A Scalable Quantum Gate-Based Implementation for Causal Hypothesis Testing

Kundu, Akash; Acharya, Tamal; Sarkar, Aritra

DOI
10.1002/qute.202300326

Publication date
2024

Document Version
Final published version

Published in
Advanced Quantum Technologies

ADVANCED QUANTUM TECHNOLOGIES
www.advquantumtech.com

A Scalable Quantum Gate-Based Implementation for Causal Hypothesis Testing

Akash Kundu, Tamal Acharya, and Aritra Sarkar*

In this work, a scalable quantum gate-based algorithm for accelerating causal inference is introduced. Specifically, the formalism of causal hypothesis testing presented in [*Nat Commun* 10, 1472 (2019)] is considered. Through the algorithm, the existing definition of error probability is generalized, which is a metric to distinguish between two competing causal hypotheses, to a practical scenario. The results on the `Qiskit` validate the predicted speedup and show that in the realistic scenario, the error probability depends on the distance between the competing hypotheses. To achieve this, the causal hypotheses are embedded as a circuit construction of the oracle. Furthermore, by assessing the complexity involved in implementing the algorithm's subcomponents, a numerical estimation of the resources required for the algorithm is offered. Finally, applications of this framework for causal inference use cases in bioinformatics and artificial general intelligence are discussed.

1. Introduction

Despite the huge success of machine learning (ML) algorithms based on deep neural networks, these systems are inscrutable black-box models. This hampers users' trust in the system and obfuscates the discovery of algorithmic biases that stem from flawed generative processes, which can be prejudicial to certain inputs (e.g., racial discrimination). Explainable artificial intelligence (XAI)[1] focuses on human understanding of the decision from the learned solution as white-box models. These models provide results that are understandable for domain experts, thus providing transparency, interpretability, and explainability. XAI algorithms provide a basis for justifying decisions, tracking and thereby verifying them, improving the algorithms, and exploring new facts. There has been relatively slow progress in XAI, despite realizing its importance as we increasingly automate critical systems. Early advances in XAI were based on symbolic reasoning systems and truth maintenance systems. To achieve causal reasoning,[2] rule-based learning and logic-based inference systems were proposed. Methods to address inherent opaque modern methods like deep learning-based neural networks and genetic algorithms include layer-wise relevance propagation and local interpretability. Other ML algorithms (e.g., decision trees, Bayesian classifiers, additive models) exist that generate interpretable models, allowing direct inspection of components such as feature weights, decision tree paths, or specific rules to understand the predictions. However, these models are not general and scalable to compete with the adoption and impact of neural networks. On the other hand, symbolic reasoning systems were abandoned owing to the difficulty in scaling these systems for a large number of parameters.

The capability of quantum computation allows us to scale symbolic reasoning models by encoding the classical rules as a superposition of quantum states or processes[3] is a core motivation in quantum explainable artificial intelligence. Quantum mechanics provide enhanced ways to identify causal links; for example, certain quantum correlations can be used to infer classical causal relationships.[4,5] This could overcome the apprehension of existing classical approaches being pursued in XAI.

In this article, we will explore how we can distinguish quantum processes by their causal structure. Specifically, we study the construction proposed in ref. [6] toward a quantum circuit implementation on the `Qiskit` quantum programming language. In doing so, we uncover i) the implementation aspects of the causal oracle, ii) the gate and qubit complexity of the full algorithm, and iii) practical case error probability, which we introduce in the paper that shows the dependence of the error probability on some distance measure between the set of processes/hypotheses being tested. While the current technology readiness level of quantum systems prevents us from demonstrating this causal reasoning within a broader application framework, we present the quantum kernel that can be readily embedded within a software pipeline for applications in bioinformatics and artificial general intelligence (AGI). In particular, it will be useful in XAI pipelines.[7,8]

A. Kundu
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
Joint Doctoral School
Silesian University of Technology
Gliwice, Poland
E-mail: akundu@iitis.pl

T. Acharya, A. Sarkar
Quantum Intelligence Research Team
Department of Quantum & Computer Engineering
Delft University of Technology
Delft 2628 CD, The Netherlands

A. Sarkar
Quantum Machine Learning Group, Quantum Computing Division
QuTech
Delft 2628 CJ, The Netherlands

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/qute.202300326

DOI: 10.1002/qute.202300326

The remaining article is organized as follows. In Section (2), we briefly review the specifics of causal reasoning and some of the well-studied techniques. A discussion on the basic concepts and quantum advantage of causal hypothesis testing is given in Section (3). In the following Section (4), we describe the problem formulation; containing the main findings of the article. Here, we define a model implementation on `Qiskit` followed by a correction factor introduced in the error probability based on our empirical results; which we call practical error probability. Finally, in Section (6), we discuss some potential use cases in bioinformatics and artificial general intelligence. The corresponding quantum resources of gates and qubits are assessed for realistic cases. Section (7) concludes the article.

2. Overview of Causal Inference

Causal inference involves examining the assumptions, study designs, and estimation strategies that enable researchers to conclude about the cause-effect relationships between data. In particular, it considers the outcomes that could manifest given exposure to each of a set of dynamics of a specific causal variable. Causal effects are defined as comparisons between these potential outcomes. Standard approaches in statistics, such as regression analysis, are concerned with quantifying how changes between two variables are associated, with no directional sense. In contrast to that, Causal inference methods determine whether changes in one variable, X, cause changes in another variable, Y, or vice versa. If X is causally related to Y, then Y's change can (at least partially) be explained in terms of X's change.

2.1. Challenges of Performing Causal Inference

Causal models are based on the idea of potential outcomes. The two major challenges in causal inference are:

a) Causation does not imply association: For example, we want to compare the impact of an academic degree on the income of a middle-aged individual. The person might have attended the academic degree or might not have. To calculate the causal effect of having an academic degree, we need to compare the output in both situations, which is not possible. This dilemma is a fundamental problem of causal inference. The challenge in causal inference is that all potential outcomes are not observed, we only observe one. Another example is described by the situation of – a black cat ran under the fence and I tripped and fell over. We could have tripped anyway. It had nothing to do with the cause of the cat running under the fence, i.e., associated events do not imply causal connection.
b) Correlation does not prove causality: Causal inference is the process of concluding a causal connection based on the occurrence of an effect. Causal inference is usually a missing data problem[9] and we tend to make assumptions to make up for the missing causes/variables. An example is a correlation between people eating ice cream and people drownings. It could indicate that eating ice cream affects drowning. The actual correlation is between the season (summer) and these otherwise unrelated things. In this case, the missing cause is the season.

2.2. Classical Techniques in Causal Inference

Causal inference involves studying systems where the measurement of one variable is believed to influence the measurement of another. This process adheres to the scientific method. The initial step is to formulate a falsifiable null hypothesis, which is then tested using statistical methods. Frequentist statistical inference assesses the probability that the observed data occurred by chance under the null hypothesis. In contrast, Bayesian inference evaluates the effect of an independent variable.

Common frameworks for causal inference include the causal pie model (component-cause),[10] Pearl's structural causal model (causal diagrams and do-calculus),[11] structural equation modelling, and Rubin's causal model (potential outcomes).[12,13] presents a more comprehensive survey of classical causal inference techniques.

The most commonly used causal models can be categorized into two types: causal Bayesian networks and structural equation models, which are closely related. The causal graph models integrate mathematics and philosophy. The mathematical components involve directed acyclic graphs (DAGs) and probability theory, particularly conditional independence; the philosophical components involve assumptions about the relationship between causation and probability.[14] An alternative approach to causal inference based on algorithmic generative models is currently gaining popularity.[15] describes the process of performing causal deconvolution using this technique. This paper talks about the different generating mechanisms by which complex data is produced. The authors introduced a universal, unsupervised, and parameter-free model-oriented approach based upon algorithmic probability that decomposes an observation into its most likely algorithmic generative sources. This is closely related to the quantum approaches discussed in the next section.

2.3. Quantum Computation and Algorithmic Generative Models

The synergy between quantum computation and algorithmic information has been studied extensively in ref. [16]. Two main directions were explored, that can be applied for causal inference.

In ref. [3] a global/objective view is presented, which involves quantum automata for algorithmic information. We develop a framework for causal inference grounded in algorithmic generative models. This technique of quantum-accelerated experimental algorithmic information theory (QEAIT) can be ubiquitously applied to diverse domains, to mention a few, genome analysis, the problem of identifying bit strings capable of self-replication is presented. We introduce a novel quantum circuit design for a quantum parallel universal linear bounded automata (QPULBA) model, facilitating the execution of classical models/programs in superposition and enabling the exploration of their properties. The automaton prepares the universal distribution as a quantum superposition state which can be queried to estimate the algorithmic properties of the causal model.

In ref. [17] the authors provide a local/subjective view that involves universal reinforcement learning in quantum environments. This theoretical framework can be applied to automated scientific modelling. A universal artificial general intelligence formalism is presented that can model quantum processes. The

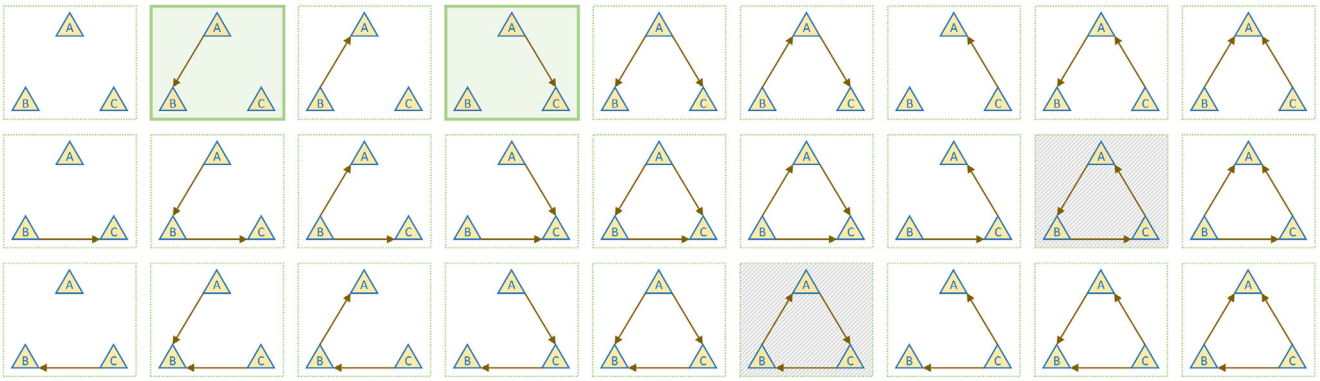


Figure 1. All possible causal relations for three variables. Blocks shaded grey have causal loops (and are typically not considered). The set of blocks shaded green indicates a case of causal hypothesis testing. For the sake of this paper, we consider only the scenarios where (1) *A* causes B and C to be independent, and (2) *A* causes C and B to be independent.

developed quantum knowledge-seeking agent (QKSA) is an evolutionary general reinforcement learning model for recursive self-improvement. It uses resource-bounded algorithmic complexity of quantum process tomography (QPT) algorithms. The cost function determining the optimal strategy is implemented as a mutating gene within a quine. The utility function for an individual agent is based on a selected quantum distance measure between the predicted and perceived environment.

These techniques motivate our research in this article. Unlike quantum/classical data-driven ML, these primitives of QEAIT and QKSA preserve the explanatory power of the model the learning converges to by exploring the space of programs on an automata model. The QPULBA model (in QEAIT) and QPT algorithms (in QKSA) can be generalized to a causal oracle and causal tomography, respectively.

Causal approaches for quantum machine learning (QML),[18,19] and quantum algorithms for causal inference[20,21] are also a related active research direction.

3. Causal Hypothesis Testing

A canonical approach in causal inference is to formulate different hypotheses on the cause–effect relations and test them against each other. This technique is typically used when there is some knowledge of the characteristic of the phenomenon that is being tested. The complete search space of directed graphs between events grows exponentially. An exhaustive example of three variable causal relations is shown in **Figure 1**. In causal hypothesis testing a subset of these graphs is considered as the set of hypotheses being tested against each other. For example, in Figure 1, we can consider, two hypotheses (shaded in green):

a) *A* causes B, C is independent
b) *A* causes C, B is independent.

3.1. Quantum Advantage in Classical Causal Hypothesis Testing

Quantum information enables a richer spectrum of causal relations that is not possible to access via classical statistics. Most research in this direction is toward exploring causality in the quantum context.[22–27] Our focus in this work is specifically using the quantum formulation to provide a computational advantage with respect to a classical technique on classical data.

This problem is studied extensively in ref. [6]. The research analyzes the task of determining the impact of a specific input variable. In the quantum version of the problem, the variables *A*, *B*, and *C* are considered as a quantum system of dimension *d* that in return satisfies either of the two following causal hypotheses Either *B* resulted from *A* via an arbitrary unitary operation, with the state of *C* being maximally mixed, or *C* resulted from *A* through an arbitrary unitary operation, with *B* being maximally mixed. The set of allowed causal relationships say $\mathcal{C}$, between input and output, depends on the physical theory, which determines the implementable mappings achievable through physical processes. In classical physics, cause-effect relationships are usually depicted through conditional probability distributions. However, in quantum theory, they are characterized by quantum channels–completely positive trace-preserving maps that transform density matrices of an input quantum system *A* into density matrices of an output. (either *B* or *C*).

Despite the fact that a cause-effect relationship can be established utilizing any unitary operation, the error probability that is attained remains

$$p_{err} = \frac{1}{2d^N} \tag{1}$$

where *N* represents the number of interventions between one instance and the next.

The error probability in Equation (1) is *d* times smaller than the classical error probability. To achieve this advantage the authors in ref. [6] make use of a universal quantum strategy by preparing a *d* particle singlet state of the form

$$|s_d\rangle = \frac{1}{\sqrt{d!}} \sum_{i_1,i_2,\ldots,i_d} \epsilon_{i_1,i_2,\ldots,i_d} |i_1\rangle, |i_2\rangle, \ldots, |i_d\rangle \tag{2}$$

where $\epsilon_{i_1,i_2,\ldots,i_d}$ is an asymmetric tensor and the sum ranges over all vectors on the computational basis. Then, each of the *d* subsystems is fed as an input to one use of $\mathcal{C}$. By repeating the experiment for *t* times, and By conducting Helstrom's minimum error measurement, one can achieve the error probability given

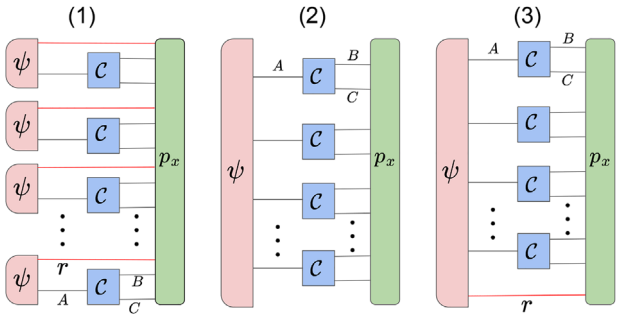


Figure 2. Here, we illustrate the three fundamental ways of constructing parallel strategy. In (1) we initialize the quantum system A along with a reference r in a state ψ. This setting is repeated for N times. Meanwhile in (2) an N size input, ψ, is provided as an input A. Similar to the previous strategy in (3) we have the same N probes of the input entangled using an additional reference r. An unknown process $\mathcal{C}$ (an arbitrary unitary) induces a causal relation between the input and the output in all settings.

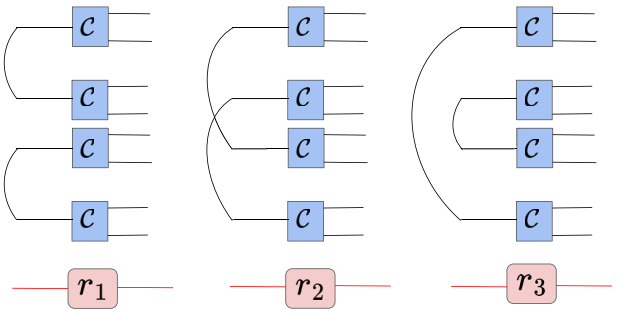


Figure 3. The different ways of partitioning a quantum system of N=4 size into d=2 groups. The three distinct configurations are superposed in the input state ψ and can be accessed by the state of the reference r.

in Equation (1) with $N = t \times d$. This exact strategy is illustrated in **Figure** 2(1).

Apart from the error probability, the discrimination rate, R, is a very crucial performance quantifier for causal hypothesis testing protocols. It can be defined as the rate at which the two causal hypotheses can be differentiated from each other. In Figure 2 for the strategy (1) and (2) the discrimination rate remains $\log d$. But slight engineering of the protocol (2) that adds a reference r (see protocol (3) of Figure 2) helps us to achieve a discrimination rate of $2 \log d$, which is twice as fast as the optimal classical strategy. The primary motive behind the introduction of the reference r lies in the fact that it helps entangle the N input probe states. To say in a more elaborate manner, in the absence of the reference we saw through strategy (2) of Figure 2 that it is optimal to partition the N input into $\frac{N}{d}$ groups and then entangle the probe with each group. Generalizing this line of thought, we consider a mechanism where the partition of the subsystems happens according to a certain configuration i if a control system is in the state $|i\rangle$. Thus, when the control system is in a superposition, the optimal input state is

$$|\psi\rangle = \frac{1}{n}\sum_{i=1}^{n}\left(|s_d\rangle^{\otimes\frac{N}{d}}\right)_i \otimes |i\rangle \tag{3}$$

where each i is labelled depending on the different ways of partitioning N indistinguishable objects into groups of d elements, the number of such ways of partitioning is n, and $|s_d\rangle^{\otimes\frac{N}{d}}$ are orthogonal states of the reference system represented through a product of $\frac{N}{d}$ singlet states. s_d is defined through Equation (2). In **Figure** 3, the different ways of dividing $N = 4$ copies of the causal process probes (C), which yields *three* different ways of entangling bi-partitions of these probes, indexed by encoding on the reference register (through r_i, $i = 1, 2, 3$).

The above observation helps us to conclude that the quantum correlation speeds up the causal hypothesis testing. However, the caveat of the approach is that it is not practically feasible due to the resource scaling of this encoding. The choice of a maximally mixed quantum channel as an alternate hypothesis prevents the step toward the practicality of the approach. In this work, we tackle this issue by easing the previously proposed general completely positive trace preserving scenario in favor of constructable causal oracles. We provide a scalable quantum algorithm that can be implemented in a gate-based quantum computer. This approach is described briefly in the next section.

4. Practical Implementation Considerations

4.1. The Problem Statement

In the general case considered in ref. [6] the intervention (i.e., C) is maximally mixed and the hypotheses are formulated as

a) A causes B i.e., $B \leftarrow UA$ and C is maximally mixed,
b) A causes C i.e., $C \leftarrow UA$ and B is maximally mixed,

where U is an arbitrary unitary operation, the authors achieve the best error probability. It is possible to implement a maximally mixed state by discarding one-half of a maximally entangled input. However, the classical causal hypotheses are given as direct acyclic graphs corresponding to constructable unitaries.

On the contrary, our formulation aims towards the practical feasibility achieved by introducing a quantum algorithm for causal hypothesis testing on the `Qiskit` quantum simulator. To construct the algorithm we consider a modified version of C, where variables A, B, and C are causally connected by the following hypotheses

a) *Hypothesis 0*: A causes B i.e., $B \leftarrow UA$ and C is independent of A,
b) *Hypothesis 1*: A causes C i.e., $C \leftarrow UA$ and B is independent of A.

we choose U as variants of `SWAP` operation and C (or B) being independent of A, which is denoted by an `Identity` ($\mathbb{I}$) operation.

As we have discussed in the previous section the parallel strategy with quantum correlated inputs through a reference, (as shown in Figure 2(3)) gives us a quantum advantage in the detection of cause and effect. Hence, we provide step-wise modification to this optimal structure to finally implement it in a quantum computer and test the *Hypothesis 0* against *Hypothesis 1* to benchmark its performance. In our implementation, we consider the input variables (causes) to be denoted by the set

$C = \{c_0, c_1, \ldots, c_{|C|-1}\}$, while the output variables (effects) are denoted by the set $E = \{e_0, e_1, \ldots, e_{|E|-1}\}$. Furthermore, we consider an equal number of input and output variables to preserve unitarity, i.e., $|C| = |E| = k$. We maintain the simplification of ref. [6], that all variables are of equal length

$$d = |c_i| = |e_j| \tag{4}$$

where $c_i \in C$ and $e_j \in E$. Thus, each variable has 2^d states. Furthermore, we consider that each effect is a permutation of only one cause, i.e., the map from C to E is a bijective function, one-to-one correspondence, or invertible function. To make our implementation more general, we consider a scenario where the access to interventions (i.e., $\mathcal{C}$) on the causes and the effects are unequal.

As a proof-of-concept, we implement a causal hypothesis testing with control over 1 cause $C = \{c_0\}$ and measurement capability over 2 potential effects $E = \{e_0, e_1\}$. These hypotheses are mutually exclusive, where, (1) $e_0 \leftarrow Uc_0$, e_1 is independent of c_0, and (2) $e_1 \leftarrow Uc_0$, e_0 is independent of c_0.

Moreover, we extend the problem statement by modifying the U to a parametric version of `Controlled-SWAP` (`CSWAP`(θ_{orc})) operation. The parameter θ_{orc} defines the strength of the `SWAP` oracle.

Describing all the necessary ingredients for causal hypothesis testing, we now head toward practical implementation.

4.2. The Implementation

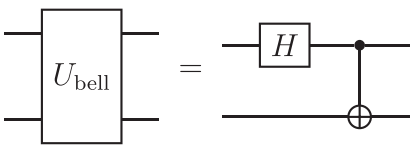
Here, we take the parallel strategy as described in Figure 2(3) and discuss the real-world implementation of its subroutines. To design a quantum circuit, all variables that are potentially inspected as causes need to be assigned a quantum information placeholder. There are a total of

$$k = \max\{|C|, |E|\} \tag{5}$$

placeholders are required for encoding cause of size $|C|$ and effect of size $|E|$. Throughout the paper, we consider, $k = 2$ sets of qubit registers that are used, namely system A and system B.

4.2.1. Preparation of $|\psi\rangle$

The very first component of the causal hypothesis testing strategy is the initial state, $|\psi\rangle$, along with the reference r. We saw in Equation (3) that the state is in a superposition of different partitions of the input and the partition depends on d. Each state in the superposed control input can be accessed by the state of a control system, which is called the reference r. As each partition forms a singlet state of the form Equation (2), we make use of the bell unitary consisting of a Hadamard (`H`) and `CNOT` gate to represent the singlet.



To describe in a simpler manner, the wires in Figure 3 that correlate with different inputs is represented by the the U_{bell}. We illustrate one of the partitions in **Figure 4** where it can be seen that we consider two qubits to represent the reference (i.e., the control system), this is because, for $N = 4$ and $d = 2$, there are three possible partitions and to trigger each configuration we need at most three linear combinations of the control in the form 00, 01, 10. This can be achieved using at least two qubits initialized with $\mathtt{H} \otimes \mathtt{H}$. The full circuit for all the combinations is illustrated through U_{per} in Figure 7.

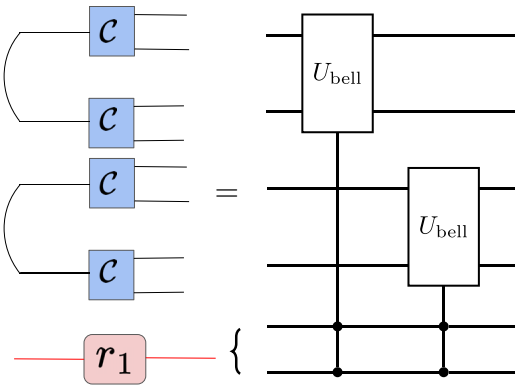


Figure 4. The practical way to implement one of the partitions of the correlated input with the help of the reference r.

4.2.2. Preparation of $\mathcal{C}$

Recalling that $\mathcal{C}$ induces a causal relationship between the input and output and in quantum, it is defined as a quantum channel that maps the input density matrices to an output. To be consistent with the *Hypothesis 0* and *Hypothesis 1*, described in Section 4, we choose $U = \mathbb{I}$ for the *Hypothesis 0* and a $U = \mathtt{SWAP}$ operation for *Hypothesis 0*. The unitary that preserves the essence of the *Hypothesis 1* is given by

$$U_{\text{orc}}^{H_1} = \text{SWAP}(A_i, B_i)^{\otimes N} \otimes \mathbb{I}(A_i)^{\otimes N} \otimes \mathbb{I}(B_i)^{\otimes N} \tag{6}$$

and shown in **Figure 5**.

Now to merge the $\mathbb{I}$ and the `SWAP` *Hypotheses* in s single unitary and to gain better control over the applied hypothesis, we modify Equation (6) by introducing an ancillary qubit initialized with $\mathtt{RX}(\theta_{\mathtt{ctrl}})$. The ancilla qubit works as the control of the SWAP that for $\theta_{\text{ctrl}} = 0$ behaves as $\mathbb{I}$ and for $\theta_{\text{ctrl}} = \pi$ gives complete `SWAP` operation. The illustration of the hypotheses is provided in **Figure 6** and given by

$$U_{\text{orc}}^{H_{0/1}}(\theta_{\text{ctrl}}) = \left[\text{CSWAP}(q_{\text{anc}}, A_i, B_i)^{\otimes N}\right] \times \left[\text{RX}(q_{\text{anc}}, \theta_{\text{ctrl}}) \otimes \mathbb{I}(A_i)^{\otimes N} \otimes \mathbb{I}(B_i)^{\otimes N}\right] \tag{7}$$

4.2.3. Initial Choice of Systems

We have already mentioned that we use $k = 2$ sets of qubit registers to encode the cause and effects through systems A and B. As our work primarily focuses on how different hypotheses affect the error probability p_{err}, we restrict ourselves to the all-zero, i.e.,

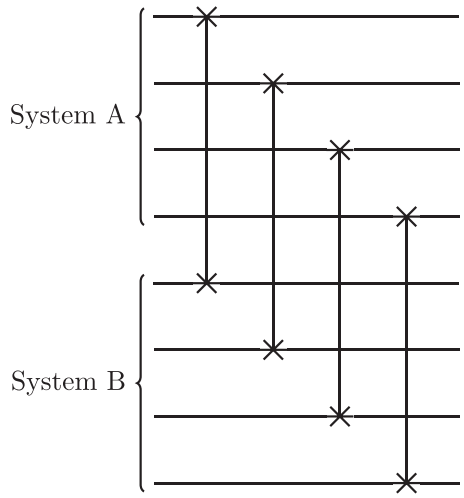


Figure 5. Quantum circuit for the SWAP oracle represented in Equation (6) for system A (the first four qubits) and B (the last four qubits) of size $N = 4$.

$|0\rangle^{\otimes N}$ initialization. But it should be noted that the causes and effects need not be the vacuum states but can be any arbitrary quantum state. To introduce this degree of freedom, we propose two unitaries U_{in}^{A} and U_{in}^{B} that randomly initialize the systems A and B.

4.2.4. Measuring the Outcome

In the original work, Helstorm's minimum error measurement[28] is used to obtain the advantage over the classical causal inference as shown in Equation (1). To calculate the distinguishing probability in computational basis, we add the probabilities of basis states that are unique to one of the qubits of effects (systems A and B).

The overall quantum circuit to perform the causal hypothesis testing is provided in **Figure 7** where (1) U_{in}^{A} and U_{in}^{B} defines the initial choice of system A, B, (2) $U_{\text{per}}^{\text{B}}$ prepares $|\psi\rangle$ (see Equation (3)), (3) The U_{orc} presents the hypothesis in Equation (7) and

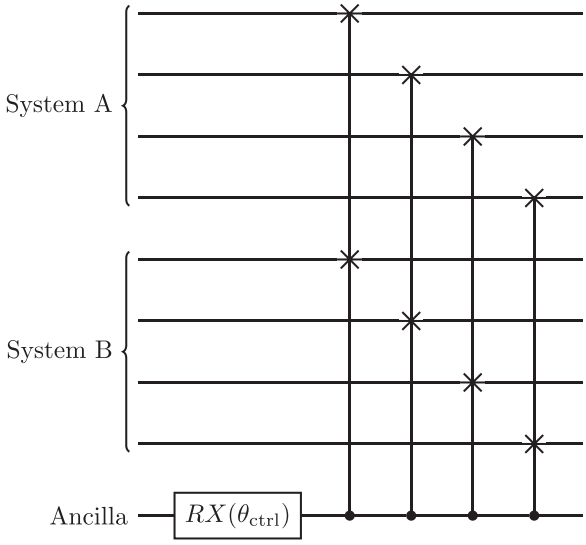


Figure 6. We illustrate the quantum circuit for the Controlled SWAP hypothesis (see Equation (7)) where the strength of the SWAP depends on an ancilla qubit and can be varied through a parameter θ_{ctrl}. The hypothesis illustrated is for system A (the first four qubits) and B (the last four qubits) of size $N = 4$.

finally (3) M_Z^A and M_Z^B are the computational base measurement to get the error probability p_{err}.

4.3. Resource Estimation

In this section, we briefly describe the resource estimation of implementing our algorithm. The number of qubits in the algorithm is

$$n + N_{\text{ref}} \tag{8}$$

where n is the number of qubits to encode the cause and effect and N_{ref} is the number of qubits in the reference state.

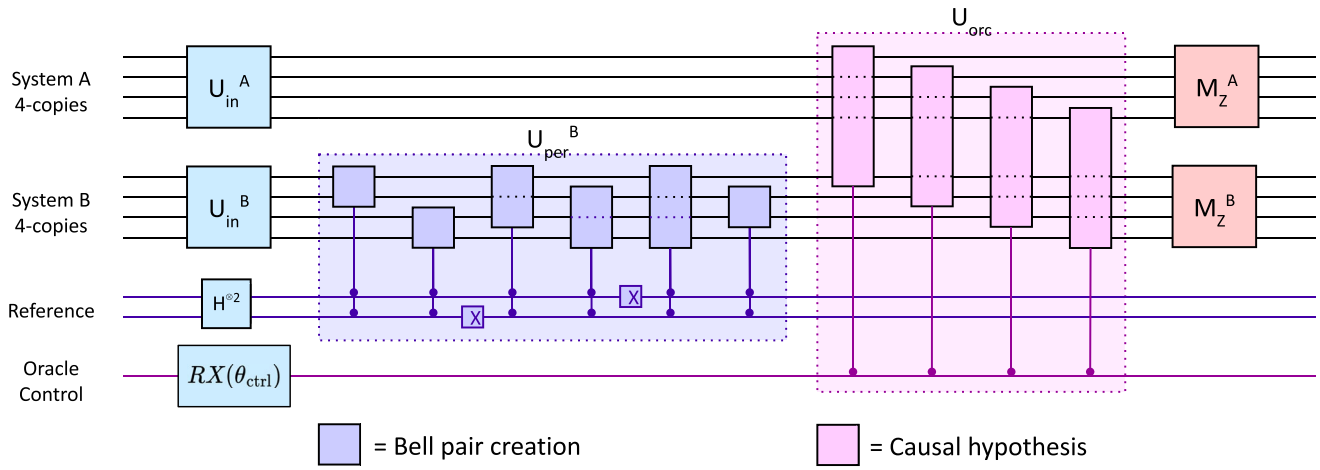


Figure 7. The illustration of complete quantum algorithm for causal hypothesis testing. In the illustration we present the U_{per} and U_{orc} for $N_{\text{A}} = N_{\text{B}} = 4$.

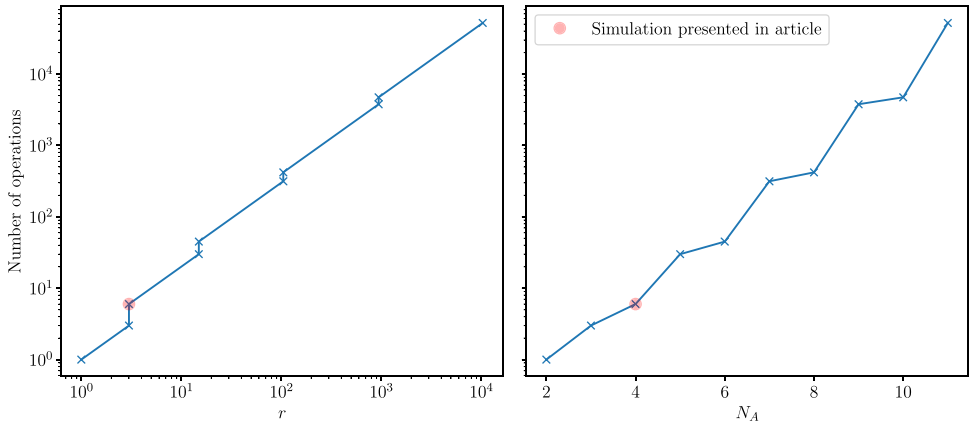


Figure 8. The number of controlled Bell unitary operations required to implement the U_{per} increases exponentially with r and N_A. On the left-hand side, we show the variation with respect to the number of linearly independent states r and on the right-hand side with the dimension of the subsystem N_A (N_B).

4.3.1. Resource Estimation of n

The number of qubits to encode the causes and effects depends on the number of causes $|C|$ (or effects $|E|$), and as C (E) denotes a set of causes (effects) the n also depends on the size of elements in C (E). In our case $|C| = |E| = k$ and each element in C (E) is of the same size i.e. $|c_i \in C| = |e_j \in E| = d$. All in all for N-qubits

$$n = N \times k \times d \tag{9}$$

4.3.2. Resource Estimation of N_{ref}

In the preparation of $|\psi\rangle$, the number of qubits depends on the total possible linearly independent pair combinations. The linearly independent pair r depends on the number of qubits required to encode n as follows

$$r = \frac{n!}{(n/2)! \times 2^{n/2}} \tag{10}$$

from the point calculating the number of qubits in the reference is straightforward. N_{ref} can be obtained as follows

$$N_{\text{ref}} = \lceil \log_2 r \rceil \tag{11}$$

4.3.3. Total Qubit Requirement

Hence, the qubit requirement for our model grows as

$$n + N_{\text{ref}} = n + \lceil \log_2 \left(\frac{n!}{\frac{n}{2}! \times 2^{\frac{n}{2}}} \right) \rceil \tag{12}$$

In the case of our implementation in Figure 7, we choose a system size $N = 4$, the number of cause (effect) $|C|$ ($|E|$) is 1 also the elements in the $|C|$ ($|E|$) of size 1 so the qubits required to encode the causes (effects) can be obtained from Equation (9) and given by $n = 4 \times 1 \times 1 = 4$. Meanwhile, the number of linearly independent pairs is calculated through Equation (10) and is $r = 4!/(2! \times 2^2) = 3$. Finally, the number of qubits in reference is obtained by ceiling r i.e., $N_{\text{ref}} = \lceil \log_2 3 \rceil = 2$.

Recalling that in ref. [6], one of the innovations introduced was to entangle the inputs for the parallel strategy instead of initializing with a tensor product state, which reduces the exponential measurement resource requirement by correlating the basis. However, the cost of implementing the entangled initialization was not analyzed.

Through our model implementation, we are able to show that the operations required to implement the U^B_{per} grow faster than exponentially with the dimension of the subsystem (N_A or N_B) as illustrated in **Figure 8** but linear in respect to the linearly independent states, i.e., r.

4.4. Practical Error Probability

While the number of controllable causes $|C|$ and the encoding length of the causes d depends on the problem formulation, the number of queries, i.e., N and thereby r, is a free parameter. It can be chosen based on the available quantum circuit resources of the number of qubits in the quantum processor, the decoherence time, and gate error probability, such that the pragmatic error remains low. It is shown in ref. [6] that using the correlated input scheme presented in Figure 2(3) we can reach an error probability

$$p_{\text{err}} = \frac{r}{2d^N}\left(1 - \sqrt{1 - r^{-2}}\right) \overset{r>>1}{\longrightarrow} \frac{1}{4rd^N} \tag{13}$$

where r is the number of linearly independent states. Equation(13) gives the limiting case error probability because, in a more general case, the error probability of causal hypothesis testing should be dependent on the two specific hypotheses.

For example, the unitary oracle as given in Equation (7) for the *Hypothesis 1* at $\theta_{\text{ctrl}} = 0$ becomes $U^{H_1}_{\text{orc}}(0) = \mathbb{I} = U^{H_0}_{\text{orc}}(0)$, in that case, the two hypothesis becomes identical and the error probability $p_{\text{err}} = 1$, which can be clearly seen in **Figures 9** and **10**. And as the *Hypothesis1* deviates from the *Hypothesis0*, which can be achieved by fine-tuning θ_{ctrl}, we see the error probability gradually decreases from 1. At $\theta_{\text{ctrl}} = \pi$ when *Hypothesis1* = SWAP the

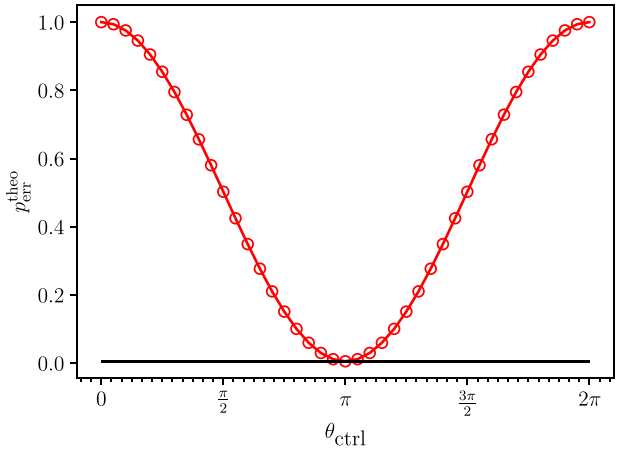


Figure 9. The error probability varies periodically with the parameter, θ_{ctrl} of the `parameterized SWAP` oracle. The black horizontal line is the error probability obtained for the `SWAP` using the definition provided in ref. [6], which is also presented in Equation 13. It can be seen that when $\theta_{\text{ctrl}} = 0$ both the 1st and the 2nd hypothesis are I hence it is the worst case scenario giving an error probability of 1. On the other hand when $\theta_{\text{ctrl}} = \pi$ the 2nd hypothesis mimics `SWAP` operation and coincides with the error probability of ref. [6].

error probability obtains the limit reached by Equation (13) (represented by the black horizontal line in Figure 9). A brief discussion of Figures 9 and 10 is provided in Section 5.

This helps us to conclude that the distance between the oracles plays a crucial role in defining the error probability. Hence, to incorporate this dependence on the relative distinguishability of the hypotheses, we introduce a correction factor proportional to the process distance (Δ) between the two oracles. This arises a

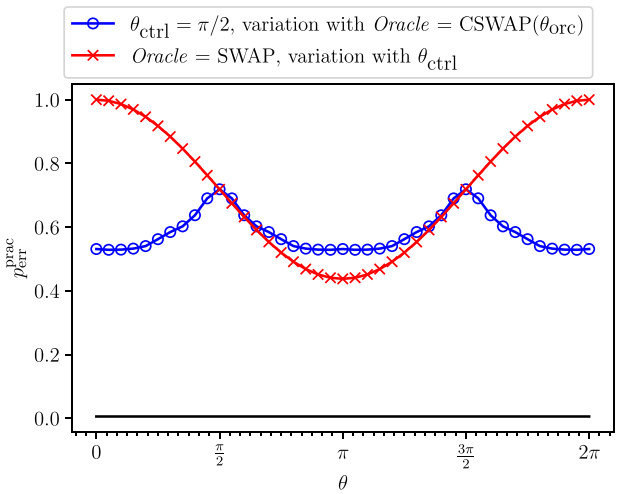


Figure 10. Illustration of variation of practical case error probability with respect to θ_{ctrl} and θ_{orc}. The blue line represents the case when $\theta_{\text{ctrl}} = \pi/2$ and we observe the variation of practical error probability with respect to the θ_{orc} and the red line is obtained by setting $\theta_{\text{orc}} = 0$, which makes the $U_{\text{orc}}^{H_{0/1}} = \text{SWAP}$ then we observe the variation with respect to the θ_{ctrl}. The black horizontal line is the error probability obtained for the `SWAP` using the definition provided in ref. [6], which is also presented in Equation (13).

modified version of Equation (13) that takes the form:

$$p_{\text{err}}^{\text{prac}} = 1 - \Delta\left[U_{\text{orc}}^{H_0}, U_{\text{orc}}^{H_1}\right](1 - p_{\text{err}}) \xrightarrow{r>>1} 1 - \Delta\left[U_{\text{orc}}^{H_0}, U_{\text{orc}}^{H_1}\right]\left(1 - \frac{1}{4rd^N}\right) \tag{14}$$

There are many choices for the distance Δ function between the hypotheses and need to be chosen based on the experimental and theoretical specifications of the application such as

1. *Trace distance* which is defined by

$$\Delta = \frac{1}{2}\text{Tr}|\rho_{\text{orc}} - \rho_{\text{orc}}^{\text{alter}}| \tag{15}$$

where ρ represents the Choi representation of the unitary.

2. *Bures distance* which defined as

$$\Delta = 2\left(1 - \sqrt{F(\rho_{\text{orc}}, \rho_{\text{orc}}^{\text{alter}})}\right) \tag{16}$$

where F quantifies the process fidelity between the Choi representation of oracles.

3. *Hilbert-Schmidt distance* which is defined by

$$\Delta = \text{Tr}\left[\left(\rho_{\text{orc}} - \rho_{\text{orc}}^{\text{alter}}\right)^2\right] \tag{17}$$

But the numerical results show that the error probability after simulating the circuit in Figure 10 i.e., practical case error probability, $P_{\text{err}}^{\text{prac}}$, coincides with *Hilbert-Schmidt distance*.

5. Numerical Results

To obtain numerical results, we consider IBM's open-source quantum computer simulator `Qiskit` was used to simulate the above-mentioned implementation. At first, we choose the oracle unitary as given in Equation (7) where q_{anc} is an ancilla qubit that allows us to control the strength of the *Hypothesis 1*. And the strength is dependent on parameter θ_{ctrl}. To obtain the specific hypothesis cases of $\mathbb{I}$ and `SWAP`, the above parameters are set to $U_{\text{orc}}^{H_0}(0)$ and $U_{\text{orc}}^{H_1}(\pi)$, respectively. The dependence of the error probability of distinguishing two hypotheses with respect to the relative difference is enquired by varying the alternate hypothesis to

$$U_{\text{orc}}^{H_{0/1}} = \begin{cases} U_{\text{orc}}^{H_0}(0) \\ U_{\text{orc}}^{H_1}(\theta_{\text{ctrl}}) \end{cases} \tag{18}$$

For the sake of better understanding, we compare the theoretical and practical case error probability, which is introduced in Equation (14). For the theoretical scenario, we numerically calculate the Equation (14) by utilizing the *Hilbert-Schmidt distance* between two hypotheses with respect to θ_{ctrl}. Meanwhile, for the practical scenario utilize the introduced causal hypothesis testing circuit in Figure 7 and get the error probability directly from the measurement outcomes.

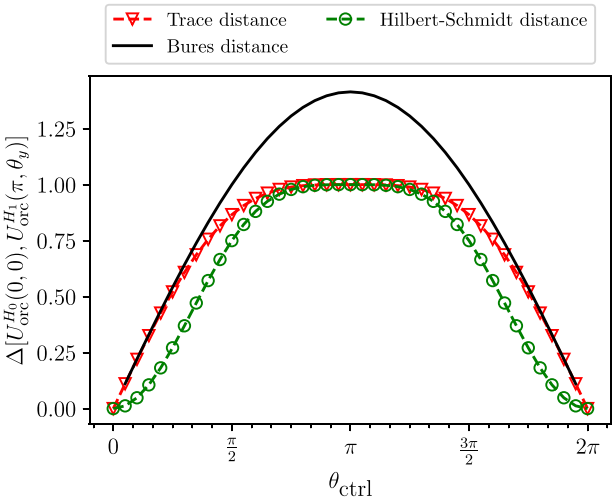


Figure 11. The variation of different distance measures with respect to the oracle angle θ_{ctrl}. It can be seen that at $\theta_{\text{ctrl}} = \pi$ the trace distance and Hilbert-Schmidt distance coincide whereas Bures distance surpasses them all.

In Figure 9, we illustrate the theoretical scenario to evaluate the error probability. The black horizontal line refers to the results corresponding to the limiting case error probability in ref. [6]. We find that when the $\theta_{\text{ctrl}} = 0$, the `RX` gate that is controlling the `SWAP` oracle is not activated. This makes the alternate hypothesis (*Hypothesis 1*) $\mathbb{I}$. As our default hypothesis is already $\mathbb{I}$, we can not distinguish between the two hypotheses by any method. This gives theoretical case error probability (i.e., $p_{\text{err}}^{\text{theo}}$) 1.0. In the same way when $\theta_{\text{ctrl}} = \pi$ the `SWAP` oracle is activated by the `RX` control and we find the minimal probability of distinguishing between `SWAP` and $\mathbb{I}$, overlapping with the results from ref. [6], in Equation (13). It is logical to examine the variation of process distance with respect to the oracle parameter θ_{ctrl}. In **Figure 11** we illustrate the distance between the null hypothesis $U_{\text{orc}}^{H_0}(0,0)$ and the alternate hypothesis $U_{\text{orc}}^{H_1}(\pi,\theta_{\text{ctrl}})$ for a class of distance measure.

For the sake of experiments, we restrict ourselves to the observation of *Trace distance*, *Bures distance*, and the *Hilbert-Schmidt distance*. It can be seen that the characteristics of *Trace* and *Hilbert-Schmidt* are similar since both of them fundamentally depend on the difference in the Choi representation of the oracles, whereas *Bures distance* depends on the process distance.

In the next experiment, we modify the unitary presented in Equation (7) to test the dependence on the error probability with other intermediary hypotheses. We generate these intermediary hypotheses by using a set of `parameterized CSWAP` gates that are generated by the decomposition implemented as 3 `iSWAP` with three interleaved `SX` gates on alternating qubits. This time the strength of the `parameterized CSWAP` depends on the angle in the ancilla qubit i.e., θ_{ctrl} (as shown in Figure 6) and the angle of the `iSWAP` i.e., θ_{orc}. The general form of the oracle is given by

$$U_{\text{orc}}^{H_{0/1}}(\theta_{\text{ctrl}},\theta_{\text{orc}}) = \left[\text{CSWAP}(\theta_{\text{orc}}, q_{\text{anc}}, A_i, B_i)^{\otimes N}\right] \times \left[\text{RX}(q_{\text{anc}},\theta_{\text{ctrl}}) \otimes \mathbb{I}(A_i)^{\otimes N} \otimes \mathbb{I}(B_i)^{\otimes N}\right] \quad (19)$$

The investigation of the hypothesis 19 is illustrated through Figure 10, where the quantum algorithm for causal hypothesis testing simulated in the `Qiskit` statevector simulator. In Figure 10, the red line shows the variation of the practical error probability with the `SWAP` gate, which is achievable if we set $\theta_{\text{orc}} = 0$. It can be seen that there is a significant difference in the results for the `SWAP` oracle presented in this results compared to the Figure 9 this is due to the practical case being implemented using the sum of probabilities of distinct states and the cumulative probability does not depend on the local phase. As it is not possible to capture the essence of the local phase that arises due to the `iSWAP` in Figure 10, we see the shift in the $p_{\text{err}}^{\text{prac}}$ compared to $p_{\text{err}}^{\text{theo}}$.

Meanwhile, the blue line shows the variation with respect to the `parameterized CSWAP` for the ancilla-control set to $\pi/2$. As expected, for the oracle angle of $\pi/2$, the `CSWAP` is the same as a `SWAP`.

We observe that the results for the theoretical case and the numerically simulated practical case error probability are distinct (see the red lines in Figures 9 and 10). This is because of two factors: i) the p_{err} was introduced purely theoretically where the authors evaluated it for a maximally mixed alternate hypothesis. To construct a maximally mixed hypothesis we need a unitary that can generate a maximally mixed state. In the practical case (numerical simulation of the quantum circuit) it is not possible to construct a maximally mixed unitary. This leads us to introduce the realistic case error probability and is presented in Equation (14). This modification is reflected through the introduction of the distance between the oracles under consideration i.e., $\Delta\left[U_{\text{orc}}^{\text{alter}}, U_{\text{orc}}\right]$. ii) the maximum distinguishing measurement as suggested in the original formulation, requires knowledge of the hypothesis and the oracles, which cannot be known apriori. While the theoretical process distance can be experimentally achieved via quantum process tomography, in the second experiment, we show the practical case error probability for Z-axis measurements without pre-rotations.

Thus, these experiments demonstrate that: i) the quantum circuit we proposed in Figure 7 can indeed be utilized for causal hypothesis testing, and ii) the success probability of distinguishing two hypotheses is dependent on the process distance between the two oracles.

6. Application Framework

With the quantum kernel presented above, in this section, we discuss an application framework in the context of two consequential applications.

6.1. Bioinformatics

Applications of causal inference are widespread in bioinformatics. Specifically, inferring a causal network is practiced in medical diagnostics and genomics.

For example, causal discovery in Alzheimer's pathophysiology is studied in ref. [29] with nine variables (13 with longitudinal data). Similarly, for detecting causal regulatory interactions between genes, tools like `Scribe-py`[30] are currently used. Exemplary use cases of (i) transcription expression dynamics hierarchy of C. elegans' early embryogenesis and (ii) core regulatory

network responsible for myelopoiesis are used for this research, with the latter graph consisting of ten nodes.

The current generation of quantum processors supports 100s of qubits and is expected to scale to 1000s in a few years. However, the challenge is the limitation of the decoherence time and gate errors, which bounds the runtime of the algorithm that can be effectively executed. For a causal graph in the order of ten causes/effects, a causal specificity bits of $d = 1$ and with $N = 100$, the estimation presented in Equation (12) is 6262 qubits. Keeping in mind the potential of near-term devices, pragmatic industrial cases of causal tomography will remain outside the reach of quantum computing in the near term.

For applications in medical diagnostics and genomics, a probabilistic graphical model encodes the assumptions about the data generation process. This probabilistic graphical model is known as a causal graph. Each application of the causal graph can be distinctly represented by a unique causal graph, which can be translated to quantum circuits via controlled rotation gates such as $\texttt{CRX}(\theta)$ where θ represents the weight of the edges of the causal graph.

6.2. Artificial General Intelligence

In the long term, quantum accelerated causal inference will benefit artificial general intelligence. Quantum accelerated AGI is still in its infancy. In ref. [31], the authors proposed the AIXI-q reinforcement learning agent empowered with quantum counting. Meanwhile, ref. [3] proposed an exhaustive enumeration of all causal oracles (or, alternatively, all bounded-size programs of a Turing machine). These techniques can develop synergies with the quantum accelerated causal tomography circuit as developed in this article.

In theoretical physics, automated science tools, specifically in the context of causal set theory, will also find the causal tomography framework of crucial use.

7. Conclusion

In this article, we extend the previously introduced causal hypothesis testing formulation[6] for the practical scenario. This led us to develop a scalable quantum gate-based algorithm that can be implemented in the available near-term quantum devices. Through the algorithm formulation, we empirically show that the limiting case error probability that is represented in ref. [6] requires modification. In our work, this modification is done by introducing process distance between the causal hypotheses, to the formulation of error probability. We term this modified version of error probability as practical case error probability, which is stricter than the limiting case. Additionally, our implementation enables an estimation of the pragmatic gate complexity of the causal tomography entangled pair indexing.

Furthermore, the proposed algorithm is implemented using the open-source quantum programming and simulation platform `Qiskit`. As in ref. [6] it is shown that the quantum advantage holds for generalized probabilistic theory so the practical case that we present here is optimal in any scenario.

Our motivation for this project is driven by the increasing focus on causal inference in ML. Besides monitoring the information flow in future quantum communication networks, as discussed in ref. [6], causal tomography is crucial for understanding the bounds of general intelligence and for bioinformatics use cases. In our future work, we aim to apply our developed quantum accelerated causal tomography framework for these applications.

Acknowledgements

This project was initiated under the QIntern 2021 project "Reinforcement Learning Agent for Quantum Foundations". Therefore, the authors would like to thank the organizers of the program and QWorld Association. The authors would also like to thank QIndia for providing us with a collaboration platform for the ideas motivating this research. A.K. was partially supported by the Polish National Science Center (NCN) under the grant agreement 2019/33/B/ST6/02011.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

causal hypothesis, causal inference, error probability, process distance

Received: September 26, 2023
Revised: June 2, 2024
Published online:

[1] R. Marcinkevičs, J. E. Vogt, *arXiv preprint arXiv:2012.01805* **2020**.
[2] R. Moraffah, M. Karami, R. Guo, A. Raglin, H. Liu, *ACM SIGKDD Explor. Newslett.* **2020**, *22*, 18.
[3] A. Sarkar, Z. Al-Ars, K. Bertels, *Appl. Sci.* **2021**, *11*, 2696.
[4] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, K. J. Resch, *Nat. Phys.* **2015**, *11*, 414.
[5] J. F. Fitzsimons, J. A. Jones, V. Vedral, *Sci. Rep.* **2015**, *5*, 1.
[6] G. Chiribella, D. Ebler, *Nat. Commun.* **2019**, *10*, 1.
[7] A. Lavin, D. Krakauer, H. Zenil, J. Gottschlich, T. Mattson, J. Brehmer, A. Anandkumar, S. Choudry, K. Rocki, A. Güneş Baydin, C. Prunkl, B. Paige, O. Isayev, E. Peterson, P. L. McMahon, J. Macke, K. Cranmer, J. Zhang, H. Wainwright, A. Hanuka, M. Veloso, S. Assefa, S. Zheng, A. Pfeffer, *arXiv preprint arXiv:2112.03235* **2021**.
[8] Y. Maruyama, in *International Conference on Artificial General Intelligence*, Springer, Berlin, Heidelberg **2021**, pp. 127–138.
[9] R. Guo, L. Cheng, J. Li, P. R. Hahn, H. Liu, *ACM Comput. Surveys (CSUR)* **2020**, *53*, 1.
[10] K. J. Rothman, S. Greenland, *Am. J. Public Health* **2005**, *95*, S144.
[11] J. Pearl, *Econometric Theory* **2003**, *19*, 675.
[12] G. W. Imbens, D. B. Rubin, *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press, Cambridge **2015**.
[13] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, A. Zhang, *ACM Trans. Knowl. Discov. Data* **2021**, *15*, 1.
[14] P. Spirtes, C. N. Glymour, R. Scheines, D. Heckerman, *Causation, prediction, and search*, MIT press, Cambridge **2000**.

[15] H. Zenil, N. A. Kiani, A. A. Zea, J. Tegnér, *Nat. Mach. Intell.* **2019**, *1*, 58.
[16] A. Sarkar, Phd thesis, Delft University of Technology, **2022**.
[17] A. Sarkar, Z. Al-Ars, H. Gandhi, K. Bertels, *arXiv preprint arXiv:2112.03643* **2021**.
[18] P. Steinmüller, T. Schulz, F. Graf, D. Herr, *arXiv preprint arXiv:2211.01441* **2022**.
[19] R. Heese, T. Gerlach, S. Mücke, S. Müller, M. Jakobs, N. Piatkowski, *arXiv preprint arXiv:2301.09138* **2023**.
[20] I. Burge, M. Barbeau, J. Garcia-Alfaro, *arXiv preprint arXiv:2301.04727* **2023**.
[21] L. Pira, C. Ferrie, *arXiv preprint arXiv:2308.11098* **2023**.
[22] F. Costa, S. Shrapnel, *New J. Phys.* **2016**, *18*, 063032.
[23] C. Giarmatzi, in *Rethinking Causality in Quantum Mechanics*, Springer, Berlin, Heidelberg **2019**, pp. 125–150.
[24] M. A. Javidian, V. Aggarwal, F. Bao, Z. Jacob, in *Quantum Information and Measurement*, Optical Society of America, Washington, D.C. **2021**, p. F2C–3.
[25] G. Bai, Y.-D. Wu, Y. Zhu, M. Hayashi, G. Chiribella, *npj Quantum Inf* **2022**, *8*, 69.
[26] G. Bai, Y.-D. Wu, Y. Zhu, M. Hayashi, G. Chiribella, *arXiv preprint arXiv:2012.01731* **2020**.
[27] J. Bavaresco, M. Murao, M. T. Quintino, *Phys. Rev. Lett.* **2021**, *127*, 200504.
[28] C. W. Helstrom, *J. Stat. Phys.* **1969**, *1*, 231.
[29] X. Shen, S. Ma, P. Vemuri, G. Simon, *Sci. Rep.* **2020**, *10*, 1.
[30] X. Qiu, A. Rahimzamani, L. Wang, B. Ren, Q. Mao, T. Durham, J. L. McFaline-Figueroa, L. Saunders, C. Trapnell, S. Kannan, *Cell Syst.* **2020**, *10*, 265.
[31] E. Catt, M. Hutter, *arXiv preprint arXiv:2005.03137* **2020**.

