

Supplementary Material for “Variational Fast Forwarding for Quantum Simulation Beyond the Coherence Time”

Supplementary Note 1. COST FUNCTION

Here we elaborate on our cost function. As noted in Sec. II A 2, our proposed cost function is the C_{LHST} function introduced in Ref. [34], defined by

$$C_{\text{LHST}}(U, V) = 1 - \frac{1}{n} \sum_{j=1}^n F_e^{(j)}. \quad (\text{S1})$$

Let us now precisely define the entanglement fidelities $F_e^{(j)}$. Consider a $2n$ -qubit system composed of the n -qubit subsystems A and B . Let A_j (B_j) denote the j -th qubit of system A (B). Let $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ denote the standard 2-qubit Bell state. Then we can write $F_e^{(j)}$ as

$$F_e^{(j)} := \text{Tr} \left(|\Phi^+\rangle \langle \Phi^+|_{A_j B_j} (\mathcal{E}_j \otimes \mathcal{I}_{B_j}) (|\Phi^+\rangle \langle \Phi^+|_{A_j B_j}) \right). \quad (\text{S2})$$

Here, $\mathcal{E}_j$ is a quantum channel that acts on qubit A_j as follows. For an arbitrary state ρ_{A_j} ,

$$\mathcal{E}_j(\rho_{A_j}) = \text{Tr}_{\bar{A}_j} \left(UV^\dagger \left(\rho_{A_j} \otimes \frac{\mathbb{1}_{\bar{A}_j}}{2^{n-1}} \right) VU^\dagger \right), \quad (\text{S3})$$

where $\bar{A}_j$ is the set of all qubits in A except for A_j .

The fact that C_{LHST} is a faithful cost function was shown in [34] by relating C_{LHST} to another cost function whose properties are more transparent. Namely, consider the function

$$C_{\text{HST}}(U, V) = 1 - \frac{1}{d^2} |\text{Tr}(UV^\dagger)|^2. \quad (\text{S4})$$

Since $\text{Tr}(UV^\dagger)$ is the Hilbert-Schmidt inner product, it is clear that $C_{\text{HST}}(U, V) = 0$ if and only if $V = U$ (up to global phase). Reference [34] then proved the following relation:

$$C_{\text{LHST}}(U, V) \leq C_{\text{HST}}(U, V) \leq n C_{\text{LHST}}(U, V). \quad (\text{S5})$$

This implies that C_{LHST} vanishes under precisely the same conditions as C_{HST} , and hence that C_{LHST} is faithful.

While the C_{HST} function has direct operational meaning in terms of the inner product between U and V , we propose to use C_{LHST} instead of C_{HST} for the following reason. In Ref. [34], it was argued that there are simple examples (e.g., when U and V are tensor-product unitaries) for which the gradient of C_{HST} vanishes exponentially with n , while the gradient of C_{LHST} is independent of n . This implies that C_{LHST} is easier to train than C_{HST} for large n , and indeed Ref. [34] confirmed this with numerical implementations for increasing values of n . Hence C_{LHST} has better scaling properties, while it also inherits the operational meaning of C_{HST} via the relation in (S5).

Supplementary Note 2. SIMULATION ERRORS

A. Linear scaling in N

Here we provide a proof of (13), which is restated in the following lemma.

Lemma 1: Suppose U_1 and U_2 are two unitary matrices, then for any positive integer N we have:

$$\|U_1^N - U_2^N\|_p \leq N \|U_1 - U_2\|_p \quad (\text{S6})$$

where $\|\dots\|_p$ denotes the Schatten p -norm.

Proof. We expand the norm of the difference of products by adding and subtracting convenient terms so that

$$\|U_1^N - U_2^N\|_p = \|U_1^N - U_1^{N-1}U_2 + U_1^{N-1}U_2 - U_1^{N-2}U_2^2 \dots + U_1U_2^{N-1} - U_2^N\|_p. \quad (\text{S7})$$

From the triangle inequality it follows that:

$$\|U_1^N - U_2^N\|_p \leq \|U_1^{N-1}(U_1 - U_2)\|_p + \|U_1^{N-2}(U_1 - U_2)U_2\|_p \dots + \|(U_1 - U_2)U_2^{N-1}\|_p. \quad (\text{S8})$$

There are a total of N terms in the above summation and as the Schatten norms are unitarily invariant, meaning that $\|UAV\| = \|A\|$ for any unitary matrices U and V , each of these N terms is equal to $\|U_1 - U_2\|$. Thus we obtain the required result. $\square$

B. Scaling of cost function with N

Here we reformulate Lemma 1 in terms of the VFF cost function. Since the latter can be efficiently estimated on a quantum computer, one can view this reformulation as a certifiable version of Lemma 1. This reformulation is derived by specializing Lemma 1 to the case of $p = 2$, i.e. the Hilbert-Schmidt norm.

We first remark that for our purposes the phases of the simulated unitaries are always global phases and therefore unphysical. (This is true since we intend for V to be implemented directly on an n -qubit system rather than on a subsystem as a subcomponent of a larger simulation.) We therefore introduce the phase-independent quantity $\tilde{\epsilon}_2(U_1, U_2)$,

$$\tilde{\epsilon}_2(U_1, U_2) := \min_{\phi} \|U_1 - \exp(i\phi)U_2\|_2, \quad (\text{S9})$$

which depends only on the Hilbert-Schmidt inner product,

$$\begin{aligned} \tilde{\epsilon}_2(U_1, U_2) &= \min_{\phi} \sqrt{2d - 2\text{Re}(\text{Tr}(U_1^\dagger e^{i\phi}U_2))} \\ &= \sqrt{2d - 2|\text{Tr}(U_1^\dagger U_2)|}. \end{aligned} \quad (\text{S10})$$

Hence this quantity can be related to C_{HST} defined in (S4) by

$$\tilde{\epsilon}_2(U_1, U_2) = \sqrt{2d \left(1 - \sqrt{1 - C_{\text{HST}}(U_1, U_2)}\right)}. \quad (\text{S11})$$

Now, let us specialize Lemma 1 to $p = 2$ and minimize over all global phase factors applied to the unitary U_2 . Using (S11), this results in:

$$1 - \sqrt{1 - C_{\text{HST}}(U_1^N, U_2^N)} \leq N^2 \left(1 - \sqrt{1 - C_{\text{HST}}(U_1, U_2)}\right). \quad (\text{S12})$$

Given that C_{LHST} is bounded by C_{HST} via (S5), the fast-forwarded C_{LHST} can similarly be bounded as

$$1 - \sqrt{1 - C_{\text{LHST}}(U^N, V^N)} \leq N^2 \left(1 - \sqrt{1 - n C_{\text{LHST}}(U, V)}\right), \quad (\text{S13})$$

where we assume $n C_{\text{LHST}} \leq 1$ and we chose $U_1 = U$ and $U_2 = V$. Equation (S13) is the exact version of (15) in the main text. Specializing to the case where the cost function C_{LHST} is small, (S13) becomes (15), i.e.,

$$C_{\text{LHST}}(U^N, V^N) \lesssim n N^2 C_{\text{LHST}}(U, V). \quad (\text{S14})$$

C. An operational termination condition

The VFF cost function is operationally meaningful by virtue of its relation to the average-case diagonalization error. Specifically, it can be shown [56, 57] that

$$C_{\text{HST}}(U_1, U_2) = \frac{d+1}{d} (1 - \overline{F}(U_1, U_2)) \quad (\text{S15})$$

where

$$\bar{F}(U_1, U_2) := \int_{\psi} |\langle \psi(U_1) | \psi(U_2) \rangle|^2 d\psi \quad (\text{S16})$$

is the average fidelity over the Haar distribution. Therefore, from Eq. (S5), C_{LHST} upper bounds the average fidelity as follows

$$C_{\text{LHST}}(U_1, U_2) \geq \frac{d+1}{nd} (1 - \bar{F}(U_1, U_2)). \quad (\text{S17})$$

This relation enables us to bound the average simulation error and hence provide a termination condition for VFF.

To derive a termination condition, we start from the bound on the total simulation error, Eq. (14), written in terms of the Hilbert-Schmidt norm ($p = 2$) and take the minimum of the total simulation error and diagonalization errors over global phases applies to V to remove their arbitrary phase dependence:

$$\tilde{\epsilon}_2(e^{-iHT}, V(\alpha, T)) \leq N(\epsilon_2^{\text{TS}}(\Delta t) + \tilde{\epsilon}_2(U(\Delta t), V(\alpha, \Delta t))). \quad (\text{S18})$$

We can then rewrite this expression in terms of the average fidelity of the simulation,

$$\bar{F}(T) := \bar{F}(e^{-iHT}, V(\alpha, T)), \quad (\text{S19})$$

and the cost function C_{HST} using (S11) and (S15):

$$G(T) \leq \frac{N}{\sqrt{2d}} \epsilon_2^{\text{TS}}(\Delta t) + N \sqrt{1 - \sqrt{1 - C_{\text{HST}}^{\text{VFF}}(\Delta t)}}. \quad (\text{S20})$$

Here, to simplify the expression, we have defined

$$G(T) := \sqrt{1 - \sqrt{1 - \frac{d+1}{d} (1 - \bar{F}(T))}}, \quad (\text{S21})$$

and $C_{\text{HST}}^{\text{VFF}}(\Delta t) := C_{\text{HST}}(U(\Delta t), V(\alpha, \Delta t))$. As the operator norm is typically used for Trotter error analysis in the quantum simulation literature [40, 41], we rewrite $\epsilon_2^{\text{TS}}(\Delta t)$ in terms of the operator norm using equivalence relation $\|X\|_2 \leq \sqrt{d} \|X\|_{\infty}$,

$$G(T) \leq \frac{N}{\sqrt{2}} \epsilon_{\infty}^{\text{TS}}(\Delta t) + N \sqrt{1 - \sqrt{1 - C_{\text{HST}}^{\text{VFF}}(\Delta t)}}. \quad (\text{S22})$$

Finally, C_{HST} is upper bounded by nC_{LHST} and therefore

$$G(T) \leq \frac{N}{\sqrt{2}} \epsilon_{\infty}^{\text{TS}}(\Delta t) + N \sqrt{1 - \sqrt{1 - nC_{\text{LHST}}^{\text{VFF}}(\Delta t)}}, \quad (\text{S23})$$

assuming that $nC_{\text{LHST}}^{\text{VFF}}(\Delta t) \leq 1$. Re-arranging terms and denoting by $\epsilon(\Delta t) := \frac{1}{\sqrt{2}} \epsilon_{\infty}^{\text{TS}}(\Delta t) + \sqrt{1 - \sqrt{1 - nC_{\text{LHST}}^{\text{VFF}}(\Delta t)}}$ we get that whenever $\epsilon(\Delta t) \leq 1/N$ then the average fidelity is bounded by

$$\bar{F}(T) \geq \frac{1}{d+1} + \frac{d}{d+1} (1 - N^2 \epsilon(\Delta t)^2)^2 \quad (\text{S24})$$

$$= 1 - \frac{N^2 \epsilon(\Delta t)^2 d}{d+1} (2 - N^2 \epsilon(\Delta t)^2) \quad (\text{S25})$$

$$\geq 1 - \frac{N^2 d}{d+1} 2\epsilon(\Delta t)^2. \quad (\text{S26})$$

One can get a more compact (but weaker) lower bound on the average fidelity by observing that

$$\sqrt{1 - \sqrt{1 - nC_{\text{LHST}}^{\text{VFF}}(\Delta t)}} \leq \sqrt{nC_{\text{LHST}}^{\text{VFF}}(\Delta t)} \quad (\text{S27})$$

which holds whenever $nC_{\text{LHST}}^{\text{VFF}}(\Delta t) \leq 1$. Therefore we get an upper bound on $\epsilon(\Delta t) \leq \frac{1}{\sqrt{2}}\epsilon_{\infty}^{\text{TS}}(\Delta t) + \sqrt{nC_{\text{LHST}}^{\text{VFF}}(\Delta t)}$ so that (S26) becomes:

$$\bar{F}(T) \geq 1 - \frac{N^2 d}{d+1} \left(\epsilon_{\infty}^{\text{TS}}(\Delta t) + \sqrt{2nC_{\text{LHST}}^{\text{VFF}}(\Delta t)} \right)^2. \quad (\text{S28})$$

After a successful optimization procedure, $nC_{\text{LHST}}^{\text{VFF}}(\Delta t)$ is expected to be small and Eq. (S26) reduces to

$$\bar{F}(T) \gtrsim 1 - \frac{N^2 d}{d+1} \left(\epsilon_{\infty}^{\text{TS}}(\Delta t) + \sqrt{nC_{\text{LHST}}^{\text{VFF}}(\Delta t)} \right)^2. \quad (\text{S29})$$

Given a fixed initial Trotter error and for a target fast-forwarding time and simulation fidelity, Eq. (S23), and Eq. (S29) its simplified approximate variant, prescribe the cost function which must be surpassed before terminating the optimization loop. Specifically, the termination condition is $C_{\text{LHST}}^{\text{VFF}} \leq C_{\text{Threshold}}$, with

$$\begin{aligned} C_{\text{Threshold}} &= \frac{1}{n} \left(1 - \left(1 - \left(\frac{1}{N} \sqrt{1 - \sqrt{1 - \frac{d+1}{d}(1 - \bar{F}(T))}} - \frac{1}{\sqrt{2}}\epsilon_{\infty}^{\text{TS}}(\Delta t) \right)^2 \right)^2 \right) \\ &\approx \frac{1}{n} \left(\frac{1}{N} \sqrt{\frac{d+1}{d}(1 - \bar{F}(T))} - \epsilon_{\infty}^{\text{TS}}(\Delta t) \right)^2. \end{aligned} \quad (\text{S30})$$

Supplementary Note 3. ESTIMATION OF ENERGY EIGENVALUES

Suppose $\{\lambda_1^{\text{exact}}, \dots, \lambda_d^{\text{exact}}\}$ are the energies corresponding to the target unitary U , and $\{\lambda_1, \dots, \lambda_d\}$ estimates extracted (using for example a time-series analysis) from the approximate unitary V obtained from the VFF algorithm. Then, using the Hoffmann-Wielandt theorem, there is an ordering of the approximate energies (i.e a permutation σ) so that

$$\sum_i |e^{i\lambda_i^{\text{exact}}} - e^{i\lambda_{\sigma(i)}}|^2 \leq \|U - V\|_2^2. \quad (\text{S31})$$

Expanding the above and introducing the additional (fixed) arbitrary global phase ϕ_0 we get

$$\sum_i 2(1 - \cos(\lambda_i^{\text{exact}} - \lambda_{\sigma(i)} + \phi_0)) \leq \|U - e^{i\phi_0} V\|_2^2, \quad (\text{S32})$$

where ϕ_0 achieves $\min_{\phi} \|U - e^{i\phi} V\|_2 = \sqrt{2d(1 - \sqrt{1 - C_{\text{HST}}(U, V)})}$, as explained in Supplementary Note 2 B. Further, the upper bound can be related to the local cost function $C_{\text{LHST}}^{\text{VFF}}$.

$$\sum_i 2(1 - \cos(\lambda_i^{\text{exact}} - \lambda_{\sigma(i)} + \phi_0)) \leq \sqrt{2d(1 - \sqrt{1 - nC_{\text{LHST}}^{\text{VFF}}(U, V)})}, \quad (\text{S33})$$

For small values of the cost function $C_{\text{LHST}}^{\text{VFF}}$ the upper bound will be small resulting in $\lambda_i^{\text{exact}} \approx \lambda_{\sigma(i)}$.

Supplementary Note 4. COST FUNCTION GRADIENT DERIVATION

Here we provide the derivation of the partial derivative of $C_{\text{LHST}}^{\text{VFF}}$ with respect to θ_k in (27). To emphasize the dependence on $\boldsymbol{\theta}$, we write

$$C_{\text{LHST}}^{\text{VFF}} = 1 - F(\boldsymbol{\theta}), \quad \text{with} \quad F(\boldsymbol{\theta}) = \text{Tr}[X(U^* \otimes W D W^\dagger) |\Phi^+\rangle \langle \Phi^+| (U^T \otimes W D^\dagger W^\dagger)], \quad (\text{S34})$$

where we have $X = \frac{1}{n} \sum_{j=1}^n |\Phi^+\rangle\langle\Phi^+|_{A_j, B_j} \otimes \mathbb{1}_{\bar{A}_j, \bar{B}_j}$ with $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Taking the partial derivative of the cost function with respect to an angle θ_k gives

$$\begin{aligned} \frac{\partial C_{\text{LHST}}^{\text{VFF}}}{\partial \theta_k} = & -\text{Tr} \left[X \left(U^* \otimes \frac{\partial W}{\partial \theta_k} D W^\dagger \right) |\Phi^+\rangle\langle\Phi^+| (U^T \otimes W D^\dagger W^\dagger) \right] \\ & -\text{Tr} \left[X (U^* \otimes W D W^\dagger) |\Phi^+\rangle\langle\Phi^+| \left(U^T \otimes W D^\dagger \frac{\partial W^\dagger}{\partial \theta_k} \right) \right] \\ & -\text{Tr} \left[X \left(U^* \otimes W D \frac{\partial W^\dagger}{\partial \theta_k} \right) |\Phi^+\rangle\langle\Phi^+| (U^T \otimes W D^\dagger W^\dagger) \right] \\ & -\text{Tr} \left[X (U^* \otimes W D W^\dagger) |\Phi^+\rangle\langle\Phi^+| \left(U^T \otimes \frac{\partial W}{\partial \theta_k} D^\dagger W^\dagger \right) \right]. \end{aligned} \quad (\text{S35})$$

The eigenvector operator, W , consists of products of Pauli rotations and can be decomposed as

$$W = W_L \exp \left(-\frac{i\theta_k \sigma_k}{2} \right) W_{R'} \equiv W_L W_R \quad (\text{S36})$$

where the operators W_L and $W_{R'}$ consist of all Pauli rotations to the left and right of the σ_k rotation respectively and we have defined $W_R = \exp(-i\theta_k \sigma_k/2) W_{R'}$ for convenience. It follows that the differential of W with respect to θ_k takes the form

$$\frac{\partial W}{\partial \theta_k} = -\frac{1}{2} i W_L \sigma_k W_R, \quad (\text{S37})$$

which on substituting into Eq. (S35) gives

$$\frac{\partial C_{\text{LHST}}^{\text{VFF}}}{\partial \theta_k} = \frac{i}{2} \left(\text{Tr} \left[X (U^* \otimes W_L) [\sigma_k, \rho_1] (U^T \otimes W_L^\dagger) \right] - \text{Tr} \left[X (U^* \otimes W D W_R^\dagger) [\sigma_k, \rho_2] (U^T \otimes W_R D^\dagger W^\dagger) \right] \right), \quad (\text{S38})$$

where we have defined

$$\rho_1 = W_R D W^\dagger |\Phi^+\rangle\langle\Phi^+| W D^\dagger W_R^\dagger \quad \text{and} \quad (\text{S39})$$

$$\rho_2 = W_L^\dagger |\Phi^+\rangle\langle\Phi^+| W_L. \quad (\text{S40})$$

Eq. (27) is now obtained directly from Eq. (S38) via the following identity, which holds for any state ρ ,

$$i[\sigma_k, \rho] = e^{i\sigma_k \pi/4} \rho e^{-i\sigma_k \pi/4} - e^{-i\sigma_k \pi/4} \rho e^{i\sigma_k \pi/4}. \quad (\text{S41})$$

Supplementary Note 5. FAST FORWARDING OF HAMILTONIANS: FROM ASYMPTOTIC TIME COMPLEXITY TO THE NISQ ERA

The aim of this section is to give a brief summary of known complexity results and open questions regarding Hamiltonians that allow fast forwarding and to discuss their implications for near-term algorithms like VFF.

A Hamiltonian H is said to be fast forwardable if it can be simulated for time T with computational resources much smaller than T . The work of Atia and Aharonov [27] establishes an equivalence between fast forwarding of Hamiltonians and performing efficient energy measurements with high precision. More specifically [27], and as sketched in Fig. S1, a family of (normalised) Hamiltonians $H(n)$ that acts on an increasing number of qubits n is fast-forwardable if for all $T \leq t(n)$ there is a circuit U with polynomial size $d(n)$ such that $\|e^{-iHT} - U\|_\infty \leq \mathcal{O}(1/\text{poly}(n))$ and $t(n)$ grows asymptotically faster than $d(n)$. In particular, whenever time grows exponentially in the number of qubits $t(n) = \mathcal{O}(2^{\Omega(n)})$, Hamiltonians satisfying this definition are said to be exponentially fast-forwardable. Specific classes of Hamiltonians have been shown to be exponentially fast-forwardable [27] including commuting Hamiltonians, quadratic fermionic systems and a family of Hamiltonians related to Shor's algorithm. Recent geometric results [58] suggest that the Sachdev-Ye-Kitaev model of chaotic systems can also be fast-forwarded.

More generally, a connection between diagonalization of Hamiltonians and exponential fast forwarding has been established. Following [27], a Hamiltonian H is said to be *quantum diagonalizable* if there exists a polynomial sized diagonalizing unitary V , and polynomial sized diagonal unitary D such that $V H V^\dagger = D$. Any such Hamiltonian

admits exponential fast forwarding. However, it remains an open question to investigate if the set of fast-forwardable Hamiltonians is larger than the set of quantum diagonalizable Hamiltonians [32]. These types of questions are similar to proving tighter lower bounds for the circuit complexity of large time simulations restricted to specific classes of Hamiltonians. Further, we emphasize that fast-forwardable Hamiltonians can generate highly non-classical behaviour. For example, in [32] a diagonalizable Hamiltonian for which V is constructed out of IQP circuits gives a quantum advantage proposal for the task of energy sampling.

The no-fast-forwarding theorem has been established in several Hamiltonian models of computation. For the unknown Hamiltonian model [27], the no-fast-forwarding theorem states that there is no generic method to exponentially fast forward an arbitrary 2-sparse row computable Hamiltonian. A similar result for the query model (where access to the Hamiltonian is achieved by querying its matrix elements) implies that for every time T there exists a (normalised, row-computable sparse) Hamiltonian whose simulation with finite precision for time T requires at least $\Omega(T)$ queries. These results establish the existence of Hamiltonians for which no generic fast forwarding occurs, without a constructive way to build non-fast forwardable Hamiltonians. As a consequence, large classes of Hamiltonians that allow universality will necessarily include families for which there is no generic way of fast-forwarding. As an example, nearest-neighbour Heisenberg interactions on a lattice give rise to a universal class of Hamiltonians.

What are the implications of these results to VFF? It is highly important to emphasize that the notion of fast forwarding technically refers to a *family* of Hamiltonians acting on an increasing number of qubits n , so it can be misleading to refer to fast forwarding of a Hamiltonian with a fixed number of qubits. Thus the question at hand is whether/how the no-fast-forwarding theorems restrict the use of VFF to simulate families of Hamiltonians.

To answer this we first note that asymptotic complexity results cannot be directly applied to finite size n and finite depth d experiments. Therefore, as it is the small and intermediary scale behaviour that determines the feasibility and practical applicability of VFF experiments for the foreseeable future, it follows that the questions of whether a family of Hamiltonians is asymptotically fast-forwardable and whether VFF can be used to perform simulation beyond the coherence time are largely independent. As sketched in Fig. S1 it is possible to conceive of families of Hamiltonians that are asymptotically non-fast-forwardable and yet, for a regime of finite n , allow for simulation beyond the coherence time. Conversely, there may exist families of Hamiltonians which are fast-forwardable but do not permit simulation beyond the coherence time, due to an unfavourable scaling at low n . In this sense, the complexity-theoretic no-fast-forwarding results are not expected to place restrictions on the practical use of VFF for NISQ simulations.

From a theoretical standpoint, the bottleneck for VFF comes from the requirement that the unitaries W and D have a-priori a fixed depth. This will affect the implementation of VFF by limiting the theoretical minimum for the optimisation error ϵ_{ML} . From an implementation point of view, the limitations of VFF were addressed in the main text with the error analysis and termination condition. The finite depth cut off will introduce an error $\delta = \min_{W,D} \|e^{-iH\Delta t} - WDW^\dagger\|$, where the minimisation occurs over all those unitaries W and D for which the total depth of $WDW^\dagger$ is at most the coherence time of the hardware device considered (i.e some fixed value). Naturally, δ will be a lower bound for the machine learning error from the optimisation step of the VFF algorithm $\delta \leq \epsilon_{ML}$. The fast forwardability of a family of Hamiltonians H acting on an increasing number of qubits n will not necessarily be enough to ensure that δ is small and scales as $\mathcal{O}(1/\text{poly}(n))$ (i.e with a similar approximation error as in the definition of fast forwardability). The same holds for the possibly weaker condition of quantum diagonalisability.

However, we could expect that the error in approximating the diagonalization with a fixed depth ansatz will exhibit better large n scaling for quantum diagonalizable Hamiltonians than for those whose diagonalisation requires exponential resources. Thus the complexity-theoretic results may place fundamental limitations on the scalability of VFF for certain Hamiltonians as we reach the fault tolerant regime.

That is not to say that Hamiltonians which are not quantum diagonalizable cannot have, for small n , a good fixed-depth ansatz approximation of their diagonalisation. We expect this to be true in particular of those Hamiltonians for which the exponential scaling of resources becomes dominant for much larger n . By contrast it will be possible that the fixed-depth ansatz approximation could not be suitable for certain Hamiltonians whose diagonalisation has a high overhead in constant factors, even if it has a polynomial asymptotic scaling.

Another issue that will affect the minimal theoretically achievable error δ for VFF is the expressivity of the fixed-depth ansatz. For a given depth, can the unitaries W and D that give the minimal δ be exactly represented by some parametrisation for the hardware-efficient ansatz (or more sophisticated ones such as UCC)? Generally, this will not necessarily be the case and this will be reflected in a higher value for δ with the restriction that one must minimise over particular fixed depth structures.

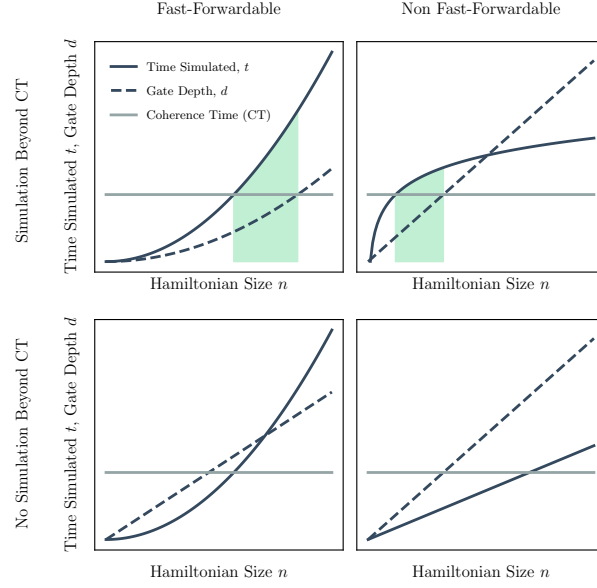


FIG. S1. Schematic highlighting the independence of the ability to fast-forward a family of Hamiltonians and the ability to use VFF to simulate a subset of those Hamiltonians beyond the coherence time. The blue lines indicate the time t that can be simulated (solid) and the corresponding gate depth d required to do so (dashed), as a function of the number of simulated qubits n . The two figures in the left column indicate fast-forwardable Hamiltonians with $t(n)$ growing asymptotically faster than $d(n)$ (i.e. the solid blue line is above the dashed blue line for large n). Correspondingly, the two figures in the right column sketch $t(n)$ and $d(n)$ for non fast-forwardable Hamiltonians. The grey line indicates the coherence time of the quantum computer, that is the gate depth after which quantum simulations are no longer possible due to the build up of errors and decoherence. The green shading indicates the region in which simulation beyond the coherence time is possible. This is the scale on which the time simulated t can be greater than the coherence time of quantum computer but the required gate depth d to perform that simulation does not exceed the coherence time (i.e. the region in which the solid blue line is above, and the dashed blue line is below, the grey line).