regression (GPR), which is the regressor of MOB-ML, to $O(N^2)$ via an iterative solver. AltBBMM also outperforms BBMM by a four-time speedup on the QM7B benchmarking dataset with negligible loss of accuracy. Third, we extend MOB-ML to learn linear-response properties by introducing the rotational-equivariant derivative kernel and generalizing the pair energy decomposition scheme. Such rotational-equivariant derivative kernel ensures the learned dipole moments to be the gradients of energy with respect to the external electric

field without explicitly including training data with non-zero electric field, and also ensures the rotational covariance requirement. The generalized pair energy framework ensures the locality of the machine learning target, which is vital for learning large systems. The accuracy and efficiency of the proposed approach are tested on various benchmark systems, and better accuracy is observed compared with all other literature methods so far.

In Chapter 3, we generalize the Cayley integrator proposed in our previous work to a family of integrators in the ring-polymer molecular dynamics (RPMD) algorithm, which is a type of the PIMD methods for simulation of molecular quantum dynamics. All such generalized integrators are strongly stable and dimension-free, which have been proven as the key properties of the previous Cayley integrator, and losing such properties leads to non-ergodicity and other pathological behaviors. We analytically evaluate such a family of integrators by equilibrium accuracy, time step stability, and convergence to equilibrium with the assumption of harmonic external potential. Within the generalized class, we find the previous Cayley integrator to be superior with respect to accuracy and efficiency for various configuration-dependent observables, although other integrators within the generalized class perform better for velocity-dependent quantities. The theoretical superiority is further verified with numerical benchmarking on both a one-dimensional quantum harmonic oscillator and a quantum-mechanical model of room-temperature liquid water.

In Chapter 4, we develop an ab initio-based semi-analytical model of the electron-phonon scattering in GaAs and study the transport behavior. The ab initio calculation of the electron-phonon scattering, especially the contribution from the two-phonon scattering, is computationally expensive, and thus several approximations such as on-shell approximation are typically applied. Such approximations were believed to be the origin of the failure to predict the characteristic non-monotonic trend of the current noise behavior, which is described by the current power spectral density (PSD). Our semi-analytical model lifts such approximations by assuming the electron-phonon scattering to be Fröhlich, which is a valid assumption in polar semiconductors like GaAs. Using the semi-analytical model, we calculate the drift velocity, scattering rates, and current PSD. Agreement within 15% for the drift velocity and 25% for the Γ valley scattering rates is found by comparing with the experimental results, which verifies the semi-analytical model. However, the monotonic PSD trend is not observed in the semi-analytical model. These results indicate that the most probable origin of the non-monotonic PSD trend is the formation of

space charge domains, instead of the intervalley electron-phonon scattering, which has been broadly believed so far.

In Chapter 5, we simulate the measurement-induced phase transition (MIPT) behavior on a superconducting quantum process. Such MIPT happens when a quantum many-body system undergoes both unitary dynamics which increase the entanglements, and measurements which decrease the entanglements. In our previous work, we observe the MIPT with up to 14 qubits on a superconducting quantum computer. However, the notorious post-selection issue leads to requirements of exponential resources, which prohibits the application to larger systems. To overcome this issue, we employ the recently introduced linear cross-entropy benchmarking, which eliminates such exponential overhead. With such linear cross-entropy benchmarking, we demonstrate the MIPT on up to 22 physical qubits with less than 8 device hours, which is significantly improved from the previous 5200 device hours. We also introduce a Clifford circuit compression technique, which allows us to investigate circuit models with all-to-all connectivities. We also compare the critical exponents, which are the key properties of phase transitions, with the theoretical predictions, and find good agreements. This work paves the way for studies of other critical phenomena on near-term quantum hardware and provides a potential benchmarking tool for quantum circuits with mid-circuit measurements.

Chapter 2

DEVELOPMENT OF MOLECULAR-ORBITAL-BASED MACHINE LEARNING FOR IMPROVED TRANSFERABILITY, SCALING, AND GENERALITY

This chapter is adapted from

1. Sun, J., Cheng, L. & Miller Thomas F., I. Molecular dipole moment learning via rotationally equivariant derivative kernels in molecular-orbital-based machine learning. *The Journal of Chemical Physics* **157,** 104109. ISSN: 0021-9606. https://doi.org/10.1063/5.0101280 (Sept. 2022).
   Contribution: J.S. participated in the conception of the project, designed and implemented the algorithm, trained the ML models, and wrote the article.

2. Husch, T., Sun, J., Cheng, L., Lee, S. J. R. & Miller Thomas F., I. Improved accuracy and transferability of molecular-orbital-based machine learning: Organics, transition-metal complexes, non-covalent interactions, and transition states. *The Journal of Chemical Physics* **154,** 064108. ISSN: 0021-9606. https://doi.org/10.1063/5.0032362 (Feb. 2021).
   Contribution: J.S. participated in the algorithm design and implementation of the improved features, and contributed to writing the article.

3. Sun, J., Cheng, L. & Miller III, T. F. Molecular energy learning using alternative blackbox matrix-matrix multiplication algorithm for exact Gaussian process. *arXiv preprint arXiv:2109.09817.* https://arxiv.org/abs/2109.09817 (2021).
   Contribution: J.S. conceptualized the project, designed and implemented the algorithm, trained the ML models, and wrote the article.

Electron correlation plays an important role in physical and chemical processes, but its quantitative description typically requires advanced ab-initio calculations. Molecular-orbital-based machine learning (MOB-ML) provides a general framework for the prediction of correlation energies at the cost of obtaining molecular orbitals. The application of Nesbet's theorem makes it possible to decompose the learning task of total correlation energies into the learning of local pair energies from pairs of molecular orbitals. Employing Gaussian Process Regression, MOB-ML is capable of reaching chemical accuracy on small molecules with dozens of training molecules. In this work, we introduce a few improvements, that extend the ability of MOB-ML from small systems to large systems, enhance its training efficiency, and allow it to learn other molecular properties. Specifically, the improved feature design automatically incorporates crucial physical constraints such as size consistency, which ensures the transferability of MOB-ML from small systems to large systems. The integration of a black-box matrix-matrix multiplication technique scales up MOB-ML training by nearly a hundredfold without compromising learning speed. Additionally, the introduction of a rotationally equivariant derivative kernel enables MOB-ML to learn general response properties with the same

efficiency as correlation energy learning. These advancements establish MOB-ML as a comprehensive, efficient, and robust framework for learning various molecular properties.

2.1 Introduction

The calculation of accurate potential energies of molecules and materials at affordable cost is at the heart of computational chemistry. While state-of-the-art ab initio electronic structure theories can yield highly accurate results, they are computationally too expensive for routine applications. Density functional theory (DFT) is computationally cheaper and has, thus, enjoyed widespread applicability. [54–56] However, DFT is hindered by a lack of systematic improvability and by an uncertain quality for many applications. [57, 58]

Applications of machine learning (ML) to electronic structure theory have grown rapidly with an increasing number of studies in a variety of chemical systems and applications [59, 60], such as directly predicting the molecular properties, developing force fields and interatomic potentials, and designing novel and efficient catalysts [61, 62], drugs [63–66], and materials [67, 68]. Important applications of machine learning include predicting chemical properties directly to reduce computational costs via supervised learning [69, 70], detecting the patterns of chemical spaces via unsupervised learning [71, 72], and proposing more suitable chemical systems via reinforcement learning [65, 73], and generative models [74, 75].

Numerous approaches have been presented in the field of machine learning for electronic structure during the last decades to aid in the learning of molecular energies and other molecular properties [25, 27, 76–114]. Among these methods, Molecular-Orbital-Based Machine Learning (MOB-ML) stands out for learning the correlation energy from molecular orbitals derived from Hartree–Fock computations. [89, 94, 95, 103, 115] The defining feature of MOB-ML is its framing of learning highly accurate correlation energies as learning a sum of orbital pair correlation energies. These orbital pair correlation energies can be individually regressed with respect to a feature vector representing the interaction of the molecular orbital pairs. Without approximation, it can be shown that such pair correlation energies add up to the correct total correlation energy for single-reference wave function methods. Phrasing the learning problem in this manner has the advantage that a given pair correlation energy, and, hence, a given feature vector, is independent of the molecular size (after a certain size threshold has been reached) because of

the inherent spatial locality of dynamic electron correlation. Utilizing these advantages, MOB-ML has shown great efficiency for highly accurate predicted molecular energies with few training data.

Despite the theoretical advantages obtained from the Nesbet theorem and the accuracy predictions on small systems, challenges remain for MOB-ML. First, MOB-ML shows a worse learnability in the transferability task of learning large molecules from small molecules [95], which hinders it from applying to more realistic systems. Second, MOB-ML employs the accurate GPR method with the price of the $O(N^3)$ computational complexity, which limits the training size to a few thousand. Considering the pair energy framework, MOB-ML training is limited to a few hundred molecules. Third, MOB-ML relies on the pair energy decomposition of the total correlation energies, which cannot be directly applied to other important molecular properties, such as linear response properties.

Three improvements are introduced to address these three challenges, respectively. To ensure the transferability to large systems, we demonstrate the importance of preserving physical constraints, including invariance conditions and size consistency, when generating the input for the machine learning model. The MOB-ML features are adjusted carefully to satisfy these physical constraints. With the new MOB-ML features, numerical improvements are demonstrated for different datasets covering total and relative energies for thermally accessible organic and transition-metal-containing molecules, non-covalent interactions, and transition-state energies. In particular, this MOB-ML model is significantly more accurate than other methods when transferred to a dataset comprising thirteen heavy atom molecules, exhibiting no loss of accuracy on a size-intensive (i.e., per-electron) basis.

To overcome the high computational complexity of GPR, we present an application of the black-box matrix-matrix multiplication (BBMM) algorithm, which has an improved computational scaling $O(N^2)$. [116, 117] An alternative implementation of BBMM (AltBBMM) is also proposed to train more efficiently (over four-fold speedup) with the same accuracy and transferability as the original BBMM implementation. The training of MOB-ML was limited to 220 molecules, and BBMM and AltBBMM scaled the training of MOB-ML up by over 30 times to 6500 molecules (more than a million pair energies). The accuracy and transferability of both algorithms are examined on the benchmark datasets of organic molecules with 7 and 13 heavy atoms.

Finally, to extend MOB-ML to learn general linear response properties featured

by dipole moments, a derivative kernel is introduced to learn the contribution of electron correlation to the dipole moments as a summation of pairwise contributions. Such derivative kernel automatically ensures the rotational equivariance of the predicted linear response properties and allows the co-training of different properties. The accuracy and efficiency of the proposed MOB-ML approach are tested on various benchmark systems, including water, fourteen small molecules, the QM9 benchmark dataset [118], and four series of peptides [106].

2.2 Review of MOB-ML

MOB-ML predicts correlation energies based on information from the molecular orbitals. [119–121] The correlation energy E^{corr} in the current study is defined as the difference between the true total electronic energy and the Hartree–Fock (HF) energy for a given basis set. Without approximation, the correlation energy is expressed as a sum over correlation energy contributions from pairs of occupied orbitals i and j, [122]

$$E^{\text{corr}} = \sum_{ij} \varepsilon_{ij}, \tag{2.1}$$

where i and j can be either canonical orbitals or non-canonical orbitals. In MOB-ML, we always use the local orbitals obtained by the Boys localization. [123] Since pair energies on local orbitals only depend on the local chemical environment, such construction ensures the accurate prediction of pair energies with a fixed size of input feature vectors.

Electronic structure theories offer different ways of approximating these pair correlation energies. For example, with the canonical orbitals, the second-order Møller-Plesset perturbation theory (MP2) correlation energy is [124]

$$\varepsilon_{ij}^{\text{MP2}} = \sum_{ab} \frac{\langle ia||jb\rangle^2}{F_{aa} + F_{bb} - F_{ii} - F_{jj}}, \tag{2.2}$$

where a, b denote virtual orbitals, F the Fock matrix in the molecular orbital basis, and $\langle ia||jb\rangle$ the anti-symmetrized exchange integral. The expression for non-canonical orbitals is more complicated, so we do not list it here. We denote a general repulsion integral over the spatial coordinates $\mathbf{x}_1, \mathbf{x}_2$ of molecular orbitals p, q, m, n following the chemist's notation as

$$\begin{aligned} [\boldsymbol{\kappa}^{pq}]_{mn} &= \langle pq|mn\rangle \\ &= \int d\mathbf{x}_1 d\mathbf{x}_2 p(\mathbf{x}_1)^* q(\mathbf{x}_1) \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|} m(\mathbf{x}_2)^* n(\mathbf{x}_2). \end{aligned} \tag{2.3}$$

The evaluation of correlation energies with post-HF methods like MP2 or coupled-cluster theory (including CCSD(T)) involves computations that exceed the cost of HF theory by orders of magnitude. By contrast, MOB-ML predicts the correlation energy at negligible cost compared to a HF calculation by machine-learning the map

$$\varepsilon_{ij} \approx \varepsilon^{\mathrm{ML}}(\mathbf{f}_{ij}), \tag{2.4}$$

where $\mathbf{f}_{ij}$ denotes the feature vector into which information on the molecular orbitals is compiled. MOB-ML is a so-called Δ machine learning (Δ-ML) model which, in this context, means that the difference between an energy calculated with a cheap (e.g., HF) and with an expensive (e.g., CCSD(T)) quantum chemical method is predicted. Other machine learning models which fall into this category are DeepHF [125] (based on a HF calculation), NeuralXC [126] (based on a DFT calculation), and OrbNet [127] (based on a semi-empirical calculation).

Following our previous work,[120] we define a canonical order of the orbitals i and j by rotating them into gerade and ungerade combinations (see Eq. (7) in Ref. [120]), creating the rotated orbitals $\tilde{i}$ and $\tilde{j}$. The feature vector $\mathbf{f}_{ij}$ assembles information on the molecular orbital interactions: (i) Orbital energies of the valence-occupied and valence-virtual orbitals F_{pp}, (ii) mean-field interaction energy of valence-occupied and valence-occupied orbitals and of valence-virtual and valence-virtual orbitals F_{pq}, (iii) Coulomb interaction of valence-occupied and valence-occupied orbitals, of valence-occupied and valence-virtual orbitals, and valence-virtual and valence-virtual orbitals $[\boldsymbol{K}^{pp}]_{qq}$, and (iv) exchange interaction of valence-occupied and valence-occupied orbitals, of valence-occupied and valence-virtual orbitals, and valence-virtual and valence-virtual orbitals $[\boldsymbol{K}^{pq}]_{pq}$. We note that all of these pieces of information enter either the MP2 or the MP3 correlation energy expressions, which helps to motivate their value within our machine learning framework. We remove repetitive information from the feature vector and separate the learning problem into the cases where (i) $i \neq j$ where we employ the feature vector as defined in Eq. (2.5) and (ii) $i = j$ where we employ the feature vector as defined in Eq. (2.6),

$$\begin{aligned}\mathbf{f}_{ij} = &\{\{F_{\tilde{i}\tilde{i}}, F_{\tilde{i}\tilde{j}}, F_{\tilde{j}\tilde{j}}\}, \{F_{\tilde{i}k}\}, \{F_{\tilde{j}k}\}, \{F_{ab}\},\\ &\{[\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{\tilde{i}\tilde{i}}, [\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{\tilde{j}\tilde{j}}, [\boldsymbol{K}^{\tilde{j}\tilde{j}}]_{\tilde{j}\tilde{j}}\}, \{[\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{kk}\}, \{[\boldsymbol{K}^{\tilde{j}\tilde{j}}]_{kk}\}, \{[\boldsymbol{K}^{\tilde{i}\tilde{i}}]_{aa}\}, \{[\boldsymbol{K}^{\tilde{j}\tilde{j}}]_{aa}\}, \{[\boldsymbol{K}^{aa}]_{bb}\},\\ &\{[\boldsymbol{K}^{\tilde{i}\tilde{j}}]_{\tilde{i}\tilde{j}}\}, \{[\boldsymbol{K}^{\tilde{i}k}]_{\tilde{i}k}\}, \{[\boldsymbol{K}^{\tilde{j}k}]_{\tilde{j}k}\}, \{[\boldsymbol{K}^{\tilde{i}a}]_{\tilde{i}a}\}, \{[\boldsymbol{K}^{\tilde{j}a}]_{\tilde{j}a}\}, \{[\boldsymbol{K}^{ab}]_{ab}\}\},\end{aligned} \tag{2.5}$$

$$\mathbf{f}_i = \{F_{ii}, \{F_{ik}\}, \{F_{ab}\}, [\boldsymbol{K}^{ii}]_{ii}, \{[\boldsymbol{K}^{ii}]_{kk}\}, \{[\boldsymbol{K}^{ii}]_{aa}\}, \{[\boldsymbol{K}^{aa}]_{bb}\}, \{[\boldsymbol{K}^{ik}]_{ik}\}, \{[\boldsymbol{K}^{ia}]_{ia}\}, \{[\boldsymbol{K}^{ab}]_{ab}\}\}. \tag{2.6}$$

Here, the index k denotes an occupied orbital other than i and j. For blocks in the feature vector that include more than one element, we specify a canonical order of the feature vector elements by the sum of the Euclidean distances between the centroids of orbital $\tilde{i}$ and p and between the centroids of orbital $\tilde{j}$ and p. A cutoff to 20 occupied orbitals and 20 virtual orbitals is applied to ensure that the feature vector $\mathbf{f}_{ij}$ always has a fixed length independent of the molecule size.

The universal functional in Eq. 2.4 is finally learned by GPR. GP [128] describes a prior distribution of random functions such that, for any finite number of possible inputs $\mathbf{X} = \{x_1, \ldots, x_n\}$, the function values $f(\mathbf{X}) = \{f(x_1), \ldots, f(x_n)\}$ has a multivariate Gaussian distribution

$$f(\mathbf{X}) \sim N(0, K(\mathbf{X}, \mathbf{X})), \tag{2.7}$$

where $K(\mathbf{X}, \mathbf{X})$ is the kernel matrix. Assuming the training data $(\mathbf{X}, \mathbf{y})$ has a Gaussian distributed noise with variance σ_n^2 (also referred as Gaussian likelihood), i.e., $\mathbf{y} \sim N(f, \Sigma_n)$, where $\Sigma_n = \sigma_n^2 I$, the GPR prediction $f(\mathbf{X}^\star)$ for the test points is a multivariate Gaussian distribution with the mean (prediction) and variance (uncertainty) as

$$\begin{aligned} \mathbb{E}[f(\mathbf{X}^\star)] &= K(\mathbf{X}^\star, \mathbf{X}) \hat{K}^{-1} \mathbf{y} \\ \mathrm{Var}[f(\mathbf{X}^\star)] &= K(\mathbf{X}^\star, \mathbf{X}) \hat{K}^{-1} K(\mathbf{X}, \mathbf{X}^\star), \end{aligned} \tag{2.8}$$

where

$$\hat{K} = K(\mathbf{X}, \mathbf{X}) + \Sigma_n. \tag{2.9}$$

Back to MOB-ML, $\mathbf{X}$ contains all feature vectors $\mathbf{f}_{ij}$, and $\mathbf{y}$ contains all pair energies ε_{ij}. In practice, the kernel function K is parameterized by some kernel parameters. All the parameters, which include the Gaussian noise variance σ_n^2 and kernel parameters, are optimized in GP training by maximizing the log marginal likelihood

$$L = -\frac{1}{2} \mathbf{y}^T \hat{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\hat{K}| - \frac{n}{2} \log 2\pi. \tag{2.10}$$

2.3 Improved feature design: Theory

Feature vector constructed by Eq. 2.5 and 2.6 includes enough information. However, they do not satisfy some important physical constraints, including the orbital-index permutation invariance and size consistency. In this section, we demonstrate the importance of these physical constraints, and introduce a set of improved feature designs so that these physical constraints are automatically satisfied.

Defining importance of feature vector elements

Careful ordering of the elements of the feature vector blocks is necessary in the current work because Gaussian process regression (GPR) is sensitive to permutation of the feature vector elements. Furthermore, the application of a Gaussian process requires that the feature vectors be of fixed length. [129]

Given the near-sighted nature of dynamical electron correlation, it is expected that only a limited number of orbital-pair interactions are important to predict the pair correlation energy with MOB-ML. To construct the fixed-length feature vector, a cutoff criterion must be introduced.[119] For some feature vector elements, a robust definition of importance is straight-forward. The spatial distance between the orbital centroids i and a is, for example, a reliable proxy for the importance of the feature vector elements $\{[\boldsymbol{\mathcal{K}}^{ii}]_{aa}\}$ of the feature vector $\mathbf{f}_i$. However, the definition of importance is less straightforward for feature vector elements that involve more than two indices. The most prominent example is the $\{[\boldsymbol{\mathcal{K}}^{ab}]_{ab}\}$ feature vector block of $\mathbf{f}_{ij}$, which contains the exchange integrals between the valence-virtual orbitals a and b and which should be sorted with respect to the importance of these integrals for the prediction of the pair correlation energy ε_{ij}. It is non-trivial to define a spatial metric which defines the importance of the feature vector elements $\{[\boldsymbol{\mathcal{K}}^{ab}]_{ab}\}$ to predict the pair correlation energy ε_{ij}; instead, we employ the the MP3 approximation for the pair correlation energy,

$$\begin{aligned}\varepsilon_{ij}^{\mathrm{MP3}} = &\frac{1}{8}\sum_{abcd}\left(t_{ij}^{ab}\right)^{*}\langle ab||cd\rangle\, t_{ij}^{cd} + \frac{1}{8}\sum_{klab}\left(t_{ij}^{ab}\right)^{*}\langle kl||ij\rangle\, t_{kl}^{ab} \\ &- \sum_{kabc}\left(t_{ij}^{ab}\right)^{*}\langle kb||ic\rangle\, t_{kj}^{ac},\end{aligned} \tag{2.11}$$

where t_{ij}^{ab} denotes the T-amplitude. Although we operate in a local molecular orbital basis, the canonical formulae are used to define the importance criterion; if we consider orbital localization as a perturbation (as in Kapuy–Møller–Plesset theory [130]), the canonical expression is the leading order term. The term we seek to attach an importance to, $\{[\boldsymbol{\mathcal{K}}^{ab}]_{ab}\}$, appears in the first term of Eq. (2.11) and all integrals necessary to compute this term are readily available as (a combination of) other feature elements, i.e., we do not incur any additional significant computational cost to obtain the importance of the feature vector elements.

The way in which we determine the importance of the $\{[\boldsymbol{\mathcal{K}}^{ab}]_{ab}\}$ elements here is an example of a more general strategy that we employ, in which the importance is assigned according to the lowest-order perturbation theory in which the features

first appear. Similar considerations have to be made for each feature vector block, all of which are specified in detail in Tables 6.1 and 6.2.

Orbital-index permutation invariance

The Fock, Coulomb, and exchange matrix elements that comprise MOB features are naturally invariant to rotation and translation of the molecule. However, some care is needed to ensure that these invariances are not lost in the construction of symmetrized MOB features. In particular, rotating the valence-occupied orbitals into gerade and ungerade combinations leads to an orbital-index permutation variance for energetically degenerate orbitals i, j because the sign of the feature vector elements $M_{\tilde{jp}}$, where $M_{\tilde{jp}}$ may be $F_{\tilde{jp}}$, $[\boldsymbol{\mathcal{K}}^{\tilde{jj}}]_{pp}$, or $[\boldsymbol{\mathcal{K}}^{\tilde{jp}}]_{\tilde{jp}}$,

$$M_{\tilde{jp}} = \frac{1}{\sqrt{2}}\left(M_{ip} - M_{jp}\right), \tag{2.12}$$

depends on the arbitrary assignment of the indices i and j. To rectify this issue, we include the absolute value of the generic feature vector element M in the feature vector instead of the signed value,

$$M_{\tilde{jp}} = \frac{1}{\sqrt{2}}\left|M_{ip} - M_{jp}\right|. \tag{2.13}$$

The corresponding equation,

$$M_{\tilde{jp}} = \frac{1}{\sqrt{2}}\left(M_{ip} + M_{jp}\right), \tag{2.14}$$

is already orbital-index permutation invariant because we chose M_{pq} $(p \neq q)$ to be positive. [120]

Size consistency

Size consistency is the formal property by which the energy of two isolated molecules equals the sum of their dimer upon infinite separation.[131, 132] In the context of MOB-ML, satisfaction of this property requires that the contributions from the diagonal feature vectors are not affected by distant, non-interacting molecules and that

$$\varepsilon^{ML}(\mathbf{f}_{ij}) = 0 \text{ for } r_{ij} = \infty \tag{2.15}$$

for contributions from the off-diagonal feature vectors. To ensure that MOB-ML exhibits size-consistency without the need for explicit training on the dimeric species, the following modifications to the feature vectors are made.

Diagonal feature vector. The feature vector as defined in Eq. (2.6) contains three blocks whose elements are independent of orbital i, $\{F_{ab}\}$, $\{[\boldsymbol{\mathcal{K}}^{aa}]_{bb}\}$, and $\{[\boldsymbol{\mathcal{K}}^{ab}]_{ab}\}$. The magnitude of these feature vector elements does not decay with an increasing distance between orbital i localized on molecule I and an orbital (for example, a) localized on molecule J. To address this issue, we multiply these feature vector elements by their estimated importance (see Section 2.3) so that they decay smoothly to zero. The other feature vector elements decay to zero when the involved orbitals are non-interacting albeit at different rates; we take the cube of feature vector elements of the type $\{[\boldsymbol{\mathcal{K}}^{pp}]_{qq}\}$ to achieve a similar decay rate for all feature vector elements in the short- to medium-range which facilitates machine learning.

Off-diagonal feature vector. We modify the off-diagonal feature vector such that $\mathbf{f}_{ij} = \mathbf{0}$ for $r_{ij} = \infty$ by first applying the newly introduced changes for $\mathbf{f}_i$ also for $\mathbf{f}_{ij}$. Further action is needed for the off-diagonal case because many feature vector elements do not decay to zero when the distance between i and j is large due to rotation of the orbitals into a gerade and an ungerade combination, e.g., $\tilde{F}_{ik} = \left|\frac{1}{\sqrt{2}}F_{ik} + \frac{1}{\sqrt{2}}F_{jk}\right| = \left|\frac{1}{\sqrt{2}}F_{ik}\right|$ for $r_{ij} = \infty, r_{jk} = \infty$. As a remedy, we apply a damping function of the form $\frac{1}{1+\frac{1}{6}(r_{ij}/r_0)^6}$ to each feature vector element. The form of this damping function is inspired by the semi-classical limit of the MP2 expression as it is also used for semi-classical dispersion corrections. [133] The damping radius, r_0, needs to be sufficiently large as to not interfere with machine learning at small r_{ij}. If a damping radius close to zero would be chosen, all off-diagonal feature vectors would be zero which nullifies the information content; however, the damping radius r_0 also should not be too large as size-consistency has to be fully learned until the off-diagonal feature vector is fully damped to zero. Therefore, we employ a damping radius in the intermediate-distance regime and we empirically found $r_0 = 5.0$ Bohr to work well. Although it could be systematically and automatically optimized in general, we simply apply $r_0 = 5.0$ Bohr throughout this work.

Lastly, we enforce that $\varepsilon^{ML}(\mathbf{0}) = 0$. The MOB features are engineered to respect this limit and would, for example, in a linear regression with a zero intercept trivially predict a zero-valued pair correlation energy without any additional training. However, the Gaussian process regression we apply in this work does not trivially yield a zero-valued pair correlation energy for a zero-valued feature vector. In the case that a training set does not include examples of zero-valued feature vectors, we need to include zero-valued feature vectors and zero-valued pair correlation ener-

gies in training to ensure that $\varepsilon^{ML}(\mathbf{0}) = 0$. For no model trained in the current study were more than 5% zero-valued feature vectors included.

The resulting MOB-ML model leads to size-consistent energy predictions to the degree to which the underlying MO generation is. It is not required that the dimer is explicitly part of training the MOB-ML model to obtain this result. The detailed definition of each feature vector block is summarized in Tables 6.1 and 6.2. We apply the feature set defined in Tables 6.1 and 6.2 consistently in this work.

2.4 Improved feature design: Results

Transferability within a molecular family

We first examine the effect of the feature vector generation strategy on the transferability of MOB-ML models within a molecular family. To this end, we revisit our alkane data set [120] which contains 1000 ethane and 1000 propane geometries as well as 100 butane and 100 isobutane geometries. We perform the transferability test outlined in Ref. [120], i.e., training a MOB-ML model on correlation energies for 50 randomly chosen ethane geometries and 20 randomly chosen propane geometries to predict the correlation energies for the 100 butane and 100 isobutane geometries (see Figure 2.1). This transferability test was repeated with 10000 different training data sets (each consisting of data for 50 ethane molecules and 20 propane molecules) to assess the training set dependence of the MOB-ML models. As suggested in Ref. [125], we consider various performance metrics to assess the prediction accuracy of the MOB-ML models: (i) the mean error (ME), (ii) the mean absolute error (MAE), (iii) the maximum absolute error (MaxAE), and (iv) the mean absolute relative error (MARE) which applies a global shift setting the mean error to zero. We report the minimum, peak, and maximum encountered MAREs in Table 2.1 alongside literature values obtained in our previous work [120], by Dick *et al.*, [126] and by Chen *et al.* [125]

In general, MOB-ML as well as NeuralXC[126] and DeePHF[125] produce MAREs well below chemical accuracy for correlation energies of butane and isobutane when trained on correlation energies of ethane and propane. Updating the feature vector generation strategy for MOB-ML results in the best peak MAREs for butane as well as for isobutane which are 0.11 kcal/mol and 0.10 kcal/mol, respectively. As in our previous work, [120] we note that the total correlation energy predictions may be

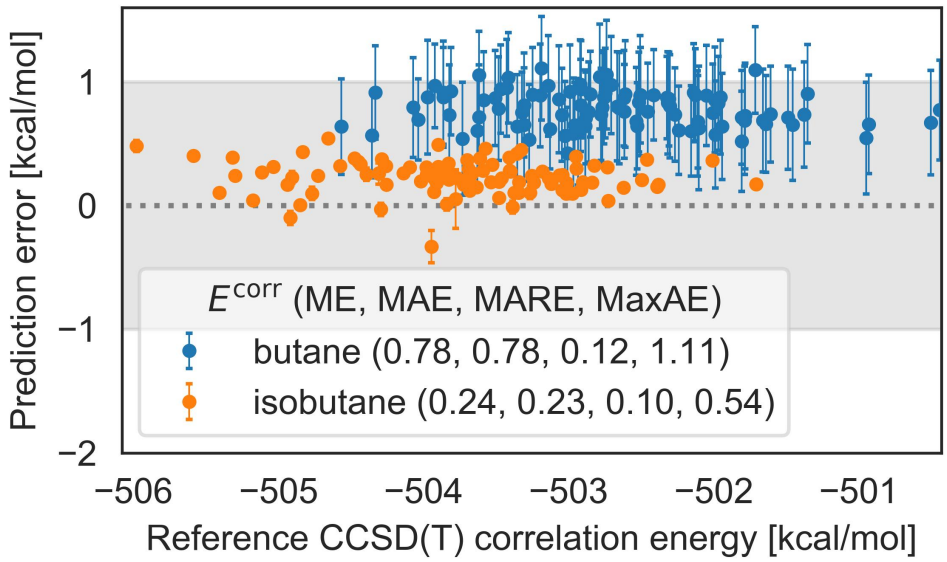


Figure 2.1: Errors in the predicted correlation energies with respect to the CCSD(T) reference values for butane and isobutane. The bar attached to each prediction error indicates the associated Gaussian process variance. The MOB-ML model used for these predictions was trained on 50 ethane and 20 propane molecules. The gray shaded area corresponds to the region where the error is smaller than chemical accuracy (1 kcal/mol).

Method	Feature set	MARE					
		Butane			Isobutane		
		min	peak	max	min	peak	max
NeuralXC[126]			0.15			0.14	
DeePHF[125]		0.06	0.11	0.43	0.07	0.13	0.53
MOB-ML	Ref. [120]		0.20			0.21	
	this work	**0.06**	**0.11**	**0.19**	**0.06**	**0.10**	**0.19**

Table 2.1: Comparison of the minimum, peak, and maximum mean absolute error after global shift (MARE) in kcal/mol for the prediction of CCSD(T) correlation energies for butane and isobutane obtained with different methods.

shifted with respect to the reference data so that the MEs for MOB-ML range from -0.92 to 2.70 kcal/mol for butane and from -0.18 to 1.02 kcal/mol for isobutane. This shift is strongly training-set dependent, which was also observed for results obtained with DeePHF [125].

The results highlight that this is an extrapolative transferability test. A considerable advantage of applying GPR in practice is that each prediction is accompanied by a Gaussian process variance which, in this case, indicates that we are in an extrapolative regime (see Figure 2.1). Here, this can be seen by comparing the variances of butane and isobutane molecules to variances of ethane and propane molecules not employed in training. The maximum variance for ethane molecules is below 0.01 kcal/mol and the one for propane molecules is below 0.06 kcal/mol. Extrapolations might be associated with quality degradation which we see, most prominently, for the mean error in butane. By contrast, other machine learning approaches like neural networks are less clear in terms of whether the predictions are in an interpolative or extrapolative regime.[134] By including the butane molecule with the largest variance in the training set (which then consists of 50 ethane, 20 propane, and 1 butane geometries) we reduce the ME from 0.78 to 0.25, MAE from 0.78 to 0.26, MaxAE from 1.11 to 0.51, and the MARE from 0.11 to 0.09 kcal/mol for butane. These results directly illustrate that MOB-ML can be systematically improved by including training data that is more similar to the test data; the improved confidence of the prediction is then also directly reflected in the associated Gaussian process variances.

As a second example, we examine the transferability of a MOB-ML model trained within a basin of a potential energy surface to the transition-state region of the same potential energy surface. We chose malonaldehyde for this case study as it has also been explored in previous machine learning studies [135]. We train a MOB-ML model on 50 thermalized malonaldehyde structures which all have the property that $d(\mathrm{O}^1\text{–H}) + d(\mathrm{O}^2\text{–H}) > 0.4$ Å (where d denotes the distance between the two nuclei) which ensures that we are sampling from the basins. We then apply this trained model to predict the correlation energies for an MP2 potential energy surface mapping out the hydrogen transfer between the two oxygen atoms (see Figure 2.2). MOB-ML produces an accurate potential energy surface for the hydrogen transfer in malonaldehyde only from information on the basins (compare left and middle left panel of Figure 2.2). The highest encountered errors on the minimum potential energy path are smaller than 1.0 kcal/mol. Unsurprisingly, the predicted minimum

energy structure ($d(O^1–H)$ = 1.00 Å, $d(O^2–H)$ = 1.63 Å) is very similar to the reference minimum energy structure ($d(O^1–H)$ = 1.00 Å, $d(O^2–H)$ = 1.64 Å). Strikingly, the predicted energy of 2.65 kcal/mol at the saddle point at $d(O^1–H) = d(O^2–H)$ = 1.22 Å differs from the reference energy by only 0.35 kcal/mol, although the MOB-ML model was not trained on any transition-state like structures. The highest errors are encountered in the high-energy regime and this region is also associated with the highest Gaussian process variance indicating low confidence in the predictions (compare middle right and right panel of Figure 2.2). The Gaussian process variance reflects the range of structures the MOB-ML model has been trained in and highlights again that we did not include transition-state-like structures in the training.

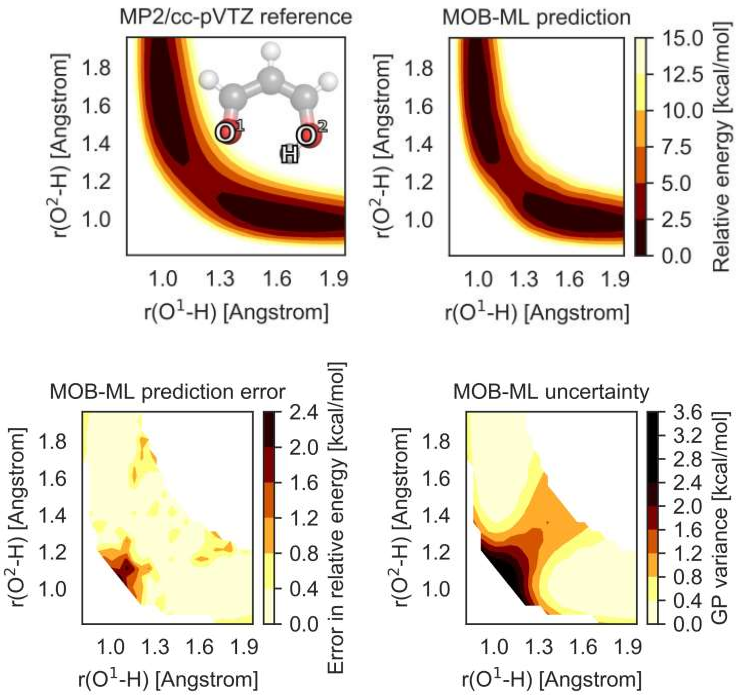


Figure 2.2: Relative energies obtained with MP2/cc-pVTZ (top left), relative energies predicted with MOB-ML (top right), the difference between the MOB-ML prediction and the reference data (bottom left), and the Gaussian process variance (bottom right) for the proton transfer in malonaldehyde as a function of the distance of the proton from the two oxygen atoms.

Transferability across organic chemistry space

The Chemical Space Project[136] computationally enumerated all possible organic molecules up to a certain number of atoms, resulting in the GDB databases.[137] In this work, we examine thermalized subsets [120] of the GDB13 data set [137] to investigate the transferability of MOB-ML models across organic chemistry space. The application of thermalized sets of molecules has the advantage that we can study the transferability of our models for chemical and conformational degrees of freedom at the same time. To test the transferability of MOB-ML across chemical space, we train our models on MP2 energies for a thermalized set of seven and fewer heavy-atom molecules (also known as QM7b-T [120]) and then we test the prediction accuracy for MP2 energies for a QM7b-T test set and for MP2 energies for a thermalized set of molecules with thirteen heavy atoms (GDB13-T [120]), as also outlined in our previous work. [120, 121] We demonstrated previously that MOB-ML learns other single-reference correlated wave function methods such as CCSD or CCSD(T) with similar efficiency.[120]

We first investigate the effect of changing the feature vector generation protocol on the QM7b-T→QM7b-T prediction task (see Figure 2.3). In Ref. [120], we found that training on about 180 structures is necessary to achieve a model with an MAE below 1 kcal/mol. The FCHL method yields an MAE below 1 kcal/mol when training on about 800 structures [138] and the DeePHF method already exhibits an MAE below 1 kcal/mol when training on their smallest chosen training set which consists of 300 structures (MAE=0.79 kcal/mol). [125] The refinements in the current work reduce the number of required training structures to reach chemical accuracy to about 100 structures when sampling randomly. This number is, however, strongly training set dependent. We can remove the training-set dependence by switching to an active learning strategy where we can achieve an MAE below 1 kcal/mol reliably with about 70 structures. In general, the MAE obtained with the active learning strategy is comparable to the smallest MAEs obtained with random sampling strategies. This has the advantage that a small number of reference data can be generated in a targeted manner.

In general, our aim is to obtain a machine learning model which reliably predicts broad swathes of chemical space. For an ML model to be of practical use, it has to be able to describe out-of-set molecules of different sizes to a similar accuracy when accuracy is measured size-intensively. [132] We probe the ability of MOB-ML to describe out-of-set molecules with a different number of electron pairs by

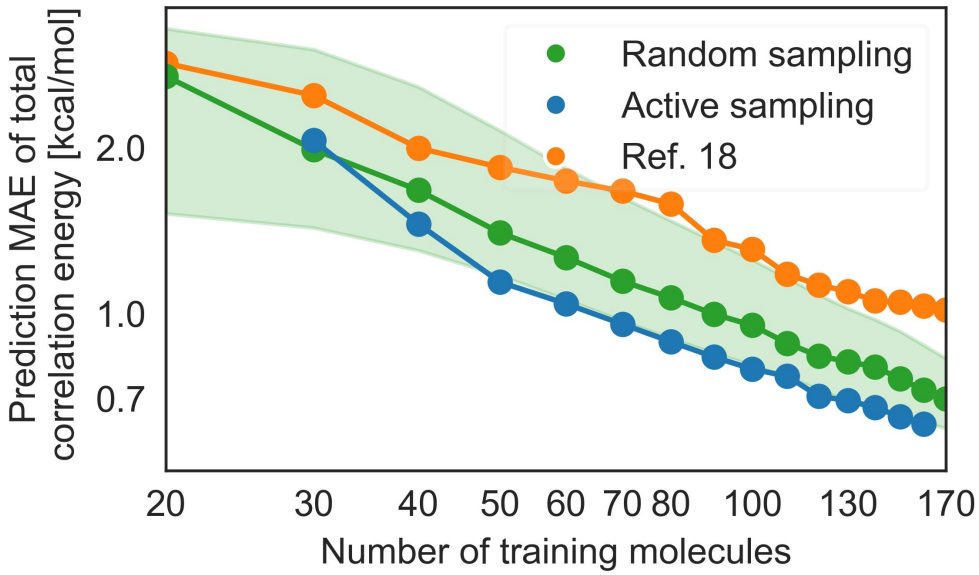


Figure 2.3: Comparison of the prediction mean absolute errors of total correlation energies for QM7b-T test molecules as a function of the number of QM7b-T molecules chosen for training for different machine learning models: MOB-ML as outlined in Ref. [120] (orange circles), MOB-ML as outlined in this work with random sampling (green circles), and MOB-ML as outlined in this work with active sampling. The green shaded area corresponds to the 90% confidence interval for the predictions obtained from 50 random samples of the training data.

applying a model trained on correlation energies for QM7b-T molecules to predict correlation energies for GDB13-T. We collect the best results published for this transfer test in the literature in Figure 2.4. Our previous best single GPR model achieved an MAE of 2.27 kcal/mol when trained on 220 randomly chosen structures. [120] The modifications in the current work now yield a single GPR model which achieves an MAE of 1.47–1.62 kcal/mol for GDB13-T when trained on 220 randomly chosen QM7b-T structures. Strikingly, MOB-ML outperforms machine learning models trained on thousands of molecules like our RC/GPR model and FCHL18 [138]. The current MOB-ML results are of an accuracy that is similar to the best reported results from DeePHF (an MAE of 1.49 kcal/mol);[125] however, MOB-ML only needs to be trained on about 3% of the molecules in the QM7b data set while DeePHF is trained on 42% to obtain comparable results (MAE of 1.52 kcal/mol for 3000 training structures). The best reported result for DeePHF (MAE of 1.49 kcal/mol) was obtained by training on 97% of the molecules of the QM7b data set. The data sets in this work focus on the extremely small data regime,

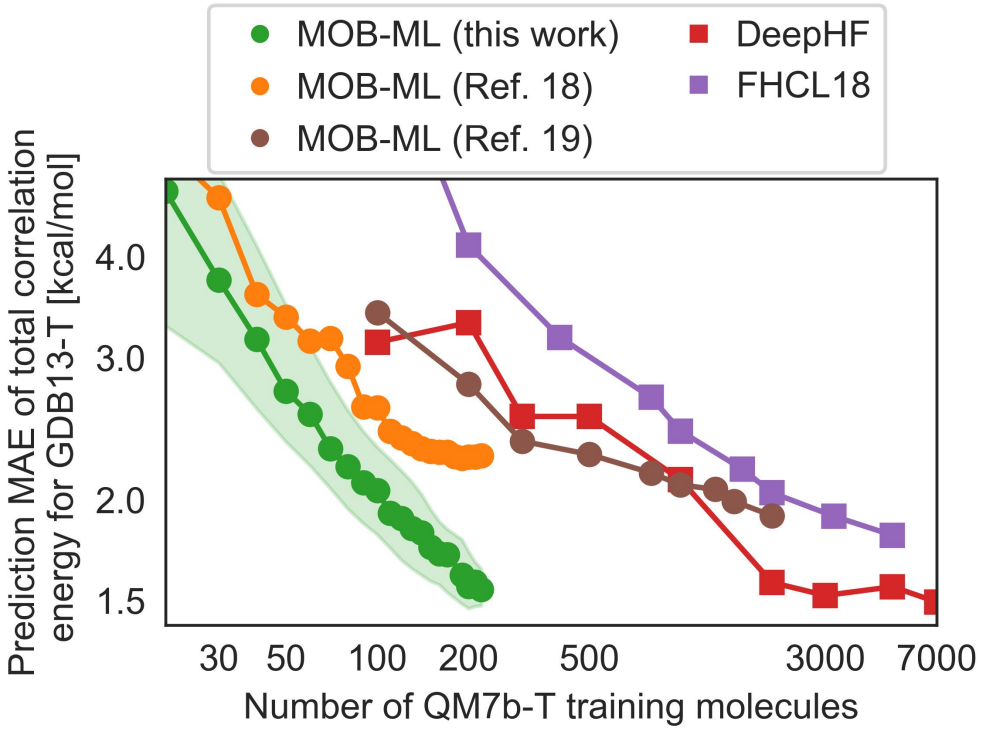


Figure 2.4: Comparison of the prediction mean absolute errors of total correlation energies for GDB13-T molecules as a function of the number of QM7b-T molecules chosen for model training for different machine learning models: MOB-ML as outlined in this work with random sampling (green circles), MOB-ML with a single GPR [120] (orange circles), MOB-ML with RC/GPR [121] (brown circles), DeePHF [125] (red squares), and FCHL18 [138] (purple squares). The green shaded area corresponds to the 90% confidence interval for the predictions obtained from 50 random samples of the training data.

whereas widely used neural net methods like SchNet or ANI-1 have instead been applied with data sets which are orders of magnitude larger. For example, a direct comparison of the learning efficiency between MOB-ML and SchNet has been provided in the context of forces.[139] We attribute the excellent transferability of MOB-ML to the fact that it focuses on the prediction of orbital-pair contributions, thereby reframing an extrapolation problem into an interpolation problem when training machine learning models on small molecules and testing them on large molecules. The pair correlation energies predicted for QM7b-T and for GDB13-T span a very similar range (0 to -20 kcal/mol), and they are predicted with a similar Gaussian process variance which we would expect in an interpolation task. The final errors for GDB13-T are larger than for QM7b-T, because the total correlation

energy is size-extensive; however, the size-intensive error per electron pair spans a comparable range for QM7b-T and for GDB13-T. This presents a significant advantage of MOB-ML over machine learning models which rely on a whole-molecule representation and creates the opportunity to study molecules of a size that are beyond the reach of accurate correlated wave function methods.

Most studies in computational chemistry require accurate relative energies rather than accurate total energies. Therefore, we also assess the errors in the relative energies for the sets of conformers for each molecule in the QM7b-T and in the GDB13-T data sets obtained with MOB-ML with respect to the MP2 reference energies (see Figure 2.5). We emphasize that MOB-ML is not explicitly trained to predict con-

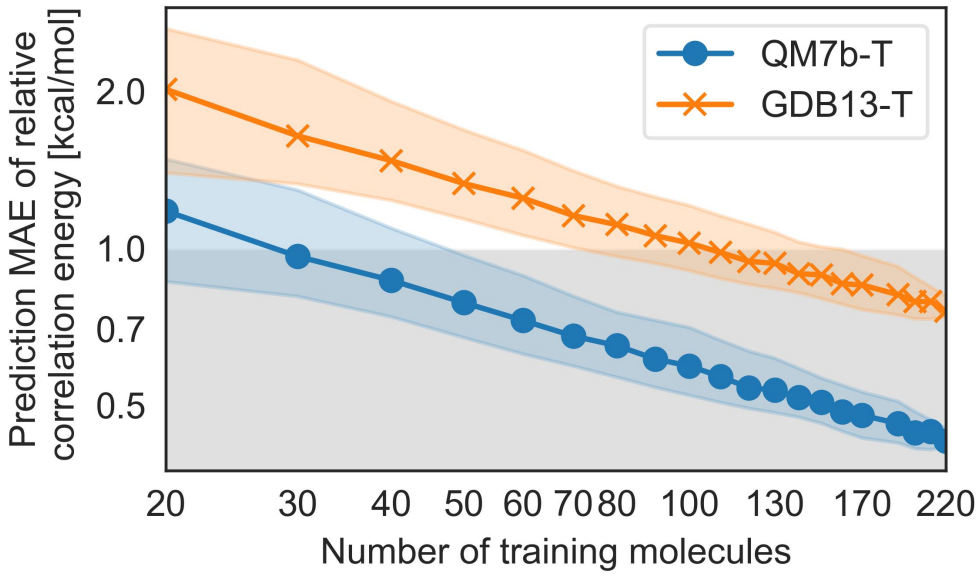


Figure 2.5: Prediction mean absolute errors for relative correlation energies as a function of the number of QM7b-T molecules chosen for model training for QM7b-T (blue circles) and for GDB13-T (orange crosses). The blue and orange shaded areas correspond to the 90% confidence interval for the predictions obtained from 50 random samples of the training data. The gray shaded area corresponds to the region where the error is smaller than chemical accuracy (1 kcal/mol).

former energies, and we include at most one conformer for each molecule in the training set. Nevertheless, MOB-ML produces on average chemically accurate relative conformer energies for QM7b-T when trained on correlation energies for only 30 randomly chosen molecules (or 0.4% of the molecules) in the QM7b set. We obtain chemically accurate relative energies for the GDB13-T data set when training on about 100 QM7b-T molecules. The prediction accuracy improves steadily

when training on more QM7b-T molecules reaching a mean MAE of 0.43 kcal/mol for the relative energies of the rest of the QM7b-T set and of 0.77 kcal/mol for the GDB13-T set.

We now present the first reported test of MOB-ML for non-covalent interactions in large molecules. To this end, we examine the backbone-backbone interaction (BBI) data set [140] which was designed to benchmark methods for the prediction of interaction energies encountered within protein fragments. Using the implementation of MOB-ML described here and using only MP2 energies for 20 randomly selected QM7b-T molecules for training, the method achieves a mean absolute error of 0.98 kcal/mol for the BBI data set (see Figure 2.6). However, these predictions are uncertain as indicated by the large Gaussian process variances associated with these data points which strongly suggest that we are now, as expected, in an extrapolative regime. We further improve the predictive capability of MOB-ML by augmenting the MOB-ML model with data from the BBI set. Specifically, we can draw on an active learning strategy and consecutively include data points until all uncertainties are below 1 kcal/mol which in this case corresponds to only two data points. This reduces the MAE to 0.28 kcal/mol for the remaining 98 data points in the BBI set. Including more reference data points would further improve the performance for this specific data set. However, this is not the focus of this work. Instead, we simply emphasize that MOB-ML is a clearly extensible strategy to accurately predict energies for large molecules and non-covalent intermolecular interactions while providing a useful estimation of confidence.

Transition-metal complexes

We finally present the first application of MOB-ML to transition-metal complexes. To this end, we train a MOB-ML model on a thermalized subset of mononuclear, octahedral transition-metal complexes introduced by Kulik and co-workers [141] which we denote as TM-T. The chosen closed-shell transition-metal complexes feature different transition metals (Fe, Co, Ni) and ligands. The ligands span the spectrochemical series from weak-field (e.g., thiocyanate) over to strong-field (e.g., carbonyl) ligands. We see in Figure 2.7 that the learning behaviour between TM-T and QM7b-T is similar when the error is measured per valence-occupied orbital. These results demonstrate that MOB-ML formalism can be straightforwardly applied outside of the organic chemistry universe without additional modifications. It is particularly notable that the learning efficiency for TM-T is comparable to that for QM7b-T, as seen in the relatively simple organic molecules in QM7b-T (Fig. 2.7).

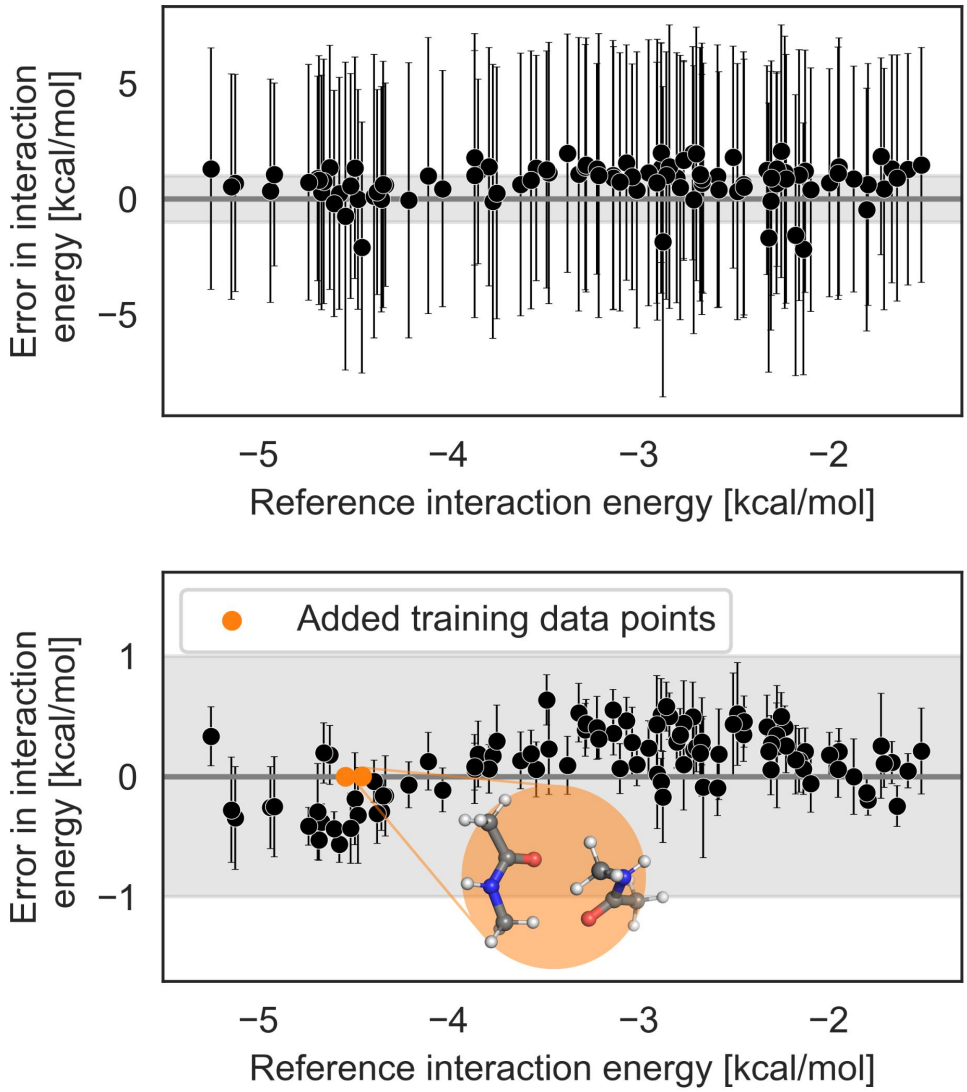


Figure 2.6: Top panel: Errors in predictions were made with a MOB-ML model trained on 20 randomly selected QM7b-T molecules with respect to reference MP2/cc-pVTZ interaction energies for the BBI data set. Bottom panel: Errors in predictions were made with a MOB-ML model trained on 20 randomly selected QM7b-T molecules and augmented with the 2 BBI data points with the largest variance (orange circles) with respect to reference MP2/cc-pVTZ interaction energies. The bar attached to each prediction error indicates the associated Gaussian process variance. The gray shaded area corresponds to the region where the error is smaller than chemical accuracy (1 kcal/mol).

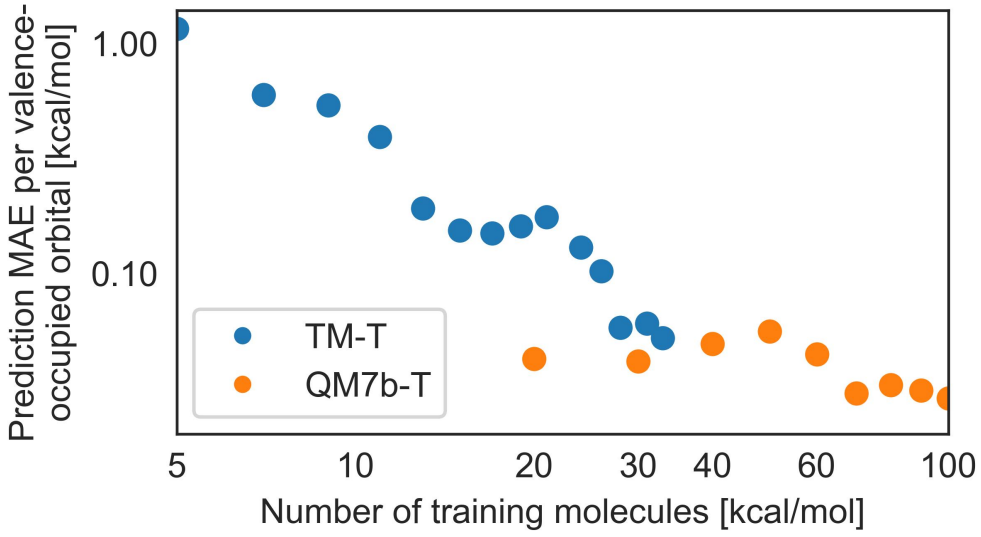


Figure 2.7: Learning curve for the prediction of MP2 correlation energies per valence-occupied orbital for transition metal complexes (TM-T) and for QM7b-T as a function of the number of structures the MOB-ML model was trained on.

We note that we do not expect MP2 theory to be quantitative for transition metal complexes.[142, 143] Instead, it provides a demonstration of the learning efficiency of MOB-ML for transition-metal complexes in the current example; and as previously demonstrated, MOB-ML learns other correlated wave function methods with similar efficiency.[119, 120]

2.5 Scaling up MOB-ML training by AltBBMM

Applying MOB-ML to the large data regime has remained challenging due to the $O(N^3)$ complexity scaling associated with the standard GPR algorithm. A strategy to reduce the complexity of GPR is to introduce a low-rank kernel approximation, which has been exploited in the Sparse Gaussian Process Regression [144] and Stochastic Variational Gaussian Processes [145] methods. However, such treatments of GPR sometimes result in a significant loss of accuracy. In contrast, Gardner et al. [116, 117] recently proposed the black box matrix-matrix multiplication (BBMM) method, which provides exact GP inference while reducing the training time complexity to $O(N^2)$ and allowing for multi-GPU usage.

In the following, we employ BBMM and a novel alternative implementation for BBMM (AltBBMM) to speed up and scale the GPR training in MOB-ML for molecular energies. We show that AltBBMM delivers more efficient training on

over 1 million pair energies without sacrificing transferability across chemical systems of different molecular sizes. The accuracy and efficiency of BBMM and AltBBMM in modeling physical problems are demonstrated by comparisons with literature results on the same datasets.

Conjugate gradient (CG)

The standard approach to obtain the predictive mean in Eq.2.8 via direct matrix inverse has a $O(N^3)$ computational complexity. In constrast, the conjugate gradient (CG) [146] algorithm can iteratively solve $\omega = \hat{K}^{-1}y$, or equivalently $\hat{K}\omega = y$ with an $O(N^2)$ cost in each iteration. CG requires only the matrix-vector multiplications (MVMs) with the kernel matrix $\hat{K}$, which is amenable to multi-GPU acceleration. In the iteration k, the solution is found in the order-k Krylov space

$$\mathscr{K} = \text{span}\left\{\hat{K}^i y \middle| i = 0,1,...k-1\right\}. \tag{2.16}$$

The solution ω^k of CG at iteration k converges to the exact solution ω^* exponentially measured by the relative residual $\frac{||\hat{K}\omega^k - y||}{||y||}$. However, the total number of iterations k_c to converge is usually very large for a kernel $\hat{K}$ with high singularity. A common way to reduce k_c is to construct a preconditioner P and then solve the equivalent equation $P^{-1}\hat{K}\omega = P^{-1}y$ such that $P^{-1}\hat{K}$ is less singular than $\hat{K}$. [147].

Block conjugate gradient (BCG) [148], as a variant of CG, can also be used to further reduce k_c. It extends CG to solve s linear equations $\hat{K}\omega_i = y_i, i = 0,1...s-1$ simultaneously. The number of linear equations s is also known as block size. In the iteration k of BCG, the solution is found in

$$\mathscr{K}^{\text{block}} = \text{span}\left\{\hat{K}^j y_i \middle| i = 0,1,...s-1, j = 0,1,...k-1\right\}. \tag{2.17}$$

By setting $y_0 = y$, and $y_i \sim N(0,I)$ for $i > 0$, BCG can converge to the same exact solution $\omega_0^* = \omega^*$ with fewer iterations since $\mathscr{K}_k \subset \mathscr{K}_k^{\text{block}}$.

Blackbox matrix-matrix multiplication (BBMM) and Alternative BBMM (AltBBMM)

BBMM [116, 117] calculates the GP inference by utilizing CG combined with the pivoted Cholesky decomposition preconditioner [149, 150]. Furthermore, a modified batched version of conjugate gradients (mBCG[1]) [116] is also proposed to estimate the marginal likelihood and its derivatives, which are required in the GP

[1] mBCG (modified batched conjugate gradients) differs from BCG (block conjugate gradient)

hyperparameter optimization. These enhancements reduce the training complexity to $O(N^2)$ in time, and $O(N)$ in memory and therefore enable the training of a million data points.

Here, we propose an alternative realization of BBMM (AltBBMM) to achieve similar accuracy as BBMM with a lower cost in molecular energy prediction applications, where a low Gaussian noise ($10^{-5}\sim10^{-8}$) is required to reach the desired accuracy. However, since the low Gaussian noise significantly increases the singularity of $\hat{K}$, CG would converge slowly or even fail to converge when the rounding errors exceed the Gaussian noise.[151] In order to speedup the CG convergence, we employ the BCG algorithm described in the previous section. The additional computational cost of BCG in each iteration is negligible compared with the kernel matrix calculations. To further improve the robustness of the convergence, we use the double-precision floating numbers in the implementation and employ the symmetric preconditioning $P^{-1/2}\hat{K}P^{-1/2}$. The Nystreoem preconditioner [147] is used as an example, but we note that better preconditioners could exist. Finally, the hyperparameters are optimized on a random subset of the entire training set in AltBBMM since the optimized hyperparameters remain similar across various training sizes for MOB-ML.

Computational details

We train all the models on random subsets of the QM7b-T dataset [89, 94, 152], which contains 7211 organic molecules with up to 7 heavy atoms. The test sets are the remaining QM7b-T molecules and the whole GDB-13-T dataset [89, 94, 152] containing 1000 organic molecules with 13 heavy atoms. The Matérn 5/2 kernel is used in all the GP trainings. We independently implement the BBMM according to the description of the mBCG and hyperparameter optimization in Ref. [116]. The symmetric Nystroem preconditioner and the block CG are used in this work. In both BBMM and AltBBMM, the rank r of the preconditioner is chosen as 10000, the BCG block size s is fixed as 50, and the BCG iterations stop when all the s relative residuals are smaller than 10^{-6}. The hyperparameters are optimized from a full GP trained on 50 random molecules. To overcome the memory limit and maximize the multi-GPU efficiencies, the kernel computations in CG are performed in 4096×4096 batches, and such computations are dynamically distributed to all the available GPUs. Additionally, we add a Gaussian noise regularization $\sigma^2_{\text{add}} = 10^{-5}$ to the optimized Gaussian noise reduce the singularity of $\hat{K}$.

Low noise regularization for accurate GP

We first demonstrate the necessity of utilizing a low noise regularization to achieve accurate predictions. We train all the offdiagonal energies (ε_o) pairs from 1000 QM7b-T molecules with different σ^2_{add} and test on the ε_o of the rest QM7b-T molecules. The training time and the prediction mean absolute error (MAE) are displayed in Table 2.2. For both BBMM and AltBBMM, regularizing with $\sigma^2_{\text{add}} = 10^{-1}$ results in a less singular $\hat{K}$ and saves half the training time, but its prediction MAE doubles when compared to the results of $\sigma^2_{\text{add}} = 10^{-5}$. Since the MOB-ML data generation is significantly more expensive than model training, we fix $\sigma^2_{\text{add}} = 10^{-5}$ for all of the following BBMM and AltBBMM experiments to achieve the most accurate model with the least amount of data.

Table 2.2: Test MAEs (kcal/mol) of offdiagonal contributions ($\sum \varepsilon_o$) in each molecule, training time (s) and memory usage (MB) by training on ε_o pairs from 1000 QM7b-T molecules (N=175,795) with different Gaussian noise regularizations.

	BBMM		AltBBMM		Memory/GPU (MB)
σ^2_{add}	Test MAE	Time (s)	Test MAE	Time (s)	
10^{-1}	0.636	1104.30	0.619	456.46	3,891
10^{-5}	0.314	2150.93	0.312	760.54	

Table 2.3: Test MAEs (kcal/mol), training time (hrs) and memory usage (MB) of BBMM and AltBBMM trained on 6500 QM7b-T molecules[a] with the same initial hyperparameters.

Algorithm	QM7b-T MAE	GDB-13-T MAE/7HA	Time (hr)	Memory/GPU (MB)
BBMM	0.185	0.490	26.52	15,359
AltBBMM	0.193	0.493	6.24	

[a] Training size of ε_o is 1,152,157 and training size of ε_d is 124,973

BBMM and AltBBMM for energies of organic molecules

We now examine the accuracy and transferability of BBMM and AltBBMM in learning QM7b-T and GDB-13-T molecular energies. The transferability of MOB-ML is assessed by the MAEs per 7 heavy atoms (MAE/7HA) of test GDB-13-T molecules predicted by the models trained on QM7b-T molecules. Table 2.3 lists the wall-clock time of training on 6500 QM7b-T molecules by BBMM and AltBBMM and the corresponding prediction MAEs on test QM7b-T and GDB-13-T molecules. Similar to the results in Table 2.2, by utilizing our AltBBMM approach,

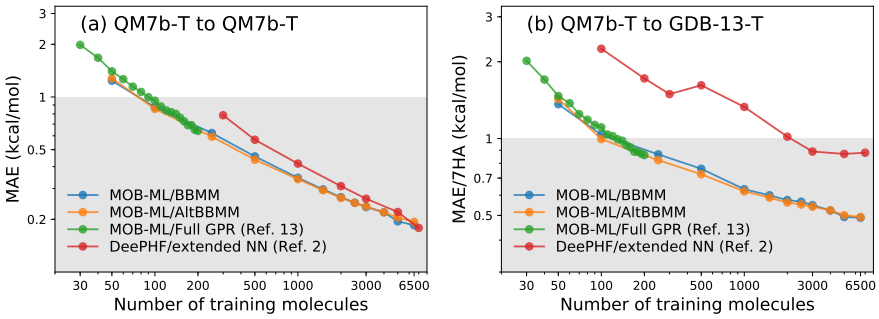


Figure 2.8: Learning curves for MOB-ML with different training protocols trained on QM7b-T and tested on (a) QM7b-T and (b) GDB-13-T. The accuracies of QM7b-T and GDB-13-T are measured by the MAEs and MAEs per 7 heavy atoms (MAE/7HA) of test molecules, respectively. We additionally plot the current best results in low and big data regimes, i.e., MOB-ML training with full GPR from Ref. [152] and the state-of-art DeePHF/extended NN from Ref. [153], respectively. The gray shaded area represents the chemical accuracy of 1 kcal/mol.

we gain a four-fold speedup in the training timings while only introducing 4% and 1% additional MAE in the prediction of QM7b-T and GDB-13-T, respectively, compared with BBMM.

In addition, we compare the performance of BBMM and AltBBMM with the results of the current most accurate literature methods, i.e., MOB-ML with full GP (MOB-ML/Full GP) [152] and DeePHF with extended neural network regressor (DeePHF/extended NN) [153]. The literature results of MOB-ML/Full GP are only available with up to 220 training molecules due to the limited memory resources. The introduction of BBMM and AltBBMM allows MOB-ML to scale up the training to 6500 molecules (over 1 million training pair energies) while retaining the accuracy and transferability compared with MOB-ML/full GP in Figure 2.8. By training on 6500 molecules, BBMM and AltBBMM reach the current best MAE/7HA for GDB-13-T as 0.490 kcal/mol and 0.493 kcal/mol, respectively. In all the cases we tested, BBMM and AltBBMM provide a better accuracy on QM7b-T and a better transferability on GDB-13-T than DeePHF/extended NN.

2.6 Extension to learning linear response properties: Theory

In the following, we extend MOB-ML to learn general time-independent linear response properties using dipole moments as an example. Following the idea of pair energies, we first introduce a decomposition of the contribution of electron correlation to the dipole moment to a summation of pairwise contributions. These pair

dipole moments can be learned as the functions of MOB features and their derivatives to electric fields via a rotationally equivariant derivative kernel in GPR. Such a framework also allows the natural co-training of energies and dipole moments.

MO Decomposition of dipole moments in MOB-ML

For a given system, the electric dipole moments $\boldsymbol{\mu}$ can be expressed as the linear response of the energy E with respect to external electric field $\boldsymbol{\mathscr{E}}$, i.e.,

$$\boldsymbol{\mu} = -\nabla_{\boldsymbol{\mathscr{E}}} E, \tag{2.18}$$

which could be further expressed as the sum of HF and correlation components:

$$\boldsymbol{\mu} = -(\nabla_{\boldsymbol{\mathscr{E}}} E^{\mathrm{HF}} + \nabla_{\boldsymbol{\mathscr{E}}} E^{\mathrm{corr}}) = \boldsymbol{\mu}^{\mathrm{HF}} + \boldsymbol{\mu}^{\mathrm{corr}}. \tag{2.19}$$

The correlation part $\boldsymbol{\mu}^{\mathrm{corr}}$ can then be decomposed on pairs of occupied orbitals similar to Eq. 2.1:

$$\boldsymbol{\mu}^{\mathrm{corr}} = -\nabla_{\boldsymbol{\mathscr{E}}} E^{\mathrm{corr}} = -\sum_{ij\in \mathrm{occ}} \nabla_{\boldsymbol{\mathscr{E}}} \varepsilon_{ij} = \sum_{ij\in \mathrm{occ}} \boldsymbol{\mu}_{ij}. \tag{2.20}$$

$\boldsymbol{\mu}_{ij}$ is referred to as pair dipole moments and is regressed by ML similar to Eq. 2.4. Compared with Eq. 2.4, we add the feature derivatives information $\nabla_{\boldsymbol{\mathscr{E}}} \boldsymbol{f}_{ij}(\phi_k)$ as part of the features motivated by $\boldsymbol{\mu}_{ij} = \nabla_{\boldsymbol{\mathscr{E}}} \varepsilon_{ij}$.

$$\boldsymbol{\mu}_{ij} \approx \boldsymbol{\mu}^{\mathrm{ML}}[\boldsymbol{f}_{ij}(\phi_k), \nabla_{\boldsymbol{\mathscr{E}}} \boldsymbol{f}_{ij}(\phi_k)]. \tag{2.21}$$

Figure 2.9 displays an example of the dipole moment decomposition on a water molecule to facilitate an understanding of this decomposition.

Feature design of dipole learning in MOB-ML

The energy feature set $\boldsymbol{f}_{ij}^{\varepsilon}$ includes enough information to model molecular energy and satisfies different invariance properties, including translational, rotational, and orbital permutational invariances. [94, 102] To efficiently model the dipole moments, we additionally include the responses of feature vector to electric field $\boldsymbol{\mathscr{E}}$, i.e., $\nabla_{\boldsymbol{\mathscr{E}}} \boldsymbol{f}_{ij}^{\varepsilon}$, in the design of dipole feature set $\boldsymbol{f}_{ij}^{\mu}$.

$$\boldsymbol{f}_{ij}^{\mu} = \{\boldsymbol{f}_{ij}^{\varepsilon}, \nabla_{\boldsymbol{\mathscr{E}}} \boldsymbol{f}_{ij}^{\varepsilon}\} \tag{2.22}$$

However, the direct definition of $\nabla_{\boldsymbol{\mathscr{E}}} \boldsymbol{f}_{ij}^{\varepsilon}$ does not satisfy translational invariance due to the dependence of the Fock matrix elements $\nabla_{\boldsymbol{\mathscr{E}}} F_{pq}$ on the positional operator matrix element $\boldsymbol{r}_{pq}$.

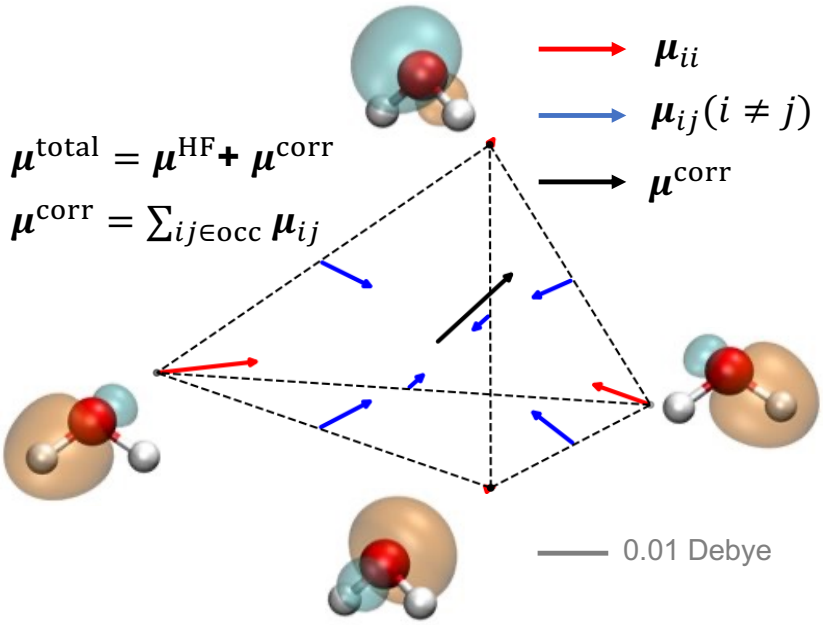


Figure 2.9: Example decomposition of dipole moments as a sum of pairwise MO contributions for a water molecule. The four vertices of the tetrahedron represent the self-interactions of four MOs (indexing as ii), and the six edges connecting the vertices represent the interactions between two MOs (indexing as ij). The pair dipoles $\boldsymbol{\mu}_{ii}$ and $\boldsymbol{\mu}_{ij}$ are shown in red and blue arrows with correct direction and scaling, respectively. Four MOs are also shown next to the corresponding vertex i. The relative length scaling is also shown using a grey line.

$$F_{pq} = h_{pq} + \sum_{k=1}^{n} (2J_{pq}^{k} - K_{pq}^{k}) + \cancel{\boldsymbol{r}_{pq} \cdot \boldsymbol{\mathscr{E}}},$$

$$\nabla_{\boldsymbol{\mathscr{E}}} F_{pq}\Big|_{\boldsymbol{\mathscr{E}}=0} = \nabla_{\boldsymbol{\mathscr{E}}} \left(h_{pq} + \sum_{k=1}^{n} (2J_{pq}^{k} - K_{pq}^{k}) \right) \Big|_{\boldsymbol{\mathscr{E}}=0} + \cancel{\nabla_{\boldsymbol{\mathscr{E}}} \boldsymbol{r}_{pq} \cdot \boldsymbol{\mathscr{E}}\Big|_{\boldsymbol{\mathscr{E}}=0}} + \cancel{\boldsymbol{r}_{pq}}.$$

where h_{pq} are the one-electron Hamiltonian matrix elements, J_{pq}^{k} and K_{pq}^{k} are Coloumb and exchange matrix elements of the kth orbital, and $\boldsymbol{r}_{pq}$ are the position operator matrix elements. The $\boldsymbol{r}_{pq} \cdot \boldsymbol{\mathscr{E}}$ and $\nabla_{\boldsymbol{\mathscr{E}}} \boldsymbol{r}_{pq} \cdot \boldsymbol{\mathscr{E}}$ terms vanish when $\boldsymbol{\mathscr{E}} = \mathbf{0}$. However, the existence of $\boldsymbol{r}_{pq}$ term in Eq. 2.23 results in a non-translataional invariant dipole feature design. Therefore, a redefinition of the derivatives of Fock matrix elements is adapted by subtracting the $\boldsymbol{r}_{pq}$ term to make dipole feature vector (Eq. 2.22) translataional invariant.

GPR with rotationally equivariant derivative kernel for ML response

Several previous studies have recognized the importance of rotational equvariance for ML framework to efficiently learn molecular dipole moments. [107, 110] For any rotational operator $\hat{U}$, a function $\boldsymbol{g} : \boldsymbol{x} \rightarrow \boldsymbol{g}(\boldsymbol{x})$ is rotationally equivariant if

$$\boldsymbol{g}(\hat{U}\boldsymbol{x}) = \hat{U}\boldsymbol{g}(\boldsymbol{x}). \tag{2.23}$$

From a physics perspective, the rotational equivariance guarantees that the predicted property will rotate correspondingly with the rotation of the system. Therefore, the rotational equivariance property is required for any tensorial molecular properties, such as force, dipole moment, and polarizability. In addition, the molecular energy is rotationally invariant, i.e., remains constant with the rotation of the system. The conditions for energy and dipole models can be formulated as follows:

$$\begin{cases} \varepsilon^{\mathrm{ML}}[\boldsymbol{f}_{ij}^{\varepsilon}] = \varepsilon^{\mathrm{ML}}[\hat{U}\boldsymbol{f}_{ij}^{\varepsilon}] \\ \boldsymbol{\mu}^{\mathrm{ML}}[\hat{U}\boldsymbol{f}_{ij}^{\mu}] = \hat{U}\boldsymbol{\mu}^{\mathrm{ML}}[\boldsymbol{f}_{ij}^{\mu}]. \end{cases} \tag{2.24}$$

For the MOB features, when applying a rotation operator $\hat{U}$, the energy and dipole feature sets satisfy the relationship $\hat{U}\boldsymbol{f}_{ij}^{\varepsilon} = \boldsymbol{f}_{ij}^{\varepsilon}$ and $\hat{U}\boldsymbol{f}_{ij}^{\mu} = \{\boldsymbol{f}_{ij}^{\varepsilon}, U\nabla_{\mathscr{E}}\boldsymbol{f}_{ij}^{\varepsilon}\}$, respectively, where U is the matrix representation of $\hat{U}$. Therefore, the MOB-ML energy model is always rotationally invariant for any regressor. However, it remains challenging and requires a special ML algorithm design to make the MOB-ML dipole model rotationally equivariant for a greater learning efficiency.

Assuming the training energy and dipole sets are $(\mathbf{X}_{\varepsilon}, \mathbf{y}_{\varepsilon}) = \{\boldsymbol{f}_{ij}^{\epsilon}, \varepsilon_{ij}\}$ and $(\mathbf{X}_{\mu}, \mathbf{y}_{\mu}) = \{\boldsymbol{f}_{ij}^{\mu}, \mu_{ij}\}$, respectively, we introduce a rotationally equivariant derivative kernel that could accurately learn dipole moment and energy separately or simultaneously.

A single-task energy model could be directly learnt using the naive GPR in Eq. 2.8 with the prior distribution

$$\varepsilon^{\mathrm{ML}}(\mathbf{X}_{\varepsilon}) \sim N(0, K_{\varepsilon}(\mathbf{X}_{\varepsilon}, \mathbf{X}_{\varepsilon})). \tag{2.25}$$

Since the derivative of a GP is also a GP [128, 154], dipole moments can be regressed by GPR with the prior distribution of $\nabla_{\mathscr{E}}\varepsilon^{\mathrm{ML}}(\mathbf{X}_{\varepsilon})$:

$$\boldsymbol{\mu}^{\mathrm{ML}}(\mathbf{X}_{\mu}) = -\nabla_{\mathscr{E}}\varepsilon^{\mathrm{ML}}(\mathbf{X}_{\varepsilon}) \sim N(0, K_{\mu}(\mathbf{X}_{\mu}, \mathbf{X}_{\mu})). \tag{2.26}$$

μ^{ML} and the corresponding kernel matrix K_{μ} could be written as follows:

$$\begin{aligned} \boldsymbol{\mu}^{\mathrm{ML}}(\mathbf{x}_{\mu}) &= -\nabla_{\mathscr{E}}\mathbf{x}_{\varepsilon} \cdot \nabla_{\mathbf{x}_{\varepsilon}}\varepsilon^{\mathrm{ML}}(\mathbf{x}_{\varepsilon}), \\ K_{\mu}(\mathbf{x}_{\mu}, \mathbf{x}'_{\mu}) &= \nabla_{\mathscr{E}}\mathbf{x}_{\varepsilon}\nabla_{\mathscr{E}}\mathbf{x}'_{\varepsilon} \cdot K_{\varepsilon}^{(1,2)}(\mathbf{x}_{\varepsilon}, \mathbf{x}'_{\varepsilon}), \end{aligned} \tag{2.27}$$

where we use symbol $\mathbf{x}$ to represent general input features, and the superscripts of K_ε represent the derivatives to the arguments, e.g. $K^{(1,2)}(x_1,x_2) = \nabla_{x_1}\nabla_{x_2}K(x_1,x_2)$. Since $\nabla_{\mathscr{E}}\mathbf{X}_\varepsilon$ will produce the derivative terms $\{\nabla_{\mathscr{E}}\boldsymbol{f}_{ij}^{\varepsilon}\}$, including these derivatives in the dipole feature set is necessary to model $\boldsymbol{\mu}^{\mathrm{ML}}$. This mathematical deduction agrees with the physical intuition discussed in Sec. 2.6.

By using the Gaussian likelihood $\mathbf{y}_\mu \sim N(\boldsymbol{\mu}^{\mathrm{ML}},\Sigma_\mu)$ with $\Sigma_\mu = \sigma_\mu^2 I$, for a set of test points $\mathbf{X}_\mu^\star$, the prediction mean and variance of the derivative kernel can be evaluated using Eq. 2.8 as

$$\begin{aligned}\mathbb{E}[\boldsymbol{\mu}^{\mathrm{ML}}(\mathbf{X}_\mu^\star)] =& K_\mu(\mathbf{X}_\mu^\star,\mathbf{X}_\mu)(K_\mu(\mathbf{X}_\mu,\mathbf{X}_\mu)+\Sigma_\mu)^{-1}\mathbf{y}_\mu \\ \mathrm{Var}[\boldsymbol{\mu}^{\mathrm{ML}}(\mathbf{X}_\mu^\star)] =& K_\mu(\mathbf{X}_\mu^\star,\mathbf{X}_\mu)(K_\mu(\mathbf{X}_\mu,\mathbf{X}_\mu)+\Sigma_\mu)^{-1} \\ & K_\mu(\mathbf{X}_\mu,\mathbf{X}_\mu^\star).\end{aligned} \tag{2.28}$$

The rotationally equivariant derivative kernel can be generalized to the multi-task learning of $\boldsymbol{\varepsilon}^{\mathrm{ML}}$ and $\boldsymbol{\mu}^{\mathrm{ML}}$ simultaneously. In such case, their joint distribution is also a GP with the predictive mean and variance as

$$\begin{aligned}\mathbb{E}\begin{bmatrix}\boldsymbol{\varepsilon}^{\mathrm{ML}}(\mathbf{X}_\mu^\star)\\ \boldsymbol{\mu}^{\mathrm{ML}}(\mathbf{X}_\mu^\star)\end{bmatrix} =& K_{\varepsilon\mu}(\mathbf{X}_\mu^\star,\mathbf{X}_\mu)(K_{\varepsilon\mu}(\mathbf{X}_\mu,\mathbf{X}_\mu)+\Sigma_{\varepsilon\mu})^{-1}\mathbf{y}_{\varepsilon\mu} \\ \mathrm{Var}\begin{bmatrix}\boldsymbol{\varepsilon}^{\mathrm{ML}}(\mathbf{X}_\mu^\star)\\ \boldsymbol{\mu}^{\mathrm{ML}}(\mathbf{X}_\mu^\star)\end{bmatrix} =& K_{\varepsilon\mu}(\mathbf{X}_\mu^\star,\mathbf{X}_\mu)(K_{\varepsilon\mu}(\mathbf{X}_\mu,\mathbf{X}_\mu)+\Sigma_{\varepsilon\mu})^{-1} \\ & K_{\varepsilon\mu}(\mathbf{X}_\mu,\mathbf{X}_\mu^\star),\end{aligned} \tag{2.29}$$

where

$$\begin{aligned}\mathbf{y}_{\varepsilon\mu} =& \begin{bmatrix}\mathbf{y}_\varepsilon\\ \mathbf{y}_\mu\end{bmatrix},\ \Sigma_{\varepsilon\mu} = \begin{bmatrix}\sigma_\varepsilon^2 I & 0\\ 0 & \sigma_\mu^2 I\end{bmatrix}, \\ K_{\varepsilon\mu}(\mathbf{x}_\mu,\mathbf{x}_\mu') =& \begin{bmatrix}K_\varepsilon(\mathbf{x}_\varepsilon,\mathbf{x}_\varepsilon') & \nabla_{\mathscr{E}}\mathbf{x}_\varepsilon'\cdot K_\varepsilon^{(2)}(\mathbf{x}_\varepsilon,\mathbf{x}_\varepsilon')\\ \nabla_{\mathscr{E}}\varepsilon\cdot K_\varepsilon^{(1)}(\mathbf{x}_\varepsilon,\mathbf{x}_\varepsilon') & \nabla_{\mathscr{E}}\mathbf{x}_\varepsilon\nabla_{\mathscr{E}}\mathbf{x}_\varepsilon'\cdot K_\varepsilon^{(1,2)}(\mathbf{x}_\varepsilon,\mathbf{x}_\varepsilon')\end{bmatrix}.\end{aligned} \tag{2.30}$$

We note that there might be other rotationally equivariant GPR frameworks that provide models with similar accuracy, they might not ensure the learnt dipole model is a derivative of the energy model.

As an analogy to the response in physics, it is desirable for the ML model of h and its response property model $\boldsymbol{g}$ to satisfy Eq. 2.31, termed as "ML response relationship".

$$\boldsymbol{g}(\boldsymbol{x}) = \nabla h(\boldsymbol{x}). \tag{2.31}$$

This relationship requires model $\boldsymbol{g}$ to be conservative (or curl-free), i.e., $\nabla \times \boldsymbol{g} = \mathbf{0}$ [155, 156]. In this study, we apply this physically driven rotationally equivariant derivative kernel formalism to satisfy this response relationship at any electric field. Both the single-task and multi-task GPR with this derivative kernel satisfy the rotational equivariance (Eq. 2.24). The proof in given in Appendix 6.2. Without any specification, we adapt the single-task framework of GPR with rotationally equivariant derivative kernel for all the following training. The performance comparisons of single-task and multi-task models are demonstrated in Sec. 2.7.

2.7 Extension to learning linear response properties: Results

The accuracy and efficiency of the proposed MOB-ML approach are tested on various benchmark systems, including water, fourteen small molecules, the QM9 benchmark dataset [118], and four series of peptides [106]. Since the QM9 dataset contains more than a hundred thousand molecules, we combine the local GPR with the Gaussian mixture model (GMM) unsupervised clustering (GMM/GPR) technique introduced in Ref. [115], and the AltBBMM technique introduced in Sec. 2.5.

Computational details

The water with random field dataset is generated by computing the data of each structure in the water dataset at finite electric field ranging from 0 to 0.05 a.u. with random directions. The structures of the QM9 dataset and the four different series of peptides are directly obtained from Ref. [118] and Ref. [106], respectively. The derivatives of the pair energies with respect to electric fields are implemented and calculated in the Molpro package following Ref. [157].

For all the datasets, we separately learn the energies and dipole moments using the energy feature set and dipole feature set, respectively. For the water and the small molecule datasets, the results for multi-task models, i.e., dipole + energy models, that learn both tasks simultaneously are also included for comparison. Table 2.4 summarizes the usage of energy and dipole features in this work. We unsupervisedly cluster the MOs represented by energy features instead of separately clustering energy and dipole feature space, i.e., the clustering models are identical for the energy and dipole learning with the same training sets. Feature selection is performed before all the models on energy labels using the random forest regression implementation in the SCIKIT-LEARN [158] package following the protocol in Ref. [94]. For the dipole learning, we use the selected energy features and their derivatives as

the selected dipole features.

Table 2.4: Usages of different feature sets in different learning models.

ML model	Feature set	Learning task
Clustering (GMM)	Energy features	All energy & dipole models
Regression (GPR)	Energy features	Energy (single-task)
	Dipole features	Dipole (single-task) Dipole + energy (multi-task)

We apply the AltBBMM algorithm as the default GP regressor, and reimplement the GMM with full covariance matrix following SCIKIT-LEARN using CUPY [159] to enable multi-GPU training. The implementations for both algorithms are available online at https://github.com/SUSYUSTC/BBMM.git. For all GPR and GMM/GPR, we employ the Matérn 5/2 kernel with white noise regularization [128]. The parameters used in training for GPR and GMM/GPR are further discussed in Supporting information.

All the results for water and small molecule datasets are collected with GPR without clustering. The local GPR with GMM unsupervised clustering is applied to scale the MOB-ML training in the QM9 dataset. In this work, we follow the same clustering protocol introduced in Ref. [115] to generate the GMM models. The GMM model is initialized by K-means clustering, and its number of clusters is automatically determined by minimizing the Bayesian information criterion. GMM could not be performed to cluster 50,000 and 110,000 QM9 molecules due to limited memory, and we thus apply the GMM model trained on 20000 QM9 molecules to approximate the clustering results of these two models. To reduce the learning costs, we also apply the same capping strategy described in Ref. [95]. For the clusters containing a large number of points, we randomly select training points with the capping size defined ahead to regress these local GPR. The capping size is 1,000,000 and 300,000 pairs for dipole and energy learning, respectively.

Dipole moment learning for small molecules via MOB-ML

To demonstrate the ability to learn dipole moments using the MOB representations, we first test the prediction accuracies of MOB-ML on water and other small molecules. Figure 2.10 displays the mean absolute errors (MAEs) of dipole moments ($|\boldsymbol{\mu}|$) and molecular energies E for water learnt by single-task and multi-task models. The sizes of the training set are varied from 2 to 100 geometries, and the test set is composed of 100 geometries not included in any training sets. MAEs at

the same order of reference data accuracy for molecular energies are achieved by training on 100 geometries using energy labels (5.54×10^{-5} kcal/mol) and dipole combined with energy labels (1.84×10^{-5} kcal/mol). Across all the training sizes, single-task and multi-task models provide similar accuracies for molecular energies, but dipole models provide much better dipole predictions than the dipole + energy model.

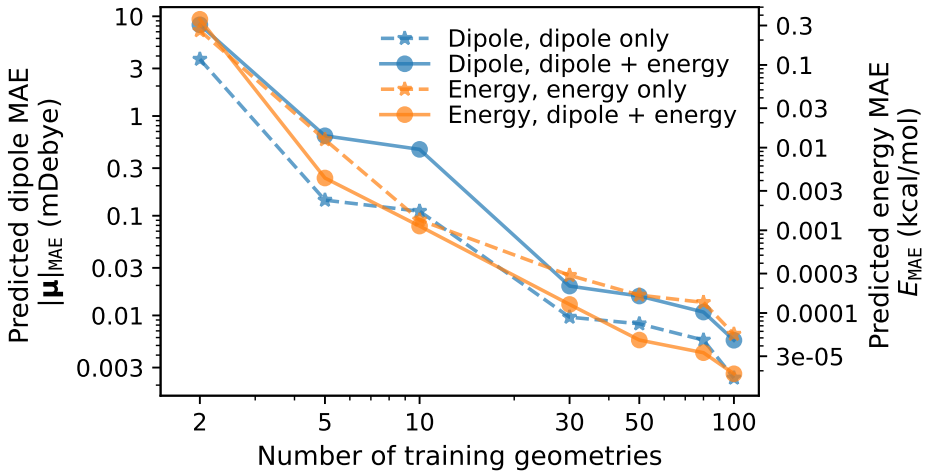


Figure 2.10: Prediction accuracies of dipole moments and energies of water using a reference theory of MP2/cc-pVTZ at zero electric field. The predicted MAEs are plotted versus the number of training geometries on a log-log scale ("learning curves"). The MOB-ML models for dipole moments and energies are constructed by training on the labels individually (dipole model and energy model) or simultaneously using multi-task learning (dipole + energy model). The primary and secondary y-axises represent the prediction MAE of dipole moments $|\boldsymbol{\mu}|_{\mathrm{err}}$ in milli-Debye (mDebye) and the prediction MAE of energy in kcal/mol, separately.

Table 2.5 lists the MAEs of dipole moments and total energies from single-task and multi-task models training on 50 geometries and testing on different 50 geometries for small molecules with different sizes. MOB-ML provides very accurate predictions for all the test molecules. Comparing the results of molecules with different molecular sizes sharing similar MO properties, such as CH_4, C_2H_6, and C_3H_8, it is clear that the larger molecules have much bigger errors than the smaller ones for both dipole and energy owing to the increasing number of pairwise contributions to the final result. The total errors should scale linearly with the increase of molecular size by summing up predicted pairwise contribution with Gaussian distributed pairwise errors. For the systems that share similar numbers of MOs, such as C_2H_4 and C_2H_6, the more rigid molecule (C_2H_4) is easier to learn for both dipole and energy.

Table 2.5: Predicted error of the MOB-ML training on dipole only or dipole and energy together on different small molecules. All the models are trained on 50 configurations and tested on the remaining 50 configurations using AltBBMM as the regressor.

System	Dipole only	Energy only	Dipole + Energy	
	$\lvert\boldsymbol{\mu}\rvert_{\text{MAE}}$ (mDebye)	E_{MAE} (kcal/mol)	$\lvert\boldsymbol{\mu}\rvert_{\text{MAE}}$ (mDebye)	E_{MAE} (kcal/mol)
CH_4	0.024	0.0005	0.040	0.003
NH_3	0.030	0.0007	0.080	0.002
HF	0.00008	0.00005	0.0001	0.00007
CO	0.004	0.00009	0.010	0.001
CH_2O	0.030	0.0005	0.049	0.003
HCN	0.052	0.0002	0.137	0.005
C_2H_4	0.151	0.004	0.260	0.011
C_2H_6	0.339	0.014	0.482	0.028
CH_3OH	1.143	0.018	1.430	0.020
CH_2F_2	2.180	0.080	3.529	0.060
C_3H_8	0.565	0.035	0.998	0.047
n-Butane	0.912	0.026	1.842	0.076
Isobutane	0.812	0.047	1.883	0.071
C_6H_6	2.433	0.039	3.403	0.053

Learning dipole moments and molecular energies simultaneously do not always provide better prediction accuracies of both dipole and energy for most of the molecules (12 out of total 14 molecules) and needs higher computational costs since it trains more points within a model. This observation indicates that dipole vectors and energies of each pair of MOs might vary independently as functions of MOB features, and therefore no mutual supervision could be provided by multi-task learning.

According to Ref.[103], an energy model could directly provide accurate force predictions by using the ML response relationship. Here we show that it is not expected to enhance the learning efficiency of the dipole model by adding energy data at a zero electric field. Figure 2.11 could help explain this observation by using the water molecule as an example. The relative water MP2/cc-pVTZ total energy is plotted as a function of one of the O-H bond lengths and the strength of the external electric field along the bond direction. We fix the bond angle and only change one of the O-H bond lengths (red curve) with $\mathscr{E} = 0$ to facilitate the understanding. The training set can be treated as samples of this simplified potential. We note that

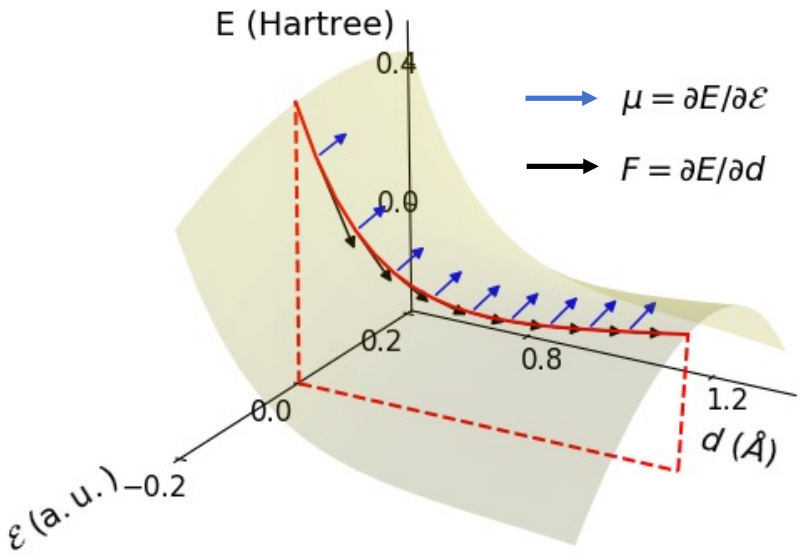


Figure 2.11: Relative MP2/cc-pVTZ total energy of a water molecule as a function of one of its O-H bond lengths d and the strength of applied electric field $\mathscr{E}$ along the bond direction. The bond angle and the other O-H bond length are fixed to the equilibrium value. The relative energy is shifted to 0 at equilibrium geometry with $\mathscr{E} = 0$. The red curve corresponds to the energy surface at $\mathscr{E} = 0$. The dashed red lines represent the projection of the red curve to d-$\mathscr{E}$ plane. The blue and black arrows illustrate the direction of the derivatives to get dipole moments and forces.

all the bond lengths and angles vary in the actual water dataset. From these data, the energy information can then directly infer the force information (black arrows) since the derivative with respect to bond length d could be estimated as differences between two sampled training points. However, since there is no change along the axis of the electric field $\mathscr{E}$, no estimated information of dipole moments (blue arrows) is available from the training set. We additionally show the results of adding data computed with random finite electric fields ranging from 0 to 0.05 a.u. (water with finite electric field dataset) in Fig. 2.12. The multi-task learning models (best MAE=3.70×10^{-4} kcal/mol) provide around twice better accuracies than the single-task energy models (best MAE=8.20×10^{-4} kcal/mol). The single-task and multi-task dipole models can reach 0.030 and 0.051 mDebye, respectively. The single-task energy models could directly provide reasonably good predictions (best MAE=0.152 mDebye) for dipole moments.

As shown in Fig. 2.12, multi-task learning could always improve the energy model but only outperform the single-task dipole model slightly with few training data. There are 400 points in the multi-task learning model (300 dipole + 100 energy

points) when training on 100 geometries. Comparing with the single-task energy and dipole models, 300 additional dipole points carry more information than 100 additional energy points in the multi-task model. Therefore, it is more beneficial to have additional supervision for energy learning than dipole learning. Furthermore, the inclusion of new tasks in the multi-task models further complicates the learning process and might even lead to deterioration in training if the original information contents are enough. Since dipole MOB features have already carried extremely high-quality information to learn dipole moments when training on 30 geometries (90 points), the few newly added energy data bring in more regression hardness than the supervision benefits. This size-dependence of performance of multi-task GPR agrees with the conclusion of Ref. [160] that a large training size might harm the performance of multi-task GPR due to the lower information quality included by the additional task, which is also supported by the various numerical experiments in Ref. [160]. Therefore, the dipole + energy model for these small molecules cannot provide better accuracies than only training on dipoles.

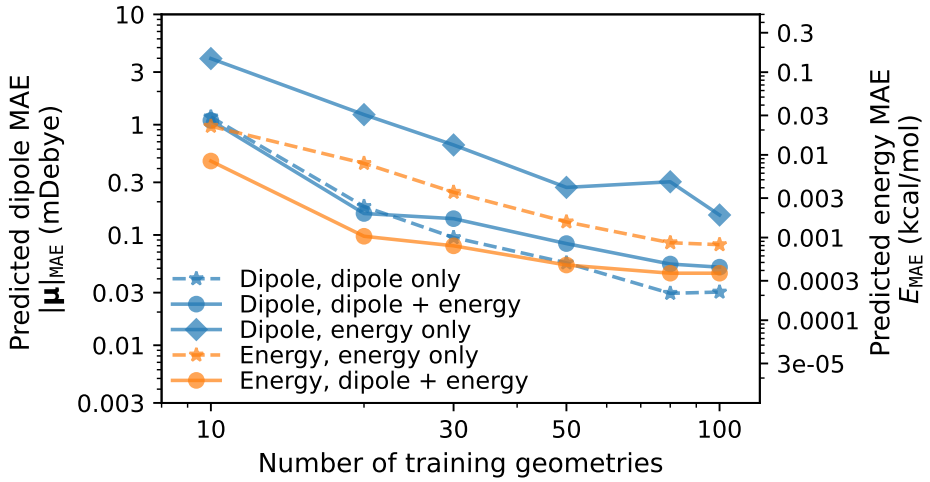


Figure 2.12: Prediction accuracies of dipole moments and energies of water with finite electric field using a reference theory of MP2/cc-pVTZ. Compared with the types of results included in Fig. 2.10, the results of predicting dipole moments directly using the model trained on energy labels only are additionally included to support the explanation in Fig. 2.11. The primary and secondary y-axises represent the prediction MAE of dipole moments $|\boldsymbol{\mu}|_{\mathrm{err}}$ in milli-Debye (mDebye) and the prediction MAE of energy in kcal/mol, separately.

MOB-ML for dipole moments and energies of organic molecules in QM9

In our previous studies, we have illustrated the excellent accuracy and transferability of MOB-ML to learn molecular energies using two thermalized organic

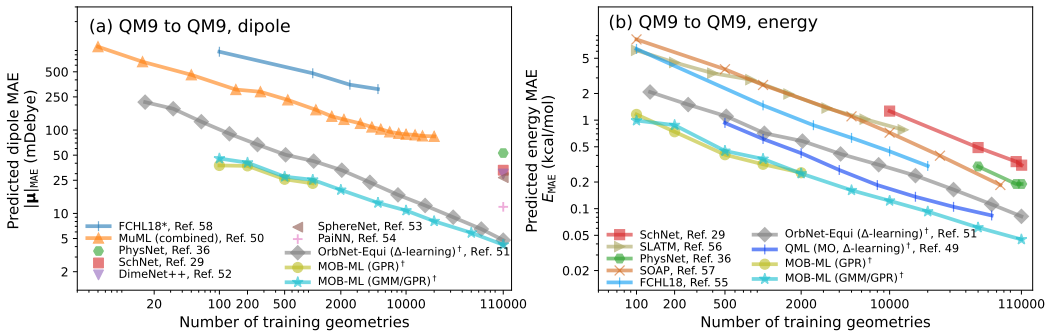


Figure 2.13: Prediction accuracies of dipole moments and energies of test QM9 molecules trained on QM9 molecules using MOB-ML. The single-task MOB-ML models for (a) dipole moments and (b) energies are constructed by training on the labels individually. For the methods that only report the prediction errors of the models training on 110,000 molecules, we plot their results as scatters with different shapes. The learning curves of other literature methods trained on QM9 properties computed using B3LYP/6-31G(2df,p) level of theory [118] are also plotted for comparison. The ones marked with daggers are the approaches using Δ-learning approaches. The prediction MAEs of dipole moments $|\boldsymbol{\mu}|_{\mathrm{err}}$ and energy are in milli-Debye (mDebye) and kcal/mol units, respectively.

molecule datasets, i.e., QM7b-T and GDB-13-T [94, 95, 102, 115]. In this study, we systematically examine the learning performance of MOB-ML for both the dipole and energy using the benchmark organic chemistry dataset QM9[118], which contains optimized structures of 133,885 molecules with up to nine heavy atoms (HAs) of C, O, N, and F. QM9 is a standard benchmark dataset that has been assessed in many different literature studies. [25, 87, 98, 105–114] Figure 2.13 displays the predicted MAEs for dipole moments (in mDebye) and energies (in kcal/mol) as functions of number of training geometries on a log-log scale (learning curves). Since our GPR regression, AltBBMM, can only train at most 1 million points, we collect the results of MOB-ML (GPR) up to training on 1,000 and 2,000 dipole moments and molecular energies, respectively. The application of GMM/GPR scales the training of dipole moments and energy to the same training size (at most 110,000 QM9 molecules) as the literature models. The test sets of MOB-ML approaches remain the same across the entire learning curve with a size of 11,843 molecules. Different literature approaches computed at the B3LYP/6-31G(2df,p) level of theory are included for comparison. Since this study is the first work providing MP2 energies and dipoles for the QM9 dataset, there are lack of direct literature comparisons testing on the same reference data.

For dipole moments in Fig. 2.13a, we compare the results from MOB-ML with

those from the state-of-the-art literature (SOTA) methods, including FCHL18* [114], MuML (combined) [106], PhysNet [25], SchNet [87], DimeNet++ [108], SphereNet [109], PaiNN [110], and OrbNet-Equi [107]. All the compared literature methods provide the predictions at B3LYP/6-31G(2df,p) level of theory. OrbNet-Equi (Δ-learning) and QML (MO, Δ-learning) use GFN-xTB and HF/STO-3G inputs, respectively, and all the rest of shown approaches use geometric inputs. MOB-ML could also be treated as a Δ-learning approach that uses HF/cc-pVTZ inputs. It is clear that both MOB-ML regressed by AltBBMM (MOB-ML(GPR)) and MOB-ML regressed by GMM clustering with local AltBBMM (MOB-ML(GMM/GPR)) outcompete other literature methods in the low-data learning regime (training set smaller than 2,000 molecules). MOB-ML/GPR and MOB-ML(GMM/GPR) achieve similar MAEs of 37.61 and 45.58 mDebye, respectively, by training on only 100 molecules. This indicates that the introduction of unsupervised clustering does not affect the accuracy of MOB-ML in learning dipole moments. Meanwhile, the second-best OrbNet-Equi learnt by Δ-learning (OrbNet-Equi (Δ-learning)) [107] requires 1,024 molecules to reach the same level of accuracy. However, OrbNet-Equi (Δ-learning) models are improved faster with an increasing number of data points with the deepest slope of the learning curve relative to other methods. When there are enough examples in the training set (110,000 training molecules), OrbNet-Equi (Δ-learning) could reach a slightly worse MAE of 4.78 mDebye than MOB-ML(GMM/GPR) (4.21 mDebye).

Similarly, the prediction errors of energies from MOB-ML approaches are compared with SchNet [87], SLATM [112], PhysNet [25], SOAP [113], FCHL18 [111], OrbNet-Equi [107], and QML [105] in Fig. 2.13b. MOB-ML (GPR) and MOB-ML (GMM/GPR) still provide the best sets of results across all the training sizes. MOB-ML (GMM/GPR) achieves accuracies of 0.99 kcal/mol and 0.045 kcal/mol with only 100 and 110,000 training molecules, respectively. Both numbers are the current best in this field. QML with an orbital based features and (Δ-learning) (QML (MO, Δ-learning)) and OrbNet-Equi (Δ-learning) are other two most accurate approaches.

The top approaches to predict dipole moments and energies, i.e., MOB-ML, OrbNet-Equi (Δ-learning), and QML (MO, Δ-learning), are further compared here from a theoretical perspective. To achieve the best accuracy, all three approaches apply the idea of "Δ-learning" by predicting the differences between low-level and high-level theories instead of directly predicting. In addition, all three approaches are orbital-

based ML approaches that adapt features related to energy matrix elements, and these features are considered to contain high-quality quantum-level information to make the ML map easier. Although the three approaches share several similarities, their differences might explain their prediction accuracy differences. Firstly, MOB-ML predicts the results from wavefunction theories using HF computations with the same basis, while OrbNet-Equi and QML (MO) predict the results from DFT using GFN-xTB and minimal basis HF computations, respectively. Since HF computed with cc-pVTZ basis set is more expensive and contains more accurate information about the orbitals than GFN-xTB and minimal basis HF, MOB-ML is more expensive in evaluation than the other two methods. MOB-ML explicitly decomposes the differences between low- and high-level theory results onto MOs and learns these pairwise contributions using GPR, while OrbNet-Equi and QML (MO) directly learn these differences by implicitly decomposing them to each kernel in kernel ridge regression (KRR) or nodes in graph neural network (GNN) by carefully designing the ML frameworks. This explicit decomposition brings an accuracy gain to MOB-ML but limits the applications of MOB-ML to decomposable properties. On the other hand, QML (MO) and OrbNet-Equi are able to predict more different molecular properties.

Timing and learning efficiency of GMM/AltBBMM with rotationally equivariant derivative kernels

Figure 2.14 displays the accuracy improvements of test QM9 molecules as functions of training costs using MOB-ML (GPR) and MOB-ML (GMM/GPR) for dipole moments and energies. These models are collected on 8 NVIDIA Tesla V100-SXM2-32GB GPUs. Since each local GPR could be regressed independently on different GPUs, GMM/GPR is a highly-parallelized approach with excellent multi-GPU speedups. For both dipole and energy learning, it is clear that GMM/GPR provides much lower training costs compared with learning without clustering. Across all the training sizes that we could train directly with GPR, GMM/GPR could achieve more than 68.5 and 21.4 times speedups for dipole and energy learning, respectively, without loss of accuracy and transferability. For example, the best GMM/GPR energy model training on 110,000 molecules only takes 20.7 hrs, but it is over seven times more expensive than the best dipole model (146.5 hrs). This is because the training points of dipole moment are three times larger than ones of energy for each molecule.

Although MOB-ML provides accurate post-HF predictions for these peptides, the

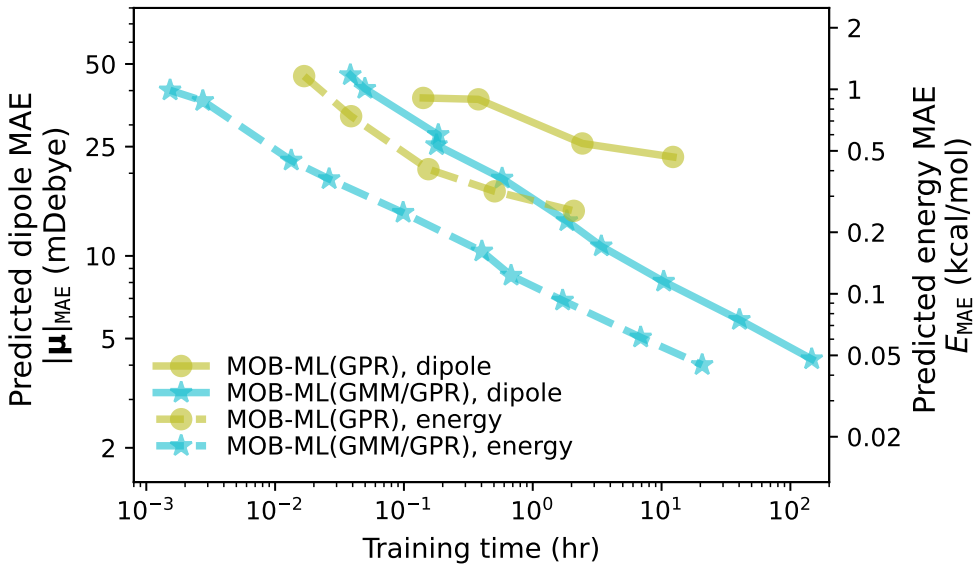


Figure 2.14: Learning costs of MOB-ML approaches for QM9 models shown in Fig. 2.13. The prediction MAEs for dipole moments and energies of QM9 test molecules are plotted as functions training hours using 8 GPUs on log-log scales. The primary axis (left axis) labels the MAEs of the dipole moments, and the secondary axis (right axis) shows the MAEs of energies. Results for dipole and energy models are plotted in solid and dashed lines, respectively. Different colors represent different learning protocols and match the ones in Fig. 2.13 The timing results only include the ML training costs and do not include HF costs.

applications of MOB-ML to predict large molecular systems are limited by the relatively high costs of HF computations ($O(N_{orb}^3 N_{occ})$ with density-fitting). Table 2.6 lists the training and evaluation time of MOB-ML (GMM/GPR) and literature approaches (with timing reported) on the QM9 dataset. The average evaluation timings of the best MOB-ML are significantly slower than other listed approaches. To reduce the evaluation costs of MOB-ML, predicting the molecular properties using the information from HF with smaller basis sets or GFN-xTB is a potential future research direction.

Predictions of dipole moments of four challenge cases

To illustrate the accuracy of MOB-ML in the actual biochemical systems, we further assess the prediction accuracy of the best MOB-ML (GMM/GPR) model on four different sets of peptides, termed as "challenging dataset". This challenging dataset is firstly introduced in Veit et al. [106], and in this study, we also included the literature results predicted by MuML model at B3LYP/daDZ level of theory [106]. B3LYP usually provides big errors for large organic molecules[161, 162], and Elst-

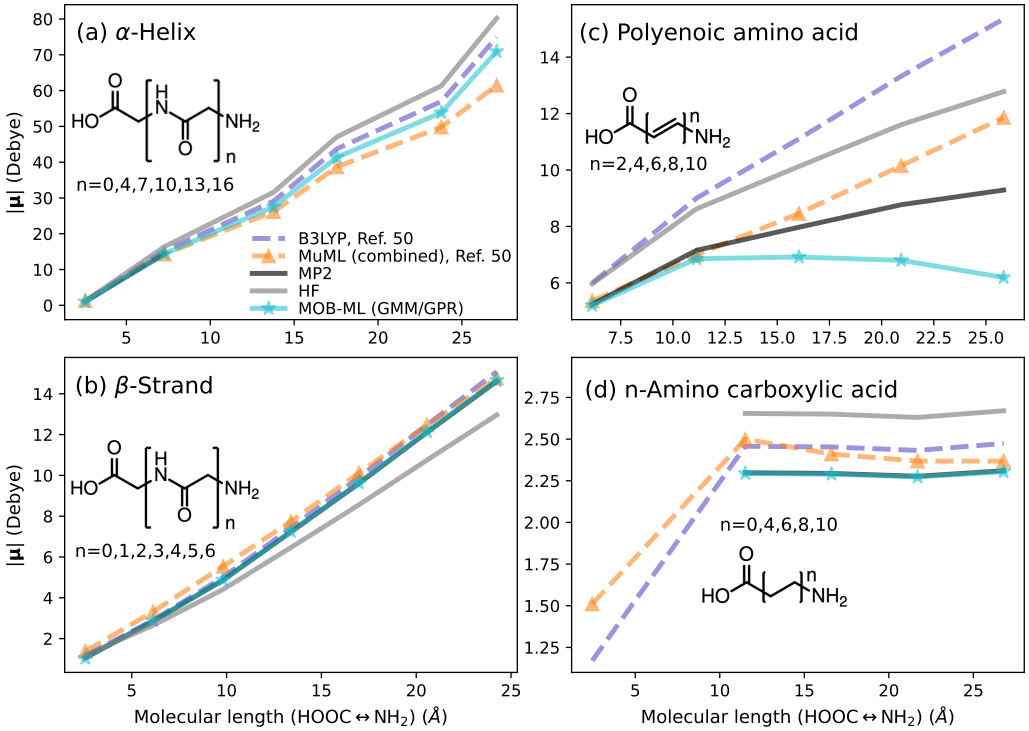


Figure 2.15: Dipole moment predictions for series of (a) α-helix, (b) β-strand, (c) polyenoic amino acid, and (d) n-amino carboxylic acid using MOB-ML and MuML. The dipole moments of different molecules are plotted versus the chain length. The MOB-ML dipole moments are computed from the best MOB-ML(GMM/GPR) trained on 110,000 QM9 molecules. The results of MuML (combined) are predicted from the MuML model trained on 5,400 QM7b molecules and extracted from Fig. 6 in Ref. [106]. The reference dipole moments computed using MP2/cc-pVTZ (for MOB-ML) and B3LYP/daDZ (for MuML) are shown in the plots, and the HF dipole moments are also provided for further discussion. In (a), the molecules with $n = 7, 10, 13, 16$ cannot be computed by MP2/cc-pVTZ, and results from other theories and ML models are shown.

Table 2.6: Training and evaluation time of different literature methods with different reference theories on the QM9 dataset. Other training information is also listed for comparison. The evaluation time is reported as the average evaluation time for each molecule.

ML approach	Task	Training size	Training time (hr)	Evaluation time (s)
MOB-ML	Energy	110000	20.7 (8 GPUs)	44.78 (1 core CPU)
	Dipole		146.5 (8 GPUs)	96.29 (1 core CPU)
QML [105]	Energy	110000	–	2.376 (24 core CPUs)
MuML[106]	Dipole	20000	1 (24 core CPUs)	0.12 (24 core CPUs)

ner et al. [163] additionally stated the systematic errors of B3LYP on glycine can be accumulated for longer peptides. A higher level of theory is needed to achieve more reliable results for peptides. In this study, we apply the MP2/cc-pVTZ MOB-ML model trained on QM9 to predict the four sets of peptides. All the true and predicted dipole moments from different theories and ML models are plotted as a function of the chain length in Fig. 2.15. Table S4 in SI summarizes the predicted and true dipole moments and energies using the best QM9 energy model (GMM/GPR trained on 110,000 molecules). The corresponding MP2, HF and MOB-ML evaluation time are also shown in Table S3. We note that MP2/cc-pVTZ is still an affordable theory for the QM9 benchmark dataset, but it is nearly impossible to obtain MP2 energy for the molecules with more than 30 heavy atoms, for instance, α-helix molecules with $n = 7, 10, 13, 16$ in the challenging dataset. Therefore, no true MP2/cc-pVTZ results are provided for these large α-helix molecules; meanwhile, MOB-ML provides reasonable dipole moment predictions for these molecules.

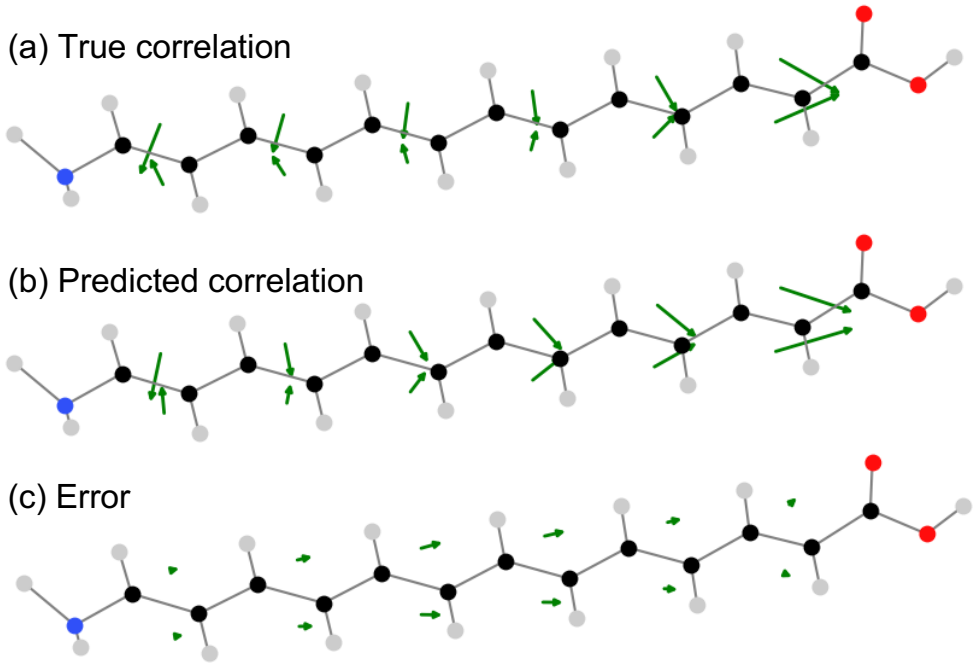


Figure 2.16: True and predicted correlation part of the dipole contribution of CC double bonds in the polyenoic amino acid with n=6. (a) True MP2/cc-pVTZ correlation part (b) predicted MP2/cc-pVTZ correlation from QM9 110k model (c) error between true and predicted correlation part of the dipole contribution of each CC double bond. The balls represent the atoms and the lines in between represent connections between atoms. The colors are C: black, H: grey, O: red, and N: blue. Since each double bond has two MOs, there are two green arrows on each double bond to represent the corresponding dipoles.

Except the large polyenonic amino acids in Fig. 2.15(c), MOB-ML provides nearly identical predicted dipole moments as MP2 for all other molecules in panels (a), (b), and (d), which indicates that the MOB-ML model for dipole moments has an excellent transferability to large molecules. MuML (combined) model also has a deteriorated accuracy for all the molecules in polyenoic amino acids. In this case, the MOB-ML model over-corrects the results from HF calculation, and the size of the errors increases dramatically with the increasing of sizes. To further investigate the source of the large deviations of MOB-ML, we additionally check the correlation part of the MP2 and predictions from MOB-ML model. Compared with the true MP2 correlation dipole, MOB-ML has errors linearly increased with the increase of the carbon chain of polyenoic amino acids and the numerical values are shown in Table 6.3 in the Appendix. The correlation part of the dipole is on the opposite direction of HF dipole. The major error source is identified to be the systematically wrong predictions for CC double bonds, and Table 6.4 in the Appendix shows the true and predicted average correlation parts of the dipoles for each type of MO using n=6 as an example. Figure 2.16 displays the true and predicted correlation dipoles for all the CC double bond MOs for the polyenoic amino acid with n=6. MOB-ML provides consistently wrong predictions of all the double bond MOs. Compared with the true values, the predictions rotate toward the right on the plot (COOH end). The errors of the CC double bonds closer to two end groups are much smaller than the ones in the middle. With n increases, all the newly added repeating units become the middle CC double bonds and have increasingly larger errors. Therefore, the predictions of MOB-ML on CC double bond MOs are accurate when n is small (n=2 or 4), but the errors increase dramatically with the increase of the molecular size when n is large.

Compared with the results of n-amino carboxylic acid, the much larger true MP2 dipole moments of the polyenoic amino acids can be attributed to the charge transfer across the entire molecule via the conjugation, which results in much larger partial charges on the end groups. This observation is supported by the partial charge study from Ref. [106]. Since the n-amino carboxylic acid molecules lose the double bonds and cannot delocalize the charges to the end groups, they have near-constant dipole moments with increasing molecular length.

2.8 Conclusions

Molecular-orbital-based machine learning (MOB-ML) provides a general framework to learn correlation energies at the cost of molecular orbital generation. In

this work, we introduce three improvements to ensure the transferability to large systems, scale up the maximum training size, and extend MOB-ML to learn general linear response properties, respectively.

First, we demonstrate that preservation of physical symmetries and constraints leads to machine-learning methods with greater learning efficiency and transferability. With the improved feature design presented in the current work, MOB-ML is shown to be highly data-efficient, which is important due to the high computational cost of generating reference correlation energies. Without ever being trained to predict relative energies, MOB-ML provides chemically accurate relative energies for QM7b-T when training on only 0.4% of the QM7b-T molecules. Furthermore, we have demonstrated that MOB-ML is not restricted to the organic chemistry space and that we are able to apply our framework out-of-the box to describe a diverse set transition-metal complexes when training on correlation energies for tens of molecules. Beyond data efficiency, MOB-ML models are shown to be very transferable across chemical space. Such transferability is demonstrated by training a MOB-ML model on QM7b-T and predicting energies for a set of molecules with thirteen heavy atoms (GDB13-T), and the best result for GDB13-T is reported by only training on 3% of QM7b-T.

Second, we introduce the BBMM algorithm and a new alternative implementation, AltBBMM, to perform the GPR training for the MOB-ML method on over a million pair energies. Even though the use of BBMM alone increases our previously attainable training-set size limit over 30 times, our newly introduced AltBBMM implementation improves this further by offering a four-fold speed-up while maintaining high accuracy. With the BBMM and AltBBMM approaches, MOB-ML models can be trained using datasets with over 6500 QM7b-T molecules, yielding the best accuracy for the QM7b-T and GDB-13-T datasets.

Third, we extend the MOB-ML framework to learn pairwise contributions of electron correlation part of dipole moments accurately and transferablely using the information computed from HF calculations. The introduction of rotationally equivariant derivative kernel for GPR leads to efficient and physical modeling of dipole moments by satisfying the significant properties of equivariance and ML response. For water and other small molecules, MOB-ML could provide more accurate predictions for the dipole moments and energies by learning two tasks separately than simultaneously. To generate a universal dipole model and energy model for organic molecules, we combine MOB-ML with the previously introduced GMM and

AltBBMM techniques, and apply it to the QM9 dataset and train the two sets of labels separately using the corresponding GPR training protocols. The final MOB-ML model could achieve accuracies of 4.21 mDebye and 0.045 kcal/mol by learning 110000 QM9 molecules for dipole moments and energies, respectively, and an accuracy of 0.99 kcal/mol by only training on 100 QM9 molecules for energies. Furthermore, MOB-ML shows comparable learnabilities compared with all other literature results for both dipole moments and molecular energies and provides accurate and transferable results to most of the tested peptides with different three-dimensional structures.

Chapter 3

A GENERALIZED CLASS OF STRONGLY STABLE AND DIMENSION-FREE T-RPMD INTEGRATORS

This chapter is adapted from

1. Rosa-Raíces, J. L., Sun, J., Bou-Rabee, N. & Miller Thomas F., I. A generalized class of strongly stable and dimension-free T-RPMD integrators. *The Journal of Chemical Physics* **154,** 024106. ISSN: 0021-9606. https://doi.org/10.1063/5.0036954 (Jan. 2021).
   Contribution: J.S. conceptualized the project, formulated the theory, and contributed to writing the article.

Recent work shows that strong stability and dimensionality freedom are essential for robust numerical integration of thermostatted ring-polymer molecular dynamics (T-RPMD) and path-integral molecular dynamics (PIMD), without which standard integrators exhibit non-ergodicity and other pathologies [J. Chem. Phys. **151**, 124103 (2019); J. Chem. Phys. **152**, 104102 (2020)]. In particular, the BCOCB scheme, obtained via Cayley modification of the standard BAOAB scheme, features a simple reparametrization of the free ring-polymer sub-step that confers strong stability and dimensionality freedom and has been shown to yield excellent numerical accuracy in condensed-phase systems with large time-steps. Here, we introduce a broader class of T-RPMD numerical integrators that exhibit strong stability and dimensionality freedom, irrespective of the Ornstein–Uhlenbeck friction schedule. In addition to considering equilibrium accuracy and time-step stability as in previous work, we evaluate the integrators on the basis of their rates of convergence to equilibrium and their efficiency at evaluating equilibrium expectation values. Within the generalized class, we find BCOCB to be superior with respect to accuracy and efficiency for various configuration-dependent observables, although other integrators within the generalized class perform better for velocity-dependent quantities. Extensive numerical evidence indicates that the stated performance guarantees hold for the strongly anharmonic case of liquid water. Both analytical and numerical results indicate that BCOCB excels over other known integrators in terms of accuracy, efficiency, and stability with respect to time-step for practical applications.

3.1 Introduction

Path-Integral Molecular Dynamics (PIMD) provides a practical and popular tool to simulate condensed-phase systems subject to strong nuclear quantum effects.[164–166] Based on the ring-polymer correspondence between quantum and classical Boltzmann statistics,[167, 168] PIMD exploits the computational methods of molecular dynamics[169–172] to approximate quantum thermodynamics and kinetics through various classical models.[173–178] Applications of PIMD include calcula-

tions of chemical reaction rates,[179, 180] diffusion coefficients,[181, 182] absorption spectra,[183, 184] solid and liquid structure,[94, 185] and equilibrium isotope effects.[186, 187]

Many numerical integration schemes for PIMD are based on a symmetric Trotter (i.e., Strang) splitting[188, 189] of the exact time-evolution operator, and feature a sub-step for free ring-polymer propagation.[190–192] Due to fast harmonic motions present in the free ring polymer, a *strongly stable* implementation of this sub-step is essential.[193, 194] Strong stability can be achieved by one of two approaches. The first approach introduces a preconditioned form of the equations of motion by modifying the ring-polymer mass matrix. Preconditioning improves the stability of the exact free ring-polymer update at the expense of consistent dynamics.[190, 192, 195–198] The second approach does not modify the ring-polymer mass matrix, leaving the dynamics non-preconditioned,[191, 199–202] and instead replaces the exact free ring-polymer update with a strongly stable approximation.[203] We apply the latter approach in the current work to Thermostatted Ring-Polymer Molecular Dynamics (T-RPMD),[200] a non-preconditioned variant of PIMD featuring an Ornstein–Uhlenbeck thermostat that approximately preserves the real-time dynamical accuracy of RPMD for quantum correlation functions of a wide range of observables.[204]

In addition to strong stability of the free ring-polymer update, another basic requirement of a numerical integrator for T-RPMD is non-zero overlap between the numerically sampled and exact ring-polymer configurational distributions in the limit of an infinite number of ring-polymer beads. Standard integrators fail to satisfy this requirement at any finite integration time-step,[205] which motivates the introduction of *dimension-free* T-RPMD schemes that allow for accurate configurational sampling with large time-stepping and arbitrarily many ring-polymer beads. We recently found that standard integrators could be made dimension-free through the introduction of a suitable strongly stable ring-polymer update,[205] and the current paper investigates this finding in much greater generality.

To this end, we introduce a function θ that defines the free ring-polymer update and deduce how the choice of θ impacts the properties and performance of the corresponding T-RPMD integrator. The case $\theta(x) = x$, i.e., θ is the identity, corresponds to the exact free ring-polymer update. Therefore, to ensure second-order accuracy, θ must approximate the identity near the origin, i.e., $\theta(0) = 0$, $\theta'(0) = 1$ and $\theta''(0) = 0$. Moreover, strong stability requires that the range of the function θ

is within $(0,\pi)$ for $x > 0$, and ergodicity and dimensionality freedom of the corresponding T-RPMD integrator impose additional requirements on θ. There are many choices of θ that fulfill the identified requirements, including $\theta(x) = 2\arctan(x/2)$ which leads to the BCOCB scheme introduced in Ref. [205]. In fact, we find that this choice of θ is superior for the estimation of configurational averages via T-RPMD from the perspectives of accuracy and efficiency, despite its poor performance with respect to the ring-polymer velocities.

3.2 Theory

T-RPMD

Consider a one-dimensional quantum particle with the Hamiltonian operator

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + V(\hat{q}) \,, \tag{3.1}$$

where m is the particle mass, $\hat{q}$ and $\hat{p}$ the position and momentum operators, and $V(\hat{q})$ a potential energy surface. Ignoring exchange statistics, the properties of this system at thermal equilibrium are encoded in the quantum partition function

$$Q = \mathrm{tr}[e^{-\beta\hat{H}}] \,, \tag{3.2}$$

where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant and T the physical temperature. Using a path-integral discretization (i.e., a Trotter factorization of the Boltzmann operator[188]), $Q = \lim_{n\to\infty} Q_n$ can be approximated by the classical partition function Q_n of a ring polymer with n beads,[167, 168]

$$Q_n = \frac{m^n}{(2\pi\hbar)^n} \int \mathrm{d}^n \boldsymbol{q} \int \mathrm{d}^n \boldsymbol{v}\, e^{-\beta H_n(\boldsymbol{q},\boldsymbol{v})} \,, \tag{3.3}$$

where $\boldsymbol{q} = \begin{bmatrix} q_0 & \dots & q_{n-1} \end{bmatrix}^{\mathrm{T}}$ is the vector of bead positions and $\boldsymbol{v}$ the corresponding vector of velocities. The ring-polymer Hamiltonian is given by

$$H_n(\boldsymbol{q},\boldsymbol{v}) = H_n^0(\boldsymbol{q},\boldsymbol{v}) + V_n^{\mathrm{ext}}(\boldsymbol{q}) \,, \tag{3.4}$$

which includes contributions from the physical potential

$$V_n^{\mathrm{ext}}(\boldsymbol{q}) = \frac{1}{n}\sum_{j=0}^{n-1} V(q_j) \tag{3.5}$$

and the free ring-polymer Hamiltonian

$$H_n^0(\boldsymbol{q},\boldsymbol{v}) = \frac{m_n}{2}\sum_{j=0}^{n-1}\left[v_j^2 + \omega_n^2 (q_{j+1} - q_j)^2\right] \,, \tag{3.6}$$

where $m_n = m/n$, $\omega_n = n/(\hbar\beta)$ and $q_n = q_0$.

T-RPMD evolves the phase $\begin{bmatrix} \boldsymbol{q}^{\mathrm{T}} & \boldsymbol{v}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ of the ring polymer as per

$$\begin{aligned} \dot{\boldsymbol{q}}(t) = \boldsymbol{v}(t)\,;\, \dot{\boldsymbol{v}}(t) = &-\boldsymbol{\Omega}^2\boldsymbol{q}(t) + m_n^{-1}\boldsymbol{F}(\boldsymbol{q}(t)) \\ &- \boldsymbol{\Gamma}\boldsymbol{v}(t) + \sqrt{2\beta^{-1}m_n^{-1}}\boldsymbol{\Gamma}^{1/2}\dot{\boldsymbol{W}}(t)\,, \end{aligned} \tag{3.7}$$

which is a coupling of the Hamitonian dynamics of $H_n(\boldsymbol{q},\boldsymbol{v})$ with a Ornstein–Uhlenbeck thermostat. In Eq. 3.7 we introduced $\boldsymbol{F}(\boldsymbol{q}) = -\nabla V_n^{\mathrm{ext}}(\boldsymbol{q})$, an n-dimensional standard Brownian motion $\boldsymbol{W}(t)$ and the $n \times n$ matrices

$$\begin{aligned} \boldsymbol{\Omega} &= \boldsymbol{U}\,\mathrm{diag}\,(0, \omega_{1,n}, \ldots, \omega_{n-1,n})\,\boldsymbol{U}^{\mathrm{T}} \text{ and} \\ \boldsymbol{\Gamma} &= \boldsymbol{U}\,\mathrm{diag}\,(0, \gamma_{1,n}, \ldots, \gamma_{n-1,n})\,\boldsymbol{U}^{\mathrm{T}}\,, \end{aligned} \tag{3.8}$$

where $\gamma_{j,n} \geq 0$ is the jth friction coefficient, $\boldsymbol{U}$ the $n \times n$ real discrete Fourier transform matrix, and the ring-polymer frequencies are given by

$$\omega_{j,n} = \begin{cases} 2\omega_n \sin\left(\frac{\pi j}{2n}\right) & \text{if } j \text{ is even}\,, \\ 2\omega_n \sin\left(\frac{\pi(j+1)}{2n}\right) & \text{else}\,. \end{cases} \tag{3.9}$$

Observe that the zero-frequency (i.e., centroid) ring-polymer mode is uncoupled from the thermostat, and the coefficients $\{\gamma_{j,n}\}_{j=1}^{n-1}$ in Eq. 3.8 constitute the friction schedule applied to the non-centroid modes.

Numerical integrators for Eq. 3.7 typically employ symmetric propagator splittings of the form[206–208]

$$\begin{aligned} e^{\Delta t\mathscr{L}_n} \approx\; & e^{a\frac{\Delta t}{2}\mathscr{O}_n} e^{\frac{\Delta t}{2}\mathscr{B}_n} e^{\frac{\Delta t}{2}\mathscr{A}_n} e^{(1-a)\Delta t\mathscr{O}_n} \\ & \times e^{\frac{\Delta t}{2}\mathscr{A}_n} e^{\frac{\Delta t}{2}\mathscr{B}_n} e^{a\frac{\Delta t}{2}\mathscr{O}_n} \quad \text{with } a \in \{0,1\}, \end{aligned} \tag{3.10}$$

where the operator $\mathscr{L}_n = \mathscr{A}_n + \mathscr{B}_n + \mathscr{O}_n$ includes contributions from the n-bead free ring-polymer motion ($\mathscr{A}_n$), the external potential ($\mathscr{B}_n$) and the thermostat ($\mathscr{O}_n$), and Δt is a sufficiently small time-step. Note that the standard microcanonical RPMD integrator is recovered in the limit of zero coupling to the thermostat,[191] and that Eq. 3.10 yields the OBABO scheme of Bussi and Parrinello [206] if $a = 1$ and the BAOAB scheme of Leimkuhler [207] if $a = 0$.

Standard implementations of the T-RPMD splittings in Eq. 3.10 use the exact free ring-polymer propagator $e^{\frac{\Delta t}{2}\mathscr{A}_n}$ to evolve the uncoupled ring-polymer modes; however, recent work by us[203] showed that such implementations exhibit poor ergodicity if large numbers n of ring-polymer beads are employed in conjunction

with large time-steps Δt, and suggested replacing the exact ring-polymer propagator with its Cayley approximation[209] for improved performance. Follow-up work[205] introduced a Cayley-modified BAOAB scheme, denoted BCOCB, and presented numerical evidence that cemented the scheme as an improvement over standard BAOAB due to its superior equilibrium accuracy and time-step stability.

Generalizing beyond the Cayley modification, the current work studies a family of modified BAOAB schemes that contains BCOCB and introduces others with similar theoretical guarantees. Specifically, the BAOAB modifications are obtained by replacing the exact free ring-polymer update in Eq. 3.10 with approximations that endow the properties listed below.

(**P1**) *Strong stability.* For a free ring polymer (i.e., for $V(q) = \text{const.}$), the integrator with $\gamma_{j,n} = 0$ is both strongly stable and second-order accurate in Δt.

(**P2**) *Free ring-polymer ergodicity.* For a free ring polymer, the integrator with $\gamma_{j,n} > 0$ is ergodic with respect to the distribution with density proportional to $e^{-\beta H_n^0(\boldsymbol{q},\boldsymbol{v})}$.

(**P3**) *Dimension-free stability.* For a harmonically confined ring polymer (i.e., for $V(q) = (\Lambda/2)\,q^2$), the integrator with $\gamma_{j,n} = 0$ is stable for any n if Δt leads to stable integration for $n = 1$.

(**P4**) *Dimension-free ergodicity.* For a harmonically confined ring polymer, the integrator with $\gamma_{j,n} > 0$ and stable Δt is ergodic with respect to its stationary distribution for any n.

(**P5**) *Dimension-free equilibrium accuracy.* For a harmonically confined ring polymer, the integrator leaves invariant an accurate approximation of the distribution with density proportional to $e^{-\frac{\beta m_n}{2}\boldsymbol{q}^{\mathrm{T}}\left(\frac{\Lambda}{m}+\boldsymbol{\Omega}^2\right)\boldsymbol{q}}$, with bounded error for any n.

To obtain integrators satisfying properties (**P1**)-(**P5**), we introduce a function θ that defines the free ring-polymer update and then construct θ accordingly. To this end, let

$$\mathscr{S}_{j,n}^{1/2} \;=\; \mathscr{Q}_{j,n}\begin{bmatrix} e^{i\theta(\omega_{j,n}\Delta t)/2} & 0 \\ 0 & e^{-i\theta(\omega_{j,n}\Delta t)/2}\end{bmatrix}\mathscr{Q}_{j,n}^{-1}\,, \tag{3.11}$$

where $\mathscr{Q}_{j,n} = \begin{bmatrix} 1 & 1 \\ i\omega_{j,n} & -i\omega_{j,n}\end{bmatrix}$ and essential properties of θ are determined in the sequel. We focus on T-RPMD schemes derived from the BAOAB splitting (i.e.,

$a = 0$ in Eq. 3.10) with the exact free ring-polymer update replaced by $\mathscr{S}^{1/2}_{j,n}$. For such schemes, an integration time-step is composed of the following sequence of sub-steps:

B: Update velocities for half a step: $\boldsymbol{v} \leftarrow \boldsymbol{v} + \frac{\Delta t}{2}\frac{\boldsymbol{F}}{m_n}$.

Convert bead Cartesian coordinates to normal modes using

$$\boldsymbol{\varrho} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{q} \qquad \text{and} \qquad \boldsymbol{\varphi} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{v}\,. \tag{3.12}$$

A: Evolve the free ring polymer in normal-mode coordinates for half a step:

$$\begin{bmatrix}\rho_j \\ \varphi_j\end{bmatrix} \leftarrow \mathscr{S}^{1/2}_{j,n}\begin{bmatrix}\rho_j \\ \varphi_j\end{bmatrix} \quad \text{for } 0 \le j \le n-1. \tag{3.13}$$

O: Perform an Ornstein–Uhlenbeck velocity update for a full time-step:

$$\varphi_j \leftarrow e^{-\gamma_{j,n}\Delta t}\varphi_j + \sqrt{\frac{1-e^{-2\gamma_{j,n}\Delta t}}{\beta m_n}}\xi_j\,, \tag{3.14}$$

where ξ_j are independent standard normal random variables and $0 \le j \le n-1$.

A: Evolve the free ring polymer in normal-mode coordinates for half a step:

$$\begin{bmatrix}\rho_j \\ \varphi_j\end{bmatrix} \leftarrow \mathscr{S}^{1/2}_{j,n}\begin{bmatrix}\rho_j \\ \varphi_j\end{bmatrix} \quad \text{for } 0 \le j \le n-1. \tag{3.15}$$

Convert back to bead Cartesian coordinates using the inverse of $\boldsymbol{U}$, which is just its transpose since $\boldsymbol{U}$ is orthogonal.

B: Update velocities for half a step: $\boldsymbol{v} \leftarrow \boldsymbol{v} + \frac{\Delta t}{2}\frac{\boldsymbol{F}}{m_n}$.

In the remainder of this section, we identify conditions on the choice of θ that imply properties (**P1**)-(**P5**) for the corresponding T-RPMD integrator. Despite our focus on BAOAB-like splittings, we describe how the conditions on θ can be adjusted to construct integrators derived from the OBABO splitting (i.e., $a = 1$ in Eq. 3.10) that satisfy properties (**P1**)-(**P5**).

Strong stability of RPMD with a constant external potential

In this section, sufficient conditions on θ are identified to satisfy property (**P1**) in Section 3.2. Let $V(q) = \text{const.}$ and $\gamma_{j,n} = 0$ for $1 \leq j \leq n-1$, corresponding to the free ring polymer. The jth normal mode $\begin{bmatrix} p_j & q_j \end{bmatrix}^{\mathrm{T}}$ satisfies

$$\begin{bmatrix} \dot{p}_j \\ \dot{q}_j \end{bmatrix} = \boldsymbol{A}_{j,n} \begin{bmatrix} p_j \\ q_j \end{bmatrix} \quad \text{where} \quad \boldsymbol{A}_{j,n} = \begin{bmatrix} 0 & 1 \\ -\omega_{j,n}^2 & 0 \end{bmatrix}. \tag{3.16}$$

In this case, the algorithm from Section 3.2 reduces to a full step of $\mathscr{S}_{j,n} \approx \exp(\Delta t \boldsymbol{A}_{j,n})$, i.e.,

$$\begin{bmatrix} p_j \\ q_j \end{bmatrix} \leftarrow \mathscr{S}_{j,n} \begin{bmatrix} p_j \\ q_j \end{bmatrix} \quad \text{for } 0 \leq j \leq n-1, \tag{3.17}$$

where $\mathscr{S}_{j,n} = \mathscr{S}_{j,n}^{1/2} \mathscr{S}_{j,n}^{1/2}$ follows from Eq. 3.11 and the function θ is such that property (**P1**) holds.

We proceed to identify sufficient conditions on θ such that the corresponding free ring-polymer update satisfies property (**P1**). First note that for any function θ such that $\theta(-x) = -\theta(x)$ for $x > 0$, the structure of $\mathscr{S}_{j,n}^{1/2}$ guarantees that the corresponding free ring-polymer update is reversible, symplectic, and preserves the free ring-polymer Hamiltonian $H_n^0(\boldsymbol{q}, \boldsymbol{v})$. Now, observe that $\mathscr{S}_{j,n}$ is exact if $\theta(x) = x$; therefore, second-order accuracy requires that θ approximates the identity near the origin, i.e.,

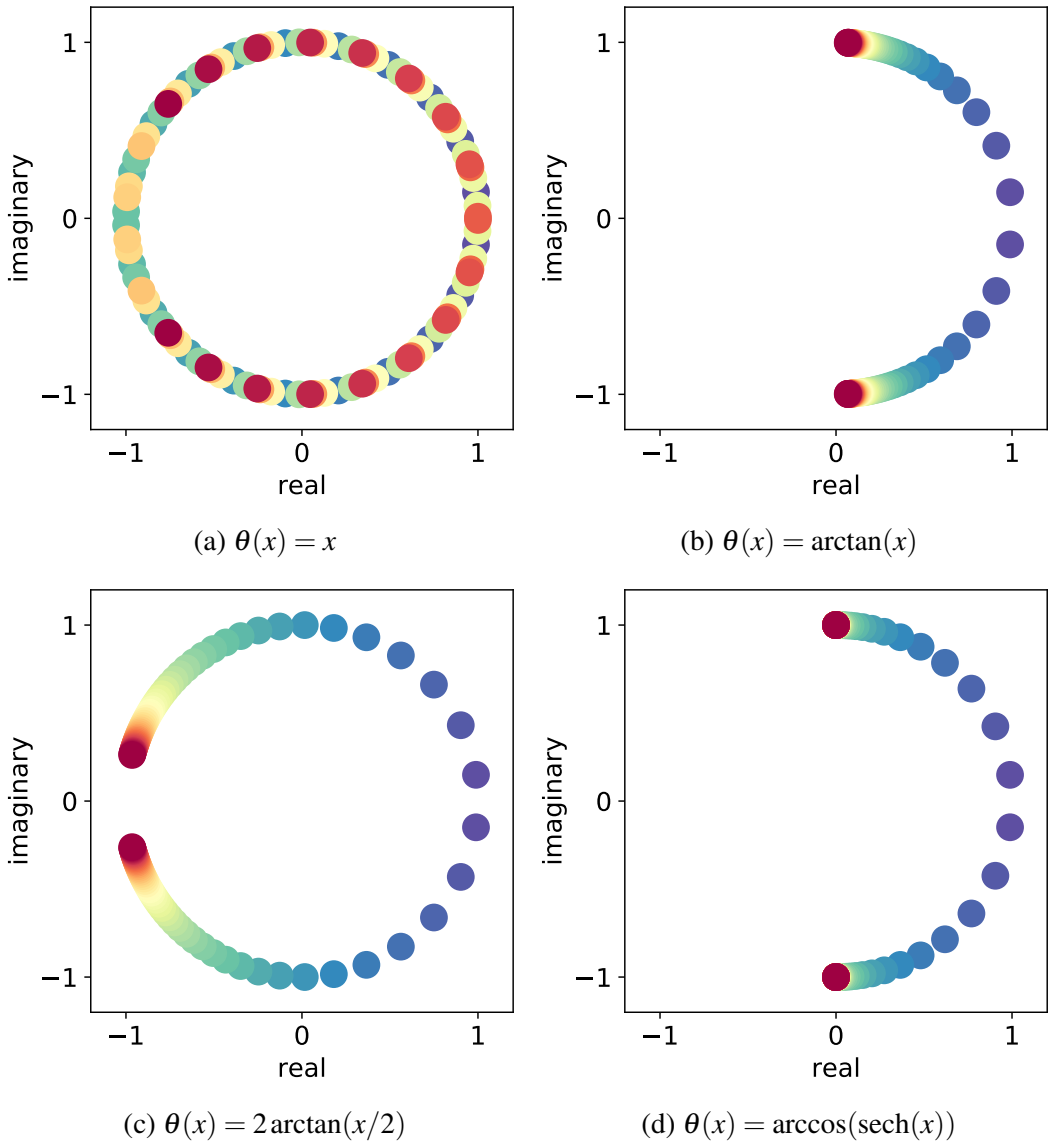
$$\theta(0) = 0,\ \theta'(0) = 1, \text{ and } \theta''(0) = 0. \tag{**C1**}$$

Moreover, strong stability follows if the eigenvalues $e^{\pm i\theta(\omega_{j,n}\Delta t)}$ of $\mathscr{S}_{j,n}$ are distinct;[203] to this end we require that

$$0 < \theta(x) < \pi \quad \text{for } x > 0. \tag{**C2**}$$

Jointly, conditions (**C1**) and (**C2**) guarantee that the update in Eq. 3.17 satisfies property (**P1**). There are many different choices of θ that obey these conditions, e.g., $\theta(x) = \arctan(x)$, $\arccos(\operatorname{sech}(x))$,[1] and $2\arctan(x/2)$. The latter choice leads to the Cayley approximation of the free ring-polymer update, as can be verified by substitution in Eq. 3.11 and comparison of the resulting $\mathscr{S}_{j,n}^{1/2}$ with Eq. 17 in Ref. [205]. Figure 3.1 compares the eigenvalues of $\mathscr{S}_{j,n}$ with $\theta(x) = x$ and several choices of θ that meet conditions (**C1**) and (**C2**).

[1] The function $\theta(x) = \arccos(\operatorname{sech}(x))$ is not differentiable at the origin and hence, strictly speaking, does not satisfy condition (**C1**). Moreover, the function has even symmetry and hence fails to yield a reversible free ring-polymer update. These formal shortcomings can be fixed by multiplying the function by $\operatorname{sign}(x)$, which we implicitly do for this and other functions θ with similar features.



(a) $\theta(x) = x$

(b) $\theta(x) = \arctan(x)$

(c) $\theta(x) = 2\arctan(x/2)$

(d) $\theta(x) = \arccos(\operatorname{sech}(x))$

Figure 3.1: Eigenvalues of $\mathscr{S}_{j,n}$ for 50 different time step sizes between 0.05 and 5.0 (evenly spaced) and fixed Matsubara frequency $\omega = 3$. The colors go from blue (smallest time step) through green and yellow to red (largest time step). In panel (a), the eigenvalues rotate around the unit circle several times, which indicates that the corresponding $\mathscr{S}_{j,n}$ is not always strongly stable. In panels (b), (c), and (d), the eigenvalues are distinct and on the unit circle; thus the corresponding $\mathscr{S}_{j,n}$ is strongly stable.

Ergodicity of T-RPMD with a constant external potential

In this section, it is shown that condition (**C2**) implies property (**P2**) in Section 3.2. Let $V(q) = \text{const.}$ and $\gamma_{j,n} > 0$ for $1 \leq j \leq n-1$, corresponding to the free ring polymer with a Ornstein–Uhlenbeck thermostat. In this case, the jth normal mode satisfies

$$\begin{bmatrix} \dot{\rho}_j \\ \dot{\varphi}_j \end{bmatrix} = (\boldsymbol{A}_{j,n} + \boldsymbol{C}_{j,n}) \begin{bmatrix} \rho_j \\ \varphi_j \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{\frac{2\gamma_{j,n}}{\beta m_n}} \dot{W}_j \end{bmatrix}, \tag{3.18}$$

where $\boldsymbol{C}_{j,n} = \begin{bmatrix} 0 & 0 \\ 0 & -\gamma_{j,n} \end{bmatrix}$ and $\dot{W}_j$ is a scalar white-noise. The solution $\begin{bmatrix} \rho_j(t) & \varphi_j(t) \end{bmatrix}^{\mathrm{T}}$ of Eq. 3.18 is an ergodic Markov process, and in the limit as $t \to \infty$, its distribution converges to the centered bivariate normal with covariance

$$\boldsymbol{\Sigma}_{j,n} = \frac{1}{\beta m_n} \begin{bmatrix} s_{j,n}^2 & 0 \\ 0 & 1 \end{bmatrix} \text{ where } s_{j,n}^2 = \frac{1}{\omega_{j,n}^2}. \tag{3.19}$$

This distribution corresponds to the jth marginal of the free ring-polymer equilibrium distribution with density proportional to $e^{-\beta H_n^0(\boldsymbol{q},\boldsymbol{v})}$.

The choice of $\gamma_{j,n} > 0$ in Eq. 3.18 determines the rate at which the associated Markov process converges to its stationary distribution if initialized away from it. When $\gamma_{j,n} < 2\omega_{j,n}$, the process is dominated by the deterministic Hamiltonian dynamics and is characterized as *underdamped*; on the other hand, when $\gamma_{j,n} > 2\omega_{j,n}$, the process is *overdamped*; and at the critical value $\gamma_{j,n} = 2\omega_{j,n}$ the process is characterized as *critically damped* and converges to equilibrium fastest.[210, 211] This analytical result motivates the so-called PILE friction schedule.[191, 200] We specialize to this schedule in the remainder of the section and set $\gamma_{j,n} = 2\omega_{j,n}$ for $1 \leq j \leq n-1$.

The BAOAB-like update in Section 3.2 applied to Eq. 3.18 can be written compactly as

$$\begin{bmatrix} \rho_j \\ \varphi_j \end{bmatrix} \leftarrow \mathscr{M}_{j,n} \begin{bmatrix} \rho_j \\ \varphi_j \end{bmatrix} + \mathscr{R}_{j,n}^{1/2} \begin{bmatrix} \xi_j \\ \eta_j \end{bmatrix} \quad \text{for } 0 \leq j \leq n-1, \tag{3.20}$$

where ξ_j and η_j are independent standard normal random variables and we have

introduced the 2×2 matrices

$$\mathscr{M}_{j,n} = \mathscr{S}_{j,n}^{1/2} \mathscr{O}_{j,n} \mathscr{S}_{j,n}^{1/2}, \quad \mathscr{O}_{j,n} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-2\omega_{j,n}\Delta t} \end{bmatrix} \text{ and}$$
$$\mathscr{R}_{j,n} = \frac{1 - e^{-4\omega_{j,n}\Delta t}}{\beta m_n} \mathscr{S}_{j,n}^{1/2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} (\mathscr{S}_{j,n}^{1/2})^{\mathrm{T}} .$$

Since $\mathscr{S}_{j,n}^{1/2}$ and the Ornstein–Uhlenbeck update are individually preservative irrespective of the chosen θ, Eq. 3.20 exactly preserves the free ring-polymer equilibrium distribution for any choice of θ that satisfies (**C1**) and (**C2**).

The ergodicity of the integrator specified by Eq. 3.20 depends entirely on the asymptotic stability of $\mathscr{M}_{j,n}$, i.e., whether or not $\|\mathscr{M}_{j,n}^k\| \to 0$ as $k \to \infty$ where $\|\cdot\|$ is a matrix norm. The matrix $\mathscr{M}_{j,n}$ is asymptotically stable if its *spectral radius* (i.e., the modulus of its largest eigenvalue) is smaller than unity,[194] which depends on

$$\det(\mathscr{M}_{j,n}) = e^{-2\omega_{j,n}\Delta t} \quad \text{and}$$
$$\mathrm{tr}(\mathscr{M}_{j,n}) = \cos(\theta(\omega_{j,n}\Delta t))(1 + e^{-2\omega_{j,n}\Delta t}) .$$

In particular, the eigenvalues of $\mathscr{M}_{j,n}$ are both inside the unit circle if and only if

$$|\mathrm{tr}(\mathscr{M}_{j,n})| < 1 + \det(\mathscr{M}_{j,n}) < 2 ; \tag{3.21}$$

a proof of this claim is provided in Appendix 7.1. This inequality reveals that condition (**C2**) implies property (**P2**). Moreover, if $\mathrm{tr}(\mathscr{M}_{j,n})^2 - 4\det(\mathscr{M}_{j,n}) \leq 0$, then the spectral radius of $\mathscr{M}_{j,n}$ is minimal and equal to $\sqrt{\det(\mathscr{M}_{j,n})} = e^{-\omega_{j,n}\Delta t}$; this occurs when $|\cos(\theta(\omega_{j,n}\Delta t))| \leq \mathrm{sech}(\omega_{j,n}\Delta t)$ for all $\omega_{j,n}\Delta t$, which holds if the function θ satisfies

$$\arccos(\mathrm{sech}(x)) \leq \theta(x) \leq \pi - \arccos(\mathrm{sech}(x)) \quad \text{for } x > 0. \tag{3.22}$$

Any choice of θ that does not satisfy Eq. 3.22 will be overdamped in some modes, in the sense that the corresponding $\mathscr{M}_{j,n}$ will have a spectral radius strictly larger than $e^{-\omega_{j,n}\Delta t}$.

The function $\theta(x) = \arccos(\mathrm{sech}(x))$ saturates the (left) inequality in Eq. 3.22 while satisfying conditions (**C1**) and (**C2**), and hence provides a strongly stable and critically damped integrator for the thermostatted free ring polymer. As an illustration, Fig. 3.2a shows that $\theta(x) = \arctan(x)$ is overdamped for all modes whereas the Cayley angle $\theta(x) = 2\arctan(x/2)$ exhibits mixed damping. In contrast, the function

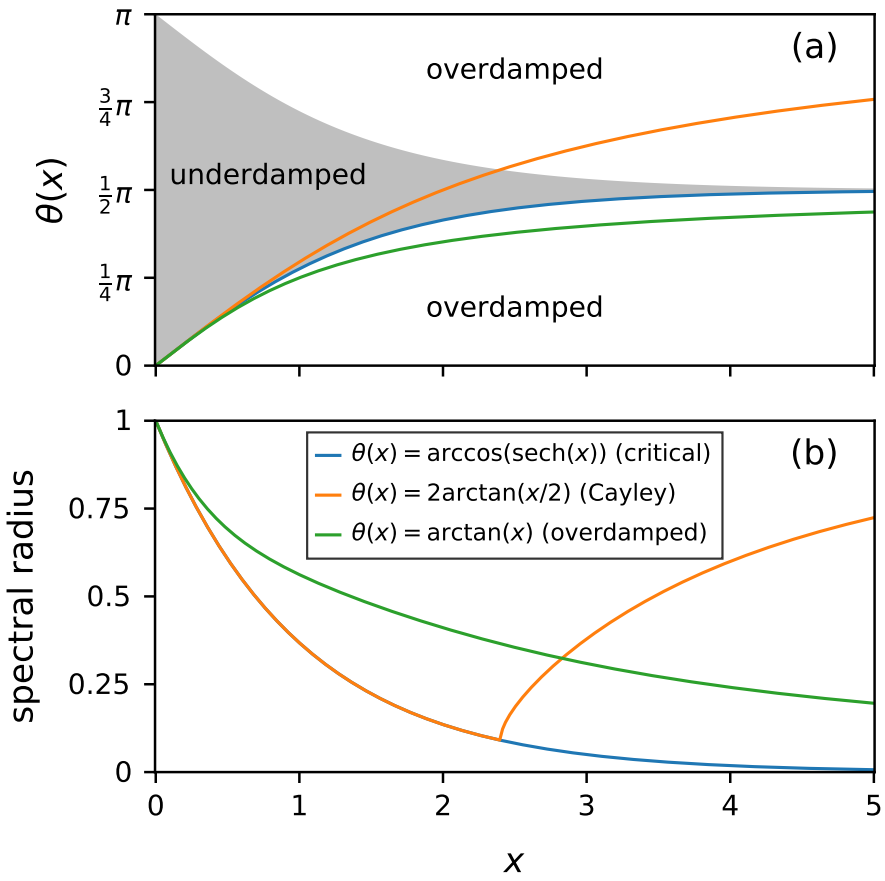


Figure 3.2: Spectral properties of the T-RPMD update for the free ring polymer for various choices of θ. Panel (a) plots the functions $\theta(x) = \arccos(\mathrm{sech}(x))$, $\arctan(x)$ and $\arctan(x/2)$, and regions of overdamping and underdamping with PILE friction, separated at the locus of points where $|\cos(\theta(x))\cosh(x)| = 1$. The gray region ($|\cos(\theta(x))\cosh(x)| < 1$) is where the dynamics is underdamped, while in the white region ($|\cos(\theta(x))\cosh(x)| > 1$) the dynamics is overdamped. The function $\theta(x) = \arctan(x)$ lies in the overdamped region for $x > 0$, whereas $\theta(x) = 2\arctan(x/2)$ is in the underdamped region for $x \lessapprox 2.4$ and in the overdamped region otherwise. The function $\theta(x) = \arccos(\mathrm{sech}(x))$, however, is critically damped for $x > 0$ and optimizes the convergence rate of the integrator. Panel (b) plots the spectral radius of $\mathscr{M}_{j,n}$ corresponding to each choice of θ as a function of x.

$\theta(x) = \arccos(\mathrm{sech}(x))$ preserves the critically damped behavior of its continuous counterpart under the PILE friction schedule. Figure 3.2b confirms that the spectral radius of $\mathscr{M}_{j,n}$ is minimal at $\theta(x) = \arccos(\mathrm{sech}(x))$ for $x > 0$; consequently, this choice of θ optimizes the convergence of the integrator to stationarity.

Conditions (**C1**) and (**C2**) also imply property (**P2**) for the OBABO-like update associated with a compliant choice of θ, because the matrices $\mathscr{S}_{j,n}^{1/2}\mathscr{O}_{j,n}\mathscr{S}_{j,n}^{1/2}$ and $\mathscr{O}_{j,n}^{1/2}\mathscr{S}_{j,n}\mathscr{O}_{j,n}^{1/2}$ have equal spectral radii.

Dimension-free stability of RPMD with a harmonic external potential

In this section, we identify a condition on θ that yields property (**P3**) in Section 3.2. Let $V(q) = (\Lambda/2)q^2$ and $\gamma_{j,n} = 0$ for $1 \leq j \leq n-1$, corresponding to the non-thermostatted ring polymer with a harmonic external potential. In this case, the jth normal mode satisfies

$$\begin{bmatrix} \dot{\rho}_j \\ \dot{\varphi}_j \end{bmatrix} = (\boldsymbol{A}_{j,n} + \boldsymbol{B}) \begin{bmatrix} \rho_j \\ \varphi_j \end{bmatrix}, \tag{3.23}$$

where $\boldsymbol{B} = \begin{bmatrix} 0 & 0 \\ -\Lambda/m & 0 \end{bmatrix}$, and conserves the Hamiltonian

$$H_{j,n}(\rho_j, \varphi_j) = \frac{m_n}{2}\left(|\varphi_j|^2 + (\omega_{j,n}^2 + \Lambda/m)|\rho_j|^2\right). \tag{3.24}$$

For this system, the BAOAB-like update in Section 3.2 reduces to

$$\begin{bmatrix} \rho_j \\ \varphi_j \end{bmatrix} \leftarrow \mathscr{M}_{j,n} \begin{bmatrix} \rho_j \\ \varphi_j \end{bmatrix} \quad \text{for } 0 \leq j \leq n-1, \tag{3.25}$$

where we have introduced the 2×2 matrices

$$\mathscr{M}_{j,n} = \mathscr{B}^{1/2}\mathscr{S}_{j,n}\mathscr{B}^{1/2} \text{ and } \mathscr{B}^{1/2} = \begin{bmatrix} 1 & 0 \\ -\Delta t(\Lambda/m)/2 & 1 \end{bmatrix}. \tag{3.26}$$

This update may be interpreted as a symplectic perturbation of the free ring-polymer update in Eq. 3.17 due to the harmonic external potential,[203] and conserves a modification of $H_{j,n}$ for choices of θ and Δt that lead to stable integration.[212]

The update in Eq. 3.25 is stable if[196]

$$\max_{0 \leq j \leq n-1} \frac{1}{2}|\mathrm{tr}(\mathscr{M}_{j,n})| = \max_{0 \leq j \leq n-1} |\mathscr{A}_{j,n}| < 1, \tag{3.27}$$

where

$$\mathscr{A}_{j,n} = \cos(\theta(\omega_{j,n}\Delta t)) - \frac{\Delta t^2(\Lambda/m)}{2}\frac{\sin(\theta(\omega_{j,n}\Delta t))}{\omega_{j,n}\Delta t}. \tag{3.28}$$

Moreover, the 0th (i.e., centroid) mode, like the single-bead ring polymer, evolves through the velocity Verlet algorithm, whose stability requires that $\Delta t^2\Lambda/m < 4$. Combining this requirement with condition (**C2**) yields a sufficient condition for Eq. 3.27 to hold at any bead number n,

$$0 < \theta(x) \leq 2\arctan(x/2) \quad \text{for } x > 0. \tag{**C3**}$$

A proof of this result is provided in Appendix 7.2. The functions $\theta(x) = 2\arctan(x/2)$, $\arctan(x)$ and $\arccos(\mathrm{sech}(x))$ all satisfy condition (**C3**), which ensures that the corresponding RPMD integrator meets property (**P3**).

Dimension-free ergodicity and equilibrium accuracy of T-RPMD with a harmonic external potential

In this section, it is shown that condition (**C3**) implies property (**P4**) in Section 3.2, and an additional condition is introduced to ensure that property (**P5**) holds. Let $V(q) = (\Lambda/2)\,q^2$ and $\gamma_{j,n} = 2\omega_{j,n}$ for $1 \le j \le n-1$. In this case, the jth normal mode satisfies

$$\begin{bmatrix} \dot{\rho}_j \\ \dot{\varphi}_j \end{bmatrix} = (\boldsymbol{A}_{j,n} + \boldsymbol{B} + \boldsymbol{C}_{j,n}) \begin{bmatrix} \rho_j \\ \varphi_j \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{\frac{4\omega_{j,n}}{\beta m_n}}\dot{W}_j \end{bmatrix} . \tag{3.29}$$

The solution $\begin{bmatrix} \rho_j(t) & \varphi_j(t) \end{bmatrix}^{\mathrm{T}}$ of Eq. 3.29 is an ergodic Markov process, and its distribution as $t \to \infty$ converges to the centered bivariate normal with covariance matrix

$$\boldsymbol{\Sigma}_{j,n} = \frac{1}{\beta m_n} \begin{bmatrix} s_{j,n}^2 & 0 \\ 0 & 1 \end{bmatrix} \text{ where } s_{j,n}^2 = \frac{1}{\Lambda/m + \omega_{j,n}^2} ; \tag{3.30}$$

the associated position-marginal is the jth marginal of the ring-polymer configurational distribution with density $e^{-\frac{\beta m_n}{2}\boldsymbol{q}^{\mathrm{T}}\left(\frac{\Lambda}{m}+\boldsymbol{\Omega}^2\right)\boldsymbol{q}}$.

For this system, the BAOAB-like update in Section 3.2 is of the same form as Eq. 3.20 with

$$\begin{aligned} \mathscr{M}_{j,n} &= \mathscr{B}^{1/2}\mathscr{S}_{j,n}^{1/2}\mathscr{O}_{j,n}\mathscr{S}_{j,n}^{1/2}\mathscr{B}^{1/2} \quad \text{and} \\ \mathscr{R}_{j,n} &= \frac{1-e^{-4\omega_{j,n}\Delta t}}{\beta m_n}\mathscr{B}^{1/2}\mathscr{S}_{j,n}^{1/2}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}(\mathscr{B}^{1/2}\mathscr{S}_{j,n}^{1/2})^{\mathrm{T}} . \end{aligned} \tag{3.31}$$

As in the case of a constant external potential, the ergodicity of this integrator depends on the spectral radius of $\mathscr{M}_{j,n}$. By Theorem 3 in Appendix 7.1 and the fact that

$$\begin{aligned} \det(\mathscr{M}_{j,n}) &= e^{-2\omega_{j,n}\Delta t} \quad \text{and} \\ \mathrm{tr}(\mathscr{M}_{j,n}) &= \mathscr{A}_{j,n}(1+e^{-2\omega_{j,n}\Delta t}) , \end{aligned}$$

it follows that condition (**C3**) gives a simple and sufficient condition for ergodicity at any bead number n and hence implies property (**P4**) for the BAOAB-like update specified by Eqs. 3.20 and 3.31. Furthermore, because the matrix $\mathscr{M}_{j,n}$ of the corresponding OBABO-like update has equal trace and determinant, condition (**C3**) also guarantees property (**P4**) in that case.[2]

[2]Condition (**C3**) may be viewed as a relaxation of the sufficient condition for ergodicity given in Eq. (18) of Ref. [205]. Indeed, condition (**C3**) implies ergodicity irrespective of the Ornstein–Uhlenbeck friction schedule, whereas Eq. (18) in Ref. [205] does not imply ergodicity for friction schedules that lead to overdamped dynamics.

If condition (**C3**) holds, the BAOAB-like update is ergodic with respect to a centered bivariate normal distribution whose covariance matrix $\boldsymbol{\Sigma}_{j,\Delta t}$ satisfies the linear equation

$$\boldsymbol{\Sigma}_{j,\Delta t} = \mathscr{M}_{j,n}\boldsymbol{\Sigma}_{j,\Delta t}\mathscr{M}_{j,n}^{\mathrm{T}} + \mathscr{R}_{j,n}\,, \tag{3.32}$$

for which the solution is

$$\boldsymbol{\Sigma}_{j,\Delta t} = \frac{1}{\beta m_n}\begin{bmatrix} s_{j,\Delta t}^2 & 0 \\ 0 & r_{j,\Delta t}^2 \end{bmatrix}, \tag{3.33}$$

where the variance in the position- and velocity-marginal is respectively $(\beta m_n)^{-1}s_{j,\Delta t}^2$ and $(\beta m_n)^{-1}r_{j,\Delta t}^2$ with

$$\begin{aligned} s_{j,\Delta t}^2 &= \left(\omega_{j,n}^2 + \frac{\Lambda}{m}\frac{\omega_{j,n}\Delta t/2}{\tan\left(\theta(\omega_{j,n}\Delta t)/2\right)}\right)^{-1} \quad \text{and} \\ r_{j,\Delta t}^2 &= 1 - \frac{\Delta t^2\Lambda}{4m}\frac{\tan\left(\theta(\omega_{j,n}\Delta t)/2\right)}{\omega_{j,n}\Delta t/2}\,. \end{aligned} \tag{3.34}$$

Because the tangent function is monotonically increasing on the range of θ specified by condition (**C3**), we have the correspondence

$$0 < s_{j,\Delta t}^2 \leq s_j^2 \quad \text{and} \quad 1 - \frac{\Delta t^2\Lambda}{4m} \leq r_{j,\Delta t}^2 < 1 \tag{3.35}$$

between the exact and numerical variances of the jth ring-polymer mode. Equation 3.34 reveals that $\theta(x) = 2\arctan(x/2)$ is the unique function that complies with condition (**C3**) and saturates the inequality $s_{j,\Delta t}^2 \leq s_j^2$ in Eq. 3.35; consequently, the corresponding BAOAB-like scheme preserves the exact position-marginal in all modes and trivially satisfies property (**P5**). The BCOCB integrator from Ref. [205] corresponds to this choice of θ and thus uniquely provides optimal equilibrium position-marginal accuracy for harmonic external potentials.

To identify other BAOAB-like schemes compliant with condition (**C3**) that satisfy property (**P5**), we examine the overlap between the numerical stationary position-marginal distribution $\mu_{n,\Delta t}$ and the exact distribution μ_n, where

$$\mu_n = \prod_{j=1}^{n-1}\mathscr{N}\left(0, \tfrac{s_j^2}{\beta m_n}\right) \quad \text{and} \quad \mu_{n,\Delta t} = \prod_{j=1}^{n-1}\mathscr{N}\left(0, \tfrac{s_{j,\Delta t}^2}{\beta m_n}\right)\,. \tag{3.36}$$

Centroid-mode marginals have been suppressed in the definitions of μ_n and $\mu_{n,\Delta t}$. A BAOAB-like scheme is dimension-free if it admits an n-independent upper bound

on the distance $d_{\mathrm{TV}}(\mu_n, \mu_{n,\Delta t})$ between μ_n and $\mu_{n,\Delta t}$, where d_{TV} is the total variation metric.[213] In particular, if we require

$$\frac{x}{1+|x|} \leq \theta(x) \leq 2\arctan(x/2) \quad \text{for } x > 0, \tag{C4}$$

then we have the dimension-free bound

$$d_{\mathrm{TV}}(\mu_n, \mu_{n,\Delta t}) < \left(\sqrt{\frac{4}{3}}\frac{\hbar\beta}{\Delta t}\right)\frac{\Delta t^2 \Lambda}{m}. \tag{3.37}$$

A proof of this claim is provided in Appendix 7.4. Condition (**C4**) ensures that any BAOAB-like integrator with a compliant choice of θ meets property (**P5**).

For OBABO-like schemes, the bound in condition (**C4**) must be tightened to guarantee non-zero overlap between μ_n and $\mu_{n,\Delta t}$ for arbitrarily large n. In particular, replacing $2\arctan(x/2)$ with $\min\{2\arctan(x/2), C\}$ for some $C \in (0, \pi)$ in the upper bound of condition (**C4**) yields a n-independent bound on $d_{\mathrm{TV}}(\mu_n, \mu_{n,\Delta t})$ for all compliant OBABO-like integrators, as can be shown through arguments similar to those in Appendix 7.4.

Jointly, conditions (**C1**)-(**C4**) specify a family of BAOAB-like schemes with dimension-free stability, ergodicity and equilibrium accuracy for applications with harmonic external potentials. Numerical results in Section 3.3 suggest that the integrators exhibit similar properties in a more realistic setting with a strongly anharmonic external potential.

Dimension-free convergence to equilibrium of T-RPMD with a harmonic external potential

Beyond ensuring ergodicity of the BAOAB-like update specified by Eqs. 3.20 and 3.31, condition (**C3**) leads to explicit dimension-free equilibration rates for compliant schemes. Theorem 6 in Appendix 7.3 proves this result in the infinite-friction limit for ring-polymer modes with arbitrarily high frequency. In detail, the theorem shows that the configurational transition kernel associated with the BAOAB-like update for the jth ring-polymer mode is contractive in the 2-Wasserstein metric[214] and equilibrates any initial mode distribution at a rate determined by the function θ, the (external) potential curvature Λ, and the (stable) time-step Δt if condition (**C3**) holds. The rate in Theorem 6, though obtained in the infinite-friction limit, holds for finite friction coefficients $\gamma_{j,n}$ leading to spectral radii $\rho(\mathscr{M}_{j,n}) \leq |\mathscr{A}_{j,n}|$, where $\mathscr{A}_{j,n}$ is defined in the display after Eq. 3.27 and $|\mathscr{A}_{j,n}| = \lim_{\gamma_{j,n}\to\infty} \rho(\mathscr{M}_{j,n})$ is the spectral radius at infinite friction.

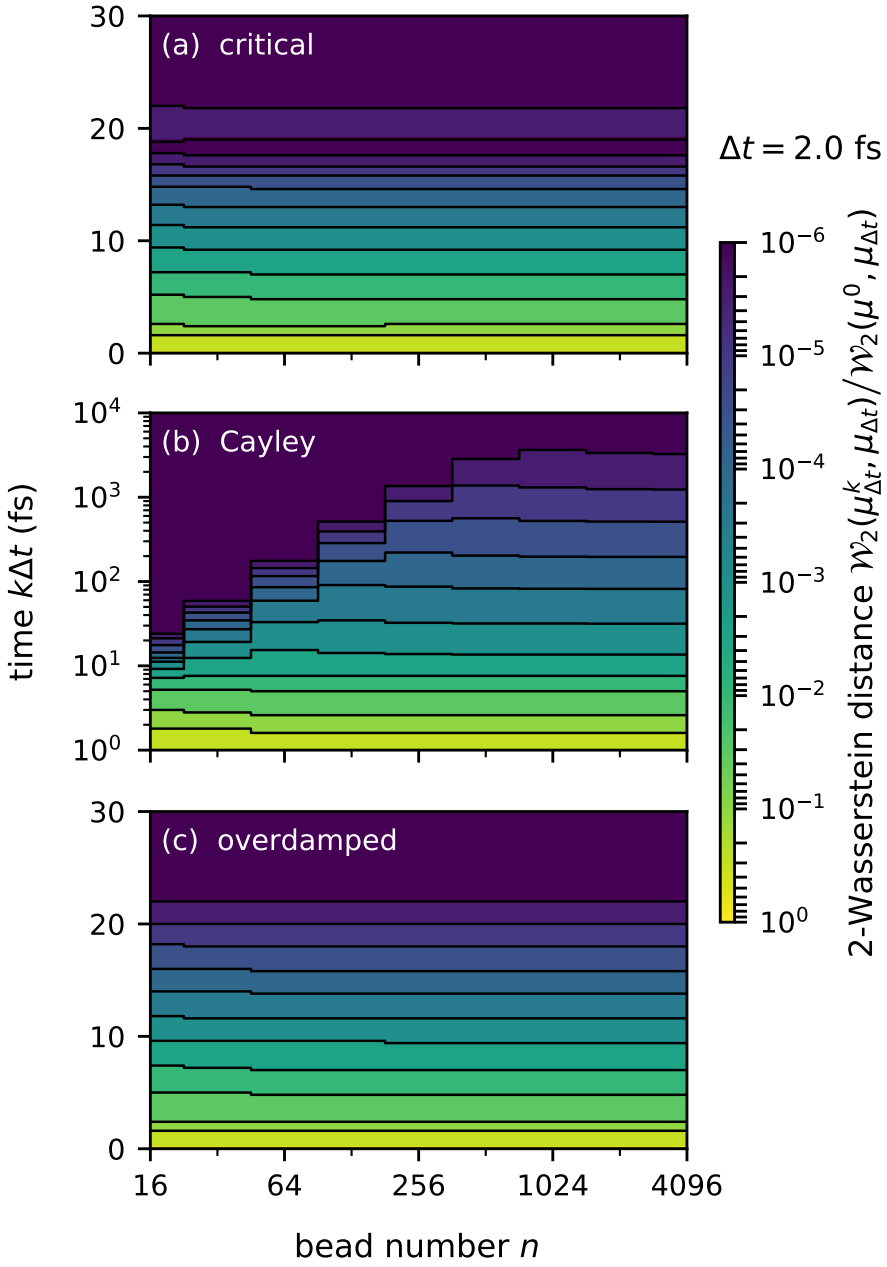


Figure 3.3: Dimension-free convergence to equilibrium of BAOAB-like T-RPMD schemes with a harmonic external potential. The physical parameters of the ring-polymer system (i.e., Λ, m, and β) are listed in Section 3.3. Panels (a), (b) and (c) plot the normalized 2-Wasserstein distance between the configurational ring-polymer distribution at stationarity and at time $k\Delta t$, as evolved via various BAOAB-like schemes from an initial point-mass distribution. Regions with darker color indicate smaller 2-Wasserstein distance to stationarity, and black lines mark iso-distance contours. The contours plateau at some value of n for all tested schemes, which checks that they exhibit dimension-free convergence as predicted by Theorem 6.

To illustrate dimension-free convergence, Fig. 3.3 plots the 2-Wasserstein distance between the stationary configurational (i.e., position-marginal) distribution $\mu_{n,\Delta t}$ and the distribution $\mu^k_{n,\Delta t}$ at the kth T-RPMD step evolved from a point mass at the origin using the schemes specified by $\theta(x) = \arccos(\operatorname{sech}(x))$ (Fig. 3.3a), $2\arctan(x/2)$ (Fig. 3.3b), and $\arctan(x)$ (Fig. 3.3c) for a range of bead numbers n. These choices of θ respectively lead to *overdamped*, *critical*, and *Cayley* evolution of the thermostatted free ring polymer under PILE friction (see Section 3.2), and are identified accordingly in Fig. 3.3. The ring-polymer system considered in Fig. 3.3 approximates the O–H stretch dynamics in liquid water at room temperature with the parameters listed in Section 3.3. Velocity-marginals were initialized as in the setting of Theorem 6 (see Appendix 7.3), and the position of the jth ring-polymer mode at time $k\Delta t$ follows a centered normal distribution with variance $(\beta m_n)^{-1}(s^k_{j,\Delta t})^2$, where

$$(s^k_{j,\Delta t})^2 = (\mathscr{M}^k_{j,n})^2_{12} + \beta m_n \sum_{\ell=0}^{k-1} \left(\mathscr{M}^\ell_{j,n} \mathscr{R}_{j,n} (\mathscr{M}^\ell_{j,n})^{\mathrm{T}}\right)_{11} \qquad \text{for } k > 0. \tag{3.38}$$

The 2-Wasserstein distances in Fig. 3.3 were evaluated using a well-known analytical result for multivariate normal distributions.[215]

Figures 3.3a and 3.3c clearly show that the critical and overdamped schemes converge at dimension-free rates, but this is less evident from Fig. 3.3b for the Cayley scheme. The latter scheme nonetheless displays an n-independent, and hence dimension-free, distance to stationarity at all times $k\Delta t > 0$, indicated by plateauing of the contour lines towards the right of Fig. 3.3b. The ladder-like pattern that precedes this plateau illustrates a transition from geometric (i.e., fast) to subgeometric (i.e., slow) convergence upon introducing higher-frequency modes into the ring polymer. The transition manifests with the Cayley scheme because of its aggressive overdamping of the high-frequency modes, which is absent in the other two schemes (see Fig. 3.2).

The example considered in this section illustrates that the equilibration timescale (e.g., the time until the 2-Wasserstein distance decays below 10^{-6}) of the Cayley scheme at large n can dramatically exceed that of other BAOAB-like schemes. Although this negative feature may render the scheme impractical for pathological applications, we find in the next section that the Cayley scheme's superior configurational sampling provides compelling justification for its preferred use in realistic settings.

3.3 Numerical results

The current section provides numerical comparisons of the BAOAB-like T-RPMD integrators in Section 3.2, on applications featuring harmonic (Section 3.3) and anharmonic (Section 3.3) external potentials. Three representative choices of θ are considered in the numerical comparisons, namely $\theta(x) = \arctan(x)$, $\arccos(\mathrm{sech}(x))$, and $2\arctan(x/2)$. These choices respectively lead to *overdamped*, *critical*, and *Cayley* evolution of the thermostatted free ring polymer under PILE friction (Section 3.2), and are identified accordingly throughout the current section. It is borne out from the numerical comparisons that the Cayley scheme exhibits superior configurational sampling among the tested schemes in both applications.

One-dimensional quantum harmonic oscillator

In the current section, we numerically integrate Eq. 3.7 with the harmonic potential $V(q) = (\Lambda/2)q^2$ using PILE friction (i.e., $\Gamma = 2\Omega$), $m = 0.95$ amu, $\sqrt{\Lambda/m} = 3886\ \mathrm{cm}^{-1}$, and $T = 298$ K. This choice of physical parameters corresponds to a harmonic approximation of the Morse contribution to the O–H bond potential in the q-TIP4P/F force field for water,[216] and sets a least upper bound for the T-RPMD stability interval at $\Delta t^{\max} = 2/\sqrt{\Lambda/m} = 2.74$ fs. The simulations reported throughout this section employ the time-step $\Delta t = 0.73 \times \Delta t^{\max} = 2.00$ fs.

Figure 3.4 compares the accuracy and efficiency of various BAOAB-like T-RPMD schemes at equilibrium as a function of the bead number n. For a description of the numerical simulation and statistical estimation procedures used to generate the numerical data (filled circles) in Fig. 3.4, the reader is referred to Appendix 7.7. Figures 3.4a and 3.4c report the mean quantum kinetic energy at equilibrium as per the primitive and virial estimators,

$$\begin{aligned} \mathsf{KE}_n^{\mathrm{pri}}(\boldsymbol{q}) &= \frac{n}{2\beta} - \sum_{j=0}^{n-1} \frac{m_n \omega_n^2}{2} (q_{j+1} - q_j)^2 \ \text{ and} \\ \mathsf{KE}_n^{\mathrm{vir}}(\boldsymbol{q}) &= \frac{1}{2\beta} + \frac{1}{2} \sum_{j=0}^{n-1} (q_j - \overline{q})\, \partial_{q_j} V_n^{\mathrm{ext}}(\boldsymbol{q}) \,, \end{aligned} \tag{3.39}$$

where $\overline{q} = \frac{1}{n}\sum_{j=0}^{n-1} q_j$ is the centroid position of the n-bead ring polymer. For these two observables, Figs. 3.4b and 3.4d quantify the equilibrium sampling efficiency of the schemes in terms of the integrated autocorrelation time (or normalized asymp-

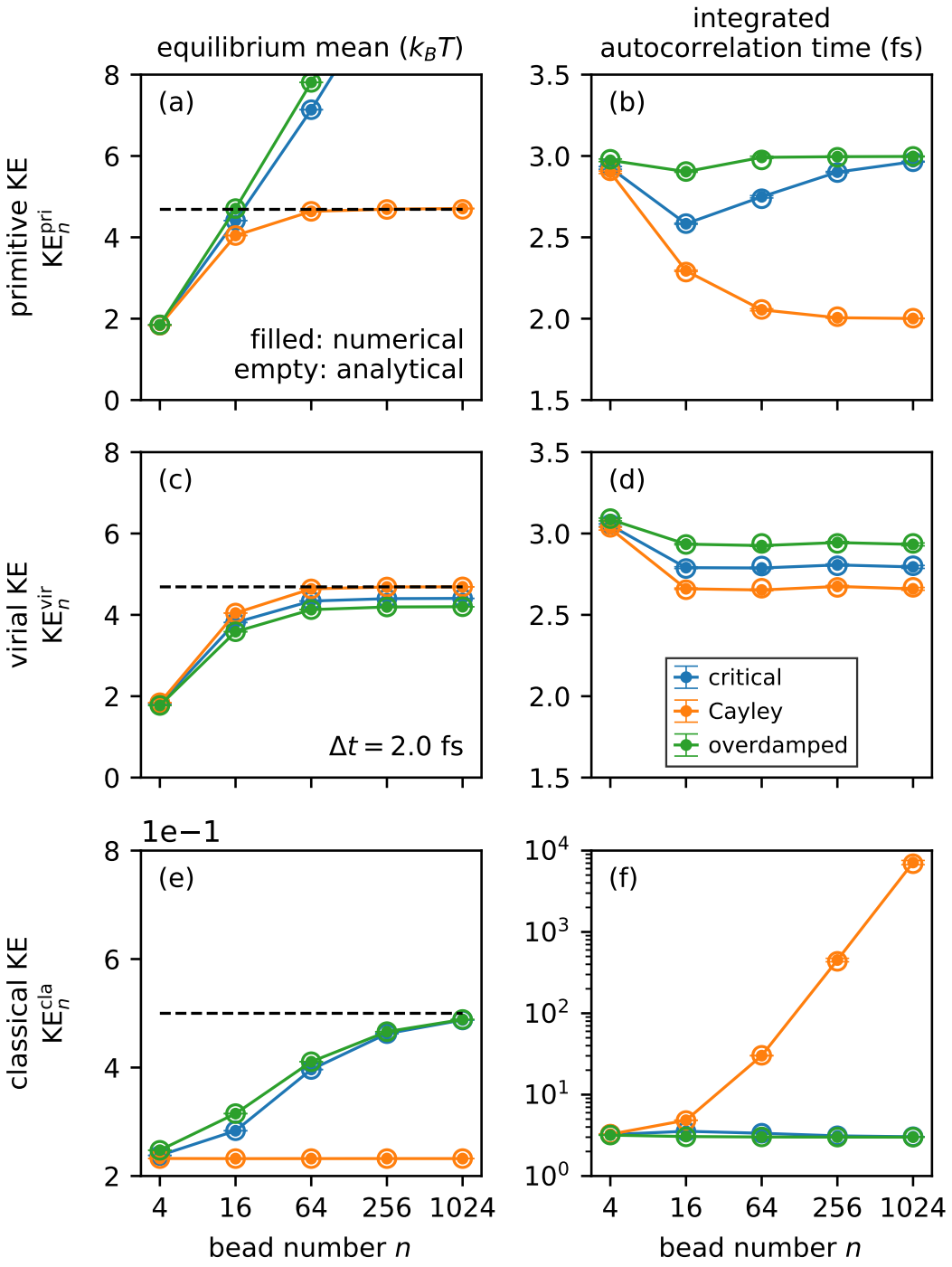


Figure 3.4: Performance at equilibrium of various BAOAB-like T-RPMD schemes applied to the one-dimensional quantum harmonic oscillator with physical parameters listed in Section 3.3. Panels (a), (c), and (e), respectively, plot the equilibrium mean primitive kinetic energy, virial kinetic energy, and non-centroid classical kinetic energy per mode as a function of bead number n; the corresponding means in the exact infinite bead limit are plotted as dashed lines. Panels (b), (d), and (f), plot the integrated autocorrelation times (Eq. 3.40) of the respective observables. Exact (resp. numerically estimated) values of the plotted quantities are shown with empty (resp. filled) circles. Numerical estimates were obtained using the protocol described in Appendix 7.7.

totic variance)[217–221]

$$\frac{\mathrm{aVar}(\mathsf{O}_n)}{\mathrm{Var}(\mathsf{O}_n)} = \frac{\lim_{K\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}\mathsf{O}_n(\boldsymbol{\xi}^{(k\Delta t)})\right)}{\mathrm{Var}\left(\mathsf{O}_n\right)} = 1+2\sum_{k=1}^{\infty}\mathrm{Cor}\left(\mathsf{O}_n(\boldsymbol{\xi}^{(0)}),\mathsf{O}_n(\boldsymbol{\xi}^{(k\Delta t)})\right), \tag{3.40}$$

where O_n is an n-bead observable, $\{\boldsymbol{\xi}^{(k\Delta t)}\}_{k=0}^{\infty} = \{(\boldsymbol{q}^{(k\Delta t)},\boldsymbol{v}^{(k\Delta t)})\}_{k=0}^{\infty}$ a stationary T-RPMD trajectory, $\mathrm{Var}(\mathsf{O}_n)$ the variance of O_n at equilibrium, and $\mathrm{Cor}(\mathsf{O}_n(\boldsymbol{\xi}^{(0)}),\mathsf{O}_n(\boldsymbol{\xi}^{(k\Delta t)}))$ the lag-$k\Delta t$ autocorrelation of O_n along the T-RPMD trajectory. The integrated autocorrelation time of O_n is interpreted as the timescale over which adjacent observations along an equilibrium trajectory become statistically uncorrelated[217–221] and is hence a measure of the efficiency of a T-RPMD scheme at estimating the mean of O_n with respect to the numerically sampled equilibrium distribution. Figures 3.4a-d show that the scheme specified by the Cayley angle (orange) outperforms others in terms of both accuracy and efficiency at estimating the equilibrium average of the quantum kinetic energy observables.

From the perspective of configurational accuracy, the optimality of the Cayley angle displayed in Figs. 3.4a and 3.4c is not surprising in light of the findings in Section 3.2. Less expected are the results in Figs. 3.4b and 3.4d, which suggest that the Cayley angle is also optimal from the standpoint of configurational sampling efficiency for the quantum kinetic energy observables in Eq. 3.39. Appendix 7.5 supports this conjecture with an analytical result for harmonic external potentials.

Figure 3.4e plots the mean classical kinetic energy at equilibrium as computed from the non-centroid ring-polymer velocities,

$$\mathsf{KE}_n^{\mathrm{cla}}(\boldsymbol{v}) = \frac{m_n}{2(n-1)}\sum_{j=0}^{n-1}(v_j^2-\bar{v}^2) \approx \frac{1}{2\beta}, \tag{3.41}$$

and Fig. 3.4f plots the corresponding integrated autocorrelation time as given by Eq. 3.40. For this observable, the equilibrium accuracy and efficiency of the Cayley scheme are significantly worse than those of the others as n increases. This is a consequence of the strongly overdamped behavior of Cayley T-RPMD at high frequencies (see Fig. 3.2), for which the integrator's ergodicity degrades as its spectral radius approaches unity. Note that this shortcoming of the Cayley scheme presents no adverse implications to the equilibrium sampling of observables that exclusively depend on the ring-polymer configuration, as confirmed by Figs. 3.4a-d.

In summary, Fig. 3.4 establishes that the T-RPMD scheme specified by the Cayley angle provides optimally accurate and efficient configurational sampling *at* equilibrium. To exploit this remarkable feature in practice, the scheme must manifest rapid converge *to* equilibrium when initialized away from it, as is necessary in most realistic applications of T-RPMD. Fortunately, Theorem 6 guarantees that any BAOAB-like scheme compliant with conditions (**C1**)-(**C4**) features a contractive configurational transition kernel for any number of ring-polymer beads, and Fig. 3.3 in Section 3.2 illustrates this fact for the quantum harmonic oscillator considered in the current section.

Room-temperature liquid water

While theoretical analysis and numerical tests of BAOAB-like T-RPMD schemes in previous sections have focused on harmonic external potentials, the current section demonstrates that the resulting insights carry over to a realistic, strongly anharmonic model of room-temperature liquid water. Our test system is a periodic box containing 32 water molecules at a temperature of 298 K and a density of 0.998 g/cm^3, with potential energy described by the q-TIP4P/F force field.[216] As in Section 3.3, we compare the performance of various BAOAB-like T-RPMD schemes for integrating the many-dimensional analogue of Eq. 3.7 with PILE friction, using the simulation time-step $\Delta t = 1.4$ fs in all simulations. Numerical tests reported in Appendix 7.6 show that this value of Δt closely approximates the upper limit of the Verlet (i.e., $n = 1$) stability interval for q-TIP4P/F liquid water. In agreement with Section 3.3, the experiments reveal that among the tested T-RPMD schemes, the Cayley scheme offers superior configurational sampling. For details on the numerical simulation and statistical estimation procedures used to generate the data presented in this section, the reader is referred to Appendix 7.8.

Figure 3.5 compares the equilibrium accuracy achieved by the tested schemes in terms of the quantum and classical kinetic energy per hydrogen atom (Figs. 3.5a, 3.5c, and 3.5i) and the intramolecular potential energy per water molecule (Figs. 3.5e and 3.5g); also plotted are the respective integrated autocorrelation times as a function of bead number n. The kinetic energy estimates in Figs. 3.5a and 3.5c exhibit similar trends to those seen in Fig. 3.4 for the one-dimensional harmonic oscillator. In particular, the T-RPMD scheme specified by the Cayley angle outperforms others in terms of quantum kinetic energy accuracy as n increases, most outstandingly with a highly accurate primitive kinetic energy estimate despite the large time-step employed. Still in close agreement with the harmonic oscillator results, Figs. 3.5b

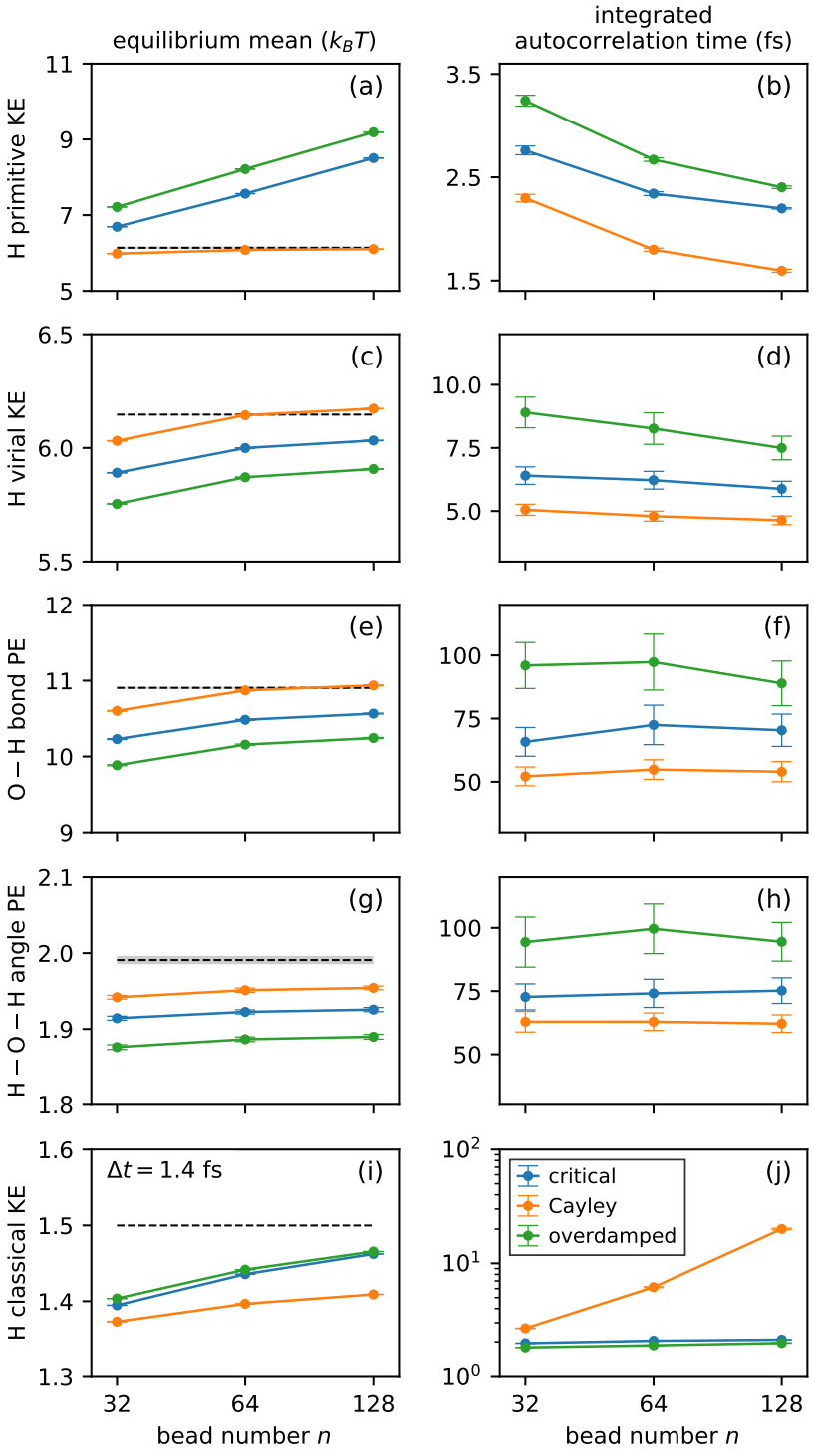


Figure 3.5: Performance of various BAOAB-like T-RPMD schemes applied to q-TIP4P/F liquid water at room temperature. As a function of the bead number n and for a 1.4-fs time-step, panels (a) and (c) plot the equilibrium kinetic energy per H atom as per the primitive and virial estimators (Eq. 3.39), and panels (b) and (d) plot the corresponding integrated autocorrelation times. Likewise, panels (e) and (g) plot the equilibrium potential energy per H_2O molecule due to the O−H-stretch and H−O−H-bend contributions, as defined in the q-TIP4P/F force field,[216] and the corresponding autocorrelation times are plotted by panels (f) and (h). Finally, panel (i) plots the classical kinetic energy per H atom computed from the non-centroid velocity estimator (Eq. 3.41), and panel (j) plots the corresponding autocorrelation time. The numerical estimates and reference results (dashed lines) were obtained using the protocols described in Appendix 7.8.

and 3.5d show that the Cayley scheme displays the shortest integrated autocorrelation time among the tested schemes for the quantum kinetic energy observables. Similar trends manifest in the intramolecular potential energy averages and their autocorrelation times (Figs. 3.5e-h), where the Cayley scheme also achieves superior accuracy and efficiency. Finally, Figs. 3.5i and 3.5j confirm that the relative performance of the compared schemes in terms of velocity-marginal sampling is qualitatively consistent with the harmonic results. Taken together, the results in Fig. 3.5 suggest that the superiority of the Cayley scheme for configurational sampling, proven in the model setting of a harmonic external potential, is also reflected in realistic applications.

In a final numerical test, Fig. 3.6 confirms that the sampling advantages of the Cayley T-RPMD scheme are obtained without downside in the estimation of dynamical quantities of typical interest. Specifically, Fig. 3.6b shows (unnormalized) infrared absorption spectra for room-temperature liquid water, computed from the 128-bead T-RPMD trajectories used to generate Fig. 3.5 using linear response theory and the T-RPMD approximation to real-time quantum dynamics.[183, 200] Linear response dictates that the absorption spectrum is proportional to $\omega^2 \tilde{\mathscr{I}}(\omega)$, where $\tilde{\mathscr{I}}(\omega) = \int_{\mathbb{R}} \mathrm{d}t\, e^{-i\omega t} \tilde{C}_{\mu\mu}(t)$ is the Fourier transform of the quantum-mechanical Kubo-transformed dipole autocorrelation function $\tilde{C}_{\mu\mu}(t)$. The latter is approximated within the T-RPMD framework[174, 181] by $\tilde{C}_{\mu\mu}(t) \approx \frac{1}{N_{\mathrm{H_2O}}} \sum_{i=1}^{N_{\mathrm{H_2O}}} \mathbb{E}\left(\overline{\mu}_i(t) \cdot \overline{\mu}_i(0)\right)$, where $N_{\mathrm{H_2O}}$ is the number of molecules in the liquid, $\overline{\mu}_i(t)$ is the bead-averaged dipole moment of molecule i at time t, and the covariance $\mathbb{E}\left(\overline{\mu}_i(t) \cdot \overline{\mu}_i(0)\right)$ is estimated from a stationary T-RPMD trajectory as indicated in Appendix 7.8. Figure 3.6a plots the T-RPMD estimates of $\tilde{C}_{\mu\mu}(t)$ leading to the absorption spectra in Fig. 3.6b. On the scale in which the absorption spectrum exhibits its key features, the spectra in Fig. 3.6b show very minor qualitative discrepancies. A similar conclusion holds for Fig. 3.6c, where the T-RPMD approximation of the Kubo-transformed velocity autocovariance function $\tilde{C}_{vv}(t) \approx \frac{1}{N_{\mathrm{H_2O}}} \sum_{i=1}^{N_{\mathrm{H_2O}}} \mathbb{E}\left(\overline{v}_i(t) \cdot \overline{v}_i(0)\right)$ is plotted for the three tested T-RPMD schemes. Collectively, these observations indicate that the accuracy of dynamical properties computed with BAOAB-like schemes is not significantly affected by the particular θ employed if conditions (**C1**)-(**C4**) in Section 3.2 are met. This result is expected due to the fact that the considered dynamical properties depend on bead-averaged (i.e., centroid-mode) coordinates, whose evolution is largely independent of the choice of θ under weak coupling between the centroid and non-centroid ring-polymer modes.

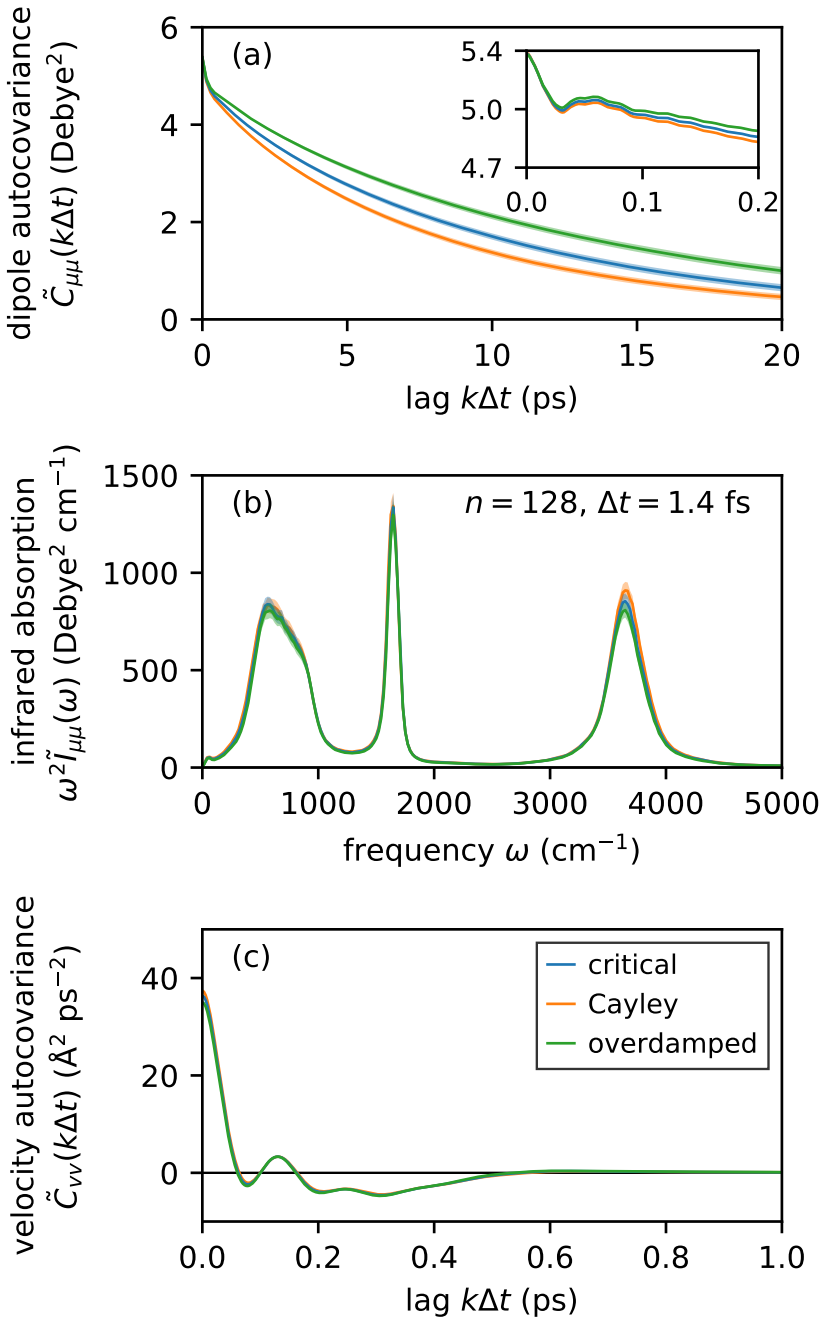


Figure 3.6: Molecular dipole autocovariance function (a), corresponding infrared absorption spectrum (b), and molecular velocity autocovariance function (c) in room-temperature liquid water for various BAOAB-like T-RPMD schemes. The plotted quantities autocovariance exhibit minor qualitative discrepancies across schemes, which suggests that all schemes compliant with conditions (**C1**)-(**C4**) exhibit comparable accuracy in the computation of dynamical properties. Numerical estimates of the autocovariance functions were obtained using the protocol described in Appendix 7.8.

3.4 Summary

Previous works showed that strong stability[203] and dimensionality freedom[205] are essential features of a robust T-RPMD integration scheme that standard integrators do not possess. A T-RPMD scheme with these features, denoted BCOCB, was introduced via a simple and inexpensive Cayley modification of the free ring-polymer update (i.e., the "A" sub-step) of the standard BAOAB integrator. The BCOCB scheme was then shown to dramatically outperform BAOAB at estimating static and dynamic properties of various systems with remarkable accuracy at unprecedented time-steps.[205]

The current work generalizes beyond the Cayley modification by introducing a simple parameterization of the free ring-polymer update and a corresponding family of strongly stable and dimension-free modifications of the BAOAB scheme. Among these schemes lies BCOCB, which is found to exhibit superior configurational sampling despite exhibiting worse accuracy and efficiency for observables that depend on the non-centroid ring-polymer velocities. This conclusion is obtained theoretically via exhaustive analysis of a harmonic model, and numerically via simulation of a realistic quantum-mechanical model of liquid water at room temperature. In this way, the current work convincingly demonstrates the superiority of the BCOCB scheme for accurate and efficient equilibrium simulation of condensed-phase systems with T-RPMD.

To conclude, we stress that implementing BCOCB or any of the new dimension-free and strongly-stable schemes leads to no additional cost, parameters or coding overhead relative to the standard BAOAB integrator. The modified integrators thus provide "turnkey" means to significantly improve the accuracy and stability of existing (T-)RPMD implementations.[222, 223]

Chapter 4

TRANSPORT AND NOISE OF HOT ELECTRONS IN GAAS USING A SEMIANALYTICAL MODEL OF TWO-PHONON POLAR OPTICAL PHONON SCATTERING

This chapter is adapted from

1. Sun, J. & Minnich, A. J. Transport and noise of hot electrons in GaAs using a semianalytical model of two-phonon polar optical phonon scattering. *Phys. Rev. B* **107,** 205201. https://link.aps.org/doi/10.1103/PhysRevB.107.205201 (20 May 2023).
   Contribution: J.S. participated in the conception of the project, designed and implemented the algorithm, performed the numerical simulation, analyzed the results, and wrote the article.

Recent ab-initio studies of electron transport in GaAs have reported that electron-phonon (e-ph) interactions beyond the lowest order play a fundamental role in charge transport and noise phenomena. Inclusion of the next-leading-order process in which an electron scatters with two phonons was found to yield good agreement for the high-field drift velocity, but the characteristic non-monotonic trend of the power spectral density of current fluctuations (PSD) with electric field was not predicted. The high computational cost of the ab-initio approach necessitated various approximations to the two-phonon scattering term, which were suggested as possible origins of the discrepancy. Here, we report a semi-analytical transport model of two-phonon electron scattering via the Fröhlich mechanism, allowing a number of the approximations in the ab-initio treatment to be lifted while retaining the accuracy to within a few percent. We compare the calculated and experimental transport and noise properties as well as scattering rates measured by photoluminescence experiments. We find quantitative agreement within 15% for the drift velocity and 25% for the Γ valley scattering rates, and agreement with the $\Gamma - L$ intervalley scattering rates within a factor of two. Considering these results and prior studies of current noise in GaAs, we conclude that the most probable origin of the non-monotonic PSD trend versus electric field is the formation of space charge domains rather than intervalley scattering as has been assumed.

4.1 Introduction

Electron transport in semiconductors is of fundamental interest and of high relevance for microelectronic devices [224–226]. The upper limit for the mobility of a semiconductor is governed by scattering of electrons by phonons. Early studies of charge transport properties employed a semi-empirical description of the band structure and electron-phonon scattering. The introduction of the Monte Carlo (MC) method allowed for the numerical simulation of transport with fewer approximations [227]. Later, full-band MC tools capable of simulating realistic device

geometries were developed, but the treatment of the e-ph scattering rates in general required fitting parameters. [228–230]. The development of the ab-initio description of the electron-phonon interactions based on density functional theory (DFT), density functional perturbation theory (DFPT) and Wannier interpolation has enabled the parameter-free computation of low-field charge transport properties such as mobility [231–233]. These methods have now been applied to a range of semiconductors, including Si [234, 235], GaN [236], GaAs [237], two-dimensional materials [238–240], and others. Recent methodology developments, including two-phonon scattering [241], quadrupole interactions [242, 243], and *GW* corrections [235, 244], have facilitated a rigorous comparison of the accepted level of theory with experiment.

The accuracy of the ab-initio calculations has been mainly based on comparison to low-field mobility values. Recent works have extended these calculations to beyond low-field transport and noise properties. [245–248] In GaAs, it was found that although the qualitative shape of the drift velocity versus electric field curve was predicted correctly compared to experiment, the magnitude of the drift velocity was overpredicted by about 50%. [247] The inclusion of the next-leading-order term of scattering involving two phonons (2ph) yielded a low-field mobility and drift velocity with substantially improved agreement. However, the characteristic nonmonotonic trend of PSD with electric field was not predicted even with the 2ph theory. Owing to the high cost of the ab-initio calculations, the treatment of 2ph processes in that work required several approximations, such as the neglect of off-shell 2ph scattering processes. Whether these neglected processes or other numerical considerations can account for the PSD discrepancy remains unknown.

Here, we introduce a semi-analytical model for both 1ph and 2ph e-ph scattering via the Fröhlich mechanism, allowing the full 2ph scattering term to be treated over the wide range of energies needed for high-field transport while introducing error on the order of only a few percent. We find that the transport and noise properties are qualitatively unchanged compared to the ab-initio calculations. The calculated scattering rates agree with those obtained from photoluminescence experiments to within 25% for the Γ valley rates and a factor of two for the Γ-L intervalley rates. Despite this degree of agreement, the qualitative discrepancy observed previously for the PSD remains. We consider the remaining approximations in the semi-analytical model and find that they are unlikely to account for the PSD discrepancy. Therefore, we conclude that the characteristic peak in the PSD with electric field most likely arises

from the formation of space charge domains rather than intervalley scattering as has been assumed in the literature. This finding has implications for the use of transport measurements to study intervalley scattering.

4.2 Theory

Overview of formalism for charge transport and noise properties

We first review the ab-initio treatment of electron transport and electronic noise using the BTE [249]. For a spatially homogeneous system with electric field $\mathscr{E}$, the electron distribution function $f_{m\boldsymbol{k}}$ is governed by

$$\frac{\partial f_{m\boldsymbol{k}}}{\partial t} + \frac{e\mathscr{E}}{\hbar} \cdot \nabla_{\boldsymbol{k}} f_{m\boldsymbol{k}} = \mathscr{I}[f_{m\boldsymbol{k}}], \tag{4.1}$$

where $f_{m\boldsymbol{k}}$ is the electron occupation in the state with band index m and wave vector $\boldsymbol{k}$, e is the fundamental charge, and $\mathscr{I}[f_{m\boldsymbol{k}}]$ is the collision term, which describes the scattering of electrons with phonons [231]. In this work, index m is dropped in all the following derivations for simplicity as only one band is relevant for electron transport in GaAs in the range of electric fields considered.

For non-degenerate electrons, the collision term can be linearized as [247]

$$\mathscr{I}[f_{\boldsymbol{k}'}] \approx -\sum_{\boldsymbol{k}} \Theta_{\boldsymbol{k}',\boldsymbol{k}} f_{\boldsymbol{k}}, \tag{4.2}$$

where $\Theta_{\boldsymbol{k}',\boldsymbol{k}}$ is e-ph collision matrix. For consistency with our prior work [247], the diagonal and off-diaonal elements of this matrix are positive and negative, respectively. The diagonal elements $\Theta_{\boldsymbol{k},\boldsymbol{k}}$ are equal to the total scattering rates as $\Theta_{\boldsymbol{k},\boldsymbol{k}} = \Gamma_{\boldsymbol{k}} = -\sum_{\boldsymbol{k}' \neq \boldsymbol{k}} \Theta_{\boldsymbol{k}',\boldsymbol{k}}$. The relative error of the linearization is on the order of $f_{\boldsymbol{k}}$ according to Eq. 3 in Ref. [247]. Due to the non-degenerate carrier concentration (10^{15} cm^{-3}) used in this work, we find that $f_{\boldsymbol{k}} \lesssim 4 \times 10^{-3}$, which implies that the linearization error can be neglected. With this linearization and a finite difference representation of the derivative operator $\nabla_{\boldsymbol{k}}$ [246], Eq. 4.1 becomes a linear partial-differential equation which can be solved by numerical linear algebra. The equation for the steady distribution function $f^s_{\boldsymbol{k}}$ is given by Eq. 6 of Ref. [246]. Steady-state mean transport properties such as drift velocity can be calculated with the appropriate Brillouin zone sum using this distribution.

The current power spectral density (PSD) is used to characterize fluctuations in occupation about the mean distribution. The PSD is defined as the Fourier transform

of the autocorrelation of the current density fluctuations (Eq. 12 of Ref. [246]). Following Ref. [246], the current PSD at frequency ω can be calculated as

$$S_{j_\alpha j_\beta}(\omega) = 2\left(\frac{2e}{\mathscr{V}_0}\right)^2 \Re\left[\sum_{\mathbf{k}} v_{k\alpha} g_{k\beta}\right], \tag{4.3}$$

where $v_{k\alpha}$ is the group velocity of the electron at state $\boldsymbol{k}$ along axis α, $\mathscr{V}_0$ is the supercell volume, and $g_{k\beta}$ is the effective distribution function [250]. Note that the effective distribution function is distinct from the e-ph matrix elements, $g_\nu(\boldsymbol{k},\boldsymbol{q})$. $g_{k\beta}$ satisfies the following equation:

$$\sum_{\boldsymbol{k}} (\Theta_{\boldsymbol{k}',\boldsymbol{k}} + \frac{e\mathscr{E}}{\hbar}\cdot \boldsymbol{D}_{\boldsymbol{k}',\boldsymbol{k}} + i\omega\delta_{\boldsymbol{k}',\boldsymbol{k}}) g_{k\alpha} = f^s_{\boldsymbol{k}'}(v_{\boldsymbol{k}',\alpha} - V_\alpha), \tag{4.4}$$

where D is the finite difference representation of $\nabla_{\boldsymbol{k}}$ (see Eq. 4 in Ref. [246]), $f^s_{\boldsymbol{k}}$ is the steady-state occupation for the state at wave vector $\boldsymbol{k}$, and $V_\alpha = \sum_{\boldsymbol{k}} v_{k\alpha} f^s_{\boldsymbol{k}} / \sum_{\boldsymbol{k}} f^s_{\boldsymbol{k}}$ is the drift velocity along axis α.

The e-ph collision matrix is obtained from perturbation theory in orders of the e-ph interaction strength. The first two orders in the expansion correspond to scattering with one phonon (1ph) and two phonons (2ph). The corresponding Feynman diagrams can be found in Fig. 1 and Fig. 2 in Ref. [241]. For 1ph scattering of non-degenerate electrons, the non-diagonal scattering matrix elements are given by [247]:

$$\Theta^{(1\mathrm{ph})}_{\boldsymbol{k}'=\boldsymbol{k}+\boldsymbol{q},\boldsymbol{k}} = -\frac{2\pi}{\hbar}\frac{1}{N}\sum_{\nu} |g_\nu(\boldsymbol{k},\boldsymbol{q})|^2 \left(\delta(\varepsilon_{\boldsymbol{k}} - \hbar\omega_{\nu\boldsymbol{q}} - \varepsilon_{\boldsymbol{k}+\boldsymbol{q}}) N_{\boldsymbol{q}\nu} + \delta(\varepsilon_{\boldsymbol{k}} + \hbar\omega_{\nu\boldsymbol{q}} - \varepsilon_{\boldsymbol{k}+\boldsymbol{q}})(N_{\boldsymbol{q}\nu}+1)\right) \tag{4.5}$$

for $\boldsymbol{k} \neq \boldsymbol{k}'$, where $g_\nu(\boldsymbol{k},\boldsymbol{q})$ is the e-ph scattering matrix element, $\varepsilon_{\boldsymbol{k}}$ is the energy of the electronic state $\boldsymbol{k}$, $\omega_{\nu\boldsymbol{q}}$ is the frequency of phonon with mode ν and wave vector $\boldsymbol{q}$, N is the total number of $\boldsymbol{k}$ points in the Brillouin zone, $N_{\boldsymbol{q}\nu}$ is the phonon occupation according to the Bose-Einstein statistics, and the two delta functions are energy conservation conditions for the emission and absorption subprocesses, respectively. We note that here we neglect the dependence of the phonon distribution on the external electric field. For sufficiently high electric fields and currents, the dissipated power from Joule heating can increase the phonon occupation above its thermal value, which is known as the hot phonon effect [251, 252]. In this work and in relevant experiments used for comparison, this effect is negligible due to the small non-degenerate carrier density ($\sim 10^{15}$ cm^{-3}).

Two-phonon scattering

The collision integral for 2ph scattering was derived in Ref. [241], and the linearized form is given in Ref. [247]. 2ph scattering can be divided into the two-emission (2e), one-emission-one-absorption (1e1a) and two-absorption (2a) subprocesses. Here, we rewrite the original formalism in Ref. [241] to facilitate the derivations in the next section. Assuming non-degenerate electron statistics in Eq. 4 in Ref. [241], we rewrite Eq. 12 in Ref. [247] by splitting $W^{(i)}$ into two parts:

$$\begin{aligned} W^{(i)} &= |\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2} + \tilde{W}_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q},\alpha_1}|^2 \\ &= \left[|\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2}|^2 + \mathrm{Re}(\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2}\tilde{W}^*_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q},\alpha_1})\right] + \left[|\tilde{W}_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q},\alpha_1}|^2 + \mathrm{Re}(\tilde{W}_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q},\alpha_1}\tilde{W}^*_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2})\right], \end{aligned} \tag{4.6}$$

where

$$\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha} = \frac{g_\nu(\boldsymbol{k},\boldsymbol{q})g_\mu(\boldsymbol{k}+\boldsymbol{q},\boldsymbol{p})}{\varepsilon_{\boldsymbol{k}'} - \varepsilon_{\boldsymbol{k}+\boldsymbol{q}} + \alpha\hbar\omega_{\nu\boldsymbol{p}} + i\eta - i\hbar\Gamma_{\boldsymbol{k}+\boldsymbol{q}}/2}, \tag{4.7}$$

where η is a positive infinitesimal, $\Gamma_{\boldsymbol{k}+\boldsymbol{q}} = \Gamma^{(1\mathrm{ph})}_{\boldsymbol{k}+\boldsymbol{q}} + \Gamma^{(2\mathrm{ph})}_{\boldsymbol{k}+\boldsymbol{q}}$ is the total scattering rate of the intermediate state $\boldsymbol{k}+\boldsymbol{q}$, and $\hbar\Gamma_{\boldsymbol{k}+\boldsymbol{q}}/2$ is the imaginary part of the self-energy. As is usually assumed for transport calculations, the Debye-Waller contribution to the self-energy is not included due to its computational difficulty.

For the 2e and 2a subprocesses, the two terms give the same contribution after the summation. For the 1e1a subprocess, the two terms in Eq. 4.6 physically represent the emission-then-absorption (a-e) and absorption-then-emission (e-a) subprocesses, respectively. Finally, we exchange the summation order of $\boldsymbol{q}$ and $\boldsymbol{p}$ in Eq. 9 in Ref. [247] for the second term of Eq. 4.6 and arrange the equations to obtain:

$$\Theta^{(2\mathrm{ph})}_{\boldsymbol{k}',\boldsymbol{k}} = -\frac{2\pi}{\hbar}\frac{1}{N^2}\sum_{\alpha_1=\pm1}\sum_{\alpha_2=\pm1}\sum_{\boldsymbol{q}+\boldsymbol{p}=\boldsymbol{k}'-\boldsymbol{k}}\sum_{\nu\mu}\tilde{\Theta}^{(\alpha_1,\alpha_2)}_{\boldsymbol{k},\boldsymbol{q}\nu,\boldsymbol{p}\mu} \tag{4.8}$$

for $\boldsymbol{k} \neq \boldsymbol{k}'$, where α_1 and α_2 indicate whether the first and second phonon is emitted ($\alpha_{1,2} = 1$) or absorbed ($\alpha_{1,2} = -1$), so that the four combinations of $\alpha_{1,2} = \pm 1$ describe the four subprocesses.

The term $\tilde{\Theta}^{(\alpha_1,\alpha_2)}_{\boldsymbol{k},\boldsymbol{q}\nu,\boldsymbol{p}\mu}$ in Eq. 4.8 is defined as

$$\begin{aligned} \tilde{\Theta}^{(\alpha_1,\alpha_2)}_{\boldsymbol{k},\boldsymbol{q}\nu,\boldsymbol{p}\mu} =& (N_{\boldsymbol{q}\nu} + \delta_{\alpha_1,1})(N_{\boldsymbol{p}\mu} + \delta_{\alpha_2,1})\left[|\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2}|^2 + \mathrm{Re}(\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2}\tilde{W}^*_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q},\alpha_1})\right] \times \\ & \delta(\varepsilon_{\boldsymbol{k}} - \varepsilon_{\boldsymbol{k}'} - \alpha_1\omega_{\nu\boldsymbol{q}} - \alpha_2\omega_{\mu\boldsymbol{p}}). \end{aligned} \tag{4.9}$$

As the 2ph scattering rates depend on the intermediate state rates, the 2ph scattering must be calculated iteratively. In each iteration, the intermediate state rates of the previous iteration are used to update the 2ph scattering matrix $\Theta^{(2\text{ph})}_{\boldsymbol{k}',\boldsymbol{k}}$ and scattering rates $\Gamma^{(2\text{ph})}_{\boldsymbol{k}}$.

Compared with 1ph scattering, 2ph scattering is much more computationally expensive, particularly for high-field transport which requires a larger energy window than for low-field transport. In Ref. [247], several approximations were made to make the computation feasible, including limiting the number of self-consistent iterations of the 2ph rates to three, restricting the off-shell extent $|\varepsilon_{\boldsymbol{k}'} - \varepsilon_{\boldsymbol{k}+\boldsymbol{q}} + \alpha\hbar\omega_{\nu\boldsymbol{p}}|$ to 25 meV, and neglecting the interference term $\mathrm{Re}(\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2}\tilde{W}^{*}_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q},\alpha_1})$ term in Eq. 4.9. Additionally, the maximum grid density that could be used was $200 \times 200 \times 200$. The effect of these approximations on the observable transport and noise properties was not assessed. In particular, the on-shell approximation neglects off-shell processes and thus underestimates the scattering rates. These approximations were mentioned as possible reasons for the PSD discrepancy in Ref. [247].

Semi-analytical model for 1ph and 2ph Γ-Γ scattering

In this section, we introduce a semi-analytical model to treat 1ph and 2ph $\Gamma - \Gamma$ intravalley scattering by the Fröhlich interaction that retains the accuracy of the ab-initio formalism to within a few percent while allowing the approximations described above to be lifted. This model is based on the fact that over the range of wavevectors considered in this study, the Γ valley in GaAs is nearly spherically symmetric, and Γ-Γ scattering can be accurately described by using only the Fröhlich interaction [253]. The model is valid only for Γ intravalley scattering because Γ-L intervalley scattering lacks an analytic description of similar accuracy. We also note that this model is valid only for materials in which the Fröhlich mechanism makes the dominant contribution to e-ph scattering.

The semi-analytical model uses the following approximations. First, the band structure is described using the Kane model [254] for a spherically symmetric, nonparabolic band. This description is accurately satisfied for the Γ valley, with the Kane model bands deviating from the ab-initio band structure by at most 7% over the range of wave vectors considered ($\sim 0.1G$, where G is the reciprocal lattice constant). Second, prior works have shown that Γ-Γ e-ph scattering in GaAs is dominated by longitudinal optical (LO) phonons via the Fröhlich interaction [255]. We

therefore neglect scattering processes involving other phonon branches and scattering mechanisms. The computed matrix elements $g_{\mathrm{LO}}(\boldsymbol{k},\boldsymbol{q})$ for Fröhlich scattering are found to exhibit negligible anisotropy so that $g_{\mathrm{LO}}(\boldsymbol{k},\boldsymbol{q}) = g_{\mathrm{LO}}(q)$, enabling an analytic form of $g^{\mathrm{fit}}_{\mathrm{LO}}(q)$ to be fit to the ab initio values as described in Sec. 4.3. In the range of wave vectors considered, this approximation is satisfied to within 3% [253]. A detailed comparison of ab-initio and semi-analytical band structure and e-ph matrix elements can be found in Appendix 8.1. Third, we take the LO phonon frequency to be a constant $\omega_{\mathrm{LO}} = 35$ meV. In the range of phonon wave vectors $q \in (0, 0.2G)$ considered here, this assumption is satisfied to within less than 0.3%.

We now discuss the treatment of 1ph and 2ph e-ph scattering based on these simplifications. The summation in Eq. 4.8 may be rewritten as an integral in the Brillouin zone over the intermediate wave vector $\boldsymbol{k}_m$ by letting $\boldsymbol{q} \to \boldsymbol{k}_m - \boldsymbol{k}$ and $\boldsymbol{p} \to \boldsymbol{k}' - \boldsymbol{k}_m$. Additionally, we exploit spherical symmetry to rewrite all the quantities in spherical coordinates as: $\Theta_{\boldsymbol{k}',\boldsymbol{k}} = \Theta(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'})$ and $\Gamma_{\boldsymbol{k}} = \Gamma_k$. After some simplifications, we obtain the 1ph and 2ph collision matrices as

$$\Theta^{(1\mathrm{ph})}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'}) = \frac{2\pi}{\hbar}\frac{1}{\Omega_{\mathrm{BZ}}}|g^{\mathrm{fit}}_{\mathrm{LO}}(|\boldsymbol{k}'-\boldsymbol{k}|)|^2 \sum_{\alpha=\pm1} A_\alpha \delta(\varepsilon_k - \alpha\hbar\omega_{\mathrm{LO}} - \varepsilon_{k'}) \quad (4.10)$$

and

$$\Theta^{(2\mathrm{ph})}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'}) = \frac{2\pi}{\hbar}\frac{1}{\Omega^2_{\mathrm{BZ}}} \sum_{\alpha_1=\pm1}\sum_{\alpha_2=\pm1} A_{\alpha_1}A_{\alpha_2}\delta(\varepsilon_k - \varepsilon_{k'} - (\alpha_1+\alpha_2)\hbar\omega_{\mathrm{LO}}) I^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'}), \quad (4.11)$$

where Ω_{BZ} is the Brillouin zone volume. A_α is the phonon occupation factor defined as

$$A_\alpha = N_{\mathrm{LO}} + \delta_{\alpha,+1}, \quad (4.12)$$

where $N_{\mathrm{LO}} = (\exp(\hbar\omega_{\mathrm{LO}}/k_BT) - 1)^{-1}$ is the LO phonon occupation, and $I^{(\alpha_1,\alpha_2)} = I_1^{(\alpha_1,\alpha_2)} + I_2^{(\alpha_1,\alpha_2)}$ is decomposed to the non-interference part $I_1^{(\alpha_1,\alpha_2)}$ and the interference part $I_2^{(\alpha_1,\alpha_2)}$:

$$\begin{aligned} I_1^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'}) &= \int |\tilde{W}_{\boldsymbol{k},\boldsymbol{k}_m-\boldsymbol{k},\boldsymbol{k}'-\boldsymbol{k}_m,\alpha_2}|^2 \mathrm{d}^3\boldsymbol{k}_m \\ &= \int \left| \frac{g^{\mathrm{fit}}_{\mathrm{LO}}(|\boldsymbol{k}_m-\boldsymbol{k}|) g^{\mathrm{fit}}_{\mathrm{LO}}(|\boldsymbol{k}'-\boldsymbol{k}_m|)}{\varepsilon_{k'} - \varepsilon_{k_m} + \alpha_2\omega_{\mathrm{LO}} + i\eta - \hbar\Gamma_{k_m}/2} \right|^2 \mathrm{d}^3\boldsymbol{k}_m, \end{aligned} \quad (4.13)$$

and

$$I_2^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'}) = \int_{\boldsymbol{q}+\boldsymbol{p}=\boldsymbol{k}'-\boldsymbol{k}} \mathrm{Re}(\tilde{W}_{\boldsymbol{k},\boldsymbol{q},\boldsymbol{p},\alpha_2}\tilde{W}^*_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q},\alpha_2})\mathrm{d}^3\boldsymbol{p}\mathrm{d}^3\boldsymbol{q}. \tag{4.14}$$

Equation 4.13 can be further simplified by writing the integration in spherical coordinates and separating the radius and angular part:

$$I_1^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'}) = \int \frac{\tilde{I}^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'},k_m)}{|\varepsilon_{k'}-\varepsilon_{k_m}+\alpha_2\omega_{\mathrm{LO}}+i\eta-\hbar\Gamma_{k_m}/2|^2}k_m^2\mathrm{d}k_m, \tag{4.15}$$

where $\tilde{I}_1^{(\alpha_1,\alpha_2)}$ is the angular part defined as

$$\tilde{I}_1^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'},k_m) = \int |g_{\mathrm{LO}}^{\mathrm{fit}}(|\boldsymbol{k}_m-\boldsymbol{k}|)g_{\mathrm{LO}}^{\mathrm{fit}}(|\boldsymbol{k}'-\boldsymbol{k}_m|)|^2\sin\theta_{\boldsymbol{k}_m}\mathrm{d}\theta_{\boldsymbol{k}_m}\mathrm{d}\phi_{\boldsymbol{k}_m}, \tag{4.16}$$

where $\theta_{\boldsymbol{k}_m}$, $\phi_{\boldsymbol{k}_m}$ are the polar angle and azimuthal angle defining the intermediate wave vector $\boldsymbol{k}_m$, respectively. Since $\tilde{I}_1^{(\alpha_1,\alpha_2)}$ is independent of the band structure and the self-energy, recomputation of this term in each 2ph iteration is not required. In practice, to significantly reduce the computational cost, $\tilde{I}_1^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'},k_m)$ is precomputed on a grid of k, k', $\theta_{\boldsymbol{k},\boldsymbol{k}'}$ and k_m before the 2ph iteration. We note that k, k', and $\theta_{\boldsymbol{k},\boldsymbol{k}'}$ are not independent of each other due to the energy conservation condition. Once $\tilde{I}_1^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'},k_m)$ is computed on a predefined grid, $\Theta^{(2\mathrm{ph})}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'})$ can be calculated according to Eqs. 4.11, 4.15 and 4.14. We note that such separation of the radius and spherical part is not valid for $\tilde{I}_2^{(\alpha_1,\alpha_2)}$, so an expensive iterative update is required. However, since $\tilde{I}_2^{(\alpha_1,\alpha_2)}$ is generally much smaller compared with $\tilde{I}_2^{(\alpha_1,\alpha_2)}$, we update $\tilde{I}_2^{(\alpha_1,\alpha_2)}$ every 10 iterations to decrease the cost of the self-consistent calculations.

To complete the 2ph iteration, the last quantities to be computed are the total 1ph and 2ph scattering rates:

$$\Gamma_k^{(\mathrm{type})} = \int \Theta^{(\mathrm{type})}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'})\mathrm{d}^3\boldsymbol{k}' = \int \Theta^{(\mathrm{type})}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'})2\pi k'^2\mathrm{d}k'\sin\theta_{\boldsymbol{k},\boldsymbol{k}'}\mathrm{d}\theta_{\boldsymbol{k},\boldsymbol{k}'}, \tag{4.17}$$

where type $=$ 1ph, 2ph indicates the type of scattering. We also perform the radius integration over k' analytically to integrate the delta functions in Eq. 4.10 and 4.11. The angular integrations in Eq. 4.17 are performed numerically; details are provided in Sec. 4.3.

The computational flow of the semi-analytical model is as follows. First, we generate a grid of k, $\theta_{\boldsymbol{k},\boldsymbol{k}'}$, k_m and calculate the corresponding k' from the energy conservation conditions for each subprocess. Second, we calculate $\Theta^{(1\mathrm{ph})}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'})$

by Eq. 4.10 and $\Gamma_k^{(1\mathrm{ph})}$ by Eq. 4.17. Then, we calculate $\tilde{I}_1^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'},k_m)$ by Eq. 4.16. Finally, we perform the self-consistent 2ph iterations using Eqs. 4.15, 4.14, 4.11 and 4.17 until convergence, where Eq. 4.14 is calculated every 10 iterations.

From the perspective of computational cost, the semi-analytical model reduces the number of integration variables in the 2ph scattering rate calculation from 9 (integration over $\boldsymbol{k}$, $\boldsymbol{k}'$, $\boldsymbol{k}_m$) to 5 (k, $\theta_{\boldsymbol{k},\boldsymbol{k}'}$, $\boldsymbol{k}_m$) due to the spherical symmetry, and avoids the recomputation of $I_1^{(\alpha_1,\alpha_2)}$ in the 2ph self-consistent iterations due to the separation of radius and angular integration in Eq. 4.15. This reduction allows for the use of a denser grid for the intermediate state integration and thereby reduces the discretization error. Therefore, the total scattering rates can be calculated with negligible discretization error compared with the ab-initio calculation. However, since the semi-analytical model is only for Γ-Γ scattering, a discretized scattering matrix ($\Theta_{\boldsymbol{k}',\boldsymbol{k}}$) is still needed to compute the drift velocity and current PSD, which are affected by Γ-L intervalley scattering. As a result, the discretization error in the final state integration cannot be avoided for the present calculations. Nevertheless, the semi-analytical model still decreases the discretization error of the intermediate state integration and treats the full 2ph scattering term. The differences between the ab-initio calculation and the semi-analytical model for Γ intravalley scattering are summarized in Table 4.1.

	Ab-initio calculation	Semi-analytical model
Final state integration	200^3	200^3 for observables Exact for scattering rates
Intermediate state integration	200^3	Exact
Processes	On-shell only	All processes included
Two-phonon iterations	3	Iterate until convergence
Interference term	Not included	Included
Computational time (CPU hours)	50000	40

Table 4.1: Comparison between the ab-initio calculation and the semi-analytical model for $\Gamma-\Gamma$ intravalley scattering. The semi-analytical model improves upon the ab-initio model in all respects except the final state integration grid density for observables, for which the same grid is used.

4.3 Computational details

Ab-initio calculations

The ab-initio calculation parameters are identical to those in our previous work [247]. In brief, electronic structure and e-ph matrix elements are computed us-

ing Density Functional Theory (DFT) and Density Functional Perturbation Theory (DFPT) with QUANTUM ESPRESSO [256, 257] with an $8 \times 8 \times 8$ coarse grid, 72 Ry plane wave energy cutoff, a relaxed lattice parameter of 5.556 Å, and a non-degenerate carrier concentration of 10^{15} cm^{-3}. Following our previous work [247], we apply a band structure correction for both the Γ valley and the L valley. For the Γ valley, we use a spherically symmetric Kane model band structure [258] with an experimental effective mass of $0.067m_e$ and a non-parabolicity of 0.64 eV^{-1} [225]. For the L valley, we shift the energy by 50 meV to achieve the experimental Γ-L valley separation of 300 meV. Wannier interpolation in PERTURBO [259] is then applied to interpolate the e-ph matrix elements to a finer grid of $200 \times 200 \times 200$. After the e-ph matrix elements are obtained, the 1ph and 2ph scattering matrices are computed according to Eqs. 4.5 and 4.8. The delta functions in Eqs. 4.5 and 4.9 are approximated by a Gaussian function with a standard deviation of 5 meV.

During the computation of scattering matrices, a phonon frequency cutoff of 2 meV is applied to neglect phonons with low frequencies. An energy cutoff of 360 meV above the conduction band minimum is used to reduce the number of k points in the Brillouin zone integration. The 2ph calculation applies the on-shell approximation by restricting the off-shell extent to 25 meV. Following Ref. [246], the linear system of equations representing the Boltzmann equation is solved by the generalized minimal residual method (GMRES).

Semi-analytical model

The band structure used in the semi-analytical model is the same as that in the ab-initio calculation. The LO phonon energy is taken to be $\omega_{\mathrm{LO}} = 35$ meV. The function $g^{\mathrm{fit}}_{\mathrm{LO}}(q)$ is obtained by a weighted averaged of $g_{\mathrm{LO}}(\boldsymbol{k}_i, \boldsymbol{p}_i)$:

$$g^{\mathrm{fit}}_{\mathrm{LO}}(q) = \frac{\sum_i g_{\mathrm{LO}}(\boldsymbol{k}_i, \boldsymbol{p}_i) \exp(-\frac{(q-p_i)^2}{2\sigma^2})}{\sum_i \exp(-\frac{(q-p_i)^2}{2\sigma^2})}, \tag{4.18}$$

where the standard deviation $\sigma = 5 \times 10^{-4}$ Ry, the summation is over all the on-shell processes, and the $g_{\mathrm{LO}}(\boldsymbol{k}_i, \boldsymbol{p}_i)$ are calculated by the Wannier interpolation. The spherical coordinates integration in Eq. 4.16 is defined such that the the $\theta_{\boldsymbol{k}_m} = 0$ direction is orthogonal with both $\boldsymbol{k}$ and $\boldsymbol{k}'$. 200 grid points are used for each $\theta_{\boldsymbol{k}_m}$ and $\phi_{\boldsymbol{k}_m}$ integration. The radius integration in Eq. 4.15 is transformed into the integration of ε_{k_m} and performed using an adaptive integration range with 120 grid points. The integration range is $(\varepsilon_{k'} + \alpha_2\hbar\omega_0 - 6\hbar\Gamma_{k_m}, \varepsilon_{k'} + \alpha_2\hbar\omega_0 + 6\hbar\Gamma_{k_m})$, corresponding to a width of $12\hbar\Gamma_{k_m}$ around the center of the Lorentzian function

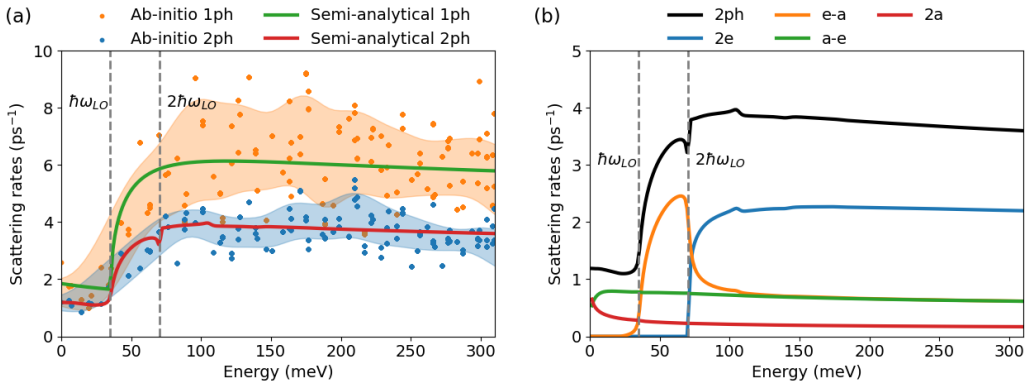


Figure 4.1: (a) 1ph and 2ph Γ-valley scattering rates at 300 K obtained from ab-initio calculations (symbols and shaded regions) and semi-analytical model (lines). Note that intervalley scattering is excluded from these rates. Due to the variations of the ab-initio rates, we apply a Gaussian smearing and plot the shaded region to indicate the region within a standard deviation. (b) Decomposition of semi-analytical 2ph rates (black solid line) to four contributing subprocesses (blue, orange, green, and red solid lines). In both (a) and (b), vertical dashed lines indicating energies of $\hbar\omega_{\mathrm{LO}}$ and $2\hbar\omega_{\mathrm{LO}}$ are plotted.

in the denominator of Eq. 4.15. The relative residual error from this choice of integration limits is estimated as $1/12^2 \approx 0.7\%$. In the final state integration of both 1ph and 2ph, the angular integrations in Eq. 4.17 are performed with 200 grid points. All the above numerical integrations are performed on uniform grids using the midpoint rule. The 2ph calculation is performed with 20 iterations, with the interference term Eq. 4.14 updated every 10 iterations. The relative difference between the 10th and 20th iterations is less than 1%, indicating convergence of the iterative process. The discretization of $\Theta(k, k', \theta_{\boldsymbol{k},\boldsymbol{k}'})$ to $\Theta_{\boldsymbol{k}',\boldsymbol{k}}$ is performed by the regular grid interpolation provided in scipy [260].

4.4 Results

■ valley scattering rates at 300 K

We first present the 1ph and 2ph Γ valley scattering rates versus energy obtained by ab-initio calculation and the semi-empirical model for GaAs at 300K in Fig. 4.1a. For both 1ph and 2ph rates, the ab-initio calculations and the semi-analytical model are in quantitative agreement. Specifically, we observe a rapid increase of the 1ph and 2ph rates at $\hbar\omega_{\mathrm{LO}} \approx 35$ meV followed by a nearly constant trend. The degree of agreement between the semi-analytical model and the mean values of the ab-initio calculation is notable considering the semi-analytical model includes off-

shell processes and the interference term, both of which are neglected in the ab-initio calculations. The agreement can be attributed in part to the cancellation of errors between the limitation on the iteration number and the on-shell approximation in the ab-initio calculation. The third iteration of the 2ph rates yields values that are overestimated from the converged value by about 9%, while the on-shell approximation and the non-interference approximation are found to underestimate the rates by around 3% and 5%, respectively. These approximations offset each other to yield good agreement between the two approaches. Overall, the contribution of off-shell processes is found to make only a minor contribution to the Γ intravalley 2ph scattering rates even up to energies of 300 meV.

The major difference between the ab-initio and the semi-analytical rates is the variation of the individual rates in the ab-initio calculation in a given energy range, which is due to the relatively low grid density used in the ab-initio calculation (see Appendix 8.1). As explained in Sec. 4.2, the semi-analytical model uses a significantly finer grid, leading to negligible variations in individual scattering rates in the same energy range. Although anisotropy could in principle lead to similar variations of the ab-initio rates, this contribution is negligible owing to the high spherical symmetry of the band structure and e-ph matrix elements (about 3% as mentioned in Sec. 4.2).

The high grid density in the semi-analytical model enables features in the scattering rates to be observed that cannot be discerned in the ab-initio calculations, including the previously mentioned rapid increase of 1ph and 2ph rates at $\hbar\omega_{\mathrm{LO}}$, and also a small but evident kink at $2\hbar\omega_{\mathrm{LO}}$ (about 70 meV). We now analyze each of these observations. For the 1ph rates, the increase at $\hbar\omega_{\mathrm{LO}}$ is because LO phonon emission from an electron may only occur above an energy of $\hbar\omega_{\mathrm{LO}}$. For the 2ph rates, the situation is more complicated due to the existence of four subprocesses (2e, e-a, a-e, 2a) in 2ph scattering. To better understand the features in the 2ph scattering rates, the scattering rates of the four subprocesses are plotted separately in Fig. 4.1b. We observe that the increase of the total 2ph rates at $\hbar\omega_{\mathrm{LO}}$ can be attributed to the e-a subprocess due to a similar reason with the emission subprocess in 1ph process, namely that the e-a subprocess requires the electron to have energy exceeding $\hbar\omega_{\mathrm{LO}}$. The kink at $2\hbar\omega_{\mathrm{LO}}$ comes from the cancellation between the increase of the 2e rates and the decrease of e-a rates. The increase of the 2e rates is due to the emission of 2 LO phonons at energies higher than $2\hbar\omega_{\mathrm{LO}}$. The decrease of the e-a rates is due to the increase of the intermediate state rates in the denomi-

nator of Eq. 4.7 (or Eq. 4.15). Specifically, for an e-a subprocess with $\varepsilon_{\boldsymbol{k}} = 2\hbar\omega_{\mathrm{LO}}$, the corresponding intermediate state has the band energy $\varepsilon_{\boldsymbol{k}+\boldsymbol{q}} = \hbar\omega_{\mathrm{LO}}$, where an increase of both $\Gamma^{(1\mathrm{ph})}$ and $\Gamma^{(2\mathrm{ph})}$ occurs as explained above. In fact, the effect of the intermediate state self-energy leads to change at any integer multiple of $\hbar\omega_{\mathrm{LO}}$, but the magnitudes are decaying with increasing energy such that they cannot be observed above a few multiples of $\hbar\omega_{\mathrm{LO}}$.

Another observation in Fig. 4.1b is the difference between the e-a rates and a-e rates asymptotically decreases to zero with increasing energy. This trend can be understood by analysis of Eq. 4.11. Specifically, in the process $\boldsymbol{k}_i \to \boldsymbol{k}_m \to \boldsymbol{k}_f$, the relative differences between norms of the state vectors k_i, k_m, k_f become small at high energies (above a few multiples the LO phonon energy), so that the factor $I^{(\alpha_1,\alpha_2)}(k,k',\theta_{\boldsymbol{k},\boldsymbol{k}'})$ becomes insensitive to the subprocess type. Thus, their magnitudes are fully determined by the phonon occupation factor $A_{\alpha_1}A_{\alpha_2}$ defined in Eq. 4.12 which satisfy $\Gamma^{(2a)}/A_{-1}^2 = \Gamma^{(e-a)}/A_{-1}A_{+1} = \Gamma^{(a-e)}/A_{-1}A_{+1} = \Gamma^{(2e)}/A_{+1}^2$. From this relationship, we find $\Gamma^{(e-a)} = \Gamma^{(a-e)}$. Furthermore, a common ratio $\Gamma^{(2a)}/\Gamma^{(e-a)} = \Gamma^{(e-a)}/\Gamma^{(2e)} = A_{-1}/A_{+1}$ can also be obtained for the subprocess rates at the high energy region. This relationship is observed in Fig. 4.1b.

Drift velocity and current PSD at 300 K

We now examine the transport and noise properties from each model. For the semi-analytical model results, the Γ-Γ block of the scattering matrix is calculated by the semi-analytical model in Eq. 4.10 and 4.11, while the Γ-L and L-L blocks are those of the ab-initio calculation. Figures 4.2a and 4.2b display the drift velocity and normalized current PSD, respectively, from the ab-initio calculations and the semi-analytical model. The experimental measurements are also plotted for comparison. In Fig. 4.2a, the ab-initio calculation and the semi-analytical model give a similar prediction of the drift velocity versus electric field up to 5 kVcm^{-1}. Both give the low-field mobility of around 7000 $\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$, which agrees with the experimental value of about 8000 $\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$ to within around 15%. The similarity between the ab-initio and semi-analytical results is expected due to the agreement of their scattering rates as discussed in Sec. 4.4.

In Fig. 4.2b, the PSD obtained from different experimental measurements reveal a non-monotonic pattern characterized by an initial decrease, followed by a marked rise around the commencement of negative differential mobility, and a subsequent decrease. The origin of this trend was explained in Ref. [247]. However, both the

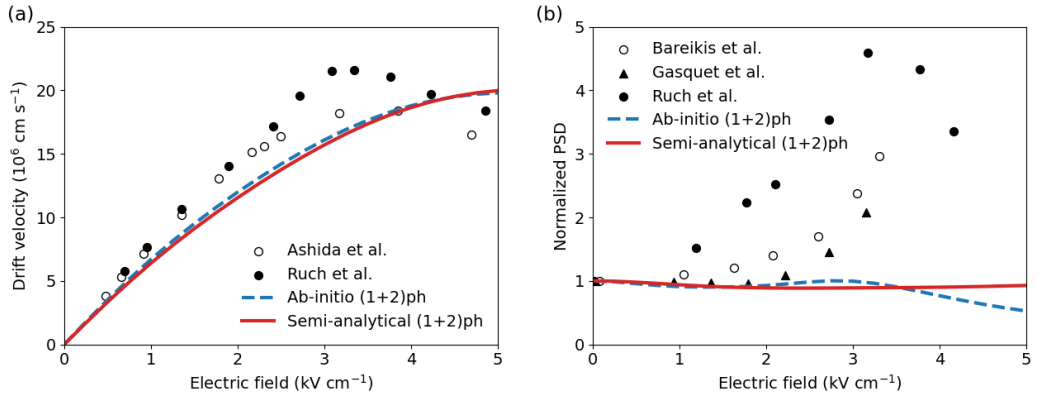


Figure 4.2: Drift velocity and normalized PSD versus electric field for the (1+2)ph results obtained by ab-initio calculation (dashed blue line) and semi-analytical model (solid red line) at a temperature of 300 K. (a) Drift velocity versus electric field. The ab-initio calculation and semi-analytical model qualitatively agree with the measurements of Ruch et al. [261] (filled circles) and Ashida et al. [262] (open circles) (b) Normalized PSD versus electric field. Values calculated using two approaches show consistent discrepancies compared with the results obtained from noise temperature and differential mobility measurements (filled circles, Ref. [263] and open circles, Ref. [264]), and from time of flight experiments (triangles, Ref. [265]).

ab-initio calculation and the semi-analytical model predict the PSD to be nearly independent of electric field and thus fail to predict the characteristic PSD peak at about 3 kVcm^{-1}.

Comparison of cryogenic ■ and ■-L scattering rates to experiment

The lifetimes of photoexcited hot electrons in GaAs have been experimentally measured at 10 K from an analysis of the linewidths of peaks from continuous-wave luminescence spectroscopy. [266] In this section, we compare the ab-initio, semi-analytical and experimental scattering rates at cryogenic temperatures. In Fig. 4.3a and 4.3b, we show the Γ and Γ-L scattering rates obtained by ab-initio calculation, the semi-analytical model, and experiment. The calculations were performed at cryogenic temperatures to enable comparison with experiment. The experimental scattering rates and error bars are converted from the corresponding lifetimes and error bars directly reported in Ref. [266]. Since the semi-analytical model is only valid for Γ-Γ scattering, only the experimental and ab-initio results are shown in Fig. 4.3b.

Figure 4.3a shows that the experimental and theoretical values for the Γ-valley scat-

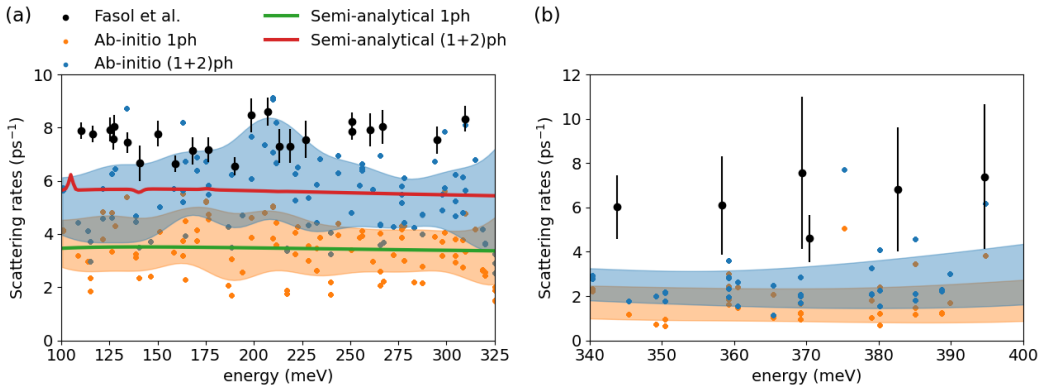


Figure 4.3: Experimental (black symbols, Ref. [266]), ab-initio 1ph (orange symbols) and (1+2)ph (blue symbols), semi-analytical 1ph (green lines) and (1+2)ph results (red lines) of (a) Γ valley scattering rates and (b) Γ-L intervalley scattering rates versus energy at helium temperatures. The (1+2)ph calculations agree better with experiment compared to the 1ph calculations in all cases. The (1+2)ph scattering rates agree with experiment to within 25% and a factor of two for Γ rates and Γ-L intervalley rates, respectively.

tering all yield a nearly constant value between 100 meV and 325 meV. The experimental rates are about 8 ps^{-1} in this energy range, while the semi-analytical 1ph and (1+2)ph calculations give about 3.5 ps^{-1} and 6 ps^{-1}, respectively. This result affirms that the 2ph scattering makes a non-negligible contribution to electron scattering in GaAs. Similarly with Sec. 4.4, the ab-initio calculations give mean values of the scattering rates that agree quantitatively with the semi-analytical model but with substantial scatter about the mean. Such observation again suggests that the approximations in the ab-initio calculations do not result in qualitative deviations.

Figure 4.3b shows the Γ-L intervalley scattering rates of the ab-initio calculations and the photoluminescence experiments. According to Ref. [266], the experimental Γ-L rates are obtained by

$$\Gamma_{\Gamma-L} = \Gamma_{\text{tot}} - \Gamma_{\Gamma-\Gamma}, \tag{4.19}$$

where $\Gamma_{\Gamma-\Gamma}$ is taken as a constant estimated by fitting the data in Fig. 4.3a. To make a consistent comparison with experiment, the ab-initio Γ-L intervalley scattering rates in Fig. 4.3b are also calculated by Eq. 4.19 instead of being directly calculated from the scattering matrix. It is found that the experimental intervalley rates are about 6 ps^{-1} in the energy range from 340 meV and 400 meV, while the ab-initio 1ph and (1+2)ph give around 2 ps^{-1} and 3 ps^{-1}, respectively. Although the additional 2ph calculation decreases the deviation from experiment results, an

underestimation of a factor of two is still observed. This discrepancy could be attributed to the need for an off-shell extent larger than 25 meV in the 2ph calculation, owing to the larger intermediate state scattering rates (appearing in the denominator of Eq. 4.7 and Eq. 4.15) at energies above 300 meV. However, at present a larger off-shell extent is computationally infeasible.

4.5 Discussion

The semi-analytical model treats the full scattering term for Γ intravalley 2ph scattering but does not qualitatively alter transport and noise properties. In particular, the marked discrepancy with the experimental PSD remains. We now examine alternate possible origins for the discrepancy.

Underestimated Γ-L intervalley scattering rates

A comparison of our computed cryogenic rates with those measured from photoluminescence experiments indicates that the Γ rates agree to within 25%, but the Γ-L intervalley rates are underestimated by around a factor of two. Despite this underestimate, prior work suggests that this effect is unlikely to reconcile the PSD discrepancy. Specifically, Monte Carlo simulations of electron transport in GaAs with a three-valley Γ-L-X model have found that increased intervalley scattering suppresses the PSD feature (see Fig. 7 of Ref. [267]). Therefore, although the possibility cannot be definitively excluded at present, including intervalley scattering processes beyond those treated already is not expected to yield improved agreement with the PSD.

Contribution from simultaneous electron-two-phonon interaction

According to Ref. [247], another possibility is that the contribution of electron-two-phonon (e-2ph) interaction [268] is not considered. Here, we make a qualitative estimation of the magnitude of this effect based on the Fröhlich mechanism for electron scattering. According to Ref. [268], the e-ph Hamiltonian up to the second order can be written as

$$H_{\text{e-ph}} = \sum_{\boldsymbol{R}\kappa} \boldsymbol{u}_{\boldsymbol{R}\kappa} \cdot \nabla V(\boldsymbol{r} - \boldsymbol{R}_\kappa) + \frac{1}{2} \sum_{\kappa} \boldsymbol{u}_{\boldsymbol{R}\kappa} \cdot \nabla\nabla V(\boldsymbol{r} - \boldsymbol{R}_\kappa) \cdot \boldsymbol{u}_{\boldsymbol{R}\kappa}, \tag{4.20}$$

where κ is the index of atom in a unit cell, $\boldsymbol{R}$ is the unit cell position, $\boldsymbol{u}_{\boldsymbol{R}\kappa}$ is the corresponding phonon-induced displacement, and V is the electron potential. In the long wavelength limit, the electric potential for the Fröhlich interaction can be obtained by assigning a point dipole to each atom [253]. For acoustic phonons with

the same displacements for atoms in the same unit cell, the net dipole moment will be zero and no scattering will occur. Similarly, the electric potential for e-2ph interaction can be obtained by assigning a point quadrupole to each atom. Following the same logic, the net quadrupole moment will be zero if the quadrupoles are induced by two acoustic phonons or two optical phonons, which means that the simultaneous e-2ph interaction based on the Fröhlich interaction can only be induced by an acoustic and optical phonon.

We estimate the order of magnitude of such simultaneous e-2ph interaction involving an acoustic and optical phonon. A full derivation can be found in the Appendix 8.2. The final estimated scattering rates in the Γ valley is

$$\Gamma^{(\mathrm{e-2ph})}(k) \sim \frac{8\pi^2}{\Omega_{\mathrm{BZ}}\hbar}\frac{k^3}{\varepsilon_k}\left(\frac{eZ}{\Omega\varepsilon_\infty}\sqrt{\frac{\hbar}{2M\omega_{\mathrm{A}}}}\sqrt{\frac{\hbar}{2M\omega_{\mathrm{O}}}}\right)^2 (N_{\mathrm{A}}+1)(N_{\mathrm{O}}+1), \tag{4.21}$$

where ω_{O} and ω_{A} are frequencies of optical and acoustic phonons at the edge of Brillouin zone, N_{O} and N_{A} are the corresponding phonon occupations, Ω_{BZ} is the Brillouin zone volume, M is the average atom mass in a unit cell, ε_∞ is the high-frequency permittivity, and Z is the Born effective charge of a single atom. For a typical k such that $\varepsilon_k = 200$ meV, the e-2ph scattering rates at 300 K can be estimated as $10^{-2.5}$ ps^{-1}, which is about 3.5 orders of magnitude smaller than the ab-initio or semi-analytical 2ph scattering rates obtained in this work. Thus, we conclude that the effect of the simultaneous e-2ph interaction based on the Fröhlich interaction can be neglected. A related effect involving strong electron-phonon interactions, polaron formation, can also likely be excluded as this effect has no dependence on electric field strength.

Space charge domains and experimental non-idealities

Finally, we consider an alternate mechanism for the PSD peak which does not rely solely on intervalley scattering. The earliest studies of negative differential resistance in GaAs arose from the observation of current instabilities at electric fields approaching a threshold value of around 3 kVcm^{-1}. [269] These instabilities were attributed to the formation of space charge domains associated with the negative differential resistance. The typical Boltzmann formalism used to describe charge transport from first principles does not include the contribution of such effects because it neglects real-space gradients and space charge effects which are essential to the instability.

Space charge instabilities manifest as current fluctuations, and therefore the nucle-

ation of space-charge domains could explain the PSD peak around the threshold field. However, an inconsistency with this explanation is that the increase in PSD begins at a field below the threshold value for NDR as in Fig 4.2. This inconsistency can be accounted for by considering the possibility that the local electric field exceeds the threshold even though the average field does not. Such a possibility was investigated theoretically by McCumber and Chynoweth [270], who found that the dipole layer generation process was sensitive to inhomogeneities such as doping fluctuations that would arise from purely random Poisson statistics. The dipole layer was found to propagate even if the average uniform field was less than the nominal threshold field based on the velocity-field characteristics.

These considerations support the explanation of the PSD peak in terms of instabilities associated with the local electric field exceeding the threshold field for negative differential resistance space charge domain formation. This finding has implications for the use of transport studies to determine intervalley scattering strength. In particular, Monte Carlo methods have been used for decades for this purpose in GaAs by interpreting transport and noise measurements, and in those simulations, noise was assumed to arise from solely intervalley scattering. Our result indicates that this approach to studying intervalley scattering is not valid because the physical origin of noise differs from that assumed in the model. Instead, photoluminescence methods which directly provide an electronic lifetime as in Ref. [266] should be employed.

4.6 Conclusions

We have introduced a semi-analytical model of 1ph and 2ph Γ intravalley scattering for electrons in GaAs which allows for prior approximations in the treatment of the 2ph term for Γ scattering to be lifted while incurring errors of a few percent. We find that the calculated transport and noise properties are qualitatively unchanged from the ab-initio values. The computed drift velocity agrees with experiment to within 15%, while agreement with measured cryogenic scattering rates are within 25% for the Γ valley scattering rates and a factor of two for Γ-L intervalley scattering. However, the qualitative discrepancy for the PSD is not improved with the semi-analytical model. Considering the totality of the evidence, our work suggests that the PSD peak mostly likely arises from space charge domain formation rather than partition noise associated with intervalley scattering, as has been assumed for many decades. This result implies that care must be taken when interpreting transport and electrical noise measurements in terms of intervalley scattering. Our findings high-

light the insights into charge transport that can be obtained from a first-principles treatment of high-field charge transport and noise properties.

Chapter 5

EXPERIMENTAL DEMONSTRATION OF SCALABLE CROSS-ENTROPY BENCHMARKING TO DETECT MEASUREMENT-INDUCED PHASE TRANSITIONS ON A SUPERCONDUCTING QUANTUM PROCESSOR

This chapter is adapted from

1. Kamakari, H., Sun, J., Li, Y., Thio, J. J., Gujarati, T. P., Fisher, M., Motta, M. & Minnich, A. J. Experimental demonstration of scalable cross-entropy benchmarking to detect measurement-induced phase transitions on a superconducting quantum processor. *arXiv preprint arXiv:2403.00938*. https://arxiv.org/abs/2403.00938 (2024).
   Contribution: J.S. designed and implemented the circuit compression algorithm, and contributed to writing the article.

Quantum systems subject to random unitary evolution and measurements at random points in spacetime exhibit entanglement phase transitions which depend on the frequency of these measurements. Past work has experimentally observed entanglement phase transitions on near-term quantum computers, but the characterization approach using entanglement entropy is not scalable due to exponential overhead of quantum state tomography and post selection. Recently, an alternative protocol to detect entanglement phase transitions using linear cross-entropy was proposed which eliminates both bottlenecks. Here, we report the demonstration of this protocol in systems with one-dimensional and all-to-all connectivities on IBM's quantum hardware on up to 22 qubits, a regime which is presently inaccessible if post-selection is required. We demonstrate a collapse of the data into a scale-invariant form with critical exponents agreeing with theory within uncertainty. Our demonstration paves the way for studies of measurement-induced entanglement phase transitions and associated critical phenomena on larger near-term quantum systems.

5.1 Introduction

Quantum systems undergoing unitary evolution in the presence of an observer making measurements (monitored quantum systems) [271–273] exhibit unique dynamics, distinct from both thermalizing closed systems [274] and conventional open quantum systems [275]. When the system is weakly monitored and subject to sufficiently entangling unitaries, initial product states typically exhibit a linear in time growth of the entanglement entropy, before evolving into steady states where the entanglement entropy admits a volume-law scaling [276–280]. In contrast, strongly monitored systems are not able to support highly entangled states, resulting in area-law entanglement scaling even at long times [281, 282]. Separating the two phases lies a phase transition, which was initially found theoretically in simplified quantum circuit models with mid-circuit measurements, and was later found to be generic to

a wide range of monitored dynamics [282–297]. Such measurement-induced phase transitions (MIPTs) have recently garnered much interests [298], in part due to multiple theoretical viewpoints one can take in describing them [299–303].

An experimental observation of MIPTs was recently demonstrated on IBM quantum hardware with up to 14 qubits [304]. By directly measuring the entanglement entropy after a comprehensive quantum tomography of the steady states, Koh et al. [304] were able to observe MIPT and confirm the competing effects of random unitaries and mid-circuit measurements. However, the experiment required over 5200 device hours and is limited in scalability due to the exponential cost of quantum state tomography and post-selection of measurement outcomes. The lifetime of superconducting qubits also puts a stringent limit on the circuit depth (as well as on system size when circuit depth scales with the number of qubits), since mid-circuit measurements can be an order of magnitude slower than two-qubit unitary gates.

To avoid mid-circuit measurements, a space-time duality mapping was introduced [305, 306] and recently implemented on Google's superconducting processor [307], where MIPT-like physics was observed in 1D unitary circuits with a reduced number of post-selections, and at the boundary of shallow 2D unitary circuits of 70 qubits without post-selection. Alternatively, order parameters based on reference qubits can be used to efficiently and scalably probe MIPTs [308], where post-selection can be avoided with an accompanying classical simulation. The use of a reference qubit to probe MIPTs has been demonstrated in trapped ion systems for Clifford circuits [309], featuring a high gate fidelity and non-local qubit connectivity. Another order parameter which can be used to probe MIPTs is the cross entropy [310], which requires no ancilla qubits and lacks the exponential overhead of post selection. Although the theoretical basis for this method has been established, a demonstration of this protocol on near-term quantum hardware has not yet been reported.

In this work, we report an experimental demonstration of the detection of MIPTs on prototypical hybrid Clifford circuit models with up to 22 physical qubits. The required circuits were executed in less than 8 device-hours on IBM superconducting devices, representing a decrease in device time by nearly two orders of magnitude compared to the approach based on measuring entanglement entropy. Moreover, a circuit compression technique allows us to investigate circuit models with all-to-all connectivity on IBM's 2D layout. From the data, we extract critical exponents which are compatible with theoretical predictions within the experimental uncertainty. This work paves the way for studies of other critical phenomena on near-

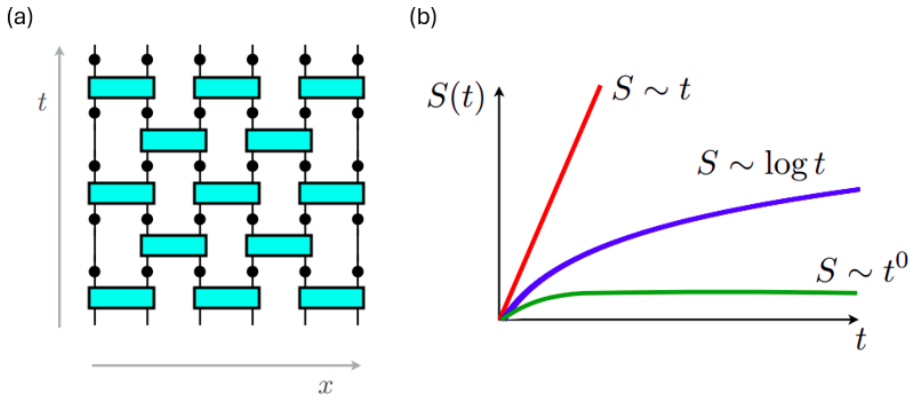


Figure 5.1: (Adapted from Ref. [271]) (a) A typical quantum circuit to characterize MIPT. The circuit is composed of alternating layers of unitary operations (bricks) and mid-circuit measurements with probabilities p (dots). (b) Asymptotic relation between entanglement entropy S and circuit depth t. In the volume phase $p < p_c$, the critical point $p = p_c$, and the area-law phase $p > p_c$, the corresponding asymptotic relations are $S \sim t$, $S \sim \log t$, and $S \sim t^0$, repsectively, where p_c is the critical measurement probability.

term quantum hardware and provides a potential benchmarking tool for quantum circuits with mid-circuit measurements.

5.2 Cross entropy benchmark for MIPT

Review of MIPT

The phenomenon of measurement-induced entanglement transitions (MIPT) comes from the competition between the unitary dynamics which tends to generate the entanglement, and measurements which tend to destroy the entanglement. [271, 272, 283] A typical example is the quantum circuit composed of alternating layers of Haar-random unitary operations and probabilistic mid-circuit measurements, where each measurement independently takes place with probability p. (see Fig 5.1(a)) Let C indicate a specific choice of the circuit, including the choice of both unitary operations and the positions of mid-circuit measurement, and $\boldsymbol{m} = (m_1, m_2, \ldots, m_N)$ be the bitstring representing the measurement outcomes, where m_j are the outcomes (0 or 1) from each mid-circuit measurement. Consider a division of the system to two subsystems A and B, and let S be the entanglement entropy between A and B. Such entanglement entropy S can be viewed as a function of C and $\boldsymbol{m}$. MIPT can be observed from the relation between the averaged value of S over C and $\boldsymbol{m}$, denoted as $\bar{S} = \langle S \rangle_{C,\boldsymbol{m}}$, and the measurement rate p in the thermodynamic limit of infinite system size. For $0 \leq p \leq 1$, there exists two distinct phases separated by a

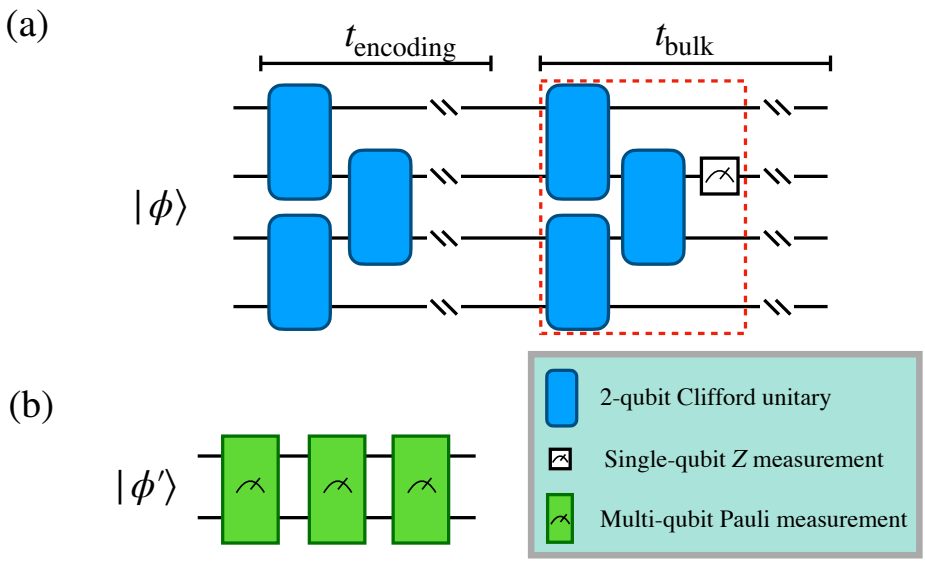


Figure 5.2: Schematic of the cross entropy benchmark protocol. (a) An L-qubit Clifford circuit used in the cross entropy benchmark protocol [310]. We choose the initial state $|\phi\rangle$ to be either $|0T\rangle^{\otimes L/2}$, where $|T\rangle$ is a magic state, or $|0\rangle^{\otimes L}$. The red box represents one layer of unitaries (and one layer of measurements in the bulk stage). (b) The compressed $L/2$-qubit circuit consisting of at most $L/2$ multi-qubit Pauli measurements. The compressed initial state is $|\phi'\rangle = |T\rangle^{\otimes L/2}$ or $|0\rangle^{\otimes L/2}$.

critical measurement rate p_c. For $p > p_c$, the system is in the area-law phase such that $\bar{S} \sim$ const, while for $p < p_c$, the system is in the volume-law phase such that $\bar{S} \sim T$. (See Fig. 5.1(b)) The latter is a generalized result of pure unitary dynamics. [311] At the exact critical point, the entanglement has a logarithmic growth with the circuit depth, i.e. $\bar{S} \sim \log T$. However, $\bar{S}$ cannot be obtained by the averaged density matrix over C and $\boldsymbol{m}$, but instead require the resolution of each individual quantum trajectory for each C and $\boldsymbol{m}$. This imposes two exponential overheads: the exponentially scaled overhead of quantum state tomography to evaluate ρ, and the exponentially–scaled overhead to post-select each measurement bitstring $\boldsymbol{m}$. Thus, direct experimental characterization of MIPT was limited to small system sizes. [304]

Cross entropy benchmark

In this work, we use the cross entropy benchmark protocol of Ref. [310] to characterize MIPT. We consider a family of random circuits, where each circuit consists of two stages: an purely unitary "encoding stage" consisting of t_{encoding} layers, and a "bulk stage" consisting of t_{bulk} layers with both unitary gates and mid-circuit measurements, see Fig. 5.2(a). For an L-qubit circuit, both stages must contain a

number of layers scaling at least linearly with L for the system to enter a steady state, particularly when the steady state has volume-law scaling of entanglement entropy. For a given circuit C with mid-circuit measurements bitstring $\boldsymbol{m}$, let the unnormalized quantum operation conditioned on mid-measurement bitstring $\boldsymbol{m}$ be $C_{\boldsymbol{m}}$, which satisfies $\sum_{\boldsymbol{m}} C_{\boldsymbol{m}}^{\dagger} C_{\boldsymbol{m}} = I$. With the initial state ρ, the corresponding unnormalized final state conditioned on the mid-measurement results $\boldsymbol{m}$ be

$$\rho_{\boldsymbol{m}} = C_{\boldsymbol{m}} \rho C_{\boldsymbol{m}}^{\dagger}, \tag{5.1}$$

and the corresponding probabilitity to obtain $\boldsymbol{m}$ be

$$p_{\boldsymbol{m}}^{\rho} = \mathrm{tr}(\rho_{\boldsymbol{m}}). \tag{5.2}$$

The protocol involves the application of the same circuit to two different initial states, ρ and σ, and a comparison between the two ensembles of measurement records. Let the probabilities of obtaining $\boldsymbol{m}$ with the initial states ρ and σ be $p_{\boldsymbol{m}}^{\rho}$ and $p_{\boldsymbol{m}}^{\sigma}$. The normalized linear cross entropy acts as a distance measure between the two probability distributions $p_{\boldsymbol{m}}^{\rho}$ and $p_{\boldsymbol{m}}^{\sigma}$, and is defined for this circuit as

$$\chi_C = \frac{\sum_{\mathbf{m}} p_{\mathbf{m}}^{\rho} p_{\mathbf{m}}^{\sigma}}{\sum_{\mathbf{m}} (p_{\mathbf{m}}^{\sigma})^2}, \tag{5.3}$$

which can be estimated by taking the sample average of $p_{\mathbf{m}}^{\sigma} / \left(\sum_{\mathbf{m}} (p_{\mathbf{m}}^{\sigma})^2\right)$ over many runs of the quantum circuit with input state ρ. We then average over random circuits C, obtaining the final cross entropy for a given measurement rate as $\chi = \mathbb{E}_C \chi_C$. As shown in [310], for $\rho \neq \sigma$ and in the absence of noise, the quantity χ acts as an order parameter which, in the thermodynamic limit, approaches 1 when the system is in the volume-law phase and approaches a constant strictly less than 1 in the area-law phase. Intuitively, χ measures the distinguishability of the two initial states by comparing mid-circuit measurement records, after the two initial states are "scrambled" by the encoding unitary. Previously, the linear cross entropy has been used as a figure of merit for random circuit sampling [312–315], or in benchmarking quantum simulators [316, 317]. For χ to be efficiently obtainable from quantum and classical hardware, the probabilities $p_{\mathbf{m}}^{\sigma}$, as well as $\sum_{\mathbf{m}} (p_{\mathbf{m}}^{\sigma})^2$, need to be efficiently classically computable. This is possible when the bulk of the circuit contains only Clifford operations and when the input state σ is a stabilizer state. The cross entropy protocol is similar in spirit to hybrid quantum-classical observables used in previous experiments [307, 309] (see also [318, 319]) and, as we will show, allows us to probe the transition and obtain critical exponents on noisy processors.

5.3 Experimental implementation

Circuit setup

This approach is implemented by our collaborators on IBM Quantum processors. The systems we considered are a 1D chain with nearest-neighbor qubit connectivity and an infinite-dimensional system with all-to-all qubit connectivity. We chose the initial L-qubit states on the quantum processor in both cases to be $\rho = |0T0T\cdots 0T\rangle\langle T0T0\cdots T0|$ with $|T\rangle = (|0\rangle + \exp(i\pi/4)|1\rangle)/\sqrt{2}$, the alternating magic state, and $\sigma = |0^{\otimes L}\rangle\langle 0^{\otimes L}|$, the all-zero state. Note that ρ is not a stabilizer state. For the alternating magic states, the number of T gates grows linearly with the number of qubits, so that an exact simulation of the circuit via either the state-vector simulation or Clifford+T simulation [320–323] requires exponential classical resources.

For all experiments, the circuits are constructed using alternating layers of unitaries and measurements. Each unitary layer consists of $L/2$ two-qubit unitary gates, sampled uniformly from the two-qubit Clifford group. For the 1D chain, the two-qubit Clifford gates are applied in a brickwork pattern on nearest-neighbor qubits. For the infinite dimensional system, $L/2$ two-qubit unitaries are applied to pairs of qubits selected uniformly at random. Each measurement layer, in both the 1D chain and the infinite-dimensional system, consists of single-qubit Z measurements occurring on each qubit with probability p. For both systems, we used an encoding ratio and bulk ratio of 3, namely $t_{\text{bulk}} = t_{\text{encoding}} = 3L$. For all experiments, 1,000 random circuits are generated for each (L, p) pair, and each circuit was run 1,000 times on the *ibm_sherbrooke* machine.

Compression of Clifford circuit with magic initial state

The resulting circuits with the above properties have as many as L^2 mid-circuit measurements, which are relatively slow operations and introduce both readout and quantum state errors, and so they cannot be executed while preserving adequate fidelity. We therefore employ a circuit compression scheme which exploits the input state being an alternating magic state and the circuit bulk being fully Clifford. The compression is based on Ref. [324] with an improvement that removes the requirement for dynamic circuits (adaptivity), instead using an efficient classical simulation and classical coin flipping.

In a particular circuit realization the unitaries and the measurements can be written

as

$$C_{\mathbf{m}} = \ldots U_3 M_{m_2} U_2 M_{m_1} U_1. \tag{5.4}$$

Here m_j is the j-th measurement outcome of the entire record, and correspondingly $M_{m_j} = (1+(-1)^{m_j} P_j)/2$ is the j-th projection operator, with P_j the Pauli operator being measured. The first step is to convert it to the so-called Pauli-based computing (PBC) model, via rewriting Eq. 5.4 by moving all unitaries past the measurements to the right as

$$C_{\mathbf{m}} = \ldots \widetilde{M}_{m_2} \widetilde{M}_{m_1}, \tag{5.5}$$

where

$$\widetilde{M}_{m_j} = \frac{1}{2}(1 + z_j \widetilde{P}_j), \quad \widetilde{P}_j = U_1^\dagger U_2^\dagger \ldots U_j^\dagger P_j U_j U_{j-1} \ldots U_1 \tag{5.6}$$

are now multi-site Pauli measurements and $z_j = (-1)^{m_j}$. In other words, the circuit is now composed of $\tilde{P}_1, \tilde{P}_2$, etc.

In such a PBC model, some of the Pauli measurements have $1/2$ probabilities of generating ± 1 outcomes, which allows us to equivalently generate classical pseudorandom numbers and rewrite the corresponding projection operation as a new Clifford operation. The detailed compression algorithm is shown in Appendix 9.1. The PBC model simplified by the above process can be turned again into a standard circuit with a reduced number of gates and measurements on a subset of the qubits. In Table 5.1 we present a summary of the quantum hardware resource requirements before and after circuit compression. The number of hardware qubits as well as the number of 2 qubit gates are both reduced by a constant factor after compression, while the average number of measurements becomes independent of the measurement rate p. We note that although the depth increases by a factor of L after circuit compression, for the system sizes in our experiments the increase in depth was not a limiting factor.

With the circuit compression, the initial state of the circuit is now $|T\rangle^{\otimes L/2}$, see Fig. 5.2(b). All circuits used in our experiments use Clifford compression, allowing us to treat up to 44 qubit systems using only 22 physical qubits.

Qubit selection

For the 1D-chain experiment with $\rho \neq \sigma$, the qubits were selected heuristically at run time as in Ref. [304]. The qubits we selected based on the one and two qubit

	Before compression	After compression
Num. hardware qubits	L	$L/2$
Average depth	$9L$	L^2
Num. 2 qubit gates	$3L^2$	$L^2/2$
Avg. num. measurements	$3L^2p$	$L/2$

Table 5.1: Hardware resources required before and after Clifford circuit compression for a fixed L and p. The number of hardware qubits, average depth, and average number of 2 qubit gates required are reduced by a constant factor after compression, whereas the average number of measurements is reduced by a factor of L and is independent of p. The values in this table apply both to the 1D system as well as the all-to-all system.

gate error rates, ε^{1q} and ε^{2q}, respectively, as well as the qubit readout error ε^{ro} rates provided by IBM in their hardware callibration data. Denoting by χ the set of qubit selections which contain all $L/2$ qubits in a connected chain, an average circuit error for circuit C is calculated as

$$\mathscr{E}_{x\in\chi}[C] = \sum_{j\in x}(\varepsilon^{1q}N_j^{1q}[C] + \varepsilon^{2q}N_j^{2q}[C] + \varepsilon^{ro}N_j^{ro}[C]), \tag{5.7}$$

where the subscript j represents the j'th qubit in the qubit set x, and the function $N_j^{1q,2q,ro}$ computes the number of single qubit gates, two qubit gates, and measurements, respectively, acting on qubit j in the circuit C. The qubit chain used in the experiment is then selected as the one which minimizes the average error over all circuits $\mathscr{C}$, $\text{argmin}_{x\in\chi}\mathbb{E}_{C\in\mathscr{C}}\mathscr{E}_x[C]$. For the all-to-all and $\rho=\sigma$ experiments, we used the same qubit layouts that were selected for the 1D-chain.

Fitting critical parameters ν and p_c by collapsing hardware data

Near the critical measurement rate p_c, the order parameter χ for different system sizes and under suitable rescaling is expected to collapse onto a single curve [271, 325, 326]. Quantitatively, this can be expressed as $\chi(L,p)$ collapsing to the same curve for all system sizes L when we suitably rescale both L and p:

$$\chi(L,p) = F\left[L^{1/\nu}(p-p_c)\right]. \tag{5.8}$$

The critical measurement rate depends on the microscopic details of the circuits, such as the encoding and bulk ratios, whereas the the critical exponent is independent of the microscopic circuit details and is the same for all systems in the same universality class [271, 325]. If the scaling function F was known, we could obtain the optimal p_c and ν, denoted by p_c^* and ν^*, by minimizing the residual sum of

squares (RSS) over all data points:

$$p_c^*, \nu^* = \arg\min_{p_c,\nu} \sum_L \sum_p \left(F\left[L^{1/\nu}(p-p_c)\right] - \chi_{\exp}(L,p) \right)^2, \tag{5.9}$$

where $\chi_{\exp}(L,p)$ is the cross entropy obtained from the experiment for a system size L and measurement rate p. When the scaling function is unknown, we still find p_c^* and ν^* by minimizing an RSS, but instead use an interpolating function for our scaling function for a fixed L, followed by symmetrization over all L in order to prevent preferential treatment of any portion of the data [304, 326]. Our approach to fitting p_c and ν follows Ref. [304] with modifications due to there being only one critical exponent in our case, versus two critical exponents in Ref. [304]. We denote by $\mathscr{L}$ the set of system sizes used in the experiment and $\mathscr{P}_L$ the set of measurement rates used for a fixed L. For each $L \in \mathscr{L}$ and $p \in \mathscr{P}_L$, we first compute the rescaled controlled variable

$$q_L(p) = L^{1/\nu}(p-p_c). \tag{5.10}$$

We then construct an interpolating function for $\chi(L,p)$ from the rescaled experimental data, which we denote by $f_L(q)$. The interpolating function is used since the q values for different values of L are different, and the RSS is taken over points with identical q values. From numerical simulations, we expect the scaling function to decrease monotonically for increasing p [310]. To preserve this monotonicity, we use a piecewise cubic Hermite polynomial implemented in SciPy to construct the interpolating function [327]. We denote the set of q_L as $\mathscr{Q}_L$, $q_L^- = \min_{\mathscr{Q}_L}$ and $q_L^+ = \max_{\mathscr{Q}_L}$. Adapting the measure of goodness of fit from References [326] and [304], we define the loss function as

$$R(\nu, p_c) = \sum_{L\in\mathscr{L}} \sum_{\substack{L'\in\mathscr{L},\\ L'\neq L}} \sum_{\substack{q\in\mathscr{Q}_{L'},\\ q_L^-\le q\le q_L^+}} \left(f_L(q) - f_{L'}(q)\right)^2. \tag{5.11}$$

In the innermost summation, we constrain q by $q_L^- \le q \le q_L^+$ in order to avoid extrapolation of $f_L(q)$. Our reported best fit values of p_c and ν are then given by

$$p_c^*, \nu^* = \arg\min_{p_c,\nu} R(\nu, p_c). \tag{5.12}$$

Following References [304, 326],the errors for ν and p_c are given by the width of

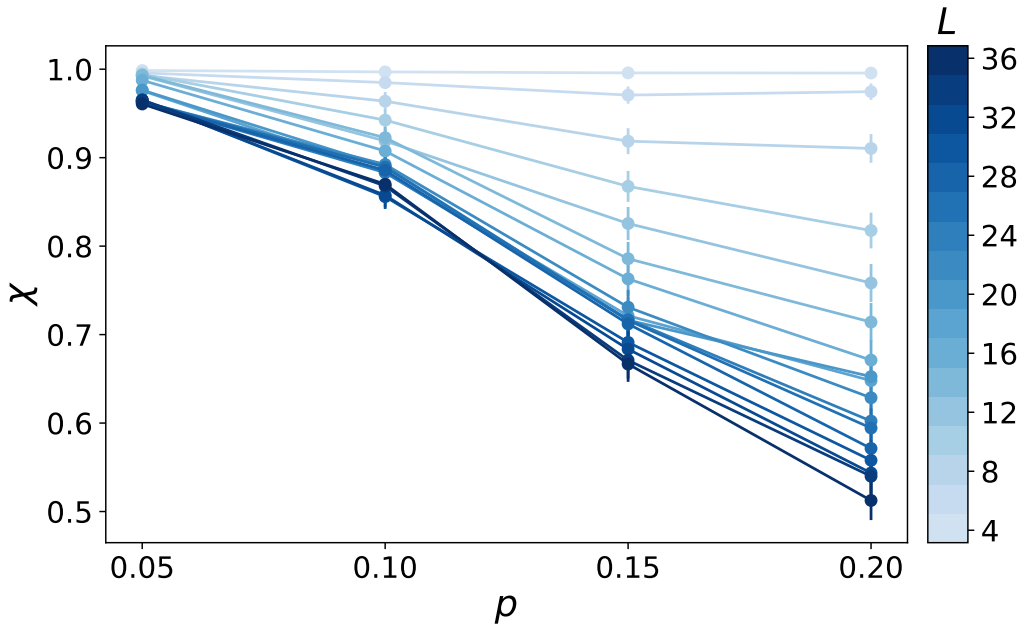


Figure 5.3: Cross entropy for identical initial states ($\rho = \sigma$) obtained from *ibm_sherbrooke* with up to 18 physical qubits (equivalent to a system size of $L = 36$ qubits before compression). The errors incurred from the physical qubits results in a cross entropy lower than the theoretical value of 1.

the minimum at level η:

$$\delta\nu_{\pm} = \eta\nu^* \left[2\log\frac{R(\nu^* \pm \eta\nu^*, p_c^*)}{R(\nu^*, p_c^*)}\right]^{-1/2} \tag{5.13}$$

$$\delta p_{c\pm} = \eta p_c^* \left[2\log\frac{R(\nu^*, p_c^* \pm \eta p_c^*)}{R(\nu^*, p_c^*)}\right]^{-1/2}. \tag{5.14}$$

Our final values of ν and p_c are then reported, setting η to the 10% level, as

$$\nu^* \pm \max(\delta\nu_+, \delta\nu_-) \tag{5.15}$$

$$p_c^* \pm \max(\delta p_{c+}, \delta p_{c-}). \tag{5.16}$$

5.4 Results for 1D connectivity

The results for 1D connectivity are obtained by our collaborators. We first present the experimental results when we set $\rho = \sigma$ to provide a benchmark of the hardware performance. The circuits are obtained from the compressed 1D circuits, but replace all $|T\rangle$ states with $|0\rangle$ states, so that the initial states on both ρ and σ are the all-zero state; see Fig. 5.2(b). In this case, since the circuits run on both the quantum and classical sides are identical, we expect to observe $\chi = 1$ for all L and for all p in the

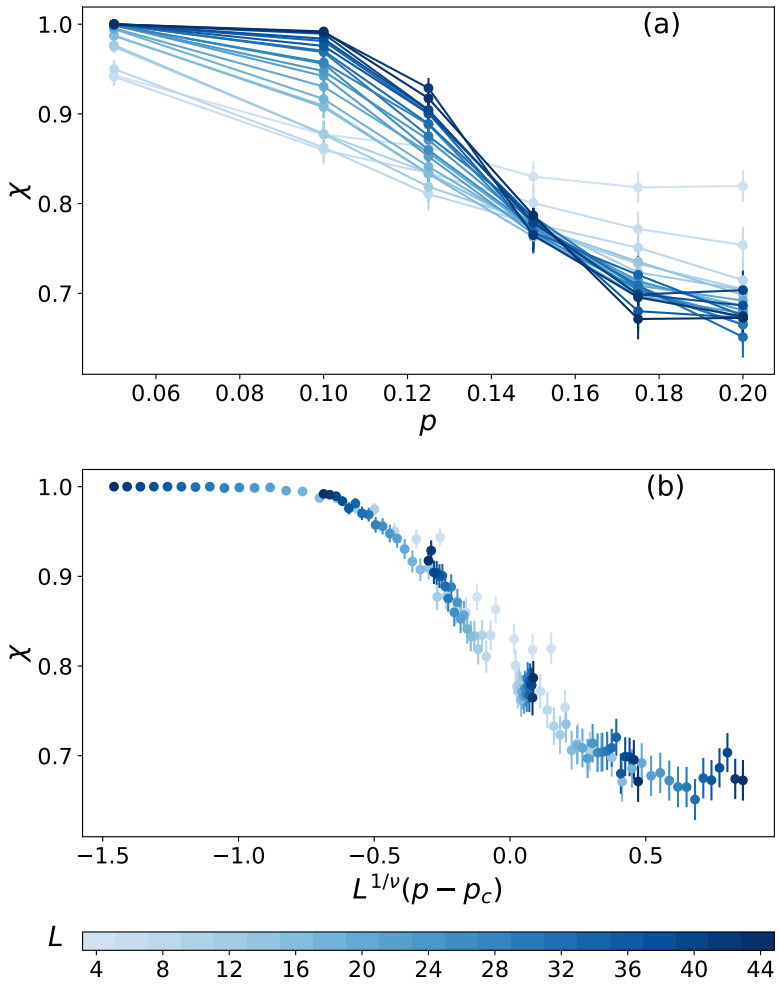


Figure 5.4: (a) Cross entropy χ for 1D chains with up to 22 physical qubits (corresponding to a system size of $L = 44$ qubits before compression) computed on *ibm_sherbrooke*. (b) Collapse of cross entropy curves near the critical point obtained by minimizing the scatter of all points to an unknown scaling function. The fitting procedure gives a critical measurement rate of $p_c = 0.14 \pm 0.01$ and critical exponent $\nu = 1.4 \pm 0.5$.

absence of any noise or hardware errors. The deviation of the cross entropy from 1 therefore provides a measure of the overall errors and noise in the circuit, which could be due to various sources such as gate errors, qubit decay and dephasing, and cross-talk from mid-circuit measurements.

Fig. 5.3 shows the cross entropy χ versus p for various L. We observe that $\chi > 0.9$ for $L > 8$ for all p, but it decreases below unity for larger L and $p \gtrsim 0.1$. That $\chi \approx 1$ for small L can be attributed to the short circuit depths of these circuits. For instance, a compressed four qubit circuit requires only two physical qubits and two

Pauli measurements, allowing them to be executed with high fidelity and resulting in a high cross entropy. For increasing L at fixed p, both uncompressed and compressed circuits contain more mid-circuit measurements.

The trends of χ with L and p in Fig. 5.3 qualitatively agree with noisy simulations (see Appendix 9.2).

We next present experimental results for the 1D chain for $\rho \neq \sigma$. Fig. 5.4 shows χ versus p for L between 4 and 44 (2 to 22 physical qubits) up to $p = 0.2$ obtained from the 127 qubit *ibm_sherbooke* device. Qualitatively, we see the expected characteristics as described below Eq. (5.3); namely, with increasing L, χ approaches unity for $p \lesssim 0.12$, while it plateaus to a constant < 1 for larger values of p. The curves for different L cross at a value of p we denote as p_c, with $p_c \in (0.15, 0.175)$.

The cross entropy χ is related to a domain wall free energy in an associated statistical mechanics model [310], and its value near the critical point depends only on the ratio of the system size and the correlation length, according to standard scaling hypotheses. We verify this hypothesis by collapsing the data from different system sizes L and measurement probabilities p with the method explained in Sec. 5.3. The resulting collapsed curve is shown in Fig. 5.4b. We obtain the critical measurement rate $p_c = 0.14 \pm 0.01$ and critical exponent $\nu = 1.4 \pm 0.5$ at the 90% confidence level. Our reported values of p_c and ν are in quantitative agreement with classical numerical calculations of uncompressed circuits in the presence of 0.1% erasure noise, where we obtained $\nu \approx 1.33$ and $p_c = 0.14$, see Appendix 9.2. For chains of fewer than 10 qubits, finite-size effect are observed as indicated by deviations from the collapsed curve as well as the plateau to a larger value for large p. Removing the smaller system sizes from the fitting did not change the values of ν and p_c within the reported uncertainties.

5.5 Results for all-to-all connectivity

The results for all-to-all connectivity are obtained by our collaborators. We finally present the experimental results for the all-to-all connectivity experiment. Compared to 1D systems, all-to-all connected systems without compression would require $O(L^3)$ SWAP gates per circuit to implement all the 2 qubit unitaries on hardware with nearest neighbor interactions. This prohibitive scaling makes all-to-all systems harder to simulate than 1D systems. With circuit compression, however, the resource requirements are the same as the 1D system since in both cases the compressed circuits have $L/2$ qubits and at most $L/2$ mid-circuit measurements.

We demonstrate this resource reduction application of circuit compression by experimentally observing an MIPT for an all-to-all connected system. Theory predicts qualitatively similar dependencies of the cross entropy on p and L as in the 1D case, but the transition is in a different universality class [328]. The initial states used in this experiment are the same as in the 1D-chain case.

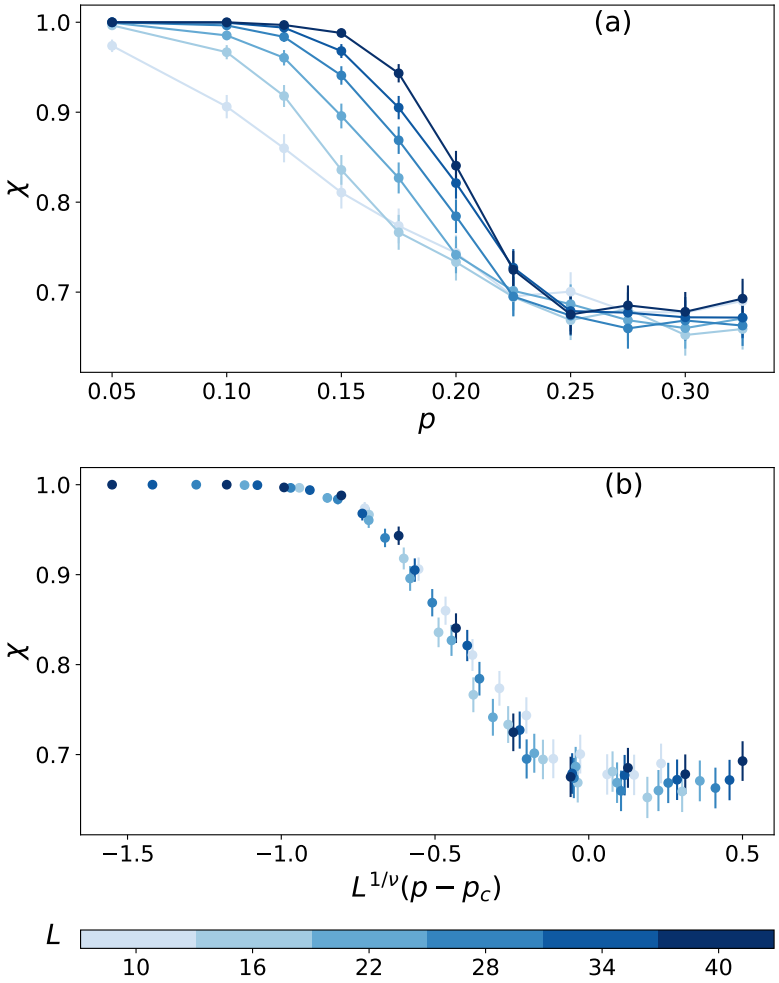


Figure 5.5: (a) Cross entropy χ for infinite-dimensional systems with up to 20 physical qubits (corresponding to a system size of L =40 qubits before compression) computed on *ibm_sherbrooke*. (b) Collapse of cross entropy curves near the critical point obtained by minimizing the scatter of all points to an unknown scaling function. The fitting procedure gives a critical measurement rate of $p_c = 0.26 \pm 0.02$ and critical exponent $\nu = 1.9 \pm 0.4$.

Fig. 5.5 shows χ for p increasing from 0.05 to 0.325. The qualitative features of χ in the all-to-all case are similar to the 1D-chain case, with larger values of χ for larger systems when $p < p_c$, crossing of all χ for different L at critical value of p, and a plateau to a constant for $p > p_c$. The critical values we extract from fitting to

the finite size scaling form Eq. (5.8) are $p_c = 0.26 \pm 0.02$ and $\nu = 1.9 \pm 0.4$ at the 90% confidence level.

A mean-field analysis of all-to-all circuits [328] predicts $\nu \approx 2.5$, and numerical simulations of uncompressed noisy circuits in the presence of a 0.1% erasure channel predicts $\nu \approx 0.8$ (see Appendix Fig. 9.7). The presence of noise evidently reduces the value of ν, and may explain why the ν obtained from experiments is lower than the theoretical value of 2.5. The increased value of p_c for the infinite-dimensional case compared to the 1D case is consistent with the intuitive picture that entanglement in a system with high connectivity is more stable to measurements that in one with low connectivity.

5.6 Resource analysis

We now discuss the resource requirements of this protocol compared to previous studies of measurement-induced phase transitions. The demonstrations reported here required fewer than 8 device-hours while retaining the hardware implementation of mid-circuit measurements. This value is more than two orders of magnitude less than the 5200 device-hours in Ref. [304] Additionally, we were able to demonstrate this protocol using 22 hardware qubits, corresponding to uncompressed systems of up to 44 qubits. This size regime is presently inaccessible to protocols requiring post-selection.

In particular, the total number of device-hours required was reduced from above 5,200 device-hours in Ref. [304] to less than 8 device-hours in our demonstration while retaining the hardware implementation of mid-circuit measurements.

With the use of circuit compression, we expect that larger systems with as many as 30 physical qubits could be accessed while still maintaining the fidelity of the current experiments. The limiting factor in the present demonstration is only the computational cost of the circuit compression which scales polynomially with system size. As this work focused on demonstrating the protocol on near-term hardware, we did not emphasize efficient implementations of the classical circuit-compression algorithm; this task could be a focus of future work. To increase to larger system sizes, the bulk and encoding ratios can also be reduced from 3, used in our experiments, to as low as 1 while still maintaining a visible phase transition.

Improvement of the experimental performance of the processor, for instance by reducing cross-talk and introducing carefully tailored dynamical decoupling sequences may also allow us to explore the phase transition in even larger systems.

Preliminary experiments including dynamical decoupling show some improvement in the fidelity obtained in the intermediate regime of 5 to 8 hardware qubits; however, for larger systems dynamical decoupling had little effect and so was not used in any of the experiments, see SI Sec. 9.6. We have also attempted a readout error mitigation for our $\rho \neq \sigma$ experiments in 1D, which did not change our results and was not applied to our data, see SI Sec. 9.6.

5.7 Discussion

Our results show that MIPTs can be studied efficiently for systems with various connectivities on near-term superconducting quantum hardware, when restricted to Clifford circuits with an arbitrary initial state. The cross-entropy protocol used in this Letter eliminates both of the exponential bottlenecks in previous studies of MIPTs on superconducting hardware [304] while preserving the mid-circuit measurements in the bulk of the circuit, providing a benchmark for the quality of mid-circuit measurements in near-term quantum hardware. In future work, this protocol may be extended to extract other critical exponents using a different circuit structure [310], or to detect other related phenomena [295, 297, 329–332].

A technique we used throughout this work is Clifford circuit compression, which takes a circuit with non-stabilizer initial states and outputs a gate-efficient representation for it. Circuit compression allows us to study systems larger than the number of available hardware qubits while minimizing the number of mid-circuit measurements, which is the slowest element of hybrid circuits. Circuit compression also makes possible the exploration of related phenomena on graphs that cannot be embedded in 2D, for example on those with all-to-all connectivity [292, 293] or on trees [285, 319]. From the cross entropy data for $\rho \neq \sigma$ we can extract critical exponents that are evidently comparable to classical simulations and theoretical predictions, even though no error mitigation techniques are applied. On the other hand, the circuit compression complicates the propagation of noise, whose analysis we leave for future work.

Comparing $\chi_{\rho=\sigma}$ (Fig. 5.3) with $\chi_{\rho\neq\sigma}$ (Fig. 5.4a), we find that the former is often visibly smaller than the latter, particularly for the larger values of p we accessed in our experiments. On the other hand, as we show in Appendix 9.2 with rigorous arguments, one has the bound $\chi_{\rho=\sigma} \geq \chi_{\rho\neq\sigma}$ in Clifford circuits with a simple noise model, namely those that can be written as stabilizer operations and their probabilistic mixtures. These include the erasure errors we use in our classical numerical

simulations. We attribute the violation of this bound to real device error, which necessarily involves e.g. coherent and non-unital noise which are not included in our simple noise model. Evidently, $\chi_{\rho=\sigma}$ is more sensitive to noise than when two different initial states are used. It will be an interesting future direction to explore the effects and the description of real device noise on the critical properties, and conversely, the extent to which a phase transition in cross entropy can be informative of experimental conditions and changes, such as dynamical decoupling, for large systems where process tomography is too costly.

5.8 Summary

In this work, we demonstrate the detection of MIPT on IBM superconducting devices by a prototypical hybrid Clifford circuit model on IBM superconducting devices. Compared with our previous work, this work employs the linear cross-entropy benchmarking protocol and overcomes the notorious post-selection issues. We also introduce a Clifford circuit compression algorithm that significantly reduces the circuit depth. With the linear cross-entropy benchmarking and the Clifford circuit compression algorithm, we realize the detection of MIPT with up to 22 physical qubits using only 8 device hours, which significantly improves from our previous work, which used over 5200 device hours. Detection of MIPT is observed on both 1D-connected and all-to-all-connected circuit models. We also compare the critical exponents with the theoretical predictions and find good agreements. This work paves the way for studies of other critical phenomena on near-term quantum hardware and provides a potential benchmarking tool for quantum circuits with mid-circuit measurements.

Chapter 6

APPENDIX OF CHAPTER 2

6.1 Feature definition

Tables 6.1 and 6.2 specify how the improved feature vector $\mathbf{f}_{ij}$ corresponding to the (i,j) pair of occupied orbitals. The feature vector is composed of several blocks listed in each row of the tables. Indices $k,l,...$ indicate occupied orbitals except for i and j, while indices a,b indicate virtual orbitals. A sorting is applied to the blocks with more than one element, and the sorting criterion of each block is listed in the tables.

Table 6.1: Diagonal part of the improved feature. The importance I is, if necessary, defined within a row.

Name	Value	Sorting Criterion
F_{ii}	$\langle i\|\hat{f}\|i\rangle$	—
$\mathbf{F}_{ik}$	$\left\|\langle i\|\hat{f}\|k\rangle\right\|$	$\langle ii\|kk\rangle^3$
$\mathbf{F}_{aa}$	$\langle a\|\hat{f}\|a\rangle \cdot I$	$I = \langle ii\|aa\rangle^3$
$\mathbf{F}_{ab}$	$\left\|\langle a\|\hat{f}\|b\rangle\right\| \cdot I$	$I = \frac{\langle ii\|aa\rangle\langle aa\|bb\rangle\langle ii\|bb\rangle}{(2F_{aa}-2F_{ii})(2F_{bb}-2F_{ii})}$
$[\boldsymbol{\mathcal{K}}^{ii}]_{ii}$	$\langle ii\|ii\rangle^3$	–
$[\boldsymbol{\mathcal{K}}^{ii}]_{kk}$	$\langle ii\|kk\rangle^3$	$\langle ii\|kk\rangle^3$
$[\boldsymbol{\mathcal{K}}^{ii}]_{aa}$	$\langle ii\|aa\rangle^3$	$\langle ii\|aa\rangle^3$
$[\boldsymbol{\mathcal{K}}^{aa}]_{aa}$	$\langle aa\|aa\rangle^3 \cdot I$	$I = \langle ii\|aa\rangle^3$
$[\boldsymbol{\mathcal{K}}^{aa}]_{bb}$	$\langle aa\|bb\rangle^3 \cdot I$	$I = \frac{\langle ii\|aa\rangle\langle aa\|bb\rangle\langle ii\|bb\rangle}{(2F_{aa}-2F_{ii})(2F_{bb}-2F_{ii})}$
$[\boldsymbol{\mathcal{K}}^{ik}]_{ik}$	$\langle ik\|ik\rangle$	$\langle ik\|ik\rangle$
$[\boldsymbol{\mathcal{K}}^{ia}]_{ia}$	$\langle ia\|ia\rangle$	$\langle ia\|ia\rangle$
$[\boldsymbol{\mathcal{K}}^{ab}]_{ab}$	$\langle ab\|ab\rangle \cdot I$	$I = \frac{\langle ia\|ia\rangle\langle ab\|ab\rangle\langle ib\|ib\rangle}{(2F_{aa}-2F_{ii})(2F_{bb}-2F_{ii})}$

Table 6.2: Off-diagonal part of the improved feature. The importance I is, if necessary, defined within a row. The damping function is defined as $G(r_{ij}) = \frac{1}{1+\frac{1}{6}(r_{ij}/r_0)^6}$

Name	Value	Sorting Criterion
$F_{\tilde{i}\tilde{i}}$	$G(r_{ij}) \cdot \left(\frac{1}{2}\langle i\vert\hat{f}\vert i\rangle + \frac{1}{2}\langle j\vert\hat{f}\vert j\rangle + \left\vert\langle i\vert\hat{f}\vert j\rangle\right\vert\right)$	—
$F_{\tilde{i}\tilde{j}}$	$G(r_{ij}) \cdot \left\vert\langle \tilde{i}\vert\hat{f}\vert\tilde{j}\rangle\right\vert$	—
$F_{\tilde{j}\tilde{j}}$	$G(r_{ij}) \cdot \left(\frac{1}{2}\langle i\vert\hat{f}\vert i\rangle + \frac{1}{2}\langle j\vert\hat{f}\vert j\rangle - \left\vert\langle i\vert\hat{f}\vert j\rangle\right\vert\right)$	—
$\mathbf{F}_{\tilde{i}k}$	$G(r_{ij}) \cdot \left(\frac{1}{\sqrt{2}}\left\vert\langle i\vert\hat{f}\vert k\rangle\right\vert + \frac{1}{\sqrt{2}}\left\vert\langle j\vert\hat{f}\vert k\rangle\right\vert\right)$	$\frac{1}{\sqrt{2}}\langle ii\vert kk\rangle^3 + \frac{1}{\sqrt{2}}\langle jj\vert kk\rangle^3$
$\mathbf{F}_{\tilde{j}k}$	$G(r_{ij}) \cdot \left\vert\frac{1}{\sqrt{2}}\left\vert\langle i\vert\hat{f}\vert k\rangle\right\vert - \frac{1}{\sqrt{2}}\left\vert\langle j\vert\hat{f}\vert k\rangle\right\vert\right\vert$	$\left\vert\frac{1}{\sqrt{2}}\langle ii\vert kk\rangle^3 - \frac{1}{\sqrt{2}}\langle jj\vert kk\rangle^3\right\vert$
$\mathbf{F}_{aa}$	$G(r_{ij}) \cdot \langle a\vert\hat{f}\vert a\rangle \cdot I$	$I = \left\vert\frac{1}{2}\langle ii\vert aa\rangle^3 - \frac{1}{2}\langle jj\vert aa\rangle^3\right\vert$
$\mathbf{F}_{ab}$	$G(r_{ij}) \cdot \left\vert\langle a\vert\hat{f}\vert b\rangle\right\vert \cdot I$	$I = \frac{(\left\vert\frac{1}{2}\langle ii\vert aa\rangle - \frac{1}{2}\langle jj\vert aa\rangle\right\vert)\langle aa\vert bb\rangle(\left\vert\frac{1}{2}\langle ii\vert bb\rangle - \frac{1}{2}\langle jj\vert bb\rangle\right\vert)}{(2F_{aa}-F_{ii}-F_{jj})(2F_{bb}-F_{ii}-F_{jj})}$
$[\boldsymbol{\kappa}^{\tilde{i}\tilde{i}}]_{\tilde{i}\tilde{i}}$	$G(r_{ij}) \left(\cdot\frac{1}{2}\langle ii\vert ii\rangle^3 + \frac{1}{2}\langle jj\vert jj\rangle^3 + \langle ii\vert jj\rangle^3\right)$	—
$[\boldsymbol{\kappa}^{\tilde{i}\tilde{i}}]_{\tilde{j}\tilde{j}}$	$G(r_{ij}) \cdot \left\vert\frac{1}{2}\langle ii\vert ii\rangle^3 - \frac{1}{2}\langle jj\vert jj\rangle^3\right\vert$	—
$[\boldsymbol{\kappa}^{\tilde{j}\tilde{j}}]_{\tilde{j}\tilde{j}}$	$G(r_{ij}) \cdot \left(\frac{1}{2}\langle ii\vert ii\rangle^3 + \frac{1}{2}\langle jj\vert jj\rangle^3 - \langle ii\vert jj\rangle^3\right)$	—
$[\boldsymbol{\kappa}^{\tilde{i}\tilde{i}}]_{kk}$	$G(r_{ij}) \cdot \left(\frac{1}{\sqrt{2}}\langle ii\vert kk\rangle^3 + \frac{1}{\sqrt{2}}\langle jj\vert kk\rangle^3\right)$	$\frac{1}{\sqrt{2}}\langle ii\vert kk\rangle^3 + \frac{1}{\sqrt{2}}\langle jj\vert kk\rangle^3$
$[\boldsymbol{\kappa}^{\tilde{j}\tilde{j}}]_{kk}$	$G(r_{ij}) \cdot \left\vert\frac{1}{\sqrt{2}}\langle ii\vert kk\rangle^3 - \frac{1}{\sqrt{2}}\langle jj\vert kk\rangle^3\right\vert$	$\left\vert\frac{1}{\sqrt{2}}\langle ii\vert kk\rangle^3 - \frac{1}{\sqrt{2}}\langle jj\vert kk\rangle^3\right\vert$
$[\boldsymbol{\kappa}^{\tilde{i}\tilde{i}}]_{aa}$	$G(r_{ij}) \cdot \left(\frac{1}{\sqrt{2}}\langle ii\vert aa\rangle^3 + \frac{1}{\sqrt{2}}\langle jj\vert aa\rangle^3\right)$	$\frac{1}{\sqrt{2}}\langle ii\vert aa\rangle^3 + \frac{1}{\sqrt{2}}\langle jj\vert aa\rangle^3$
$[\boldsymbol{\kappa}^{\tilde{j}\tilde{j}}]_{aa}$	$G(r_{ij}) \cdot \left\vert\frac{1}{\sqrt{2}}\langle ii\vert aa\rangle^3 - \frac{1}{\sqrt{2}}\langle jj\vert aa\rangle^3\right\vert$	$\left\vert\frac{1}{\sqrt{2}}\langle ii\vert aa\rangle^3 - \frac{1}{\sqrt{2}}\langle jj\vert aa\rangle^3\right\vert$
$[\boldsymbol{\kappa}^{aa}]_{aa}$	$G(r_{ij}) \cdot \langle aa\vert aa\rangle^3 \cdot I$	$I = \left\vert\frac{1}{2}\langle ii\vert aa\rangle^3 - \frac{1}{2}\langle jj\vert aa\rangle^3\right\vert$
$[\boldsymbol{\kappa}^{aa}]_{bb}$	$G(r_{ij}) \cdot \langle aa\vert bb\rangle^3 \cdot I$	$I = \frac{(\left\vert\frac{1}{2}\langle ii\vert aa\rangle - \frac{1}{2}\langle jj\vert aa\rangle\right\vert)\langle aa\vert bb\rangle(\left\vert\frac{1}{2}\langle ii\vert bb\rangle - \frac{1}{2}\langle jj\vert bb\rangle\right\vert)}{(2F_{aa}-F_{ii}-F_{jj})(2F_{bb}-F_{ii}-F_{jj})}$
$[\boldsymbol{\kappa}^{\tilde{i}\tilde{j}}]_{\tilde{i}\tilde{j}}$	$G(r_{ij}) \cdot \left\vert\frac{1}{2}\langle ii\vert ii\rangle - \frac{1}{2}\langle jj\vert jj\rangle\right\vert$	—
$[\boldsymbol{\kappa}^{\tilde{i}k}]_{\tilde{i}k}$	$G(r_{ij}) \cdot \left(\frac{1}{\sqrt{2}}\langle ik\vert ik\rangle + \frac{1}{\sqrt{2}}\langle jk\vert jk\rangle\right)$	$\frac{1}{\sqrt{2}}\langle ik\vert ik\rangle + \frac{1}{\sqrt{2}}\langle jk\vert jk\rangle$
$[\boldsymbol{\kappa}^{\tilde{j}k}]_{\tilde{j}k}$	$G(r_{ij}) \cdot \left\vert\frac{1}{\sqrt{2}}\langle ik\vert ik\rangle - \frac{1}{\sqrt{2}}\langle jk\vert jk\rangle\right\vert$	$\left\vert\frac{1}{\sqrt{2}}\langle ik\vert ik\rangle - \frac{1}{\sqrt{2}}\langle jk\vert jk\rangle\right\vert$
$[\boldsymbol{\kappa}^{\tilde{i}a}]_{\tilde{i}a}$	$G(r_{ij}) \cdot \left(\frac{1}{\sqrt{2}}\langle ia\vert ia\rangle + \frac{1}{\sqrt{2}}\langle jk\vert jk\rangle\right)$	$\frac{1}{\sqrt{2}}\langle ia\vert ia\rangle + \frac{1}{\sqrt{2}}\langle jk\vert jk\rangle$
$[\boldsymbol{\kappa}^{\tilde{j}a}]_{\tilde{j}a}$	$G(r_{ij}) \cdot \left\vert\frac{1}{\sqrt{2}}\langle ia\vert ia\rangle - \frac{1}{\sqrt{2}}\langle ja\vert ja\rangle\right\vert$	$\left\vert\frac{1}{\sqrt{2}}\langle ia\vert ia\rangle - \frac{1}{\sqrt{2}}\langle ja\vert ja\rangle\right\vert$
$[\boldsymbol{\kappa}^{ab}]_{ab}$	$G(r_{ij}) \cdot \langle ab\vert ab\rangle \cdot I$	$I = \frac{(\left\vert\frac{1}{2}\langle ia\vert ia\rangle - \frac{1}{2}\langle ja\vert ja\rangle\right\vert)\langle ab\vert ab\rangle(\left\vert\frac{1}{2}\langle ib\vert ib\rangle - \frac{1}{2}\langle jb\vert jb\rangle\right\vert)}{(2F_{aa}-F_{ii}-F_{jj})(2F_{bb}-F_{ii}-F_{jj})}$

6.2 Rotational equivariance of dipole model

In this section, we prove that rotational equivariance is satisfied by both the single-task and multi-task GPR model with derivatives, respectively.

We first prove the rotational equivariance for the single-task GPR model.

Theorem 1. *For an n-dimensional external variable $\boldsymbol{\alpha} \in \mathbb{R}^n$, we train single-task*

GPR with derivatives using features $\mathbf{X}_c = (\mathbf{X}, \nabla_{\boldsymbol{\alpha}}\mathbf{X})$ *and labels* $\nabla_{\boldsymbol{\alpha}}\mathbf{y}$ *and generate the predicted distribution* $\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star)$ *on test features* $\mathbf{X}_c^\star$*. Assume that the transformation of features and labels under rotation operator* $\hat{U}$ *is given by*

$$\hat{U}\mathbf{X} = \mathbf{X}, \hat{U}\nabla_{\boldsymbol{\alpha}}\mathbf{X} = U\nabla_{\boldsymbol{\alpha}}\mathbf{X}, \hat{U}\mathbf{y} = \mathbf{y}, \hat{U}\nabla_{\boldsymbol{\alpha}}\mathbf{y} = U\nabla_{\boldsymbol{\alpha}}\mathbf{y}, \tag{6.1}$$

we show that $\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star)$ *satisfies rotational equivariance:*

$$\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) = U\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star). \tag{6.2}$$

Proof. In single-task GPR with derivatives, the kernel matrix element of $x_c = (x, \nabla_{\boldsymbol{\alpha}} x)$, $x_c' = (x', \nabla_{\boldsymbol{\alpha}} x')$ is

$$K_s(x_c, x_c') = \nabla_{\boldsymbol{\alpha}} x \nabla_x \nabla_{x'} K(x, x')(\nabla_{\boldsymbol{\alpha}} x')^T. \tag{6.3}$$

Let the Gaussian likelihood covariance to be Σ_s. The prediction on $\mathbf{X}_c^\star$ is a Gaussian distributed function $\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star)$ with the mean and variance

$$\mathbb{E}[\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star)] = K_s(\mathbf{X}_c^\star, \mathbf{X}_c)(K_s(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_s)^{-1}\nabla_{\boldsymbol{\alpha}}\mathbf{y}, \tag{6.4}$$

$$\mathrm{Var}[\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star)] = K_s(\mathbf{X}_c^\star, \mathbf{X}_c)(K_s(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_s)^{-1} K_s(\mathbf{X}_c, \mathbf{X}_c^\star). \tag{6.5}$$

So the rotational equivariance is equivalent to

$$\mathbb{E}[\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star)] = U\mathbb{E}[\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star)],$$
$$\mathrm{Var}[\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star)] = U\mathrm{Var}[\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star)]U^T.$$

Note that

$$K_s(\hat{U}\mathbf{X}_c^\star, \mathbf{X}_c) = U\nabla_{\boldsymbol{\alpha}}\mathbf{X}^\star\nabla_{x^\star}\nabla_{\mathbf{X}} K(\mathbf{X}^\star, \mathbf{X})(\nabla_{\boldsymbol{\alpha}}\mathbf{X})^T = UK_s(\mathbf{X}_c^\star, \mathbf{X}_c), \tag{6.6}$$

$$K_s(\mathbf{X}_c, \hat{U}\mathbf{X}_c^\star) = \nabla_{\boldsymbol{\alpha}}\mathbf{X}\nabla_{\mathbf{X}}\nabla_{x^\star} K(\mathbf{X}, \mathbf{X}^\star)(U\nabla_{\boldsymbol{\alpha}}\mathbf{X}^\star)^T = K_s(\mathbf{X}_c, \mathbf{X}_c^\star)U^T. \tag{6.7}$$

Thus we have

$$\mathbb{E}[\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star)] = K_s(\hat{U}\mathbf{X}_c^\star, \mathbf{X}_c)(K_x(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_s)^{-1}\nabla_{\boldsymbol{\alpha}}\mathbf{y} \tag{6.8}$$

$$= UK_s(\mathbf{X}_c^\star, \mathbf{X}_c)(K_x(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_s)^{-1}\nabla_{\boldsymbol{\alpha}}\mathbf{y} \tag{6.9}$$

$$= U\mathbb{E}[\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star)] \tag{6.10}$$

and

$$\mathrm{Var}[\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star)] = K_s(\hat{U}\mathbf{X}_c^\star, \mathbf{X}_c)(K_x(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_s)^{-1} K_s(\mathbf{X}_c, \hat{U}\mathbf{X}_c^\star) \tag{6.11}$$

$$= UK_s(\mathbf{X}_c^\star, \mathbf{X}_c)(K_x(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_s)^{-1} K_s(\mathbf{X}_c, \mathbf{X}_c^\star)U^T \tag{6.12}$$

$$= U\mathrm{Var}[\nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star)]U^T. \tag{6.13}$$

which finish the proof. □

In MOB-ML dipole learning, the features $\boldsymbol{f}_{ij}^{\boldsymbol{\mu}} = \{\boldsymbol{f}_{ij}^{\varepsilon}, \nabla_{\mathscr{E}} \boldsymbol{f}_{ij}^{\varepsilon}\}$ and the pair dipole moment labels satisfy the condition in Eq. 6.1, and thus the prediction $\boldsymbol{\mu}^{\mathrm{ML}}$ is rotational equivariant.

We then prove the rotational equivariance for the multi-task GPR model.

Theorem 2. *For an n-dimensional external variable $\boldsymbol{\alpha} \in \mathbb{R}^n$, we train multi-task GPR with derivatives using features $\mathbf{X}_c = (\mathbf{X}, \nabla_{\boldsymbol{\alpha}}\mathbf{X})$ and labels $\mathbf{y}_c = (\mathbf{y}, \nabla_{\boldsymbol{\alpha}}\mathbf{y})$ and generate the predicted distribution $\begin{bmatrix} f(\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star) \end{bmatrix}$ on test features $\mathbf{X}_c^\star$. Assume that the transformation of features and labels under rotation operator $\hat{U}$ is given by*

$$\hat{U}\mathbf{X} = \mathbf{X}, \hat{U}\nabla_{\boldsymbol{\alpha}}\mathbf{X} = U\nabla_{\boldsymbol{\alpha}}\mathbf{X}, \hat{U}\mathbf{y} = \mathbf{y}, \hat{U}\nabla_{\boldsymbol{\alpha}}\mathbf{y} = U\nabla_{\boldsymbol{\alpha}}\mathbf{y}, \tag{6.14}$$

we show that $f(\mathbf{X}_c^\star)$ satisfies rotaional invariance and $\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star)$ satisfies rotational equivariance:

$$\begin{bmatrix} f(\hat{U}\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) \end{bmatrix} = \begin{bmatrix} f(\mathbf{X}_c^\star) \\ U\nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star) \end{bmatrix}. \tag{6.15}$$

Proof. In multi-task GPR with derivatives, the kernel matrix element for $x_c = (x, \nabla_{\boldsymbol{\alpha}} x)$, $x_c' = (x', \nabla_{\boldsymbol{\alpha}} x')$ is

$$K_m(x_c, x_c') = \begin{bmatrix} K(x,x') & \nabla_{x'}K(x,x')(\nabla_{\boldsymbol{\alpha}}x')^T \\ \nabla_{\boldsymbol{\alpha}}x\nabla_x K(x,x') & \nabla_{\boldsymbol{\alpha}}x\nabla_x\nabla_{x'}K(x,x')(\nabla_{\boldsymbol{\alpha}}x')^T \end{bmatrix}. \tag{6.16}$$

Let the Gaussian likelihood covariance to be Σ_m. The prediction of $\mathbf{X}_c^\star$ is a Gaussian distributed function $\begin{bmatrix} f(\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star) \end{bmatrix}$ with the mean and variance

$$\mathbb{E}[\begin{bmatrix} f(\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star) \end{bmatrix}] = K_m(\mathbf{X}_c^\star, \mathbf{X}_c)(K_m(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_m)^{-1}\mathbf{y}_c, \tag{6.17}$$

$$\mathrm{Var}[\begin{bmatrix} f(\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star) \end{bmatrix}] = K_m(\mathbf{X}_c^\star, \mathbf{X}_c)(K_m(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_m)^{-1}K_m(\mathbf{X}_c, \mathbf{X}_c^\star). \tag{6.18}$$

So the rotational equivariance is equivalent to

$$\mathbb{E}[\begin{bmatrix} f(\hat{U}\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) \end{bmatrix}] = \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} \mathbb{E}[\begin{bmatrix} f(\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star) \end{bmatrix}], \tag{6.19}$$

$$\mathrm{Var}[\begin{bmatrix} f(\hat{U}\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) \end{bmatrix}] = \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} \mathrm{Var}[\begin{bmatrix} f(\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\mathbf{X}_c^\star) \end{bmatrix}] \begin{bmatrix} I & 0 \\ 0 & U^T \end{bmatrix}. \tag{6.20}$$

Note that

$$K_m(\hat{U}\mathbf{X}_c^\star, \mathbf{X}_c) = \begin{bmatrix} K(\mathbf{X}^\star, \mathbf{X}) & \nabla_{\mathbf{X}} K(\mathbf{X}^\star, \mathbf{X})(\nabla_{\boldsymbol{\alpha}}\mathbf{X})^T \\ U\nabla_{\boldsymbol{\alpha}}\mathbf{X}^\star\nabla_{x^\star}K(\mathbf{X}^\star, \mathbf{X}) & U\nabla_{\boldsymbol{\alpha}}\mathbf{X}^\star\nabla_{x^\star}\nabla_{\mathbf{X}}K(\mathbf{X}^\star, \mathbf{X})(\nabla_{\boldsymbol{\alpha}}\mathbf{X})^T \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} K_m(\mathbf{X}_c^\star, \mathbf{X}_c), \tag{6.21}$$

$$K_m(\mathbf{X}_c, \hat{U}\mathbf{X}_c^\star) = \begin{bmatrix} K(\mathbf{X}, \mathbf{X}^\star) & \nabla_{x^\star} K(\mathbf{X}, \mathbf{X}^\star)(U\nabla_{\boldsymbol{\alpha}}\mathbf{X}^\star)^T \\ \nabla_{\boldsymbol{\alpha}}\mathbf{X}\nabla_{\mathbf{X}}K(\mathbf{X}, \mathbf{X}^\star) & \nabla_{\boldsymbol{\alpha}}\mathbf{X}\nabla_{\mathbf{X}}\nabla_{x^\star}K(\mathbf{X}, \mathbf{X}^\star)(U\nabla_{\boldsymbol{\alpha}}\mathbf{X}^\star)^T \end{bmatrix} = K_m(\mathbf{X}_c, \mathbf{X}_c^\star) \begin{bmatrix} I & 0 \\ 0 & U^T \end{bmatrix}. \tag{6.22}$$

Thus we have

$$\mathbb{E}[\begin{bmatrix} f(\hat{U}\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) \end{bmatrix}] = K_m(\hat{U}\mathbf{X}_c^\star, \mathbf{X}_c)(K_m(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_m)^{-1}\mathbf{y}_c, \tag{6.23}$$

$$= \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} K_m(\mathbf{X}_c^\star, \mathbf{X}_c)(K_m(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_m)^{-1}\mathbf{y}_c, \tag{6.24}$$

$$= \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} \mathbb{E}[\begin{bmatrix} f(\hat{U}\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) \end{bmatrix}], \tag{6.25}$$

and

$$\mathrm{Var}[\begin{bmatrix} f(\hat{U}\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) \end{bmatrix}] = K_m(\hat{U}\mathbf{X}_c^\star, \mathbf{X}_c)(K_m(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_m)^{-1}K_m(\mathbf{X}_c, \hat{U}\mathbf{X}_c^\star), \tag{6.26}$$

$$= \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} K_s(\mathbf{X}_c^\star, \mathbf{X}_c)(K_m(\mathbf{X}_c, \mathbf{X}_c) + \Sigma_m)^{-1}K_m(\mathbf{X}_c, \mathbf{X}_c^\star) \begin{bmatrix} I & 0 \\ 0 & U^T \end{bmatrix}, \tag{6.27}$$

$$= \begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix} \mathrm{Var}[\begin{bmatrix} f(\hat{U}\mathbf{X}_c^\star) \\ \nabla_{\boldsymbol{\alpha}} f(\hat{U}\mathbf{X}_c^\star) \end{bmatrix}] \begin{bmatrix} I & 0 \\ 0 & U^T \end{bmatrix}, \tag{6.28}$$

which finish the proof. □

In MOB-ML energy+dipole learning, the features $\boldsymbol{f}_{ij}^{\boldsymbol{\mu}} = \{\boldsymbol{f}_{ij}^{\varepsilon}, \nabla_{\mathscr{E}}\boldsymbol{f}_{ij}^{\varepsilon}\}$ and the labels (pair energies and pair dipoles) satisfy the condition in Eq. 6.14, and thus the energy prediction $\varepsilon^{\mathrm{ML}}$ is rotational invariant and the dipole prediction $\boldsymbol{\mu}^{\mathrm{ML}}$ is rotational equivariant.

6.3 Analysis of MOBML prediction for polyenoic amino acid

The true and predicted contribution of dipole moments of polyenoic amino acid is reported in Table 6.3, and the values on each orbital pairs are reported in Table 6.4.

Table 6.3: True and predicted contribution (debye) of polyenoic amino acid with different n values. The model is trained on QM9 110K molecules with derivative kernel. The corresponding true and predicted dipoles are shown in Figure 7c in the main text.

n	True correlation	Predicted correlation
2	-0.7241	-0.7590
4	-1.4535	-1.7515
6	-2.1242	-3.1819
8	-2.7921	-4.7708
10	-3.4290	-6.5512

Table 6.4: Average pair-wise dipole true and predicted contribution (debye) of each type of MO for the correlation part of the dipole of polyenoic amino acid with n=6. The correlation parts are computed using true MP2 as a reference (on the direction of true MP2). The off-diagonal pairs have already been summed up and included in the contributions. The CC double bond has a significantly large error. The model is trained on QM9 110K molecules with derivative kernel. The corresponding true and predicted dipoles are shown in Figure 7c in the main text.

Connection	Bond order	True correlation	Predicted correlation
CC	1	0.1985	0.2162
CC	**2**	**-0.1946**	**-0.0466**
CN	1	0.4573	0.4793
CO	1	-0.2568	-0.2742
CO	2	-0.6463	-0.6508
CH	1	-0.0113	-0.0119
NH	1	-0.0843	-0.0897
OH	1	0.1780	0.1570
N	–	0.9734	0.9591
O	–	-0.1056	-0.1233

Chapter 7

APPENDIX OF CHAPTER 3

7.1 Necessary and sufficient condition for eigenvalues of a 2×2 real matrix to be inside the unit circle

This section provides a proof of the standard result used in Sections 3.2 and 3.2 to infer ergodicity of the T-RPMD update for free and harmonically-confined ring polymers.

Theorem 3. *The spectral radius of a 2×2 real matrix $\boldsymbol{M}$ is strictly less than one if and only if*

$$|tr(\boldsymbol{M})| < 1 + \det(\boldsymbol{M}) < 2 \,. \tag{7.1}$$

Fig. 7.1 plots eigenvalue pairs λ_1, λ_2 that satisfy Eq. 7.1 for a fixed value of $\det(\boldsymbol{M}) = \lambda_1 \lambda_2$. Note that the spectral radius of $\boldsymbol{M}$ is minimized when λ_1 and λ_2 are on the circle with radius $r = \sqrt{\det(\boldsymbol{M})}$.

Proof. Let λ_1, λ_2 be the (possibly complex) eigenvalues of $\boldsymbol{M}$. By definition, the spectral radius of $\boldsymbol{M}$ is $\max(|\lambda_1|, |\lambda_2|) =: \rho$. Since $\boldsymbol{M}$ is real, both $\mathrm{tr}(\boldsymbol{M}) = \lambda_1 + \lambda_2$ and $\det(\boldsymbol{M}) = \lambda_1 \lambda_2$ are real. Thus, either:

1. λ_1, λ_2 are a complex conjugate pair; or,
2. λ_1, λ_2 are both real.

In the first case, $\lambda_1 = a + ib$ and $\lambda_2 = a - ib$ for some real numbers a and b with $b \neq 0$, and hence, $\det(\boldsymbol{M}) = \lambda_1 \lambda_2 = a^2 + b^2 > 0$, and $\rho = |\lambda_1| = |\lambda_2| = \sqrt{a^2 + b^2}$, i.e., the eigenvalues lie on the circle with radius $\rho = \sqrt{a^2 + b^2} = \sqrt{\det(\boldsymbol{M})}$. In this case, the first inequality in Eq. 7.1 holds since $b \neq 0$ implies

$$|\mathrm{tr}(\boldsymbol{M})| = 2|a| < 2\rho \leq 1 + \rho^2 = 1 + \det(\boldsymbol{M}) \,. \tag{7.2}$$

Hence, Eq. 7.1 is equivalent to $1 + \det(\boldsymbol{M}) < 2$ or $\rho < 1$.

In the second case, λ_1, λ_2 are both real, and the condition $|\mathrm{tr}(\boldsymbol{M})| < 1 + \det(\boldsymbol{M})$ is equivalent to

$$\begin{aligned} 1 + \lambda_1 \lambda_2 + \lambda_1 + \lambda_2 &= (1 + \lambda_1)(1 + \lambda_2) > 0 \text{ and }, \\ 1 + \lambda_1 \lambda_2 - \lambda_1 - \lambda_2 &= (1 - \lambda_1)(1 - \lambda_2) > 0 \,. \end{aligned} \tag{7.3}$$

Together with $\det(\boldsymbol{M}) = \lambda_1 \lambda_2 < 1$, these conditions are equivalent to $\rho = \max(|\lambda_1|, |\lambda_2|) < 1$. □

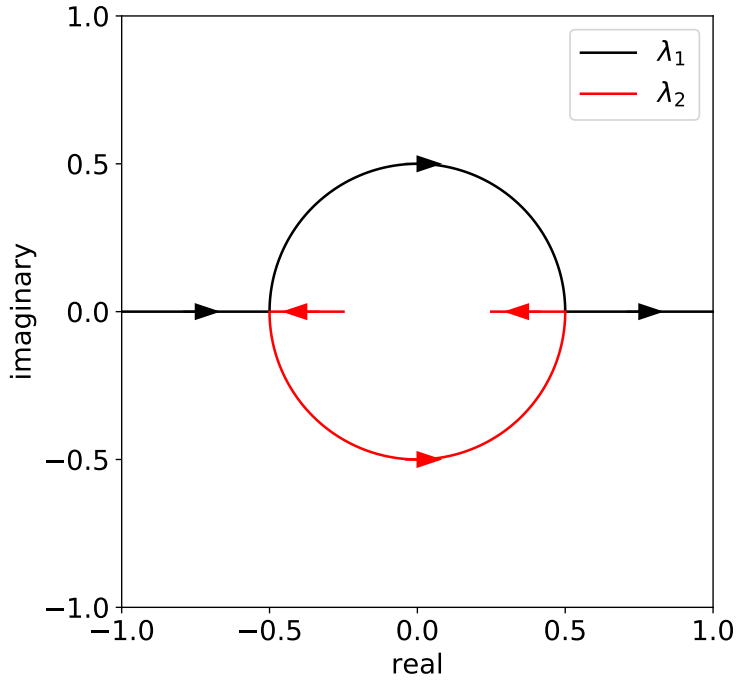


Figure 7.1: This figure plots all possible eigenvalue pairs λ_1, λ_2 of a matrix M that satisfies Eq. 7.1 with $\det(M) = \lambda_1\lambda_2 = 1/4$. The eigenvalue pairs either lie on the circle with radius $r = 1/2$ or are both real, and in the former case, the spectral radius of M is minimal.

7.2 Stability condition for harmonic external potentials

This section proves that condition (**C3**) implies property (**P3**), as claimed in Section 3.2. For notational brevity, we define

$$A(x) \;:=\; \cos(\theta(x)) - \frac{\Delta t^2(\Lambda/m)}{2}\frac{\sin(\theta(x))}{x}\,. \tag{7.4}$$

Note that $A(x)$ is equal to $\mathscr{A}_{j,n}$ in the display under Eq. 3.27 if $x = \omega_{j,n}\Delta t$.

Theorem 4. *For any $\alpha^\star > 0$, (A2) implies (A1).*

(A1) *For all $\Lambda \geq 0$, $m > 0$ and $\Delta t > 0$ satisfying $\Delta t^2\Lambda/m < \alpha^\star$, the function θ satisfies*

$$|A(x)| < 1 \quad \textit{for } x > 0\,. \tag{7.5}$$

(A2) *The function θ satisfies:*

$$0 < \theta(x) < 2\arctan(2x/\alpha^\star) \quad \textit{for } x > 0\,. \tag{7.6}$$

Proof. Let $\alpha = \Delta t^2(\Lambda/m)$. For notational brevity, define

$$\phi_\alpha(x) := \arctan(\alpha/(2x)) \quad \text{for } x > 0\,. \tag{7.7}$$

By the harmonic addition identity

$$\cos(\theta) - \tan(\phi_\alpha)\sin(\theta) = \frac{\cos(\theta+\phi_\alpha)}{\cos(\phi_\alpha)}\,, \tag{7.8}$$

note that (A1) can be rewritten as

$$\left|\frac{\cos(\theta(x)+\phi_\alpha(x))}{\cos(\phi_\alpha(x))}\right| < 1 \quad \text{for } x>0,\ 0<\alpha<\alpha^\star\,. \tag{7.9}$$

For $0<\theta(x)<\pi$, Eq. 7.9 holds if and only if

$$\phi_\alpha(x) < \theta(x)+\phi_\alpha(x) < \pi - \phi_\alpha(x)\,, \tag{7.10}$$

which can be rewritten as

$$0 < \theta(x) < 2\arctan(2x/\alpha)\,, \tag{7.11}$$

where we used the identity

$$\pi - 2\arctan(x) = 2\arctan(1/x) \quad \text{valid for } x>0\,. \tag{7.12}$$

Since arctan is monotone increasing, and $0<\alpha<\alpha^\star$ by assumption, we may conclude that

$$0 < \theta(x) < 2\arctan(2x/\alpha^\star) < 2\arctan(2x/\alpha)\,. \tag{7.13}$$

Thus, if (A2) holds, then Eq. 7.11 holds and therefore (A1) holds. □

Fix $\varepsilon \in (0,1)$. Since Theorem 4 is true for arbitrary $\alpha^\star$, if we take $\alpha^\star = 4-\varepsilon$, then the theorem holds with $\Delta t^2\Lambda/m < 4-\varepsilon$ in Theorem 4 (A1), and $\theta(x) < 2\arctan(2x/(4-\varepsilon))$ in Theorem 4 (A2). Since $\varepsilon>0$ is arbitrary, and arctan is monotone increasing, we can conclude that the theorem holds with $\Delta t^2\Lambda/m<4$ and $\theta(x)\le 2\arctan(x/2)$. Summarizing,

Corollary 5. *Suppose that the function θ satisfies*

$$0 < \theta(x) \le 2\arctan(x/2) \quad \textit{for } x>0\,. \tag{7.14}$$

Then for all $\Lambda\ge 0$, $m>0$ and $\Delta t>0$ satisfying $\Delta t^2\Lambda/m<4$, we have

$$|A(x)| \;<\; 1 \quad \textit{for } x>0\,. \tag{7.15}$$

7.3 Dimension-free quantitative contraction rate for harmonic external potentials in the infinite-friction limit

In the infinite-friction limit, Eq. 3.31 simplifies to

$$\mathscr{M}_{j,n} = \mathscr{B}^{1/2}\mathscr{S}_{j,n}^{1/2}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\mathscr{S}_{j,n}^{1/2}\mathscr{B}^{1/2} \text{ and}$$

$$\mathscr{R}_{j,n} = \frac{1}{\beta m_n}\mathscr{B}^{1/2}\mathscr{S}_{j,n}^{1/2}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}(\mathscr{B}^{1/2}\mathscr{S}_{j,n}^{1/2})^{\mathrm{T}}\,.$$

The kth step of the corresponding T-RPMD integrator can be written compactly as

$$\begin{bmatrix}\rho_j^{(k)}\\ \varphi_j^{(k)}\end{bmatrix} = \mathscr{M}_{j,n}\begin{bmatrix}\rho_j^{(k-1)}\\ \varphi_j^{(k-1)}\end{bmatrix} + \mathscr{R}_{j,n}^{1/2}\begin{bmatrix}\xi_j^{(k-1)}\\ \eta_j^{(k-1)}\end{bmatrix}\,, \tag{7.16}$$

where $\xi_j^{(k-1)}$ and $\eta_j^{(k-1)}$ are independent standard normal random variables. Suppose that the initial velocity is drawn from the Maxwell–Boltzmann distribution, i.e., $\varphi_j^{(0)} \sim \mathscr{N}(0,(\beta m_n)^{-1})$ and the initial position is drawn from an arbitrary distribution μ_j on $\mathbb{R}$, i.e., $\rho_j^{(0)} \sim \mu_j$. Let $p_{j,n}^k$ denote the k-step transition kernel of the position-marginal, i.e., $\mu_j p_{j,n}^k$ is the probability distribution of $\rho_j^{(k)}$ with $\rho_j^{(0)} \sim \mu_j$.

The next theorem shows that starting from any two initial distributions μ_j and ν_j on $\mathbb{R}$, the distance between the distributions $\mu_j p_{j,n}^k$ and $\nu_j p_{j,n}^k$ is contractive. We quantify the distance between these distributions in terms of the 2-Wasserstein metric. For two probability distributions μ and ν on $\mathbb{R}$, the 2-Wasserstein distance between μ and ν is defined as

$$\mathscr{W}_2(\mu,\nu) = \left(\inf_{\substack{X\sim\mu\\ Y\sim\nu}} \mathbb{E}(|X-Y|^2)\right)^{1/2}, \tag{7.17}$$

where the infimum is taken over all bivariate random variables (X,Y) such that $X\sim\mu$ and $Y\sim\nu$.[214]

Theorem 6. *Suppose that the function θ satisfies*

$$0 < \theta(x) \le 2\arctan(x/2) \quad \textit{for } x > 0\,. \tag{7.18}$$

Then for all $k>1$, $\Lambda\ge 0$, $m>0$ and $\Delta t>0$ satisfying $\Delta t^2\Lambda/m<4$, and for all initial distributions μ_j and ν_j on $\mathbb{R}$,

$$\mathscr{W}_2(\mu_j p_{j,n}^k, \nu_j p_{j,n}^k) \le \begin{cases} A(\omega_{j,n}\Delta t)^{k-1}\mathscr{W}_2(\mu_j,\nu_j) & \textit{if } A(\omega_{j,n}\Delta t)>0,\\ \frac{1}{2}\frac{1}{k-1}\mathscr{W}_2(\mu_j,\nu_j) & \textit{else.}\end{cases} \tag{7.19}$$

Proof. In the infinite-friction limit, the eigenvalues of $\mathscr{M}_{j,n}$ are $\{0, A(\omega_{j,n}\Delta t)\}$, where $A(x)$ is defined in Appendix 7.2. Let $\rho_j^{(0)} \sim \mu_j$ and $\tilde{\rho}_j^{(0)} \sim \nu_j$ be an optimal coupling of μ_j and ν_j, i.e., $\mathscr{W}_2(\mu_j, \nu_j) = \mathbb{E}(|\rho_j^{(0)} - \tilde{\rho}_j^{(0)}|^2)^{1/2}$. Conditional on $\rho_j^{(0)}$ and $\tilde{\rho}_j^{(0)}$, $\rho_j^{(k)}$ and $\tilde{\rho}_j^{(k)}$ are Gaussian random variables with equal variances, but different means. By a well-known result for the 2-Wasserstein distance between Gaussian distributions,[215]

$$\begin{aligned}
&\mathscr{W}_2(\mu_j p_{j,n}^k, \nu_j p_{j,n}^k)^2 \\
&= |A(\omega_{j,n}\Delta t)|^{2(k-1)} (\mathscr{M}_{j,n})_{11}^2 \mathscr{W}_2(\mu_j, \nu_j)^2 \\
&= |A(\omega_{j,n}\Delta t)|^{2(k-1)} \frac{(1 + A(\omega_{j,n}\Delta t))^2}{4} \mathscr{W}_2(\mu_j, \nu_j)^2 \,,
\end{aligned} \tag{7.20}$$

where we used $(\mathscr{M}_{j,n})_{11} = (1 + A(\omega_{j,n}\Delta t))/2$.

Now we distinguish between two cases. In the case where $A(\omega_{j,n}\Delta t) > 0$, we obtain the required result since $|A(\omega_{j,n}\Delta t)| < 1$ by Corollary 5, and therefore

$$\frac{(1 + A(\omega_{j,n}\Delta t))^2}{4} \leq 1 \,. \tag{7.21}$$

Otherwise, for $-1 < A(\omega_{j,n}\Delta t) \leq 0$ the quantity $|A(\omega_{j,n}\Delta t)|^{2(k-1)}(1 + A(\omega_{j,n}\Delta t))^2$ is maximized at $(-1 + 1/k)^{2k}(k-1)^{-2}$, and therefore

$$|A(\omega_{j,n}\Delta t)|^{2(k-1)} \frac{(1 + A(\omega_{j,n}\Delta t))^2}{4} \leq \frac{1}{4(k-1)^2} \,. \tag{7.22}$$

Inserting Eq. 7.21 and Eq. 7.22 into Eq. 7.20, and then taking square roots, gives the required result. □

7.4 Total variation bound on the equilibrium accuracy error for harmonic external potentials

In this section, we show that Eq. 3.37 follows from conditions (**C1**)-(**C4**) in the setting of Section 3.2. It is helpful to recall the quantities

$$\omega_j = \lim_{n\to\infty} \omega_{j,n} = \begin{cases} \frac{\pi j}{\hbar\beta} & \text{if } j \text{ is even} \,, \\ \frac{\pi(j+1)}{\hbar\beta} & \text{else} \,. \end{cases} \tag{7.23}$$

In the following, $\mu_{j,\Delta t}$ and μ_j respectively denote the jth factor of the product distributions $\mu_{n,\Delta t}$ and μ_n introduced in Section 3.2.

Theorem 7. *Suppose that the function θ satisfies conditions* (**C1**)-(**C4**). *Then for all $\Lambda \geq 0$, $m > 0$ and $\Delta t > 0$ satisfying $\Delta t^2 \Lambda / m < 4$, the total variation distance between μ_n and $\mu_{n,\Delta t}$ is bounded as in Eq. 3.37.*

Proof. Subadditivity of the total variation distance d_{TV} between product distributions and its equivalence with the Hellinger distance[213] d_{H} lead to the inequalities

$$\begin{aligned} d_{\mathrm{TV}}(\mu_n, \mu_{n,\Delta t})^2 &\leq \sum_{j=1}^{n-1} d_{\mathrm{TV}}(\mu_j, \mu_{j,\Delta t})^2 \\ &\leq \sum_{j=1}^{n-1} 2\, d_{\mathrm{H}}(\mu_j, \mu_{j,\Delta t})^2 \leq \sum_{j=1}^{n-1} \frac{2(s_j - s_{j,\Delta t})^2}{(s_j^2 + s_{j,\Delta t}^2)} \\ &\leq \sum_{j=1}^{n-1} \left(1 - \frac{s_j}{s_{j,\Delta t}}\right)^2 \leq \sum_{j=1}^{n-1} \left(1 - \frac{s_j^2}{s_{j,\Delta t}^2}\right)^2 , \end{aligned} \tag{7.24}$$

where the second-to-last step uses Eq. 3.35 and the last step uses the elementary inequality $(1-x^2)^2 \geq (1-x)^2$ valid for all $x \geq 0$.

Since $\tan(\cdot)$ increases superlinearly on the interval $(0, \pi)$, we have $\theta(x)/2 \leq \tan(\theta(x)/2) \leq x/2$ for $x > 0$, where the second inequality uses (**C3**). Consequently, the jth summand in Eq. 7.24 admits the bound

$$\begin{aligned} \left(1 - \frac{s_j^2}{s_{j,\Delta t}^2}\right)^2 &= \left(\frac{\Lambda/m}{\omega_{j,n}^2 + \Lambda/m}\left(\frac{\omega_{j,n}\Delta t/2}{\tan\left(\theta(\omega_{j,n}\Delta t)/2\right)} - 1\right)\right)^2 \\ &\leq \left(\frac{\Delta t^2 \Lambda/m}{(\omega_{j,n}\Delta t)^2}\left(\frac{\omega_{j,n}\Delta t}{\theta(\omega_{j,n}\Delta t)} - 1\right)\right)^2 \\ &\leq \left(\frac{\Delta t^2 \Lambda}{m}\right)^2 \frac{1}{(\omega_{j,n}\Delta t)^2} , \end{aligned} \tag{7.25}$$

where the last line uses the lower bound in (**C4**). Using that for any even positive integer n

$$\sum_{j=1}^{n-1} \frac{1}{\omega_{j,n}^2} < \lim_{n\to\infty} \sum_{j=1}^{n-1} \frac{1}{\omega_{j,n}^2} = \sum_{j=1}^{\infty} \frac{1}{\omega_j^2} < \left(\frac{\hbar\beta}{\pi}\right)^2 \sum_{j=1}^{\infty} \frac{2}{j^2} , \tag{7.26}$$

where we used Eq. 7.23, the bound in Eq. 7.24 becomes

$$d_{\mathrm{TV}}(\mu_n, \mu_{n,\Delta t})^2 < \left(\frac{\Delta t^2 \Lambda}{m}\right)^2 \left(\frac{\hbar\beta}{\pi\Delta t}\right)^2 \sum_{j=1}^{\infty} \frac{2}{j^2} . \tag{7.27}$$

Taking square roots and using the Riemann zeta function[333] to evaluate the infinite sum yields Eq. 3.37. □

7.5 Asymptotic variance of kinetic energy observables for harmonic external potentials in the infinite-friction limit

In Section 3.3, Figs. 3.4b and 3.4d show that the T-RPMD scheme specified by $\theta(x) = 2\arctan(x/2)$, which coincides with the Cayley-modified BAOAB scheme

introduced in Ref. [205], provides the smallest integrated autocorrelation time (Eq. 3.40) for quantum kinetic energy observables (Eq. 3.39) among several schemes with properties (**P1**)-(**P5**). In this section, we show that this scheme minimizes an upper bound (Eq. 7.32) on the integrated autocorrelation time of the quantum kinetic energy among all dimension-free and strongly-stable BAOAB-like schemes for harmonic external potentials.

To this end, note that for a n-bead thermostatted ring polymer with external potential $V_n^{\text{ext}}(\boldsymbol{q}) = \frac{\Lambda}{2n}|\boldsymbol{q}|^2$, Eq. 3.39 can be rewritten

$$\begin{aligned} \mathsf{KE}_n^{\text{pri}}(\boldsymbol{\varrho}) &= \frac{n}{2\beta} - \sum_{j=1}^{n-1} \frac{m_n \omega_{j,n}^2}{2} \rho_j^2 \text{ and} \\ \mathsf{KE}_n^{\text{vir}}(\boldsymbol{\varrho}) &= \frac{1}{2\beta} + \sum_{j=1}^{n-1} \frac{\Lambda}{2n} \rho_j^2, \end{aligned} \tag{7.28}$$

where $\boldsymbol{\varrho}$ is defined in Eq. 3.12. In the following, we denote both observables in Eq. 7.28 as KE_n and distinguish between the two as needed.

To control the integrated autocorrelation time of KE_n, we need the stationary autocorrelation $\text{Cor}\big(\mathsf{KE}_n(\boldsymbol{\varrho}^{(0)}), \mathsf{KE}_n(\boldsymbol{\varrho}^{(k\Delta t)})\big)$ for $k \geq 0$. Note that the distributions of $\boldsymbol{\varrho}^{(k\Delta t)}$ and $\boldsymbol{\varrho}^{(0)}$ are equal by stationarity, and that components $(\boldsymbol{\rho}_j)_{j=0}^{n-1}$ are uncorrelated in a harmonic external potential. Thus,

$$\begin{aligned} &\text{Cor}\big(\mathsf{KE}_n(\boldsymbol{\varrho}^{(0)}), \mathsf{KE}_n(\boldsymbol{\varrho}^{(k\Delta t)})\big) \\ &\qquad = \sum_{j=1}^{n-1} \chi_{j,n} \text{Cor}\big(|\boldsymbol{\rho}_j^{(0)}|^2, |\boldsymbol{\rho}_j^{(k\Delta t)}|^2\big) \,, \end{aligned}$$

where

$$\chi_{j,n} = \frac{\kappa_{j,n}^2 \text{Var}\big(|\boldsymbol{\rho}_j^{(0)}|^2\big)}{\sum_{i=1}^{n-1} \kappa_{i,n}^2 \text{Var}\big(|\boldsymbol{\rho}_i^{(0)}|^2\big)} \tag{7.29}$$

and

$$\kappa_{j,n} = \begin{cases} \frac{m_n \omega_{j,n}^2}{2} & \text{for } \mathsf{KE}_n^{\text{pri}} \,, \\ \frac{\Lambda}{2n} & \text{for } \mathsf{KE}_n^{\text{vir}} \,. \end{cases} \tag{7.30}$$

If the evolution of the ring polymer is governed by the BAOAB-like update in Eq. 3.20, then the jth mode satisfies

$$\begin{aligned} \text{Cor}\big(|\boldsymbol{\rho}_j^{(0)}|^2, |\boldsymbol{\rho}_j^{(k\Delta t)}|^2\big) &= \frac{\text{Cov}\big(|\boldsymbol{\rho}_j^{(0)}|^2, |\boldsymbol{\rho}_j^{(k\Delta t)}|^2\big)}{\text{Var}\big(|\boldsymbol{\rho}_j^{(0)}|^2\big)} \\ &= (\mathscr{M}_{j,n}^k)_{11}^2 \,, \end{aligned} \tag{7.31}$$

where we used that the phase $\begin{bmatrix}\rho_j^{(k\Delta t)} & \varphi_j^{(k\Delta t)}\end{bmatrix}^{\mathrm{T}}$ follows a centered Gaussian distribution with covariance given in Eq. 3.33 for all $k \geq 0$. Therefore, in the infinite-friction limit where $\mathscr{M}_{j,n}$ is given in Appendix 7.3, the integrated autocorrelation time of KE_n evaluates to

$$\begin{aligned}\frac{\mathrm{aVar}(\mathsf{KE}_n)}{\mathrm{Var}(\mathsf{KE}_n)} &= 1 + 2\sum_{j=1}^{n-1} \chi_{j,n} \sum_{k=1}^{\infty} (\mathscr{M}_{j,n}^k)_{11}^2 \\ &\leq 1 + \frac{1}{2} \max_{1\leq j\leq n-1} \left| \frac{1+A(\omega_{j,n}\Delta t)}{1-A(\omega_{j,n}\Delta t)} \right| , \end{aligned} \tag{7.32}$$

where simplification of $(\mathscr{M}_{j,n}^k)_{11}$ was aided by the Cayley–Hamilton theorem for 2×2 matrices,[334] $A(x)$ is defined in Appendix 7.2, and in the last line we used that $\sum_{j=1}^{n-1} \chi_{j,n} = 1$. Eq. 7.32 states that the integrated autocorrelation time of KE_n can only be as small as that of the component $|\rho_j|^2$ exhibiting the slowest uncorrelation at stationarity.

Having derived Eq. 7.32, we now prove our claim for this section. Let $x := \omega_{j,n}\Delta t > 0$ and $\alpha := \Delta t^2 \Lambda/m \in (0,4)$. For fixed x and α, the function $A(x) := \cos(\theta(x)) - \frac{\alpha}{2x}\sin(\theta(x))$ monotonically decreases toward -1 as the angle $\theta(x)$ increases toward π. Consequently, the function $\left|\big(1+A(x)\big)/\big(1-A(x)\big)\right|$ decreases (toward 0) as $\theta(x)$ increases (toward π), but condition (**C3**) requires $\theta(x) \leq 2\arctan(x/2)$ to achieve stable evolution. Therefore, because it yields the largest stable angle, the choice $\theta(x) = 2\arctan(x/2)$ (i.e., the Cayley angle) minimizes the upper bound in Eq. 7.32.

A similar argument can be made to support the conjecture, suggested by Fig. 3.4f, that the non-centroid velocity estimator for the classical kinetic energy $\mathsf{KE}_n^{\mathrm{cla}}$ in Eq. 3.41, equivalently written

$$\mathsf{KE}_n^{\mathrm{cla}}(\varphi) = \frac{m_n}{2(n-1)} \sum_{j=1}^{n-1} \varphi_j^2 \tag{7.33}$$

with φ defined in Eq. 3.12, exhibits a *maximal* integrated autocorrelation time if the Cayley angle $\theta(x) = 2\arctan(x/2)$ is used. Indeed, the integrated autocorrelation time of this estimator is bounded by

$$\frac{\mathrm{aVar}(\mathsf{KE}_n^{\mathrm{cla}})}{\mathrm{Var}(\mathsf{KE}_n^{\mathrm{cla}})} \leq 1 + \frac{1}{2} \max_{1\leq j\leq n-1} \left| \frac{1-A(\omega_{j,n}\Delta t)}{1+A(\omega_{j,n}\Delta t)} \right| , \tag{7.34}$$

where the function $\left|\big(1-A(x)\big)/\big(1+A(x)\big)\right|$ monotonically increases as $\theta(x)$ approaches the largest stable (i.e., Cayley) angle for fixed x and α.

To conclude, we note that the conclusions of this section hold for arbitrary friction schedules despite our use of the infinite-friction limit in Eqs. 7.32 and 7.34.

7.6 Stability interval calibration for liquid water simulations

This section describes the computational procedure used to identify $\Delta t = 1.4$ fs as close to the upper bound of the stability interval of T-RPMD applied to q-TIP4P/F liquid water at 298 K and 0.998 g/cm^3. The procedure consisted of integrating an ensemble of 10^4 thermally initialized T-RPMD trajectories using the algorithm outlined in Section 3.2 in its single-bead realization (identical to velocity Verlet in classical MD[171]), and counting the fraction of trajectories that remained within an energy sublevel (i.e., did not exhibit detectable energy drift) throughout their duration for each tested time-step. A time-step was deemed stable if 99% or more of the ensemble remained in an energy sublevel throughout a 50-picosecond time period. A range of time-steps was tested, and the fraction of stable trajectories at each time-step is reported in Fig. 7.2.

To avoid initialization bias in the stability interval estimation, thermalized initial phase-points were generated with a Metropolized Markov-chain Monte Carlo sampler targeted at the equilibrium configurational distribution of the liquid. Specifically, a randomized Hamiltonian Monte Carlo[196, 335] (rHMC) simulation of sufficient length was used to thermalize a crystalline configuration of the system at the target density, and 10^2 configurations were extracted from well-separated points along the rHMC trajectory. Each of these approximately independent draws from the equilibrium *configurational* distribution of the liquid at the target physical conditions was subsequently paired with 10^2 independent velocities drawn from the corresponding Maxwell–Boltzmann distribution, yielding 10^4 approximately independent draws from the *phase space* distribution of the classical liquid at thermal equilibrium.

7.7 One-dimensional quantum harmonic oscillator

Numerical equilibrium averages and integrated autocorrelation times for the quantum harmonic oscillator were estimated by averaging over a 10-nanosecond T-RPMD trajectory integrated using the algorithm listed in Section 3.2, and initialized at an exact sample from the numerical stationary distribution (listed for the jth ring-polymer mode in Eq. 3.34) corresponding to the physical parameters (i.e., Λ, m, and β) and simulation parameters (i.e., n, Δt, and the function θ) listed in Section 3.3. Specifically, the statistics reported in Fig. 3.4 were obtained by partitioning

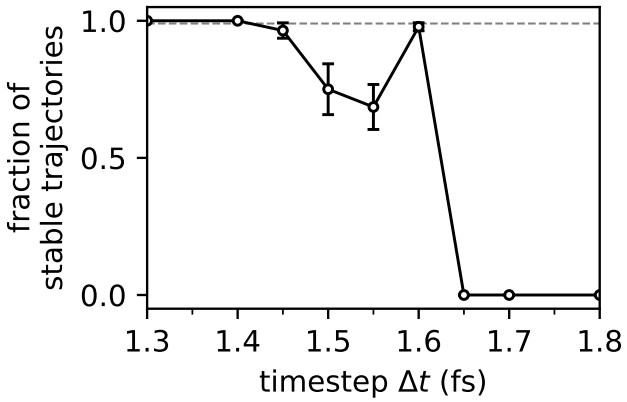


Figure 7.2: Stability interval calibration for q-TIP4P/F room-temperature liquid water simulations. Data points correspond to the fraction of thermally initialized single-bead T-RPMD trajectories that remained stable over a 50-picosecond period at the respective integration time-step Δt. Error bars correspond to the standard error of the fraction of stable trajectories across initialization points with different configurations. The gray dashed line marks the $\geq 99\%$ threshold for deeming a time-step *stable*, which no time-step beyond $\Delta t = 1.4$ fs reaches.

the T-RPMD trajectory into 10 disjoint blocks, estimating the equilibrium average and autocorrelation time within each block, and computing the sample mean and standard error among the resulting block estimates with 1000 bootstrap resamples.

We now describe the formulas and methods used to obtain block estimates for the equilibrium mean and integrated autocorrelation time. The equilibrium average μ_{O_n} of observable O_n within each block of the partitioned T-RPMD trajectory was estimated using the standard estimator[336]

$$\hat{\mu}_{\mathsf{O}_n} = \frac{1}{K}\sum_{k=0}^{K-1} \mathsf{O}_n^{(k\Delta t)} \,, \tag{7.35}$$

where K is the number of steps in the block (i.e., the block size) and $\mathsf{O}_n^{(k\Delta t)}$ the value of O_n at the kth step within the block. Similarly, the lag-$k\Delta t$ autocovariance $C_{\mathsf{O}_n}(k\Delta t)$ was estimated using[336]

$$\hat{C}_{\mathsf{O}_n}(k\Delta t) = \sum_{\ell=0}^{K-k-1} \frac{\left(\mathsf{O}_n^{(\ell\Delta t)} - \hat{\mu}_{\mathsf{O}_n}\right)\left(\mathsf{O}_n^{((\ell+k)\Delta t)} - \hat{\mu}_{\mathsf{O}_n}\right)}{K-k} \tag{7.36}$$

for $0 \leq k\Delta t \leq (K-1)\Delta t = 1$ ns. The integrated autocorrelation time was subsequently estimated using[218, 336]

$$\frac{\widehat{\mathrm{aVar}_{\mathsf{O}_n}}}{\mathrm{Var}_{\mathsf{O}_n}}(M) = 1 + 2\sum_{k=1}^{M} \frac{\hat{C}_{\mathsf{O}_n}(k\Delta t)}{\hat{C}_{\mathsf{O}_n}(0)} \,, \tag{7.37}$$

where $0 < M \leq K$ is a suitable cutoff. The choice of M is nontrivial, as it carries a trade-off between bias (more pronounced at small M) and variance (more pronounced at large M).[218] To choose M judiciously, we follow the *automatic windowing* (AW) method described in Appendix C of Ref. [337]. The AW method dictates that M should correspond to the smallest lag that satisfies the inequality

$$M \geq c \frac{\widehat{\mathrm{aVar}_{\mathsf{O}_n}}}{\mathrm{Var}_{\mathsf{O}_n}}(M) \,, \tag{7.38}$$

where the parameter $c > 0$ dictates the variance-bias trade-off in place of M, and is chosen as large as possible to reduce the bias of the estimator for a given variance threshold.

Fig. 7.3 illustrates usage of the AW method for integrated autocorrelation time estimation, using trajectory data generated by the T-RPMD scheme with $\theta(x) = 2\arctan(x/2)$ at $n = 64$ beads and $\Delta t = 2.0$ fs, and focusing on the observables $\mathsf{KE}_n^{\mathrm{pri}}$ (black), $\mathsf{KE}_n^{\mathrm{vir}}$ (red), and $\mathsf{KE}_n^{\mathrm{cla}}$ (cyan) introduced in Section 3.3. The estimated integrated autocorrelation times are plotted with solid lines in Fig. 7.3a for various values of c, and the corresponding cutoffs M are plotted in Fig. 7.3b. Exact integrated autocorrelation times are plotted with dashed lines in Fig. 7.3a. Note that as c (and thus M) increases, the estimates converge to the corresponding exact values at the expense of a larger variance, which can nonetheless be controlled by adjusting the block size K.

7.8 Room-temperature liquid water

The equilibrium averages and integrated autocorrelation times reported in Fig. 3.5 were obtained by averaging over 10-nanosecond T-RPMD trajectories integrated for each considered bead number n, time-step Δt, and function θ. All trajectories were initialized at an approximate sample from the corresponding numerical equilibrium distribution, obtained by thermalizing for 20 picoseconds a classical (i.e., $n = 1$) configuration of the system into the n-bead ring-polymer phase space. The reference equilibrium averages plotted with dashed lines in Fig. 3.5 were obtained by averaging over a one-nanosecond, 256-bead staging PIMD[190] trajectory integrated at a 0.1-fs time-step with the mass and friction parameters recommended in Ref. [192], and initialized with the same protocol used for the T-RPMD simulations.

The observables considered in Fig. 3.5 measure properties per H atom or per H_2O molecule, and thus the reported values are averages over estimates obtained for each simulated moiety. The equilibrium mean and integrated autocorrelation time

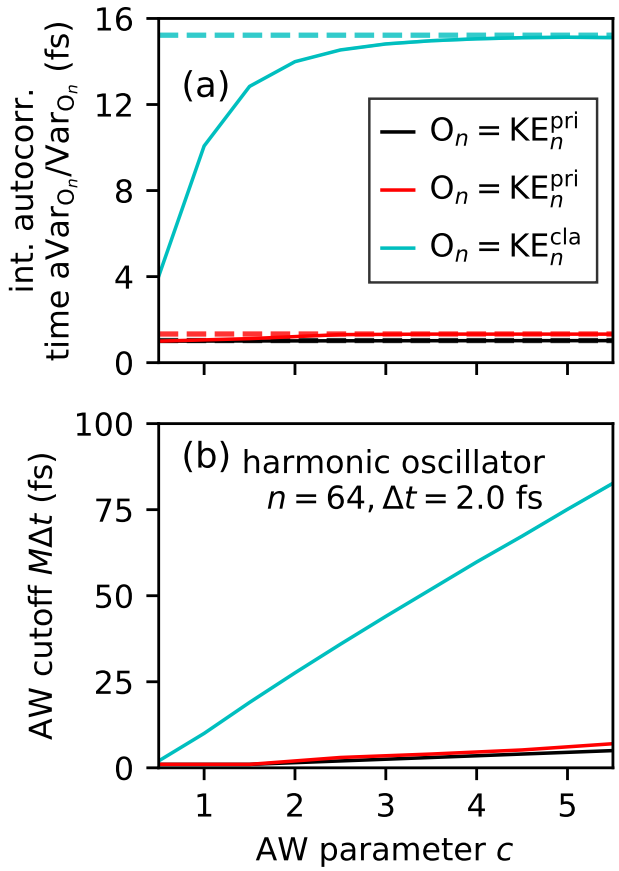


Figure 7.3: Integrated autocorrelation times of several observables of the one-dimensional harmonic oscillator in Section 3.3, estimated with the AW method. Trajectory data for the estimates was generated using the T-RPMD scheme with $\theta(x) = 2\arctan(x/2)$ at $n = 64$ beads and $\Delta t = 2.0$ fs, and processed as described in the current section. Estimated (resp. exact) integrated autocorrelation times for observables $\mathsf{KE}_n^{\mathrm{pri}}$ (black), $\mathsf{KE}_n^{\mathrm{vir}}$ (red), and $\mathsf{KE}_n^{\mathrm{cla}}$ (cyan) are shown in solid (resp. dashed) lines in panel (a) as a function of the windowing parameter c. Panel (b) plots the cutoffs determined by the choice of c for the three observables, where the linear relation between $M\Delta t$ and c at large values of the latter corroborates the non-spurious convergence of the autocorrelation time estimates.

of observable O_n for each moiety was estimated by partitioning the trajectory of the moiety into 10 disjoint 1-nanosecond blocks, evaluating Eqs. 7.35 and 7.37 within each block, and determining the sample mean and standard error among the block estimates with 1000 bootstrap resamples. The AW method[337] was applied to choose a cutoff lag $M \leq 1$ ns in Eq. 7.37, as illustrated in Fig. 7.3 for the harmonic oscillator application.

The T-RPMD trajectories used to generate Fig. 3.5 also yielded Fig. 3.6, where panels (a) and (c) plot autocovariance functions of the form $\frac{1}{N_{\mathrm{H_2O}}}\sum_{i=1}^{N_{\mathrm{H_2O}}}\mathbb{E}\big(\bar{\mathsf{O}}_i(0)\cdot\bar{\mathsf{O}}_i(k\Delta t)\big)$, where $N_{\mathrm{H_2O}}=32$ is the number of simulated H_2O molecules and $\bar{\mathsf{O}}_i(k\Delta t)$ is the bead-averaged value of observable O (e.g., the molecular dipole moment or center-of-mass velocity) on the ith molecule at time $k\Delta t$ along a stationary T-RPMD trajectory. The autocovariance $\mathbb{E}\big(\bar{\mathsf{O}}_i(0)\cdot\bar{\mathsf{O}}_i(t)\big)$ was estimated for the lags $k\Delta t$ shown in Fig. 3.6 by

$$\mathbb{E}\big(\bar{\mathsf{O}}_i(0)\cdot\bar{\mathsf{O}}_i(k\Delta t)\big) \approx \sum_{\ell=0}^{K-k-1}\frac{\bar{\mathsf{O}}_i^{(\ell\Delta t)}\cdot\bar{\mathsf{O}}_i^{((\ell+k)\Delta t)}}{K-k}\,, \tag{7.39}$$

where $K\Delta t = 1$ ns is the length of each block in a partitioned 10-nanosecond T-RPMD trajectory. As with the results in Fig. 3.5, statistics for each molecule were obtained from block estimates via bootstrapping, and Figs. 3.6a and 3.6c report molecule-averaged statistics.

Fig. 7.4 validates the 20-picosecond thermalization interval used to initialize the trajectories that generated Figs. 3.5 and 3.6. In detail, Figs. 7.4a and 7.4b (resp., Figs. 7.4c and 7.4d) plot the non-equilibrium mean of the primitive and virial quantum kinetic energy per H atom (resp. the mean $\mathrm{O}-\mathrm{H}$ bond and $\mathrm{H}-\mathrm{O}-\mathrm{H}$ angle potential energy per water molecule) as it approaches the equilibrium value in Figs. 3.5a and 3.5c (resp., Figs. 3.5e and 3.5g) for a 64-bead ring polymer at a 1.4 fs time-step with the considered choices of θ. At each time $k\Delta t$ within the 20-picosecond interval, the non-equilibrium mean is estimated by averaging across 1000 independent trajectories initialized at a point-mass distribution on the n-bead ring-polymer phase space centered at the classical (i.e., $n=1$) sample used to initialize the reported simulations. Within statistical uncertainty, the non-equilibrium mean for each observable converges to its equilibrium value within the 20-picosecond interval at visually indistinguishable rates across the tested choices of θ.

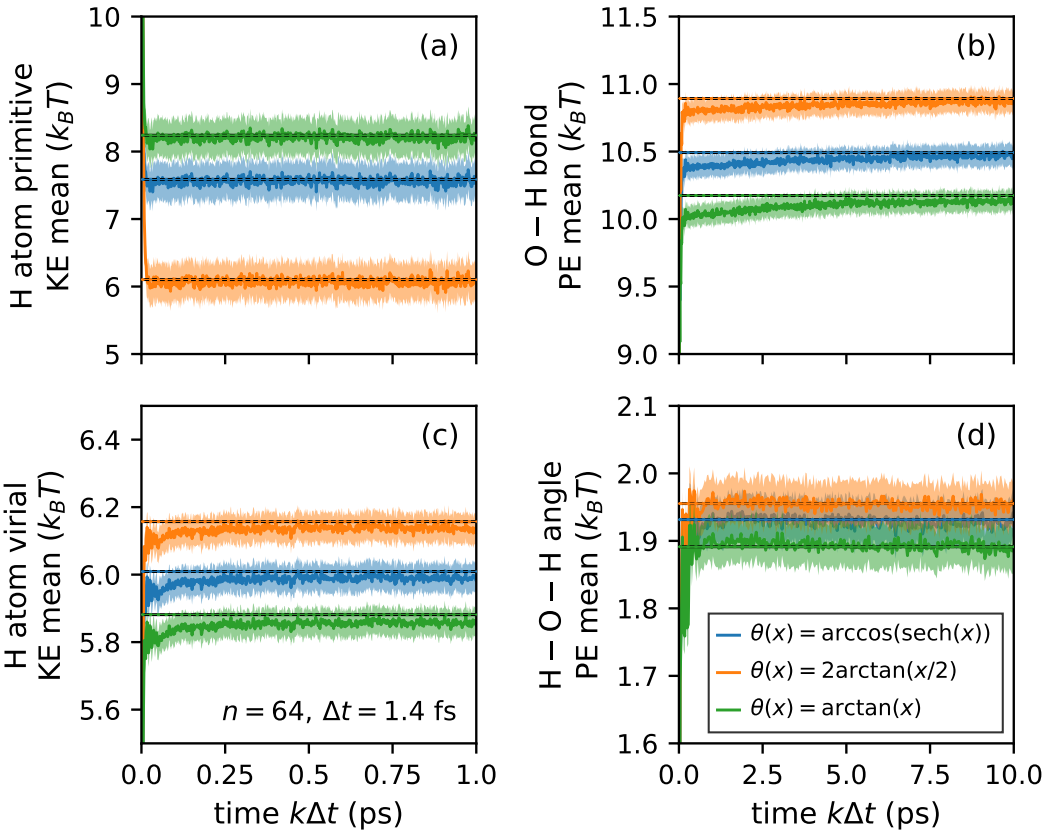


Figure 7.4: Convergence to equilibrium of the BAOAB-like schemes considered in Section 3.3 with $n = 64$ ring-polymer beads and a $\Delta t = 1.4$ fs time-step. With respect to the non-equilibrium 64-bead configurational distribution evolved from a point mass at a classical (i.e., $n = 1$) configuration, panels (a) and (c) plot the mean kinetic energy per H atom for the n-bead system as per the primitive and virial estimators, respectively, for times up to 1.0 ps. Panels (b) and (d), respectively, plot the non-equilibrium mean O−H-bond and H−O−H-angle potential energy per q-TIP4P/F water molecule,[216] for times up to 10 ps. The lightly shaded interval around each curve corresponds to the standard error of the estimated non-equilibrium mean, computed with 1,000 bootstrap resamples from a sample of 1,000 independent trajectories.

Chapter 8

APPENDIX OF CHAPTER 4

8.1 Comparison of band structure and e-ph matrix elements

Here, we compare the band structure and e-ph matrix elements in Γ valley for the ab-initio calculation and the semi-analytical model. In Fig. 8.1a, we show the Γ valley band structure below 360 meV obtained in the ab-initio DFT calculation and the Kane model [258] used in the semi-analytical model. The Kane model uses an experimental effective mass of $0.067m_e$ and a non-parabolicity of 0.64 eV^{-1}. The figure shows that the anisotropy of the DFT band structure is negligible below 200 meV, and the maximum anisotropy remains less than 7% below 360 meV. The Kane model uses an isotropic experimental effective mass and non-parabolicity, which results in a slightly different band structure, especially for small k. The difference between DFT and Kane model band structure at low energy is because the Kane model uses the experimental effective mass $0.067m_e$ instead of the ab-initio effective mass $0.056m_e$.

In Fig. 8.1b, we show the density of states (DOS) obtained from the ab-initio calculation and the semi-analytical model. The two DOS agree in overall trend, but the ab-initio DOS has large variations in energy due to the finite sampling of the Brillouin zone (fine grid density of $200 \times 200 \times 200$). Comparing Figs. 8.1b and 4.1a, the variations of the DOS in the ab-initio calculation can account for the variations in scattering rates.

In Fig. 8.1c, we plot the absolute values of ab-initio e-ph scattering matrix elements $|g_{\text{LO}}(\boldsymbol{k},\boldsymbol{q})|$ for LO phonons involved in the on-shell Γ-Γ scattering versus the phonon wavevector norm q in the range of $q < 0.2G$. We see good agreement for $q < 0.05G$ and slightly larger variations for large $q > 0.05G$. In order to quantify the errors, we compute the relative error by

$$\Delta_{\text{err}} = \frac{\sum_{(\boldsymbol{k},\boldsymbol{q})\in\text{on-shell}} \left| |g_{\text{LO}}(\boldsymbol{k},\boldsymbol{q})| - |g^{\text{fit}}_{\text{LO}}(q)| \right|}{\sum_{(\boldsymbol{k},\boldsymbol{q})\in\text{on-shell}} |g_{\text{LO}}(\boldsymbol{k},\boldsymbol{q})|}. \tag{8.1}$$

We finally find $\Delta_{\text{err}} < 3\%$.

8.2 Derivation of simultaneous electron-two-phonon scattering rates

We provide a derivation for estimation of the simultaneous e-2ph scattering rates given in Sec. 4.5. Consider a crystal in which each primitive unit cell has atoms with charge Z_κ at position $\boldsymbol{\tau}_\kappa$, $\boldsymbol{R}$ is the lattice vector, and $\boldsymbol{G}$ is the reciprocal lattice constant. The lattice displacement $u_{\boldsymbol{R}\kappa}$ is decomposed using normal modes:

$$u_{\boldsymbol{R}\kappa} = \sum_{\boldsymbol{q}\nu} \boldsymbol{u}_{\boldsymbol{q}\kappa\nu} e^{i\boldsymbol{q}\cdot\boldsymbol{R}}, \tag{8.2}$$

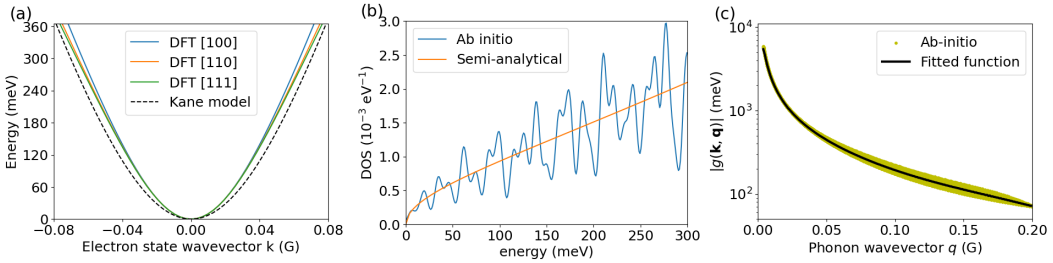


Figure 8.1: (a): Γ valley band structure in the ab-initio DFT calculation and the semi-analytical Kane model below 360 meV. For the DFT band structure the [100], [110], [111] directions are plotted. The DFT band structure exhibits only slight anisotropy. (b): Electronic DOS versus energy in the ab-initio calculation and the semi-analytical model. The ab-initio DOS has large variations due to the finite grid density. (c): Absolute values of on-shell e-ph scattering matrix elements $|g(\boldsymbol{k},\boldsymbol{q})|$ for LO phonons versus q in the range of $q < 0.2G$ of the Γ-Γ scattering process in the ab-initio calculation along with the fitted relation (Eq. 4.18) used in the semi-analytical model.

where

$$\boldsymbol{u}_{\boldsymbol{q}\kappa\nu} = \sqrt{\frac{\hbar}{2NM_\kappa\omega_{\boldsymbol{q}\nu}}}\boldsymbol{e}_{\boldsymbol{q}\kappa\nu}(b^\dagger_{\boldsymbol{q}\nu} + b_{-\boldsymbol{q}\nu}), \tag{8.3}$$

N is the number of unit cells in a supercell, M_κ is the mass of atom κ, ν is the phonon mode, $\boldsymbol{e}_{\boldsymbol{q}\kappa\nu}$ is the phonon polarization unit vector, $b^\dagger_{\boldsymbol{q}\nu}$, and $b_{\boldsymbol{q}\nu}$ are creation and annihilation operators of phonon $\boldsymbol{q}\nu$, respectively.

The Coulomb potential energy of an electron generated by point charge Z at position $\boldsymbol{R}_\kappa = \boldsymbol{R} + \boldsymbol{\tau}_\kappa$ is

$$V(\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\tau}_\kappa) = -\frac{Z_\kappa e^2}{4\pi\varepsilon_\infty|\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\tau}_\kappa|}. \tag{8.4}$$

It will be convenient to rewrite $V(\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\tau}_\kappa)$ in reciprocal space by Fourier transformation:

$$V(\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\tau}_\kappa) = -\frac{1}{N\Omega}Z_\kappa\sum_{\boldsymbol{q}}\sum_{\boldsymbol{G}}V(\boldsymbol{q}+\boldsymbol{G})e^{i(\boldsymbol{q}+\boldsymbol{G})\cdot(\boldsymbol{r}-\boldsymbol{R}-\boldsymbol{\tau}_\kappa)}, \tag{8.5}$$

where Ω is the primitive unit cell volume, and

$$V(\boldsymbol{q}+\boldsymbol{G}) = \frac{e^2}{\varepsilon_\infty(\boldsymbol{q}+\boldsymbol{G})^2}. \tag{8.6}$$

Following Ref. [268], the electron-two-phonon Hamiltonian is

$$H(\boldsymbol{r}) = \frac{1}{2}\sum_{\boldsymbol{R}\kappa}\boldsymbol{u}_{\boldsymbol{R}\kappa}\cdot\nabla\nabla V(\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\tau}_\kappa)\cdot\boldsymbol{u}_{\boldsymbol{R}\kappa}. \tag{8.7}$$

From Eq. 8.5, we can obtain

$$\nabla\nabla V(\boldsymbol{r}-\boldsymbol{R}-\boldsymbol{\tau}_\kappa) = Z_\kappa(\boldsymbol{k}+\boldsymbol{G})(\boldsymbol{k}+\boldsymbol{G})\sum_{\boldsymbol{k}}\sum_{\boldsymbol{G}} V(\boldsymbol{k}+\boldsymbol{G})e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot(\boldsymbol{r}-\boldsymbol{R}-\boldsymbol{\tau}_\kappa)}, \quad (8.8)$$

where the product of the vectors in this and following equations is defined as the outer product. By using Eq. 8.2 and Eq. 8.8, we have

$$\begin{aligned}
H(\boldsymbol{r}) &= \frac{1}{2\Omega}\sum_{\boldsymbol{R}\kappa} Z_\kappa \boldsymbol{u}_{\boldsymbol{R}\kappa}\cdot\nabla\nabla V(\boldsymbol{r}-\boldsymbol{R}-\boldsymbol{\tau}_\kappa)\cdot\boldsymbol{u}_{\boldsymbol{R}\kappa} \\
&= \frac{1}{N\Omega}\sum_{\boldsymbol{R}\kappa}\sum_{\boldsymbol{q}\nu<\boldsymbol{q}'\nu'} Z_\kappa \sum_{\boldsymbol{k}}\sum_{\boldsymbol{G}} \boldsymbol{u}_{\boldsymbol{q}\kappa\nu}\cdot(\boldsymbol{k}+\boldsymbol{G})\boldsymbol{u}_{\boldsymbol{q}'\kappa\nu'}\cdot(\boldsymbol{k}+\boldsymbol{G})V(\boldsymbol{k}+\boldsymbol{G})e^{i(\boldsymbol{q}+\boldsymbol{q}')\cdot\boldsymbol{R}}e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot(\boldsymbol{r}-\boldsymbol{R}-\boldsymbol{\tau}_\kappa)} \\
&= \frac{1}{N\Omega}\sum_{\boldsymbol{G}\kappa}\sum_{\boldsymbol{q}\nu<\boldsymbol{q}'\nu'} Z_\kappa \boldsymbol{u}_{\boldsymbol{q}\kappa\nu}\cdot(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})\boldsymbol{u}_{\boldsymbol{q}'\kappa\nu'}\cdot(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})V(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})e^{i(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})\cdot(\boldsymbol{r}-\boldsymbol{\tau}_\kappa)} \\
&= \frac{e^2}{\Omega\varepsilon_\infty}\sum_{\boldsymbol{G}\kappa}\sum_{\boldsymbol{q}\nu<\boldsymbol{q}'\nu'} Z_\kappa \boldsymbol{u}_{\boldsymbol{q}\kappa\nu}\cdot\frac{(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})}{|(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})|^2}\cdot\boldsymbol{u}_{\boldsymbol{q}'\kappa\nu'}e^{i(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})\cdot(\boldsymbol{r}-\boldsymbol{\tau}_\kappa)} \\
&\sim \frac{e^2}{\Omega\varepsilon_\infty}\sum_{\boldsymbol{G}\kappa}\sum_{\boldsymbol{q}\nu<\boldsymbol{q}'\nu'} Z_\kappa u_{\boldsymbol{q}\kappa\nu}u_{\boldsymbol{q}'\kappa\nu'}e^{i(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})\cdot(\boldsymbol{r}-\boldsymbol{\tau}_\kappa)}(b^\dagger_{\boldsymbol{q}\nu}+b_{-\boldsymbol{q}\nu})(b^\dagger_{\boldsymbol{q}'\nu'}+b_{-\boldsymbol{q}'\nu'}) \\
&\sim \frac{e^2}{\Omega\varepsilon_\infty}\sum_{\boldsymbol{q}\nu<\boldsymbol{q}'\nu'}\sum_{\kappa} Z_\kappa u_{\boldsymbol{q}\kappa\nu}u_{\boldsymbol{q}'\kappa\nu'}e^{i(\boldsymbol{q}+\boldsymbol{q}')\cdot(\boldsymbol{r}-\boldsymbol{\tau}_\kappa)}(b^\dagger_{\boldsymbol{q}\nu}+b_{-\boldsymbol{q}\nu})(b^\dagger_{\boldsymbol{q}'\nu'}+b_{-\boldsymbol{q}'\nu'}),
\end{aligned} \quad (8.9)$$

where $u_{\boldsymbol{q}\kappa\nu} = \sqrt{\hbar/2M_\kappa\omega_q}$ is the amplitude of the phonon displacement,

$$\sum_{\boldsymbol{R}} e^{i(\boldsymbol{q}+\boldsymbol{q}'-\boldsymbol{k}-\boldsymbol{G})\cdot\boldsymbol{R}} = N\delta_{\boldsymbol{k},\boldsymbol{q}+\boldsymbol{q}'} \quad (8.10)$$

was used in the third line,

$$\frac{(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})}{|(\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{G})|^2} \sim 1 \quad (8.11)$$

was used in the fourth line, and the contributions from $\boldsymbol{G}\neq 0$ are neglected from $\sum_{\boldsymbol{G}}$ in the last line.

We observe that the quadrupole moment $\sum_\kappa Z_\kappa \boldsymbol{u}_{\boldsymbol{q}\kappa\nu}\boldsymbol{u}_{\boldsymbol{q}'\kappa\nu'}$ appears in the third line of Eq. 8.9. For a polar material like GaAs that each unit cell has two atoms with opposite charge and similar mass, we can perform a $\mathbb{Z}_2$ symmetry analysis for the quadrupole moment. Since Z_κ has odd symmetry for the two atoms in a unit cell, the phonon modes must be one of odd symmetry (optical mode) and one of even symmetry (acoustic mode) to avoid cancellation.

The electron-phonon scattering matrix elements can be written as

$$g^{(2)}_{\nu\nu'}(\boldsymbol{q},\boldsymbol{p}) \sim \frac{Z_\kappa e^2}{\Omega\varepsilon_\infty}u_{\boldsymbol{q}\kappa\nu}u_{\boldsymbol{p}\kappa\nu'}, \quad (8.12)$$

where the phase factor is neglected.

Similar to Eq. 4.11 and Eq. 4.13, we can derive the e-2ph collision matrix element for a specific phonon mode and subprocess type as

$$\Theta_{\boldsymbol{k}',\boldsymbol{k}} \sim \frac{2\pi}{\hbar}\frac{1}{\Omega_{\mathrm{BZ}}^2}\delta(\varepsilon_{k'}-\varepsilon_k-\Delta E)\int_{\boldsymbol{q}+\boldsymbol{p}=\boldsymbol{k}'-\boldsymbol{k}} N_q N_p \left|g^{(2)}(\boldsymbol{q},\boldsymbol{p})\right|^2 d^3\boldsymbol{p}. \tag{8.13}$$

In the following, we assume that phonon $\boldsymbol{q}$ is optical and phonon $\boldsymbol{p}$ is acoustic. We additionally assume that optical phonons have no dispersion and acoustic phonons have linear dispersion with velocity s such that

$$\begin{aligned} \omega_q &= \omega_{\mathrm{O}} \\ \omega_p &\sim sp \sim \frac{p}{k_{\max}}\omega_{\mathrm{A}}, \end{aligned} \tag{8.14}$$

where ω_{O} and ω_{A} are phonon frequencies at the edge of the Brillouin zone for optical and acoustic phonons, respectively. Since the frequencies of transverse and longitudinal phonons at the edge of the Brillouin zone are of the same magnitude here we neglect their difference. Using the phonon dispersion relation assumed above, we then have

$$\begin{aligned} N_q &\sim N(\omega_{\mathrm{O}}) \\ N_p &\sim N(\omega_{\mathrm{A}})\frac{N(\omega_p)}{N(\omega_A)} \sim N(\omega_{\mathrm{A}})\frac{\omega_A}{\omega_p} \sim N(\omega_{\mathrm{A}})\frac{k_{\max}}{p}, \end{aligned} \tag{8.15}$$

where in the second line we assume the temperature is not too low in the sense of $\beta\hbar\omega \lesssim 1$ so that $N(\omega) = \frac{1}{e^{\beta\hbar\omega}-1} \sim \frac{1}{\beta\hbar\omega}$, and $k_{\max}$ is the wave vector the edge of the Brillouin zone.

Similarly, we have

$$g^{(2)}(\boldsymbol{q},\boldsymbol{p}) \sim g^{(2)}(k_{\max},k_{\max})\sqrt{\frac{k_{\max}}{p}}, \tag{8.16}$$

where

$$g^{(2)}(k_{\max},k_{\max}) = \frac{eZ}{\Omega\varepsilon}\sqrt{\frac{\hbar}{2M\omega_{\mathrm{A}}}}\sqrt{\frac{\hbar}{2M\omega_{\mathrm{O}}}}. \tag{8.17}$$

We can then calculate $\Theta_{\boldsymbol{k}',\boldsymbol{k}}$ from Eq. 8.13 as

$$\begin{aligned} \Theta_{\boldsymbol{k}',\boldsymbol{k}} &\sim \frac{2\pi}{\hbar}\frac{1}{\Omega_{\mathrm{BZ}}^2}\left|g^{(2)}(k_{\max},k_{\max})\right|^2\delta(\varepsilon_{k'}-\varepsilon_k-\Delta E)A_{\alpha_1}(\omega_{\mathrm{O}})A_{\alpha_2}(\omega_{\mathrm{A}})\int\left(\frac{k_{\max}}{p}\right)^2 d^3\boldsymbol{p} \\ &\sim \frac{2\pi}{\hbar}\frac{1}{\Omega_{\mathrm{BZ}}^2}\left|g^{(2)}(k_{\max},k_{\max})\right|^2\delta(\varepsilon_{k'}-\varepsilon_k-\Delta E)A_{\alpha_1}(\omega_{\mathrm{O}})A_{\alpha_2}(\omega_{\mathrm{A}})\Omega_{\mathrm{BZ}} \\ &= \frac{2\pi}{\hbar}\frac{1}{\Omega_{\mathrm{BZ}}}\left|g^{(2)}(k_{\max},k_{\max})\right|^2\delta(\varepsilon_{k'}-\varepsilon_k-\Delta E)A_{\alpha_1}(\omega_{\mathrm{O}})A_{\alpha_2}(\omega_{\mathrm{A}}), \end{aligned} \tag{8.18}$$

where Ω_{BZ} is the Brillouin zone volume, $\alpha_{1,2}$ indicates whether a phonon is absorbed or emitted, and

$$A_\alpha(\omega) = N(\omega) + \delta_{\alpha,+1}. \tag{8.19}$$

The scattering rate can be calculated by integrating $\Theta_{\boldsymbol{k}',\boldsymbol{k}}$ over $\boldsymbol{k}'$:

$$\begin{aligned}
\Gamma_k &= \int \Theta_{\boldsymbol{k}',\boldsymbol{k}} d^3\boldsymbol{k}' \\
&\sim \int \frac{2\pi}{\hbar}\frac{1}{\Omega_{\mathrm{BZ}}}\left|g^{(2)}(k_{\max},k_{\max})\right|^2 \delta(\varepsilon_{k'}-\varepsilon_k-\Delta E) A_{\alpha_1}(\omega_{\mathrm{O}}) A_{\alpha_2}(\omega_{\mathrm{A}}) 4\pi k'^2 dk' \\
&= \frac{8\pi^2}{\hbar\Omega_{\mathrm{BZ}}}\left|g^{(2)}(k_{\max},k_{\max})\right|^2 A_{\alpha_1}(\omega_{\mathrm{O}}) A_{\alpha_2}(\omega_{\mathrm{A}}) k'^2 \frac{dk'}{d\varepsilon_{k'}} \\
&\sim \frac{8\pi^2}{\hbar\Omega_{\mathrm{BZ}}}\left|g^{(2)}(k_{\max},k_{\max})\right|^2 A_{\alpha_1}(\omega_{\mathrm{O}}) A_{\alpha_2}(\omega_{\mathrm{A}}) \frac{k^3}{\varepsilon_k} \\
&\sim \frac{8\pi^2}{\hbar}\left(\frac{k}{k_{\max}}\right)^3 \left|g^{(2)}(k_{\max},k_{\max})\right|^2 A_{\alpha_1}(\omega_{\mathrm{O}}) A_{\alpha_2}(\omega_{\mathrm{A}}) \frac{1}{\varepsilon_k},
\end{aligned} \tag{8.20}$$

where $k' \sim k$ is assumed since the phonon energy (≤ 35 meV) is low compared with the energy range we are considering (~ 200 meV).

Considering GaAs at temperature ~ 300 K and an energy of about 200 meV (corresponding to $\frac{k}{k_{\max}} \sim 0.05$), we have:

$$\begin{aligned}
A_{\alpha_1}(\omega_{\mathrm{O}}) \sim A_{\alpha_2}(\omega_{\mathrm{A}}) &\sim 1 \\
\left(\frac{k}{k_{\max}}\right)^3 &\sim 0.05^3 \sim 10^{-4} \\
\varepsilon_k &\sim 0.2\mathrm{eV} \sim 10^{-2}\mathrm{Ry} \\
\left|g^{(2)}(k_{\max},k_{\max})\right|^2 &\sim 10^{-4}\mathrm{Ry}.
\end{aligned} \tag{8.21}$$

Thus

$$\Gamma_k \sim \frac{8\pi^2}{\hbar}\left(\frac{k}{k_{\max}}\right)^3 \left|g^{(2)}(k_{\max},k_{\max})\right|^2 A_{\alpha_1}(\omega_{\mathrm{O}}) A_{\alpha_2}(\omega_{\mathrm{A}}) \frac{1}{\varepsilon_k} \sim 10^{-8}\mathrm{Ry} \sim 10^{-4}\mathrm{ps}^{-1} \tag{8.22}$$

Considering 3×3 phonon polarizations and 2×2 subprocess types, the total scattering rate is about $\Gamma_k^{(\mathrm{total})} \sim 36\Gamma_k \sim 10^{-2.5}\mathrm{ps}^{-1}$, which is about 3.5 magnitudes lower than the 2ph rates studied in this work.

Chapter 9

APPENDIX OF CHAPTER 5

9.1 Compression of Clifford circuit with magic initial state

Here we describe the Clifford-based compression algorithm we use to reduce the required number of physical qubits by a factor of two, as well as to reduce the total number of mid-circuit measurements to equal the number of physical qubits. The compression is based on Ref. [324] with an improvement that removes the requirement for dynamic circuits (adaptivity), instead using an efficient classical simulation and classical coin flipping. Here, we first summarize the compression algorithm stated in Ref. [324], and then explain how to remove the adaptivity.

In a particular circuit realization the unitaries and the measurements can be written as

$$C_{\mathbf{m}} = \ldots U_3 M_{m_2} U_2 M_{m_1} U_1. \tag{9.1}$$

Here m_j is the j-th measurement outcome of the entire record, and correspondingly $M_{m_j} = (1+(-1)^{m_j} P_j)/2$ is the j-th projection operator, with P_j the Pauli operator being measured. Moving all unitaries past the measurements to the right, we can equivalently write

$$C_{\mathbf{m}} = \ldots \widetilde{M}_{m_2} \widetilde{M}_{m_1}, \tag{9.2}$$

where

$$\widetilde{M}_{m_j} = \frac{1}{2}(1+z_j \widetilde{P}_j), \quad \widetilde{P}_j = U_1^\dagger U_2^\dagger \ldots U_j^\dagger P_j U_j U_{j-1} \ldots U_1 \tag{9.3}$$

are now multi-site Pauli measurements and $z_j = (-1)^{m_j}$.

Let $A = \{1,\ldots,k\}$, and $B = \{k+1,\ldots,N\}$. Following Ref. [324], we state without proof that the following algorithm correctly samples an output bitstring of the circuit C on a input state in the new basis, with input states of the form $|\psi\rangle = |\phi_A\rangle \otimes |0_B^{\otimes N-k}\rangle$.

1. Initialize the quantum state $|\phi_A\rangle$, define the initial stabilizer group $\mathscr{S} = \langle Z_{k+1}, \ldots Z_N \rangle$, and let the Pauli operators be $\{\widetilde{P}_j\}$.

2. Consider each $\widetilde{P}_j$ in increasing order of j. For each j there are three possible cases:

   a) $\widetilde{P}_j \in \mathscr{S}$. In this case the measurement result is deterministic, and can be classically computed and we do not need to update the state or $\mathscr{S}$.

b) $\widetilde{P}_j \notin \mathscr{S}$, and it anticommutes with at least one element $Q \in \mathscr{S}$. In this case, the measurement result of $\widetilde{P}_j$ is equally likely $z_j = \pm 1$. We can flip a classical coin to sample z_j. Further, we need to account for the change in the state, which can be shown to be

$$|\phi\rangle \to V_j(z_j)\,|\psi\rangle\,, \tag{9.4}$$

where $V_j(z_j)$ is a Clifford unitary operator

$$V_j(z_j) = \frac{1}{\sqrt{2}}(Q + z_j \widetilde{P}_j). \tag{9.5}$$

Instead of evolving the state and updating $\mathscr{S}$, we adopt the Heisenberg picture and modify all subsequent measurements $\widetilde{P}_{k>j}$ as follows:

$$\widetilde{P}_k \to V_j(z_j)^\dagger \widetilde{P}_k V_j(z_j), \quad \forall k > j. \tag{9.6}$$

c) $\widetilde{P}_j \notin \mathscr{S}$, and it commutes with all elements of $\mathscr{S}$. It then necessarily commutes with $Z_{k+1}, \ldots, Z_N$ since these stabilizers are permanent, as we can check at the end of the algorithm (see comment 2 below). It follows that $\widetilde{P}_j$ only contains the identity operator or the Pauli Z operator on B. We can then consider a truncated Pauli operator that is supported only on A,

$$\widetilde{P}_j^A \coloneqq \eta_j \cdot \widetilde{P}_j|_A, \tag{9.7}$$

where $\widetilde{P}_j|_A$ is the restriction of $\widetilde{P}_j$ on A, and the sign $\eta_j = \pm 1$ can be chosen such that for any state $|\phi_A\rangle$ we have

$$\langle \phi_A | \widetilde{P}_j^A | \phi_A \rangle = \langle \phi_A \otimes 0_B^{\otimes N-k} | \widetilde{P}_j | \phi_A \otimes 0_B^{\otimes N-k} \rangle\,. \tag{9.8}$$

The measurement of $\widetilde{P}_j$ on the joint system AB can therefore be faithfully simulated by a measurement of $\widetilde{P}_j^A$ on just A. We perform this measurement on the state $|\phi_A\rangle$, update the state accordingly and record the measurement result z_j'. We then update the stabilizer group as

$$\mathscr{S} \to \langle \mathscr{S}, z_j' \widetilde{P}_j^A \rangle. \tag{9.9}$$

We see that in this algorithm

1. Cases (1) and (2) can be accounted for by classical simulation, and only in case (3) a quantum operation on $|\phi_A\rangle$ needs to be performed.

2. The stabilizer group $\mathscr{S}$ gets augmented only in case (3), and can be augmented at most k times. Once an operator is added into $\mathscr{S}$, it will remain in $\mathscr{S}$ until the algorithm terminates.

In this way, a given sequence of multi-site measurements can be simulated by a "compressed circuit" with at most k multi-site measurements on A, as well as classical coin flips, up to a polynomial time overhead.

A technical problem of the above algorithm is that the update of the stabilizer group $\mathscr{S}$ in case (c) depends on the quantum measurement result z'_j. Not knowing z'_j before the circuit execution will lead to the lack of knowledge of the sign of $Q \in \mathscr{S}$ in case (b) if occuring after the update of $\mathscr{S}$ due to case (c). Here we show the adaptivity can be removed by proving that the effect of flipping signs of z'_j or Q can be captured by classical postprocessing.

In order to prove it, we first notice that $Q \to -Q$ is equivalent to $z_j \to -z_j$ in Eq. 9.5 ($V \to -V$ has no effect on Eq. 9.6). We additionally notice that

$$V_j(-z_j) = QV_j(z_j)Q, \tag{9.10}$$

so that for any $k > j$,

$$\begin{aligned} V_j(-z_j)^\dagger \widetilde{P}_k V_j(-z_j) &= QV_j(z_j)^\dagger Q\widetilde{P}_k QV_j(z_j)Q \\ &= \lambda_{Q,\widetilde{P}_k} QV_j(z_j)^\dagger \widetilde{P}_k V_j(z_j)Q \\ &= \lambda_{Q,\widetilde{P}_k}\lambda_{Q,V_j(z_j)^\dagger \widetilde{P}_k V_j(z_j)} V_j(z_j)^\dagger \widetilde{P}_k V_j(z_j), \end{aligned} \tag{9.11}$$

where we have defined the commutator of Pauli operators A, B

$$AB = \lambda_{A,B} BA. \tag{9.12}$$

Eq. (9.11) implies that flipping measurement results z'_j at most result in sign changes of the subsequent measurements operators $\widetilde{P}_{k>j}$, and such sign dependence can be classically captured. In practice, we can first determine the form of each Pauli operator to be measured on A in the compressed circuit, and assume they all have +1 sign; the adativity can be re-introduced in post-processing, by flipping the measurement results appropriately.

9.2 Simulated noisy data for the 1D circuit

In this section, we provide classical numerical simulations as a reference for 1D circuit experimental data presented in the main text.

We first choose $\rho = \sigma = (|0\rangle\langle 0|)^{\otimes L}$, as in Fig. 5.3. In the circuit, we insert an erasure channel at each spacetime location of the ρ-circuit with probability $q = 0.1\%$, while keeping the σ-circuit noiseless. The erasure channel replaces the local density matrix with a maximally mixed one, and upon averaging over random circuit realizations becomes a weak depolarizing channel,

$$\mathscr{E}_x(\rho) = (1-q)\rho + q\left[\left(\frac{\mathbb{1}}{2}\right)_x \otimes \mathrm{tr}_x\rho\right]. \tag{9.13}$$

The results are shown in Fig. 9.1(a), where we see a decrease in $\chi^{\mathrm{noisy}}_{\rho=\sigma}$ when either L or p is increased. This trend is *qualitatively* consistent with what we observe in Fig. 5.3. The data can be fitted to the following functional form:

$$\chi^{\mathrm{noisy}}_{\rho=\sigma} \propto \exp\left[-\alpha(p,q)\cdot L^2\right], \tag{9.14}$$

where $\alpha(p,q)$ is a nonzero coefficient depending on p and q, see Fig. 9.1(b). As we will explain below, this form can be motivated from a statistical mechanics picture, see Eq. (9.24). However, this functional form is inconsistent with the experimental data, see Fig. 9.1(c).

Next we consider the $\rho \neq \sigma$ case, but instead with stabilizer initial states $\rho = \frac{1}{2^L}\mathbb{1}$ and $\sigma = (|0\rangle\langle 0|)^{\otimes L}$ to facilitate efficient classical simulation. In Fig. 9.2(a), we present numerical results obtained from a noiseless simulation. The overall trend of the results are in qualitative agreement with those in Fig. 5.4. The data collapse in Fig. 9.2(b) is performed with $p_c = 0.16$ and $\nu = 1.33$, as consistent with Ref. [310].

We also perform a noisy simulation for $\rho = \frac{1}{2^L}\mathbb{1}$ and $\sigma = (|0\rangle\langle 0|)^{\otimes L}$, where we insert an erasure channel at each spacetime location of the ρ-circuit with probability

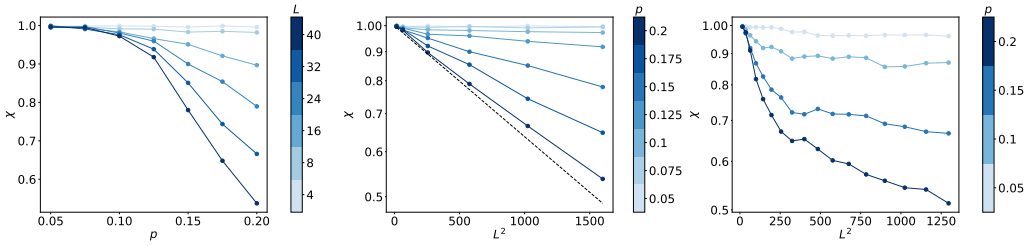


Figure 9.1: (Left) Results from noisy numerical simulations of Clifford circuits in 1D, for system sizes $L \leq 40$. We take the initial states $\rho = \sigma = (|0\rangle\langle 0|)^{\otimes L}$ as in Fig. 5.3, and randomly insert an erasure channel at each spacetime location of the ρ-circuit with probability $q = 0.1\%$. (Middle) We find the data consistent with the functional form in Eq.(9.14). (Right) Experimentally obtained χ. The non-linear behaviour may be caused due to coherent errors or other noise sources not captured by an erasure channel.

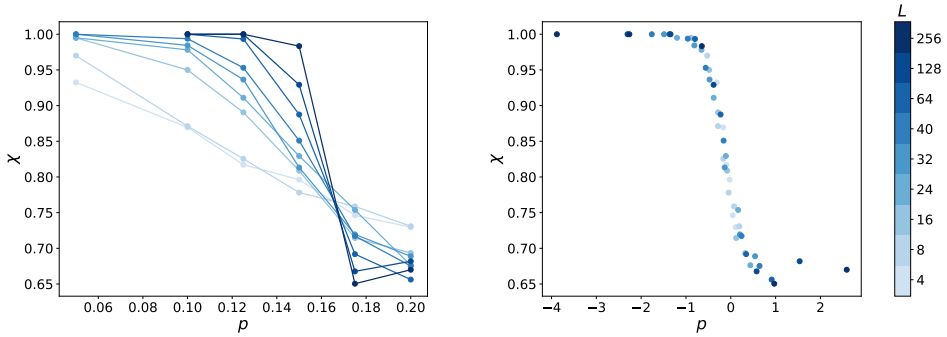


Figure 9.2: (Left) Results from noiseless numerical simulations of Clifford circuits in 1D, for system sizes $L \leq 256$. In our simulation, we take $\rho = \frac{1}{2^L}\mathbb{1}$ and $\sigma = (|0\rangle\langle 0|)^{\otimes L}$, as in Ref. [310]. (Right) When fitting the data to the scaling form in Eq. (5.8), we obtain $p_c \approx 0.16$ and $\nu \approx 1.33$, as consistent with Ref. [310].

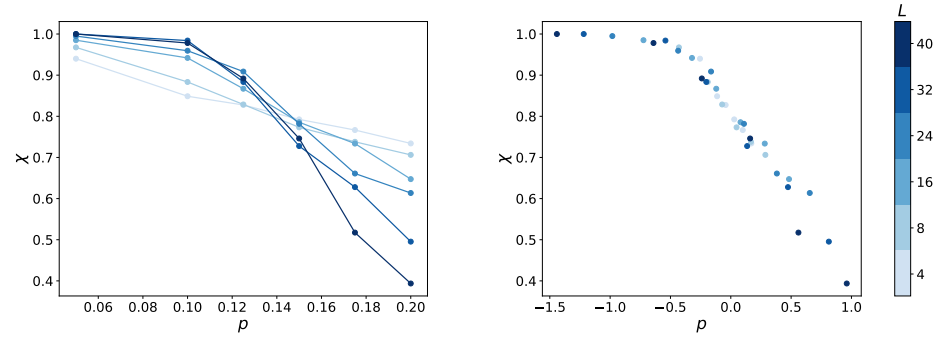


Figure 9.3: (Left) Results from noisy numerical simulations of Clifford circuits in 1D, for system sizes $L \leq 40$. We take the same initial states ρ and σ as in Fig. 9.2, and randomly insert an erasure channel at each spacetime location of the ρ-circuit with probability $q = 0.1\%$. (Right) When fitting the data to the scaling form in Eq. (5.8), we use $p_c \approx 0.14$ and $\nu \approx 1.33$ as obtained from Fig. 5.4 in the main text, where we find consistency.

$q = 0.1\%$. The numerical results are shown in Fig. 9.3. As we anticipate from statistical mechanics arguments (see Ref. [310] and below), for any finite noise rate, the cross entropy will be suppressed to zero for all value of p, in the thermodynamic limit. For small system sizes (before the cross entropy is reduced to zero) the curves will instead appear to cross at a smaller value of p_c. Indeed, the best fit for p_c has now shifted to a smaller value, $p_c \approx 0.14$ (whereas we use the same value for ν), close to the one used for fitting in the main text.

The qualitative behavior the results in Fig. 9.3 can be understood from a mapping to statistical mechanics models, which we briefly describe here. (We refer the reader

to Ref. [310] and references therein for further details.) Recall that

$$\chi := \mathbb{E}_C \chi_C = \mathbb{E}_C \frac{\sum_{\mathbf{m}} p^{\rho}_{\mathbf{m}} p^{\sigma}_{\mathbf{m}}}{\sum_{\mathbf{m}} (p^{\sigma}_{\mathbf{m}})^2} = \mathbb{E}_C \frac{\sum_{\mathbf{m}} \mathrm{tr}[C_{\mathbf{m}}(\rho)] \cdot \mathrm{tr}[C_{\mathbf{m}}(\sigma)]}{\sum_{\mathbf{m}} (\mathrm{tr}[C_{\mathbf{m}}(\sigma)])^2} = \mathbb{E}_C \frac{\sum_{\mathbf{m}} \mathrm{tr}[C^{\otimes 2}_{\mathbf{m}}(\rho \otimes \sigma)]}{\sum_{\mathbf{m}} \mathrm{tr}[C^{\otimes 2}_{\mathbf{m}}(\sigma \otimes \sigma)]}. \tag{9.15}$$

Here $C_m(\rho)$ denotes the resultant state when unitaries and projective measurements (labeled by the measurement record $\boldsymbol{m}$) from C are applied to the initial state ρ.[1] It is easier to study the following proxy quantity, which is an approximation of χ by averaging the numerator and the denominator separately over C,

$$\overline{\chi} = \frac{\mathbb{E}_C \sum_{\mathbf{m}} \mathrm{tr}[C^{\otimes 2}_{\mathbf{m}}(\rho \otimes \sigma)]}{\mathbb{E}_C \sum_{\mathbf{m}} \mathrm{tr}[C^{\otimes 2}_{\mathbf{m}}(\sigma \otimes \sigma)]}. \tag{9.16}$$

For C a brickwork circuit with local 2-qubit random unitary gates forming a 2-design, the averages can be performed. As a result, the numerator and the denominator will both take the form of a partition function of the Ising model on a triangular lattice, where the Boltzmann weights can be explicitly written down ([279, 300, 338–340]). The two partition functions are identical in the bulk, and only differ in their boundary conditions (coming from the difference in initial states). Following Ref. [310], we denote them $Z_{\rho\neq\sigma}$ and $Z_{\rho=\sigma}$, respectively.

In all our circuits we choose ρ and σ to be tensor products of onsite density matrices, and let them be different states. We also take the circuit to have a purely-unitary "encoding" stage without measurements, before measurements take place (see Fig. 5.2 of the main text). Within these circuits, $\overline{\chi} = Z_{\rho\neq\sigma}/Z_{\rho=\sigma}$ corresponds to the partition function ratio shown in Fig. 9.4(a). Each term lives in a rectangular geometry, with the lower half an Ising model at zero temperature (corresponding to the encoding stage), and the upper half at finite temperature [310]. The blue color denotes a "+" boundary condition, and the yellow color denotes a "−" one. The numerator $Z_{\rho\neq\sigma}$ has a boundary condition where both the top and bottom spins are fixed to be $+$, whereas $Z_{\rho=\sigma}$ has an additional contribution where the bottom boundary condition is also "−". Thus,

$$\overline{\chi}_{\rho\neq\sigma} = \frac{Z_{\rho\neq\sigma}}{Z_{\rho=\sigma}} = \frac{1}{1 + Z_{+-}/Z_{++}}. \tag{9.17}$$

The $p < p_c$ phase of circuit maps to the the ferromagnetic phase of the Ising magnet, where $-\ln(Z_{+-}/Z_{++})$ is the free energy of a horizontal domain wall separating the

[1] This notation is different from Ref. [310] to accommodate possible appearances of quantum channels.

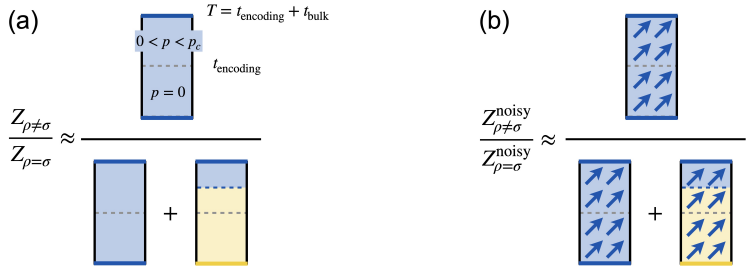


Figure 9.4: Mapping $\overline{\chi}$ defined in Eq. (9.17) to quantities in an effective Ising model, when the circuit is (a) noiseless and (b) noisy. See the text for more details. In both figures the blue color represents spins pointing in the "+" direction, the yellow color represents spins pointing in the "−" direction, and the black color represents a "free" boundary condition, where the spins can point in either direction.

bottom and the top (see Fig. 9.4(a)), which diverges with L, therefore $\overline{\chi} \to 1$. On the other hand, in the $p > p_c$ "paramagnetic" phase the domain wall free energy vanishes, so $Z_{+-}/Z_{++} \to 1$ and $\overline{\chi} \to 1/2$. We see that the numerical value of $\overline{\chi}$ in the $p > p_c$ phase differs from our numerical results, due to the annealed average.

The Ising picture is also useful for a qualitative understanding of the behavior of linear cross entropy in the presense of noise. For simplicity, we take the the noise to be a random erasure at each spacetime location. The cross entropy now reads

$$\chi := \mathbb{E}_{C,\mathscr{N}} \chi_{C,\mathscr{N}} = \mathbb{E}_{C,\mathscr{N}} \frac{\sum_{\mathbf{m}} \mathrm{tr}[(C'_{\mathbf{m}} \otimes C_{\mathbf{m}})(\rho \otimes \sigma)]}{\sum_{\mathbf{m}} \mathrm{tr}[C_{\mathbf{m}}^{\otimes 2}(\sigma \otimes \sigma)]}. \tag{9.18}$$

Here the circuit C' is obtained from C by inserting erasure noise (denoted $\mathscr{N}$) at random spacetime locations, which in general turns pure states into mixed states. Similarly, we define

$$\overline{\chi}^{\text{noisy}}_{\rho\neq\sigma} = \frac{\mathbb{E}_{C,\mathscr{N}} \sum_{\mathbf{m}} \mathrm{tr}[(C'_{\mathbf{m}} \otimes C_{\mathbf{m}})(\rho \otimes \sigma)]}{\mathbb{E}_C \sum_{\mathbf{m}} \mathrm{tr}[C_{\mathbf{m}}^{\otimes 2}(\sigma \otimes \sigma)]} = \frac{Z^{\text{noisy}}_{\rho\neq\sigma}}{Z_{\rho=\sigma}}. \tag{9.19}$$

This quantity is similar to our of experimental data in Fig. 5.4. We can also consider the following ratio:

$$\overline{\chi}^{\text{noisy}}_{\rho=\sigma} = \frac{Z^{\text{noisy}}_{\rho=\sigma}}{Z_{\rho=\sigma}}, \tag{9.20}$$

which approaches 1 as the noise rate vanishes, and is similar to Fig. 5.3. Both $Z^{\text{noisy}}_{\rho\neq\sigma}$ and $Z^{\text{noisy}}_{\rho=\sigma}$ can be obtained from their noiseless versions by applying a "magnetic field" everywhere in the system favoring the "+" direction and penalizing the "−" direction. More precisely, its effect can be captured by an additional term to the

energy function of the Ising model,

$$E[\{s\},h] = E[\{s\},h=0] + h\sum_j \delta_{s_j,-1} = E[\{s\},h=0] + \frac{h}{2}\sum_j (1-s_j), \tag{9.21}$$

where $Z[h] = \mathrm{tr}_{\{s\}} e^{-\beta E[\{s\},h]}$ is the Ising partition function, and h is the strength of the field (proportional to the strength of the noise). The field breaks the Ising symmetry and destroys the phase transition. For Eq. (9.20), we write the partition function in the numerator as follows:

$$Z^{\text{noisy}}_{\rho=\sigma} = e^{-Vf(h)}, \tag{9.22}$$

where $V \propto L^2$ is the circuit volume, and $f(h)$ is the free energy density when the field is applied to the magnet. Regardless of the phase the Ising magnet is in, a finite magnetization density $m(h)$ will appear, as a response to a small but finite h. The free energy density can then be approximated as

$$f(h) = f(h=0) + \frac{h}{2}(1-m(h)). \tag{9.23}$$

Therefore, we have

$$\overline{\chi}^{\text{noisy}}_{\rho=\sigma} = \frac{Z^{\text{noisy}}_{\rho=\sigma}}{Z_{\rho=\sigma}} \propto \exp\left[-\text{const}\cdot\frac{h}{2}(1-m(h))\cdot L^2\right]. \tag{9.24}$$

This is consistent with the functional form in Eq. (9.14) and numerical results Fig. 9.1(b). In particular, we observe in Fig. 9.1(b) a increasing rate of the exponential decay for a fixed noise rate and increasing p, corresponding to a smaller magnetization $m(h)$ as we raise the temperature.

A similar exponential dependence on L^2 is expected for $\overline{\chi}^{\text{noisy}}_{\rho\neq\sigma}$. The dependences will cancel if we take their ratio. As we illustrate in Fig. 9.4(b), their ratio should always be upper bounded by 1,

$$\frac{\overline{\chi}^{\text{noisy}}_{\rho\neq\sigma}}{\overline{\chi}^{\text{noisy}}_{\rho=\sigma}} = \frac{Z^{\text{noisy}}_{\rho\neq\sigma}}{Z^{\text{noisy}}_{\rho=\sigma}} = \frac{1}{1+Z_{+-}(h>0)/Z_{++}(h>0)} \leq 1, \tag{9.25}$$

as the Ising partition functions remain positive under the erasure channel. To complement the statistical mechanics approach, below in Sec. 9.4 we give a rigorous derivation of an upper bound of χ (rather than its proxy $\overline{\chi}$) in circuits with stabilizer operations. We also discuss the the apparent violation of the upper bound by experimental data.

9.3 Simulated noisy data for the all-to-all circuit

We perform classical numerical simulations for circuits with all-to-all connectivity, taking the same initial states as our 1D simulations. The results are shown in Fig. 9.5, 9.7. We fit both noiseless and noisy data to the scaling form in Eq. (5.8).

From the noiseless simulation (Fig. 9.5) of $L \leq 256$ we obtain fits $p_c \approx 0.33$ and $\nu \approx 2.50$. In particular, the critical exponent $\nu \approx 2.50$ agrees with a mean-field analysis as well as numerical simulations from Ref. [328]. We also observe that if we only include data from $L \leq 40$, then both the parameters here $(p_c, \nu) \approx (0.33, 2.50)$, and the best fits obtained from experimental data $(p_c, \nu) \approx (0.26, 1.90)$ (see Fig. 5.5), will result in high quality data collapses (data not shown). This is consistent with our observation of a large uncertainty in the fitting parameters in our experimental data from Fig. 5.5. Indeed, collapsing the experimental data from Fig. 5.5 with the theoretical value $\nu = 2.5$, we find reasonable agreement (see Fig. 9.6), even though $\nu = 2.5$ lies outside the 90% confidence interval, $\nu = 1.9 \pm 0.4$, as obtained from fitting procedures in Sec. 5.3.

On the other hand, from our noisy data at noise rate $q = 0.1\%$, we obtain $p_c \approx 0.20$ and $\nu \approx 0.80$, see Fig. 9.7. Recall that the same noise model and noise rate produced Fig. 9.3, which are comparable to experimental results in 1D. This suggests that noise affects the data strongly in all-to-all connectivity, and our experimental data cannot be fully captured by the simple simulated noise model.

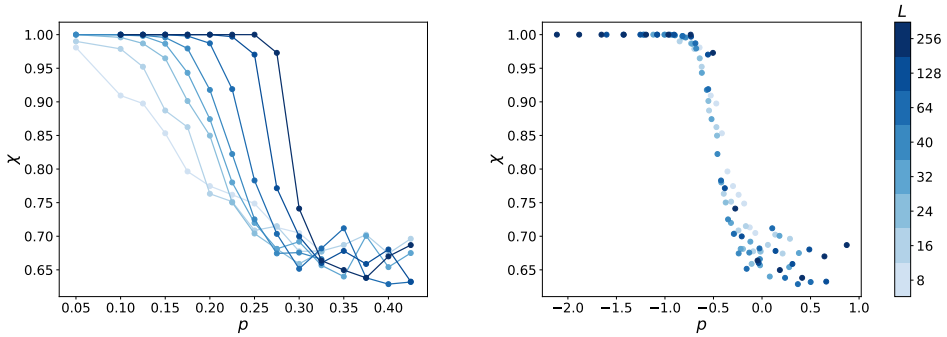


Figure 9.5: (Left) Results from noiseless numerical simulations of Clifford circuits with all-to-all connectivity, for system sizes $L \leq 256$. In our simulation, we take $\rho = \frac{1}{2^L}\mathbb{1}$ and $\sigma = (|0\rangle\langle 0|)^{\otimes L}$, identical to our choices in Fig. 9.2. (Right) When fitting the data to the scaling form in Eq. (5.8), we obtain $p_c \approx 0.33$ and $\nu \approx 2.50$.

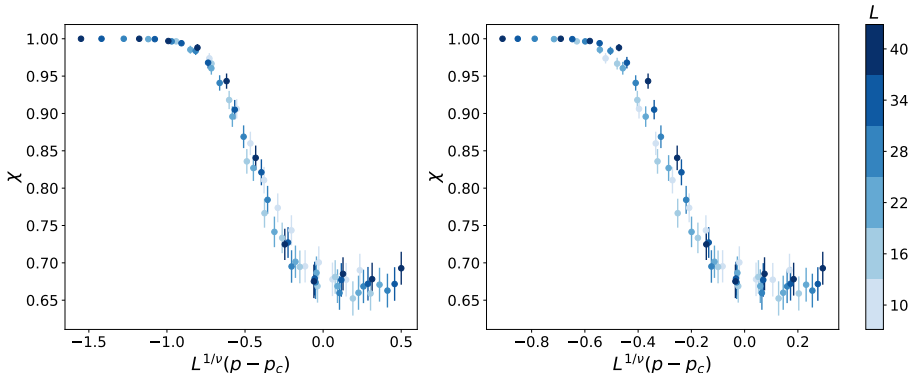


Figure 9.6: (Left) Data collapse for the experimentally obtained cross entropy when fitting both ν and p_c. (Right) Data collapse for the experimentally obtained cross entropy when setting ν to its theoretical value of $\nu = 2.5$.

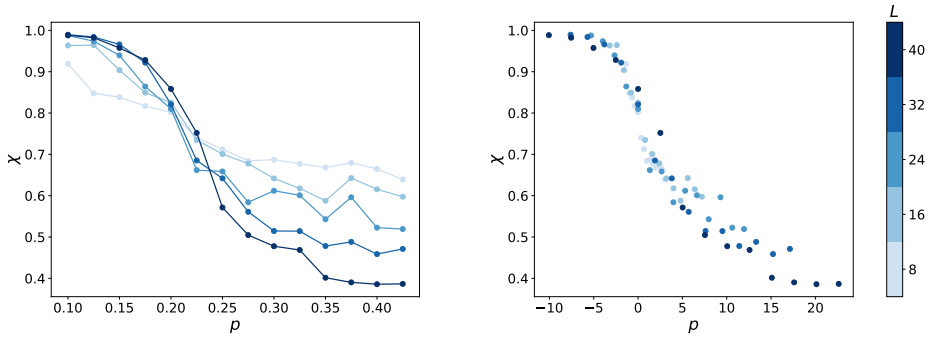


Figure 9.7: (Left) Results from noisy numerical simulations of Clifford circuits with all-to-all connectivity, for system sizes $L \leq 40$. We take the same initial states ρ and σ as in Fig. 9.5, and randomly insert an erasure channel at each spacetime location of the ρ-circuit with probability 0.1%. (Right) When fitting the data to the scaling form in Eq. (5.8), we find $p_c \approx 0.20$ and $\nu \approx 0.80$.

9.4 Effect of Pauli noise

Our stat mech picture in Appendix 9.2 suggests that $\chi^{\text{noisy}}_{\rho\neq\sigma} \leq \chi^{\text{noisy}}_{\rho=\sigma}$ (see Eq. (9.25)), while our experimental results clearly violate this relation, compare $\chi_{\rho=\sigma}$ (Fig. 5.3) with $\chi_{\rho\neq\sigma}$ (Fig. 5.4a). To get a better handle on this, we formalize the following characterization of linear cross entropy in stabilizer circuits (where the circuit architecture is arbitrary).

Proposition 8. *Let ρ and σ be two stabilizer states, which are in general different from each other. Consider a "noiseless" Clifford circuit C, composed of arbitrary Clifford unitaries and arbitrary Pauli measurements; and a noisy Clifford circuit C' obtained from C by injecting a number of stabilizer channels[2] at arbitrary space-*

[2]With stabilizer channels we mean channels that can be represented as stabilizer operations, of

time locations. Define

$$\chi(C',\rho|C,\sigma)=\frac{\sum_{\mathbf{m}}\mathrm{tr}[C'_m(\rho)]\cdot\mathrm{tr}[C_m(\sigma)].}{\sum_{\mathbf{m}}(\mathrm{tr}[C_m(\sigma)])^2}. \tag{9.26}$$

We have the following inequality between the linear cross entropies:

$$\chi(C',\rho|C,\sigma)\leq\chi(C',\sigma|C,\sigma). \tag{9.27}$$

Proof. We first adopt a purified representation for the Pauli measurements, see Appendix S2 of [310]. For each measurement of Pauli operator P in the circuit, we can introduce an additional register qubit, and apply a controlled Clifford unitary operator acting on the register qubit as well as qubits being measured, followed by a dephasing channel on the register, to simulate the effect of that measurement. In effect, at the end of the time evolution we have the following joint stabilizer states on the physical qubits Q and the register qubits R,

$$\rho_{QR}^{C'}=\sum_{\mathbf{m}}C'_{\mathbf{m}}(\rho)\otimes|\mathbf{m}\rangle\langle\mathbf{m}|_R, \tag{9.28}$$

$$\sigma_{QR}^{C'}=\sum_{\mathbf{m}}C'_{\mathbf{m}}(\sigma)\otimes|\mathbf{m}\rangle\langle\mathbf{m}|_R, \tag{9.29}$$

$$\sigma_{QR}^{C}=\sum_{\mathbf{m}}C_{\mathbf{m}}(\sigma)\otimes|\mathbf{m}\rangle\langle\mathbf{m}|_R. \tag{9.30}$$

With this representation, we have

$$\chi(C',\sigma|C,\sigma)=\frac{\mathrm{tr}[\sigma_R^{C'}\cdot\sigma_R^{C}]}{\mathrm{tr}[(\sigma_R^{C})^2]}, \tag{9.31}$$

$$\chi(C',\rho|C,\sigma)=\frac{\mathrm{tr}[\rho_R^{C'}\cdot\sigma_R^{C}]}{\mathrm{tr}[(\sigma_R^{C})^2]}, \tag{9.32}$$

where $\rho_R^{C'}$, $\sigma_R^{C'}$, and σ_R^{C} are reduced state of $\rho_{QR}^{C'}$, $\sigma_{QR}^{C'}$, and σ_{QR}^{C} on R, respectively.

Denote by $\mathscr{S}(\rho)$ the stabilizer group corresponding to a stabilizer state ρ. By induction, one can show that (Lemma 1, see below)

$$\mathscr{S}(\sigma_{QR}^{C'})\subseteq\mathscr{S}(\sigma_{QR}^{C}), \tag{9.33}$$

due to that C' is obtained from C by additional stabilizer channels. Such channels can only eliminate elements from the stabilizer group.

---

the form $\mathscr{E}(\cdot)=\frac{1}{2}(\cdot)+\frac{1}{2}P(\cdot)P$ with Pauli operator P (e.g. biased erasure errors), or their compositions (e.g. erasure errors).

By similar reasoning, we also have (Lemma 2, see below)

$$\mathscr{S}(\rho_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C}) \subseteq \mathscr{S}(\sigma_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C}) = \mathscr{S}(\sigma_{QR}^{C'}). \tag{9.34}$$

We can then calculate

$$\chi(C',\rho|C,\sigma) = \frac{\mathrm{tr}[\rho_R^{C'} \cdot \sigma_R^{C}]}{\mathrm{tr}[(\sigma_R^{C})^2]} \leq \frac{|\mathscr{S}(\rho_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C})|}{|\mathscr{S}(\sigma_{QR}^{C})|} \leq \frac{|\mathscr{S}(\sigma_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C})|}{|\mathscr{S}(\sigma_{QR}^{C})|} = \chi(C',\sigma|C,\sigma), \tag{9.35}$$

where we use the following result [310]:

$$\mathrm{tr}[\rho_R \sigma_R] = \frac{1}{2^{2|R|}} \sum_{g \in \mathscr{S}_{\rho_R}} \sum_{h \in \mathscr{S}_{\sigma_R}} \mathrm{tr}[gh] = \begin{cases} 2^{-|R|} |\mathscr{S}_{\rho_R} \cap \mathscr{S}_{\sigma_R}|, & -1 \notin \mathscr{S}_{\rho_R} \cdot \mathscr{S}_{\sigma_R} \\ 0, & -1 \in \mathscr{S}_{\rho_R} \cdot \mathscr{S}_{\sigma_R} \end{cases} \tag{9.36}$$

for any two stabilizer states on R. □

Lemma 9. $\mathscr{S}(\sigma_{QR}^{C'}) \subseteq \mathscr{S}(\sigma_{QR}^{C})$.

Proof. We can show this by induction on the quantum operations appearing in the circuit C' and C, which are Clifford unitaries and stabilizer channels. For brevity, we adopt the shorthand notation $\mathscr{S}_t'$ and $\mathscr{S}_t$ to denote the stabilizer groups of the instantaneous states for the two circuits C' and C after t quantum operations shared between C and C' are applied.

- At initialization, before any operation is applied, the two states are equal, so that $\mathscr{S}_0' = \mathscr{S}_0 \subseteq \mathscr{S}_0$.
- If the next operation is a unitary shared between C and C', we have that

  $$\mathscr{S}_t' \to \mathscr{S}_{t+1}' = U \mathscr{S}_t' U^\dagger, \quad \mathscr{S}_t \to \mathscr{S}_{t+1} = U \mathscr{S}_{t+1} U^\dagger. \tag{9.37}$$

  The inclusion is prevserved.
- If the next operation is a stabilizer channel shared between C and C' we have

  $$\mathscr{E}(\cdot) = \frac{1}{2}(\cdot) + \frac{1}{2} P(\cdot) P, \tag{9.38}$$

then

$$\mathscr{S}'_t \to \mathscr{S}'_{t+1} = \{g \in \mathscr{S}' : gP = Pg\} \subseteq \mathscr{S}'_t, \tag{9.39}$$

$$\mathscr{S}_t \to \mathscr{S}_{t+1} = \{g \in \mathscr{S} : gP = Pg\} \subseteq \mathscr{S}_t. \tag{9.40}$$

The inclusion is also prevserved.

- If the next operation is a stabilizer channel that is only in C' but not in C, $\mathscr{S}$ remains unchanged, and we have

$$\mathscr{S}'_t \to \{g \in \mathscr{S}'_t : gP = Pg\} \subseteq \mathscr{S}'_t \subseteq \mathscr{S}_t. \tag{9.41}$$

By induction, we conclude that $\mathscr{S}(\sigma_{QR}^{C'}) \subseteq \mathscr{S}(\sigma_{QR}^{C})$ after all operations are applied. □

Lemma 10. $\mathscr{S}(\rho_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C}) \subseteq \mathscr{S}(\sigma_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C})$.

Proof. The idea is similar to the proof of Lemma 1. Let the corresponding stabilizer groups at time t be denoted $\mathscr{S}'_\rho$, $\mathscr{S}'_\sigma$, and $\mathscr{S}_\sigma$, respectively. We want to show that $\mathscr{S}'_\rho \cap \mathscr{S}_\sigma \subseteq \mathscr{S}'_\sigma \cap \mathscr{S}_\sigma$ at all times, by induction. This property is true at initialization since $\mathscr{S}'_\sigma = \mathscr{S}_\sigma$ to start with. The preservation of this property under Clifford unitaries and quantum operations can also be straightforwardly verified. □

The inequality Eq. (9.27) applies to each sample from the ensemble of stabilizer circuits we considered in our numerical simulation above. Eq. (9.27) also applies to a slighlty broader class of error channels beyond stabilizers, e.g. if each additional stabilizer channel in C' (but not in C) is replaced by a probabilistic mixture of stabilizer channels, since linear cross entropy is linear in each channel, see Eq. (9.26). These include weak depolarizing or weak dephasing noise, as is usually assumed in the literature of random circuit sampling [313, 341–343].

Furthermore, we argue that the same inequality holds for a compressed circuit, as obtained from the algorithm in Sec. 5.3. Recall that the compression algorithm returns a circuit composed of measurements of Pauli operators only. In particular, a Pauli measurement in the uncompressed circuit either maps to another Pauli measurement in the compressed circuit, or to a classical coin flip. In the compressed circuit, the sign of a later Pauli operators might depend on earlier measurement results and/or earlier coin flip results, therefore "adaptive". To remove such adaptivity, we showed that one can simply assume that all Pauli operators (to be measured)

have the sign $+1$, and the adaptivity can be equivalently achieved by postprocessing the measurement results and the results of the classical coin flips. In particular, the postprocessing takes the form of an $\mathbb{F}_2$-linear map on $(\mathbb{F}_2)^N$, where N is the number measurements in the uncompressed circuit (which is equal to the number of measurements plus the number of coin flips in the compressed circuit). Thus the compressed circuit would seem to have more structure than the circuits C we considered above. To make our results applicable, we note that

- The classical coin flips can be simulated by initializing a corresponding register qubit in the maximally mixed state. We treat the other Pauli measurements as usual, by introducing an register qubit for each. This way, we have N register qubits in total.
- The "postprocessing" map can be realized by a Clifford unitary on the N register qubits, which is also diagonal in the computational basis.

This way, the compressed circuits can also be recasted as a stabilizer circuit with Clifford unitaries and stabilizer channels. For such a circuit C, and a "noisy" version C' obtained from C by inserting stabilizer channels (which can be seen as a crude approximation of the hardware experiments we carried out), we will also have

$$\chi(C',\rho|C,\sigma) \leq \chi(C',\sigma|C,\sigma). \tag{9.42}$$

The differences between the setup in our Proposition and the experiment are (i) the initial state ρ in our experiments is taken to be a nonstabilizer state, and (ii) the noise in our experiments is not simply of the form of a stabilizer channel (or their probabilistic mixture). The violation of the inequality Eq. (9.27) by our experimental results in Figs. 5.3,5.4 may thus be attributed to non-stabilizerness of the initial state, realistic error models (which necessarily involve coherent and non-unital noise, as well as read-out error), or a combination thereof. While it is easy to construct contrived example Clifford circuits with adversarial coherent noise that show violations of the bound (e.g. a unitary that exchanges the two initial state), we have not been able to find natural and physically relevant examples that can closely approximate the experimental data. It will be an interesting future direction to explore the effects and the description of non-stabilizer noise channels on MIPT, and conversely, the extent to which many-body phenomena in random circuits can be informative of noise.

We also state a result similar in spirit to Proposition 1, which applies to other choices of initial states. The key condition to Proposition 1 is the relation $\mathscr{S}(\rho_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C}) \subseteq \mathscr{S}(\sigma_{QR}^{C'}) \cap \mathscr{S}(\sigma_{QR}^{C})$, which is preserved throughout the time evolution. It is straightforward to a similar condition in the following scenario.

Proposition 2. Let $\rho_{1,2}$ and σ be three stabilizer states, which are in general different from each other. Let ρ_1 be obtainable from ρ_2 via stabilizer channels (or their probabilistic mixtures). Consider a "noiseless" Clifford circuit C and a noisy Clifford circuit C' specified the same way as in Proposition 1. We have the following inequality between the linear cross entropies:

$$\chi(C', \rho_1 | C, \sigma) \leq \chi(C', \rho_2 | C, \sigma). \tag{9.43}$$

9.5 Calculation of error bars

In order for the linear cross entropy to be a scalable probe for measurement induced phase transitions, the number of circuits and and circuit evaluations required for a given (L, p) pair must be polynomial in L, p, and $1/\varepsilon$, the error in estimating $\chi(L, p)$ from multiple samples. As shown in Reference [310], the number of samples can in fact be taken to be independent of L and p, and exhibits a linear dependence on N in $1/\varepsilon$, where N is the number of circuits used. We can see this dependence explicitly in the calculation of the error bars reported in the main text, shown in the following.

For a given (L, p) pair, we use N randomly generated circuits and execute each circuit M times on IBM's quantum hardware, resulting in M different measurement outcomes. We calculate the cross entropy for each circuit i as

$$\chi_i = \frac{1}{M} \sum_{j=1}^{M} x_{ij}, \tag{9.44}$$

where x_{ij} is the j'th measurement bit string for the i'th circuit and is defined as

$$x_{ij} = \begin{cases} 1, \text{ if } x_{ij} \text{ can occur on } \sigma_i \\ 0, \text{ if } x_{ij} \text{ cannot occur on } \sigma_i \end{cases}. \tag{9.45}$$

Here, σ_i is the σ circuit corresponding to the i'th ρ circuit. We next calculate the standard error of the mean as

$$\varepsilon_i = \frac{\hat{s}_i}{\sqrt{M}}, \quad \hat{s}_i^2 = \frac{1}{M-1} \sum_{j=1}^{M} (x_{ij} - \chi_i)^2. \tag{9.46}$$

We then compute the final estimate of the cross entropy as $\bar{\chi} = (1/N)\sum_{i=1}^{N} \chi_i$. The variance of $\bar{\chi}$ is given by

$$\varepsilon^2 = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i^2 \tag{9.47}$$

and the error bars reported in all figures are given by $\bar{\chi} \pm 1.96\varepsilon$, representing the 95% confidence interval for the estimate of χ.

9.6 Error mitigation for hardware experiments

Here we consider to error mitigation techniques, i.e. dynamical decoupling (DD), and readout error mitigation (ROEM).

DD is a quantum control technique employed in quantum computing to mitigate errors by taking advantage of time-dependent pulses [344–350]. In its simplest form, DD is implemented by sequences of X control pulses, whose effect is to protect qubits from decoherence due to low-frequency system-environment coupling. Here, we applied sequences of two X pulses (as in Ramsey echo experiments) to idle qubits. In Figure 9.8, we illustrate the impact of DD on the cross entropy, focusing on the $\rho = \sigma$ case. As seen, for $L \simeq 10$, DD increases the cross entropy towards the exact value of $\chi = 1$. However, the increase in χ is of order 0.01 whereas the difference between χ and 1 is of order 0.1 and, furthermore, it becomes less pronounced for $L \simeq 18$.

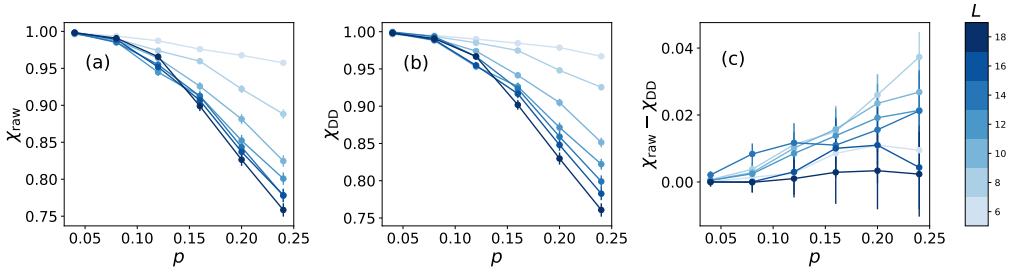


Figure 9.8: Cross entropy χ for chains of $L = 6$ to $L = 18$ qubits, with initial states $\rho = \sigma$, computed without (a) and with (b) dynamical decoupling, and difference between these two quantities (c).

ROEM is a standard technique to compensate for errors incurred during qubit readout [351, 352]. We tested ROEM for small systems of up to $L = 14$ (7 physical qubits) and observed negligible differences between the readout error mitigated cross entropies and the unmitigated cross entropies, see Figure 9.9. Due to the negligible effects of ROEM, we did not use ROEM for any of the results presented in the main text.

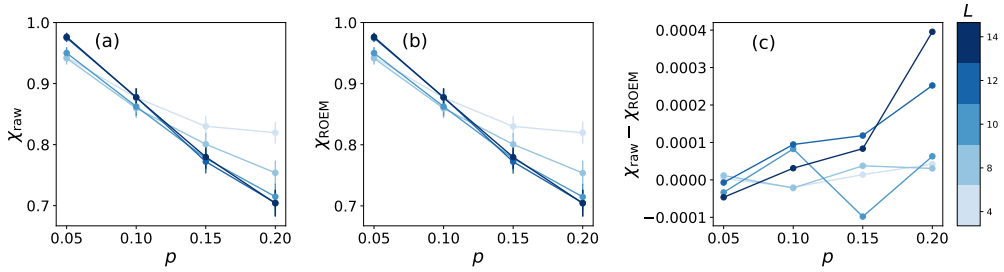


Figure 9.9: Effects of readout error mitigation on cross entropy for systems with up to 7 physical qubits. (a) The raw cross entropies without ROEM. (b) The cross entropies with ROEM applied. (c) The difference $\chi_{\text{raw}} - \chi_{\text{ROEM}}$, which shows that the differences between the raw and ROEM cross entropies are significantly smaller than the error bars for the raw cross entropies.

BIBLIOGRAPHY

1. Lee, J. G. *Computational materials science: an introduction* (CRC press, 2016).
2. Hafner, J. Atomic-scale computational materials science. *Acta Materialia* **48,** 71–92 (2000).
3. Blunt, N. S., Camps, J., Crawford, O., Izsák, R., Leontica, S., Mirani, A., Moylett, A. E., Scivier, S. A., Sunderhauf, C., Schopf, P., *et al.* Perspective on the current state-of-the-art of quantum computing for drug discovery applications. *Journal of Chemical Theory and Computation* **18,** 7001–7023 (2022).
4. Heifetz, A. *Quantum mechanics in drug discovery* (Springer, 2020).
5. Preskill, J. Quantum computing and the entanglement frontier. *arXiv preprint arXiv:1203.5813* (2012).
6. Sakurai, J. J. & Napolitano, J. *Modern quantum mechanics* (Cambridge University Press, 2020).
7. Kempe, J., Kitaev, A. & Regev, O. The complexity of the local Hamiltonian problem. *Siam journal on computing* **35,** 1070–1097 (2006).
8. Aharonov, D. & Naveh, T. Quantum NP-a survey. *arXiv preprint quant-ph/0210077* (2002).
9. Born, M. & Heisenberg, W. Zur quantentheorie der molekeln. *Original Scientific Papers Wissenschaftliche Originalarbeiten,* 216–246 (1985).
10. Yarkony, D. R. Diabolical conical intersections. *Reviews of Modern Physics* **68,** 985 (1996).
11. Domcke, W., Yarkony, D. & Köppel, H. *Conical intersections: theory, computation and experiment* (World Scientific, 2011).
12. Tapavicza, E., Bellchambers, G. D., Vincent, J. C. & Furche, F. Ab initio non-adiabatic molecular dynamics. *Physical Chemistry Chemical Physics* **15,** 18336–18348 (2013).
13. Baer, M. Introduction to the theory of electronic non-adiabatic coupling terms in molecular systems. *Physics Reports* **358,** 75–142 (2002).
14. Szabo, A. & Ostlund, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory* (Courier Corporation, 2012).
15. Boys, S. F. Electronic wave functions-I. A general method of calculation for the stationary states of any molecular system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **200,** 542–554 (1950).

16. Jørgensen, P. *Second quantization-based methods in quantum chemistry* (Elsevier, 2012).

17. Slater, J. C. A simplification of the Hartree-Fock method. *Physical review* **81,** 385 (1951).

18. Slater, J. C. The theory of complex spectra. *Physical review* **34,** 1293 (1929).

19. Møller, C. & Plesset, M. S. Note on an approximation treatment for many-electron systems. *Physical review* **46,** 618 (1934).

20. Shavitt, I. in *Methods of electronic structure theory* 189–275 (Springer, 1977).

21. Bartlett, R. J. & Musiał, M. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics* **79,** 291–352 (2007).

22. Bartlett, R. J. & Stanton, J. F. Applications of Post-Hartree—Fock Methods: A Tutorial. *Reviews in computational chemistry,* 65–169 (1994).

23. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters* **98,** 146401 (2007).

24. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **148** (2018).

25. Unke, O. T. & Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation* **15,** 3678–3693 (2019).

26. Bogojeski, M., Vogt-Maranto, L., Tuckerman, M. E., Müller, K.-R. & Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nature communications* **11,** 5223 (2020).

27. Hermann, J., Schätzle, Z. & Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nature Chemistry* **12,** 891–897 (2020).

28. Kitaev, A. Y. Quantum measurements and the Abelian stabilizer problem. *arXiv preprint quant-ph/9511026* (1995).

29. Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.-H., Zhou, X.-Q., Love, P. J., Aspuru-Guzik, A. & O'brien, J. L. A variational eigenvalue solver on a photonic quantum processor. *Nature communications* **5,** 4213 (2014).

30. Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nature Reviews Chemistry* **2,** 0109 (2018).

31. Westheimer, F. The magnitude of the primary kinetic isotope effect for compounds of hydrogen and deuterium. *Chemical reviews* **61,** 265–273 (1961).

32. Marx, D. & Parrinello, M. Ab initio path integral molecular dynamics: Basic ideas. *The Journal of chemical physics* **104,** 4077–4082 (1996).

33. Meyer, H.-D., Gatti, F. & Worth, G. A. *Multidimensional quantum dynamics: MCTDH theory and applications* (John Wiley & Sons, 2009).

34. Giustino, F. Electron-phonon interactions from first principles. *Reviews of Modern Physics* **89,** 015003 (2017).

35. Mahan, G. D. *Many-particle physics* (Springer Science & Business Media, 2013).

36. Askerov, B. M. *Electron transport phenomena in semiconductors* (World scientific, 1994).

37. Quan, Y., Yue, S. & Liao, B. Impact of electron-phonon interaction on thermal transport: A review. *Nanoscale and Microscale Thermophysical Engineering* **25,** 73–90 (2021).

38. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Microscopic theory of superconductivity. *Physical Review* **106,** 162 (1957).

39. Harris, S. *An introduction to the theory of the Boltzmann equation* (Courier Corporation, 2004).

40. Mitra, A. Quantum quench dynamics. *Annual Review of Condensed Matter Physics* **9,** 245–259 (2018).

41. Eisert, J., Friesdorf, M. & Gogolin, C. Quantum many-body systems out of equilibrium. *Nature Physics* **11,** 124–130 (2015).

42. Schaller, G. *Open quantum systems far from equilibrium* (Springer, 2014).

43. Nandkishore, R. & Huse, D. A. Many-body localization and thermalization in quantum statistical mechanics. *Annu. Rev. Condens. Matter Phys.* **6,** 15–38 (2015).

44. Chiorescu, I., Nakamura, Y., Harmans, C. M. & Mooij, J. Coherent quantum dynamics of a superconducting flux qubit. *Science* **299,** 1869–1871 (2003).

45. Agarwal, G. S. Fluctuation-dissipation theorems for systems in non-thermal equilibrium and applications. *Zeitschrift für Physik A Hadrons and nuclei* **252,** 25–38 (1972).

46. Itano, W. M., Heinzen, D. J., Bollinger, J. J. & Wineland, D. J. Quantum zeno effect. *Physical Review A* **41,** 2295 (1990).

47. Skinner, B., Ruhman, J. & Nahum, A. Measurement-induced phase transitions in the dynamics of entanglement. *Physical Review X* **9,** 031009 (2019).

48. Bao, Y., Choi, S. & Altman, E. Theory of the phase transition in random unitary circuits with measurements. *Physical Review B* **101,** 104301 (2020).

49. Orús, R. Tensor networks for complex quantum systems. *Nature Reviews Physics* **1,** 538–550 (2019).

50. Von der Linden, W. A quantum Monte Carlo approach to many-body physics. *Physics Reports* **220,** 53–162 (1992).

51. Aoki, H., Tsuji, N., Eckstein, M., Kollar, M., Oka, T. & Werner, P. Nonequilibrium dynamical mean-field theory and its applications. *Reviews of Modern Physics* **86,** 779–837 (2014).

52. Smith, A., Kim, M., Pollmann, F. & Knolle, J. Simulating quantum many-body dynamics on a current digital quantum computer. *npj Quantum Information* **5,** 106 (2019).

53. Potter, A. C. & Vasseur, R. in *Entanglement in Spin Chains: From Theory to Quantum Technology Applications* 211–249 (Springer, 2022).

54. Sholl, D. S. & Steckel, J. A. *Density functional theory: a practical introduction* (John Wiley & Sons, 2022).

55. Verma, P. & Truhlar, D. G. Status and challenges of density functional theory. *Trends in Chemistry* **2,** 302–318 (2020).

56. Burke, K. Perspective on density functional theory. *The Journal of chemical physics* **136** (2012).

57. Cohen, A. J., Mori-Sánchez, P. & Yang, W. Challenges for density functional theory. *Chemical reviews* **112,** 289–320 (2012).

58. Bryenton, K. R., Adeleke, A. A., Dale, S. G. & Johnson, E. R. Delocalization error: The greatest outstanding challenge in density-functional theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **13,** e1631 (2023).

59. Von Lilienfeld, O. A. & Burke, K. Retrospective on a decade of machine learning for chemical discovery. *Nat. Commun.* **11,** 1–4 (2020).

60. Westermayr, J., Gastegger, M., Schütt, K. T. & Maurer, R. J. Perspective on integrating machine learning into computational chemistry and materials science. *J. Chem. Phys.* **154,** 230903 (2021).

61. Freeze, J. G., Kelly, H. R. & Batista, V. S. Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. *Chem. Rev.* **119,** 6595–6612 (2019).

62. Elton, D. C., Boukouvalas, Z., Fuge, M. D. & Chung, P. W. Deep learning for molecular design—a review of the state of the art. *Mol. Syst. Des. Eng.* **4,** 828–849 (2019).

63. Von Lilienfeld, O. A., Lins, R. D. & Rothlisberger, U. Variational particle number approach for rational compound design. *Phys. Rev. Lett.* **95,** 153002 (2005).

64. Von Lilienfeld, O. A. & Tuckerman, M. Alchemical variations of intermolecular energies according to molecular grand-canonical ensemble density functional theory. *J. Chem. Theory Comput.* **3,** 1083–1090 (2007).

65. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4,** eaap7885. ISSN: 2375-2548. http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aap7885 (July 2018).

66. Yang, X., Wang, Y., Byrne, R., Schneider, G. & Yang, S. Concepts of artificial intelligence for computer-assisted drug discovery. *Chem. Rev.* **119,** 10520–10594 (2019).

67. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4,** 331–348 (2019).

68. Tran, A., Tranchida, J., Wildey, T. & Thompson, A. P. Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys. *J. Chem. Phys.* **153,** 074705 (2020).

69. Grisafi, A., Fabrizio, A., Meyer, B., Wilkins, D. M., Corminboeuf, C. & Ceriotti, M. Transferable Machine-Learning Model of the Electron Density. *ACS Cent. Sci.* **5,** 57–64 (2019).

70. Manzhos, S. & Carrington Jr, T. Neural network potential energy surfaces for small molecules and reactions. *Chem. Rev.* **121,** 10187–10217 (2020).

71. Ceriotti, M. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *J. Chem. Phys.* **150,** 150901 (2019).

72. Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G. & Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571,** 95–98 (2019).

73. Zhou, Z., Kearnes, S., Li, L., Zare, R. N. & Riley, P. Optimization of molecules via deep reinforcement learning. *Sci. Rep.* **9,** 1–10 (2019).

74. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361,** 360–365. ISSN: 1095-9203. http://www.ncbi.nlm.nih.gov/pubmed/30049875 (July 2018).

75. Schwalbe-Koda, D. & Gómez-Bombarelli, R. in *Machine Learning Meets Quantum Physics* 445–467 (Springer, 2020).

76. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104,** 136403. ISSN: 0031-9007. https://link.aps.org/doi/10.1103/PhysRevLett.104.136403 (Apr. 2010).

77. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108,** 58301 (2012).

78. Montavon, G., Rupp, M., Gobre, V., Vazquez-Mayagoitia, A., Hansen, K., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **15,** 95003 (2013).

79. Hansen, K., Montavon, G., Biegler, F., Fazli, S., Rupp, M., Scheffler, M., von Lilienfeld, O. A., Tkatchenko, A. & Müller, K.-R. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.* **9,** 3404 (2013).

80. Gasparotto, P. & Ceriotti, M. Recognizing molecular patterns by machine learning: An agnostic structural definition of the hydrogen bond. *J. Chem. Phys.* **141,** 174110 (2014).

81. Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Big data meets quantum chemistry approximations: the Δ-machine learning approach. *Journal of chemical theory and computation* **11,** 2087–2096 (2015).

82. Brockherde, F., Vogt, L., Li, L., Tuckerman, M. E., Burke, K. & Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **8,** 872 (2017).

83. Kearnes, S., McCloskey, K., Berndl, M., Pande, V. & Riley, P. Molecular graph convolutions: Moving beyond fingerprints. *J. Comput. Aided Mol. Des.* **30,** 595 (2016).

84. Paesani, F. Getting the right answers for the right reasons: toward predictive molecular simulations of water with many-body potential energy functions. *Acc. Chem. Res.* **49,** 1844 (2016).

85. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145,** 170901. ISSN: 0021-9606. http://aip.scitation.org/doi/10.1063/1.4966192 (Nov. 2016).

86. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K.-R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8,** 13890 (2017).

87. Schütt, K., Kindermans, P.-J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A. & Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **30** (2017).

88. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8,** 3192–3203. ISSN: 20416539 (2017).

89. Welborn, M., Cheng, L. & Miller III, T. F. Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **14,** 4772–4779. http://pubs.acs.org/doi/10.1021/acs.jctc.8b00636 (Sept. 2018).

90. Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K. & Pande, V. MoleculeNet: A benchmark for molecular machine learning. *Chem. Sci.* **9,** 513 (2018).

91. Nguyen, T. T., Székely, E., Imbalzano, G., Behler, J., Csányi, G., Ceriotti, M., Götz, A. W. & Paesani, F. Comparison of permutationally invariant polynomials, neural networks, and Gaussian approximation potentials in representing water interactions through many-body expansions. *J. Chem. Phys.* **148,** 241725 (2018).

92. Yao, K., Herr, J. E., Toth, D. W., McKintyre, R. & Parkhill, J. The TensorMol-0.1 model chemistry: A neural network augmented with long-range physics. *Chem. Sci.* **9,** 2261–2269. ISSN: 20416539 (2018).

93. Fujikake, S., Deringer, V. L., Lee, T. H., Krynski, M., Elliott, S. R. & Csányi, G. Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures. *J. Chem. Phys.* **148,** 241714. ISSN: 0021-9606. http://aip.scitation.org/doi/10.1063/1.5016317 (June 2018).

94. Cheng, B., Engel, E. A., Behler, J., Dellago, C. & Ceriotti, M. Ab initio thermodynamics of liquid and solid water. *Proceedings of the National Academy of Sciences* **116,** 1110–1115 (2019).

95. Cheng, L., Kovachki, N. B., Welborn, M. & Miller III, T. F. Regression Clustering for Improved Accuracy and Training Costs with Molecular-Orbital-Based Machine Learning. *J. Chem. Theory Comput.* **15,** 6668–6677 (2019).

96. Dick, S. & Fernandez-Serra, M. Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Commun.* **11,** 1–10 (2020).

97. Chen, Y., Zhang, L., Wang, H. & E, W. Ground State Energy Functional with Hartree–Fock Efficiency and Chemical Accuracy. *J. Phys. Chem. A* **124,** 7155–7165 (2020).

98. Qiao, Z., Welborn, M., Anandkumar, A., Manby, F. R. & Miller III, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **153,** 124111 (2020).

99. Qiao, Z., Ding, F., Welborn, M., Bygrave, P. J., Smith, D. G., Anandkumar, A., Manby, F. R. & Miller III, T. F. Multi-task learning for electronic structure to predict and explore molecular potential energy surfaces. *arXiv preprint arXiv:2011.02680* (2020).

100. Deringer, V. L., Bernstein, N., Csányi, G., Ben Mahmoud, C., Ceriotti, M., Wilson, M., Drabold, D. A. & Elliott, S. R. Origins of structural and electronic transitions in disordered silicon. *Nature* **589,** 59–64 (2021).

101. Christensen, A. S., Sirumalla, S. K., Qiao, Z., O'Connor, M. B., Smith, D. G., Ding, F., Bygrave, P. J., Anandkumar, A., Welborn, M., Manby, F. R., *et al.* OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy. *J. Chem. Phys.* **155,** 204103 (2021).

102. Husch, T., Sun, J., Cheng, L., Lee, S. J. & Miller III, T. F. Improved accuracy and transferability of molecular-orbital-based machine learning: Organics, transition-metal complexes, non-covalent interactions, and transition states. *J. Chem. Phys.* **154,** 064108 (2021).

103. Lee, S. J. R., Husch, T., Ding, F. & Miller III, T. F. Analytical gradients for molecular-orbital-based machine learning. *J. Chem. Phys.* **154,** 124120 (2021).

104. Sun, J., Cheng, L. & Miller III, T. F. *Molecular Energy Learning Using Alternative Blackbox Matrix-Matrix Multiplication Algorithm for Exact Gaussian Process* in *NeurIPS 2021 AI for Science Workshop* (2021). https://openreview.net/forum?id=lyJ9BRKUzms.

105. Karandashev, K. & von Lilienfeld, O. A. An orbital-based representation for accurate quantum machine learning. *J. Chem. Phys.* **156,** 114101. https://doi.org/10.1063/5.0083301 (2022).

106. Veit, M., Wilkins, D. M., Yang, Y., DiStasio, R. A. & Ceriotti, M. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J. Chem. Phys.* **153,** 024113 (2020).

107. Qiao, Z., Christensen, A. S., Welborn, M., Manby, F. R., Anandkumar, A. & Miller, T. F. Informing Geometric Deep Learning with Electronic Interactions to Accelerate Quantum Chemistry. *arXiv preprint arXiv:2011.02680* (2021).

108. Klicpera, J., Giri, S., Margraf, J. T. & Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115* (2020).

109. Liu, Y., Wang, L., Liu, M., Lin, Y., Zhang, X., Oztekin, B. & Ji, S. *Spherical Message Passing for 3D Molecular Graphs* in *International Conference on Learning Representations* (2022). https://openreview.net/forum?id=givsRXsOt9r.

110. Schütt, K., Unke, O. & Gastegger, M. *Equivariant message passing for the prediction of tensorial properties and molecular spectra* in *Proceedings of the 38th International Conference on Machine Learning* (eds Meila, M. & Zhang, T.) **139** (PMLR, 18–24 Jul 2021), 9377–9388. https://proceedings.mlr.press/v139/schutt21a.html.

111. Faber, F. A., Christensen, A. S., Huang, B. & Von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **148,** 241717 (2018).

112. Huang, B. & von Lilienfeld, O. A. Quantum machine learning using atom-in-molecule-based fragments selected on the fly. *Nat. Chem.* **12,** 945–951 (2020).

113. Bartók, A. P., De, S., Poelking, C., Bernstein, N., Kermode, J. R., Csányi, G. & Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **3,** e1701816 (2017).

114. Christensen, A. S., Faber, F. A. & von Lilienfeld, O. A. Operators in quantum machine learning: Response properties in chemical space. *J. Chem. Phys.* **150,** 064105 (2019).

115. Cheng, L., Sun, J. & Miller III, T. F. Improved Accuracy of Molecular Energy Learning via Unsupervised Clustering for Organic Chemical Space with Molecular-orbital-based Machine Learning. *arXiv preprint arXiv:2204.09831* (2022).

116. Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q. & Wilson, A. G. *GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration* in *Advances in Neural Information Processing Systems* (2018), 7576–7586.

117. Wang, K., Pleiss, G., Gardner, J., Tyree, S., Weinberger, K. Q. & Wilson, A. G. *Exact Gaussian Processes on a Million Data Points* in *Advances in Neural Information Processing Systems* **32** (Curran Associates, Inc., 2019), 14648–14659.

118. Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1,** 1–7 (2014).

119. Welborn, M., Cheng, L. & Miller III, T. F. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **14,** 4772–4779 (2018).

120. Cheng, L., Welborn, M., Christensen, A. S. & Miller III, T. F. A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules. *J. Chem. Phys.* **150,** 131103 (2019).

121. Cheng, L., Kovachki, N. B., Welborn, M. & Miller III, T. F. Regression Clustering for Improved Accuracy and Training Costs with Molecular-Orbital-Based Machine Learning. *J. Chem. Theory Comput.* **15,** 6668–6677 (2019).

122. Nesbet, R. K. Brueckner's Theory and the Method of Superposition of Configurations. *Phys. Rev.* **109,** 1632–1638 (1958).

123. Boys, S. F. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **32,** 296–299. ISSN: 00346861 (1960).

124. Møller, C. & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **46,** 618–622 (1934).

125. Chen, Y., Zhang, L., Wang, H. & E, W. Ground state energy functional with Hartree-Fock efficiency and chemical accuracy. *J. Phys. Chem. A* **124,** 7155–7165 (2020).

126. Dick, S. & Fernandez-Serra, M. Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Comm.* **11,** 3509 (2020).

127. Qiao, Z., Welborn, M., Anandkumar, A., Manby, F. R. & Miller III, T. F. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *arXiv:2007.08026 [physics]* (2020).

128. Rasmussen, C. E. & Williams, C. K. I. *Gaussian processes for machine learning* http://www.gaussianprocess.org/gpml/chapters/RW.pdf (MIT Press, Cambridge, MA, 2006).

129. Rasmussen, C. E. & Williams, C. K. I. *Gaussian processes for machine learning* Publication Title: Gaussian processes for machine learning. http://www.gaussianprocess.org/gpml/chapters/RW.pdf (MIT Press, Cambridge, MA, 2006).

130. Kapuy, E., Csépes, Z. & Kozmutza, C. Application of the many-body perturbation theory by using localized orbitals. *Int. J. Quantum Chem.* **23,** 981 (1983).

131. Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **32,** 359–401 (1981).

132. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry* 261–265. ISBN: 0486691861 (Dover, Mineola, 1996).

133. Grimme, S., Hansen, A., Brandenburg, J. G. & Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **116,** 5105–5154 (2016).

134. Hirschfeld, L., Swanson, K., Yang, K., Barzilay, R. & Coley, C. W. Uncertainty Quantification Using Neural Networks for Molecular Property Prediction. *J. Chem. Inf. Model* **60,** 3770–3780 (2020).

135. Brockherde, F., Vogt, L., Li, L., Tuckerman, M. E., Burke, K. & Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Comm.* **8,** 872 (2017).

136. Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **48,** 722–730 (2015).

137. Blum, L. C. & Reymond, J.-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **131,** 8732 (2009).

138. Christensen, A. S., Bratholm, L. A., Faber, F. A. & Anatole von Lilienfeld, O. FCHL revisited: Faster and more accurate quantum machine learning. *J. Chem. Phys.* **152,** 044107 (2020).

139. Lee, S. J. R., Husch, T., Ding, F. & Miller III, T. F. Analytical Gradients for Molecular-Orbital-Based Machine Learning. *arXiv:2012.08899* (2020).

140. Burns, L. A., Faver, J. C., Zheng, Z., Marshall, M. S., Smith, D. G. A., Vanommeslaeghe, K., MacKerell, A. D., Merz, K. M. & Sherrill, C. D. The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **147,** 161727 (2017).

141. Janet, J. P. & Kulik, H. J. Predicting electronic structure properties oftransition metal complexes with neural networks. *Chem. Sci.* **8,** 5137–5152 (2017).

142. Weymuth, T., Couzijn, E. P. A., Chen, P. & Reiher, M. New Benchmark Set of Transition-Metal Coordination Reactions for the Assessment of Density Functionals. *J. Chem. Theory ut.* **10,** 3092–3103 (2014).

143. Husch, T., Freitag, L. & Reiher, M. Calculation of Ligand Dissociation Energies in Large Transition-Metal Complexes. *J. Chem. Theory Comput.* **14,** 2456–2468 (2018).

144. Titsias, M. *Variational learning of inducing variables in sparse Gaussian processes* in *Artificial intelligence and statistics* (2009), 567–574.

145. Hensman, J., Fusi, N. & Lawrence, N. D. *Gaussian processes for Big data* in *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence* (2013), 282–290.

146. Van der Vorst, H. A. *Iterative Krylov methods for large linear systems* **13** (Cambridge University Press, 2003).

147. Cutajar, K., Osborne, M., Cunningham, J. & Filippone, M. *Preconditioning kernel matrices* in *International Conference on Machine Learning* (2016), 2529–2538.

148. O'Leary, D. P. The block conjugate gradient algorithm and related methods. *Linear Algebra Its Appl.* **29,** 293–322 (1980).

149. Bach, F. *Sharp analysis of low-rank kernel matrix approximations* in *Conference on Learning Theory* (2013), 185–209.

150. Harbrecht, H., Peters, M. & Schneider, R. On the low-rank approximation by the pivoted Cholesky decomposition. *Appl. Numer. Math.* **62,** 428–440 (2012).

151. Greenbaum, A. *Iterative methods for solving linear systems* (SIAM, 1997).

152. Husch, T., Sun, J., Cheng, L., Lee, S. J. & Miller III, T. F. Improved Accuracy and Transferability of Molecular-orbital-based Machine Learning: Organics, Transition-metal Complexes, Non-covalent Interactions, and Transition States. *J. Chem. Phys.* **154,** 064108 (2021).

153. Chen, Y., Zhang, L., Wang, H. & E, W. Ground State Energy Functional with Hartree–Fock Efficiency and Chemical Accuracy. *J. Phys. Chem. A* **124,** 7155–7165 (2020).

154. Parzen, E. *Stochastic processes* (SIAM, 1999).

155. Marsden, J. E. & Tromba, A. *Vector calculus* (Macmillan, 2003).

156. Chmiela, S., Tkatchenko, A., Sauceda, H. E., Poltavsky, I., Schütt, K. T. & Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3,** e1603015 (2017).

157. Pulay, P. & Saebø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theor. Chim. Acta* **69,** 357–368 (1986).

158. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. Scikit-learn: machine learning in python (v0.21.2). *J. Mach. Learn. Res.* **12,** 2825 (2011).

159. Okuta, R., Unno, Y., Nishino, D., Hido, S. & Loomis, C. *CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations* in *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)* (2017). http://learningsys.org/nips17/assets/papers/paper_16.pdf.

160. Liu, H., Cai, J. & Ong, Y.-S. Remarks on multi-output Gaussian process regression. *Knowl. Based Syst.* **144,** 102–121 (2018).

161. Schwabe, T. & Grimme, S. Theoretical thermodynamics for large molecules: walking the thin line between accuracy and computational cost. *Acc. Chem. Res.* **41,** 569–579 (2008).

162. Tirado-Rives, J. & Jorgensen, W. L. Performance of B3LYP density functional methods for a large set of organic molecules. *J. Chem. Theory Comput.* **4,** 297–306 (2008).

163. Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., Suhai, S. & Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58,** 7260 (1998).

164. Parrinello, M. & Rahman, A. Study of an F center in molten KCl. *The Journal of Chemical Physics* **80,** 860–867 (1984).

165. Habershon, S., Manolopoulos, D. E., Markland, T. E. & Miller III, T. F. Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space. *Annual Reviews of Physical Chemistry* **64,** 387–413 (2013).

166. Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nature Reviews Chemistry* **2,** 0109 (2018).

167. Feynman, R. P. *Quantum mechanics and path integrals* 1st (McGraw-Hill, 1965).

168. Chandler, D. & Wolynes, P. G. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *The Journal of Chemical Physics* **74,** 4078–4095 (1981).

169. Frenkel, D. & Smit, B. *Understanding molecular simulation: From algorithms to applications* 2nd (Academic Press, 2002).

170. Rapaport, D. C. *The art of molecular dynamics simulation* 2nd (Cambridge University Press, UK, 2004).

171. Leimkuhler, B. & Matthews, C. *Molecular Dynamics* 1st (Springer International Publishing, 2015).

172. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* 2nd (Oxford University Press, 2017).

173. Cao, J. & Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties. *The Journal of Chemical Physics* **100,** 5106–5117 (1994).

174. Craig, I. R. & Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *The Journal of Chemical Physics* **121,** 3368–3373 (2004).

175. Liu, J. Path integral Liouville dynamics for thermal equilibrium systems. *The Journal of Chemical Physics* **140,** 224107 (2014).

176. Hele, T. J. H., Willatt, M. J., Muolo, A. & Althorpe, S. C. Boltzmann-conserving classical dynamics in quantum time-correlation functions: "Matsubara dynamics". *The Journal of Chemical Physics* **142,** 134103 (2015).

177. Hele, T. J. H., Willatt, M. J., Muolo, A. & Althorpe, S. C. Communication: Relation of centroid molecular dynamics and ring-polymer molecular dynamics to exact quantum dynamics. *The Journal of Chemical Physics* **142,** 191101 (2015).

178. Cendagorta, J. R., Bačić, Z. & Tuckerman, M. E. An open-chain imaginary-time path-integral sampling approach to the calculation of approximate symmetrized quantum time correlation functions. *The Journal of Chemical Physics* **148,** 102340 (2018).

179. Craig, I. R. & Manolopoulos, D. E. Chemical reaction rates from ring polymer molecular dynamics. *The Journal of Chemical Physics* **122,** 084106 (2005).

180. Craig, I. R. & Manolopoulos, D. E. A refined ring polymer molecular dynamics theory of chemical reaction rates. *The Journal of Chemical Physics* **123,** 034102 (2005).

181. Miller III, T. F. & Manolopoulos, D. E. Quantum diffusion in liquid water from ring polymer molecular dynamics. *The Journal of Chemical Physics* **123,** 154504 (2005).

182. Miller III, T. F. & Manolopoulos, D. E. Quantum diffusion in liquid parahydrogen from ring-polymer molecular dynamics. *The Journal of Chemical Physics* **122,** 184503 (2005).

183. Habershon, S., Fanourgakis, G. S. & Manolopoulos, D. E. Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water. *The Journal of Chemical Physics* **129,** 074501 (2008).

184. Witt, A., Ivanov, S. D., Shiga, M., Forbert, H. & Marx, D. On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy. *The Journal of Chemical Physics* **130,** 194510 (2009).

185. Morrone, J. A. & Car, R. Nuclear quantum effects in water. *Physical Review Letters* **101,** 017801 (1 2008).

186. Zimmermann, T. & Vaníček, J. Path integral evaluation of equilibrium isotope effects. *The Journal of Chemical Physics* **131,** 024111 (2009).

187. Eldridge, D. L., Korol, R., Lloyd, M. K., Turner, A. C., Webb, M. A., Miller III, T. F. & Stolper, D. A. Comparison of experimental vs theoretical abundances of $^{13}CH_3D$ and $^{12}CH_2D_2$ for isotopically equilibrated systems from 1 to 500 °C. *ACS Earth and Space Chemistry* **3,** 2747–2764 (2019).

188. Trotter, H. F. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society* **10,** 545–551 (1959).

189. Strang, G. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis* **5,** 506–517 (1968).

190. Tuckerman, M. E., Berne, B. J., Martyna, G. J. & Klein, M. L. Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *The Journal of Chemical Physics* **99,** 2796–2808 (1993).

191. Ceriotti, M., Parrinello, M., Markland, T. E. & Manolopoulos, D. E. Efficient stochastic thermostatting of path integral molecular dynamics. *The Journal of Chemical Physics* **133,** 124104 (2010).

192. Liu, J., Li, D. & Liu, X. A simple and accurate algorithm for path integral molecular dynamics with the Langevin thermostat. *The Journal of Chemical Physics* **145,** 024103 (2016).

193. Calvo, M. P. & Sanz-Serna, J. M. Instabilities and inaccuracies in the integration of highly oscillatory problems. *SIAM Journal on Scientific Computing* **31,** 1653–1677 (2009).

194. Arnol'd, V. I. *Mathematical methods of classical mechanics* 2nd (Springer Science & Business Media, 2013).

195. Minary, P., Martyna, G. J. & Tuckerman, M. E. Algorithms and novel applications based on the isokinetic ensemble. I. Biophysical and path integral molecular dynamics. *The Journal of Chemical Physics* **118,** 2510–2526 (2003).

196. Bou-Rabee, N. & Sanz-Serna, J. M. Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numerica* **27,** 113–206 (2018).

197. Bou-Rabee, N. & Eberle, A. Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions. *Stochastics and Partial Differential Equations: Analysis and Computations* (2020).

198. Lu, J., Lu, Y. & Zhou, Z. Continuum limit and preconditioned Langevin sampling of the path integral molecular dynamics. *Journal of Computational Physics* **423,** 109788 (2020).

199. Ceriotti, M., Manolopoulos, D. E. & Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *The Journal of Chemical Physics* **134,** 084104 (2011).

200. Rossi, M., Ceriotti, M. & Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. *The Journal of Chemical Physics* **140,** 234116 (2014).

201. Zhang, Z., Liu, X., Chen, Z., Zheng, H., Yan, K. & Liu, J. A unified thermostat scheme for efficient configurational sampling for classical/quantum canonical ensembles via molecular dynamics. *The Journal of Chemical Physics* **147,** 034109 (2017).

202. Rossi, M., Kapil, V. & Ceriotti, M. Fine tuning classical and quantum molecular dynamics using a generalized Langevin equation. *Journal of Chemical Physics* **148,** 102301 (2018).

203. Korol, R., Bou-Rabee, N. & Miller III, T. F. Cayley modification for strongly stable path-integral and ring-polymer molecular dynamics. *The Journal of Chemical Physics* **151,** 124103 (2019).

204. Braams, B. J. & Manolopoulos, D. E. On the short-time limit of ring polymer molecular dynamics. *The Journal of Chemical Physics* **125,** 124105 (2006).

205. Korol, R., Rosa-Raíces, J. L., Bou-Rabee, N. & Miller III, T. F. Dimension-free path-integral molecular dynamics without preconditioning. *The Journal of Chemical Physics* **152,** 104102 (2020).

206. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126,** 014101 (2007).

207. Leimkuhler, B. & Matthews, C. Rational construction of stochastic numerical methods for molecular sampling. *Applied Mathematics Research eXpress* **2013,** 34–56 (2013).

208. Bou-Rabee, N. Time integrators for molecular dynamics. *Entropy* **16,** 138–162 (2014).

209. Bou-Rabee, N. Cayley splitting for second-order Langevin stochastic partial differential equations. *Preprint.* arXiv: 1707.05603 (2017).

210. Metafune, G., Pallara, D. & Priola, E. Spectrum of Ornstein-Uhlenbeck operators in L^p spaces with respect to invariant measures. *Journal of Functional Analysis* **196,** 40–60 (2002).

211. Pavliotis, G. A. *Stochastic Processes and Applications* (Springer, 2014).

212. Sanz-Serna, J. M. & Calvo, M. P. *Numerical Hamiltonian problems* 1st (Chapman & Hall, 1994).

213. Gibbs, A. L. & Su, F. E. On choosing and bounding probability metrics. *International Statistical Review* **70,** 419–435 (2002).

214. Villani, C. *Optimal transport: Old and new* (Springer Science & Business Media, 2008).

215. Givens, C. R. & Shortt, R. M. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal* **31,** 231–240 (1984).

216. Habershon, S., Markland, T. E. & Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *The Journal of Chemical Physics* **131,** 024501 (2009).

217. Geyer, C. J. Practical Markov chain Monte Carlo. *Statistical Science* **7,** 473–483 (1992).

218. Sokal, A. in *Functional integration: Basics and applications* (eds DeWitt-Morette, C., Cartier, P. & Folacci, A.) 131–192 (Springer Science & Business Media, Boston, MA, 1997).

219. Asmussen, S. & Glynn, P. W. *Stochastic simulation: Algorithms and analysis* 1st (Springer, 2007).

220. Skeel, R. D. & Fang, Y. Comparing Markov chain samplers for molecular simulation. *Entropy* **19,** 561 (2017).

221. Fang, Y., Cao, Y. & Skeel, R. D. Quasi-reliable estimates of effective sample size. *preprint.* arXiv: 1705.03831 (2017).

222. Suleimanov, Y. V., Allen, J. W. & Green, W. H. RPMDrate: Bimolecular chemical reaction rates from ring polymer molecular dynamics. *Computer Physics Communications* **184,** 833–840 (2013).

223. Kapil, V., Rossi, M., Marsalek, O., Petraglia, R., Litman, Y., Spura, T., Cheng, B., Cuzzocrea, A., Meißner, R. H., Wilkins, D. M., Helfrecht, B. A., Juda, P., Bienvenue, S. P., Fang, W., Kessler, J., Poltavsky, I., Vandenbrande, S., Wieme, J., Corminboeuf, C., Kühne, T. D., Manolopoulos, D. E., Markland, T. E., Richardson, J. O., Tkatchenko, A., Tribello, G. A., Van Speybroeck, V. & Ceriotti, M. i-PI 2.0: A universal force engine for advanced molecular simulations. *Computer Physics Communications* **236,** 214–223 (2019).

224. Sze, S. M. & Ng, K. K. *Physics of semiconductor devices* 3rd ed. (John wiley & sons, 2007).

225. Lundstrom, M. *Fundamentals of Carrier Transport* (Cambridge University Press, Cambridge, 2000).

226. Hartnagel, H., Katilius, R. & Matulionis, A. *Microwave Noise in Semiconductor Devices* chap. 8 (John Wiley & Sons, New York, 2001).

227. Jacoboni, C. & Reggiani, L. The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Rev. Mod. Phys.* **55,** 645–705. https://link.aps.org/doi/10.1103/RevModPhys.55.645 (3 July 1983).

228. *Monte Carlo Device Simulation: Full Band and Beyond* (ed Hess, K.) ISBN: 978-1-4615-4026-7 (Springer Science + Business Media, LLC, 1991).

229. Fischetti, M. Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. I. Homogeneous transport. *IEEE Transactions on Electron Devices* **38,** 634–649 (1991).

230. Fischetti, M. & Laux, S. Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. II. Submicrometer MOSFET's. *IEEE transactions on electron devices* **38,** 650–660. ISSN: 00189383. http://ieeexplore.ieee.org/document/75177/ (2023) (Mar. 1991).

231. Bernardi, M. First-principles dynamics of electrons and phonons. *European Physical Journal B* **89** (2016).

232. Giustino, F. Electron-phonon interactions from first principles. *Rev. Mod. Phys.* **89,** 015003 (1 Feb. 2017).

233. Poncé, S., Li, W., Reichardt, S. & Giustino, F. First-principles calculations of charge carrier mobility and conductivity in bulk semiconductors and two-dimensional materials. *Reports on Progress in Physics* **83,** 036501. https://doi.org/10.1088/1361-6633/ab6a43 (Feb. 2020).

234. Restrepo, O. D., Varga, K. & Pantelides, S. T. First-principles calculations of electron mobilities in silicon: Phonon and Coulomb scattering. *Appl. Phys. Lett.* **94,** 212103. https://doi.org/10.1063/1.3147189 (2009).

235. Poncé, S., Margine, E. R. & Giustino, F. Towards predictive many-body calculations of phonon-limited carrier mobilities in semiconductors. *Phys. Rev. B* **97,** 121201. https://link.aps.org/doi/10.1103/PhysRevB.97.121201 (12 Mar. 2018).

236. Poncé, S., Jena, D. & Giustino, F. Route to High Hole Mobility in GaN via Reversal of Crystal-Field Splitting. *Phys. Rev. Lett.* **123,** 096602. https://link.aps.org/doi/10.1103/PhysRevLett.123.096602 (9 Aug. 2019).

237. Zhou, J.-J. & Bernardi, M. Ab initio electron mobility and polar phonon scattering in GaAs. *Phys. Rev. B* **94,** 201201. https://link.aps.org/doi/10.1103/PhysRevB.94.201201 (20 Nov. 2016).

238. Borysenko, K. M., Mullen, J. T., Barry, E. A., Paul, S., Semenov, Y. G., Zavada, J. M., Nardelli, M. B. & Kim, K. W. First-principles analysis of electron-phonon interactions in graphene. *Phys. Rev. B* **81,** 121412. https://link.aps.org/doi/10.1103/PhysRevB.81.121412 (12 Mar. 2010).

239. Kaasbjerg, K., Thygesen, K. S. & Jacobsen, K. W. Phonon-limited mobility in n-type single-layer MoS_2 from first principles. *Phys. Rev. B* **85,** 115317. https://link.aps.org/doi/10.1103/PhysRevB.85.115317 (11 Mar. 2012).

240. Cheng, L., Zhang, C. & Liu, Y. Why Two-Dimensional Semiconductors Generally Have Low Electron Mobility. *Phys. Rev. Lett.* **125,** 177701. https://link.aps.org/doi/10.1103/PhysRevLett.125.177701 (17 Oct. 2020).

241. Lee, N.-E., Zhou, J.-J., Chen, H.-Y. & Bernardi, M. Ab initio electron-two-phonon scattering in GaAs from next-to-leading order perturbation theory. *Nat. Comm.* **11,** 1607. https://www.nature.com/articles/s41467-020-15339-0 (1 Mar. 2020).

242. Brunin, G., Miranda, H. P. C., Giantomassi, M., Royo, M., Stengel, M., Verstraete, M. J., Gonze, X., Rignanese, G.-M. & Hautier, G. Electron-Phonon beyond Fröhlich: Dynamical Quadrupoles in Polar and Covalent Solids. *Phys. Rev. Lett.* **125,** 136601. https://link.aps.org/doi/10.1103/PhysRevLett.125.136601 (13 Sept. 2020).

243. Park, J., Zhou, J.-J., Jhalani, V. A., Dreyer, C. E. & Bernardi, M. Long-range quadrupole electron-phonon interaction from first principles. *Phys. Rev. B* **102,** 125203. https://link.aps.org/doi/10.1103/PhysRevB.102.125203 (12 Sept. 2020).

244. Li, Z., Antonius, G., Wu, M., da Jornada, F. H. & Louie, S. G. Electron-Phonon Coupling from Ab Initio Linear-Response Theory within the *GW* Method: Correlation-Enhanced Interactions and Superconductivity in $Ba_{1-x}K_xBiO_3$. *Phys. Rev. Lett.* **122,** 186402. https://link.aps.org/doi/10.1103/PhysRevLett.122.186402 (18 May 2019).

245. Hatanpää, B., Choi, A. Y., Cheng, P. S. & Minnich, A. J. Two-phonon scattering in non-polar semiconductors: a first-principles study of warm electron transport in Si. *arXiv preprint arXiv:2207.11376* (2022).

246. Choi, A. Y., Cheng, P. S., Hatanpää, B. & Minnich, A. J. Electronic noise of warm electrons in semiconductors from first principles. *Phys. Rev. Materials* **5,** 044603. https://doi.org/10.1103/PhysRevMaterials.5.044603 (4 Apr. 2021).

247. Cheng, P. S., Sun, S.-N., Choi, A. Y. & Minnich, A. J. High-field transport and hot electron noise in GaAs from first principles: role of two-phonon scattering. *arXiv preprint arXiv:2201.11912* (2022).

248. Catherall, D. S. & Minnich, A. J. High-field charge transport and noise in p-Si from first principles. *Physical Review B* **107,** 035201 (2023).

249. Hamaguchi, C. & Hamaguchi, C. *Basic semiconductor physics* (Springer, 2010).

250. Rustagi, A. & Stanton, C. J. Hot-electron noise properties of graphene-like systems. *Phys. Rev. B* **90,** 245424. https://link.aps.org/doi/10.1103/PhysRevB.90.245424 (24 Dec. 2014).

251. Ferry, D. K., Barker, J. R. & Jacobini, C. *Physics of nonlinear transport in semiconductors* (Springer Science & Business Media, 2012).

252. Paranjape, V. & Paranjape, B. Phonon excitation by electric field in semiconductors. *Physical Review* **166,** 757 (1968).

253. Verdi, C. & Giustino, F. Fröhlich electron-phonon vertex from first principles. *Physical review letters* **115,** 176401 (2015).

254. Kane, E. O. Band structure of indium antimonide. *Journal of Physics and Chemistry of Solids* **1,** 249–261 (1957).

255. Fröhlich, H. Electrons in lattice fields. *Advances in Physics* **3,** 325–361 (1954).

256. Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., Corso, A. D., de Gironcoli, S., Fabris, S., Fratesi, G., Gebauer, R., Gerstmann, U., Gougoussis, C., Kokalj, A., Lazzeri, M., Martin-Samos, L., Marzari, N., Mauri, F., Mazzarello, R., Paolini, S., Pasquarello, A., Paulatto, L., Sbraccia, C., Scandolo, S., Sclauzero, G., Seitsonen, A. P., Smogunov, A., Umari, P. & Wentzcovitch, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter* **21,** 395502 (Sept. 2009).

257. Giannozzi, P., Andreussi, O., Brumme, T., Bunau, O., Nardelli, M. B., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Cococcioni, M., Colonna, N., Carnimeo, I., Corso, A. D., de Gironcoli, S., Delugas, P., DiStasio, R. A., Ferretti, A., Floris, A., Fratesi, G., Fugallo, G., Gebauer, R., Gerstmann, U.,

Giustino, F., Gorni, T., Jia, J., Kawamura, M., Ko, H.-Y., Kokalj, A., Küçükbenli, E., Lazzeri, M., Marsili, M., Marzari, N., Mauri, F., Nguyen, N. L., Nguyen, H.-V., Otero-de-la-Roza, A., Paulatto, L., Poncé, S., Rocca, D., Sabatini, R., Santra, B., Schlipf, M., Seitsonen, A. P., Smogunov, A., Timrov, I., Thonhauser, T., Umari, P., Vast, N., Wu, X. & Baroni, S. Advanced capabilities for materials modelling with Quantum ESPRESSO. *Journal of Physics: Condensed Matter* **29,** 465901 (Oct. 2017).

258. Conwell, E. M. & Vassell, M. O. High-Field Transport in- Type GaAs. *Phys. Rev.* **166,** 797–821. ISSN: 0031-899X. https://link.aps.org/doi/10.1103/PhysRev.166.797 (Feb. 1968).

259. Zhou, J.-J., Park, J., Lu, I.-T., Maliyov, I., Tong, X. & Bernardi, M. Perturbo: A software package for ab initio electron–phonon interactions, charge transport and ultrafast dynamics. *Computer Physics Communications* **264,** 107970. ISSN: 00104655. https://linkinghub.elsevier.com/retrieve/pii/S0010465521000837 (2021) (July 2021).

260. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. & SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 261–272 (2020).

261. Ruch, J. G. & Kino, G. S. MEASUREMENT OF THE VELOCITY-FIELD CHARACTERISTIC OF GALLIUM ARSENIDE. *Appl. Phys. Lett.* **10,** 40–42. https://doi.org/10.1063/1.1754837 (1967).

262. Ashida, K., Inoue, M., Shirafuji, J. & Inuishi, Y. Energy Relaxation Effect of Hot Electrons in GaAs. *Journal of the Physical Society of Japan* **37,** 408–414. https://doi.org/10.1143/JPSJ.37.408 (1974).

263. Gasquet, D., de Murcia, M., Nougier, J. & Gontrand, C. Transport parameters of hot electrons in GaAs at 300 K. *Physica B+C* **134,** 264–268. ISSN: 0378-4363. https://www.sciencedirect.com/science/article/pii/0378436385903535 (1985).

264. Bareikis, V., Viktoravicius, V., Galdikas, A. & Miliusyte, R. Microwave Noise and Constant of the Coupling Between the Valleys Gamma and L in a Three-Valley Model of GaAs. *Sov. Phys. Semicond.* **14,** 847–849 (1980).

265. Ruch, J. G. & Kino, G. S. Transport Properties of GaAs. *Phys. Rev.* **174,** 921–931. https://link.aps.org/doi/10.1103/PhysRev.174.921 (3 Oct. 1968).

266. Fasol, G., Hackenberg, W., Hughes, H. P., Ploog, K., Bauser, E. & Kano, H. Continuous-wave spectroscopy of femtosecond carrier scattering in GaAs. *Phys. Rev. B* **41,** 1461–1478. https://link.aps.org/doi/10.1103/PhysRevB.41.1461 (3 Jan. 1990).

267. Požela, J. & Reklaitis, A. Electron transport properties in GaAs at high electric fields. *Solid-State Electronics* **23,** 927–933 (1980).

268. Kocevar, P. in *Physics of nonlinear transport in semiconductors* (eds Ferry, D. K., Barker, J. R. & Jacoboni, C.) 167–174 (Springer US, Boston, MA, 1980). ISBN: 978-1-4684-3638-9. http://link.springer.com/10.1007/978-1-4684-3638-9%5C_7 (2021).

269. Gunn, J. B. Instabilities of Current in III–V Semiconductors. *IBM Journal of Research and Development* **8,** 141–159 (1964).

270. McCumber, D. & Chynoweth, A. Theory of negative-conductance amplification and of Gunn instabilities in "two-valley" semiconductors. *IEEE transactions on electron devices* **ED-13,** 4–21. ISSN: 0018-9383. http://ieeexplore.ieee.org/document/1474219/ (2021) (Jan. 1966).

271. Skinner, B., Ruhman, J. & Nahum, A. Measurement-Induced Phase Transitions in the Dynamics of Entanglement. *Phys. Rev. X* **9,** 031009 (July 2019).

272. Chan, A., Nandkishore, R. M., Pretko, M. & Smith, G. Unitary-projective entanglement dynamics. *Physical Review B* **99,** 224307 (June 2019).

273. Li, Y., Chen, X. & Fisher, M. P. A. Quantum Zeno effect and the many-body entanglement transition. *Physical Review B* **98,** 205136 (Nov. 2018).

274. Nandkishore, R. & Huse, D. A. Many-Body Localization and Thermalization in Quantum Statistical Mechanics. *Annual Review of Condensed Matter Physics* **6,** 15–38. arXiv: 1404.0686 [cond-mat.stat-mech] (Mar. 2015).

275. Breuer, H.-P. & Petruccione, F. *The theory of open quantum systems* (Oxford University Press, 2002).

276. Calabrese, P. & Cardy, J. Evolution of entanglement entropy in one-dimensional systems. *Journal of Statistical Mechanics: Theory and Experiment* **4,** 04010. eprint: cond-mat/0503393 (Apr. 2005).

277. Calabrese, P. & Cardy, J. Quantum quenches in extended systems. *Journal of Statistical Mechanics: Theory and Experiment* **6,** 06008. arXiv: 0704.1880 [cond-mat.stat-mech] (June 2007).

278. Kim, H. & Huse, D. A. Ballistic Spreading of Entanglement in a Diffusive Nonintegrable System. *Physical Review Letters* **111,** 127205. arXiv: 1306.4306 [quant-ph] (Sept. 2013).

279. Nahum, A., Vijay, S. & Haah, J. Operator Spreading in Random Unitary Circuits. *Phys. Rev. X* **8,** 021014 (Apr. 2018).

280. von Keyserlingk, C. W., Rakovszky, T., Pollmann, F. & Sondhi, S. L. Operator Hydrodynamics, OTOCs, and Entanglement Growth in Systems without Conservation Laws. *Phys. Rev. X* **8,** 021013. arXiv: 1705.08910 [cond-mat.str-el] (Apr. 2018).

281. Misra, B. & Sudarshan, E. G. The Zeno's paradox in quantum theory. *Journal of Mathematical Physics* **18,** 756–763. https://doi.org/10.1063/1.523304 (1977).

282. Cao, X., Tilloy, A. & De Luca, A. Entanglement in a fermion chain under continuous monitoring. *SciPost Physics* **7,** 024 (Aug. 2019).

283. Li, Y., Chen, X. & Fisher, M. P. A. Measurement-driven entanglement transition in hybrid quantum circuits. *Physical Review B* **100,** 134306 (Oct. 2019).

284. Nahum, A. & Skinner, B. Entanglement and dynamics of diffusion-annihilation processes with Majorana defects. *Phys. Rev. Research* **2,** 023288 (June 2020).

285. Lopez-Piqueres, J., Ware, B. & Vasseur, R. Mean-field entanglement transitions in random tree tensor networks. *Phys. Rev. B* **102,** 064202. https://link.aps.org/doi/10.1103/PhysRevB.102.064202 (6 Aug. 2020).

286. Lavasani, A., Alavirad, Y. & Barkeshli, M. Measurement-induced topological entanglement transitions in symmetric random quantum circuits. *Nature Physics* **17,** 342–347 (Jan. 2021).

287. Sang, S. & Hsieh, T. H. Measurement-protected quantum phases. *Phys. Rev. Research* **3,** 023200 (June 2021).

288. Ippoliti, M., Gullans, M. J., Gopalakrishnan, S., Huse, D. A. & Khemani, V. Entanglement Phase Transitions in Measurement-Only Dynamics. *Phys. Rev. X* **11,** 011030 (Jan. 2021).

289. Chen, X., Li, Y., Fisher, M. P. A. & Lucas, A. Emergent conformal symmetry in nonunitary random dynamics of free fermions. *Phys. Rev. Research* **2,** 033017 (July 2020).

290. Fuji, Y. & Ashida, Y. Measurement-induced quantum criticality under continuous monitoring. *Phys. Rev. B* **102,** 054302. https://link.aps.org/doi/10.1103/PhysRevB.102.054302 (5 Aug. 2020).

291. Alberton, O., Buchhold, M. & Diehl, S. Entanglement Transition in a Monitored Free-Fermion Chain: From Extended Criticality to Area Law. *Physical Review Letters* **126,** 170602 (Apr. 2021).

292. Vijay, S. *Measurement-Driven Phase Transition within a Volume-Law Entangled Phase* 2020. arXiv: 2005.03052 [quant-ph].

293. Nahum, A., Roy, S., Skinner, B. & Ruhman, J. Measurement and Entanglement Phase Transitions in All-To-All Quantum Circuits, on Quantum Trees, and in Landau-Ginsburg Theory. *PRX Quantum* **2,** 010352. https://link.aps.org/doi/10.1103/PRXQuantum.2.010352 (1 Mar. 2021).

294. Bao, Y., Choi, S. & Altman, E. Symmetry enriched phases of quantum circuits. *Annals of Physics* **435,** 168618. ISSN: 0003-4916. http://dx.doi.org/10.1016/j.aop.2021.168618 (Dec. 2021).

295. Agrawal, U., Zabalo, A., Chen, K., Wilson, J. H., Potter, A. C., Pixley, J. H., Gopalakrishnan, S. & Vasseur, R. Entanglement and Charge-Sharpening Transitions in U(1) Symmetric Monitored Quantum Circuits. *Phys. Rev. X* **12,** 041002. https://link.aps.org/doi/10.1103/PhysRevX.12.041002 (4 Oct. 2022).

296. Barratt, F., Agrawal, U., Gopalakrishnan, S., Huse, D. A., Vasseur, R. & Potter, A. C. Field Theory of Charge Sharpening in Symmetric Monitored Quantum Circuits. *Phys. Rev. Lett.* **129,** 120604. https://link.aps.org/doi/10.1103/PhysRevLett.129.120604 (12 Sept. 2022).

297. Agrawal, U., Lopez-Piqueres, J., Vasseur, R., Gopalakrishnan, S. & Potter, A. C. *Observing quantum measurement collapse as a learnability phase transition* 2023. arXiv: 2311.00058 [quant-ph].

298. Fisher, M. P. A., Khemani, V., Nahum, A. & Vijay, S. Random Quantum Circuits. *Annual Review of Condensed Matter Physics* **14,** 335–379. ISSN: 1947-5462. http://dx.doi.org/10.1146/annurev-conmatphys-031720-030658 (Mar. 2023).

299. Choi, S., Bao, Y., Qi, X.-L. & Altman, E. Quantum Error Correction in Scrambling Dynamics and Measurement-Induced Phase Transition. *Physical Review Letters* **125,** 030505 (July 2020).

300. Bao, Y., Choi, S. & Altman, E. Theory of the phase transition in random unitary circuits with measurements. *Physical Review B* **101,** 104301 (Mar. 2020).

301. Gullans, M. J. & Huse, D. A. Dynamical Purification Phase Transition Induced by Quantum Measurements. *Phys. Rev. X* **10,** 041020 (Oct. 2020).

302. Gullans, M. J., Krastanov, S., Huse, D. A., Jiang, L. & Flammia, S. T. Quantum coding with low-depth random circuits. *arXiv e-prints,* arXiv:2010.09775. arXiv: 2010.09775 [quant-ph] (Oct. 2020).

303. Napp, J., La Placa, R. L., Dalzell, A. M., Brandao, F. G. S. L. & Harrow, A. W. Efficient classical simulation of random shallow 2D quantum circuits. *arXiv e-prints,* arXiv:2001.00021. arXiv: 2001.00021 [quant-ph] (Dec. 2019).

304. Koh, J. M., Sun, S.-N., Motta, M. & Minnich, A. J. Measurement-induced entanglement phase transition on a superconducting quantum processor with mid-circuit readout. *Nature Physics* **19,** 1314–1319. https://doi.org/10.1038/s41567-023-02076-6 (2023).

305. Ippoliti, M. & Khemani, V. Postselection-free entanglement dynamics via spacetime duality. *Physical Review Letters* **126,** 060501. https://doi.org/10.1103/PhysRevLett.126.060501 (2021).

306. Lu, T.-C. & Grover, T. Spacetime duality between localization transitions and measurement-induced transitions. *PRX Quantum* **2,** 040319. https://doi.org/10.1103/PRXQuantum.2.040319 (2021).

307. Hoke, J. C. *et al.* Measurement-induced entanglement and teleportation on a noisy quantum processor. *Nature* **622,** 481–486. https://doi.org/10.1038/s41586-023-06505-7 (2023).

308. Gullans, M. J. & Huse, D. A. Scalable probes of measurement-induced criticality. *Physical review letters* **125,** 070606. https://doi.org/10.1103/PhysRevLett.125.070606 (2020).

309. Noel, C., Niroula, P., Zhu, D., Risinger, A., Egan, L., Biswas, D., Cetina, M., Gorshkov, A. V., Gullans, M. J., Huse, D. A. & Monroe, C. Measurement-induced quantum phases realized in a trapped-ion quantum computer. *Nature Physics* **18,** 760–764. ISSN: 1745-2481. http://dx.doi.org/10.1038/s41567-022-01619-7 (June 2022).

310. Li, Y., Zou, Y., Glorioso, P., Altman, E. & Fisher, M. P. A. Cross entropy benchmark for measurement-induced phase transitions. *Physical Review Letters* **130,** 220404. https://doi.org/10.1103/PhysRevLett.130.220404 (2023).

311. Nahum, A., Ruhman, J., Vijay, S. & Haah, J. Quantum Entanglement Growth under Random Unitary Dynamics. *Phys. Rev. X* **7,** 031016. arXiv: 1608.06950 [cond-mat.stat-mech] (July 2017).

312. Boixo, S., Isakov, S. V., Smelyanskiy, V. N., Babbush, R., Ding, N., Jiang, Z., Bremner, M. J., Martinis, J. M. & Neven, H. Characterizing quantum supremacy in near-term devices. *Nature Physics* **14,** 595–600. ISSN: 1745-2481. http://dx.doi.org/10.1038/s41567-018-0124-x (Apr. 2018).

313. Arute, F. *et al.* Quantum supremacy using a programmable superconducting processor. *Nature* **574,** 505–510 (Oct. 2019).

314. Dalzell, A. M., Hunter-Jones, N. & Brandao, F. G. S. L. *Random quantum circuits transform local noise into global white noise* 2021. arXiv: 2111.14907 [quant-ph].

315. Gao, X., Kalinowski, M., Chou, C.-N., Lukin, M. D., Barak, B. & Choi, S. *Limitations of Linear Cross-Entropy as a Measure for Quantum Advantage* 2021. arXiv: 2112.01657 [quant-ph].

316. Choi, J., Shaw, A. L., Madjarov, I. S., Xie, X., Finkelstein, R., Covey, J. P., Cotler, J. S., Mark, D. K., Huang, H.-Y., Kale, A., *et al.* Preparing random states and benchmarking with many-body quantum chaos. *Nature* **613,** 468–473. http://dx.doi.org/10.1038/s41586-022-05442-1 (2023).

317. Mark, D. K., Choi, J., Shaw, A. L., Endres, M. & Choi, S. Benchmarking Quantum Simulators Using Ergodic Quantum Dynamics. *Phys. Rev. Lett.* **131,** 110601. https://link.aps.org/doi/10.1103/PhysRevLett.131.110601 (11 Sept. 2023).

318. Garratt, S. J., Weinstein, Z. & Altman, E. Measurements Conspire Nonlocally to Restructure Critical Quantum States. *Phys. Rev. X* **13,** 021026. https://link.aps.org/doi/10.1103/PhysRevX.13.021026 (2 May 2023).

319. Feng, X., Skinner, B. & Nahum, A. *Measurement-induced phase transitions on dynamical quantum trees* 2022. arXiv: 2210.07264 [cond-mat.stat-mech].

320. Jozsa, R. & Van Den Nest, M. Classical simulation complexity of extended clifford circuits. *Quantum Info. Comput.* **14,** 633–648. ISSN: 1533-7146 (May 2014).

321. Koh, D. E. *Further extensions of Clifford circuits and their classical simulation complexities* 2015. arXiv: 1512.07892 [quant-ph].

322. Bouland, A., Fitzsimons, J. F. & Koh, D. E. *Complexity Classification of Conjugated Clifford Circuits* 2017. arXiv: 1709.01805 [quant-ph].

323. Bravyi, S. & Gosset, D. Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates. *Physical Review Letters* **116,** 250501 (June 2016).

324. Yoganathan, M., Jozsa, R. & Strelchuk, S. Quantum advantage of unitary Clifford circuits with magic state inputs. *Proceedings of the Royal Society A* **475,** 20180427. https://doi.org/10.1098/rspa.2018.0427 (2019).

325. Stanley, H. E. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Reviews of modern physics* **71,** S358. https://doi.org/10.1103/RevModPhys.71.S358 (1999).

326. Bhattacharjee, S. M. & Seno, F. A measure of data collapse for scaling. *Journal of Physics A: Mathematical and General* **34,** 6375. https://dx.doi.org/10.1088/0305-4470/34/33/302 (2001).

327. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. & SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 261–272 (2020).

328. Nahum, A., Roy, S., Skinner, B. & Ruhman, J. Measurement and entanglement phase transitions in all-to-all quantum circuits, on quantum trees, and in Landau-Ginsburg theory. *PRX Quantum* **2,** 010352. https://doi.org/10.1103/PRXQuantum.2.010352 (2021).

329. Sang, S. & Hsieh, T. H. Measurement-protected quantum phases. *Physical Review Research* **3,** 023200. https://doi.org/10.1103/PhysRevResearch.3.023200 (2021).

330. Lavasani, A., Alavirad, Y. & Barkeshli, M. Measurement-induced topological entanglement transitions in symmetric random quantum circuits. *Nature Physics* **17,** 342–347. https://doi.org/10.1038/s41567-020-01112-z (2021).

331. Tikhanovskaya, M., Lavasani, A., Fisher, M. P. A. & Vijay, S. *Universality of the cross entropy in $\mathbb{Z}_2$ symmetric monitored quantum circuits* 2023. arXiv: 2306.00058 [quant-ph].

332. Lovas, I., Agrawal, U. & Vijay, S. *Quantum Coding Transitions in the Presence of Boundary Dissipation* 2023. arXiv: 2304.02664 [quant-ph].

333. Abramowitz, M. & Stegun, I. A. *Handbook of mathematical functions, with formulas, graphs, and mathematical tables* 1st (Dover Publications, Inc., USA, 1965).

334. Andreescu, T. *Essential linear algebra with applications* 1st (Birkhäuser Basel, 2016).

335. Bou-Rabee, N. & Sanz-Serna, J. M. Randomized Hamiltonian Monte Carlo. *Annals of Applied Probability* **27,** 2159–2194 (2017).

336. Priestly, M. B. *Spectral analysis and time series* 1st (Academic Press, UK, 1981).

337. Madras, N. & Sokal, A. The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *Journal of Statistical Physics* **50,** 109–186 (1988).

338. Jian, C.-M., You, Y.-Z., Vasseur, R. & Ludwig, A. W. W. Measurement-induced criticality in random quantum circuits. *Physical Review B* **101,** 104302 (Mar. 2020).

339. Zhou, T. & Nahum, A. Emergent statistical mechanics of entanglement in random unitary circuits. *Physical Review B* **99,** 174205 (May 2019).

340. Zhou, T. & Nahum, A. Entanglement Membrane in Chaotic Many-Body Systems. *Phys. Rev. X* **10,** 031066. https://link.aps.org/doi/10.1103/PhysRevX.10.031066 (3 Sept. 2020).

341. Bouland, A., Fefferman, B., Landau, Z. & Liu, Y. *Noise and the Frontier of Quantum Supremacy* in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Feb. 2022). http://dx.doi.org/10.1109/focs52979.2021.00127.

342. Ware, B., Deshpande, A., Hangleiter, D., Niroula, P., Fefferman, B., Gorshkov, A. V. & Gullans, M. J. *A sharp phase transition in linear cross-entropy benchmarking* 2023. arXiv: 2305.04954 [quant-ph].

343. A. Morvan *et. al. Phase transition in Random Circuit Sampling* 2023. arXiv: 2304.11119 [quant-ph].

344. Viola, L. & Lloyd, S. Dynamical suppression of decoherence in two-state quantum systems. **58,** 2733 (1998).

345. Kofman, A. & Kurizki, G. Universal dynamical control of quantum mechanical decay: modulation of the coupling to the continuum. **87,** 270405 (2001).

346. Biercuk, M. J., Uys, H., VanDevender, A. P., Shiga, N., Itano, W. M. & Bollinger, J. J. Optimized dynamical decoupling in a model quantum memory. *Nature* **458,** 996–1000 (2009).

347. Rost, B., Jones, B., Vyushkova, M., Ali, A., Cullip, C., Vyushkov, A. & Nabrzyski, J. Simulation of thermal relaxation in spin chemistry systems on a quantum computer using inherent qubit decoherence. *arXiv:2001.00794* (2020).

348. Niu, S. & Todri-Sanial, A. Effects of dynamical decoupling and pulse-level optimizations on ibm quantum computers. *IEEE Trans. Quantum Eng.* **3,** 1–10 (2022).

349. Niu, S. & Todri-Sanial, A. Analyzing strategies for dynamical decoupling insertion on IBM quantum computer. *arXiv:2204.14251* (2022).

350. Ezzell, N., Pokharel, B., Tewala, L., Quiroz, G. & Lidar, D. A. Dynamical decoupling for superconducting qubits: a performance survey. *arXiv:2207.03670* (2022).

351. Bravyi, S., Sheldon, S., Kandala, A., Mckay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Physical Review A* **103,** 042605. https://doi.org/10.1103/PhysRevA.103.042605 (2021).

352. Nation, P. D., Kang, H., Sundaresan, N. & Gambetta, J. M. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* **2,** 040326. https://doi.org/10.1103/PRXQuantum.2.040326 (2021).