Variational Consistent Histories: A Hybrid Algorithm for Quantum Foundations

Andrew Arrasmith,1, 2 Lukasz Cincio,1 Andrew T. Sornborger,2 Wojciech H. Zurek,1 and Patrick J. Coles1

1Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
2Department of Physics, University of California Davis, Davis, CA 95616, USA.

While quantum computers are predicted to have many commercial applications, less attention has been given to their potential for resolving foundational issues in quantum mechanics. Here we focus on quantum computers’ utility for the Consistent Histories formalism, which has previously been employed to study quantum cosmology, quantum paradoxes, and the quantum-to-classical transition. We present a variational hybrid quantum-classical algorithm for finding consistent histories, which should revitalize interest in this formalism by allowing classically impossible calculations to be performed. In our algorithm, the quantum computer evaluates the decoherence functional (with exponential speedup in both the number of qubits and the number of times in the history), and a classical optimizer adjusts the history parameters to improve consistency. We implement our algorithm on a cloud quantum computer to find consistent histories for a spin in a magnetic field, and on a simulator to observe the emergence of classicality for a chiral molecule.

The foundations of quantum mechanics (QM) have been debated for the past century [1, 2], including topics such as the EPR paradox, hidden-variable theories, Bell’s Theorem, Born’s rule, and the role of measurements in QM. This also includes the quantum-to-classical transition, i.e., the emergence of classical behavior (objectivity, irreversibility, lack of interference, etc.) from quantum laws [3–5].

The Consistent Histories (CH) formalism was introduced by Griffiths, Omnès, Gell-Mann, and Hartle to address some (though not all) of the aforementioned issues [6–8]. One inventor considered CH to be “the Copenhagen interpretation done right” [6], as it resolves some of the paradoxes of quantum mechanics by enforcing strict rules for logical reasoning with quantum systems. In this formalism, the Copenhagen interpretation’s focus on measurements as the origin of probabilities is replaced by probabilities for sequences of events (histories) to occur, and hence by avoiding measurements it avoids the measurement problem. The sets of histories whose probabilities are additive (as the histories do not interfere with each other) are considered to be consistent and are thus the only ones able to be reasoned about in terms of classical probability and logic [7].

Regardless of one’s opinion of the philosophical interpretation (on which this paper is agnostic), this computational framework has proven useful in applications such as attempting to solve the cosmological measure problem [9, 10], understanding quantum jumps [11], and evaluating the arrival time for particles at a detector [12–14]. One of the main reasons that this framework has not received more attention and use is that carrying out the calculations for non-trivial cases can be difficult [11]. While numerical approaches have been attempted [15, 16], they require exponentially scaling resources as either the number of times considered or the system size grows. This makes classical numerical approaches unusable for any but the simplest cases.

Here we present a scalable algorithm for the CH formalism that achieves an exponential speedup over classical methods both in terms of the system size and the number of times considered. It will allow exploration beyond toy models, such as the quantum-to-classical transition in mesoscopic quantum systems. We expect this to revitalize interest in the CH approach to quantum mechanics by increasing its practical utility.

Our algorithm is a variational hybrid quantum-classical algorithm (VHQCA). With the impending arrival of the first noisy intermediate-scale quantum computers [17], the field of VHQCAs, which make the most of short quantum circuits combined with classical optimizers, has been taking off. VHQCAs have now been demonstrated for a myriad of tasks ranging from factoring to finding ground states, among others [18–25]. The VHQCA framework potentially brings the practical applications of quantum computers years closer to fruition. Hence, useful implementations of our algorithm will be feasible on near-term quantum devices.

Below we introduce our algorithm, present experimental results from implementing our method on IBM’s superconducting qubit quantum processor as well as on a simulator, and then discuss future applications.

**CONSISTENT HISTORIES BACKGROUND**

In the CH framework [26–28], a history \( \mathcal{Y}^\alpha \) is a sequence of properties (i.e., projectors onto the appropriate subspaces) at a succession of times \( t_1 < t_2 < \ldots < t_k \),

\[
\mathcal{Y}^\alpha = (P_1^\alpha, P_2^\alpha, \ldots, P_k^\alpha),
\]

where \( P_j^\alpha \) is chosen from a set \( P_j \) of projectors that sum to the identity at time \( t_j \). For example, for a photon passing through a sequence of diffraction gratings and then striking a screen, a history could be the photon passed through one slit in the first grating, another slit in the second, and so on. Clearly, we find interference between such histories unless there is some sense in which the photon’s path has been recorded. Since there is interference, we cannot add the probabilities of the different
histories classically and expect to correctly predict where the photon strikes the screen.

The CH framework provides tools for determining when a family (i.e., a set that sums to the multi-time identity operator) of histories \( F = \{ Y^{\alpha} \} \) exhibits interference, which is not always obvious. In this framework, one defines the so-called class operator

\[
C^{\alpha} = P^{\alpha}_{k}(t)P^{\alpha-1}_{k-1}(t) \ldots P^{\alpha}_{1}(t),
\]

which is the time-ordered product of the projection operators (now in the Heisenberg picture and hence explicitly time dependent) in history \( Y^{\alpha} \). If the system is initially described by a density matrix \( \rho \), the degree of interference or overlap between histories \( Y^{\alpha} \) and \( Y^{\alpha'} \) is

\[
D(\alpha, \alpha') = \text{Tr} \left( C^{\alpha} \rho C^{\alpha'} \right).
\]

This quantity is called the decoherence functional. The consistency condition for a family of histories \( F \) is then

\[
D(\alpha, \alpha') = 0, \quad \forall \alpha \neq \alpha'.
\]

If and only if this condition holds do we say that \( D(\alpha, \alpha) \) is the probability for history \( Y^{\alpha} \). Since we are presenting a numerical algorithm, it will also be useful to consider approximate consistency, where we merely insist that the interference is small in the following sense:

\[
|D(\alpha, \alpha')|^2 \leq \epsilon^2 D(\alpha, \alpha)D(\alpha', \alpha'), \quad \forall \alpha \neq \alpha',
\]

which guarantees that probability sum rules for \( F \) are satisfied within an error of \( \epsilon \) [29].

To study consistency arising purely from decoherence (i.e., records in the environment), researchers have proposed a functional that instead takes a partial trace over \( E \), which is (a subsystem of) the environment [30, 31]:

\[
D_{pt}(\alpha, \alpha') = \text{Tr}_{E} \left( C^{\alpha} \rho C^{\alpha'} \right).
\]

With this modification, the consistency condition is

\[
D_{pt}(\alpha, \alpha') = 0, \quad \forall \alpha \neq \alpha',
\]

where 0 is the zero matrix. Instead of only signifying the lack of interference, partial-trace consistency singles out whether or not the records of the histories in the environment interfere. Note that the full-trace condition of Eq. (4) is satisfied when this partial-trace consistency is satisfied, but the converse does not hold [30].

With this formalism in hand, we can now see why classical numerical schemes for CH have faced difficulty. For example, consider histories of a collection of \( n \) spin-1/2 particles for \( k \) time steps, depicted in Fig. 1. The number of histories is \( 2^{nk} \); and hence there are \( \approx 2^{2nk} \) decoherence functional elements. Furthermore, evaluating each decoherence functional element \( D(\alpha, \alpha') \) requires the equivalent of a Hamiltonian simulation of the system, i.e., the multiplication of \( 2^n \times 2^n \) matrices. This means modern clusters would take centuries to evaluate the consistency of a family of histories with \( k = 2 \) time steps and \( n = 10 \) spins. Given this limitation, we can see why, for the most part, only toy models have been analyzed in this framework thus far.

**HYBRID ALGORITHM FOR FINDING CONSISTENT HISTORIES**

We refer to our VHQCA as Variational Consistent Histories (VCH), see Fig. 2. VCH takes as its inputs a physical model (i.e., an initial state \( \rho \) and a Hamiltonian \( H \)) and some ansatz for the types of projectors to consider. It outputs: (1) a family \( F \) of histories that is (approximately) full and/or partial trace consistent in the form of projection operators prepared on a quantum computer, (2) the probabilities of the most likely histories \( Y^{\alpha} \) in \( F \), and (3) a bound on the consistency parameter \( \epsilon \).

VCH involves a parameter optimization loop, where a quantum computer evaluates a cost function that quantifies the family’s inconsistency, while a classical optimizer adjusts the family (i.e., varies the projector parameters) to reduce the cost. Classical optimizers for VHQCA are actively being investigated [25, 32], and one is free to choose the classical optimizer on an empirical basis.

To compute the cost, note that the elements of the decoherence functional form a positive semi-definite matrix with trace one. In VCH, we exploit this property to encode \( \mathcal{D} \) in a quantum state \( \sigma^{A} \), whose matrix elements are \( \langle \alpha | \sigma^{A} | \alpha' \rangle = D(\alpha, \alpha') \). Step b of Fig. 2 shows a quantum circuit that prepares \( \sigma^{A} \). This circuit transforms an initial state \( \rho \otimes |0\rangle\langle 0| \) on systems SA, where S simulates the physical system of interest and A is an ancilla system, into a state \( \sigma^{SA} \) whose marginal is \( \sigma^{A} \). For the full trace consistency, we introduce a global measure of the (in)consistency that quantifies how far \( \sigma^{A} \) is from being...
prepare the projectors for any history in the probabilities of the most likely histories in time. The optimal parameters are then used to compute the denotations to multi-qubit systems, non-trivial environment depicted for a one-qubit system, the SM discusses the generations of observables (i.e., avoiding full state tomography) to learn important information about the histories.

where $D_{\text{HS}}$ is the Hilbert-Schmidt distance and $Z^A(\sigma^A)$ is the dephased (all off-diagonal elements set to zero) version of $\sigma^A$. This quantity goes to zero if and only if $\mathcal{F}$ is consistent. For the partial trace case, we arrive at a similar cost function but with $\sigma^A$ replaced by $\sigma^{SA}$:

$$C_{\text{pt}} := \sum_{\alpha \neq \alpha'} \| D_{\text{pt}}(\alpha, \alpha') \|^2_{\text{HS}} = D_{\text{HS}}(\sigma^{SA}, Z^A(\sigma^{SA})).$$

Here the notation $Z^A(\sigma^{SA})$ indicates that the dephasing operation only acts on system $A$, and the absolute squares of Eq. (8) have been generalized to Hilbert-Schmidt norms, $\| M \|^2_{\text{HS}} := \text{Tr}(M^\dagger M)$. In the Methods section, we present quantum circuits that compute these cost functions from two copies of $\sigma^A$ or $\sigma^{SA}$. Derivations of the second equalities in Eq. (8) and Eq. (9) can be found in the Supplementary Material (SM). We remark that alternative cost functions may be useful, for example, to penalize families $\mathcal{F}$ with high entropy (see Methods) or to obtain a larger cost gradient by employing local instead of global observables (see Ref. [25]).

The parameter optimization loop results in an approximately consistent family, $\mathcal{F}$, of histories, where the consistency parameter $\epsilon$ is upper bounded in terms of the final cost (see Methods). In Step c in Fig. 2, we then generate the probabilities for the most likely histories by repeatedly preparing $\sigma^A$ and measuring in the standard basis, where the measurement frequencies give the probabilities. Step d shows how one prepares the set of projection operators for any given history in $\mathcal{F}$. These projectors can then be characterized with an efficient number of observables (i.e., avoiding full state tomography) to learn important information about the histories.

Let us discuss the scaling of VCH. With the potential exceptions of the Hamiltonian evolution and the projection operators, the complexity of our quantum circuits (i.e., the gate count, circuit depth, and total number of required qubits) scales linearly with both the system size $n$ and the number of times $k$ considered. The complexity of Hamiltonian evolution to some accuracy is problem dependent, but we typically expect polynomial scaling in $n$ for physical systems with properties like translational symmetry [33]. On the other hand, we consider the circuit depth for preparing the history projectors to be a refinement parameter. One can begin with a short-depth ansatz for the projectors and incrementally increase the depth to refine the ansatz, potentially improving the approximate consistency. We therefore expect the overall scaling of our quantum circuits to be polynomial in $n$ and $k$ for the anticipated use cases of VCH.

The complexity of minimizing our non-convex cost function is unknown, which is typical for VHQCAs. As classical methods for finding consistent families also involve optimizing over some parameterization for the projectors, classical methods also need to deal with this optimization complexity issue.

While the number of required repetitions of the probability readout step can scale inefficiently in $n$ and $k$ for
certain families of histories, we assume that minimizing the cost outputs a family $F$ for which the probability readout step is efficient. (See Methods for elaboration on this point.)

This scaling behavior means that for systems that can be tractably simulated on a quantum computer and whose properties of interest are simple to implement, we achieve an exponential speedup and reduction in the needed resources as compared to classical approaches to this problem.

**EXPERIMENTAL IMPLEMENTATIONS**

**Quantum Hardware**

We now present an experimental demonstration of VCH on a cloud quantum computer. See the SM for further details on this implementation. We examine a spin-1/2 particle in a magnetic field $B\hat{z}$, whose Hamiltonian is $H = -\gamma B \sigma_z$. For this model, Fig. 3 shows the landscape of the cost in (8) for a simulator and for the ibmqx5 quantum processor [34]. Several minima found by running VCH on ibmqx5 are superimposed on the landscape. These minima correspond reasonably well to theoretically consistent families, and hence this represents a successful proof-of-principle implementation of VCH.

**Simulator**

To highlight applications that will be possible on future hardware, we simulate VCH to observe the quantum-to-classical transition for a chiral molecule [35, 36]. It has been modeled as a two level system where the right $|R\rangle$ and left $|L\rangle$ chirality states are described as $|R\rangle/|L\rangle = (|+\rangle/|\rangle) = \frac{1}{\sqrt{2}} (|0\rangle \pm |1\rangle)$ [36]. A chiral molecule in isolation would tunnel between $|R\rangle$ and $|L\rangle$, but we consider the molecule to be in a gas, where collisions with other molecules convey information about the molecule’s chirality to its environment. This information transfer is modeled by a rotation by angle $\theta_x$ about the $x$ axis of an environment qubit, controlled on the system’s chirality, and for simplicity we suppose such collisions are evenly spaced at five points in time. (See the SM for further details.) We then consider simple families of stationary histories [36], where the projector set corresponds to the same basis at all five times (just after a collision occurs). Letting $\theta_z$ be the precession angle due to tunneling in the time between collisions, we can then explore the competition between decoherence and tunneling. Figure 4 shows our results for this model. Notably we observe the transition from a quantum regime, where the chirality is not consistent, to a classical regime, where the chirality is both consistent and stable over time.

![Cost Landscape (Simulator)](image)

**FIG. 3.** Consistency of three-time histories for a spin-1/2 particle in a magnetic field, with initial state $\rho = |+\rangle|+\rangle$. The full-trace cost landscape, $C(\phi_1, \phi_2)$, is plotted as a function of the azimuths, $\phi_1$ and $\phi_2$, of the first and second projection bases, which we constrained to the $xy$ plane of the Bloch sphere. The point $(0, 0)$ corresponds to both projections being along the $x$ axis. Consistency is expected everywhere along certain vertical lines ($\phi_1 = 2 + n\pi$ rad), as they correspond to histories where the initial state is one of the projectors after the first time step, so there are no branches to interfere in the second time step. In addition, some slope-one lines ($\phi_2 = \phi_1 + (2 + n\pi)$ rad) should be consistent, as they correspond to histories where the second projectors are the same as the first after time evolution, so no interference occurs in the second time step. Indeed, one can see valleys in the cost landscapes for these vertical and slope-one lines, when the cost is quantified on a simulator a and on the ibmqx5 quantum computer b. Note that negative cost values are possible due to finite statistics. The white “x” symbols in b mark some of the non-unique minima that the VCH algorithm found.

**DISCUSSION**

Making it possible to apply the tools and concepts of quantum foundations to a wide array of physical situations, as VCH will, is an important step for our
understanding of the physical world. Specifically by providing an exponential speedup and reduction in resources over classical methods, VCH will provide a way to study phenomena including the quantum-to-classical transition \cite{30,31,37}, dynamics of quantum phase transitions \cite{38}, quantum biological processes \cite{39}, conformational changes \cite{40}, and many other complex phenomena that so far have been computationally intractable. In addition, our work highlights the synergy of two distinct fields, quantum foundations and quantum computational algorithms, and hopefully will inspire further research into their intersection.

\section*{METHODS}

\subsection*{A. Evaluation of the Cost}

Figure 5 shows the circuits for computing the full trace cost (partial trace cost) from two copies of $\sigma^A$ ($\sigma^{SA}$). Note that both costs can be written as a difference of purities:

\begin{align}
C &= \text{Tr}((\sigma^A)^2) - \text{Tr}(Z^A(\sigma^A)^2) \\
C_{pt} &= \text{Tr}((\sigma^{SA})^2) - \text{Tr}(Z^A(\sigma^{SA})^2) .
\end{align}

\begin{align}
C &= \text{Tr}((\sigma^A)^2) - \text{Tr}(Z^A(\sigma^A)^2) \\
C_{pt} &= \text{Tr}((\sigma^{SA})^2) - \text{Tr}(Z^A(\sigma^{SA})^2) .
\end{align}

The $\text{Tr}((\sigma^A)^2)$ and $\text{Tr}((\sigma^{SA})^2)$ terms are computed via the Swap Test, with a depth-two circuit and classical post-processing that scales linearly in the number of qubits \cite{41,42}. A similar but even simpler circuit, called the Diagonalized Inner Product (DIP) Test \cite{25}, calculates the $\text{Tr}(Z^A(\sigma^A)^2)$ term with a depth one circuit and no post-processing. Finally, the $\text{Tr}(Z^A(\sigma^{SA})^2)$ term is evaluated with the Partial-DIP (PDIP) Test \cite{25}, a depth-two circuit that is a hybridization of the Swap Test and the DIP Test.

\subsection*{B. Precision of probability readout}

One does not know \textit{a priori} how many histories will be characterized in the probability readout step (Fig. 2c). Due to statistical noise, the probability of histories with greater probability will be determined with greater relative precision than those with lesser probability. Hence, it is reasonable to set a precision (or statistical noise) threshold, $\epsilon_{\text{max}}$. Let $N_{\text{readout}}$ be the number of repetitions of the probability readout circuit. Then, histories $\lambda^\alpha$ whose bitstring $\alpha$ occurs with frequency $f_\alpha < \sqrt{N_{\text{readout}}}/\epsilon_{\text{max}}$ should be ignored, since their probabilities $p(\alpha) = f_\alpha/N_{\text{readout}}$ were not characterized with the desired precision. We separate $\mathcal{F}$ into the set $\mathcal{F}_c$ of histories whose probabilities are above the precision threshold (which we previously referred to loosely as the most likely
where corresponds to the approximate consistency condition from Eq. (5). Hence, probability sum rules for these two histories are satisfied within error $\epsilon_{\alpha, \alpha'}$, which can be calculated from Eq. (14) for histories in $\mathcal{F}_c$ since the probabilities are known for these histories.

Next, consider histories in $\mathcal{F}$. As we do not have enough information to differentiate these histories, we advocate combining the elements of $\mathcal{F}_c$ into a single coarse-grained history $\mathcal{Y}_\gamma$.

Let $\mathcal{Y}_\beta$ be the least likely history in $\mathcal{F}_c$. Then defining $\delta^2 = D(\gamma, \gamma)/D(\beta, \beta)$, we can make use of the positive semi-definite property of $\sigma^A$ to write:

$$|D(\gamma, \beta)|^2 \leq D(\gamma, \gamma)D(\beta, \beta) = \delta^2 D(\beta, \beta)^2. \quad (16)$$

Since $\mathcal{Y}_\beta$ is the least likely history in $\mathcal{F}_c$, this expression then lets us bound the error on the probability sum rule (giving a weaker approximate consistency condition [29]) between $\mathcal{Y}_\gamma$ and any $\mathcal{Y}_\alpha \in \mathcal{F}_c$ as:

$$|D(\gamma, \alpha)| \leq \delta D(\alpha, \alpha') \leq \delta (D(\gamma, \gamma) + D(\alpha, \alpha')). \quad (17)$$

It is then possible to characterize the approximate consistency of the histories of $\mathcal{F}$ pairwise with $\epsilon_{\alpha, \alpha'}$ and $\delta$. Alternatively, to give an upper bound on the overall consistency $\epsilon$, we take the greatest of these pairwise bounds:

$$\epsilon \leq \max(\{\epsilon_{\alpha, \alpha'}\} \cup \{\delta\}). \quad (18)$$

For those applications where we are working with the partial trace consistency, the notion of approximate consistency is somewhat more obscured. In order to generate probabilities and bound $\epsilon$, we therefore recommend evaluating the full trace cost function at the minimum found with the partial trace cost. This approach is helpful since any partial trace consistent family will also be full trace consistent and the partial trace consistency does not directly allow one to discuss probabilities in the same way. Taking this approach allows us to then directly utilize the approximate consistency framework above.

**ACKNOWLEDGMENTS**

This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of High Energy Physics, and also by the U.S. DOE, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, Condensed Matter Theory Program. All authors acknowledge support from the LDRD program at Los Alamos National Laboratory (LANL). LC was also supported by the DOE through the J. Robert Oppenheimer fellowship. ATS and PJC additionally acknowledge support from the LANL ASC Beyond Moore’s Law project. Finally, WHZ acknowledges partial support by the Foundational Questions Institute grant FQXi-1821, and Franklin Fetzer Fund, a donor advised fund of the Silicon Valley Community Foundation.
[1] J. A. Wheeler and W. H. Zurek, editors. *Quantum Theory and Measurement (Princeton Series in Physics).* Princeton University Press, 2016.

[2] G. Auletta. *Foundations and Interpretation of Quantum Mechanics.* World Scientific, 2000.

[3] E. Joos and H. D. Zeh. The emergence of classical properties through interaction with the environment. *Zeitschrift für Physik B Condensed Matter,* 59(2):223–243, 1985.

[4] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.,* 75:715–775, 2003.

[5] M. A. Schlosshauer. *Decoherence: and the quantum-to-classical transition.* Springer Science & Business Media, 2007.

[6] R. B. Griffiths. Consistent histories and the interpretation of quantum mechanics. *Journal of Statistical Physics,* 36(1):219–272, 1984.

[7] R. Omnès. Logical reformulation of quantum mechanics. *Journal of Statistical Physics,* 53(3):893–932, 1988.

[8] M. Gell-Mann and J. B. Hartle. Quantum Mechanics in the Light of Quantum Cosmology. In *Proceedings of the 3rd International Symposium Foundations of Quantum Mechanics in the Light of New Technology,* 1989.

[9] J. Hartle and T. Hertog. One bubble to rule them all. *Phys. Rev. D,* 95:123502, 2017.

[10] S. Lloyd. Decoherent histories approach to the cosmological measure problem. *arXiv:1608.05672.*

[11] T. A. Brun. Quantum jumps as decoherent histories. *Phys. Rev. Lett.,* 78:1833–1837, 1997.

[12] J. J. Halliwell and J. M. Yearsley. Quantum arrival time formula from decoherent histories. *Physics Letters A,* 374(2):154 – 157, 2009.

[13] J. J. Halliwell and J. M. Yearsley. Arrival times, complex potentials, and decoherent histories. *Phys. Rev. A,* 79:062101, 2009.

[14] C. Anastopoulos and N. Savvidou. Time of arrival and localization of relativistic particles. *arXiv:1807.06533.*

[15] H.-J. Poehl. How to calculate decoherence matrices numerically. *Physica A: Statistical Mechanics and its Applications,* 213(3):435 – 449, 1995.

[16] D. Schmidtke and J. Gemmer. Numerical evidence for approximate consistency and Kovarikovianity of some quantum histories in a class of finite closed spin systems. *Phys. Rev. E,* 93:012125, 2016.

[17] J. Preskill. Quantum Computing in the NISQ era and beyond. *Quantum,* 2:79, 2018.

[18] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications,* 5:4213, 2014.

[19] E. R. Anschuetz, J. P. Olson, A. Aspuru-Guzik, and Y. Cao. Variational quantum factoring. *arXiv:1808.08927.*

[20] E. Farhi, J. Goldstone, and S. Gutmann. A quantum approximate optimization algorithm. *arXiv:1411.4028.*

[21] J. Romero, J. P. Olson, and A. Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology,* 2(4):045001, 2017.

[22] Y. Li and S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X,* 7(2):021050, 2017.

[23] P. D. Johnson, J. Romero, J. Olson, Y. Cao, and A. Aspuru-Guzik. QVECTOR: an algorithm for device-tailored quantum error correction. *arXiv:1711.02249.*

[24] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles. Quantum assisted quantum compiling. *arXiv:1807.08080.*

[25] R. LaRose, A. Tikku, É. O’Neel-Judy, L. Cincio, and P. J. Coles. Variational quantum state diagonalization. *arXiv:1810.10506.*

[26] R. B. Griffiths. *Consistent Quantum Theory.* Cambridge University Press, 2001.

[27] J. J. Halliwell. A review of the decoherent histories approach to quantum mechanics. *Annals of the New York Academy of Sciences,* 755(1):726–740, 1995.

[28] P. C. Hohenberg. Colloquium: An introduction to consistent quantum theory. *Rev. Mod. Phys.,* 82:2835–2844, 2010.

[29] H. F. Dowker and J. J. Halliwell. Quantum mechanics of history: The decoherence functional in quantum mechanics. *Phys. Rev. D,* 46:1580–1609, 1992.

[30] C. J. Riedel, W. H. Zurek, and M. Zwolak. Objective past of a quantum universe: Redundant records of consistent histories. *Phys. Rev. A,* 93:032126, 2016.

[31] J. Finkelstein. Definition of decoherence. *Phys. Rev. D,* 47:5430–5433, 1993.

[32] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics,* 18(2):023023, 2016.

[33] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma. Simulating hamiltonian dynamics with a truncated taylor series. *Phys. Rev. Lett.,* 114:090502, 2015.

[34] IBM Q 16 Rueschlikon backend specification. https://github.com/Qiskit/qiskit-backend-information/tree/master/backends/rueschlikon/V1, 2018.

[35] J. Trost and K. Hornberger. Hund’s paradox and the collisional stabilization of chiral molecules. *Phys. Rev. Lett.,* 103:023202, 2009.

[36] P. J. Coles, V. Gheorghiu, and R. B. Griffiths. Consistent histories for tunneling molecules subject to collisional decoherence. *Phys. Rev. A,* 86:042111, 2012.

[37] J. P. Paz and W. H. Zurek. Environment-induced decoherence, classicality, and consistency of quantum histories. *Phys. Rev. D,* 48:2728–2738, 1993.

[38] W. H. Zurek, U. Dorner, and P. Zoller. Dynamics of a Quantum Phase Transition. *Phys. Rev. Lett.,* 95:105701, 2005.

[39] M. Allegra, P. Giorda, and S. Lloyd. Global coherence of quantum evolutions based on decoherent histories: Theory and application to photosynthetic quantum energy transport. *Phys. Rev. A,* 93:042312, 2016.

[40] H. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hernández, and W. Yang. Quantum mechanics simulation of protein dynamics on long timescale. *Phys. Rev. Lett.,* 103:023202, 2009.

[41] J. C. Garcia-Escartin and P. Chamorro-Posada. Swap test and Hong-Ou-Mandel effect are equivalent. *Physical Review A,* 87(5):052330, 2013.

[42] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles.
Learning the quantum algorithm for state overlap. *New Journal of Physics*, 20(11):113022, 2018.

[43] M. Gell-Mann and J. B. Hartle. Classical equations for quantum systems. *Phys. Rev. D*, 47:3345–3382, 1993.
Appendix A: Generalizations

Here we discuss various generalizations of the circuits shown in the main text, which presented our VCH algorithm for the special case of branch-independent histories of a one-qubit system $S$ with no environment $E$.

1. Multi-Qubit Systems

The circuits in the main text showed systems $S$ composed of a single qubit. The generalization to multi-qubit systems is straightforward. We must discuss the generalizations of both the state preparation circuit in Fig. 2 as well as the cost evaluation circuits in Fig. 5.

Figure S.1 illustrates how the state preparation circuit generalizes to multi-qubit systems. In particular, this figure shows how a portion of state preparation circuit (the portion that entangles the system to the ancillas) generalizes for the case of a fine-grained set of projectors. (Note that the case of a coarse-grained set of projectors is discussed in the next subsection.)

The cost evaluation circuits in Fig. 5 generalize as follows. For fine-grained histories, one needs $n$ ancillas for each time step and hence a total of $nk$ ancillas. The circuits in Fig. 5 shown for $k$ ancillas generalize in a straightforward way, where now one has $nk$ ancilla systems. In addition, the circuits in Fig. 5b also involve the $S$ system, and hence all $n$ qubits in $S$ must be included in this circuit. Again, these $n$ qubits are included in the most straightforward way (in the same way that the single qubit $S$ system appears in the circuits in Fig. 5b).

2. Coarse Grained Histories

Multi-qubit systems $S$ allow for non-trivial coarse-grained histories. In such families of histories, the sets $P_j$ are composed of projectors whose ranks are possibly greater than one. We remark that coarse-grained histories are often important to the study of macroscopic systems and the quantum-to-classical transition. VCH can easily be adapted to study coarse-grained histories as follows.

For each time $t_j$, one should decide (prior to running VCH) what projector ranks that one is interested in. VCH will then optimize over sets of projectors with these particular ranks. The projector ranks can therefore be viewed as hyperparameters, i.e., parameters that one fixes for a given run of VCH.

For instance, suppose $S$ is composed of a pair of spins. In this case, Fig. S.2 shows two examples of the state preparation circuit for a single time step. In the first example, Fig. S.2a, we consider a projector set that contains two rank-two projectors revealing whether the spins were aligned or anti-aligned. In the second example, Fig. S.2b, we consider a projector set that contains a rank-three and a rank-one projector that respectively indicate whether the spins are in the triplet states or the the singlet state. Note that the ranks of the projectors are determined by the gate that entangles the system to the ancilla, which is a single CNOT gate in Fig. S.2a and a Toffoli gate in Fig. S.2b. Hence the choice of the projector ranks (mentioned in the previous paragraph) translates into a choice of gate sequence that entangles the system to the ancilla.

3. Nontrivial Environments

For many applications of VCH, (e.g., the chiral molecule example in the main text) it will be helpful to explicitly model an environment $E$. We can think of this case as a particular choice of coarse graining where the projectors we consider only act on a subsystem of our
model (the S system) and do not directly record any information about E. Note that the Hamiltonian evolution involves both S and E, as shown in Fig. S.3.

![Diagram](image)

**FIG. S.3.** Simple example with an environment E. The projectors still only act on S, but the evolution includes both S and E.

4. **Branch Dependent Histories**

A final generalization that we consider are families of branch dependent histories [43], or histories where the projector set at a given time may depend on the properties of the system at earlier points in the histories. VCH can accommodate these histories, as follows.

The basic idea is that the unitary gate $B_j$ that determines the projector set at time $t_j$ now becomes a controlled unitary. Specifically, the control system(s) for $B_j$ are (potentially) all the ancilla qubits associated with times $t_i < t_j$. So the choice of projector set at some time is influenced by the ancillas states for earlier times.

*Figure S.4* shows an example of what this looks like, for the special case of only two times. In this figure, if the first ancilla is in the $|0\rangle$ state ($|1\rangle$ state), then the $B_2$ unitary ($B_2^\dagger B_2$ unitary) is applied at the second time step. For more general cases, the $B_2^\dagger$ unitary shown here might be replaced by a sequence of controlled unitaries controlled by different ancilla qubits.

![Diagram](image)

**FIG. S.4.** Example implementation of a branch dependent projector set in our state preparation circuit. In this circuit, depending upon the result for $t_1$, either $B_2$ or the product $B_2^\dagger B_2$ defines the projector set for the second time.

**Appendix B: Generalized state preparation**

We now present the details of our generalized state preparation circuit (as shown in Fig. S.5) and show that $\sigma^{SA}$ and $\sigma^A$ have the properties we claim in the main text. Note that our treatment here includes all of the generalizations discussed above in Appendix A. We then apply the gate sequence associated with the $P_i$ projector set, which includes $B_1$, a multi-qubit gate that entangles $S$ and $A$ (which we refer to as the “entangling gate”), and then $B_1^\dagger$. This gives the state:

$$\sum_{\alpha_1,\alpha'_1} \left[ P_{11}^{\alpha_1} \rho^{SE} P_{11}^{\alpha'_1} \right] \otimes \left[ |\alpha_1\rangle \langle \alpha'_1| \otimes |0\rangle \langle 0| \right]^A.$$  \hspace{1cm} (B1)

Note that the system and ancilla are (possibly) entangled at this point.

Next in our state preparation circuit is the time evolution from $t_1$ to $t_2$, given by $e^{-iH\Delta t_{1,2}}$. This is followed by the gate sequence associated with $P_2$, which in general may be branch dependent. The resulting state is

$$\sum_{\alpha_1,\alpha'_1,\alpha_2,\alpha'_2} \left[ P_{22}^{\alpha_2}(\alpha_1) e^{-iH\Delta t_{1,2}} P_{11}^{\alpha_1} \rho^{SE} P_{11}^{\alpha'_1} e^{iH\Delta t_{1,2}} \right]$$

$$\otimes \left[ |\alpha_1\rangle \langle \alpha'_1| \otimes |\alpha_2\rangle \langle \alpha'_2| \right] \otimes |0\rangle \langle 0|^A,$$  \hspace{1cm} (B2)

where the notation $P_{22}^{\alpha_2}(\alpha_1)$ indicates that the second projector set depends on $\alpha_1$. Repeating this state evolution until we have applied the gate sequences associated with all $k$ projector sets (and switching to the Heisenberg picture), we end up with

$$\sum_{\alpha,\alpha'} \left[ P_{k}^{\alpha_k}(t_k) \cdots P_{2}^{\alpha_2}(t_2) P_{1}^{\alpha_1}(t_1) \rho^{SE} \right]$$

$$\otimes \left[ |\alpha_1\rangle \langle \alpha'_1| \otimes |\alpha_2\rangle \langle \alpha'_2| \otimes \cdots \otimes |\alpha_k\rangle \langle \alpha'_k| \right]$$

$$= \sum_{\alpha,\alpha'} C^{\alpha_\alpha'} c^{\alpha_-} \otimes |\langle \alpha\rangle \langle \alpha'| \rangle|^A.$$  \hspace{1cm} (B3)

Note that we have suppressed explicit branch dependence here to simplify notation. Branch dependence does not alter the formalism except to make the later projectors functions of the earlier $\alpha_i$’s, so our treatment remains fully general.

If we then trace out the environment (which in the circuit means not measuring it) we are then left with $\sigma^{SA}$:

$$\sigma^{SA} = \sum_{\alpha,\alpha'} \text{Tr}_E(C^{\alpha} \rho^{SE} C^{\alpha_-}) \otimes |\langle \alpha\rangle \langle \alpha'| \rangle|^A.$$  \hspace{1cm} (B4)

By examining Eq. (B4), we can see that $(1 \otimes |\langle \alpha|\rangle) \sigma^{SA} (1 \otimes |\langle \alpha|\rangle)$ is precisely $D_{pt}(\alpha,\alpha') = \text{Tr}_E(C^{\alpha} \rho^{SE} c^{\alpha_-})$. Further, if we similarly trace over the system S, we get:

$$\sigma^A = \sum_{\alpha,\alpha'} \text{Tr}(C^{\alpha} \rho^{SE} c^{\alpha_-}) |\langle \alpha\rangle \langle \alpha'| \rangle|^A.$$  \hspace{1cm} (B5)

We can thus see that we have prepared a density matrix whose elements are $D(\alpha,\alpha') = \text{Tr}(C^{\alpha} \rho^{SE} c^{\alpha_-})$, as claimed in the main text.
Appendix C: Derivation of Cost Functions

1. Full trace cost

Let us now derive the equivalence stated in the definition of our full trace cost function, Eq. (8). Starting with the definition of $C$ we have:

\[
C := \sum_{\alpha \neq \alpha'} |\mathcal{D}(\alpha, \alpha')|^2 = \sum_{\alpha \neq \alpha'} \langle \alpha | \sigma^A | \alpha' \rangle \langle \alpha' | \sigma^A | \alpha \rangle \\
= \sum_{\alpha \neq \alpha'} \text{Tr} (\langle \alpha | \sigma^A | \alpha' \rangle \langle \alpha' | \sigma^A | \alpha \rangle) \\
= \sum_{\alpha, \alpha'} \text{Tr} (\langle \alpha | \sigma^A | \alpha' \rangle \langle \alpha' | \sigma^A | \alpha \rangle) \\
- \sum_{\alpha} \text{Tr} (\langle \alpha | \sigma^A | \alpha \rangle \langle \alpha' | \sigma^A | \alpha \rangle) \\
= \text{Tr}(\sigma^A)^2 - \text{Tr}(Z^A(\sigma^A)^2) \\
= D_{\text{HS}}(\sigma^A, Z^A(\sigma^A)). \tag{C1}
\]

Therefore, the circuits we use to calculate $\text{Tr}(\sigma^A)^2$ and $\text{Tr}(Z^A(\sigma^A)^2)$ implement this cost function as claimed.

2. Partial trace cost

Arriving at the expression for the partial trace cost function (Eq. (9)) is similar if slightly more complicated:

\[
C_{\text{pt}} := \sum_{\alpha \neq \alpha'} \|\mathcal{D}_{\text{pt}}(\alpha, \alpha')\|_{\text{HS}}^2 = \sum_{\alpha \neq \alpha'} \text{Tr} (\mathcal{D}_{\text{pt}}(\alpha, \alpha') \mathcal{D}_{\text{pt}}(\alpha, \alpha')^\dagger) \\
= \sum_{\alpha \neq \alpha'} \text{Tr} \left( (I \otimes |\alpha\rangle)(I \otimes |\alpha\rangle)^\dagger \sigma^A \right) \\
= \sum_{\alpha \neq \alpha'} \text{Tr} \left( (I \otimes |\alpha\rangle)(I \otimes |\alpha\rangle)^\dagger \sigma^A (I \otimes |\alpha\rangle) \right) \\
= \sum_{\alpha \neq \alpha'} \text{Tr} \left( (I \otimes |\alpha\rangle)(I \otimes |\alpha\rangle)^\dagger \sigma^A (I \otimes |\alpha\rangle)^\dagger \sigma^A \right) \\
= \sum_{\alpha \neq \alpha'} \text{Tr} \left( (I \otimes |\alpha\rangle)(I \otimes |\alpha\rangle)^\dagger \sigma^A (I \otimes |\alpha\rangle \sigma^A) \right) \\
- \sum_{\alpha} \text{Tr} \left( (I \otimes |\alpha\rangle)(I \otimes |\alpha\rangle)^\dagger \sigma^A \sigma^A \right) \\
= \text{Tr}(\sigma^A)^2 - \text{Tr}(Z^A(\sigma^A)^2) \\
= D_{\text{HS}}(\sigma^A, Z^A(\sigma^A)). \tag{C2}
\]

As with the full trace cost function, the circuits we use to calculate $\text{Tr}(\sigma^A)^2$ and $\text{Tr}(Z^A(\sigma^A)^2)$ thus implement this cost function as claimed.

Appendix D: Reading out the Decoherence Functional Elements

While VCH avoids the need to compute the exponentially many $\mathcal{D}(\alpha, \alpha')$s in order to determine the consistency of a family $F$, we do have the ability to efficiently read out any particular $\mathcal{D}(\alpha, \alpha')$ if desired. Figure S.6 shows the circuit that one can use to read the real and/or imaginary parts of $\mathcal{D}(\alpha, \alpha')$ out for $\alpha \neq \alpha'$. The post-processing is similar to that of the Swap test [41, 42], except that we add a conditional statement.
FIG. S.6. Circuit to read out $D(\alpha, \alpha')$. The controlled $U(\alpha, \alpha')$ prepares the state $|\alpha\rangle$ on the B registers when the control qubit in the state $|0\rangle$ and $|\alpha'\rangle$ when the control qubit is in the state $|1\rangle$, so the combination of the Hadamard gate on C and the controlled $U(\alpha, \alpha')$ prepares a superposition of the histories. The $z$-rotation in the green box is excluded when we calculate the real part of $D(\alpha, \alpha')$ and included when we calculate the imaginary part. The post processing is described in the text.

When we exclude the $z$-rotation, conditioned on the control qubit C being measured in the state $|0\rangle$ we perform the Swap test between the A and B registers to get:

$$R_0 = \text{Tr} \left( \sigma^A \left[ \frac{1}{2} (|\alpha\rangle + |\alpha'\rangle) (\langle \alpha | + \langle \alpha' |) \right] \right)$$

$$= \frac{1}{2} \left( D(\alpha, \alpha) + D(\alpha', \alpha') + D(\alpha, \alpha') + D(\alpha', \alpha) \right)$$

$$= \frac{1}{2} \left( D(\alpha, \alpha) + D(\alpha', \alpha') \right) + \text{Re}(D(\alpha, \alpha')). \quad \text{(D1)}$$

If we instead condition on C being measured in the state $|1\rangle$ we perform the Swap test between the A and B registers to get:

$$R_1 = \text{Tr} \left( \sigma^A \left[ \frac{1}{2} (|\alpha\rangle - |\alpha'\rangle) (\langle \alpha | - \langle \alpha' |) \right] \right)$$

$$= \frac{1}{2} \left( D(\alpha, \alpha) + D(\alpha', \alpha') \right) - \text{Re}(D(\alpha, \alpha')). \quad \text{(D2)}$$

Our method therefore separates the output based on the result of measuring C, and then performs the usual Swap test post processing on each partition of the output counts to get $R_0$ and $R_1$. Finally, we combine these to get:

$$\text{Re}(D(\alpha, \alpha')) = \frac{1}{2} (R_0 - R_1). \quad \text{(D3)}$$

Instead including that $z$-rotation gives us

$$I_0 = \text{Tr} \left( \sigma^A \left[ \frac{1}{2} (|\alpha\rangle + i|\alpha'\rangle) (\langle \alpha | - i\langle \alpha' |) \right] \right)$$

$$= \frac{1}{2} \left( D(\alpha, \alpha) + D(\alpha', \alpha') - iD(\alpha, \alpha') + iD(\alpha', \alpha) \right)$$

$$= \frac{1}{2} \left( D(\alpha, \alpha) + D(\alpha', \alpha') \right) - \text{Im}(D(\alpha, \alpha')). \quad \text{(D4)}$$

conditioned on C being measured in the state $|0\rangle$. Similarly, conditioned on C being measured in the state $|1\rangle$ we find:

$$I_1 = \text{Tr} \left( \sigma^A \left[ \frac{1}{2} (|\alpha\rangle - i|\alpha'\rangle) (\langle \alpha | + i\langle \alpha' |) \right] \right)$$

$$= \frac{1}{2} \left( D(\alpha, \alpha) + D(\alpha', \alpha') + iD(\alpha, \alpha') - iD(\alpha', \alpha) \right)$$

$$= \frac{1}{2} \left( D(\alpha, \alpha) + D(\alpha', \alpha') \right) + \text{Im}(D(\alpha, \alpha')). \quad \text{(D5)}$$

Again, we combine these to get:

$$\text{Im}(D(\alpha, \alpha')) = \frac{1}{2} (I_1 - I_0) \quad \text{(D6)}$$

We also note that the controlled $U(\alpha, \alpha')$ we have made use of here can be implemented with depth that scales linearly in the number of bits by which $|\alpha\rangle$ and $|\alpha'\rangle$ differ. This is accomplished by acting with $X$ gates on all of the registers where the bit-string associated with $|\alpha\rangle$ is 1 followed by CNOT gates from C to each of the registers where the bit-strings for $|\alpha\rangle$ and $|\alpha'\rangle$ differ.

Finally, we comment that reading out $D(\alpha, \alpha)$ is simpler than the general case as we merely have to prepare $|\alpha\rangle/|\alpha\rangle$ (which consists of a single layer of $X$ gates) on the B registers and perform the Swap test, without any need for or reference to C.

Appendix E: Implementation Circuits

1. Spin in a Magnetic Field

For our simulations of the spin-1/2 particle in a magnetic field, Fig. S.7 shows the quantum circuit that was used on the simulator and IBM’s ibmqx5 processor to perform the cost minimization and to generate the cost landscape plots (shown in Fig. 3).

2. Chiral Molecule

Figure S.8 shows the quantum circuit that was used on a simulator to map the cost function landscapes for the chiral molecule (shown in Fig 4). The tunneling between the chirality states was modeled as a rotation about the $z$-axis by an angle $\theta_z$. We considered the chiral molecule to be in a gas, and hence its environment is composed of other surrounding molecules that may collide with the molecule of interest. Our model for these collision interactions was implemented by performing a rotation around the $x$-axis by an angle $\theta_x$ (which determines the interaction strength) on an environmental qubit representing the colliding molecule, controlled by the chirality of the molecule of interest.
**FIG. S.7.** Quantum circuit that we employed to evaluate the cost functions for the spin in a magnetic field. The wires labeled \( S \) represent the copies of the spin and those labeled \( A \) represent the ancillas. Note that this circuit prepares two copies of \( \sigma^A \). The gates and measurements inside the solid green box are only included to calculate \( \text{Tr}((\sigma^A)^2) \), as without them this is the circuit to calculate \( \text{Tr}(Z^A(\sigma^A)^2) \).

**FIG. S.8.** Quantum circuit that we employed to evaluate the cost functions for the chiral molecule example in the main text. The wires labeled \( S \) represent the chirality degree of freedom of the molecule, \( E \) represents the environment (other surrounding molecules), and \( A \) represents. Note that this circuit prepares two copies of \( \sigma^{SA} \) (and hence \( \sigma^A \)). The gates and measurements inside the blue dashed boxes are only included when we are evaluating the partial trace cost function (i.e., when working with \( \sigma^{SA} \) rather than \( \sigma^A \)). The gates and measurements inside the solid green box are only included when calculating \( \text{Tr}((\sigma^A)^2) \) or \( \text{Tr}((\sigma^{SA})^2) \), and otherwise the circuit calculates \( \text{Tr}(Z^A(\sigma^A)^2) \) or \( \text{Tr}(Z^A(\sigma^{SA})^2) \).