
Cambridge Quantum

The Problem with
Grover-Rudolph State
Preparation for Quantum
Monte-Carlo Integration

by Steven Herbert

Cambridge Quantum
Department of Computer Science and Technology
University of Cambridge, United Kingdom
STEVEN HERBERT
steven.herbert@cambridgequantum.com

Cambridge Quantum
Terrington House
13-15 Hills Road
Cambridge CB2 1NL
United Kingdom

Published by Cambridge Quantum

06 January 2021

The Problem with Grover-Rudolph State Preparation for Quantum Monte-Carlo Integration

Steven Herbert*

Cambridge Quantum Computing Ltd, 9a Bridge Street, Cambridge, CB2 1UB, UK and
Department of Computer Science and Technology, University of Cambridge, UK

We prove that there is no quantum speed-up when using quantum Monte-Carlo integration to estimate the mean (and other moments) of analytically-defined log-concave probability distributions prepared as quantum states using the Grover-Rudolph method.

I. INTRODUCTION

A frequent, and much relied-upon claim in quantum finance and quantum data-science literature is that there is an automatic quadratic quantum advantage when performing Monte-Carlo integration to approximate the expectation of some probability distribution, if the distribution is efficiently integrable [1–9]. In this paper we show that this is not true in general. The crucial factor is that the quadratic speed-up arises from comparing classical complexity to quantum *query* complexity, and if we take into account the additional operations needed to build increasingly precise oracles for the quantum Monte-Carlo (i.e., as the overall accuracy required increases), then we find that there is no quantum advantage.

To see this, we can start by looking at the essential problem that Monte-Carlo integration solves. If we have some probability density function $p(x)$, for which we want to calculate the expectation then we must calculate the integral:

$$\mathbb{E}(x) = \int_x xp(x) dx \quad (1)$$

However, should this integral not be analytically calculable, then we may resort to Monte-Carlo integration. This entails sampling from $p(x)$ a number (say N_s) times, and averaging the samples to approximate the expectation:

$$\mathbb{E}(x) \approx \frac{1}{N_s} \sum_{j=1}^{N_s} x_j \quad (2)$$

where x_j are samples from $p(x)$. In this case, the approximation error is well-defined: the root mean squared error (RMSE) is proportional to $1/\sqrt{N_s}$ [10].

If, however, we are in possession of some quantum circuit that prepares the n -qubit state:

$$|\psi_n\rangle = \sum_{i \in \{0,1\}^n} \sqrt{p_i} |i\rangle \quad (3)$$

where $\{p_i\}$ is a 2^n -point discretisation of $p(x)$, then we can use the *quantum* Monte-Carlo to achieve RMSE that

decays as $\Theta(1/N_q)$, where N_q is the number of times that the circuit preparing $|\psi_n\rangle$ is queried as part of an oracle circuit [11–16]. That is, a quadratic speed-up in query complexity, where a classical query involves one sample from $p(x)$, and a quantum query involves one oracle call to the state preparation circuit.

This quadratic quantum advantage can indeed be realised if we are in possession of a circuit that *exactly* encodes $|\psi_n\rangle$. However, in general, the quantum circuit will only *approximately* prepare $|\psi_n\rangle$, and furthermore in order to more accurately prepare $|\psi_n\rangle$, the process of state preparation will itself generally be more computationally expensive.

In this paper we address the case where the Grover-Rudolph method [17] is used to prepare $|\psi_n\rangle$ encoding some log-concave distribution [18]. Notably, many very commonly-used distributions such as the exponential and normal families are log-concave. The Grover-Rudolph method takes $\Theta(n)$ steps to prepare a discrete distribution with $N = 2^n$ points, and thus is efficient in this sense – with the important caveat that $p(x)$ need to be efficiently integrable. The authors claim that the efficiency holds, even when the integration is performed numerically, so long as an *efficient* approximate integration algorithm exists. This claim is, of course, almost a tautology, but there is an important subtlety here: In their paper, Grover and Rudolph implicitly take “efficiently” to mean “with only a polynomial time overhead in the problem size”, which would obviously preserve any *exponential* speed-ups in algorithms that take $|\psi_n\rangle$ as an input. However, in the case where quantum Monte-Carlo takes $|\psi_n\rangle$ as an input, only a *quadratic* speed-up is available, and so it is necessary to demand an accordingly stricter standard for what “efficiently integrable” should actually mean. The central claim of this paper is that when classical Monte-Carlo integration is used in the state preparation circuit then this *is not* sufficiently efficient to preserve the quadratic quantum advantage in quantum Monte-Carlo.

II. PRELIMINARIES

For simplicity, we assume that the distribution of interest has been discretised over $N = 2^n$ points (for some integer n) and shifted and scaled such that its domain is $\{0, 1, \dots, 2^n - 1\}$. We let μ be the mean of $p(x)$ and

* Steven.Herbert@cambridgequantum.com

$\hat{\mu} = \mathbb{E}(x)$ be an estimate of μ . As mentioned in Section I the natural measure of error for Monte-Carlo integration is the RMSE, defined:

$$\hat{\epsilon} = \sqrt{\mathbb{E}((\hat{\mu} - \mu)^2)} \quad (4)$$

Grover-Rudolph State Preparation

The Grover-Rudolph method constructs the state $|\psi\rangle$ iteratively, at each step adding a qubit and doubling the number of points in the discrete probability distribution. Since each point in the discrete distribution corresponds to an interval, the new distribution is obtained by splitting each interval in two; the allocation of probability mass to the left and right halves is computed by integration. Moreover, the quantum-nature of the algorithm allows this to happen for all points in superposition (i.e., in a single step). This is achieved by applying the rotation which has the following effect on the i^{th} computational basis state:

$$|\theta_i\rangle |i\rangle |0\rangle \rightarrow |\theta_i\rangle |i\rangle (\cos \theta_i |0\rangle + \sin \theta_i |1\rangle) \quad (5)$$

where:

$$\theta_i = \arccos \sqrt{f(i)} \quad (6)$$

and:

$$f(i) = \frac{\int_{x_L^i}^{\frac{x_R^i + x_L^i}{2}} p(x) dx}{\int_{x_L^i}^{x_R^i} p(x) dx} \quad (7)$$

where x_L^i and x_R^i are the left- and right-hand boundaries of the i^{th} interval respectively. The very first Grover-Rudolph iteration is an unconditional rotation, whose circuit is shown in Fig. 1(a). The circuit for the subsequent iterations is shown in Fig. 1(b) – in particular, the iterations of the Grover-Rudolph method are indexed “ m ”, and Fig. 1(b) shows the circuit for adding the $(m+1)^{\text{th}}$ qubit to the state.

The Grover-Rudolph paper simply states that this procedure readily generalises to higher dimensions, however we give an explicit construction to this effect in Appendix A. To simplify the following analysis, we consider only a univariate probability distribution, however we note that Monte-Carlo integration is unlikely to be the method of numerical integration of choice in this setting. Therefore in Appendix A we also briefly discuss why the following main result will trivially extend to the case of using quantum Monte-Carlo integration to estimate the mean of a multivariate distribution, when the Grover-Rudolph method is used to prepare the state encoding that distribution.

III. MAIN RESULT

As stated in Section I, the RMSE of a classical Monte-Carlo estimate of the mean is proportional to $\frac{1}{\sqrt{N_s}}$, which

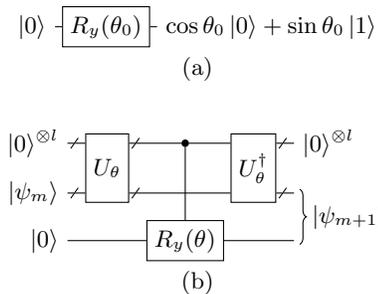


FIG. 1. Quantum circuits for the Grover-Rudolph method: (a) for the first iteration; and (b) for subsequent iterations, where $|0\rangle^{\otimes l} |i\rangle \xrightarrow{U_\theta} |\theta_i\rangle |i\rangle$. Note that l is the number of qubits needed to represent θ to appropriate accuracy; and also that the controlled $R_y(\theta)$ gate shown is shorthand for a series of $R_y(\theta)$ gates with fixed rotation angles, each conditionally controlled by the various qubits in the θ register, such that the overall action is a rotation by θ .

corresponds to a sample complexity $N_s \in \Theta\left(\frac{1}{\epsilon^2}\right)$. We show in Appendix B that, when implemented on a digital classical computer this corresponds to a computational complexity of:

$$\tilde{\Theta}\left(\frac{1}{\epsilon^2}\right) \quad (8)$$

That is, the computational complexity is only a negligible poly-logarithmic factor greater than the sample complexity (as indicated by the tilde). We now show that there is no quantum advantage if the Grover-Rudolph method is used to prepare a state for quantum Monte-Carlo.

Theorem 1. *To achieve a RMSE of $\hat{\epsilon}$ using an unbiased quantum Monte-Carlo estimation method requires $\tilde{\Omega}\left(\frac{1}{\epsilon^2}\right)$ operations when the Grover-Rudolph method is used to prepare some log-concave distribution as a quantum state.*

Remark. *The requirement that the quantum Monte-Carlo estimation is unbiased is included purely for technical reasons in the following proof. In practice, even a biased estimator would yield no advantage, unless that bias were to be correlated with the error of an ostensibly independent classical Monte-Carlo estimation of some aspect of the distribution, in such a way the two errors cancel out. It is, however, worth noting that quantum Monte-Carlo techniques based on phase-estimation are likely to be unbiased in general; and also the NISQ-appropriate algorithm “amplitude estimation without phase estimation” claims to be asymptotically unbiased [12].*

Proof. The proof strategy is thus: firstly, we assume an error is present in the preparation of $|\psi_n\rangle$ and show the effect of this on the RMSE; and secondly we demonstrate this such a state-preparation error is present when the Grover-Rudolph method is used, and give its scaling.

To assess the convergence rate of quantum Monte-Carlo, we make the conservative assumption that error is

only incurred in the first Grover-Rudolph iteration, i.e., the circuit in Fig. 1(a) performs a rotation by an erroneous angle θ_0 . Given that the first iteration of Grover-Rudolph calculates the amount of probability mass that is distributed over the left-hand half of the distribution, it is convenient in the following analysis to consider the error in the probability mass function directly. Formally, we let:

$$p_l = \sum_{i=0}^{2^{n-1}-1} p_i, \quad (9)$$

and ϵ_l be the error incurred in the first Grover-Rudolph iteration, such that rather than preparing state (3), we instead prepare the erroneous state:

$$|\psi'\rangle = \sum_{i=0}^{2^{n-1}-1} \sqrt{\frac{p_i(p_l + \epsilon_l)}{p_l}} |i\rangle + \sum_{i=2^{n-1}}^{2^n-1} \sqrt{\frac{p_i(1-p_l - \epsilon_l)}{1-p_l}} |i\rangle \quad (10)$$

We can now evaluate the mean, μ' , of the erroneous distribution that has been encoded in the state, $|\psi'\rangle$:

$$\begin{aligned} \mu' &= \sum_{i=0}^{2^{n-1}-1} i \frac{p_i(p_l + \epsilon_l)}{p_l} + \sum_{i=2^{n-1}}^{2^n-1} i \frac{p_i(1-p_l - \epsilon_l)}{1-p_l} \\ &= \sum_{i=0}^{2^n-1} i p_i + \epsilon_l \left(\sum_{i=0}^{2^{n-1}-1} i \frac{p_i}{p_l} - \sum_{i=2^{n-1}}^{2^n-1} i \frac{p_i}{1-p_l} \right) \\ &= \mu - k\epsilon_l \end{aligned} \quad (11)$$

where k is the mean of the left-hand half of the discretised distribution minus the mean of the right-hand half of the discretised distribution. That is, a non-zero constant that depends only on the distribution being encoded and not the sampling process. We can see that the error, ϵ_l , in the first iteration of the Grover-Rudolph state preparation method is inherited as a proportional term in the error of the distribution mean. It follows that when performing the quantum Monte-Carlo, we are actually sampling from an erroneous distribution whose mean errs from the mean of the actual distribution of interest by a factor proportional to ϵ_l . Thus, we can consider the overall mean squared error, which is equal to:

$$\begin{aligned} \hat{\epsilon}^2 &= \mathbb{E} \left((\mu - \hat{\mu}(N_q))^2 \right) \\ &= \mathbb{E} \left((\mu' + k\epsilon_l - \hat{\mu}(N_q))^2 \right) \\ &= \mathbb{E} \left((\mu' - \hat{\mu}(N_q))^2 \right) + k^2 \mathbb{E}(\epsilon_l^2) \\ &\quad + \mathbb{E}_{\epsilon_l} (k\epsilon_l \mathbb{E}_{\hat{\mu}|\epsilon_l} (\mu' - \hat{\mu}(N_q))) \\ &= \mathbb{E} \left((\mu' - \hat{\mu}(N_q))^2 \right) + k^2 \mathbb{E}(\epsilon_l^2) \end{aligned} \quad (12)$$

where $\hat{\mu}(N_q)$ is the estimate of the mean obtained by performing quantum Monte-Carlo such that the state preparation circuit is queried N_q times. This derivation uses the fact that ϵ_l is itself a random variable, and that the

quantum estimate is unbiased (i.e., by the previous assumption), so $\mathbb{E}(\mu' - \hat{\mu}(N_q)) = 0$. Thus the effect of (12) is to separate the error pertaining to the construction of the state-preparation circuit, and the error pertaining to the inaccuracy of the quantum Monte-Carlo itself. We know that the first term in the sum in (12) can be suppressed as $\Theta(1/N_q^2)$. Thus we turn our attention to the second term, which relates to the state preparation.

Grover and Rudolph specify that log-concave probability distributions should be numerically integrated using the technique described by Applegate and Kannan [18] which concerns the general case of numerically evaluating the integral over some region of the support of the probability mass. This technique essentially involves dividing the region into halves, and performing Monte-Carlo integration to evaluate the ratio of probability mass in each half – and continuing recursively (choosing the half with greatest probability mass to further divide each time), until the region is sufficiently small that the probability mass is essentially constant therein. However, we can easily see that this approach can be tweaked slightly in the case in which the integral of the entire area sums to one – as is the case for the *first* iteration of Grover-Rudolph. In this case, the simpler method of sampling N'_s times and then counting up the proportion that fall in the region of interest can be used – which is manifestly less computationally complex for given accuracy. We have that the proportion of random samples that will fall in the left-hand half of the distribution is expected to be p_l . Letting $X \sim p(x)$ be random samples from the distribution of interest, and \mathcal{L} be the region covered by the left-hand half of the distribution, we can see that the estimate, \hat{p}_l of p_l corresponds to a binomial distribution:

$$\hat{p}_l \sim \mathcal{B}(N'_s, p_l) \quad (13)$$

We can thus bound the mean squared error, $\mathbb{E}(\epsilon_l^2)$, of this Monte-Carlo estimate of p_l :

$$\begin{aligned} \mathbb{E}(\epsilon_l^2) &= \mathbb{E} \left((p_l - \hat{p}_l)^2 \right) \\ &= \mathbb{E} \left(\left(p_l - \frac{1}{N'_s} \sum_{X \in \mathcal{L}} 1 \right)^2 \right) \\ &= \frac{1}{(N'_s)^2} \mathbb{E} \left(\left(N'_s p_l - \sum_{X \in \mathcal{L}} 1 \right)^2 \right) \\ &= \frac{1}{(N'_s)^2} N'_s p_l (1 - p_l) \\ &= \frac{p_l(1-p_l)}{N'_s} \end{aligned} \quad (14)$$

So we can see that, whilst the error in the quantum Monte-Carlo is suppressed as $\Theta(1/N_q^2)$, the error in the classical Monte-Carlo required to prepare the state used in the quantum Monte-Carlo is only suppressed as $\Theta(1/N'_s)$. Returning to (12), we can see that the classical Monte-Carlo sampling required to prepare the state sufficiently accurately is therefore the bottleneck, thus giving

us an overall computational complexity of:

$$\tilde{\Omega} \left(\frac{1}{\epsilon^2} \right) \quad (15)$$

where Ω is used because this is a lower-bound: we have considered only the error in the first Grover-Rudolph iteration; and furthermore it is $\tilde{\Omega}$, because the classical complexity of the Monte-Carlo sampling will be subject to the same poly-logarithmic overhead as identified in Appendix B. \square

IV. DISCUSSION

At its core, the result we show in this paper is the simple observation that, if classical sampling is required to prepare a quantum state which encodes a probability distribution, then the best estimate of the mean that we can obtain is the average of the samples themselves: putting these samples into and out-of a quantum computer cannot, in and of itself, beget more accuracy. In particular, we have shown that there is no quantum advantage when the Grover-Rudolph method is used to prepare the state for quantum Monte-Carlo.

For simplicity and definiteness, the analysis in this paper has considered a very restricted error model, which means that the computational overhead required to build a correspondingly more accurate state-preparation circuit as the desired estimation accuracy grows falls solely into a single classical pre-processing step to calculate θ_0 and so the quantum complexity does not increase. In reality, however, the circuit depth of the state-preparation circuit *will* grow with the desired accuracy. This is because the Grover-Rudolph method specifies that the Monte-Carlo integration required in the state preparation should be performed using a

register of qubits initialised in a random computational basis state, meaning that the numerical integration is to be achieved using quantum rather than classical operations (in fact this is a necessity in all but the first Grover-Rudolph iteration). It follows that the depth of the circuit U_θ in Fig. 1(b) will grow with the desired accuracy.

This conclusion adds theoretical basis to the emerging picture that Grover-Rudolph state preparation is too costly in practice [6, Tab. 1]. This does not, however, condemn the whole idea of using quantum Monte-Carlo to achieve quantum speed-ups in data-science. For example, even though the Grover-Rudolph paper is very heavily cited, there are alternative ways to prepare the state (eg. Refs. [6, 19]), which may yield a quantum speed-up. Moreover, in cases in which we do not simply want to evaluate the expectation of the probability distribution itself, but rather to first apply function to the random variables before the expectation is to be taken, then there could be a quantum speed-up overall if this function that is applied to the random variables is sufficiently complex. Finally, it is worth noting that, Grover-Rudolph also applies to situations in which the function *can* be analytically integrated, which may occur even when the expectation *cannot* be (that is, functions for which $\int p(x)dx$ between appropriate limits is analytically calculable, but where $\int xp(x)dx$ isn't): in this case, the absence of classical Monte-Carlo in the state-preparation circuit means that the quadratic quantum advantage will be upheld, even when considering computational rather than query complexity.

ACKNOWLEDGEMENT

The author thanks Ross Duncan, Cristina Cirstoiu and Luciana Henaut for their comments and suggestions.

-
- [1] P. Rebentrost, B. Gupt, and T. R. Bromley, "Quantum computational finance: Monte carlo pricing of financial derivatives," *Physical Review A*, vol. 98, no. 2, Aug 2018. [Online]. Available: <http://dx.doi.org/10.1103/PhysRevA.98.022321>
 - [2] S. Woerner and D. J. Egger, "Quantum risk analysis," *npj Quantum Information*, vol. 5, no. 1, Feb 2019. [Online]. Available: <http://dx.doi.org/10.1038/s41534-019-0130-6>
 - [3] N. Stamatopoulos, D. J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner, "Option pricing using quantum computers," *Quantum*, vol. 4, p. 291, Jul 2020. [Online]. Available: <http://dx.doi.org/10.22331/q-2020-07-06-291>
 - [4] R. Orús, S. Mugel, and E. Lizaso, "Quantum computing for finance: Overview and prospects," *Reviews in Physics*, vol. 4, p. 100028, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405428318300571>
 - [5] D. J. Egger, R. Garcia Gutierrez, J. Cahue Mestre, and S. Woerner, "Credit risk analysis using quantum computers," *IEEE Transactions on Computers*, pp. 1–1, 2020.
 - [6] S. Chakrabarti, R. Krishnakumar, G. Mazzola, N. Stamatopoulos, S. Woerner, and W. J. Zeng, "A threshold for quantum advantage in derivative pricing," 2020.
 - [7] P. Rebentrost and S. Lloyd, "Quantum computational finance: quantum algorithm for portfolio optimization," 2018.
 - [8] K. Kaneko, K. Miyamoto, N. Takeda, and K. Yoshino, "Quantum pricing with a smile: Implementation of local volatility model on quantum computer," 2020.
 - [9] D. J. Egger, C. Gambella, J. Marecek, S. McFaddin, M. Mevissen, R. Raymond, A. Simonetto, S. Woerner, and E. Yndurain, "Quantum computing for finance: State-of-the-art and future prospects," *IEEE Transactions on Quantum Engineering*, vol. 1, pp. 1–24, 2020.

- [10] R. E. Caflisch, “Monte Carlo and quasi-Monte Carlo methods,” *Acta Numerica*, vol. 7, p. 1–49, 1998.
- [11] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, “Quantum amplitude amplification and estimation,” *Quantum Computation and Information*, p. 53–74, 2002. [Online]. Available: <http://dx.doi.org/10.1090/conm/305/05215>
- [12] Y. Suzuki, S. Uno, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto, “Amplitude estimation without phase estimation,” *Quantum Information Processing*, vol. 19, no. 2, Jan 2020. [Online]. Available: <http://dx.doi.org/10.1007/s11128-019-2565-2>
- [13] D. Grinko, J. Gacon, C. Zoufal, and S. Woerner, “Iterative quantum amplitude estimation,” *npj Quantum Information*, vol. 7, no. 1, Mar 2021. [Online]. Available: <http://dx.doi.org/10.1038/s41534-021-00379-1>
- [14] A. Montanaro, “Quantum speedup of Monte Carlo methods,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2181, p. 20150301, 2015.
- [15] S. Aaronson and P. Rall, “Quantum approximate counting, simplified,” *Symposium on Simplicity in Algorithms*, p. 24–32, Jan 2020. [Online]. Available: <http://dx.doi.org/10.1137/1.9781611976014.5>
- [16] K. Nakaji, “Faster amplitude estimation,” 2020.
- [17] L. Grover and T. Rudolph, “Creating superpositions that correspond to efficiently integrable probability distributions,” 2002.
- [18] D. Applegate and R. Kannan, “Sampling and integration of near log-concave functions,” in *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, ser. STOC ’91. New York, NY, USA: Association for Computing Machinery, 1991, p. 156–163. [Online]. Available: <https://doi.org/10.1145/103418.103439>
- [19] A. Carrera Vazquez and S. Woerner, “Efficient state preparation for quantum amplitude estimation,” *Physical Review Applied*, vol. 15, no. 3, Mar 2021. [Online]. Available: <http://dx.doi.org/10.1103/PhysRevApplied.15.034027>

Appendix A: Grover-Rudolph for Multivariate probability distributions

Suppose we have a d -dimensional multivariate probability distribution, $p(x^{(1)}, x^{(2)}, \dots, x^{(d)})$. Assuming each of these dimensions has 2^n equally spaced points of probability mass, our aim is to use the Grover-Rudolph method a number of times to prepare a state of the form:

$$|\psi\rangle = \sum_{x^{(1)} \dots x^{(d)}} \sqrt{p(x^{(1)}, \dots, x^{(d)})} |x^{(1)} \dots x^{(d)}\rangle \quad (\text{A1})$$

which thus encodes the multivariate probability distribution by using d registers, each of n qubits, to represent the points of probability mass. The Grover-Rudolph method dictates that p should be efficiently integrable, which in the multivariate case we may take to mean that we can efficiently compute:

$$\int_{x_1^{(1)}}^{x_u^{(1)}} \dots \int_{x_1^{(d)}}^{x_u^{(d)}} p(x^{(1)}, \dots, x^{(d)}) dx^{(1)} \dots dx^{(d)} \quad (\text{A2})$$

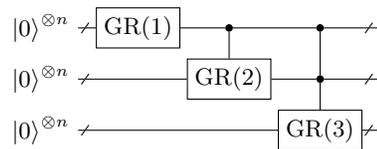


FIG. 2. Grover-Rudolph state preparation for tri-variate distribution. GR(1) denotes the preparation of the marginal distribution of the first dimension of the multivariate distribution (as per normal univariate case of Grover-Rudolph); GR(2) denotes the use of Grover-Rudolph to prepare the conditional distribution of the second dimension given the first (with the third dimension marginalised out) – the conditioning in the integral achieved by the quantum control; and GR(3) denotes the use of Grover-Rudolph to prepare the conditional distribution of the third dimension given the first and second.

for any limits of integration. In particular, we can compute the integral over the entire support of each dimension, except for one dimension where we set the limits we want – thus achieving integration of the marginal distribution of the dimension in question. For example, if we wish to integrate the marginal distribution of the first dimension, we get:

$$\begin{aligned} & \int_{x_1^{(1)}}^{x_u^{(1)}} p(x^{(1)}) dx^{(1)} \\ &= \int_{x_1^{(1)}}^{x_u^{(1)}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x^{(1)}, \dots, x^{(d)}) dx^{(1)} dx^{(2)} \dots dx^{(d)} \end{aligned} \quad (\text{A3})$$

where $p(x^{(1)})$ is the marginal distribution of $x^{(1)}$. In order to use the Grover-Rudolph method to prepare a quantum state encoding p , we further require that efficient integration of some conditional distributions is possible. In particular, for any d' such that $1 \leq d' < d$ we must be able to efficiently compute the integral:

$$\int_{x_1^{(d'+1)}}^{x_u^{(d'+1)}} \dots \int_{x_1^{(d)}}^{x_u^{(d)}} p(x^{(d'+1)}, \dots, x^{(d)} | x^{(1)}, \dots, x^{(d')}) dx^{(d'+1)} \dots dx^{(d)} \quad (\text{A4})$$

once again, including the cases where some of the integrals are over the entire domain. We can thus prepare the joint distribution by noticing the decomposition:

$$\begin{aligned} p(x^{(1)}, x^{(2)}, \dots, x^{(d)}) &= p(x^{(1)}) p(x^{(2)} | x^{(1)}) \\ &\dots p'(x^{(d)} | x^{(1)}, \dots, x^{(2)}) \end{aligned} \quad (\text{A5})$$

and then using Grover-Rudolph a total of d times to prepare first the marginal distribution of $x^{(1)}$, then the conditional distribution of $x^{(2)}$ given $x^{(1)}$ (marginalising out all other variables), through to the conditional distribution of $x^{(d)}$ given $x^{(d-1)} \dots x^{(1)}$. These d uses of the

Grover-Rudolph method can be performed coherently in superposition, in exactly the same way that the Grover-Rudolph algorithm itself works. Figure 2 gives the circuit diagram to achieve this for the trivariate case.

As preparing a d -dimensional multivariate probability distribution simply amounts to d uses of the Grover-Rudolph method, we can see that our main result presented for the case of univariate probability distributions will extend to any multivariate probability distribution where numerical integration is used.

Appendix B: Complexity of Classical Monte-Carlo

In this section we show that, when classical Monte-Carlo is performed on a digital computer, the number of bits required is only poly-logarithmic in the reciprocal of the desired RMSE, and furthermore that all operations have complexity that is polynomial in the *number of bits*. Therefore, performing the classical Monte-Carlo on a digital computer only introduces a poly-logarithmic overhead in complexity compared to the sample complexity.

We start with the definition of the RMSE:

$$\hat{\epsilon}^2 = \mathbb{E} \left(\left(\mu - \frac{1}{N_s} \sum_{j=1}^{N_s} X_j \right)^2 \right) \quad (\text{B1})$$

However, when performed on a digital computer, X_j will not be samples from the distribution of interest, but rather an approximate version thereof. This approximation arises because of two inaccuracies introduced by performing the sampling on a digital computer. Firstly, the samples will not be drawn from the (continuous) distribution itself, but rather from some discretised version.

Let the support of $p(x)$ be between $x_l \leq x < x_u$, where the distribution is shifted such that $x_l \geq 0$, and also let $|\frac{dp(x)}{dx}| \leq \beta$ for some constant β in the region $x_l \leq x < x_u$. Treating $p(x)$ as having only finite support is justified as it can be thought of as the result of a pre-truncating step that is necessary in both the classical and quantum cases.

Next, we must define a *discretised* version of $p(x)$ with N_d equally spaced intervals between x_l and x_u , for simplicity we let N_d be a power of 2 and also let $\Delta x = x_u - x_l$.

$$X = \left\{ x | x = x_l + i \frac{\Delta x}{N_d} \text{ for } i = 0 \dots N_d - 1 \right\} \quad (\text{B2})$$

From this, we can define the mean of the discretised distribution:

$$\mu_d = \frac{1}{N_d} \sum_{i=0}^{N_d-1} X_i p(X_i) \quad (\text{B3})$$

This enables us to establish the discretisation error incurred when estimating the error on a digital classical or

quantum computer:

$$\mu_d - \beta/2 \left(\frac{\Delta x}{N_d} \right)^2 \leq \mu \leq \mu_d + \beta/2 \left(\frac{\Delta x}{N_d} \right)^2 \quad (\text{B4})$$

This is because the definition of X is such that a probability mass function is formed by taking the left-Riemann sum of $p(x)$. So it follows that the maximum discrepancy for each region in the sum is upper-bounded by the area of a right-angle triangle of base $\frac{\Delta x}{N}$ and height $\beta \frac{\Delta x}{N}$. Thus, in the worst case, the probability of each region errs by this maximum amount, and so the mean itself errs by the same. From this we can upper-bound the discretisation error, ϵ_d :

$$\epsilon_d = |\mu - \mu_d| \leq \beta/2 \left(\frac{\Delta x}{N_d} \right)^2 \quad (\text{B5})$$

We now turn to the second error which is introduced by performing the classical Monte-Carlo on a digital computer, namely the sampling error. To ensure consistency with the quantum Monte-Carlo case, we consider the sampling technique described by Applegate and Kannan [18] according to which sampling from $p(x)$ requires

$$\mathcal{O} \left(\log \frac{1}{\epsilon_{s_{max}}} + \log N_d \right) \quad (\text{B6})$$

operations to achieve a maximum sampling error of $\epsilon_{s_{max}}$.

Thus the discrepancy in the expectation of the samples, μ_s , and the actual mean of the discretised distribution, μ_d , can be bounded by considering the case in which each of the discretised probability masses errs by the maximum amount $\epsilon_{s_{max}}$, thus meaning that the entire distribution is inflated (or deflated) by a factor $1 + \epsilon_{s_{max}}$ (or $1 - \epsilon_{s_{max}}$) and so the mean is increases (or decreased) accordingly. So we get:

$$(1 - \epsilon_{s_{max}})\mu_d < \mu_s < (1 + \epsilon_{s_{max}})\mu_d \quad (\text{B7})$$

Of course, this is a bounding case, as it will remain the case that the probabilities in the distribution with sampling errors must still sum to one, hence why they are strict inequalities in (B7), as these bounds cannot actually be saturated.

Returning to (B1), we can now see that we are sampling from a distribution whose mean differs from the mean of the (continuous) distribution of interest by the discretisation and sampling errors (ϵ_d and ϵ_s respectively), and we define $\mu' = \mu + \epsilon_d + \epsilon_s$ such that it is the mean of the distribution we are sampling from. We

can therefore re-express (B1):

$$\begin{aligned}
\hat{\epsilon}^2 &= \mathbb{E} \left(\left(\mu' + \epsilon_d + \epsilon_s - \frac{1}{N_s} \sum_{j=1}^{N_s} X_j \right)^2 \right) \\
&= \mathbb{E} \left(\left(\mu' - \frac{1}{N_s} \sum_{j=1}^{N_s} X_j \right)^2 \right) + (\epsilon_d + \epsilon_s)^2 \\
&\quad + 2(\epsilon_d + \epsilon_s) \mathbb{E} \left(\mu' - \frac{1}{N_s} \sum_{s=1}^{N_s} X_j \right) \\
&= \mathbb{E} \left(\left(\mu' - \frac{1}{N_s} \sum_{j=1}^{N_s} X_j \right)^2 \right) + (\epsilon_d + \epsilon_s)^2 \\
&\leq \mathbb{E} \left(\left(\mu' - \frac{1}{N_s} \sum_{j=1}^{N_s} X_j \right)^2 \right) + (\epsilon_{d_{max}} + \epsilon_{s_{max}})^2 \quad (\text{B8})
\end{aligned}$$

Thus we have expressed the RMSE as the actual error that is suppressed by the Monte-Carlo integration (i.e., the first term in the right-hand side (RHS) of (B8)), and also the errors pertaining to the discretisation and sampling (i.e., the second term in the RHS of (B8)).

The first term is the error suppressed by classical Monte-Carlo, i.e., such that $\mathbb{E} \left(\left(\mu' - \frac{1}{N_s} \sum_{j=1}^{N_s} X_j \right)^2 \right) \in$

$\Theta\left(\frac{1}{N_s}\right)$; more pertinently, we can see from the second term in (B8) that to achieve RMSE $\hat{\epsilon}$ it suffices to set up the digital computation such that $\epsilon_{s_{max}}, \epsilon_{d_{max}} \in \Theta(\hat{\epsilon})$. From (B5) and (B6), we have that

$$N_d \in \text{Poly} \left(\frac{1}{\hat{\epsilon}} \right) \quad (\text{B9})$$

suffices to achieve this. Finally, we recall that N_d is the number of intervals in the discretised distribution, and thus letting $n = \log_2 N$ we have that $\mathcal{O}(n)$ bits are required. All of the operations required in classical Monte-Carlo, such as summing up samples, are polynomial in the number of bits (noting that the complexity of generating a single sample, (B6), incorporates operations on the number of bits required to digitally express a sample to the required accuracy) and so the computational overhead is

$$\text{Poly}(n) = \text{Poly} \log N = \text{Poly} \log \left(\frac{1}{\hat{\epsilon}} \right) \quad (\text{B10})$$

Thus, when we include the sample complexity, we get an overall computational complexity

$$\mathcal{O} \left(\left(\frac{1}{\hat{\epsilon}} \right)^2 \text{Poly} \log \left(\frac{1}{\hat{\epsilon}} \right) \right) = \tilde{\mathcal{O}} \left(\left(\frac{1}{\hat{\epsilon}} \right)^2 \right) \quad (\text{B11})$$