



**HAL**  
open science

# Quantum Distance Calculation for $\epsilon$ -Graph Construction

Naomi Mona Chmielewski, Nina Amini, Paulin Jacquot, Joseph Mikael

► **To cite this version:**

Naomi Mona Chmielewski, Nina Amini, Paulin Jacquot, Joseph Mikael. Quantum Distance Calculation for  $\epsilon$ -Graph Construction. IEEE Quantum Week 2023, Sep 2023, Washington, United States. 10.1109/QCE57702.2023.00010 . hal-04118977

**HAL Id: hal-04118977**

**<https://hal.science/hal-04118977>**

Submitted on 6 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Quantum Distance Calculation for $\varepsilon$ -Graph Construction

Naomi Mona Chmielewski  
OSIRIS/L2S

EDF Lab/CentraleSupélec, Université Paris-Saclay CNRS, CentraleSupélec, Université Paris-Saclay.  
Palaiseau, France  
naomi.chmielewski@edf.fr

Nina Amini  
L2S

Gif sur Yvette, France  
nina.amini@centralesupelec.fr

Paulin Jacquot  
OSIRIS

EDF Lab  
Palaiseau, France  
paulin.jacquot@edf.fr

Joseph Mikael

OSIRIS

EDF Lab

Palaiseau, France

joseph.mikael@edf.fr

**Abstract**—In machine learning and particularly in topological data analysis,  $\varepsilon$ -graphs are important tools but are generally hard to compute as the distance calculation between  $n$  points takes time  $\mathcal{O}(n^2)$  classically. Recently, quantum approaches for calculating distances between  $n$  quantum states have been proposed, taking advantage of quantum superposition and entanglement. We investigate the potential for quantum advantage in the case of quantum distance calculation for computing  $\varepsilon$ -graphs. We show that, relying on existing quantum multi-state SWAP test based algorithms, the query complexity for correctly identifying (with a given probability) that two points are not  $\varepsilon$ -neighbours is at least  $\mathcal{O}(n^3/\ln n)$ , showing that this approach, if used directly for  $\varepsilon$ -graph construction, does not bring a computational advantage when compared to a classical approach.

**Index Terms**—quantum algorithms, topological data analysis, epsilon graphs, swap test

## I. INTRODUCTION

Nearest-neighbour or similarity graphs are an important tool in machine learning. They are used in collaborative filtering for recommendation systems [1], clustering [2] and pattern recognition [3].  $\varepsilon$ -nearest neighbour graphs in particular can be used in the construction of the Vietoris-Rips complex, an important step in Topological Data Analysis (TDA) [4], which has been shown to have potential quantum advantages in certain cases [5], [6].

In order to construct an  $\varepsilon$ -graph from a point cloud with  $n$  points, we need a way to calculate euclidean distances between points. Classically, the distance calculation for the construction of an  $\varepsilon$ -graph can be done in time  $\mathcal{O}(n \log n)$  on average, using the *kd-tree* algorithm [7] which only scans relevant areas in the embedding space. Constructing the full distance matrix between all  $n$  inputs takes time  $\mathcal{O}(n^2)$ .

This work is supported by the Agence Nationale de la Recherche projects Q-COAST ANR-19-CE48-0003, QUACO ANR-17-CE40-0007, and IGNITION ANR-21-CE47-0015, by the project NEASQC that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951821 and by EDF Lab's OSIRIS department which is gratefully acknowledged.

Quantumly, distance calculation between two quantum states is commonly done using the SWAP-test [8], requiring one CSWAP gate and one ancillary qubit. In [9], the authors propose a method to calculate the distances between all  $n$  input states using only  $\mathcal{O}(n)$  CSWAP gates and  $\mathcal{O}(\log n)$  ancillaries, suggesting an interesting alternative to the classical methods.

The authors in [10] propose a modification of the multi-state SWAP test from [9], potentially reducing the number of circuit repetitions by combining multiple SWAP tests in parallel, while increasing the number of CSWAP gates in the circuit to  $\mathcal{O}(n \log n)$ .

In [11], the authors suggest a quantum algorithm for finding the nearest neighbour of a query input state by combining amplitude estimation and the Dürr-Hoyer minimisation algorithm [12]. The algorithm in [11] can be used to find the  $k$  nearest neighbours,  $k \in \mathbb{N}$  by repeating the process  $k$  times while removing all previously found nearest neighbours from the data. This approach does not easily extend to  $\varepsilon$ -nearest neighbour identification.

In this paper, we show that the number of oracle calls necessary to construct an  $\varepsilon$ -graph using the quantum algorithm proposed in [9] to  $(1 - \gamma)$  correctly identify that two points are not  $\varepsilon$ -neighbours is at least  $\mathcal{O}(n^3/\ln n)$ , where  $n$  is the number of input states.

This paper is organised as follows: in section II, we recall the definition of an  $\varepsilon$ -graph for our purposes as well as the standard SWAP test and summarise the algorithm developed in [9]. We also propose a simple extension of the algorithm to  $d$ -dimensional inputs. We give our main results for the query and gate complexity of the algorithm proposed in [9] in section III by establishing sharp lower bounds for the number of oracle calls necessary to create  $\varepsilon$ -graphs. Section IV provides a discussion on the presented results.

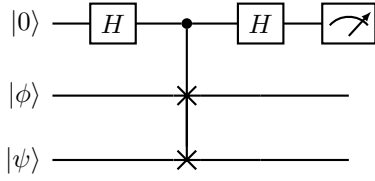


Fig. 1: Quantum circuit representing the standard SWAP test

## II. DISTANCE CALCULATION IN QUANTUM PARALLEL

### A. Classical Distance Calculation

*Definition 2.1 ( $\varepsilon$ -graph):* Given a finite set of  $d$ -dimensional points  $S \subseteq \mathbb{R}^d$  of size  $|S| = n$  and scale  $\varepsilon > 0$ , the  $\varepsilon$ -graph is an undirected graph  $G_\varepsilon = (V, E_\varepsilon)$  where  $V = S$  and

$$E_\varepsilon = \{\{u, v\} \mid \delta(u, v) < \varepsilon, u \neq v \in S\}$$

where  $\delta$  is the euclidean metric.

Thus, to construct an  $\varepsilon$ -graph from a given set  $S$ , for each  $u \in S$  we must find all points  $v_j \in S$  that are at a distance  $d(u, v_j) < \varepsilon$ . One way to do this is to use a *brute-force* algorithm that calculates all  $n(n-1)/2$  distances between the  $n$  points and selects pairs that are within  $\varepsilon$ -distance. This algorithm takes  $\mathcal{O}(n^2)$  time and is linear in the dimensionality  $d$ .

For large  $n$ , there are more efficient algorithms such as the *kd-tree* which takes time  $\mathcal{O}(n \log n)$  on average by sorting the data into a tree and searching only subsections of that tree [7]. This algorithm thus does not generally calculate all  $n(n-1)/2$  possible distances. Note that the dimensionality dependence of the kd-tree algorithm is at least  $\mathcal{O}(2^d)$  due to the curse of dimensionality, making the algorithm less suitable for high-dimensional data [13].

As the number  $n$  of input points tends to be extremely high, it would be favourable to reduce the time dependence in  $n$  while also reducing the dependence in  $d$ . A naive approach to try and reduce the complexity could be to use quantum algorithms that estimate distances in parallel. As the number of CSWAP gates in the algorithm proposed in [9] is  $\mathcal{O}(n)$ , and a simple extension to  $d$ -dimensional input states only leads to a linear increase in the number of CSWAP gates, this might be achieved if the query complexity is sufficiently low. In our analysis, we find that the query complexity is too high to achieve any improvement over classical exact algorithms.

### B. SWAP-Test

The common way to calculate the distance between two quantum states  $|\phi\rangle, |\psi\rangle$  of dimension  $d = 1$  is done via the SWAP test [8] (see Figure 1).

The resulting state is then

$$\frac{1}{2} |0\rangle (|\phi, \psi\rangle + |\psi, \phi\rangle) + \frac{1}{2} |1\rangle (|\phi, \psi\rangle - |\psi, \phi\rangle)$$

so that at the end, the probability of measuring 0 is

$$\begin{aligned} & \frac{1}{2} (\langle \phi | \langle \psi | + \langle \psi | \langle \phi |) \frac{1}{2} (|\phi\rangle |\psi\rangle + |\psi\rangle |\phi\rangle) \\ &= \frac{1}{2} + \frac{1}{2} |\langle \psi | \phi \rangle|^2. \end{aligned}$$

Thus,  $\mathbb{P}(0) = 1$  if and only if  $|\psi\rangle$  and  $|\phi\rangle$  are parallel, and  $\mathbb{P}(0) = 1/2$  if and only if  $|\psi\rangle$  and  $|\phi\rangle$  are orthogonal. Note in particular that  $\mathbb{P}(0) \equiv p \in [1/2, 1]$ . The way to estimate the distance is then to repeat the circuit  $N$  times and to take  $\hat{p} = 1 - \frac{1}{N} \sum_{i=1}^N X_i$  as the estimate for the probability, where  $X_i$  is a Bernoulli random variable with parameter  $q = 1 - p$ , representing the  $i$ th measurement outcome of the circuit. Finally, the distance is estimated by noting that for normalised vectors, we have

$$|\phi - \psi| = \sqrt{2(1 - |\langle \phi | \psi \rangle|)} = \sqrt{2(1 - \sqrt{2p - 1})}.$$

In the case where  $d > 1$ , write  $|\phi\rangle = |\phi_1\rangle |\phi_2\rangle \cdots |\phi_d\rangle$  and  $|\psi\rangle = |\psi_1\rangle |\psi_2\rangle \cdots |\psi_d\rangle$ . Then, one CSWAP gate is applied to each pair  $|\phi_i\rangle, |\psi_i\rangle$  while the rest of circuit remains the same. Unless otherwise specified, we assume  $d = 1$  for the remainder of the paper, as the explicit dependence in  $d$  is linear and all results are easily extended to  $d > 1$ .

### C. Naive Extension to $n$ States

A simple way to extend the SWAP test to  $n$  input states is to create one SWAP test circuit for each of the  $n(n-1)/2$  possible distances. This circuit requires  $n(n-1)/2$  CSWAP gates. Since the ancilla qubit is measured at the end of each circuit run, we can reuse the same qubit for the ancillary state. We show in section III that the number of necessary circuit repetitions to  $(1 - \gamma)$  correctly identify that two points  $|\phi_i\rangle, |\phi_j\rangle, i, j \in \{1, \dots, n\}$  are not  $\varepsilon$ -neighbours is  $\mathcal{O}(n^2)$  for this circuit design, meaning that the total number of CSWAP gates is  $\mathcal{O}(n^4)$  for any  $\gamma \in (0, 1)$ .

### D. Multi-State SWAP Test

The authors in [9] propose a recursive gate arrangement to calculate all distances between  $n$  quantum states using  $\mathcal{O}(n)$  CSWAP gates and  $\mathcal{O}(\log n)$  ancillary qubits.

The algorithm is based on a unitary  $\mathcal{U}_4$  constructed from 3 CSWAP gates and 3 ancillary qubits and is applied to 4 input states (see Figure 2). The circuit puts all possible pairs between the 4 input states into the first two input registers in superposition.

For  $n = 2^k$  input states, the circuit  $\mathcal{U}_n$  that moves all possible pairs into the first two input registers is constructed recursively from the circuit  $\mathcal{U}_4$ , using  $3n/2 - 3$  CSWAP gates and  $d_n = 3 \log_2(n/2)$  ancillary qubits (see Figure 4).

This circuit can be used for any  $n$  inputs by simply padding the remaining input register with  $|0\rangle$ s.

To finish the multi-state SWAP test, one additional ancillary qubit is added, and a final CSWAP gate and Hadamard gate are implemented, analogous to the two-state SWAP test. As in the two-state SWAP test, the top ancilla qubit is measured, and to extract information on all state-pairs, an additional measurement is done on the  $d_n$  ancilla qubits (see Figure 6).

The quantity to be estimated is then  $p_{0ij}$  for  $i, j \in \{1, \dots, n\}$  where  $p_{0ij}$  designates the probability of measuring the ancillary state  $|0ij\rangle$ , where  $|ij\rangle$  is shorthand for the  $d_n$ -dimensional ancillary basis state associated with the couple

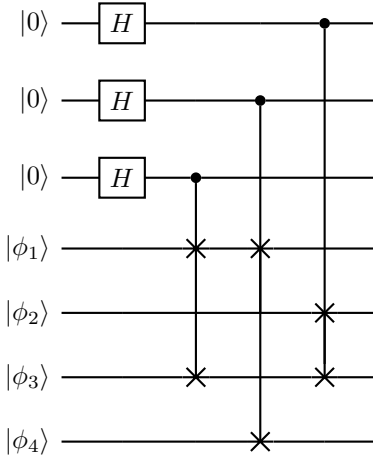


Fig. 2: Quantum circuit representing  $\mathcal{U}_4$  (from [9])

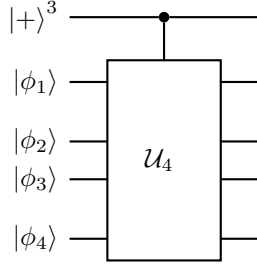


Fig. 3: Compact representation of  $\mathcal{U}_4$

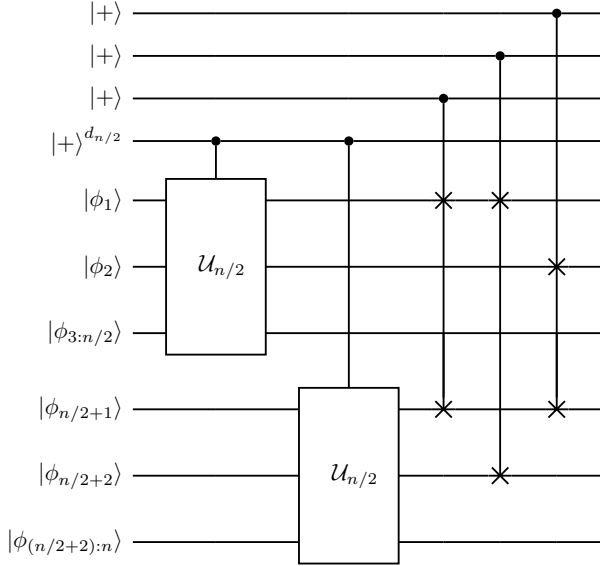


Fig. 4: Quantum circuit representing  $\mathcal{U}_n$  (from [9])

$(|\phi_i\rangle, |\phi_j\rangle)$ . As the authors state in [9, Theorem 2.2], the probability can be expressed as

$$p_{0ij} = \frac{1 + |\langle \phi_i | \phi_j \rangle|^2}{2^{d_n}}$$

where  $d_n$  is the number of ancillary qubits  $d_n = 3 \log(n/2)$ ,

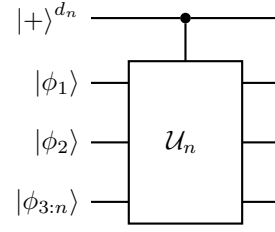


Fig. 5: Compact representation of  $\mathcal{U}_n$

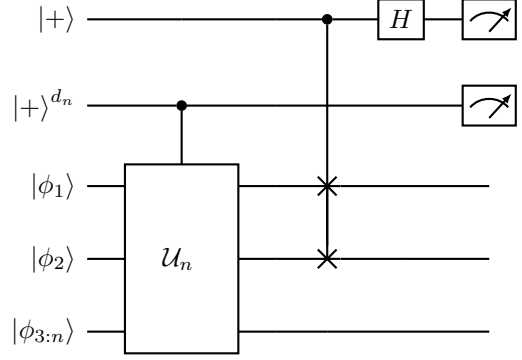


Fig. 6: Final SWAP test and measurement (from [9])

thus

$$p_{0ij} = \frac{2^3(1 + |\langle \phi_i | \phi_j \rangle|^2)}{n^3}. \quad (1)$$

### E. Extension to Multi-Dimensional Inputs

Although the authors only detail their algorithm for one-dimensional inputs, a natural extension to higher dimensional inputs  $d > 1$  can be done analogous to the simple SWAP test, adding one additional CSWAP gate per dimension for each CSWAP gate present in the base circuit  $\mathcal{U}_4$  as well as one additional CSWAP gate per dimension in the final part before the measurement (see Fig 7). This would lead to a linear increase in CSWAP gates, more specifically, the number of gates for  $n$  inputs of dimension  $d$  is  $(3n/2 - 3)d$ , and the number of ancillary states does not change.

## III. QUERY COMPLEXITY OF THE MULTI-STATE SWAP TEST FOR $\epsilon$ -GRAPH CONSTRUCTION

Denote by  $\mathcal{O}_n$  an oracle that generates  $n$  quantum states  $|\phi_1\rangle, \dots, |\phi_n\rangle$ , for example a qRAM [14]. In [9], the authors establish an upper bound on the number of oracle calls necessary to obtain  $\gamma$ -close estimates of all distances, where  $\gamma$ -close refers to the expectation of the difference between the estimates and the true distances being at most  $\gamma$ . The theorem is stated for a larger class of algorithms that includes  $\mathcal{U}_n$ . We will only focus on  $\mathcal{U}_n$  in this paper. Adapting their notations to fit ours, we note  $\mu = \{|\langle \phi_i | \phi_j \rangle|^2\}_{i < j=1}^n$  the ground truth overlaps between all states (instead of  $\delta$ ),  $\hat{\mu} = \{\mu_{ij}\}_{i < j=1}^n$

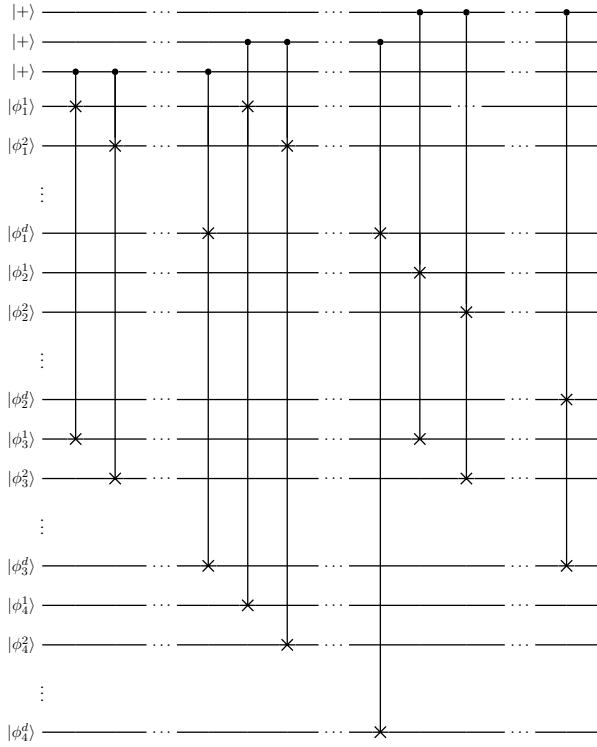


Fig. 7: Quantum circuit representing  $\mathcal{U}_4$  for  $d > 1$

the estimate of the overlaps (instead of  $\hat{\delta}$ ),  $n$  the number of input states (instead of  $m$ ),  $\gamma \in (0, 1)$  the precision parameter (instead of  $\epsilon$ ), and noting that  $2^{2d_n} = n^6/2^6$ , [9, Theorem 2.3] states the following:

*Theorem 1 ([9]):* The circuit  $\mathcal{U}_n$  needs at most  $\mathcal{O}(n^6/\gamma^2)$  calls to the oracle  $\mathcal{O}_n$  to obtain an estimate  $\hat{\mu}$  of  $\mu = \{|\langle \phi_i | \phi_j \rangle|^2\}_{i < j=1}^n$  such that  $\mathbb{E}[\|\hat{\mu} - \mu\|_2] \leq \gamma$ , where  $\|\hat{\mu} - \mu\|_2^2 \equiv \sum_{i < j} (\hat{\mu}_{ij} - \mu_{ij})^2$ .

Note in particular that this implies a number  $\mathcal{O}(n^7/\gamma^2)$  of CSWAP gates.

To calculate the  $\epsilon$ -graph, it is not necessary to calculate all distances exactly. In fact, it suffices to decide whether the distance between any two states is likely smaller than  $\epsilon$ .

Define

$$\delta \equiv \mathbb{1}_{|\phi - \psi| < \epsilon} = \mathbb{1}_{p > \frac{(1 - \epsilon^2/2)^2 + 1}{2}}$$

where  $\epsilon \in \mathbb{R}_+$ . For ease of notation, write

$$\alpha_\epsilon = \frac{(1 - \epsilon^2/2)^2 + 1}{2}.$$

Then,  $X$  is the binomial random variable  $\sim \text{Bin}(N, q)$  defined by  $X = \sum_{i=1}^N X_i$  where  $X_i$  is a Bernoulli random variable  $\sim \text{Ber}(q)$ ,  $q = \mathbb{P}(X_i = 1) = 1 - p$ ,  $p \in [1/2, 1)$ . Denote by  $\hat{p} = 1 - \sum_{i=1}^N X_i/N$  the estimator of  $p$ . A *false negative* is identified when  $\hat{p} \leq \alpha_\epsilon$  when in reality,  $p > \alpha_\epsilon$ . The probability of such a false negative is given by

$$\begin{aligned} \xi_p(N, \alpha_\epsilon) &\equiv \mathbb{P}(\hat{p} \leq \alpha_\epsilon) = \mathbb{P}(X \geq N(1 - \alpha_\epsilon)) \\ &= \sum_{i=\lceil N(1 - \alpha_\epsilon) \rceil}^N \binom{N}{i} (1-p)^i p^{N-i}. \end{aligned}$$

Our main result is stated in the following proposition:

*Proposition 1:* For any  $\gamma \in (0, 1)$ , the number of circuit repetitions necessary to  $(1 - \gamma)$ -correctly identify that two points are not  $\epsilon$ -neighbours is at least  $\mathcal{O}\left(\frac{n^3}{\ln n}\right)$ . In particular, this implies a total number of CSWAP gates of at least  $\mathcal{O}\left(\frac{n^4}{\ln n}\right)$ .

To prove this proposition, we will be using the following lemma and corollary. First let us define

$$N(\gamma) = \frac{\ln(1/\gamma)}{(1 - \alpha_\epsilon) \ln\left(\frac{1 - \alpha_\epsilon}{1 - p}\right) + \alpha_\epsilon \ln\left(\frac{\alpha_\epsilon}{p}\right)}. \quad (2)$$

*Lemma 1:* There exists a  $\tilde{\gamma} \in (0, 1)$  as well as tuples  $(\tilde{\alpha}_\epsilon, \tilde{p})$  such that for  $\tilde{N} \equiv N(\tilde{\gamma})$ ,

$$\xi_{\tilde{p}}(\tilde{N}, \tilde{\alpha}_\epsilon) = \tilde{\gamma}.$$

*Proof:* The Chernoff-Hoeffding inequality [15, Lemma 4.7.2] states that for  $0 < q < \lambda < 1$ ,  $\mathbb{P}(X \geq N\lambda) \leq \exp\{-N\text{KL}(\lambda||q)\}$  where  $\text{KL}(\lambda||q) = \lambda \ln\left(\frac{\lambda}{q}\right) + (1 - \lambda) \ln\left(\frac{1 - \lambda}{1 - q}\right)$  is the Kullback-Leibler divergence between two Bernoulli variables with parameters  $\lambda$  and  $q$ . Replacing  $\lambda$  with  $1 - \alpha_\epsilon$  and noting that  $q = 1 - p$ , we find that for  $0 < \alpha_\epsilon < p < 1$ ,

$$\xi_p(N, \alpha_\epsilon) \leq \exp\{-N\text{KL}(\alpha_\epsilon||p)\}. \quad (3)$$

Thus, for any  $\gamma \in (0, 1)$  we have  $\xi_p(N, \alpha_\epsilon) \leq \gamma$  for any  $N \geq N(\gamma)$ . Note that the KL divergence is non-negative and for our case non-zero (as  $\alpha_\epsilon < p$ ), so  $N(\gamma)$  is well defined.

We have the lower bound [15, Lemma 4.7.1 and 4.7.2]

$$\xi_p(N, \alpha_\epsilon) \geq \frac{1}{\sqrt{2N}} \exp\{-N\text{KL}(\alpha_\epsilon||p)\} \quad (4)$$

for any  $N \in \mathbb{N}^*$ ,  $\alpha_\epsilon, p$ . Plugging in the expression for  $N(\gamma)$  from (2), we find that

$$\begin{aligned} \xi_p(N(\gamma), \alpha_\epsilon) &\geq \frac{1}{\sqrt{2 \frac{\ln(1/\gamma)}{\text{KL}(\alpha_\epsilon||p)}}} \exp\left\{\frac{\ln(\gamma)}{\text{KL}(\alpha_\epsilon||p)} \text{KL}(\alpha_\epsilon||p)\right\} \\ &= \frac{\gamma}{\sqrt{2 \frac{\ln(1/\gamma)}{\text{KL}(\alpha_\epsilon||p)}}}. \end{aligned} \quad (5)$$

Setting the rightmost term in equation (5) equal to  $\gamma$  will give us conditions on the values of  $\text{KL}(\alpha_\epsilon||p)$  for which the bound is sharp. We find that for

$$\text{KL}(\alpha_\epsilon||p) = 2 \ln(1/\gamma)$$

$\xi_p(N(\gamma), \alpha_\varepsilon) = \gamma$ . Note that for  $\alpha_\varepsilon$  small and  $p$  large, the KL-divergence explodes, so that for a large enough difference between  $\alpha_\varepsilon$  and  $p$ , the equality holds for any  $\gamma$  which concludes the proof. If we restrict the analysis to more common cases where  $p$  is reasonably <sup>1</sup> far from 1, we find that for example for the tuple  $(\tilde{\alpha}_\varepsilon = 0.5, \tilde{p} = 0.9)$ , the KL-divergence is 0.51, meaning that for  $\tilde{\gamma} \approx 0.78$ , and  $(\tilde{\alpha}_\varepsilon, \tilde{p})$  as above,  $\xi_{\tilde{p}}(N(\tilde{\gamma}), \tilde{\alpha}_\varepsilon) = \tilde{\gamma}$ .

*Corollary 1:* There exist  $\tilde{\gamma} \in (0, 1)$  as well as tuples  $(\tilde{\alpha}_\varepsilon, \tilde{p})$  such that for any  $N \leq N(\tilde{\gamma}) \equiv \tilde{N}$  and for any  $\gamma \in (0, 1)$  we can find reasonable  $(\alpha_\varepsilon, p)$  such that  $\xi_p(N, \alpha_\varepsilon) \equiv \mathbb{P}(X \geq N(1 - \alpha_\varepsilon)) \geq \gamma$ .

In other words, there exists a minimal number  $\tilde{N}$  of circuit repetitions such that for any precision parameter  $\gamma$ , there are couples  $(\alpha_\varepsilon, p)$  such that the probability of identifying a false negative is greater than  $\gamma$  for any number of circuit repetitions less than  $\tilde{N}$ .

*Proof:* This follows directly from Lemma 1 by noting that for all  $\gamma \geq \tilde{\gamma}$  as found in Lemma 1, reasonable tuples  $(\alpha_\varepsilon, p)$  exist that fulfill the equality  $\xi_p(N(\gamma), \alpha_\varepsilon) = \gamma$ , and for  $\gamma \leq \tilde{\gamma}$ , the number of repetitions to attain  $\gamma$  error must be at least  $\tilde{N}$ .

*Proof of Proposition 1:* From Corollary 1 we know that for any  $\gamma \in (0, 1)$ , the number of circuit repetitions necessary to  $(1 - \gamma)$ -correctly identify that two points are not  $\varepsilon$ -neighbours is at least  $\tilde{N} = N(\tilde{\gamma})$ . Note that at the end of circuit  $\mathcal{U}_n$  the measurement outcome of the  $d_n$  ancillary states is going to be one of the basis states of the ancilla register. Each basis state is associated with one pair of the input states, and measuring 0 in the top most ancillary qubit is equivalent to measuring 0 in the classic SWAP test. Adopting the notation from [9], we denote by  $p_{0ij}$  the probability of measuring the ancillary state  $|0ij\rangle$  as in (1). Note that  $p_{0ij} \in [2^3/n^3, 2^4/n^3]$ ,  $n \in \{2^k, k \geq 2\}$ . We can then model one circuit run as a Bernoulli variable which returns the measurement outcome associated with the couple  $(|\phi_i\rangle, |\phi_j\rangle)$  with probability  $p_{0ij}$ , and any other outcome with probability  $1 - p_{0ij}$ . We can thus replace  $p$  from the previous sections with  $p_{0ij}$ .

Write

$$\delta \equiv \mathbb{1}_{|\phi - \psi| \leq \varepsilon} = \mathbb{1}_{p_{0ij} \geq \frac{[(1 - \varepsilon^2/2)^2 + 1]2^3}{n^3}}$$

where  $\varepsilon \in \mathbb{R}_+$ . Then,

$$\alpha_\varepsilon = \frac{[(1 - \varepsilon^2/2)^2 + 1]2^3}{n^3}$$

and for a false negative to be less than  $\gamma$ -probable, we need

$$N \geq \tilde{N} = \mathcal{O}\left(\frac{n^3 \ln(1/\tilde{\gamma})}{\ln n}\right)$$

circuit repetitions to  $(1 - \gamma)$  correctly identify that states  $(|\phi_i\rangle, |\phi_j\rangle)$  are not  $\varepsilon$ -neighbours. Since this is a lower bound

<sup>1</sup>By "reasonable" we mean values of  $\alpha_\varepsilon$  and  $p$  that are likely to occur in real experiments.

for only one particular couple  $(|\phi_i\rangle, |\phi_j\rangle)$ , and each circuit run only returns a measurement associated with exactly one couple, the total number of circuit runs to  $(1 - \gamma)$ -correctly identify all non- $\varepsilon$ -neighbours is at least  $\mathcal{O}\left(\frac{n^3}{\ln n}\right)$ .

*Remark 1:* Note that by combining the lower and upper bound from (4) and (3), it becomes evident that the upper bound in (3) is sharp asymptotically as  $N \rightarrow \infty$ .

*Remark 2:* From the above calculations, we see that the query complexity for all possible pairs using the standard SWAP test is at least  $\mathcal{O}(n^2 \ln(1/\tilde{\gamma}))$ . Since the naive multi-state SWAP test proposed in section II-C consists of running  $\mathcal{O}(n^2)$  independent circuits (one for each possible distance), this implies a total number of CSWAP gates of  $\mathcal{O}(n^4)$ . The number of ancillary qubits is  $\mathcal{O}(1)$ , as opposed to  $\mathcal{O}(\log n)$  for circuit  $\mathcal{U}_n$ .

*Remark 3:* The results of section III are not directly comparable to the classical methods mentioned in section II-A. While the brute-force algorithm exactly calculates all distances, and the kd-tree algorithm exactly calculates all relevant distances, the quantum method analysed in section III is only concerned with approximately identifying  $\varepsilon$ -neighbours.

#### IV. DISCUSSION

We have analysed the query and gate complexity of the multi-state SWAP test proposed in [9] for the purposes of creating an  $\varepsilon$ -graph, which are  $\mathcal{O}\left(\frac{n^3}{\ln n}\right)$  and  $\mathcal{O}\left(\frac{n^4}{\ln n}\right)$  respectively. Comparing these complexities directly to the time complexities of more common classical algorithms such as brute-force and kd-tree does not immediately yield a quantum advantage. This result makes the particular circuit  $\mathcal{U}_n$  less suitable than other classical algorithms for classical data for the construction of an  $\varepsilon$ -graph, however the circuit might still be adapted for quantum data.

A possible research avenue would be the search for better performing quantum distance algorithms. Although the algorithm from [9] might not be adapted for  $\varepsilon$ -graph construction, it could still be useful for other applications, in particular if the application calls for the preparation of all possible quantum state pairs in superposition in only two registers.

#### REFERENCES

- [1] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*. Springer New York, NY, second ed., 2016.
- [2] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 1996.
- [3] M. N. Murty and V. S. Devi, *Undergraduate Topics in Computer Science - Pattern Recognition, An Algorithmic Approach*. Springer University Press, 2011.
- [4] A. Zomorodian, "Fast construction of the Vietoris-rips complex," *Computers & Graphics*, no. 34, pp. 263–271, 2010.
- [5] I. Y. Akhalwaya, S. Ubaru, K. L. Clarkson, M. S. Squillante, V. Jejjala, Y.-H. He, K. Naidoo, V. Kalantzis, and L. Horesh, "Towards quantum advantage on noisy quantum computers," *arXiv*, 2022. unpublished.
- [6] S. Lloyd, S. Garnerone, and P. Zanardi, "Quantum algorithms for topological and geometric analysis of data," *Nature Communications*, vol. 7, no. 1, 2016.

- [7] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software*, vol. 3, no. 3, 1977.
- [8] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, "Quantum fingerprinting," *Physical Review Letters*, vol. 87, no. 16, 2001.
- [9] X. Gitiiaux, I. Morris, M. Emelianenko, and M. Tian, "Swap test for an arbitrary number of quantum states," *Quantum Information Processing*, vol. 21, no. 10, 2022.
- [10] W. Liu, H.-W. Yin, Z.-R. Wang, and W.-Q. Fan, "Multi-state swap test algorithm," *arXiv*, 2022. unpublished.
- [11] N. Wiebe, A. Kapoor, and K. Svore, "Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning," *Quantum Information and Computation*, vol. 15, no. 3&4, 2015.
- [12] C. Dürr and P. Hoyer, "A quantum algorithm for finding the minimum," *arXiv*, 1996. unpublished.
- [13] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," *Journal of the ACM*, vol. 45, no. 6, 1998.
- [14] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum random access memory," *Physical Review Letters*, vol. 100, no. 16, 2008.
- [15] R. B. Ash, *Information Theory*. Dover Publications, Inc., 1990. ISBN 0-486-66521-6.