

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/366212677>

Importance sampling for stochastic quantum simulations

Preprint · December 2022

DOI: 10.48550/arXiv.2212.05952

CITATIONS

0

3 authors:



Oriel Kiss

University of Geneva

8 PUBLICATIONS 6 CITATIONS

SEE PROFILE



Michele Grossi

CERN

34 PUBLICATIONS 254 CITATIONS

SEE PROFILE



Alessandro Roggero

Università degli Studi di Trento

59 PUBLICATIONS 758 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



W Boson Polarization Studies for Vector Boson Scattering at LHC: from Classical Approaches to Quantum Computing [View project](#)

Importance sampling for stochastic quantum simulations

Oriel Kiss^{1,2}, Michele Grossi¹, and Alessandro Roggero^{3,4}

¹European Organization for Nuclear Research (CERN), Geneva 1211, Switzerland

²Department of Nuclear and Particle Physics, University of Geneva, Geneva 1211, Switzerland

³Physics Department, University of Trento, Via Sommarive 14, I-38123 Trento, Italy

⁴INFN-TIFPA Trento Institute of Fundamental Physics and Applications, Trento, Italy

Simulating complex quantum systems is a promising task for digital quantum computers. However, the depth of popular product formulas scales with the number of summands in the Hamiltonian, which can therefore be challenging to implement on near-term as well as fault-tolerant devices. An efficient solution is given by the stochastic compilation protocol known as qDrift, which builds random product formulas by sampling from the Hamiltonian according to the magnitude of their coefficients. In this work, we unify the qDrift protocol with importance sampling, allowing us to sample from arbitrary distributions while controlling both the bias as well as the statistical fluctuations. We show that the simulation cost can be reduced while achieving the same accuracy by considering the individual simulation cost during the sampling stage.

Moreover, we incorporate recent work on composite channel and compute rigorous bounds on the bias and variance showing how to choose the number of samples, experiments, and time steps for a given target accuracy. These results lead to a more efficient implementation of the qDrift protocol, both with and without the use of composite channels. Theoretical results are confirmed by numerical simulations performed on a lattice nuclear effective field theory.

1 Introduction

The simulation of quantum systems is one of the most promising applications for quantum computers [1]. Hence, the exponential scaling of the Hilbert space, or the infamous sign problem in Monte Carlo techniques [2], makes it a notoriously difficult task for classical devices in most cases. On the other hand, the resource requirements for quantum simulations are only subject to polynomial growth in many practical circumstances, as in the simulation of local Hamiltonians [3]. Hence, quantum computers offer a natural paradigm for Hamiltonian simulations, with numer-

ous applications in nuclear [4–6] and condensed matter physics [7–10], quantum field theory [11–13] and quantum chemistry [14–16]. For instance, quantum simulations have been applied to the computation of energy levels via quantum phase estimation [17], chemical reaction rates predictions [18], correlation functions [7, 8, 19, 20], neutrino oscillations [21, 22] and scattering experiments [23–25].

Given a Hamiltonian H written as the sum of L multi-qubit operators, typically expressed as Pauli strings, the solution of the time-independent Schrödinger equation is obtained through the exponentiation of the former. One of the most popular and straightforward techniques to compute this matrix exponential is given by product formulas, such as Trotterization and higher order Trotter-Suzuki decomposition [26–29], due to their simplicity and high performance in practice. One main drawback of product formulas is that their gate count increases proportionally to the number of summands in the Hamiltonian. Even if the asymptotic scaling is favorable, the pre-factor might be significant enough [23, 30, 31] to create problems in practice. Hence, deep circuits usually exceed the coherence time of noisy intermediate scale quantum (NISQ) [32] devices while having a significant runtime on error-corrected ones. We refer to [33] and [34] for informative reviews, where the former is theoretical, and the latter focuses on practical applications on NISQ devices.

Randomization has proven to be an important tool for improving the accuracy and efficiency of product formulas. Childs et al. [35] achieved better gate-complexity by randomizing the ordering of the terms in the Hamiltonian, while Faehrmann et al. [36] increased the order of the product formula by averaging over different time slices. More recently, Cho et al. [37] doubled the order of product formulas by introducing random corrections. Hence, stochasticity can turn coherent errors into incoherent ones, making the error scaling behave as a random walk [38–41]. Even if those improvements are considerable, none of these methods address the scaling with L , which can be a large pre-factor for many relevant situations.

For this reason, Campbell [42] introduced qDrift, a protocol to build product formulas by sampling over the coefficients, whose length does not depend specif-

Oriel Kiss: oriel.kiss@cern.ch

ically on L , but on the square of the simulation time t^2 and on the norm of the Hamiltonian λ . Chen et al. [43] improved the error bound on the bias and showed that one experiment of the qDrift protocol converges exponentially fast towards its expected value. Finally, Ouyang et al. [15] combined the advantages of qDrift and first-order Trotter formula to simulate the Hamiltonian through sparsification.

The main contributions of this paper are the generalization of the qDrift protocol for arbitrary sampling distribution and the expansion of the results from Ref. [43] to multiple qDrift executions and composite channels with multiples Trotter steps. Sampling from arbitrary distributions has numerous benefits, such as allowing for a direct reduction of the actual cost of the simulations in terms of native gates or expanding qDrift to situations where it might be difficult to sample directly from the coefficients. For instance, we propose an alternative sampling distribution, which decreases the simulation cost, such as the total CNOT count. Moreover, this paper gives a rigorous understanding of the behavior of qDrift and composite channels with multiple experiments. We show that qDrift can be efficiently parallelized on multiples devices, and we give a rigorous formula for choosing the number of qDrift samples N , experiments M , and Trotter steps r , for a given accuracy ϵ and simulation time t .

Even if we restrict ourselves to time-independent problems, we note that the present work can be directly applied to time-dependent situations through the continuous qDrift [44] protocol. We note that alternative and more refined techniques exist, using extra ancillary qubits and complex gadgets, such as qubitization and linear combination of unitaries [45–48]. They usually offer better asymptotic scaling but are more challenging to implement in practice and fall outside this paper’s scope.

The relevant background and notations are covered in Section 2, with Section 2.2 recalling the qDrift protocol. We introduce importance sampling for stochastic quantum simulations in Section 3.1, while rigorous bounds on the bias, variance, and fluctuation are shown in Section 3.2. Section 3.3 unifies the importance sampled qDrift with composite channels, mixing Trotter and qDrift product formulas. Section 4 focuses on practical applications that benefit from this general framework, such as cost reduction and Hamiltonian partitioning. Finally, numerical simulations are performed in Section 5. Rigorous proofs of the stated theorems can be found in the appendices A and B for the results and applications sections, respectively.

2 Preliminaries

In this section, we introduce the background and notations adopted throughout this paper. We denote by

$\|\cdot\|$ the spectral norm, $\|\cdot\|_1$ the trace norm and by

$$\|\mathcal{U} - \mathcal{E}\|_\diamond := \max_\rho (\|(\mathcal{U} \otimes \mathbb{1}_k - \mathcal{E} \otimes \mathbb{1}_k)\rho\|_1), \quad (1)$$

the diamond norm distance between two quantum channels \mathcal{U} and \mathcal{E} . We note that the maximum is taken over all $(n+k)$ -qubit states, where n is the dimension of \mathcal{E} and $k \geq 1$. In the remaining of this paper, we consider a time-independent n -qubit Hamiltonian, in the form of a $(2^n, 2^n)$ Hermitian matrix with the following decomposition into L summands

$$H = \sum_{l=1}^L h_l H_l, \quad (2)$$

with $\|H_l\| = 1$ and $h_l > 0$. We denote $\lambda = \sum_l h_l$ the norm of the Hamiltonian and $\Lambda := \max_l(h_l)$.

2.1 Deterministic Trotter product formulas

Given a Hamiltonian H , the first order Trotter product formula is built by exponentiating all the individual terms as

$$U(t) = \prod_l e^{-ith_l H_l}. \quad (3)$$

The usual technique for long-time simulations is to split the time into r fragments and to repeatedly apply $U(t/r)$, which are known as Trotter steps. Analytical work [35] shows that the Trotter error ϵ is upper bounded by

$$\epsilon \leq \frac{L^2 \Lambda^2 t^2}{2r} e^{\Lambda t L/r}. \quad (4)$$

Better error scaling can be achieved by considering higher order product formula, which typically requires a symmetric extension or randomization, leading to deeper circuits. Tighter error bounds with commutator scaling for general product formulas have been found by Childs et al. [49]. Despite their simplicity, Trotter formulas are performing surprisingly well and often much better than the predicted bounds, making them the default choice in many situations, including early fault tolerant quantum hardware.

Since we will be dealing with quantum channels, it is useful to recall how time evolution translates in the density matrix formalism. Given a unitary operator $U(t)$ and a density matrix ρ , the time evolution is performed as

$$\rho(t) = \mathcal{U}(t)[\rho] = U(t)\rho U^\dagger(t), \quad (5)$$

where $\mathcal{U}(t)$ is the unitary channel of $U(t)$.

2.2 The qDrift protocol

One of the main drawbacks of Trotter-Suzuki decompositions is their gate count. Hence, every term in the Hamiltonian, see Eq. (2), must be simulated sequentially, leading to deep circuits in many relevant

use cases. The resulting quantum circuits are therefore heavily affected by noise on NISQ devices and are time-consuming on fault-tolerant hardware, motivating the search for more efficient algorithms. Campbell [42] remarked that the gate count can be significantly reduced in some regimes by considering the relative importance of each term H_j in the Hamiltonian, given by the corresponding coefficient h_j . The qDrift protocol builds an approximate channel $\mathcal{E}(t; N, M)$, which is randomly constructed by sampling terms from the Hamiltonian according to the magnitude of the coefficients. We call t the simulation time, N the number of qDrift samples, and M the number of qDrift experiments. In practice, a qDrift experiment is sampled from the product distribution $p_N(\mathbf{j}) = \lambda^{-N} \prod_{k=1}^N h_{j_k}$, where each term is sampled with probability $p(j) = h_j/\lambda$, and $\mathbf{j} = (j_1, j_2, \dots, j_N)$ a multi-index, leading to

$$V_{\mathbf{j}}(t) = \prod_{k=1}^N e^{-i\tau_{j_k} H_{j_k}}, \quad (6)$$

for appropriately chosen time steps τ_{j_k} . The qDrift channel is ultimately built as the arithmetic average of the M individual experiments

$$\mathcal{E}(t; N, M)[\rho] = \frac{1}{M} \sum_m \left[V_{j_m} \rho V_{j_m}^\dagger \right], \quad (7)$$

as summarised in Algorithm 1. In the asymptotic limit of infinite experiments, this will then converge to the following average qDrift channel

$$\begin{aligned} \bar{\mathcal{E}}(t; N)[\rho] &= \mathbb{E}_{p_N} [\mathcal{E}(t; N, 1)[\rho]] \\ &= \sum_{j_1=1}^L \cdots \sum_{j_N=1}^L p_N(\mathbf{j}) V_{\mathbf{j}}(t) \rho V_{\mathbf{j}}^\dagger(t). \end{aligned} \quad (8)$$

The special case with $N = 1$ leads to

$$\bar{\mathcal{E}}(t; 1)[\rho] = \sum_{j=1}^L p(j) e^{-i\tau_j H_j} \rho e^{i\tau_j H_j}, \quad (9)$$

which we will call the deterministic qDrift channel.

The strength of each unitary is fixed to a constant $\tau_j := \tau = t\lambda/N$, which is chosen such that the qDrift channel is equal to the first order Trotter product formula, up to first order in the Taylor expansion, with an upper bound on the diamond norm given by

$$\|\mathcal{U}(t) - \bar{\mathcal{E}}(t; N)\|_\diamond \leq \frac{2\lambda^2 t^2}{N} e^{2\lambda t/N}, \quad (10)$$

We note that this bound has been improved [43] to

$$\|\mathcal{U}(t) - \bar{\mathcal{E}}(t; N)\|_\diamond \leq \frac{2\lambda^2 t^2}{N}. \quad (11)$$

The deterministic channel can be used to parallelize a Trotter product formula for small time scales, where

Algorithm 1: QDRIFT

Input: Hamiltonian $H = \sum_{l=1}^L h_l H_l$ with interaction strength $\lambda = \sum_l h_l$, $h_l > 0$ and $\|H_l\| = 1$, total simulation time t , number of samples N , number of experiments M and initial state ρ .

```

 $m \leftarrow 1$ ;
 $\mathcal{E} \leftarrow 0$ ;
while  $m \leq M$  do
   $V \leftarrow \mathbb{1}$ ;
   $n \leftarrow 1$ ;
  while  $n \leq N$  do
    sample  $j \sim p_j = h_j/\lambda$ ;
     $\tau_j \leftarrow t\lambda/N$ ;
     $V \leftarrow e^{-iH_j\tau_j} V$ ;
     $n \leftarrow n + 1$ ;
  end
   $\mathcal{E}[\rho] \leftarrow \mathcal{E}[\rho] + V\rho V^\dagger$ ;
   $m \leftarrow m + 1$ ;
end
 $\rho \leftarrow \frac{1}{M} \mathcal{E}[\rho]$ ;
Output: final state  $\mathcal{E}(t; N, M)[\rho]$ 

```

every H_l is simulated simultaneously and independently on different quantum devices, or by using different qubits of the same device. This scheme essentially trades the circuit's depth with measurements. Expectations values can then be obtained by post-processing. However, since the number of circuits grows exponentially with the number of slicing steps r , this application is impractical for long-time scales or large L , but could benefit NISQ devices which often fail in this regime due to the noise.

2.3 Importance sampling

Importance sampling is a useful technique to compute expectation values

$$\mathbb{E}_p[f(x)] = \sum_x p(x) f(x), \quad (12)$$

when the distribution $p(x)$ is difficult to sample from, or as a way to reduce the variance. Frequently used in Monte Carlo integration, the trick is to sample from an alternative, in some cases considerably easier, distribution $q(x) > 0$ and re-weight accordingly

$$\mathbb{E}_p[f(x)] = \sum_x q(x) \frac{p(x)}{q(x)} f(x) \equiv \mathbb{E}_q[\omega(x) f(x)], \quad (13)$$

with $\omega(x) := p(x)/q(x)$ is the re-weighting factor. We can easily see that this gives us an unbiased estimator of the expectation value of interest, while the variance depends on the choice of $q(x)$. With an adequate choice, the variance can be significantly reduced, leading to less expensive calculations. We guide the reader to [50] for an informative review on the topic.

3 Results

The first results presented in this section, are the application of importance sampling to the qDrift protocol and the computation of the corresponding bias, variance and fluctuation bounds. The second set of results unifies this framework with composite channels. We will begin by introducing the importance sampled qDrift.

3.1 Importance sampled qDrift

To better understand the paradigm shift, we adopt the Liouvillian representation of a unitary channel e^{-iHt}

$$\mathcal{E}(t)[\rho] = e^{iHt} \rho e^{-iHt} \equiv e^{t\mathcal{L}}(\rho) = \sum_{n=0}^{\infty} \frac{t^n \mathcal{L}^n(\rho)}{n!}, \quad (14)$$

with

$$\mathcal{L}(\rho) = i(H\rho - \rho H) = i[H, \rho]. \quad (15)$$

We first write the qDrift channel, sampled from an arbitrary distribution $q(j)$ as

$$\begin{aligned} \bar{\mathcal{E}}_q(t; 1)[\rho] &= \sum_j q(j) e^{-i\tau_j H_j} \rho e^{i\tau_j H_j} \\ &\equiv \sum_j q(j) e^{\tau_j \mathcal{L}_j}(\rho) \\ &= \left(1 + \sum_j q(j) \tau_j \mathcal{L}_j + \sum_{n=2}^{\infty} \sum_j q(j) \tau_j^n \mathcal{L}_j^n \right) (\rho). \end{aligned} \quad (16)$$

We then choose τ_j so that we match the ideal channel to linear order, that is

$$\begin{aligned} \sum_j q(j) \tau_j \mathcal{L}_j &= \frac{t}{N} \sum_j h_j \mathcal{L}_j \\ &= \frac{t\lambda}{N} \sum_j \frac{h_j}{\lambda} \mathcal{L}_j \\ &= \frac{t\lambda}{N} \sum_j p_j \mathcal{L}_j. \end{aligned} \quad (17)$$

Since the bias at the second order in t/N cannot be matched with any choice of $q(j)$, we focus on the part which is linear in time. For ease of notation, we incorporate the constant factor inside the generators as

$$\widetilde{\mathcal{L}}_j = \frac{t\lambda}{N} \mathcal{L}_j. \quad (18)$$

The expectation value of a qDrift sample is then in the right form and can be written as

$$\begin{aligned} \mathbb{E}_p[\widetilde{\mathcal{L}}_j] &= \sum_j p_j \widetilde{\mathcal{L}}_j \\ &= \sum_j q_j \frac{p_j}{q_j} \widetilde{\mathcal{L}}_j \\ &= \sum_j q_j \omega_j \widetilde{\mathcal{L}}_j \\ &= \mathbb{E}_q[\omega_j \widetilde{\mathcal{L}}_j]. \end{aligned} \quad (19)$$

Therefore, the channel can be written as

$$\begin{aligned} \bar{\mathcal{E}}_q(t; 1)[\rho] &= \sum_j q_j \exp\left(\omega_j \frac{t\lambda}{N} \mathcal{L}_j\right) \\ &= \sum_j q_j \exp\left(\frac{t}{N} \frac{h_j}{q_j} \mathcal{L}_j\right), \end{aligned} \quad (20)$$

where we used the explicit expression for p_j , and find

$$\tau_j = \frac{th_j}{Nq_j}. \quad (21)$$

We note that, contrary to the standard qDrift, the strength of each unitary τ_j is now dependent on j .

3.2 Bias, variance and fluctuation bounds

In this section, we compute the bias, concentration and fluctuation bounds for the importance sampled qDrift channel, as a function of the sampling distribution $q(j)$, simulation time t , number of samples N and number of experiments M . We remark that for $q(j) = p(j)$, we recover the usual bounds, meaning that our framework is a natural extension of the standard qDrift implementation. For the ease of notations, all minimum values of N , M and r , which are integer numbers per definition, are given in function of ϵ , t and λ , which may not lead to integer value numbers. For practical situations, we should round them up to the next integer. Borrowing the proof strategy from [43, Proposition 3.2], we will now provide an upper bound on the error of the bias. We refer to Appendix A for the complete proofs of the presented results.

Theorem 1 (Bias error bound). *Let $\mathcal{U}(t)$ be the unitary channel of a first-order Trotter product formula, $\bar{\mathcal{E}}_q(t; N)$ an average qDrift channel with importance sampling and $\omega(j) = p(j)/q(j)$ the re-weighting factor. The diamond norm distance between these two channels for $N = 1$ is then upper bounded by*

$$\|\mathcal{U}(t) - \bar{\mathcal{E}}_q(t; 1)\|_{\diamond} \leq t^2 \lambda^2 (1 + \mathbb{E}_p[\omega(j)]), \quad (22)$$

leading to the following result

$$\|\mathcal{U}(t) - \bar{\mathcal{E}}_q(t; N)\|_{\diamond} \leq \frac{t^2 \lambda^2}{N} (1 + \mathbb{E}_p[\omega(j)]) \quad (23)$$

for an arbitrary N .

We are now able to understand that importance sampling can lead to an increase in the number of qDrift samples N at fixed accuracy ϵ . In fact

$$N_q = \left\lceil \frac{t^2 \lambda^2}{\epsilon} (1 + \mathbb{E}_p[\omega(j)]) \right\rceil \geq \left\lceil \frac{t^2 \lambda^2}{\epsilon} (1 + 1) \right\rceil = N_p, \quad (24)$$

implying $N_q \geq N_p$. Therefore, the standard qDrift channel will always requires a smaller number of samples, but we will argue in the next section that the total simulation cost can still be reduced with importance sampling, without sacrificing accuracy, since we can favorise the sampling of cheaper circuits.

Now that we have generalized the error bound on the bias, we need to understand how a finite importance sampled qDrift channel concentrates around its expectation value. Hence, this will provide an estimate of M and N , for a given accuracy ϵ and simulation time t .

Theorem 2 (Concentration bound). *Let $\mathcal{E}_q(t; N, M)$ be a finite importance sampled qDrift channel on n qubits and V_j instances of the NM unitaries that make up the channel. Their concentration around their expectation value can then be upper bounded $\forall \epsilon \in [0, 4t\lambda]$ as follows*

$$\Pr \left[\left\| \frac{1}{M} \sum_{m=1}^M \prod_{k=1}^N V_{k_m} - \mathbb{E}_q[V_j]^N \right\| \geq \epsilon/2 \right] \leq 2^{n+1} \exp \left\{ - \frac{NM\epsilon^2}{11t^2\lambda^2(1 + \max_k \omega(k))^2} \right\}. \quad (25)$$

In order to guarantee an approximation error $\epsilon/2$ with probability at least $1 - \delta$, it is then sufficient to take

$$NM = 11 \frac{t^2 \lambda^2}{\epsilon^2} \left(1 + \max_k \omega(k) \right)^2 (n+1) \log \left(\frac{2}{\delta} \right). \quad (26)$$

This theorem gives us two important pieces of information. First, we learn that we can distribute the resource budget across M and N , and that, for fixed accuracy ϵ , the channel converges exponentially fast towards the deterministic qDrift with their product. Moreover, the qDrift channel can be efficiently simulated in parallel, since it concentrates exponentially fast in M . This trade-off between circuit's depth for an increase in measurements has some advantages, such as reducing hardware errors on NISQ devices and shorter real-time simulation on fault-tolerant ones. Secondly, we are now aware that the qDrift results present in the literature also hold for any distribution $q(j)$, as long as their ratio is similar, i.e., if $\omega(j) \approx 1$. This enables the design of alternative distributions, which for example, produce less expensive circuits or are easier to sample from. For instance, these results can be directly transported to the continuous qDrift protocol [44], where the continuous distribution is replaced with an easier one. In [44] the authors already

proposed to use a more readily available distribution using norm upperbounds instead of norms. Our results here makes it easier to characterize both the bias and the variance change induced by such a choice. We will see in the next section how to choose $q(j)$ to obtain a guaranteed reduction in the simulation cost.

Finally, we compute the bound of the expected fluctuations around the true evolution, following the strategy of [43, Proposition 3.4].

Corollary 1 (Fluctuation bound). *Let H be a n -qubit Hamiltonian, $q(j)$ an arbitrary distribution, t the simulation time, N a fixed number of qDrift samples, and M a fixed number of qDrift experiments. Set $\mathcal{U}_H[\rho] = U_H \rho U_H^\dagger$ (with $U_H = e^{-iHt}$) and take the importance sampled qDrift channel $\mathcal{E}_q(t; N, M)$. We have*

$$\begin{aligned} \mathbb{E} [\| \mathcal{E}_q(t; N, M) - \mathcal{U}_H \|_\diamond] &\leq 2 \frac{t^2 \lambda^2}{N} (1 + \mathbb{E}_p[\omega]) \\ &+ \alpha \frac{nt\lambda}{NM} \left(1 + \max_k \omega(k) \right) \\ &+ \alpha \sqrt{\frac{n}{NM}} t\lambda \left(1 + \max_k \omega(k) \right), \end{aligned} \quad (27)$$

with α being a numerical constant depending on H .

We now have a better understanding of how to choose N and M for a particular distribution $q(j)$ and desired expected accuracy, in diamond norm, ϵ . In particular, if we choose N to control the bias to ϵ/κ , with $\kappa > 1$, as follows

$$N = 2\kappa \frac{t^2 \lambda^2}{\epsilon} (1 + \mathbb{E}_p[\omega]), \quad (28)$$

we then need to choose M so that

$$\alpha t\lambda \left(1 + \max_k \omega(k) \right) \left(\frac{n}{NM} + \sqrt{\frac{n}{NM}} \right) \leq \frac{\kappa - 1}{\kappa} \epsilon. \quad (29)$$

Using the choice for N above, we have

$$t\lambda \left(1 + \max_k \omega(k) \right) = \sqrt{N \frac{\epsilon}{2\kappa}} \frac{1 + \max_k \omega(k)}{\sqrt{1 + \mathbb{E}_p[\omega]}}, \quad (30)$$

resulting in the condition

$$\alpha \sqrt{\frac{n}{2M}} \frac{1 + \max_k \omega(k)}{\sqrt{1 + \mathbb{E}_p[\omega]}} \left(1 + \sqrt{\frac{n}{NM}} \right) \leq (\kappa - 1) \sqrt{\frac{\epsilon}{\kappa}}. \quad (31)$$

Since $N \geq 1$, for $M \geq n$ we obtain

$$M = \frac{n}{\epsilon} \frac{2\alpha^2 \kappa}{(\kappa - 1)^2} \frac{(1 + \max_k \omega(k))^2}{1 + \mathbb{E}_p[\omega]}. \quad (32)$$

The parameter κ can be chosen to reduce N as much as possible while not making M diverge. In general, for a given κ and fixed choice for the distribution $q(j)$, we have the following asymptotic scaling

$$N = \mathcal{O} \left(\frac{t^2 \lambda^2}{\epsilon} \right) \quad M = \mathcal{O} \left(\frac{n}{\epsilon} \right). \quad (33)$$

We then see that, thanks to the concentration bound in Theorem 2 depending on NM , the number of qDrift experiments scales better than $\mathcal{O}(1/\epsilon^2)$ that one would naively expect from shot noise.

3.3 Composite channels

We recall that the main goal of this paper is to reduce the actual implementation cost of random product formulas when running on quantum hardware. We remark that our framework can be naturally embedded in the context of composite channels [51–53], and use the deterministic nature of Trotter product formula to further reduce the cost. Following [51], we will first introduce composite channels, unify them with our importance sampling scheme, and compute the relevant bounds on the bias, variance and fluctuation.

Given a partition of the Hamiltonian, a composite channel is a composition of channels, which are used to simulate the different terms in the partition. For simplicity, we will only consider the case where H is

split into two parts $H = A + B$ with decompositions

$$A = \sum_{i \in I^A} a_i A_i \quad B = \sum_{i \in I^B} b_i B_i, \quad (34)$$

with $\|A_i\| = \|B_i\| = 1$ and I^A and I^B set of indices for each part, and use a deterministic formula for A and a stochastic one for B . To better understand this paradigm, we first perform an outer first order Trotter decomposition

$$e^{itH} \rho e^{-itH} := e^{itA} e^{itB} \rho e^{-itB} e^{-itA} + E_{A,B}(t), \quad (35)$$

where $E_{A,B}$ is the error term. We then take $\tilde{\mathcal{U}}_A(t)$ to be an approximation to the unitary channel $\mathcal{U}_A(t)[\rho] = e^{itA} \rho e^{-itA}$ performing evolution under A (we consider here a first order product formula) and $\mathcal{E}_q^B(t; N, M)$ the importance sampled qDrift channel for the B term. If we define the composite channel as $\Lambda_q(t; N, M) = \tilde{\mathcal{U}}_A(t) \circ \mathcal{E}_q^B(t; N, M)$, and its corresponding average channel $\bar{\Lambda}_q(t; N)$, one can show that its diamond norm distance from the ideal channel \mathcal{U}_H can be bounded as follows (cf. [51])

$$\begin{aligned} \epsilon &:= \|\mathcal{U}_H(t) - \bar{\Lambda}_q(t; N_B)\|_\diamond \\ &\leq \|\mathcal{U}_A(t) - \tilde{\mathcal{U}}_A(t)\|_\diamond + \|\mathcal{U}_B(t) - \bar{\mathcal{E}}_q^B(t; N_B)\|_\diamond + \|E_{A,B}(t)\|_\diamond \\ &\leq t^2 \left(\sum_{i < j} a_i a_j \| [A_i, A_j] \| + \frac{1}{2} \sum_{ij} a_i b_j \| [A_i, B_j] \| + \frac{\lambda_B^2 (1 + \mathbb{E}_p[\omega(j)])}{N_B} \right). \end{aligned} \quad (36)$$

If we split the total time into r segments and we use the union bound as usual, we find

$$\left\| \mathcal{U}_H(t) - \bar{\Lambda}_q\left(\frac{t}{r}; N_B\right) \right\|_\diamond^{or} \leq \frac{t^2}{r} \left(\sum_{i < j} a_i a_j \| [A_i, A_j] \| + \frac{1}{2} \sum_{ij} a_i b_j \| [A_i, B_j] \| + \frac{\lambda_B^2 (1 + \mathbb{E}_p[\omega(j)])}{N_B} \right). \quad (37)$$

Apart from the use of a general importance sample qDrift, and the use of the improved bound Eq. (11) obtained in [43] for the error in the qDrift channel, this is the same result obtain already in [51]. Where we depart from their scheme is in the fact that we consider the possibility that evolution under different terms in the expansion could have a different gate cost. If we denote by C_j^A and C_j^B the cost of the term j in either A or B , we can bound the total cost of the composite channel as follows

$$\begin{aligned} C(\epsilon, t) &\leq r \left(\sum_{l \in I^A} C_l^A + N_B \max_{j \in B} (C_j^B) \right) \\ &\leq \left(\sum_{l \in I^A} C_l^A + N_B \max_{j \in B} (C_j^B) \right) \frac{t^2}{\epsilon} \left(\sum_{i < j} a_i a_j \| [A_i, A_j] \| + \frac{1}{2} \sum_{ij} a_i b_j \| [A_i, B_j] \| + \frac{\lambda_B^2 (1 + \mathbb{E}_p[\omega(j)])}{N_B} \right) \end{aligned} \quad (38)$$

The average cost instead can be estimated as

$$\mathbb{E}_q[C(\epsilon, t)] \leq (C_{tot}^A + N_B \mathbb{E}_q[C^B]) \frac{t^2}{\epsilon} \left(\Gamma_{\text{comm}}^{A,B} + \frac{\lambda_B^2 (1 + \mathbb{E}_p[\omega(j)])}{N_B} \right), \quad (39)$$

where we have denoted by $C_{tot}^A = \sum_{l \in I^A} C_l^A$ the total cost for A and by $\Gamma_{\text{comm}}^{A,B}$ the contribution containing the commutators among the A terms and between the A and B partitions.

As was pointed out already in [51], the number of

qDrift samples N_B per segment is now a completely

free parameter. Following the same strategy adopted there, namely finding explicitly the minimum of the cost, one obtains an optimal number of samples as

$$N_B = \lambda_B \sqrt{\frac{1 + \mathbb{E}_p[\omega(j)]}{\mathbb{E}_q[C^B]} \frac{C_{tot}^A}{\Gamma_{comm}^{A,B}}}. \quad (40)$$

With this choice, the expected cost becomes

$$\mathbb{E}_q[C(\epsilon, t)] \leq \frac{t^2}{\epsilon} \left(\sqrt{\Gamma_{comm}^{A,B} C_{tot}^A} + \lambda_B \sqrt{\mathbb{E}_q[C^B] (1 + \mathbb{E}_p[\omega(j)])} \right)^2. \quad (41)$$

Unfortunately we cannot use directly the result for the concentration bound in Theorem 2 since r compositions of the channel $\Lambda_q(t; N, M)$ will require M^r experiments to implement. It is thus convenient to introduce another channel defined as

$$\Omega_q(t; N, M, r)[\rho] = \frac{1}{M} \sum_{m=1}^M \Lambda_q^m \left(\frac{t}{r}; N, 1 \right)^{\circ r} [\rho] \quad (42)$$

obtained by averaging M channels composed r times each. The superscript m in Λ_q^m is used to indicate that for every experiment indexed by m the channel uses a different sample of Nr indices. Note that in the limit of infinite experiments $\bar{\Omega}_q(t; N, r) = \bar{\Lambda}_q \left(\frac{t}{r}; N_B \right)^{\circ r}$ and we can still use Eq. (37) to control the bias. For the fluctuations around the average we instead have

Theorem 3 (Concentration bound for composite channels). *Let $\Omega_q(t; N, M, r)$ be a composite channel on n qubits for the Hamiltonian $H = A + B$ employing an approximate unitary $\tilde{U}_A \approx e^{-i \frac{t}{r} A}$ for the time evolution under the term A and total time t/r , a finite importance sampled qDrift channel $\mathcal{E}_q(t; N, 1)$ to approximate evolution under B and r steps using unitaries $W_j = e^{-i \frac{\tau_j}{r} B_j}$ with $\tau_j = (tb_j)/(Nq_j)$. Its concentration around its expectation value can be upper bounded $\forall \epsilon \in [0, 4t\lambda_B]$ as*

$$\Pr \left[\left\| \frac{1}{M} \left[\prod_{s=r}^1 \left(\tilde{U}_A \prod_{k=N}^1 W_{j_{ksm}} \right) \right] - \left(\tilde{U}_A \mathbb{E}[W]^N \right)^r \right\| \geq \frac{\epsilon}{2} \right] \leq 2^{n+1} \exp \left\{ - \frac{NM r \epsilon^2}{11 t^2 \lambda_B^2 (1 + \max_k \omega(k))^2} \right\}. \quad (43)$$

In order to guarantee an approximation error $\epsilon/2$ with probability at least $1 - \delta$, it is then sufficient to take

$$NM r = 11 \frac{t^2 \lambda_B^2}{\epsilon^2} \left(1 + \max_k \omega(k) \right)^2 (n+1) \log \left(\frac{2}{\delta} \right). \quad (44)$$

Using the new version of the concentration bound Theorem 3 we can also provide an estimate for the expected error of the composite channel.

Corollary 2 (Fluctuation bound for composite channels). *Let $H = A + B$ be a n -qubit Hamiltonian with decomposition as in Eq. (34), $q(j)$ an arbitrary distribution, t the simulation time, N a fixed number of qDrift samples, and M a fixed number of qDrift experiments. Take $\mathcal{U}_H(t)[\rho] = U_H(t) \rho U_H^\dagger(t)$ (with $U_H(t) = e^{-iHt}$), $\tilde{U}_A(t)$ a first order Trotter approximation of the channel $\mathcal{U}_A(t)$, $\mathcal{E}_q^B(t; N, M)$ the importance sampled qDrift channel for the B term and $\Omega_q(t; N, M, r)$ the importance sampled composite channel. We then have*

$$\begin{aligned} \mathbb{E} [\| \Omega_q(t; N, M, r) - \mathcal{U}_H \|_\diamond] \leq & 2 \frac{t^2}{r} \left(\Gamma_{comm}^{A,B} + \frac{\lambda_B^2}{N} (1 + \mathbb{E}_p[\omega]) \right) \\ & + \alpha \frac{nt\lambda_B}{NM r} \left(1 + \max_k \omega(k) \right) \\ & + \alpha \sqrt{\frac{n}{NM r}} t \lambda_B \left(1 + \max_k \omega(k) \right), \end{aligned} \quad (45)$$

where the parameter

$$\begin{aligned} \Gamma_{comm}^{A,B} = & \sum_{i < j} a_i a_j \| [A_i, A_j] \| \\ & + \frac{1}{2} \sum_{ij} a_i b_j \| [A_i, B_j] \|, \end{aligned} \quad (46)$$

contains the dependence on commutators.

If we now want to ensure the expected error to be less than ϵ we can take a number of steps given by

$$r = 2\kappa \frac{t^2}{\epsilon} \left(\Gamma_{comm}^{A,B} + \frac{\lambda_B^2}{N} (1 + \mathbb{E}_p[\omega]) \right), \quad (47)$$

for some $\kappa > 1$ together with $N = N_B$ from Eq. (40) and a number of experiments given by

$$M = \mu_q \frac{n}{\epsilon} \frac{2\alpha^2 \kappa}{(\kappa - 1)^2} \frac{\lambda_B}{\sqrt{C_{tot}^A}}, \quad (48)$$

where the dependence on $q(j)$ is absorbed in

$$\mu_q = \frac{(1 + \max_k \omega(k))^2 \sqrt{\mathbb{E}_q[C^B] / (1 + \mathbb{E}_p[\omega])}}{\sqrt{\Gamma_{comm}^{A,B} C_{tot}^A + \lambda_B \sqrt{(1 + \mathbb{E}_p[\omega]) \mathbb{E}_q[C^B]}}}. \quad (49)$$

A similar derivation to the one carried out here could be used for more general cases, discussed already in [51], when one employs a higher-order Trotter formula for the outer breakup in Eq. (35) or in the simulation of the evolution under A as well as protocols where the latter is implemented using qubitization. Further improvements obtained by using randomization [35, 36, 43] could also be accommodated.

Having unified the composite channel framework from [51] with the previous result about importance sampling described in this paper, we will now look into more concrete applications.

4 Applications

4.1 Reduction of the simulation cost

The main idea of this paper is to obtain a computational advantage by choosing the probability distribution $q(j)$ in order to reduce the total simulation cost. For instance, we assign a weight $C_j > 0$ to each term H_j , representing the number of resources required for its simulation, and choose

$$q_c(j) = \frac{h_j}{C_j \lambda_c}, \quad \lambda_c = \sum_l \frac{h_l}{C_l}. \quad (50)$$

We will denote this sampling strategy with the subscript c , while remarking that the standard qDrift protocol is recovered when the cost is constant $C_j = C$. This framework can be subsumed under the umbrella of importance sampling where the goal is a reduction in the integrated computational cost instead of the variance. The choice of the cost C_j is then dictated by the restrictions from the hardware, and is particularly advantageous if the distribution of the cost has a large variance. We first show that the expected cost of one qDrift sample is always lower when sampling from $q_c(j)$ with respect to $p(j)$, using Jensen's inequality. The complete derivation of the following results can be found in Appendix B.

Lemma 1 (Jensen's inequality [54]). *Let X be an integrable random variable and $\varphi(x) : \mathbb{R} \rightarrow \mathbb{R}$ a convex function. We then have the following inequality:*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)] \quad (51)$$

Corollary 3. *The expected cost of an importance sampled qDrift channel with $N = 1$ samples and $q(j) = q_c(j)$ is always lower than for the standard qDrift protocol, i.e., we have*

$$\mathbb{E}_{q_c}[C] \leq \mathbb{E}_p[C]. \quad (52)$$

Even if, on average, one qDrift sample is cheaper when sampling from this alternative distribution, this alone might not be enough to claim a reduction in the total simulation cost. This is because, due to Theorem 1, the standard qDrift channel needs less samples at fixed accuracy than an importance sampled one. However, we can show that the cost reduction holds in general, as formulated in the following theorem.

Theorem 4 (Cost reduction - pure qDrift). *Let $N_p(\epsilon, t)$ and $N_{q_c}(\epsilon, t)$ be the number of qDrift samples for the two distributions $p(j) = h_j/\lambda$ and $q_c(j) = h_j/(\lambda_c C_j)$ for a given target precision ϵ and propagation time t . The expected cost of the important sampled qDrift channel is then always smaller than the standard one*

$$N_{q_c}(\epsilon, t) \mathbb{E}_{q_c}[C] \leq N_p(\epsilon, t) \mathbb{E}_p[C]. \quad (53)$$

The number of experiments is instead increased as

$$M_{q_c}(\epsilon) = M_p(\epsilon) \frac{(1 + \mathbb{E}_p[1/C] \max_j C_j)^2}{1 + \mathbb{E}_p[1/C] \mathbb{E}_p[C]}, \quad (54)$$

and independent on the total evolution time t .

We therefore see that, for any simulation using pure qDrift, the importance sampling procedure described in this work guarantees a saving in the cost at the price of increasing the number of independent experiments. A similar saving can also be shown to hold when employing composite channels.

Theorem 5 (Cost reduction - composite channel). *Let $C_p(\epsilon, t)$ and $C_{q_c}(\epsilon, t)$ be the expected cost to implement the composite channels $\Omega_p(t; N, M, r)$ and $\Omega_{q_c}(t; N, M, r)$ using two distributions $p(j) = h_j/\lambda$ and $q_c(j) = h_j/(\lambda_c C_j)$ for a given target precision ϵ and propagation time t . Then the following holds*

$$C_{q_c}(\epsilon, t) \leq C_p(\epsilon, t). \quad (55)$$

The number of experiments is instead increased, $M_{q_c}(\epsilon) \geq M_p(\epsilon)$ but retaining the same scaling with error ϵ and system size n and also independent on t .

We note that the specific definition of the cost is specific to each application and hardware. For example, on NISQ devices, the cost is dominated by entangling two-qubit gates, and we might then take their number to define the cost C . This choice automatically considers the structure of the device, such as its connectivity and the particular application. For example, when simulating physical systems in second quantization, the importance sampling scheme proposed here will give a lower probability to terms with large Jordan-Wigner strings [55]. On the other hand, for applications involving error-corrected devices, one would be tempted to choose the number of T gate instead to define the cost C , as they typically take time to be fabricated and are the main bottleneck in the fault-tolerant regime [56]. The situation is, however, less straightforward in this case since the decomposition in terms of T -gates can depend, in general, on the choice of sampling probability $q(j)$. For our cost reduction scheme to work, we have made the common assumption that the time evolution under the individual terms H_j can be fast-forwarded; a typical example is when H_j are tensor products of Pauli operators. In these cases, the T cost is directly associated with the implementation of a single qubit z -rotation with an angle determined by the time step τ_j and therefore on $q(j)$ itself (cf. Eq. (21)) making it difficult to obtain a good candidate for the coefficients C_j . One possibility would be to empirically optimize the sampling distribution $q(j)$, for example, using a Monte Carlo approach or genetic optimization [57], in order to obtain as many time steps τ_j as possible equal to integer multiples of $\pi/8$ while preventing the average weight

$\mathbb{E}_p[\omega(j)]$ to grow too much. There is, however, a different setup where the importance sampling strategy from Eq. (50) could be employed directly to reduce the overall T count. One can take a decomposition where the terms H_j forming the Hamiltonian are still fast-forwardable but allow a decomposition in Clifford+T gates whose complexity does not depend on the time (see e.g. [58] for possible candidates). In this case, sampling from $q_c(j)$ will also guarantee a cost reduction.

4.2 Choice of the partitioning

Composite channels offer great versatility by combining deterministic and random product formulas, but their performance greatly depends on the adopted scheme to partition the Hamiltonian. By inspecting the optimized expression of the expected cost in Eq. (41) one can already see that a composite channel might be helpful in situations where λ_B is very small (as already noticed in [51]). For instance, even in the particular case where the terms in B all commute with each other, if λ_B is sufficiently small, the cost depends directly only on the cost of implementing the terms in A with the same pre-factor, we would have for a direct first-order Trotter simulation. Thanks to our ability to directly take into account the cost of individual terms, we can also see that another use case is whenever $\mathbb{E}_p[C^B] \ll C_{tot}^B$ which is possible in situations where the cost of individual terms within B have a significant variation. In addition, for these situations, the distribution $q_c(j)$ from Eq. (50) guarantees a further reduction of the expected cost at the expense of an increase in the required number M of experiments that need to be carried out.

These properties frequently arise in simulations of Effective Field Theories, where higher-order corrections to the interaction are suppressed. In this context, a particularly convenient situation is when an accidental symmetry forces some low-energy constants to take unnaturally small values. In nuclear physics, for instance, typical examples of this phenomenon are the $SU(4)$ Wigner symmetry in systems composed by neutrons and protons [59, 60] and its generalization to $SU(16)$ in the presence of hyperons [61]. For simulations of low energy reactions with nuclei, one could then consider the $SU(4)$ symmetric potential in the deterministic part (as worked out e.g., in [23]) and add symmetry-breaking terms in a stochastic manner using a qDrift channel. Another example would be the inclusion of effective range effects, absent in purely contact interactions, to improve the accuracy in bulk neutron matter [62] or to provide the required stability to medium mass nuclei [63, 64]. For situations where the physically relevant value of λ_B is not sufficiently small to allow for considerable savings in cost, it could also be possible to rely on extrapolation techniques, e.g., eigenvector continuation [65], to

study the system at reduced values of λ_B followed by an extrapolation.

More generally, one can also employ more direct optimizations of the splitting by attempting to minimize the cost directly, possibly at the same time as the optimization of the weights in the importance sampling distribution to be used for the stochastic portion of the algorithm. To this end, schemes like the probabilistic partitioning scheme introduced in Ref. [51] could prove extremely valuable in enabling substantial savings for any given Hamiltonian one is interested in.

5 Numerical Simulation

This section presents an application of importance sampling and composite channels for the quantum simulation of a model inspired by a pionless lattice effective field theory [66], in particular, a simple toy model for a triton introduced in [23]. We take $A = 2$ dynamical nucleons together with a static one (infinite mass) fixed on the first site of a 2x2 lattice with periodic boundary conditions. We will consider the static nucleon to be the proton while the two dynamical ones will be neutrons in two different spin states ($N_f = 2$). This model is simple enough to be easily simulated, yet contains much of the leading order contributions to the interaction and can thus provide valuable information about light nuclei and their response functions [67]. The model is equivalent to a two dimensional ($d = 2$) Hubbard model with a kinetic hopping term

$$H_{kin} = -t \sum_{f=1}^{N_f} \sum_{\langle i,j \rangle} c_{i,f}^\dagger c_{j,f}, \quad (56)$$

together with two and three body interactions

$$\begin{aligned} H_{int} = & U \sum_{i=1}^{N_f} \sum_{f,f'} n_{i,f} n_{i,f'} + V \sum_{i=1}^{N_f} \sum_{f < f' < f''} n_{i,f} n_{i,f'} n_{i,f''} \\ & + U \sum_{f=1}^{N_f} n_{1,f} + V \sum_{f < f'}^{N_f} n_{1,f} n_{1,f'}, \end{aligned} \quad (57)$$

with U the two-body interaction strength and V the three-body one. We recall that the fermionic operator $c_{i,f}$ destroys a particle of the species f on site i , $c_{i,f}^\dagger$ is the corresponding creation operator and $n_{i,f} = c_{i,f}^\dagger c_{i,f}$ the number operator. The Hamiltonian becomes particularly simple in first quantization because of the small size of the lattice. By using two qubits to encode the position of each nucleon using the following encoding strategy

$$|1\rangle \equiv |00\rangle \quad |2\rangle \equiv |01\rangle \quad |3\rangle \equiv |10\rangle \quad |4\rangle \equiv |11\rangle, \quad (58)$$

this model can be expressed in the Pauli basis as

$$\begin{aligned}
H = & 8t + \frac{3U}{4} + \frac{V}{16} - 2t \sum_{i=1}^4 X_i \\
& + \frac{V}{16} \left(\sum_{i < j < k} Z_i Z_j Z_k + Z_1 Z_4 + Z_2 Z_3 \right) \\
& + \left(\frac{U}{4} + \frac{V}{16} \right) \cdot \left(\sum_i Z_i + Z_1 Z_2 + Z_1 Z_3 \right. \\
& \left. + Z_2 Z_4 + Z_3 Z_4 + Z_1 Z_2 Z_3 Z_4 \right),
\end{aligned} \tag{59}$$

where X_k, Y_k, Z_k are the corresponding Pauli matrices acting on qubit k . More details about the conversion can be found in the original work [23].

Instead of using realistic coefficients from experiments, we split the first quantized Hamiltonian into two parts A and B , and define our Hamiltonian as

$$H^{(j)} = a \sum_{i \in I^A} A_i^{(j)} + b \sum_{i \in I^B} B_i^{(j)}, \tag{60}$$

where the subscript j denotes different splitting strategies and I^X a set of indices for X . We choose uniform coefficients inside each term of the partition in order to have a better control over the error bounds, as well as two different Hamiltonian models. The first is defined through the following separate contributions

$$\begin{aligned}
A^{(0)} = & \sum_{k=1}^4 X_k + Z_1 Z_4 + Z_2 Z_3 \\
& + \sum_{1 \leq i < j < k \leq 4} Z_i Z_j Z_k
\end{aligned} \tag{61}$$

$$\begin{aligned}
B^{(0)} = & \sum_{k=1}^4 Z_k + Z_1 Z_2 + Z_1 Z_3 \\
& + Z_2 Z_4 + Z_3 Z_4 + Z_1 Z_2 Z_3 Z_4.
\end{aligned} \tag{62}$$

The second one is instead given as

$$\begin{aligned}
A^{(1)} = & \sum_{k=1}^4 X_k + Z_1 + Z_1 Z_4 \\
& + Z_2 Z_3 + Z_2 Z_4 + Z_1 Z_2 Z_4
\end{aligned} \tag{63}$$

$$\begin{aligned}
B^{(1)} = & \sum_{k=2}^4 Z_k + Z_1 Z_2 + Z_1 Z_3 + Z_3 Z_4 + Z_1 Z_2 Z_3 \\
& + Z_2 Z_3 Z_4 + Z_1 Z_3 Z_4 + Z_1 Z_2 Z_3 Z_4.
\end{aligned} \tag{64}$$

For simplicity, we will set $a = 1$ and express everything, i.e., the coefficient b and simulation time t in units of a .

The first Hamiltonian follows the perturbation theoretical approach, where A is the simpler model H in

generator	cost	generator	cost
X_k	0.1	$Z_1 Z_2$	6
Z_k	0.1	$Z_3 Z_4$	6
$Z_1 Z_4$	2	$Z_1 Z_2 Z_3 Z_4$	6
$Z_2 Z_4$	2	$Z_1 Z_3 Z_4$	8
$Z_2 Z_3$	2	$Z_1 Z_2 Z_3$	8
$Z_1 Z_2 Z_4$	4	$Z_1 Z_3$	10
$Z_2 Z_3 Z_4$	4		

Table 1: Implementation cost for the different generators appearing in the two considered Hamiltonians.

	model 0	model 1
$\sum_{i \in I^A} C_i$	28.4	10.5
$\sum_{i \in I^B} C_i$	30.4	48.3
$\mathbb{E}_{q_c}[\omega^B]$	15.43	15.02
$\mathbb{E}_p[C^B]$	3.38	4.83
$\mathbb{E}_{q_c}[C^B]$	0.22	0.32
$N_p \mathbb{E}_p[C^B] \cdot \frac{\epsilon}{t^2 \lambda^2}$	6.76	9.66
$N_{q_c} \mathbb{E}_{q_c}[C^B] \cdot \frac{\epsilon}{t^2 \lambda^2}$	3.6	5.15

Table 2: Expectation value of the cost of simulating the two terms in the partition A and B for the two different models $j = 0, j = 1$, using either first order Trotterization or (importance sampled) qDrift.

the $U = -4V$ configuration, where most of the coefficients are zero, and B describes a small perturbation making the system more realistic. A deterministic first order Trotter product formula is used to simulate the bulk of the system A , while a qDrift channel handles the contribution from the perturbation B .

The second Hamiltonian is chosen instead in order to minimize the expected cost of each circuit and exemplifies the role of importance sampling. The A part contains the terms that require small resources to be simulated on the quantum devices, i.e., which have a small cost C in terms of number of two-qubits CNOT gates, while B is composed of the most difficult terms, with some easy single-qubit rotations in order to diminish the expected cost.

We assume a constant cost of 0.1 for the one-qubit operations and of one for every CNOT gate appearing in the multi-qubit terms, after transpilation. We will assume a linear connectivity 1423 and neglect any compilation optimization obtained through gate cancellation from neighboring operations for individual sampled circuits. The cost of each generator is displayed in Table 1, while the expected cost for the two different systems for Trotter, plain and importance sampled composite channels, can be seen in Table 2. More precisely, we compute the deterministic cost, the expectation value of the cost per step, and the total

cost at fixed accuracy ϵ and time t . We can notice that importance sampling is able to diminish the total simulation cost by a factor of two compared to the plain qDrift and that the use of qDrift channel allows a reduction of an order of magnitude in cost compared to Trotterization.

We perform quantum simulation of both models $j = 0$ and $j = 1$ on a noiseless simulator for two different simulation times $t = 0.05a$ and $t = 0.1a$ and different values of the strength coefficient $b \in [0.005, 0.5]$, using a composite channel $\Omega_q(t; N = 1, M, r = 1)$. Part A is evolved using one step of the first order Trotter formula, while part B is evolved using a qDrift channel with $N = 1$ step. We consider two different sampling distributions: the standard one $q = p$ and $q = q_c$ with the cost defined as above. The diamond norm, see Eq. (1), is displayed in Figure 1 against b/a , where the top line shows the computations for the model $j = 0$ and the bottom one for the model $j = 1$. The diamond distance to the ideal simulation is displayed in dashed black for the full Trotterized channel, in red for the deterministic qDrift $\bar{\mathcal{E}}(t; N = 1)$ channel, in blue (yellow) dashed line for the standard ($q = p$) composite channel with $M = 1$ ($M = 10$) and in violet (green) dots for the importance sampled ($q = q_c$) composite channel with $M = 1$ ($M = 10$). Finally, panels (c) and (f) on the last column show the histograms of the cost from the sampled circuits obtained with a (importance sampled) composite qDrift channel with $M = 50$, while the error bars correspond to one standard deviation over 50 repetitions of the whole experiment. We observe that the errors of the different composite channels are close to each others and are matching the full Trotter channel, except at $b \approx 0.05$ where Trotterization is slightly better. Most importantly, we obtain a reduction in the required amount of resources of a factor 1.8 (2) for the model $j = 0$ with the (importance sampled) composite channel and 3.8 (5) for model $j = 1$, without sacrificing precision, compared to the full Trotterization. Moreover, the cost for the simulation of the B part alone is reduced of one (two) order of magnitude with the pure (importance sampled) qDrift channel compared to direct first order Trotterization, see also Table 2.

6 Conclusions

This work generalizes past results on concentration bounds of random product formulae [43], allowing for the introduction of a generic importance sampling scheme. We provide a rigorous characterization of the protocol's bias and statistical fluctuations, extending previous results on qDrift to quantify the algorithm's sample complexity. In particular, we showed that a qDrift channel concentrates exponentially fast in NM around its expectation value, where N is the number of qDrift samples and M the number of experiments, thus allowing an efficient allocation between quan-

tum resources (controlled by N) and classical ones (controlled by M). These results allow for parallel controllable simulations of a qDrift channel on multiple quantum devices while keeping each circuit shallow enough to mitigate the noise and run time in the NISQ and fault-tolerant era, respectively. Moreover, by incorporating the individual implementation cost for evolution under each of the Hamiltonian terms in a suitable sampling distribution $q_c(j)$, we show that importance sampling obtains a guaranteed total cost reduction, in terms of hardware native cost such as the number of CNOT gates, leading the way to a more straightforward implementation of qDrift in the near term. Under reasonable assumptions, similar cost savings can also be obtained when one wants to reduce the number of T gates, which is beneficial for error-corrected devices.

In addition, we extend our result to consider composite channels [51], where the Hamiltonian is partitioned into A and B , which are simulated separately with a deterministic method (such as a Trotter-Suzuki product formula) and a qDrift channel respectively. We show that the same importance sampling distribution $q_c(j)$ can also be employed in these cases to reduce the quantum resources required for the implementation. In the typical situation where evolution under different terms in the total Hamiltonian incurs different implementation costs, the explicit inclusion of this information in our construction opens the possibility to improve further the savings that can be achieved by using composite channels by optimizing the partitioning schemes [51]. In general, it could be profitable to handle Hamiltonian terms, which are expensive but have small norms in a stochastic way using qDrift. We propose different concrete applications within nuclear physics that may benefit from such an approach from an Effective Field Theory perspective.

Finally, the theoretical results are illustrated through numerical simulations of a simple model of a triton on a (2×2) lattice in first quantization. We find that a significant cost reduction (5x) can be obtained using composite channels and importance sampling without sacrificing accuracy. The approach is robust for different strength values between the two parts of the partition. Due to the quadratic scaling with simulation time of the qDrift part of the scheme, the protocol is particularly well suited for applications that require relatively short evolution times, e.g., protocols to measure observables by signal processing [68–70].

The example importance sampling strategy described in this work has the advantage of being simple and providing a guaranteed cost reduction, but it might not be optimal for some specific problems. We leave for future work exploration of more direct numerical optimizations of the sampling distribution and the partitioning scheme for constructing composite channels.

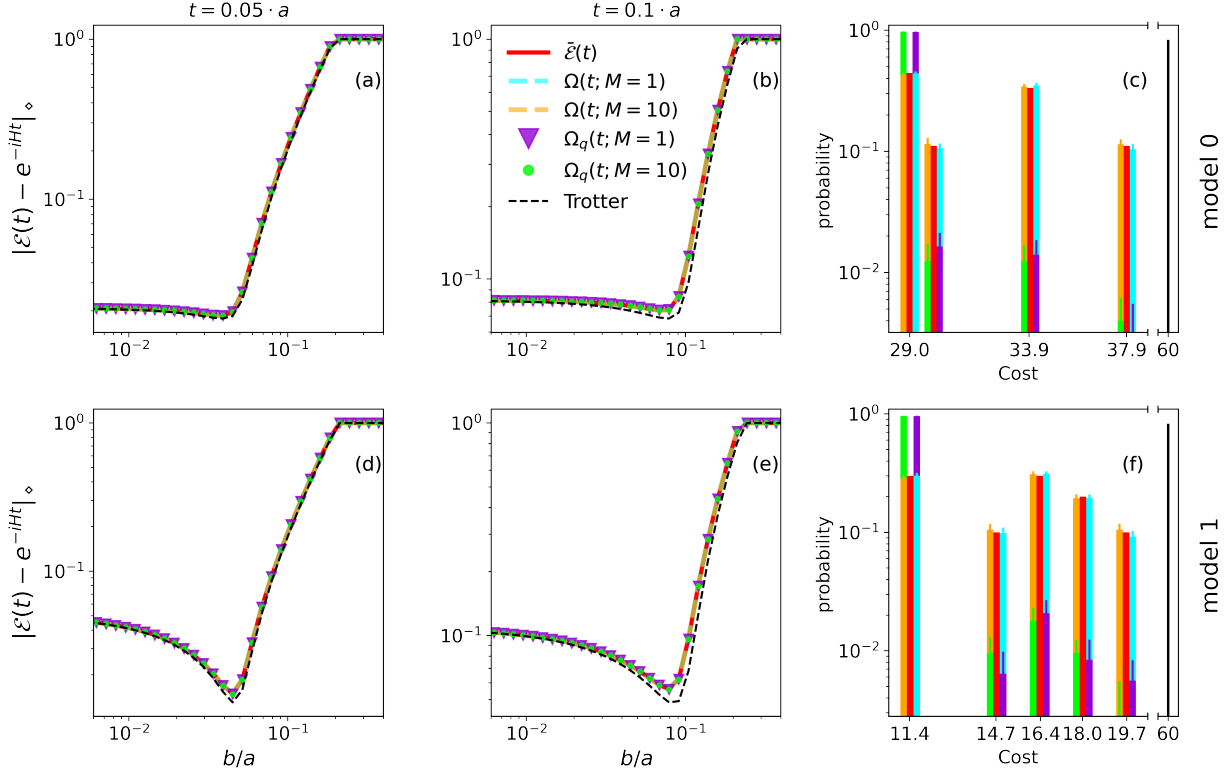


Figure 1: Diamond norm [(a), (b), (d), (e)] against the coefficient ratio b/a and histograms of the cost of the quantum circuits [(c), (f)] in terms of CNOT gates for different composite channels (deterministic $\tilde{\mathcal{E}}(t)$ in red, standard $\Omega(t; N=1, M, r=1)$ in dashed lines and importance sampled $\Omega_q(t; N=1, M, r=1)$ with $q = q_c$ in dots) for the two different models H^0 (upper line) and H^1 (bottom line), and simulation times ($t = 0.05a$ for the first column and $t = 0.1a$ for the second one).

Acknowledgements

OK and MG are supported by CERN Quantum Technology Initiative. AR is funded by the European Union under Horizon Europe Programme - Grant Agreement 101080086 — NeQST. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of CERN, the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither CERN, the European Union nor the granting authority can be held responsible for them.

References

- [1] R.P. Feynman. Simulating physics with computers. *Int J Theor Phys*, 21:467–488, 1982. DOI: <https://doi.org/10.1007/BF02650179>.
- [2] Matthias Troyer and Uwe-Jens Wiese. Computational complexity and fundamental limitations to fermionic quantum monte carlo simulations. *Phys. Rev. Lett.*, 94:170201, May 2005. DOI: [10.1103/PhysRevLett.94.170201](https://doi.org/10.1103/PhysRevLett.94.170201). URL <https://link.aps.org/doi/10.1103/PhysRevLett.94.170201>.
- [3] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

DOI: [10.1126/science.273.5278.1073](https://doi.org/10.1126/science.273.5278.1073). URL <https://www.science.org/doi/abs/10.1126/science.273.5278.1073>.

- [4] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski. Cloud quantum computing of an atomic nucleus. *Phys. Rev. Lett.*, 120:210501, May 2018. DOI: [10.1103/PhysRevLett.120.210501](https://doi.org/10.1103/PhysRevLett.120.210501). URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.210501>.
- [5] Alessandro Roggero and Joseph Carlson. Dynamic linear response quantum algorithm. *Phys. Rev. C*, 100:034610, Sep 2019. DOI: [10.1103/PhysRevC.100.034610](https://doi.org/10.1103/PhysRevC.100.034610). URL <https://link.aps.org/doi/10.1103/PhysRevC.100.034610>.
- [6] Oriel Kiss, Michele Grossi, Pavel Lougovski, Federico Sanchez, Sofia Vallecorsa, and Thomas Papenbrock. Quantum computing of the ${}^6\text{Li}$ nucleus via ordered unitary coupled clusters. *Phys. Rev. C*, 106:034325, Sep 2022. DOI: [10.1103/PhysRevC.106.034325](https://doi.org/10.1103/PhysRevC.106.034325). URL <https://link.aps.org/doi/10.1103/PhysRevC.106.034325>.
- [7] W Hofstetter and T Qin. Quantum simulation of strongly correlated condensed matter systems. *Journal of Physics B: Atomic, Molecu-*

- lar and Optical Physics*, 51(8):082001, mar 2018. DOI: [10.1088/1361-6455/aaa31b](https://doi.org/10.1088/1361-6455/aaa31b). URL <https://doi.org/10.1088/1361-6455/aaa31b>.
- [8] Nathan Keenan, Niall Robertson, Tara Murphy, Sergiy Zhuk, and John Gool. Evidence of kardar-parisi-zhang scaling on a digital quantum simulator, 2022. URL <https://arxiv.org/abs/2208.12243>.
- [9] Michele Grossi, Oriel Kiss, Francesco De Luca, Carlo Zollo, Ian Gremese, and Antonio Mandarino. Finite-size criticality in fully connected spin models on superconducting quantum hardware, 2022. URL <https://arxiv.org/abs/2208.02731>.
- [10] Maxime Dupont and Joel E. Moore. Quantum criticality using a superconducting quantum processor. *Phys. Rev. B*, 106:L041109, Jul 2022. DOI: [10.1103/PhysRevB.106.L041109](https://doi.org/10.1103/PhysRevB.106.L041109). URL <https://link.aps.org/doi/10.1103/PhysRevB.106.L041109>.
- [11] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum algorithms for quantum field theories. *Science*, 336(6085):1130–1133, 2012. DOI: [10.1126/science.1217069](https://doi.org/10.1126/science.1217069). URL <https://www.science.org/doi/abs/10.1126/science.1217069>.
- [12] Alexander F. Shaw, Pavel Lougovski, Jesse R. Stryker, and Nathan Wiebe. Quantum Algorithms for Simulating the Lattice Schwinger Model. *Quantum*, 4:306, August 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-08-10-306](https://doi.org/10.22331/q-2020-08-10-306). URL <https://doi.org/10.22331/q-2020-08-10-306>.
- [13] Natalie Klco, Alessandro Roggero, and Martin J Savage. Standard model physics and the digital quantum revolution: thoughts about the interface. *Reports on Progress in Physics*, 85(6):064301, may 2022. DOI: [10.1088/1361-6633/ac58a4](https://doi.org/10.1088/1361-6633/ac58a4). URL <https://dx.doi.org/10.1088/1361-6633/ac58a4>.
- [14] Yuan Su, Hsin-Yuan Huang, and Earl T. Campbell. Nearly tight Trotterization of interacting electrons. *Quantum*, 5:495, July 2021. ISSN 2521-327X. DOI: [10.22331/q-2021-07-05-495](https://doi.org/10.22331/q-2021-07-05-495). URL <https://doi.org/10.22331/q-2021-07-05-495>.
- [15] Yingkai Ouyang, David R. White, and Earl T. Campbell. Compilation by stochastic Hamiltonian sparsification. *Quantum*, 4:235, February 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-02-27-235](https://doi.org/10.22331/q-2020-02-27-235). URL <https://doi.org/10.22331/q-2020-02-27-235>.
- [16] Luis A. Martínez-Martínez, Tzu-Ching Yen, and Artur F. Izmaylov. Assessment of various hamiltonian partitionings for the electronic structure problem on a quantum computer using the trotter approximation, 2022. URL <https://arxiv.org/abs/2210.10189>.
- [17] Daniel S. Abrams and Seth Lloyd. Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. *Phys. Rev. Lett.*, 83:5162–5165, Dec 1999. DOI: [10.1103/PhysRevLett.83.5162](https://doi.org/10.1103/PhysRevLett.83.5162). URL <https://link.aps.org/doi/10.1103/PhysRevLett.83.5162>.
- [18] Zhaokai Li, Xiaomei Liu, Hefeng Wang, Sahel Ashhab, Jiangyu Cui, Hongwei Chen, Xinhua Peng, and Jiangfeng Du. Quantum simulation of resonant transitions for solving the eigenproblem of an effective water hamiltonian. *Phys. Rev. Lett.*, 122:090504, Mar 2019. DOI: [10.1103/PhysRevLett.122.090504](https://doi.org/10.1103/PhysRevLett.122.090504). URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.090504>.
- [19] A. Baroni, J. Carlson, R. Gupta, Andy C. Y. Li, G. N. Perdue, and A. Roggero. Nuclear two point correlation functions on a quantum computer. *Phys. Rev. D*, 105:074503, Apr 2022. DOI: [10.1103/PhysRevD.105.074503](https://doi.org/10.1103/PhysRevD.105.074503). URL <https://link.aps.org/doi/10.1103/PhysRevD.105.074503>.
- [20] A. Chiesa, F. Tacchino, M. Grossi, P. Santini, I. Tavernelli, D. Gerace, and S. Carretta. Quantum hardware simulating four-dimensional inelastic neutron scattering. *Nat. Phys.*, 15:455–459, 2019. DOI: <https://doi.org/10.1038/s41567-019-0437-4>.
- [21] Benjamin Hall, Alessandro Roggero, Alessandro Baroni, and Joseph Carlson. Simulation of collective neutrino oscillations on a quantum computer. *Phys. Rev. D*, 104:063009, Sep 2021. DOI: [10.1103/PhysRevD.104.063009](https://doi.org/10.1103/PhysRevD.104.063009). URL <https://link.aps.org/doi/10.1103/PhysRevD.104.063009>.
- [22] Valentina Amitrano, Alessandro Roggero, Piero Luchi, Francesco Turro, Luca Vespucci, and Francesco Pederiva. Trapped-ion quantum simulation of collective neutrino oscillations, 2022. URL <https://arxiv.org/abs/2207.03189>.
- [23] Alessandro Roggero, Andy C. Y. Li, Joseph Carlson, Rajan Gupta, and Gabriel N. Perdue. Quantum computing for neutrino-nucleus scattering. *Phys. Rev. D*, 101:074038, Apr 2020. DOI: [10.1103/PhysRevD.101.074038](https://doi.org/10.1103/PhysRevD.101.074038). URL <https://link.aps.org/doi/10.1103/PhysRevD.101.074038>.
- [24] Weijie Du, James P. Vary, Xingbo Zhao, and Wei Zuo. Quantum simulation of nuclear inelastic scattering. *Phys. Rev. A*, 104:012611, Jul 2021. DOI: [10.1103/PhysRevA.104.012611](https://doi.org/10.1103/PhysRevA.104.012611). URL <https://link.aps.org/doi/10.1103/PhysRevA.104.012611>.
- [25] Marc Illa and Martin J. Savage. Multi-neutrino entanglement and correlations in dense neutrino systems, 2022. URL <https://arxiv.org/abs/2210.08656>.
- [26] Masuo Suzuki. General theory of fractal path

- integrals with applications to many-body theories and statistical physics. *Journal of Mathematical Physics*, 32(2):400–407, 1991. DOI: [10.1063/1.529425](https://doi.org/10.1063/1.529425). URL <https://doi.org/10.1063/1.529425>.
- [27] Masuo Suzuki. Fractal decomposition of exponential operators with applications to many-body theories and monte carlo simulations. *Physics Letters A*, 146(6): 319–323, 1990. ISSN 0375-9601. DOI: [https://doi.org/10.1016/0375-9601\(90\)90962-N](https://doi.org/10.1016/0375-9601(90)90962-N). URL <https://www.sciencedirect.com/science/article/pii/037596019090962N>.
- [28] Nathan Wiebe, Dominic Berry, Peter Høyer, and Barry C Sanders. Higher order decompositions of ordered operator exponentials. *J. Phys. A: Math. Theor.*, 43(6): 065203, jan 2010. DOI: [10.1088/1751-8113/43/6/065203](https://dx.doi.org/10.1088/1751-8113/43/6/065203). URL <https://dx.doi.org/10.1088/1751-8113/43/6/065203>.
- [29] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu. Theory of trotter error with commutator scaling. *Phys. Rev. X*, 11:011020, Feb 2021. DOI: [10.1103/PhysRevX.11.011020](https://link.aps.org/doi/10.1103/PhysRevX.11.011020). URL <https://link.aps.org/doi/10.1103/PhysRevX.11.011020>.
- [30] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. Gate-count estimates for performing quantum chemistry on small quantum computers. *Phys. Rev. A*, 90:022305, Aug 2014. DOI: [10.1103/PhysRevA.90.022305](https://link.aps.org/doi/10.1103/PhysRevA.90.022305). URL <https://link.aps.org/doi/10.1103/PhysRevA.90.022305>.
- [31] Markus Reiher, Nathan Wiebe, Krysta M. Svore, Dave Wecker, and Matthias Troyer. Elucidating reaction mechanisms on quantum computers. *Proceedings of the National Academy of Sciences*, 114(29):7555–7560, 2017. DOI: [10.1073/pnas.1619152114](https://www.pnas.org/doi/abs/10.1073/pnas.1619152114). URL <https://www.pnas.org/doi/abs/10.1073/pnas.1619152114>.
- [32] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. ISSN 2521-327X. DOI: [10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79). URL <https://doi.org/10.22331/q-2018-08-06-79>.
- [33] I. M. Georgescu, S. Ashhab, and Franco Nori. Quantum simulation. *Rev. Mod. Phys.*, 86: 153–185, Mar 2014. DOI: [10.1103/RevModPhys.86.153](https://link.aps.org/doi/10.1103/RevModPhys.86.153). URL <https://link.aps.org/doi/10.1103/RevModPhys.86.153>.
- [34] Kaelyn J. Ferris, A. J. Rasmusson, Nicholas T. Bronn, and Olivia Lanes. Quantum simulation on noisy superconducting quantum computers, 2022. URL <https://arxiv.org/abs/2209.02795>.
- [35] Andrew M. Childs, Aaron Ostrander, and Yuan Su. Faster quantum simulation by randomization. *Quantum*, 3, 2019. ISSN 2521-327X. DOI: [10.22331/q-2019-09-02-182](https://doi.org/10.22331/q-2019-09-02-182). URL <https://doi.org/10.22331/q-2019-09-02-182>.
- [36] Paul K. Faehrmann, Mark Steudtner, Richard Kueng, Mária Kieferová, and Jens Eisert. Randomizing multi-product formulas for Hamiltonian simulation. *Quantum*, 6:806, September 2022. ISSN 2521-327X. DOI: [10.22331/q-2022-09-19-806](https://doi.org/10.22331/q-2022-09-19-806). URL <https://doi.org/10.22331/q-2022-09-19-806>.
- [37] Chien Hung Cho, Dominic W. Berry, and Min-Hsiu Hsieh. Doubling the order of approximation via the randomized product formula, 2022. URL <https://arxiv.org/abs/2210.11281>.
- [38] George C. Knee and William J. Munro. Optimal trotterization in universal quantum simulators under faulty control. *Phys. Rev. A*, 91:052327, May 2015. DOI: [10.1103/PhysRevA.91.052327](https://link.aps.org/doi/10.1103/PhysRevA.91.052327). URL <https://link.aps.org/doi/10.1103/PhysRevA.91.052327>.
- [39] Joel J. Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A*, 94:052325, Nov 2016. DOI: [10.1103/PhysRevA.94.052325](https://link.aps.org/doi/10.1103/PhysRevA.94.052325). URL <https://link.aps.org/doi/10.1103/PhysRevA.94.052325>.
- [40] David Poulin, Angie Qarry, Rolando Somma, and Frank Verstraete. Quantum simulation of time-dependent hamiltonians and the convenient illusion of hilbert space. *Phys. Rev. Lett.*, 106:170501, Apr 2011. DOI: [10.1103/PhysRevLett.106.170501](https://link.aps.org/doi/10.1103/PhysRevLett.106.170501). URL <https://link.aps.org/doi/10.1103/PhysRevLett.106.170501>.
- [41] Minh C. Tran, Yuan Su, Daniel Carney, and Jacob M. Taylor. Faster digital quantum simulation by symmetry protection. *PRX Quantum*, 2:010323, Feb 2021. DOI: [10.1103/PRXQuantum.2.010323](https://link.aps.org/doi/10.1103/PRXQuantum.2.010323). URL <https://link.aps.org/doi/10.1103/PRXQuantum.2.010323>.
- [42] Earl Campbell. Random compiler for fast hamiltonian simulation. *Phys. Rev. Lett.*, 123:070503, Aug 2019. DOI: [10.1103/PhysRevLett.123.070503](https://link.aps.org/doi/10.1103/PhysRevLett.123.070503). URL <https://link.aps.org/doi/10.1103/PhysRevLett.123.070503>.
- [43] Chi-Fang Chen, Hsin-Yuan Huang, Richard Kueng, and Joel A. Tropp. Concentration for random product formulas. *PRX Quantum*, 2: 040305, Oct 2021. DOI: [10.1103/PRXQuantum.2.040305](https://link.aps.org/doi/10.1103/PRXQuantum.2.040305). URL <https://link.aps.org/doi/10.1103/PRXQuantum.2.040305>.
- [44] Dominic W. Berry, Andrew M. Childs, Yuan Su, Xin Wang, and Nathan Wiebe. Time-dependent Hamiltonian simulation with L^1 -norm scaling. *Quantum*, 4:254, April 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-04-20-254](https://doi.org/10.22331/q-2020-04-20-254). URL <https://doi.org/10.22331/q-2020-04-20-254>.
- [45] Andrew M. Childs and Nathan Wiebe. Hamil-

- tonian simulation using linear combinations of unitary operations. *Quantum Information and Computation*, 12(11&12):0901–0924, 2012. DOI: <https://doi.org/10.26421/QIC12.11-12-1>.
- [46] Dominic W. Berry and Andrew M. Childs. Black-box hamiltonian simulation and unitary implementation. *Quantum Info. Comput.*, 12(1-2): 29–62, jan 2012. ISSN 1533-7146. URL <https://dl.acm.org/doi/10.5555/2231036.2231040>.
- [47] Guang Hao Low and Isaac L. Chuang. Hamiltonian Simulation by Qubitization. *Quantum*, 3:163, July 2019. ISSN 2521-327X. DOI: [10.22331/q-2019-07-12-163](https://doi.org/10.22331/q-2019-07-12-163). URL <https://doi.org/10.22331/q-2019-07-12-163>.
- [48] Ryan Babbush, Craig Gidney, Dominic W. Berry, Nathan Wiebe, Jarrod McClean, Alexandru Paler, Austin Fowler, and Hartmut Neven. Encoding electronic spectra in quantum circuits with linear t complexity. *Phys. Rev. X*, 8:041015, Oct 2018. DOI: [10.1103/PhysRevX.8.041015](https://doi.org/10.1103/PhysRevX.8.041015). URL <https://link.aps.org/doi/10.1103/PhysRevX.8.041015>.
- [49] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu. Theory of trotter error with commutator scaling. *Phys. Rev. X*, 11:011020, Feb 2021. DOI: [10.1103/PhysRevX.11.011020](https://doi.org/10.1103/PhysRevX.11.011020). URL <https://link.aps.org/doi/10.1103/PhysRevX.11.011020>.
- [50] Surya T. Tokdar and Robert E. Kass. Importance sampling: a review. *WIREs Computational Statistics*, 2(1):54–60, 2010. DOI: <https://doi.org/10.1002/wics.56>. URL <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.56>.
- [51] Matthew Hagan and Nathan Wiebe. Composite quantum simulations, 2022. URL <https://arxiv.org/abs/2206.06409>.
- [52] Shi Jin and Xiantao Li. A partially random trotter algorithm for quantum hamiltonian simulations, 2021. URL <https://arxiv.org/abs/2109.07987>.
- [53] Abhishek Rajput, Alessandro Roggero, and Nathan Wiebe. Hybridized Methods for Quantum Simulation in the Interaction Picture. *Quantum*, 6:780, August 2022. ISSN 2521-327X. DOI: [10.22331/q-2022-08-17-780](https://doi.org/10.22331/q-2022-08-17-780). URL <https://doi.org/10.22331/q-2022-08-17-780>.
- [54] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906. DOI: [10.1007/BF02418571](https://doi.org/10.1007/BF02418571).
- [55] P. Jordan and E. Wigner. Über das paulische Äquivalenzverbot. *Z. Physik*, 47:631–651, 1928. DOI: <https://doi.org/10.1007/BF01331938>.
- [56] C. Chamberland and Noh K. Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits. *npj Quantum Inf*, 6:91, 2020. DOI: <https://doi.org/10.1038/s41534-020-00319-5>.
- [57] Sourabh Katoch, Sumit Singh Chauhan, and Vijay Kumar. A review on genetic algorithm: past, present, and future. *Multimed Tools Appl*, 80:8091–8126, 2021. DOI: <https://doi.org/10.1007/s11042-020-10139-6>.
- [58] Shouzhen Gu, Rolando D. Somma, and Burak Şahinoğlu. Fast-forwarding quantum evolution. *Quantum*, 5:577, November 2021. ISSN 2521-327X. DOI: [10.22331/q-2021-11-15-577](https://doi.org/10.22331/q-2021-11-15-577). URL <https://doi.org/10.22331/q-2021-11-15-577>.
- [59] E. Wigner. On the consequences of the symmetry of the nuclear hamiltonian on the spectroscopy of nuclei. *Phys. Rev.*, 51:106–119, Jan 1937. DOI: [10.1103/PhysRev.51.106](https://doi.org/10.1103/PhysRev.51.106). URL <https://link.aps.org/doi/10.1103/PhysRev.51.106>.
- [60] David B. Kaplan and Martin J. Savage. The spin-flavor dependence of nuclear forces from large-n qcd. *Physics Letters B*, 365(1):244–251, 1996. ISSN 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(95\)01277-X](https://doi.org/10.1016/0370-2693(95)01277-X). URL <https://www.sciencedirect.com/science/article/pii/037026939501277X>.
- [61] Michael L. Wagman, Frank Winter, Emmanuel Chang, Zohreh Davoudi, William Detmold, Kostas Orginos, Martin J. Savage, and Phiala E. Shanahan. Baryon-baryon interactions and spin-flavor symmetry from lattice quantum chromodynamics. *Phys. Rev. D*, 96:114510, Dec 2017. DOI: [10.1103/PhysRevD.96.114510](https://doi.org/10.1103/PhysRevD.96.114510). URL <https://link.aps.org/doi/10.1103/PhysRevD.96.114510>.
- [62] Andrei Alexandru, Paulo Bedaque, Evan Berkowitz, and Neill C. Warrington. Structure factors of neutron matter at finite temperature. *Phys. Rev. Lett.*, 126:132701, Apr 2021. DOI: [10.1103/PhysRevLett.126.132701](https://doi.org/10.1103/PhysRevLett.126.132701). URL <https://link.aps.org/doi/10.1103/PhysRevLett.126.132701>.
- [63] L. Contessi, A. Lovato, F. Pederiva, A. Roggero, J. Kirscher, and U. van Kolck. Ground-state properties of 4He and 16O extrapolated from lattice qcd with pionless eft. *Physics Letters B*, 772:839–848, 2017. ISSN 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2017.07.048>. URL <https://www.sciencedirect.com/science/article/pii/S0370269317306044>.
- [64] Bing-Nan Lu, Ning Li, Serdar Elhatisari, Dean Lee, Evgeny Epelbaum, and Ulf-G. Meißner. Essential elements for nuclear binding. *Physics Letters B*, 797:134863, 2019. ISSN 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2019.134863>. URL <https://www.sciencedirect.com/science/article/pii/S0370269319305775>.
- [65] Dillon Frame, Rongzheng He, Ilse Ipsen, Daniel Lee, Dean Lee, and Ermal Rrapaj. Eigenvector

- continuation with subspace learning. *Phys. Rev. Lett.*, 121:032501, Jul 2018. DOI: [10.1103/PhysRevLett.121.032501](https://doi.org/10.1103/PhysRevLett.121.032501). URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.032501>.
- [66] Paulo F. Bedaque and Ubirajara van Kolck. Effective field theory for few-nucleon systems. *Annual Review of Nuclear and Particle Science*, 52(1):339–396, 2002. DOI: [10.1146/annurev.nucl.52.050102.090637](https://doi.org/10.1146/annurev.nucl.52.050102.090637). URL <https://doi.org/10.1146/annurev.nucl.52.050102.090637>.
- [67] S. Pastore, J. Carlson, S. Gandolfi, R. Schiavilla, and R. B. Wiringa. Quasielastic lepton scattering and back-to-back nucleons in the short-time approximation. *Phys. Rev. C*, 101:044612, Apr 2020. DOI: [10.1103/PhysRevC.101.044612](https://doi.org/10.1103/PhysRevC.101.044612). URL <https://link.aps.org/doi/10.1103/PhysRevC.101.044612>.
- [68] Raffaele Santagati, Jianwei Wang, Antonio A. Gentile, Stefano Paesani, Nathan Wiebe, Jarrod R. McClean, Sam Morley-Short, Peter J. Shadbolt, Damien Bonneau, Joshua W. Silverstone, David P. Tew, Xiaoqi Zhou, Jeremy L. O’Brien, and Mark G. Thompson. Witnessing eigenstates for quantum simulation of hamiltonian spectra. *Science Advances*, 4(1):eaap9646, 2018. DOI: [10.1126/sciadv.aap9646](https://doi.org/10.1126/sciadv.aap9646). URL <https://www.science.org/doi/abs/10.1126/sciadv.aap9646>.
- [69] A. Roggero and A. Baroni. Short-depth circuits for efficient expectation-value estimation. *Phys. Rev. A*, 101:022328, Feb 2020. DOI: [10.1103/PhysRevA.101.022328](https://doi.org/10.1103/PhysRevA.101.022328). URL <https://link.aps.org/doi/10.1103/PhysRevA.101.022328>.
- [70] Thomas E. O’Brien, Stefano Polla, Nicholas C. Rubin, William J. Huggins, Sam McArdle, Sergio Boixo, Jarrod R. McClean, and Ryan Babbush. Error mitigation via verified phase estimation. *PRX Quantum*, 2:020317, May 2021. DOI: [10.1103/PRXQuantum.2.020317](https://doi.org/10.1103/PRXQuantum.2.020317). URL <https://link.aps.org/doi/10.1103/PRXQuantum.2.020317>.
- [71] Joel Tropp. Freedman’s inequality for matrix martingales. *Electronic Communications in Probability*, 16(none):262 – 270, 2011. DOI: [10.1214/ECP.v16-1624](https://doi.org/10.1214/ECP.v16-1624). URL <https://doi.org/10.1214/ECP.v16-1624>.

A Proofs of the main results

In this section, we present rigorous proofs of the theorems stated in the main results Section 3.2 and Section 3.3. Before being able to provide them, we first need to state two lemmas from Ref. [43].

Lemma 2. Let $\mathcal{U}(\rho) = U\rho U^\dagger$ and $\mathcal{V}(\rho) = V\rho V^\dagger$ be unitary channels, we then have

$$\|\mathcal{U} - \mathcal{V}\|_\diamond \leq 2\|U - V\|. \quad (65)$$

The results carries over to ensembles (p_k, V_k) of unitary channels with weights $p_k \geq 0$ for which $\sum_k p_k = 1$.

$$\left\| \mathcal{U} - \sum_k p_k \mathcal{V}_k \right\|_\diamond \leq 2 \left\| U - \sum_k p_k V_k \right\|. \quad (66)$$

Lemma 3. Let X be hermitian. We then have the zero-th order bound $\|\exp\{iX\} - \mathbb{1}\| \leq \|X\|$ and the first-order bound $\|\exp\{iX\} - iX - \mathbb{1}\| \leq \frac{1}{2}\|X\|^2$

We are now in a position to prove Theorem 1, that we will recall for the ease of the reader.

Theorem 1 (Bias error bound). Let $\mathcal{U}(t)$ be a first-order Trotter product channel, $\bar{\mathcal{E}}_q(t; N)$ an average q Drift channel with importance sampling and $\omega(j) = p(j)/q(j)$ the re-weighting factor. The diamond norm distance between these two channels for $N = 1$ is then upper bounded by

$$\|\mathcal{U}(t) - \bar{\mathcal{E}}_q(t; 1)\|_\diamond \leq t^2 \lambda^2 (1 + \mathbb{E}_p[\omega(j)]) , \quad (67)$$

leading to the following result

$$\|\mathcal{U}(t) - \bar{\mathcal{E}}_q(t; N)\|_\diamond \leq \frac{t^2 \lambda^2}{N} (1 + \mathbb{E}_p[\omega(j)]) . \quad (68)$$

Proof of Theorem 1. We first note that the Hamiltonian H can be written as the following expectation value

$$H = \sum_j h_j H_j = \lambda \mathbb{E}_p[H_j] = \lambda \mathbb{E}_q[\omega(j) H_j], \quad (69)$$

with $\omega(j) = h_j/(\lambda q(j))$ and therefore

$$\begin{aligned}
U(t) &= e^{-itH} \\
&= e^{-it\lambda\mathbb{E}_p[H_j]} \\
&= e^{-it\lambda\mathbb{E}_q[\omega(j)H_j]} \\
&= e^{-i\mathbb{E}_q[X_j]},
\end{aligned} \tag{70}$$

with $X_j(t) = \frac{th_j}{q(j)}H_j$. By noting that

$$\mathbb{E}_q[X_j(t)] \leq \sum_j q(j)X_j(t) = tH, \tag{71}$$

we obtain the following bound

$$\|\mathbb{E}_q[X_j(t)]\| = t \sum_j h_j \|H_j\| = \lambda t, \tag{72}$$

while

$$\|X_j(t)\| = \left(\frac{th_j}{q(j)}\right) \|H_j\| = \frac{th_j}{q(j)}. \tag{73}$$

If we denote $V(t) = \exp\{-iX(t)\}$ and the corresponding channel as $\mathcal{V}(t)[\rho] = V(t)\rho V(t)^\dagger$, we can express the deterministic qDrift channel with importance sampling as (cf. Eq. (9) in the main text)

$$\bar{\mathcal{E}}_q(t; 1)[\rho] = \sum_j q(j)V(t)\rho V(t)^\dagger = \mathbb{E}_q[\mathcal{V}(t)]. \tag{74}$$

We now observe that

$$\begin{aligned}
\|\mathcal{U}(t) - \bar{\mathcal{E}}_q(t; 1)\|_\diamond &= \|\mathcal{U}(t) - \mathbb{E}_q[\mathcal{V}(t)]\|_\diamond \leq 2\|U(t) - \mathbb{E}_q[V(t)]\| \\
&= 2\left\|e^{-i\mathbb{E}_q[X(t)]} - \mathbb{1} + i\mathbb{E}_q[X(t)] + \mathbb{E}_q\left[\mathbb{1} - iX(t) - e^{-iX(t)}\right]\right\| \\
&\leq 2\left\|e^{-i\mathbb{E}_q[X(t)]} - \mathbb{1} + i\mathbb{E}_q[X(t)]\right\| + 2\mathbb{E}_q\left[\left\|e^{-iX(t)} - \mathbb{1} + iX(t)\right\|\right] \\
&\leq \|\mathbb{E}_q[X(t)]\|^2 + \mathbb{E}_q[\|X(t)\|^2] \\
&\leq (t\lambda)^2 + \mathbb{E}_q\left[\left(\frac{th_j}{q(j)}\right)^2\right] \\
&= (t\lambda)^2 (1 + \mathbb{E}_q[\omega^2(j)]) \\
&= (t\lambda)^2 (1 + \mathbb{E}_p[\omega(j)]),
\end{aligned} \tag{75}$$

which was the first result what we set out to show. The result for the general average channel with $N > 1$ follows by first considering the fact that we can obtain $\bar{\mathcal{E}}_q(t; 1)$ by N compositions

$$\bar{\mathcal{E}}_q(t; N) = \bar{\mathcal{E}}_q\left(\frac{t}{N}; 1\right) \circ \dots \circ \bar{\mathcal{E}}_q\left(\frac{t}{N}; 1\right) = \bar{\mathcal{E}}_q\left(\frac{t}{N}; 1\right)^{\circ N} = \mathbb{E}_q\left[\mathcal{V}\left(\frac{t}{N}\right)\right]^{\circ N}. \tag{76}$$

Following [43] we then decompose the total evolution time t into N steps of duration t/N to find

$$\begin{aligned}
\|\mathcal{U}(t) - \bar{\mathcal{E}}_q(t; N)\|_\diamond &= \left\|\mathcal{U}\left(\frac{t}{N}\right)^{\circ N} - \bar{\mathcal{E}}_q\left(\frac{t}{N}; 1\right)^{\circ N}\right\|_\diamond \\
&= \left\|\mathcal{U}\left(\frac{t}{N}\right)^{\circ N} - \mathbb{E}_q\left[\mathcal{V}\left(\frac{t}{N}\right)\right]^{\circ N}\right\|_\diamond \\
&\leq 2\left\|U\left(\frac{t}{N}\right)^N - \mathbb{E}_q\left[V\left(\frac{t}{N}\right)\right]^N\right\| \\
&\leq 2N\left\|U\left(\frac{t}{N}\right) - \mathbb{E}_q\left[V\left(\frac{t}{N}\right)\right]\right\| \\
&\leq \frac{t^2\lambda^2}{N} (1 + \mathbb{E}_p[\omega(j)]),
\end{aligned} \tag{77}$$

where we used Lemma 2 for the second line, the union bound for the third and Eq. (75) for the last step. \square

Now that we have generalized the bias error bound, we need to understand how a finite importance sampled qDrift channel concentrates around its expected value. This will provide us with an estimate of M and N for a given accuracy ϵ . We will use the martingale formalism and rely on Ref. [71] for a more in-depth consideration.

Definition 1 (martingale). *Consider a filtration of the master sigma algebra $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}$. A martingale is a sequence $\{B_0, B_1, \dots\}$ of random variable satisfying*

$$1. \sigma(B_k) \subset \mathcal{F}_k \quad (\text{causality})$$

$$2. \mathbb{E}[B_k | B_{k-1} \dots B_0] = B_{k-1} \quad (\text{status quo}).$$

The intuition one may have is to think of k as a time index and \mathcal{F}_k to contain all events determined by the past up to time k . The causality requirement states that the present B_k may only depend on the past (B_{k-1}, \dots, B_0) , and the status quo conditions formulate that, on average, today is the same as yesterday.

Before proving the theorem, we need to state another result from Ref. [43, Corollary 3.4].

Corollary 4. *Let $\{B_k : k = 0, \dots, N\} \subset \mathbb{M}_{d \times d}$ be a matrix martingale. Assume that the associated difference $C_k := B_k - B_{k-1}$ obeys $\|C_k\| \leq R$ and its conditional variance $\left\| \sum_{k=0}^N \mathbb{E}[C_k C_k^\dagger | C_{k-1} \dots C_0] \right\| \leq v$ almost surely. Then $\forall \tau \geq 0$, we have*

$$\Pr[\|B_N - B_0\| \leq \tau] \geq 2d \exp\left\{ \frac{-\tau^2/2}{v + R\tau/3} \right\}. \quad (78)$$

In order to show Theorem 2, we will generalize the construction from Ref. [43] of a suitable interpolating martingale. Let's start by introducing the unitaries $V_j = e^{-i\tau_j H_j}$, for which $\mathbb{E}_q[V_j] = \mathbb{E}_q[V]$ independently on j , and consider the situation where we take a set of M separate samples of the N indices forming the product formula resulting in M distinct martingales of the form

$$B_k^m = \mathbb{E}_q[V]^{N-k} \prod_{r=k}^1 V_r^m, \quad (79)$$

with $m \in \{1, 2, \dots, M\}$ and V_j^m the j -th unitary in the m -th sample. For technical reasons, we also define, for every value of k , $B_k^0 = 0$ as well as $B_k^m = 0$ for all $m > M$. The causality condition in Definition 1 is automatically satisfied since $\forall m$ B_k^m is completely determined by the random samples V_1^m, \dots, V_k^m obtained up to the k -th step. The second condition can be checked explicitly, in fact

$$\mathbb{E}[B_{k+1}^m | B_k^m, \dots, B_0^m] = \mathbb{E}_q[V]^{N-k-1} \mathbb{E}_q[V_{k+1}^m] \prod_{r=k}^1 V_r^m = B_k^m. \quad (80)$$

The generalized interpolating martingale needed for our construction can then be defined for $j \in \{0, 1, \dots, NM\}$ as follows

$$D_j = \sum_{m=\lfloor j/N \rfloor + 2}^M B_0^m + B_{j\%N}^{\lfloor j/N \rfloor + 1} + \sum_{m=1}^{\lfloor j/N \rfloor} B_N^m, \quad (81)$$

where we denote with $\lfloor a/b \rfloor$ the entire division from a by b and $a\%b$ the rest, i.e., the integer value a modulo b . It is straightforward to see that the sum of two independent martingales is also a martingale, making D_j a valid martingale. We have that the first element is given by

$$D_0 = \sum_{m=1}^M B_0^m = \sum_{m=1}^M \mathbb{E}_q[V]^N = M \mathbb{E}_q[V]^N, \quad (82)$$

while the last element, corresponding to $j = NM$, is given instead by

$$D_{NM} = \sum_{m=1}^M B_N^m = \sum_{m=1}^M \prod_{r=N}^1 V_r^m. \quad (83)$$

For the special case $M = 1$ we recover the construction presented in Ref. [43]. The elements of the associated difference sequence are given, for $j \in \{1, 2, \dots, NM\}$, by

$$\begin{aligned}
C_j &= D_j - D_{j-1} = B_{j\%N}^{\lfloor j/N \rfloor + 1} - B_{j\%N-1}^{\lfloor j/N \rfloor + 1} \\
&:= B_{b_j}^{a_j} - B_{b_j-1}^{a_j} \\
&= \mathbb{E}_q[V]^{N-b_j} \prod_{r=b_j}^1 V_r^{a_j} - \mathbb{E}_q[V]^{N-b_j+1} \prod_{r=b_j-1}^1 V_r^{a_j} \\
&= \mathbb{E}_q[V]^{N-b_j} \left(V_{b_j}^{a_j} - \mathbb{E}_q[V] \right) \prod_{r=b_j-1}^1 V_r^{a_j},
\end{aligned} \tag{84}$$

where, for ease of notation, we have introduced $a_j = \lfloor j/N \rfloor + 1$ and $b_j = j\%N$ in the second line. Since both V_k and $\mathbb{E}_q[V]$ are bounded ($\|V_k\| \leq 1$ almost surely and $\|\mathbb{E}_q[V]\| \leq 1$) we can find the following bound

$$\begin{aligned}
\|C_j\| &= \left\| V_{b_j}^{a_j} - \mathbb{E}_q[V] \right\| \\
&\leq \left\| e^{-i\tau_{b_j} H_{b_j}} - \mathbb{1} \right\| + \left\| \mathbb{1} - \mathbb{E}_q[e^{-i\tau_k H_k}] \right\| \\
&\leq \left\| e^{-i\tau_{b_j} H_{b_j}} - \mathbb{1} \right\| + \mathbb{E}_q \left[\left\| \mathbb{1} - e^{-i\tau_k H_k} \right\| \right],
\end{aligned} \tag{85}$$

where we used the triangle inequality in the second line and Jensen's inequality in the last. Furthermore, since

$$\|\tau_k H_k\| = \frac{t h_k}{N q(k)} \leq \frac{t\lambda}{N} \max_k \omega(k), \tag{86}$$

almost surely, using Lemma 3, it also holds almost surely that

$$\begin{aligned}
\|C_j\| &\leq \|\tau_k H_k\| + \mathbb{E}_q[\|\tau_k H_k\|] \\
&\leq \frac{t\lambda}{N} \max_k \omega(k) + \mathbb{E}_q[\tau_k] \\
&= \frac{t\lambda}{N} \left(1 + \max_k \omega(k) \right).
\end{aligned} \tag{87}$$

Finally, in order to control the variance, we use

$$\left\| \sum_{j=1}^{NM} \mathbb{E}[C_j C_j^\dagger | C_{j-1} \dots C_0] \right\| \leq NM \max_j \|C_j\|^2 = M \frac{t^2 \lambda^2}{N} \left(1 + \max_k \omega(k) \right)^2. \tag{88}$$

We will now use this construction to show Theorem 2, which is reformulated below.

Theorem 2 (Concentration bound). *Let $\mathcal{E}_q(t; N, M)$ be a finite importance sampled qDrift channel on n qubits and V_j instances of the NM unitaries that compose the channel. Their concentration around their expectation value can be upper bounded $\forall \epsilon \in [0, 4t\lambda]$ as*

$$\Pr \left[\left\| \frac{1}{M} \sum_m \prod_{k=N}^1 V_{k_m} - \mathbb{E}_q[V_j]^N \right\| \geq \epsilon/2 \right] \leq 2^{n+1} \exp \left\{ - \frac{NM\epsilon^2}{11t^2\lambda^2(1 + \max_k \omega(k))^2} \right\}. \tag{89}$$

In order to guarantee an approximation error $\epsilon/2$ with probability at least $1 - \delta$, it is then sufficient to take

$$NM = 11 \frac{t^2 \lambda^2}{\epsilon^2} \left(1 + \max_k \omega(k) \right)^2 (n+1) \log \left(\frac{2}{\delta} \right). \tag{90}$$

Proof of Theorem 2. Using the results in Eq. (87) and Eq. (88), we see that the parameters R and v from Corollary 4 can be chosen as

$$R = \frac{t\lambda}{N} \left(1 + \max_k \omega(k) \right), \quad v = MNR^2. \tag{91}$$

Using Corollary 4, we can show that

$$\begin{aligned}
\Pr \left[\left\| \frac{1}{M} \sum_m \prod_{k=N}^1 V_{k_m} - \mathbb{E}_q [V_j]^N \right\| \geq \tau \right] &= \Pr \left[\left\| \frac{1}{M} \sum_m \prod_{k=N}^1 V_{k_m} - \mathbb{E}_q \left[\frac{1}{M} \sum_m \prod_{k=N}^1 V_{k_m} \right] \right\| \geq \tau \right] \\
&= \Pr \left[\sum_{m=1}^M \prod_{k=N}^1 V_{k_m} - \mathbb{E}_q \left[\sum_{m=1}^M \prod_{k=N}^1 V_{k_m} \right] \geq M\tau \right] \\
&= \Pr [\|D_{NM} - D_0\| \geq M\tau] \\
&\leq 2^{n+1} \exp \left\{ -\frac{M^2 \tau^2 / 2}{v + MR\tau/3} \right\} \\
&= 2^{n+1} \exp \left\{ -\frac{3M\tau^2}{6NR^2 + 2R\tau} \right\}.
\end{aligned} \tag{92}$$

As in Ref. [43], for $\tau \leq NR$ we consider the simpler bound

$$\begin{aligned}
\Pr \left[\left\| \frac{1}{M} \sum_m \prod_{k=N}^1 V_{k_m} - \mathbb{E} [V_j]^N \right\| \geq \tau \right] &\leq 2^{n+1} \exp \left\{ -\frac{3M\tau^2}{8NR^2} \right\} \\
&= 2^{n+1} \exp \left\{ -\frac{3NM\tau^2}{8(t\lambda(1 + \max_k \omega(k)))^2} \right\}.
\end{aligned} \tag{93}$$

A looser sufficient condition $\tau \leq 2\lambda t$ can be obtained by noticing that $\max_k \omega(k) \geq 1$, and the equality only holds when all the weights are the same. Substituting $\tau = \epsilon/2$, using $32/3 \leq 11$ and Lemma 2, we obtain the theorem statement. \square

Finally, these results can be used to compute the expected fluctuation bound, see Corollary 1.

Corollary 1 (Fluctuation bound). *Let H be a n -qubit Hamiltonian, $q(j)$ an arbitrary distribution, t the simulation time, N a fixed number of qDrift samples, and M a fixed number of qDrift experiment. Set $\mathcal{U}_H[\rho] = U_H \rho U_H^\dagger$ (with $U_H = e^{-iHt}$) and take the importance sampled qDrift channel $\mathcal{E}_q(t; N, M)$. We then have*

$$\mathbb{E}_q [\|\mathcal{E}_q(t; N, M) - \mathcal{U}_H\|_\diamond] \leq 2 \frac{t^2 \lambda^2}{N} (1 + \mathbb{E}_p[\omega]) + \alpha \frac{nt\lambda}{NM} \left(1 + \max_k \omega(k) \right) + \alpha \sqrt{\frac{n}{NM}} t\lambda \left(1 + \max_k \omega(k) \right). \tag{94}$$

Proof. We prove the corollary by relating the diamond norm distance to the operator norm for ensembles of unitary channels, see Lemma 2, and then using the triangle inequality

$$\begin{aligned}
\mathbb{E}_q [\|\mathcal{E}_q(t; N, M) - \mathcal{U}_H\|_\diamond] &\leq 2 \mathbb{E}_q \left[\left\| \frac{1}{M} \sum_{m=1}^M \prod_{k=N}^1 V_{m_k} - U_H \right\| \right] \\
&\leq 2 \|U_H - \mathbb{E}_q[V]^N\| + 2 \mathbb{E}_q \left[\left\| \frac{1}{M} \sum_{m=1}^M \prod_{k=N}^1 V_{m_k} - \mathbb{E}_q[V]^N \right\| \right].
\end{aligned} \tag{95}$$

The first term can be bounded by using Theorem 1, while for the second we have

$$\begin{aligned}
2 \mathbb{E}_q \left[\left\| \frac{1}{M} \sum_{m=1}^M \prod_{k=N}^1 V_{m_k} - \mathbb{E}_q[V]^N \right\| \right] &= 2 \int_0^\infty \Pr \left(\left\| \frac{1}{M} \sum_{m=1}^M \prod_{k=N}^1 V_{m_k} - \mathbb{E}_q[V]^N \right\| \geq \tau \right) d\tau \\
&\leq 2 \int_0^\infty \min \left(1, 2 \cdot 2^n e^{-\frac{3M\tau^2}{6NR^2 + 2R\tau}} \right) d\tau \\
&\leq \frac{\alpha}{2} \max \left(\sqrt{\frac{n}{NM}} t\lambda(1 + \max_j \omega(j)), \frac{nt\lambda(1 + \max_j \omega(j))}{NM} \right) \\
&\leq \alpha \left(\sqrt{\frac{n}{NM}} t\lambda(1 + \max_j \omega(j)) + \frac{nt\lambda(1 + \max_j \omega(j))}{NM} \right)
\end{aligned} \tag{96}$$

As in Ref. [43], the integral is evaluated by cutting it into two parts. The first, with a contribution of nearly one, when the denominator in the exponent is bigger than the numerator, i.e., $\tau \leq \max \left(\sqrt{\frac{2n}{M}} R, \frac{2Rn}{3M} \right)$, where α suppresses any constant. The contribution for larger τ is marginal and of order $\mathcal{O} \left(\max \left(\sqrt{\frac{2}{M}} R, \frac{2R}{3M} \right) \right)$. \square

When employing composite channels, as introduced by Ref. [51], to simulate Hamiltonians formed by two contributions $H = A + B$, we need to generalize the result from Theorem 2 from the channel $\mathcal{E}_q(t; N, M)$ to the following one (cf. Eq. (42) in the main text)

$$\Omega_q(t; N, M, r) = \frac{1}{M} \sum_{m=1}^M \left(\tilde{\mathcal{U}}_A \left(\frac{t}{r} \right) \circ \mathcal{E}_q^B \left(\frac{t}{r}; N, 1 \right) \right) \circ \dots \circ \left(\tilde{\mathcal{U}}_A \left(\frac{t}{r} \right) \circ \mathcal{E}_q^B \left(\frac{t}{r}; N, 1 \right) \right), \quad (97)$$

with r outer compositions. Here $\mathcal{E}_q^B(t; N, 1)$ is an importance sampled qDrift channel for approximating the evolution under the B term in the Hamiltonian with only one experiment ($M = 1$) and $\tilde{\mathcal{U}}_A(t)$ is a unitary channel that approximates $\mathcal{U}_A(t)[\rho] = e^{-itA} \rho e^{itA}$. In the main text we considered a first-order Trotter approximation but the next theorem applies to more general cases, including $\tilde{\mathcal{U}}_A$ being an arbitrary unitary matrix.

Theorem 3 (Concentration bound for composite channels). *Let $\Omega_q(t; N, M, r)$ be a composite channel on n qubits for the Hamiltonian $H = A + B$ employing an approximate unitary $\tilde{\mathcal{U}}_A \approx e^{-i\frac{t}{r}A}$ for the time evolution under the term A and total time t/r , a finite importance sampled qDrift channel $\mathcal{E}_q(t; N, 1)$ to approximate evolution under B and r steps using unitaries $W_j = e^{-i\frac{\tau_j}{r}B_j}$ with $\tau_j = (tb_j)/(Nq_j)$. Its concentration around its expectation value can be upper bounded $\forall \epsilon \in [0, 4t\lambda_B]$ as*

$$\Pr \left[\left\| \frac{1}{M} \left[\prod_{s=r}^1 \left(\tilde{\mathcal{U}}_A \prod_{k=N}^1 W_{j_{ksm}} \right) \right] - \left(\tilde{\mathcal{U}}_A \mathbb{E}[W]^N \right)^r \right\| \geq \epsilon/2 \right] \leq 2^{n+1} \exp \left\{ -\frac{NM r \epsilon^2}{11 t^2 \lambda_B^2 (1 + \max_k \omega(k))^2} \right\}. \quad (98)$$

In order to guarantee an approximation error $\epsilon/2$ with probability at least $1 - \delta$, it is then sufficient to take

$$NM r = 11 \frac{t^2 \lambda_B^2}{\epsilon^2} \left(1 + \max_k \omega(k) \right)^2 (n+1) \log \left(\frac{2}{\delta} \right). \quad (99)$$

Proof. The result can be shown by extending the martingales introduced for the proof of the concentration bound Theorem 2. We start by noticing that $\mathbb{E}_q[W_j] = \mathbb{E}_q[W]$ independent on j . For each of the M experiments we consider r sets of N indices for the unitaries forming the product and denote by W_{sj}^m the j -th unitary on the s -th block for the m -th experiment. We can then use them to generalize the martingale from Eq. (79) to

$$\mathcal{B}_k^m = \left(\prod_{s=r}^{s_k+1} \tilde{\mathcal{U}}_A \mathbb{E}_q[W]^N \right) \left(\tilde{\mathcal{U}}_A \mathbb{E}_q[W]^{N-j_k} \prod_{j=j_k}^1 W_{s_k j}^m \right) \left(\prod_{s=s_k-1}^1 \tilde{\mathcal{U}}_A \prod_{j=N}^1 W_{sj}^m \right), \quad (100)$$

for $k = \{0, 1, \dots, Nr\}$. Here we have defined the indices as $s_k = \lfloor (k-1)/N \rfloor + 1$, $j_0 = 0$ while for $k > 0$ $j_k = (k-1)\%N + 1$. The different definition of j_0 is required in order to accommodate the edge case $k = 0$. We also require that $\mathcal{B}_k^0 = 0$ as well as $\mathcal{B}_k^m = 0$ for $m > M$ as was done before for the B_k^m martingale. The causality condition in Definition 1 is automatically satisfied since $\forall m$ \mathcal{B}_k^m is completely determined by the random samples $W_{11}^m, \dots, W_{s_k j_k}^m$ obtained up to the k -th step. The second condition can be checked explicitly, for $\lfloor (k)/N \rfloor = \lfloor (k-1)/N \rfloor$, implying that $s_{k+1} = s_k$ and $j_{k+1} = j_k + 1$, we have

$$\begin{aligned} & \mathbb{E} [\mathcal{B}_{k+1}^m | \mathcal{B}_k^m, \dots, \mathcal{B}_0^m] \\ &= \left(\prod_{s=r}^{s_k+1} \tilde{\mathcal{U}}_A \mathbb{E}_q[W]^N \right) \left(\tilde{\mathcal{U}}_A \mathbb{E}_q[W]^{N-j_{k+1}} \mathbb{E}_q[W_{s_k j_{k+1}}] \prod_{j=j_{k+1}-1}^1 W_{s_k j}^m \right) \left(\prod_{s=s_k-1}^1 \tilde{\mathcal{U}}_A \prod_{j=N}^1 W_{sj}^m \right) \\ &= \left(\prod_{s=r}^{s_k+1} \tilde{\mathcal{U}}_A \mathbb{E}_q[W]^N \right) \left(\tilde{\mathcal{U}}_A \mathbb{E}_q[W]^{N-j_k} \prod_{j=j_k}^1 W_{s_k j}^m \right) \left(\prod_{s=s_k-1}^1 \tilde{\mathcal{U}}_A \prod_{j=N}^1 W_{sj}^m \right) \\ &= \mathcal{B}_k^m \end{aligned} \quad (101)$$

while, for $\lfloor k/N \rfloor = \lfloor (k-1)/N \rfloor + 1$, i.e., $k = nN$ for some integer n and therefore $s_{k+1} = s_k + 1$ together with

$j_k = N$ and $j_{k+1} = 1$, we have the following

$$\begin{aligned}
& \mathbb{E} [\mathcal{B}_{k+1}^m | \mathcal{B}_k^m, \dots, \mathcal{B}_0^m] \\
&= \left(\prod_{s=r}^{s_{k+1}+1} \tilde{U}_A \mathbb{E}_q[W]^N \right) \left(\tilde{U}_A \mathbb{E}_q[W]^{N-j_{k+1}} \mathbb{E}_q[W_{s_k j_{k+1}}] \prod_{j=j_{k+1}-1}^1 W_{s_k j}^m \right) \left(\prod_{s=s_{k+1}-1}^1 \tilde{U}_A \prod_{j=N}^1 W_{s j}^m \right) \\
&= \left(\prod_{s=r}^{s_{k+1}+1} \tilde{U}_A \mathbb{E}_q[W]^N \right) \left(\tilde{U}_A \mathbb{E}[W]^N \right) \left(\tilde{U}_A \prod_{j=N}^1 W_{(s_{k+1}-1)j}^m \right) \left(\prod_{s=s_{k+1}-2}^1 \tilde{U}_A \prod_{j=N}^1 W_{s j}^m \right) \\
&= \left(\prod_{s=r}^{s_k+1} \tilde{U}_A \mathbb{E}_q[W]^N \right) \left(\tilde{U}_A \mathbb{E}_q[W]^{N-j_k} \prod_{j=j_k}^1 W_{s_k j}^m \right) \left(\prod_{s=s_k-1}^1 \tilde{U}_A \prod_{j=N}^1 W_{s j}^m \right) \\
&= \mathcal{B}_k^m,
\end{aligned} \tag{102}$$

by noting that $N - j_k = 0$.

Finally, the new interpolating martingale for the different experiments takes the following form

$$\mathcal{D}_j = \sum_{m=\lfloor j/N \rfloor + 2}^M \mathcal{B}_0^m + \mathcal{B}_{j \% N}^{\lfloor j/N \rfloor + 1} + \sum_{m=1}^{\lfloor j/N \rfloor} \mathcal{B}_N^m, \tag{103}$$

with the special case for $j = 0$ and $j = NM$ given, similarly to before, by

$$\mathcal{D}_0 = \sum_{m=1}^M \mathcal{B}_0^m = M \left(\prod_{s=r}^1 \tilde{U}_A \mathbb{E}_q[W]^N \right) \quad \mathcal{D}_{NM} = \sum_{m=1}^M \mathcal{B}_N^m = \sum_{m=1}^M \left(\prod_{s=r}^1 \tilde{U}_A \prod_{j=N}^1 W_{s j}^m \right). \tag{104}$$

Since between consecutive indices the martingale \mathcal{D}_j changes by only a single unitary, the difference sequence $\mathcal{C}_j = \mathcal{D}_j - \mathcal{D}_{j-1}$ have similar properties as C_j in Eq. (84) above. In particular

$$\|\mathcal{C}_j\| \leq \left\| \frac{\tau_{k_j}}{r} B_{k_j} \right\| + \mathbb{E}_q \left[\left\| \frac{\tau_k}{r} B_k \right\| \right] \leq \frac{t\lambda_B}{Nr} \left(1 + \max_k \omega(k) \right), \tag{105}$$

using the same strategy employed to arrive at Eq. (87). For the variance instead

$$\left\| \sum_{j=1}^{NM} \mathbb{E}[\mathcal{C}_j \mathcal{C}_j^\dagger | \mathcal{C}_{j-1} \dots \mathcal{C}_0] \right\| \leq NM \max_j \|\mathcal{C}_j\|^2 = M \frac{t^2 \lambda_B^2}{Nr} \left(1 + \max_k \omega(k) \right)^2 \tag{106}$$

The result follows then by taking the parameters R and v from Corollary 4 as

$$R = \frac{t\lambda_B}{Nr} \left(1 + \max_k \omega(k) \right), \quad v = MNrR^2, \tag{107}$$

and using Corollary 4 as was done to show Theorem 2. \square

We are now in a position to show an upperbound for the expected error of a composite channel

Corollary 2 (Fluctuation bound for composite channels). *Let $H = A + B$ be a n -qubit Hamiltonian with decomposition as in Eq. (34), $q(j)$ an arbitrary distribution, t the simulation time, N a fixed number of qDrift samples, and M a fixed number of qDrift experiments. Take $\mathcal{U}_H(t)[\rho] = U_H(t)\rho U_H^\dagger(t)$ (with $U_H(t) = e^{-iHt}$), $\tilde{\mathcal{U}}_A(t)$ a first order Trotter approximation of the channel $\mathcal{U}_A(t)$, $\mathcal{E}_q^B(t; N, M)$ the importance sampled qDrift channel for the B term and $\Omega_q(t; N, M, r)$ the importance sampled composite channel. We then have*

$$\begin{aligned}
\mathbb{E} [\|\Omega_q(t; N, M, r) - \mathcal{U}_H\|_\diamond] &\leq 2 \frac{t^2}{r} \left(\Gamma_{comm}^{A,B} + \frac{\lambda_B^2}{N} (1 + \mathbb{E}_p[\omega]) \right) \\
&\quad + \alpha \frac{nt\lambda_B}{NM} \left(1 + \max_k \omega(k) \right) + \alpha \sqrt{\frac{n}{NM}} t\lambda_B \left(1 + \max_k \omega(k) \right),
\end{aligned} \tag{108}$$

where the parameter

$$\Gamma_{comm}^{A,B} = \sum_{i < j} a_i a_j \| [A_i, A_j] \| + \frac{1}{2} \sum_{ij} a_i b_j \| [A_i, B_j] \|, \tag{109}$$

contains the dependence on commutators.

Proof. We prove the corollary in the same way as we did Corollary 1 by relating the diamond norm distance to the operator norm for ensembles of unitary channels, see Lemma 2, and then using the triangle inequality

$$\begin{aligned} \mathbb{E} [\| \Omega_q(t; N, M, r) - \mathcal{U}_H \|_\diamond] &\leq 2\mathbb{E} \left[\left\| \frac{1}{M} \sum_{m=1}^M \prod_{s=r}^1 \prod_{k=N}^1 \tilde{U}_A W_{j_{ksm}} - U_H \right\| \right] \\ &\leq 2 \left\| U_H - \left(\tilde{U}_A \mathbb{E}_q[W]^N \right)^r \right\| + 2\mathbb{E}_q \left[\left\| \frac{1}{M} \sum_{m=1}^M \prod_{s=r}^1 \prod_{k=N}^1 \tilde{U}_A W_{j_{ksm}} - \left(\tilde{U}_A \mathbb{E}_q[W]^N \right)^r \right\| \right]. \end{aligned} \quad (110)$$

The first term can be bounded by using Theorem 1 and the error bounds for composite channel from Eq. (37) to

$$2 \left\| U_H - \mathbb{E}_q \left[\left(\tilde{U}_A \mathbb{E}_q[W]^N \right)^r \right] \right\| \leq 2 \frac{t^2}{r} \left(\Gamma_{\text{comm}}^{A,B} + \frac{\lambda_B^2}{N} (1 + \mathbb{E}_p[\omega]) \right), \quad (111)$$

while for the second we have

$$\begin{aligned} &\mathbb{E}_q \left\| \frac{1}{M} \sum_{m=1}^M \prod_{s=r}^1 \prod_{k=N}^1 \tilde{U}_A W_{m_{ks}} - \left(\tilde{U}_A \mathbb{E}_q[W]^N \right)^r \right\| \\ &= 2 \int_0^\infty \Pr \left(\left\| \frac{1}{M} \sum_{m=1}^M \prod_{s=r}^1 \prod_{k=N}^1 \tilde{U}_A W_{j_{mks}} - \left(\tilde{U}_A \mathbb{E}_q[W]^N \right)^r \right\| \geq \tau \right) d\tau \\ &\leq 2 \int_0^\infty \min \left(1, 2^{n+1} e^{-\frac{3M\tau^2}{6NrR^2+2R\tau}} \right) d\tau \\ &\leq \frac{\alpha}{2} \max \left(\sqrt{\frac{n}{NM r}} t\lambda_B (1 + \max_k \omega(k)), \frac{nt\lambda_B (1 + \max_k \omega(k))}{NM r} \right) \\ &\leq \alpha \left(\sqrt{\frac{n}{NM r}} t\lambda_B (1 + \max_k \omega(k)) + \frac{nt\lambda_B (1 + \max_k \omega(k))}{NM r} \right), \end{aligned} \quad (112)$$

with $R = \frac{t\lambda_B}{Nr} (1 + \max_k \omega(k))$ from above. As in Ref. [43], the integral is evaluated by cutting it into two parts. The first, with a contribution of nearly one, when the denominator in the exponent is bigger than the numerator, i.e., $\tau \leq \max \left(\sqrt{\frac{2n}{M}} R, \frac{2Rn}{3M} \right)$, where α suppresses any constant. The contribution for larger τ is marginal and of order $\mathcal{O} \left(\max \left(\sqrt{\frac{2}{M}} R, \frac{2R}{3M} \right) \right)$. □

B Proofs for the cost reduction

In this section, we will recall and show the results for the particular distribution $q_c(j) = h_j/C_j$, where C_j is the implementation cost of the corresponding term.

Corollary 3. *The expected cost of an important sampled qDrift channel with $N = 1$ sample and $q(j) = q_c(j)$ is always lower than for the standard qDrift*

$$\mathbb{E}_{q_c}[C] \leq \mathbb{E}_p[C]. \quad (113)$$

Proof of Corollary 3.

$$\mathbb{E}_{q_c}[C] = \sum_{j=1}^L q(j) C_j = \frac{\sum_{j=1}^L \frac{h_j}{C_j} C_j}{\sum_{j=1}^L \frac{h_j}{C_j}} = \frac{\sum_{j=1}^L h_j}{\sum_{j=1}^L \frac{h_j}{C_j}} = \frac{1}{\mathbb{E}_p[1/C]} \leq \mathbb{E}_p[C], \quad (114)$$

where we used in the last step Jensen's inequality with $\varphi(C) = 1/C$, which is convex for real positive numbers $C > 0$. □

This result shows that, on average, unitaries sampled from $q(j) = q_c(j)$ are cheaper to implement than from $q(j) = p(j)$. It remains to be shown if the total implementation cost at fixed accuracy is also reduced with this choice of distribution.

Theorem 4 (Cost reduction - pure qDrift). *Let N_p and N_{q_c} be the number of qDrift samples for the two distributions $p(j) = h_j/\lambda$ and $q_c(j) = h_j/(\lambda_c C_j)$ for a given target precision ϵ . The expected cost of the importance sampled qDrift channel is then always smaller than the standard one*

$$N_{q_c} \mathbb{E}_{q_c}[C] \leq N_p \mathbb{E}_p[C]. \quad (115)$$

The number of experiments is instead increased as

$$M_{q_c}(\epsilon, t) = M_p(\epsilon, t) \frac{(1 + \mathbb{E}_p[1/C] \max_j C_j)^2}{1 + \mathbb{E}_p[1/C] \mathbb{E}_p[C]} \quad (116)$$

and, more particularly, independent on the total evolution time t .

Proof of Theorem 4. The factor between N_p and N_{q_c} , see Eq. (24), can be expanded as follows

$$\begin{aligned} \mathbb{E}_p[\omega(j)] &= \sum_j \frac{h_j}{\lambda} \omega(j) = \sum_j \lambda_c \frac{h_j C_j}{\lambda^2} \\ &= \mathbb{E}_p[C] \frac{\lambda_c}{\lambda} = \mathbb{E}_p[C] \mathbb{E}_p[1/C]. \end{aligned} \quad (117)$$

Using this result, together with Corollary 3 and Theorem 1, we find that a sufficient choice for N_{q_c} to guarantee error ϵ over a total time t is at least

$$N_{q_c} = \frac{t^2 \lambda^2}{\epsilon} (1 + \mathbb{E}_p[\omega(j)]), \quad (118)$$

which translates into an average total cost of

$$C_{q_c} = \frac{t^2 \lambda^2}{\epsilon} (1 + \mathbb{E}_p[\omega(j)]) \mathbb{E}_{q_c}[C] = \frac{t^2 \lambda^2}{\epsilon} \frac{1 + \mathbb{E}_p[C] \mathbb{E}_p[1/C]}{\mathbb{E}_p[1/C]} \quad (119)$$

On the other hand, using regular qDrift we have

$$C_p = 2 \frac{t^2 \lambda^2}{\epsilon} \mathbb{E}_p[C]. \quad (120)$$

In order to guarantee a cost reduction we then need

$$\frac{1 + \mathbb{E}_p[C] \mathbb{E}_p[1/C]}{\mathbb{E}_p[1/C]} \leq 2 \mathbb{E}_p[C], \quad (121)$$

or equivalently

$$\mathbb{E}_p[C] \mathbb{E}_p[1/C] \geq 1. \quad (122)$$

This is always satisfied, as one can easily show using Jensen's inequality. The result on the increase in the number of experiments follows instead directly from the definition of the distribution $q_c(j)$ and the sufficient condition Eq. (32) derived from Corollary 1. \square

We will finally show that the cost reduction is also retained when considering composite channels.

Theorem 5 (Cost reduction - composite channel). *Let $C_p(\epsilon, t)$ and $C_{q_c}(\epsilon, t)$ be the expected cost to implement the composite channels $\Omega_p(t; N, M, r)$ and $\Omega_{q_c}(t; N, M, r)$ using two distributions $p(j) = h_j/\lambda$ and $q_c(j) = h_j/(\lambda_c C_j)$ for a given target precision ϵ and propagation time t . Then the following holds*

$$\mathbb{E}_{q_c}[C(\epsilon, t)] \leq \mathbb{E}_p[C(\epsilon, t)]. \quad (123)$$

The number of experiments is instead increased, $M_{q_c}(\epsilon) \geq M_p(\epsilon)$

but retaining the same scaling with error ϵ and system size n and also independent on the total evolution time t .

Proof. As from Eq. (41), we know that the cost of the composite channel can be written as

$$C(\epsilon, t) = \frac{t^2}{\epsilon} \left(\sqrt{\Gamma_{\text{comm}}^{A,B} C_{\text{tot}}^A} + \lambda_B \sqrt{\mathbb{E}_{q_c}[C^B] (1 + \mathbb{E}_p[\omega(j)])} \right)^2. \quad (124)$$

Since the first term is independent from the choice of the sampling distribution, we only need to consider the second one which also appears in Theorem 4 and can therefore be bounded as

$$\mathbb{E}_{q_c} [C^B] (1 + \mathbb{E}_p[\omega(j)]) \leq 2\mathbb{E}_p [C^B] . \quad (125)$$

For the number of experiments instead, using Eq. (48) and Eq. (49), for the $p(j)$ distribution we have

$$M_p(\epsilon) = \frac{n}{\epsilon} \frac{2\alpha^2\kappa}{(\kappa-1)^2} \frac{\lambda_B}{\sqrt{C_{tot}^A}} \frac{2^{3/2}\sqrt{\mathbb{E}_p[C^B]}}{\sqrt{\Gamma_{\text{comm}}^{A,B}C_{tot}^A + \lambda_B\sqrt{2\mathbb{E}_p[C^B]}}}, \quad (126)$$

while for the importance sampled distribution $q_c(j)$ we find

$$\begin{aligned} M_{q_c}(\epsilon) &= \frac{n}{\epsilon} \frac{2\alpha^2\kappa}{(\kappa-1)^2} \frac{\lambda_B}{\sqrt{C_{tot}^A}} \frac{(1 + \mathbb{E}_p[1/C^B] \max_j C_j^B)^2 \sqrt{\frac{1}{(1+\mathbb{E}_p[1/C^B]\mathbb{E}_p[C^B])\mathbb{E}_p[1/C^B]}}}{\sqrt{\Gamma_{\text{comm}}^{A,B}C_{tot}^A + \lambda_B\sqrt{(1 + \mathbb{E}_p[1/C^B]\mathbb{E}_p[C^B])/\mathbb{E}_p[1/C^B]}}} \\ &\geq \frac{n}{\epsilon} \frac{2\alpha^2\kappa}{(\kappa-1)^2} \frac{\lambda_B}{\sqrt{C_{tot}^A}} \frac{(1 + \mathbb{E}_p[1/C^B] \max_j C_j^B)^2 \sqrt{\frac{1}{(1+\mathbb{E}_p[1/C^B]\mathbb{E}_p[C^B])\mathbb{E}_p[1/C^B]}}}{\sqrt{\Gamma_{\text{comm}}^{A,B}C_{tot}^A + \lambda_B\sqrt{2\mathbb{E}_p[C^B]}}} \\ &\geq \frac{n}{\epsilon} \frac{2\alpha^2\kappa}{(\kappa-1)^2} \frac{\lambda_B}{\sqrt{C_{tot}^A}} \frac{2\sqrt{1 + \mathbb{E}_p[1/C^B] \max_j C_j^B} \sqrt{\mathbb{E}_p[C^B]}}{\sqrt{\Gamma_{\text{comm}}^{A,B}C_{tot}^A + \lambda_B\sqrt{2\mathbb{E}_p[C^B]}}} \\ &\geq \frac{n}{\epsilon} \frac{2\alpha^2\kappa}{(\kappa-1)^2} \frac{\lambda_B}{\sqrt{C_{tot}^A}} \frac{2^{3/2}\sqrt{\mathbb{E}_p[C^B]}}{\sqrt{\Gamma_{\text{comm}}^{A,B}C_{tot}^A + \lambda_B\sqrt{2\mathbb{E}_p[C^B]}}} = M_p(\epsilon), \end{aligned} \quad (127)$$

where the second line is obtained by remarking that, due to Jensen's inequality,

$$\frac{1 + \mathbb{E}_p[C^B]\mathbb{E}_p[1/C^B]}{\mathbb{E}_p[1/C^B]} \leq 2\mathbb{E}_p[C^B] . \quad (128)$$

In order to get to the third line we used the inequality

$$\frac{(1 + \mathbb{E}_p[1/C^B] \max_j C_j^B)^3}{1 + \mathbb{E}_p[1/C^B]\mathbb{E}_p[C^B]} \geq 4\mathbb{E}_p[1/C^B]\mathbb{E}_p[C^B] , \quad (129)$$

and the last inequality follows from $1 + \mathbb{E}_p[1/C^B] \max_j C_j^B \geq 2$. \square