

# Quantum Algorithms for Compositional Text Processing

Tuomas Laakkonen, Konstantinos Meichanetzidis, Bob Coecke

Quantinuum, 17 Beaumont Street, Oxford OX1 2NA, United Kingdom

{tuomas.laakkonen, k.mei, bob.coecke}@quantinuum.com

Quantum computing and AI have found a fruitful intersection in the field of natural language processing. We focus on the recently proposed DisCoCirc framework for natural language, and propose a quantum adaptation, QDisCoCirc. This is motivated by a compositional approach to rendering AI interpretable: the behavior of the whole can be understood in terms of the behavior of parts, and the way they are put together.

For the model-native primitive operation of text similarity, we derive quantum algorithms for fault-tolerant quantum computers to solve the task of question-answering within QDisCoCirc, and show that this is BQP-hard; note that we do not consider the complexity of question-answering in other natural language processing models. Assuming widely-held conjectures, implementing the proposed model classically would require super-polynomial resources. Therefore, it could provide a meaningful demonstration of the power of practical quantum processors.

The model construction builds on previous work in compositional quantum natural language processing. Word embeddings are encoded as parameterized quantum circuits, and compositionality here means that the quantum circuits compose according to the linguistic structure of the text. We outline a method for evaluating the model on near-term quantum processors, and elsewhere we report on a recent implementation of this on quantum hardware.

In addition, we adapt a quantum algorithm for the closest vector problem to obtain a Grover-like speedup in the fault-tolerant regime for our model. This provides an unconditional quadratic speedup over any classical algorithm in certain circumstances, which we will verify empirically in future work.

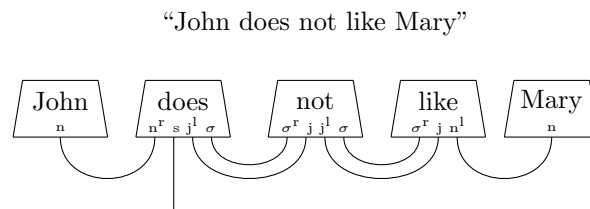
## 1 Introduction

Artificial intelligence permeates a wide range of areas of activity, from academia to industrial applications, with natural language processing (NLP) standing center stage. In parallel, quantum computing has seen a surge in development, and practical quantum processors are reaching scales where small-scale quantum algorithms are becoming feasible. The merging of these two fields has given rise to the young area of research on quantum natural language processing (QNLP). Here we focus on the ‘first wave’ of QNLP, starting around 2016 [14, 28, 37, 40, 54], although there are also other approaches [51, 52]. One important feature that distinguishes that first wave from other work, and from the broad field of quantum machine learning [6, 21, 46], is that it provides a path towards *explainability* and *interpretability*.

Despite all of the impressive advancements of contemporary artificial intelligence – which is synonymous with machine learning methods – these advancements have been achieved by training black-box deep-learning models on large amounts of data. In order to arrive at general applicability and high performance, what is often compromised by such setups, at least in terms of model architecture, is exactly explainability and interpretability. When things go wrong unexpectedly, you typically don’t know why. Therefore, several attempts have been put forward recently to render black-box deep-learning models more explainable and interpretable. For example, methods for *post-hoc* mechanistic interpretability have recently formed an active area of research [9]. Conversely, first-wave QNLP made use of an earlier

quantum-inspired model for language [11, 16] that had been crafted precisely with the features of explainability and interpretability in mind. These features were achieved through *compositionality*.

For us, compositionality means that the behavior of the whole can be understood in terms of the behavior of parts, and the way they are put together [13]. In the case of language this includes linguistic meaning as well as linguistic structure such as grammar and coreference. An early compositional framework for natural language that combined linguistic meaning and linguistic structure was DisCoCat [16, 24, 29, 44]. Somewhat surprisingly, this was enabled by the foundational research program of categorical quantum mechanics [1, 15]: DisCoCat emerged from the observation that the category-theoretic structure of categorical quantum mechanics perfectly matched Lambek’s model of linguistic structure in terms of pregroups [31, 32]. In this way, linguistic structure mediates how word-meanings interact in order to form phrase- and sentence-meanings – see [11, 14] for more details. Just as it is the case in categorical quantum mechanics, sentences in DisCoCat take the form of diagrams:



where the boxes capture word-meanings and the wires represent the grammatical interactions between those word-meanings, yielding the meaning of the sentence as a whole. The fact that DisCoCat essentially was derived from a quantum mechanical formalism, enabled one to use it to easily craft quantum NLP models, which we will refer to as QDisCoCat [14, 39, 54]. This was the foundation of the ‘first wave’ of QNLP.

QDisCoCat models have been implemented on quantum hardware for the task of sentence classification [28, 37, 40], but they have some shortcomings. Firstly, QDisCoCat diagrams have to be converted into quantum circuits in order to fit them on a quantum computer, and consequently, quantum algorithms derived from them require post-selection – no lower bounds are known for the post-selection probability, so these algorithms may require exponential time in the worst case.<sup>1</sup> A related issue is that the reliance of DisCoCat on ‘categorical grammars’ [30, 31], which cannot natively represent texts larger than individual sentences. Both problems are addressed in the DisCoCirc framework [12, 49], where sentences are represented by circuits, and these circuits can be further composed to represent texts. We call these *text circuits*. This bypasses the need for conversion of diagrams into circuits, making post-selection unnecessary. Besides these, DisCoCirc enjoys a number other advantages that are discussed in [49], and has been studied within a wide variety of contexts [19, 43].

In this work, we propose QDisCoCirc, an adaptation of DisCoCirc which provides *efficient quantum algorithms* for several NLP tasks within the model. In particular, in Section 4 we show that there exists an NLP task, namely question-answering, which is BQP-hard to solve within the QDisCoCirc framework. Therefore, evaluating this model can provide a demonstration of the power of practical quantum processors, as simulating the model classically would require super-polynomial resources (assuming widely-held hardness conjectures). Furthermore, this would be less abstract compared to experiments such as those based on random circuit sampling [2, 18], which are convincing demonstrations of quantum hardware, but do not attempt to solve a widely applicable task. Importantly, we do not claim that the

<sup>1</sup>Sample-efficient sentence-level QNLP models can be constructed [25], but they are not compositional in the above sense.

underlying NLP task of question-answering cannot be solved efficiently with a classical machine, only that the QDisCoCirc model in particular would be hard to evaluate in this case.

We believe that QDisCoCirc models are worth considering because of the interpretability and explainability they afford. However, other concrete implementations of the DisCoCirc framework could also be proposed – for example, based on classical neural networks – and in any quantum model, it is necessary to justify why it should be preferred over a classical model, given the extreme difficulty of constructing quantum devices. We argue that a quantum model is most natural for this framework. DisCoCirc, via DisCoCat, is ultimately derived from pregroup grammars [32], where the composition of spaces is their tensor product. This has been argued as essential to fully capture the semantics of language [16, 24]. In fact, we have the following result:

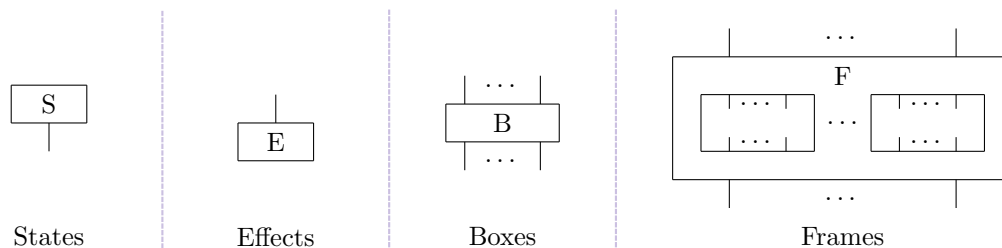
**Proposition** (Informal, [17, Theorem 2.1] & [15, Theorem 5.49]). *If a model has the same abstract structure as pregroup grammars, composition of spaces must be analogous to the tensor product.*

This can be made precise in several ways, as discussed in [15, 17]. This fixes the compositional structure of our model, but leaves the underlying Hilbert space undefined. As is customary in machine learning, we take this to be numerical vectors with the usual dot product, and in this case these models become tensor networks. Since large general tensor networks are not feasible to evaluate, both for quantum and classical devices, we must pick a subset which can be efficiently evaluated. We choose to base our model on quantum circuits, which represent the largest such subset currently known [7].

Finally, in Section 5 we give an algorithm that achieves for QDisCoCirc a Grover-like quadratic speedup over any classical algorithm for tasks such as text similarity, under certain technical assumptions about the dataset which can be verified empirically. This algorithm improves over the equivalent algorithm for QDisCoCat [54], which considers a technological regime where fault-tolerant scalable quantum computers and large quantum random access memories (QRAM) are available, while this is unlikely to be feasible in the near term. In contrast, our algorithm uses exponentially fewer bits of QRAM. Additionally, in Section 3 we present algorithms that do not use a QRAM, but are quadratically slower than the algorithm given in Section 5. They require much fewer quantum resources, and so are practical to implement on near-term machines. In Section 3, we show how text circuits can be mapped into parameterized quantum circuits to realize these algorithms.

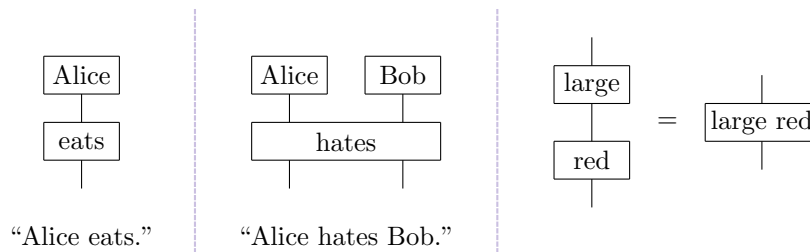
## 2 DisCoCirc

In the DisCoCirc framework, texts are the result of the composition of sentences, analogously to how sentences are compositions of words. Texts are represented by text circuits, which are read top to bottom. Boxes represent processes, and every box has input wires and output wires. In such circuits, only connectivity matters and so the information flow is explicit in the circuit. A circuit is composed of *generators*, or atomic processes, each of which has a linguistic interpretation – see [35] for more information. Text circuits are composed of the following set of generators, which we call *states*, *effects*, *boxes*, and *frames*:

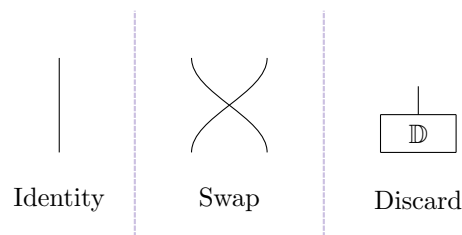


Generators can be composed – sequential composition of processes can be done by connecting the output wires of one process with the input wires of another process, and parallel composition is performed by placing boxes side by side. An effect is to be understood as a ‘test’, or the inverse, of a state. Applying an effect to a state, by joining all of their wires, results in a circuit with no open wires, which we call a *scalar*.

In this approach to modeling text as circuits, nouns are ‘first class citizens’, and are represented as states. The idea is that noun states are carried along wires, and so all wires here carry the same type, the noun type. Boxes are processes that transform states. Examples of these are adjectives or intransitive and transitive verbs, that act on nouns and pairs of nouns respectively. In general, generators with compatible shapes can be serially composed:

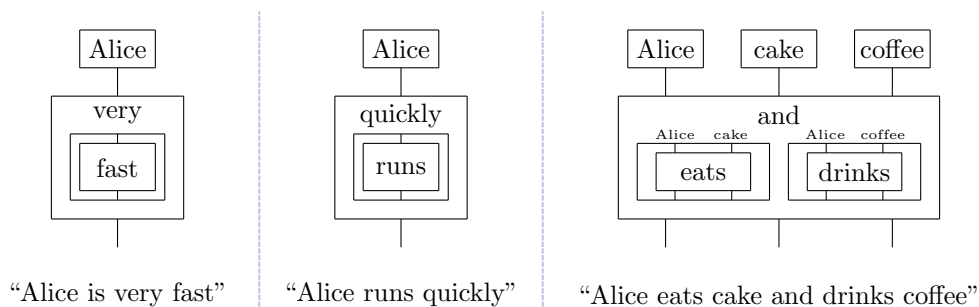


We also have some special generators – the identity, the swap, and the discard effect:

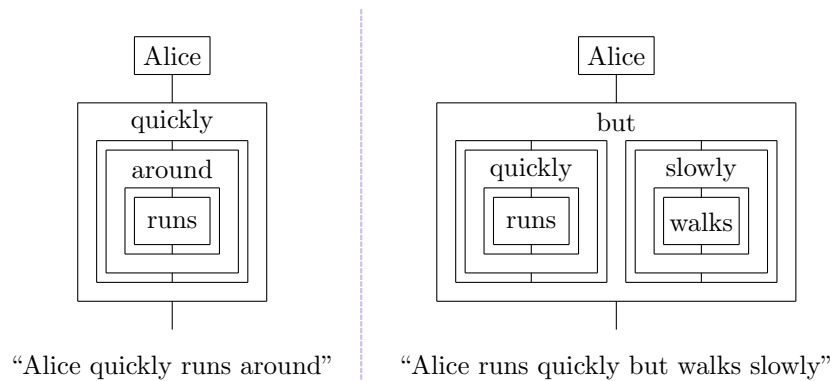


These interact with each other in a well-behaved way [15], so that *only connectivity matters* – we can bend and move wires freely so that equality of circuits is considered up to connectivity between inputs and outputs of generators. We consider any circuit to be state, effect, box, or frame if it has a compatible pattern of input and output wires. Additionally, we associate each generator with an inverse, which when composed serially yields the identity. Inverses for composite circuits are constructed by mirroring the whole circuit vertically and replacing every individual generator with its inverse.

Frames are super-maps that transform boxes, for example, intensifiers that act on adjectives, adverbs that act on verbs, or conjunctions that act on phrases:



Note that not all wires may occur in all the holes in each frame. In this case, the interior wires can be labeled with exterior wires to which they correspond (note that this need not be a single wire). Frames can also be serially composed:



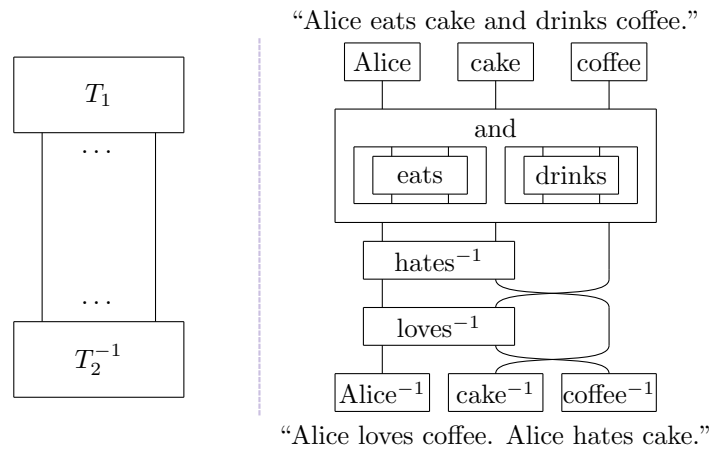
The composition of the generators of text circuits is not free; not every composition is allowed – for instance, wires may not pass through frames. Compositions are restricted in the sense that text circuit are the output of a text-to-circuit *parser* [35], which for this work we assume as given – the parser generates parse trees for each sentence that are converted to text circuits, and the wires corresponding to common nouns are connected using coreference resolution so that generators concerning the same nouns throughout the text act on the same wires in the circuit. Arbitrary text circuits are *local*, as the number of input and output wires of every box is upper bounded by a constant that is independent of the text size. Furthermore, we assume that text circuits derived from a text via the parser are acyclic.

## 2.1 Native NLP Tasks

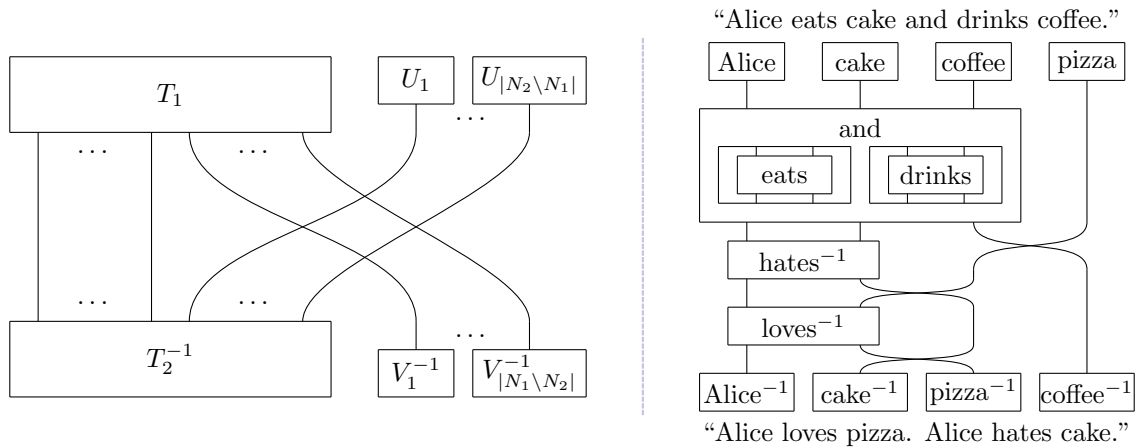
The range of NLP tasks is wide and diverse. Here, we consider tasks that are *native* to DisCoCirc. That is, we formulate simple NLP tasks using the compositional structure of the model. The key idea is that text circuits constructed by the parser pipeline are states which carry the information of the text. Then this information is retrieved by constructing appropriate *tests* in terms of effects. In general, the effects are constructed as (inverses of) text circuits themselves. The primitive operation that we will use to define NLP tasks is that of *similarity*, or overlap, which is tested by serially composing effects onto states. In this way, we can represent each task as a single, scalar, text circuit.

First we consider the text-level analogue of sentence similarity as presented in the DisCoCat framework [54], ie *text similarity*. Suppose  $T_1$  and  $T_2$  are text circuits defining states obtained from the parser, defined on the wires generated by same set of nouns  $N$ . We form a scalar text circuit corresponding to

their similarity by serially composing one with the inverse of the other. For example:

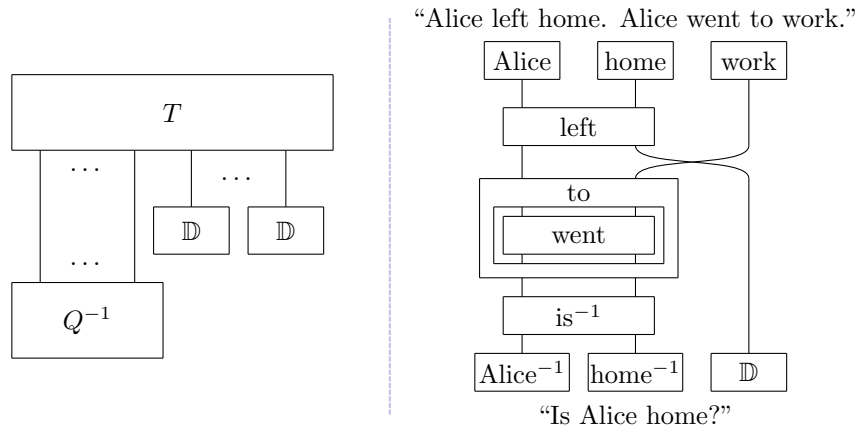


In the case that the circuits are not defined on the same set of nouns, we can still compute the similarity. Suppose  $T_1$  and  $T_2$  be defined on the sets of nouns  $N_1$  and  $N_2$ . Let  $U_1, U_2, \dots, U_{|N_2 \setminus N_1|}$  be the noun states present in  $T_2$  but not  $T_1$ , and similarly  $V_1, V_2, \dots, V_{|N_1 \setminus N_2|}$  be the noun states present in  $T_1$  but not  $T_2$ . By parallel composing  $T_1$  with  $U_i$  and  $T_2$  with  $V_i$ , we form two text circuits that are defined on the same wires. For example:



While this task compares the whole of both texts, we can extract more specific information about a text by only comparing against a subset of the nouns that we wish to know about. We call this task *question answering*. In this task, we want to ask a question about a subset of the nouns in the text. To do this, we formulate the question as a statement (for example, ‘Is Alice home?’ becomes ‘Alice is home.’) and then we compute the similarity of the text with the inverse of the question in the same way as above, except

that we discard the noun wires in the text that are not in the subset used by the question. For example:



In the case of questions which have several different possible answers, we formulate such a question statement and thus text circuit for each possible answer. For example, ‘Is Alice home?’ becomes a pair of question statements ‘Alice is home.’ and ‘Alice is not home.’ Then we compare the scalar text circuits formed in each case to determine the answer to the question.

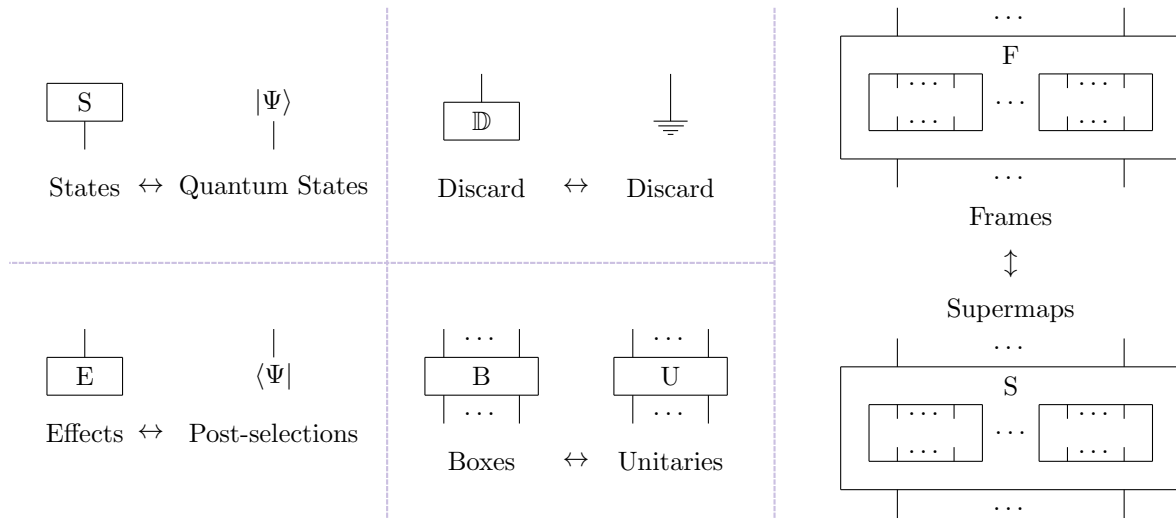
In Section 6, we show a collection of other model-native NLP tasks that enjoy a natural representation as text circuits. In general, DisCoCirc itself does not prescribe a particular semantics for any of these tasks (for instance, we leave the way that scalar circuits are computed or compared unspecified) — it provides a method for constructing the text circuits representing the input texts, as well as an intuitive framework for designing circuits corresponding to a particular task. Moreover, once an implementation of the model is decided, the word-embeddings that form the boxes will be derived to solve a specific task. The aim of DisCoCirc is only to incorporate linguistic structure into the model, so it is not imperative that these task-circuits are defined rigorously (for example, as a statement of discourse representation theory), but rather that they are ‘intuitively correct’ so that the imparted structure is helpful in practice. In the next section, we will construct a specific implementation of DisCoCirc that fills in these details.

### 3 QDisCoCirc Models

To construct a concrete DisCoCirc model that performs a model-native NLP task for any potential text and task, we apply a structure-preserving map to the corresponding text circuit. This is achieved by defining a map that is applied to every generator individually, in a way that respects the relations defined on the generators. As discussed in the introduction, a natural choice of space in which such circuits may exist is tensor networks. As tensor networks are hard to evaluate in general, we propose QDisCoCirc, a family of models based on quantum circuits, which are a restricted family of tensor networks, and can be efficiently evaluated on a quantum computer.

In QDisCoCirc, text circuits are translated to quantum circuits with postselections by replacing each generator with a quantum operation or state and each wire with some fixed number of qubits. Further-

more, the inverses of generators are given by their adjoints. The generators are mapped as follows:



The specific form of supermaps obtained via this mapping will be discussed shortly. We identify post-selected quantum circuit diagrams with the density matrices they represent (which are not normalized in general); thus, scalar text circuits are indeed transformed to scalars representing the probability of successfully post-selecting the given circuit.

To solve the model-native tasks defined above using this model, we observe that all the scalar text circuits for the tasks take the form of  $A$  serially composed with  $B^{-1}$  for some circuits  $A$  and  $B$  (this is true also for the tasks presented in Section 6). This corresponds to a text similarity operation. Through the quantum semantic functor, this operation maps to quantum state overlap. According to the Born rule, these result to expressions of the form  $F(\Psi, \Phi) = |\langle\Psi | \Phi\rangle|^2$  for some quantum states  $\Psi$  and  $\Phi$  in the pure case, and in the general case  $F(\rho, \sigma) = \text{tr}(\rho\sigma)$  for density matrices  $\rho$  and  $\sigma$ .

However, in the case that both  $\rho$  and  $\sigma$  are mixed, this does not satisfy the properties we would expect from a similarity operation, such as  $F(\rho, \rho) = 1$  and  $F(\rho, \sigma) < 1$  if  $\rho \neq \sigma$ . There are a few quantities [33] such as the Uhlmann-Josza fidelity  $\text{tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2$  which do satisfy these properties for mixed states, but unfortunately they cannot be computed efficiently using a quantum computer. Therefore, we will ensure text circuits obtained from the parser pipeline are translated to pure quantum states, so that for the tasks defined previously, *at most one* of the states representing  $A$  and  $B$  is mixed.

Since the parser does not generate discards, we just need to make sure none of the generators are mapped to mixed operations. This is trivially true for states, effects, and boxes, but frames are more problematic. Since states map to quantum states and boxes to unitaries, frames become maps between unitaries – supermaps – by analogy. However, some frames in text circuits may contain more wires for one of their arguments than are input and output to the frame – this arises in many cases, for instance in reflexive structures. Furthermore, general quantum supermaps can be written as pairs of unitaries conjugating their arguments, along with a side channel consisting of some auxiliary qubits. In either case, we cannot use auxiliary qubits or wires because these would then be subsequently discarded, making the state mixed. Therefore, we instead constrain define our mapping from text circuits such that these situations do not occur, see Appendix A for details.



### 3.1 Solving Model-Native Tasks

After decomposing the frames in a text circuit, the resulting QDisCoCirc quantum circuit will contain only quantum states and unitaries labeled with the words from which they originated. For each task, we assume we are given these states and unitaries as quantum circuits, which we call *quantum word embeddings*, such that the circuits defined for the task are evaluated as expected. We will consider in future work how to pick word embeddings such that this is the case. In particular, for text-similarity tasks, that the degree of semantic similarity between the two texts is indicated by  $F(\rho, \sigma) = \text{tr}(\rho\sigma)$ . We can solve tasks based on this primitive using the *swap test* [10]:

**Theorem 1.** *The text-similarity primitive of QDisCoCirc models between two text circuits  $A$  and  $B$  can be approximated to precision  $\varepsilon$  in polynomial time on a quantum computer. In particular, given  $\varepsilon > 0$  and  $\rho, \sigma$  the quantum states derived from  $A$  and  $B$  respectively, the quantity  $\text{tr}(\rho\sigma)$  can be determined to precision  $\varepsilon$  with failure probability  $\delta$  by a quantum circuit of size*

$$O\left(\frac{(|A| + |B|)\log(\frac{1}{\delta})}{\varepsilon^2}\right)$$

where  $|A|$  and  $|B|$  are the number of elementary quantum gates comprising  $\rho$  and  $\sigma$ .

*Proof.* We construct a quantum circuit based on the swap test [10, Figure 1]. This gives a circuit of size  $O(|A| + |B|)$  for which the first qubit has measurement probability  $\frac{1}{2} + \frac{\text{tr}(\rho\sigma)}{2}$ . By a Hoeffding bound,  $O(\log(\frac{1}{\delta})\varepsilon^{-2})$  samples of this circuit suffice to estimate  $\text{tr}(\rho\sigma)$  to precision  $\varepsilon$  with failure probability  $\delta$ .  $\square$

To solve the question-answering task in the QDisCoCirc model, we are given a text  $T$  and a set of questions  $Q_i$  and wish to find the  $Q_i$  which maximizes  $\text{tr}(\rho_T, \rho_{Q_i})$ . We can do this by evaluating  $\text{tr}(\rho_T, \rho_{Q_i})$  for each  $i$  using the method above, and then taking the maximum of these. We show below that this also runs in polynomial time on a quantum computer. To compare this algorithm to classical algorithms, it is important to consider the scaling of  $\varepsilon$  carefully. The best-known classical method to calculate  $\text{tr}(\rho\sigma)$  would at least require performing statevector simulation of the quantum circuit. In this case, we would expect a total time of at least  $O((|A| + |B|)N^n \text{polylog}(\varepsilon^{-1}))$  and memory requirement of  $O(N^n \log(\varepsilon^{-1}))$ , where  $\log(N)$  is the number of qubits per wire, and  $n$  is the number of wires in the text circuits – this is an exponentially worse dependency on  $N$  and  $n$ .

On the other hand, the scaling in terms of  $\varepsilon$  is much better in the classical case. For instance, consider question answering. If the number of nouns referenced in the question is constant  $c$ , then only  $c \log(N)$  qubits will not be discarded. In this case, if the quantum states corresponding to the questions are sufficiently uniformly distributed then  $\varepsilon = O(N^{-c})$  should suffice to resolve the maximum similarity, by an argument of Wiebe et al [53], which gives exponentially better dependence in  $n$ . However, considering the case of full text-text similarity with no discarding, the quantities  $\text{tr}(\rho\sigma)$  may grow as small as  $N^{-n}$  in general, and so if small relative error is needed,  $\varepsilon = O(N^{-n})$  may be required, which makes the complexity at least as bad as the classical algorithm. Therefore, for any given dataset and word embeddings, it would be important to analyze the scaling of  $\varepsilon$  in terms of  $N$  and  $n$ , either analytically or empirically.

## 4 Hardness for Question Answering

We will now consider the problem of question answering in the QDisCoCirc framework more formally. In particular, we will show that for the appropriate choices of texts and word embeddings, question

Embeddings \ Texts	Texts			
		<i>Worst-Case</i>	<i>Average-Case</i>	<i>Typical</i>
<i>Worst-Case</i>		BQP-hard (Thm.3)	BQP-hard (Thm.6)	Empirical only
<i>Average-Case</i>		BQP-hard (Thm.4)	Unknown	Empirical only
<i>Typical</i>		Testable (Thm.5)	Unknown	Empirical only

Table 1: The hardness results shown in this paper. The embeddings axis refers to the choice of word embeddings  $V$ , whereas the texts axis refers to the choice of context  $T$  and questions  $\{Q_i\}$ . ‘Average-case’ hardness refers to a ‘natural’ distribution - for embeddings this is the Haar distribution, whereas for texts we define a distribution. ‘Typical’ hardness refers to texts and embeddings taken from actual human-generated data. Since typical texts are always bounded by definition, only empirical evidence can be given for hardness in this case.

answering is BQP-hard, and that there is an efficient quantum algorithm to solve it. Note these results only apply to question answering strictly as specified by the QDisCoCirc framework, not as a general NLP task. We call this task QDISCO-CIRC-QA to emphasize the distinction. Proofs for the theorems in this section are given in Appendix B.

**Definition 1.** *The problem QDISCO-CIRC-QA is defined as follows: given a set of word embeddings  $V$ , a context text  $T$ , and a set of  $k$  question texts  $\{Q_i\}$ , determine any  $j$  such that*

$$\left| \text{tr}(\rho_T(\rho_{Q_j} \otimes I)) - \max_i \text{tr}(\rho_T(\rho_{Q_i} \otimes I)) \right| < \varepsilon$$

where  $\rho_T = U_T|0\rangle\langle 0|U_T^\dagger$ ,  $\rho_{Q_i} = U_{Q_i}|0\rangle\langle 0|U_{Q_i}^\dagger$ , and  $U_T, U_{Q_i}$  are the QDisCoCirc text circuits generated from  $T$  and  $Q_i$  respectively over  $V$ .

**Theorem 2.** *QDISCO-CIRC-QA can be solved on a quantum computer in time*

$$O\left(\frac{k \log\left(\frac{k}{\delta}\right) |V|_w (|T| + \max_i |Q_i|)}{\varepsilon^2}\right)$$

with precision  $\varepsilon$  and failure probability  $\delta$ , where  $|V|_w$  is the maximum number of elementary gates of any word embedding in  $V$ , and  $|T|, |Q_i|$  are the number of elementary gates in  $U_T$  and  $U_{Q_i}$  respectively.

Note that if the separation between the values  $\text{tr}(\rho_T(\rho_{Q_i} \otimes I))$  is smaller than  $\varepsilon$ , then the formulation of QDISCO-CIRC-QA does not guarantee that the ‘correct’ answer (in the sense of maximum similarity) will be returned. We assume that word embeddings exist such that  $\varepsilon$  need not be too small, as discussed in Section 3. Now we tackle the hardness results in the first column of Table 1 - that is, we will show that for some sets of word embeddings  $V$  it is possible to construct context texts  $T$  and questions  $\{Q_i\}$  for which QDISCO-CIRC-QA is BQP-hard.

**Theorem 3.** *Suppose that a set of word embeddings  $V$  satisfies the following:*

1. *The operations of  $V$  use one qubit for each input wire,*
2.  *$V$  contains arbitrarily many proper nouns,*
3.  *$V$  contains at least two adjectives that generate a dense subset of  $SU(2)$ ,*
4.  *$V$  contains at least one transitive verb that is entangling*

*then for any fixed  $\varepsilon < \frac{1}{7}$ , QDISCO-CIRC-QA is BQP-hard.*

Supposing that we have word embeddings with enough nouns (and thus enough qubits) to perform the reduction in the previous theorem, the particular unitaries in the word embeddings aren't important – Haar-random word embeddings suffice with high probability. This shows that QDISCOCIRC-QA is hard in the average-case over word embeddings.

**Theorem 4.** *Given a set of word embeddings  $V$ , suppose that operations in  $V$  are independent Haar-random unitaries. Then conditions three and four of Theorem 3 are almost surely satisfied for all word embeddings containing at least two adjectives and one transitive verb.*

Moreover, suppose we are given a set of word embeddings (for instance, that can solve a particular task), then we give a method to test if QDISCOCIRC-QA is hard in this case.

**Theorem 5.** *Given a specific set of word embeddings  $V$  it is possible to check numerically that the conditions of Theorem 3 are satisfied.*

To show average-case hardness of QDISCOCIRC-QA in terms of texts (second column in Table 1), we will first define a distribution over QDisCoCirc circuits of a given size, and then adapt Theorem 3 to show that we can draw polynomially sized texts  $T$ ,  $\{Q_i\}$  from this distribution and construct a set of word embeddings  $V$  such that solving the corresponding question answering instance is BQP-hard. Note that this distribution is not particularly representative of human-generated texts.

**Definition 2.** *For every size  $k$ , we define a distribution  $\mathcal{D}_k$  of text circuits with  $k$  unique states:*

- *There are no frames, and there are  $\Omega(k^\gamma)$  boxes with high probability for some  $\gamma > 4$ .*
- *The label of each box is drawn independently from a distribution of words that obey Heap's law and Zipf's law exactly.*
- *The number of inputs to each box is drawn independently from a distribution  $A$  such that  $P(A = 2)$  is nonzero and each input to a box is selected uniformly randomly from all possible wires.*

We are now ready to show the reduction. However, to make this work we need the number of words that only occur once in a given text (*hapax legomena*) to be bounded below by a polynomial in the total size of the text. Clearly, this cannot hold for all sizes for any finite vocabulary. For typically-sized texts, this assumption is justified statistically by Heap's law and Zipf's law [34, 45] — it has been shown that for large corpora, roughly half of the vocabulary are hapax legomena, and hence their number scales like  $\propto N^k$  for some  $k \approx 0.5$  [3]. However, since we are considering the limit of arbitrarily-sized texts which do not exist in practice (all human texts are bounded), we require that this behavior can be extrapolated.

**Theorem 6.** *There exists a polynomial  $f(m, n)$ , such that given an oracle to solve QDISCOCIRC-QA instances with arbitrary word embeddings and text circuits drawn from  $\mathcal{D}_{f(m, n)}$ , we can perform arbitrary quantum computations with  $m$  2-local gates on  $n$  qubits with high probability in polynomial time. That is, QDISCOCIRC-QA for worst-case word embeddings is average-case BQP-hard over texts drawn from  $\mathcal{D}_{f(m, n)}$ .*

Since QDISCOCIRC-QA is BQP-hard, we expect that it is not easy to solve with a classical computer, and that quantum algorithms to solve this ought to provide super-polynomial speedups. Therefore, supposing that (a) there exists a set of word embeddings for which a QDisCoCirc model is accurate on a dataset of interest, (b) the resulting quantum circuits are not easy to simulate classically (in the sense of Theorem 5), (c) a quantum device exists with sufficient fidelity to provide meaningful results, and (d) the precision  $\epsilon$  required is not too high (in the sense of Section 3.1), then it would be possible to perform an experiment on a quantum device which is hard to simulate classically but solves a task of interest. However, it is important to note that for many datasets this task can be solved classically by other means (notwithstanding the arguments presented in favor of QDisCoCirc in Section 1), so this would *not* constitute a demonstration of quantum advantage over *every* classical algorithm.

## 5 Further Quadratic Speedups

In this section, we present a quantum algorithm for obtaining polynomial speedups for the task of text question answering that we presented in Sections 3 and 4. Specifically, our results are obtained by adapting and improving on the approach of the work of Zeng and Coecke [54]. These algorithms provide a quadratic improvement over the simpler swap test algorithm, as well as being quadratically faster than any classical algorithm for solving the same task. While they require a QRAM [23], it is used to store exponentially fewer bits as compared to [54].

Basheer et al [5] give a quantum algorithm for accelerating the  $k$ -nearest neighbors problem as follows. Given the vectors  $v_0 \in \mathbb{C}^N$  and  $\{v_1, \dots, v_k\} \subset \mathbb{C}^N$  such that  $\|v_i\| = 1$  and quantum amplitude encoding oracles  $\mathcal{V}|0\rangle = |v_0\rangle$  and  $\mathcal{W}|i\rangle|0\rangle = |i\rangle|v_i\rangle$  with gate complexities  $T_V$  and  $T_W$ , then there is a quantum algorithm to find  $i = \arg \max_{j \geq 1} |\langle v_0 | v_j \rangle|$  with error at most  $\varepsilon$  and fixed positive success probability in time  $\tilde{O}\left(\sqrt{k}\varepsilon^{-1}(T_V + T_W + \log^2 N)\right)$ . The algorithm takes quantum states, encodes the fidelity via a partial swap test, then digitizes this with the quantum analog-to-digital conversion algorithm of Mitarai et al [41] before finding the minimum via Dürr-Høyer search [22]. In the case that  $\mathcal{V}(|0\rangle\langle 0|) = \rho_0$  and  $\mathcal{W}(|i\rangle\langle i| \otimes |0\rangle\langle 0|) = |i\rangle\langle i| \otimes \rho_i$  are actually operators preparing mixed states, then the quantity calculated will be  $i = \arg \max_{j \geq 1} \text{tr}(\rho_0 \rho_i)$ .

### 5.1 Application to DisCoCirc

Since DisCoCirc is based on a circuit architecture, it is natural to work with the *word embeddings* described as *quantum circuits*, and construct the amplitude encoding of texts directly as quantum circuits. We will apply the nearest-neighbor algorithm to the following task: given a text  $T_0$  and a set of texts  $T_i$  for  $k \geq i > 0$ , find  $i$  such that  $T_0$  and  $T_i$  are the most similar, and thus that maximizes  $\text{tr}(\rho_0 \rho_i)$ , where  $\rho_i$  is the density matrix corresponding to the QDisCoCirc circuit for  $T_i$ . This extends both the question-answering and text-similarity tasks given in Section 3.

Unlike the previous section where we assumed that the boxes in QDisCoCirc were arbitrary unitaries, we now suppose that each box with  $d$  wires is a particular instance of a parameterized quantum circuit ansatz. Let us define the following hyperparameters: (a) each noun wire is mapped to a set of qubits with Hilbert space dimension  $N$ , (b) each QDisCoCirc circuit considered has at most  $w$  unitaries and  $n$  noun wires total, (c) each unitary in the quantum circuit acts on at most  $d$  wires, (d) there are  $V$  different word embeddings given, (e) the number of gates in the ansatz for  $d$  wires of dimension  $N$  scales as  $A_g(N, d)$ , (f) the number of parameters in the ansatz for  $d$  wires of dimension  $N$  scales as  $A_p(N, d)$ , and (g) each parameter for the ansatz is stored with  $P$  bits of precision.

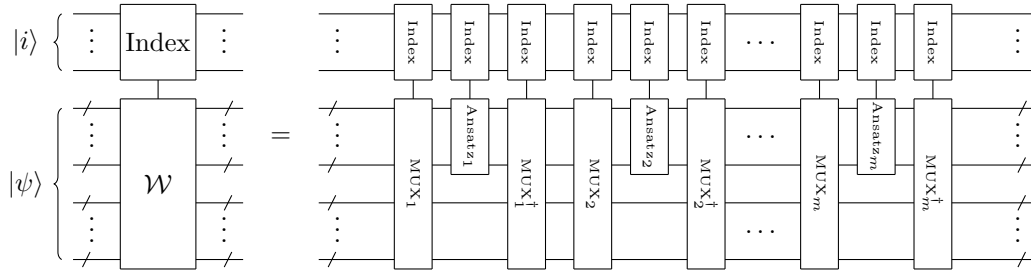
To solve this task, we can straightforwardly apply the quantum  $k$ -nearest neighbors algorithm discussed above, with  $\mathcal{V}(|0\rangle\langle 0|) = \rho_0$  and  $\mathcal{W}(|i\rangle\langle i| \otimes |0\rangle\langle 0|) = |i\rangle\langle i| \otimes \rho_i$ . Clearly,  $\mathcal{V}$  is given by the quantum circuit preparing  $\rho_0$ , however for  $\mathcal{W}$  we need a significantly more complicated circuit. This is explicitly constructed in Appendix C, and requires the following QRAM oracles:

- $\text{Param}_{ij}|k\rangle|0\rangle = |k\rangle|\phi_{ijk}\rangle$  where  $\phi_{ijk}$  is a  $P$ -bit fixed-precision binary encoding of the  $j$ th parameter for the ansatz of the  $i$ th unitary in the circuit preparing  $\rho_k$ ,
- $\text{Width}_i|k\rangle|0\rangle = |k\rangle|W_{ik}\rangle$  where  $W_{ik}$  is a binary encoding of the width of the  $i$ th unitary in the circuit preparing  $\rho_k$ ,
- $\text{Index}_{ij}|k\rangle|0\rangle = |k\rangle|I_{ijk}\rangle$  where  $I_{ijk}$  is an encoding of the position of the  $j$ th argument to unitary  $i$  in the circuit preparing  $\rho_k$ . Note that the encoding is not a mapping from zero-indexed position to binary, but it is of size  $O(\log(w))$  – see Appendix C.

Each of these oracles can be implemented using a series of calls (one for each bit) to bucket-brigade QRAMs of size  $k$ . Therefore,  $O(kw(PA_p(N, d) + \log(d) + d \log(w)))$  bits of QRAM is required in total, each taking time  $O(\log(k))$  to query.

If we assume that  $A_p(N, d) = O(\text{poly}(\log(N), d))$  (this is true for any of the widely-used ansatze defined by Sim et al [47], for example), that  $w = O(\text{poly}(n))$  (which is justified by Heap's law), and that  $P$  and  $d$  are fixed, then this is  $O(k \text{poly}(n, \log(N)))$  bits total. By comparison, adapting the strategy employed in [54] to QDisCoCirc would require at most  $O(kN^n)$  bits total, so our approach provides a significant saving.

The oracle for  $\mathcal{W}$  is constructed by alternating layers of multiplexers, which bring the qubits for each unitary to the top of the circuit, and parameterized ansatze to implement each unitary. Each of these load parameters from QRAM to specify their behavior. The overall construction is as follows:



Using this, we derive the following result:

**Theorem 7.** *Given a text  $T_0$  and a set of texts  $T_i$  for  $k \geq i > 0$ , we can find  $i$  that maximizes  $\text{tr}(\rho_0 \rho_i)$ , with error at most  $\epsilon$ , in time*

$$\tilde{O}\left(\frac{\sqrt{k}}{\epsilon}(wd \log(k)(n \log(N) + PA_g(N, d)) + \log^2(N))\right)$$

on a quantum computer with a constant success probability, where  $\rho_i$  is the density matrix corresponding to the QDisCoCirc quantum circuit for  $T_i$ .

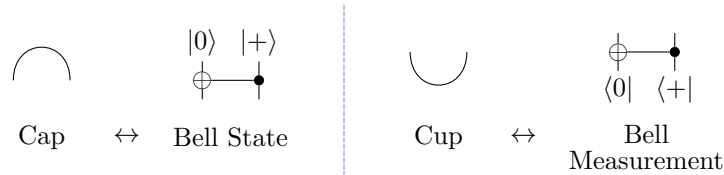
*Proof.* We apply the quantum  $k$ -nearest neighbor algorithm of Basheer et al [5] with  $\mathcal{V}$  and  $\mathcal{W}$  defined above. These can be implemented with  $O(A_g(N, d))$  and  $O(wd \log(k)(n \log(N) + PA_g(N, d)))$  gates respectively, see Appendix C for details. We require controlled versions of these, but this only increases the gate count by a constant factor.  $\square$

Comparing this to the naive swap test-based algorithm for solving the question-answering task defined in Section 4, it is clear that this method achieves a quadratic speedup both in  $k$  and  $\epsilon$ , up to extra logarithmic factors needed to construct the oracles. Comparing against a classical algorithm, note that for any given  $\epsilon$ , this algorithm is better than any classical algorithm, which clearly must be at least  $O(k)$ , and so for large values of  $k$  we would expect to see a speedup as compared to any classical algorithm to solve the same task. However, as discussed in Section 3, the classical algorithm scales much better in terms of  $\epsilon$ , namely  $O(\log(\frac{1}{\epsilon}))$  vs  $O(\epsilon^{-1})$ , and so the same concerns about the dependence on  $N$  and  $n$  apply here. Additionally for this algorithm, for a given dataset, it is necessary to analyze the scaling of  $\epsilon$  in terms of  $k$  – for instance, if  $\epsilon = o(k^{-\frac{1}{2}})$  is required, then there would be no quadratic speedup in  $k$  compared to the classical algorithm.

## 6 Other Model-Native Tasks

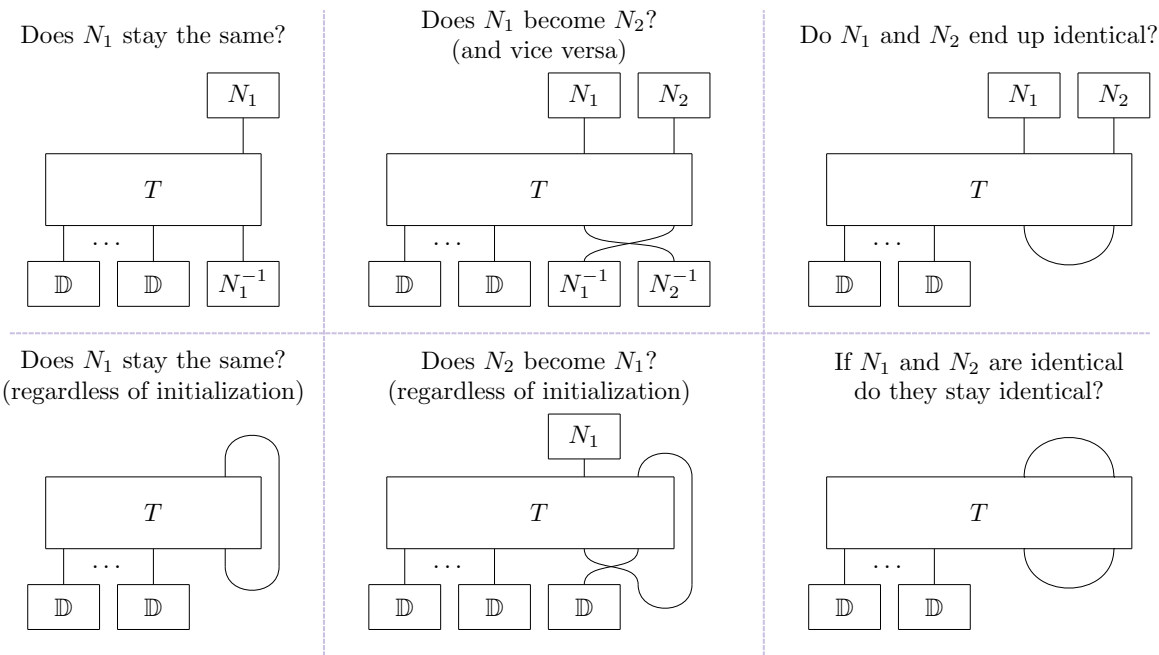
In the previous sections, we have considered only the fundamental primitive of text-similarity, and its local version that we call question-answering. However, we can also propose some simple model-native text processing tasks regarding ‘character development’. All tasks we introduce here have a natural representation as text circuits in the DisCoCirc framework.

To do so, we introduce two new generators: the cap and cup. The cap semantically represents a joint state of two nouns which are unspecified but identical. Likewise, the cup is the corresponding effect that tests for this situation.



Their translations into QDisCoCirc are naturally defined as the Bell state and measurement respectively. These generators are both self-inverse, and obey certain relations along with the identity and swap generators to ensure that they are well-behaved [15] - in particular, to maintain the property that only connectivity of the diagram matters.

We can use these to construct several tasks. For example, measuring a ‘character arc’ — comparing the similarity of a noun state wire before and after it has been acted on by a text, either with itself or between a pair of nouns. We can also form a diagrammatic trace using caps, cups, and identity wires, which we can use to measure a similar quantity in a way that is independent of the initial noun state (i.e. how much *any* noun changes if it was taking that particular role in the text). Let  $T$  be a text and  $N_1$  and  $N_2$  be two nouns present in it, then we can form the following text circuits:



Further, we can have combinations of the above, by applying noun effects and traces in the same circuit. This is to be understood as testing for properties in conjunction. From this, we see the motivation

for including cups, caps, and inverses in our framework – it allows us to represent more intuitively some natural classes of tasks as a single text circuit. Additionally, while the results in Sections 4 and 5 are tailored to the sentence-similarity and question-answering tasks, we would expect that similar theorems hold for all the tasks defined here as well.

## 7 Discussion and Future Work

In this paper we focused on the DisCoCirc framework because of its explainability and interpretability in compositional terms. For the primitive task of question-answering within the QDisCoCirc model, we showed BQP-hardness. Further, we adapted known quantum algorithms to achieve for QDisCoCirc a Grover-like quadratic speedup over any classical algorithm for tasks such as text similarity, under certain technical conditions.

Recalling that the way in which we constructed QDisCoCirc is by applying a map from text circuits to quantum circuits, other variants of this mapping could be explored in the future. For example, there are families of quantum circuits that are *classically efficiently simulable*, such as those constructed from nearest-neighbor match-gates [27], which have been used successfully to numerically optimize large quantum circuits [38] and could be used as ‘warm start’ initializations for QDisCoCirc models. Viewing quantum circuits as special cases of tensor networks, we can also consider models based on tensor networks (although they would be costly to compute). This would constitute compositional tensor-network machine learning, which is an emerging subfield in machine learning [26, 48].

Overall, our approach can be viewed as providing sparse structural priors to NLP models, potentially aiding generalization performance or allowing more efficient training. Such an approach could also provide a structure that may guide how mechanistic interpretability methods can be applied [36]. Most importantly, we prioritized transparent model-building, with performance and broad applicability as a secondary target.

The results here are presented in a way compatible with the parser introduced by Liu, Shaikh, Rodatz, Yeung, and Coecke [35], which mechanically transforms a subset of English text into DisCoCirc text circuits. Given a different parser, the definition of the mapping from text circuits to quantum circuits might need modification. However, the results of Section 4 and 5 regarding quantum algorithms would still be applicable. These results concern fault-tolerant quantum computers (either via deep quantum circuits, or by the assumption of quantum random access memories).

Focusing on more near-term hardware, and given access to a parser and state-of-the-art near-term quantum processors [4], allows us to perform heuristic explorations based on the algorithm in Section 3.1. In a separate paper [20], we report on performing tasks such as question-answering within the noisy intermediate-scale quantum regime. We train the quantum word embeddings that encode the semantics of texts in-task, as was done in previous work with DisCoCat [37, 40]. However, it is also promising to experiment with training these in a task-agnostic way.

Furthermore, QDisCoCirc allows us to *put to the test* the principle of *compositionality* with respect to training, i.e. to quantify to what extent a compositional model aids generalization to larger instances after being trained on instances up to a certain size. This setup is motivated by the possibility that a compositional approach to quantum machine learning with parameterized circuits bypasses the fundamental problem posed by barren plateaus [42]. Our results in [20] are very promising in this respect.

**Acknowledgments** We thank Matty Hoban for discussions on complexity theory, Harry Buhman, Richie Yeung, Luuk Coopmans and Ilyas Khan for comments on the manuscript, the QDisCoCirc team

(Saskia Bruhn, Tiffany Duneau, Gabriel Matos, Anna Pearson, and Katerina Saiti) for helpful discussions refining the model, and the parser-pipeline team (Jonathon Liu, Razin Shaikh, and Benjamin Rodatz) for answering all of our questions.

## References

- [1] S. Abramsky & B. Coecke (2004): *A categorical semantics of quantum protocols*. In: *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science (LICS)*, pp. 415–425, doi:10.48550/arXiv.quant-ph/0402130. arXiv:quant-ph/0402130.
- [2] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven & John M. Martinis (2019): *Quantum Supremacy Using a Programmable Superconducting Processor*. *Nature* 574(7779), pp. 505–510, doi:10.1038/s41586-019-1666-5.
- [3] R Harald Baayen (2001): *Word frequency distributions*. 18, Springer Science & Business Media, doi:10.1007/978-94-010-0844-0.
- [4] Charles H. Baldwin, Karl Mayer, Natalie C. Brown, Ciarán Ryan-Anderson & David Hayes (2022): *Re-examining the quantum volume test: Ideal distributions, compiler optimizations, confidence intervals, and scalable resource estimations*. *Quantum* 6, p. 707, doi:10.22331/q-2022-05-09-707.
- [5] Afrad Basheer, A. Afham & Sandeep K. Goyal (2021): *Quantum k-nearest neighbors algorithm*, doi:10.48550/arXiv.2003.09187. arXiv:2003.09187.
- [6] Marcello Benedetti, Erika Lloyd, Stefan Sack & Mattia Fiorentini (2019): *Parameterized quantum circuits as machine learning models*. *Quantum Science and Technology* 4(4), p. 043001, doi:10.1088/2058-9565/ab4eb5.
- [7] Ethan Bernstein & Umesh Vazirani (1997): *Quantum Complexity Theory*. *SIAM Journal on Computing* 26(5), pp. 1411–1473, doi:10.1145/167088.167097.
- [8] Adam Bouland & Tudor Giurgica-Tiron (2021): *Efficient Universal Quantum Compilation: An Inverse-free Solovay-Kitaev Algorithm*, doi:10.48550/arXiv.2112.02040. arXiv:2112.02040.
- [9] Trenton Bricken, Adly Templeton, Joshua Batson, Brian Chen, Adam Jermyn, Tom Conerly, Nick Turner, Cem Anil, Carson Denison, Amanda Askell, Robert Lasenby, Yifan Wu, Shauna Kravec, Nicholas Schiefer, Tim Maxwell, Nicholas Joseph, Zac Hatfield-Dodds, Alex Tamkin, Karina Nguyen, Brayden McLean, Josiah E Burke, Tristan Hume, Shan Carter, Tom Henighan & Christopher Olah (2023): *Towards Monosemanticity: Decomposing Language Models With Dictionary Learning*. *Transformer Circuits Thread*. <https://transformer-circuits.pub/2023/monosemantic-features/index.html>.
- [10] H. Buhrman, R. Cleve, J. Watrous & R. De Wolf (2001): *Quantum fingerprinting*. *Physical Review Letters* 87(16), p. 167902, doi:10.1103/PhysRevLett.87.167902.
- [11] S. Clark, B. Coecke, E. Grefenstette, S. Pulman & M. Sadrzadeh (2014): *A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure*. *Malaysian Journal of Mathematical Sciences* 8, pp. 15–25, doi:10.48550/arXiv.1305.0556. ArXiv:1305.0556.



- [12] B. Coecke (2019): *The mathematics of text structure*, doi:10.48550/arXiv.1904.03478. ArXiv:1904.03478.
- [13] B. Coecke (2021): *Compositionality as we see it, everywhere around us*. arXiv preprint arXiv:2110.05327, doi:10.48550/arXiv.2110.05327.
- [14] B. Coecke, G. de Felice, K. Meichanetzidis & A. Toumi (2020): *Foundations for Near-Term Quantum Natural Language Processing*, doi:10.48550/arXiv.2012.03755. ArXiv preprint arXiv:2012.03755.
- [15] B. Coecke & A. Kissinger (2017): *Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press, doi:10.1017/9781316219317.
- [16] B. Coecke, M. Sadrzadeh & S. Clark (2010): *Mathematical foundations for a compositional distributional model of meaning*. In J. van Benthem, M. Moortgat & W. Buszkowski, editors: *A Festschrift for Jim Lambek, Linguistic Analysis* 36, pp. 345–384, doi:10.48550/arXiv.1003.4394. Arxiv:1003.4394.
- [17] Bob Coecke, Fabrizio Genovese, Stefano Gogioso, Dan Marsden & Robin Piedeleu (2018): *Uniqueness of Composition in Quantum Theory and Linguistics*. *Electronic Proceedings in Theoretical Computer Science* 266, pp. 249–257, doi:10.4204/eptcs.266.17.
- [18] Matthew DeCross, Reza Haghshenas, Minzhao Liu, Yuri Alexeev, Charles H. Baldwin, John P. Bartolotta, Matthew Bohn, Eli Chertkov, Jonhas Colina, Davide DelVento, Joan M. Dreiling, Cameron Foltz, John P. Gaebler, Thomas M. Gatterman, Christopher N. Gilbreth, Johnnie Gray, Dan Gresh, Nathan Hewitt, Ross B. Hutson, Jacob Johansen, Dominic Lucchetti, Danylo Lykov, Ivaylo S. Madjarov, Karl Mayer, Michael Mills, Pradeep Niroula, Enrico Rinaldi, Peter E. Siegfried, Bruce G. Tiemann, Curtis Volin, James Walker, Ruslan Shaydulín, Marco Pistoia, Steven. A. Moses, David Hayes, Brian Neyenhuis, Russell P. Stutz & Michael Foss-Feig (2024): *The computational power of random quantum circuits in arbitrary geometries*, doi:10.48550/arXiv.2406.02501. arXiv:2406.02501.
- [19] Tiffany Duneau (2021): *Solving Logical Puzzles in DisCoCirc*. *Journal of Cognitive Science* 22(3), pp. 355 – 389. Available at <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85125423576&partnerID=40&md5=4c6db945056372334795ff799ea0f56f>. Type: Article.
- [20] Tiffany Duneau, Saskia Bruhn, Gabriel Matos, Tuomas Laakkonen, Katerina Saiti, Anna Pearson, Konstantinos Meichanetzidis & Bob Coecke: *Scalable and interpretable quantum natural language processing: an implementation on trapped ions*. In preparation.
- [21] Vedran Dunjko & Hans J Briegel (2018): *Machine learning & artificial intelligence in the quantum domain: a review of recent progress*. *Reports on Progress in Physics* 81(7), doi:10.1088/1361-6633/aab406. Available at <https://dx.doi.org/10.1088/1361-6633/aab406>.
- [22] Christoph Durr & Peter Hoyer (1999): *A Quantum Algorithm for Finding the Minimum*, doi:10.48550/arXiv.quant-ph/9607014. arXiv:quant-ph/9607014.
- [23] V. Giovannetti, S. Lloyd & L. Maccone (2008): *Quantum random access memory*. *Physical review letters* 100(16), p. 160501, doi:10.1103/PhysRevLett.100.160501.
- [24] E. Grefenstette & M. Sadrzadeh (2011): *Experimental Support for a Categorical Compositional Distributional Model of Meaning*. In: *The 2014 Conference on Empirical Methods on Natural Language Processing*., pp. 1394–1404, doi:10.48550/arXiv.1106.4058. ArXiv:1106.4058.
- [25] Carys Harvey, Richie Yeung & Konstantinos Meichanetzidis (2023): *Sequence Processing with Quantum Tensor Networks*, doi:10.48550/arXiv.2308.07865. arXiv:2308.07865.
- [26] William Huggins, Piyush Patil, Bradley Mitchell, K Birgitta Whaley & E Miles Stoudenmire (2019): *Towards quantum machine learning with tensor networks*. *Quantum Science and Technology* 4(2), p. 024001, doi:10.1088/2058-9565/aaea94.
- [27] Richard Jozsa & Akimasa Miyake (2008): *Matchgates and classical simulation of quantum circuits*. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 464(2100), p. 3089–3106, doi:10.1098/rspa.2008.0189.
- [28] D. Kartsaklis, I. Fan, R. Yeung, A. Pearson, R. Lorenz, A. Toumi, G. de Felice, K. Meichanetzidis, S. Clark & B. Coecke (2021): *lambeq: An Efficient High-Level Python Library for Quantum NLP*. arXiv preprint arXiv:2110.04236, doi:10.48550/arXiv.2110.04236.

- [29] D. Kartsaklis & M. Sadrzadeh (2013): *Prior disambiguation of word tensors for constructing Sentence vectors*. In: *The 2013 Conference on Empirical Methods on Natural Language Processing.*, ACL, pp. 1590–1601. <https://aclanthology.org/D13-1166/>.
- [30] J. Lambek (1958): *The mathematics of sentence structure*. *American Mathematics Monthly* 65, doi:10.1080/00029890.1958.11989160.
- [31] J. Lambek (1999): *Type grammar revisited*. *Logical Aspects of Computational Linguistics* 1582, doi:10.1007/3-540-48975-4\_1.
- [32] J. Lambek (2008): *From Word to Sentence*. Polimetrica, Milan. ISBN: 8876991174. <https://www.math.mcgill.ca/barr/lambek/pdf/2008lambek.pdf>.
- [33] Yeong-Cherng Liang, Yu-Hao Yeh, Paulo E M F Mendonça, Run Yan Teh, Margaret D Reid & Peter D Drummond (2019): *Quantum fidelity measures for mixed states*. *Reports on Progress in Physics* 82(7), p. 076001, doi:10.1088/1361-6633/ab1ca4.
- [34] Henry Lin & Max Tegmark (2017): *Critical Behavior in Physics and Probabilistic Formal Languages*. *Entropy* 19(7), p. 299, doi:10.3390/e19070299.
- [35] Jonathon Liu, Razin A. Shaikh, Benjamin Rodatz, Richie Yeung & Bob Coecke (2023): *A Pipeline For Discourse Circuits From CCG*, doi:10.48550/arXiv.2311.17892. arXiv:2311.17892.
- [36] Ziming Liu, Eric Gan & Max Tegmark (2023): *Seeing is Believing: Brain-Inspired Modular Training for Mechanistic Interpretability*, doi:10.48550/arXiv.2305.08746. arXiv:2305.08746.
- [37] Robin Lorenz, Anna Pearson, Konstantinos Meichanetzidis, Dimitri Kartsaklis & Bob Coecke (2023): *QNL in Practice: Running Compositional Models of Meaning on a Quantum Computer*. *Journal of Artificial Intelligence Research* 76, p. 1305–1342, doi:10.1613/jair.1.14329.
- [38] Gabriel Matos, Chris N. Self, Zlatko Papić, Konstantinos Meichanetzidis & Henrik Dreyer (2023): *Characterization of variational quantum algorithms using free fermions*. *Quantum* 7, p. 966, doi:10.22331/q-2023-03-30-966.
- [39] Konstantinos Meichanetzidis, Stefano Gogioso, Giovanni de Felice, Nicolò Chiappori, Alexis Toumi & Bob Coecke (2021): *Quantum Natural Language Processing on Near-Term Quantum Computers*. *Electronic Proceedings in Theoretical Computer Science* 340, p. 213–229, doi:10.4204/eptcs.340.11.
- [40] Konstantinos Meichanetzidis, Alexis Toumi, Giovanni de Felice & Bob Coecke (2023): *Grammar-aware sentence classification on quantum computers*. *Quantum Machine Intelligence* 5(1), doi:10.1007/s42484-023-00097-1.
- [41] Kosuke Mitarai, Masahiro Kitagawa & Keisuke Fujii (2019): *Quantum analog-digital conversion*. *Phys. Rev. A* 99, p. 012301, doi:10.1103/PhysRevA.99.012301. Available at <https://link.aps.org/doi/10.1103/PhysRevA.99.012301>.
- [42] Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, Alexander F. Kemper, Carlos Ortiz Marrero, Martin Larocca & M. Cerezo (2023): *A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits*, doi:10.48550/arXiv.2309.09342. arXiv:2309.09342.
- [43] Benjamin Rodatz, Razin A. Shaikh & Lia Yeh (2021): *Conversational Negation using Worldly Context in Compositional Distributional Semantics*, doi:10.48550/arXiv.2105.05748. arXiv:2105.05748.
- [44] M. Sadrzadeh, S. Clark & B. Coecke (2013): *The Frobenius anatomy of word meanings I: subject and object relative pronouns*. *Journal of Logic and Computation* 23, pp. 1293–1317, doi:10.1093/logcom/ext044. ArXiv:1404.5278.
- [45] Yukie Sano, Hideki Takayasu & Misako Takayasu (2012): *Zipf’s Law and Heaps’ Law Can Predict the Size of Potential Words*. *Progress of Theoretical Physics Supplement* 194, p. 202–209, doi:10.1143/ptps.194.202.
- [46] Maria Schuld, Ilya Sinayskiy & Francesco Petruccione (2014): *An introduction to quantum machine learning*. *Contemporary Physics* 56(2), p. 172–185, doi:10.1080/00107514.2014.964942.

- [47] Sukin Sim, Peter D. Johnson & Alán Aspuru-Guzik (2019): *Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms*. *Advanced Quantum Technologies* 2(12), doi:10.1002/qute.201900070.
- [48] Andrei Tomut, Saeed S. Jahromi, Sukhbinder Singh, Faysal Ishtiaq, Cesar Muñoz, Prabdeep Singh Bajaj, Ali Elborady, Gianni del Bimbo, Mehrazin Alizadeh, David Montero, Pablo Martin-Ramiro, Muhammad Ibrahim, Oussama Tahiri Alaoui, John Malcolm, Samuel Mugel & Roman Orus (2024): *CompactifAI: Extreme Compression of Large Language Models using Quantum-Inspired Tensor Networks*, doi:10.48550/arXiv.2401.14109. arXiv:2401.14109.
- [49] V. Wang-Mascianica, J. Liu & B. Coecke (2023): *Distilling Text into Circuits*. arXiv preprint arXiv:2301.10595, doi:10.48550/arXiv.2301.10595.
- [50] Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin & Tomas Mikolov (2015): *Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks*, doi:10.48550/arXiv.1502.05698. arXiv:1502.05698.
- [51] Dominic Widdows, Willie Aboumrad, Dohun Kim, Sayonee Ray & Jonathan Mei (2024): *Natural Language, AI, and Quantum Computing in 2024: Research Ingredients and Directions in QNLP*. arXiv preprint arXiv:2403.19758, doi:10.48550/arXiv.2403.19758.
- [52] Dominic Widdows, Aaranya Alexander, Daiwei Zhu, Chase Zimmerman & Arunava Majumder (2024): *Near-term advances in quantum natural language processing*. *Annals of Mathematics and Artificial Intelligence*, doi:10.1007/s10472-024-09940-y.
- [53] N. Wiebe, A. Kapoor & K. M. Svore (2015): *Quantum nearest-neighbour algorithms for machine learning*. *Quantum Information and Computation* 15, pp. 318–358, doi:10.26421/QIC15.3-4-7.
- [54] William Zeng & Bob Coecke (2016): *Quantum Algorithms for Compositional Natural Language Processing*. *Electronic Proceedings in Theoretical Computer Science* 221, p. 67–75, doi:10.4204/eptcs.221.8.
- [55] Jun Zhang, Jiri Vala, Shankar Sastry & K. Birgitta Whaley (2003): *Exact Two-Qubit Universal Quantum Circuit*. *Physical Review Letters* 91(2), doi:10.1103/physrevlett.91.027903.

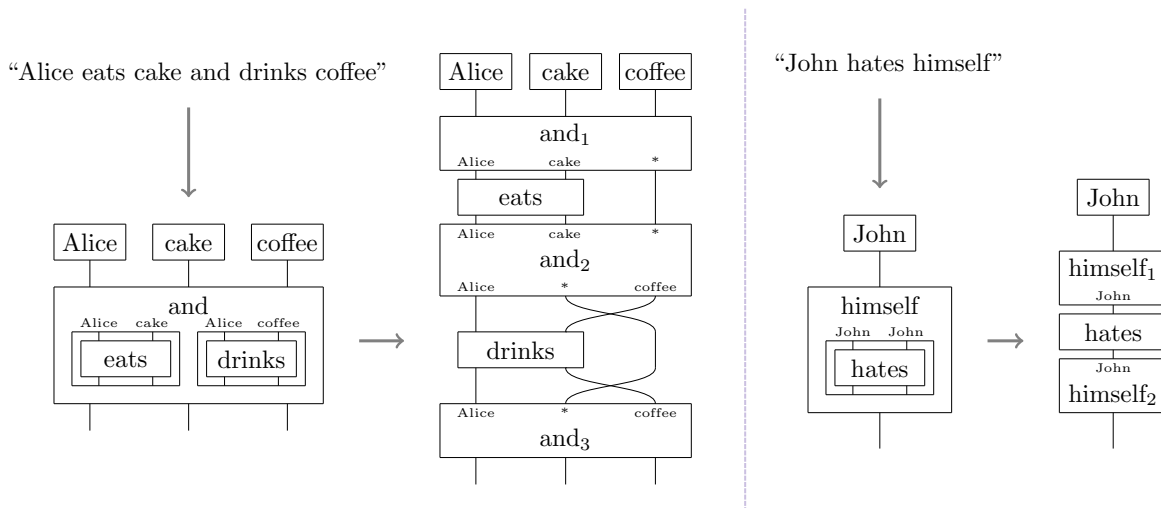
## A Decomposing Frames

Our semantic functor maps states, effects, and boxes to pure quantum operations trivially. However, frames, which are mapped to quantum supermaps, in general, need some special treatment. Some frames in text circuits may contain more wires for one of their arguments than are input and output to the frame – this arises in many cases, for instance in reflexive structures. We cannot represent these extra wires as auxiliary qubits because these would then be subsequently discarded, making the state mixed. In general, quantum supermaps can be written as pairs of unitaries conjugating their arguments, along with a side channel consisting of some auxiliary qubits. However, this conflicts with our no-ancilla requirement.

Therefore, we will instead *constrain* our representation of frames as follows:

- We lay the frame out vertically, with unitaries before the first argument, between every argument, and after the last argument.
- For every wire in each argument, we map it to one of the wires of the unitary: if that wire corresponds to an unused noun wire that is input to the frame, we map it to that wire. Otherwise, we map it to any unused noun wire. If there are none, then we delete the wire.
- After this mapping, we substitute the argument between the unitaries of the frame, and any unused wires are connected directly between the unitaries of the frame to serve as side-channels.

In order to resolve which wires correspond to which nouns (and thus where to assign each wire of the frame arguments), we rely on coindexing information provided by the parser, which within sentences is provided by finding the grammatical 'head' of each phrase, and between sentences is provided by a coreference tool [35]. To minimize the number of wires that are deleted, we perform this mapping for each argument independently, not respecting the wire assignments in previous arguments. This may create boxes that are ungrammatical, but we assign them unitaries in the same way as other boxes, so this is not an issue.



This illustrates two examples of the sandwich construction. In each, the wires inside the unitaries making up the frame are labeled with which nouns they are representing, and wires that are being used as side channels are labeled with '\*'. In the first example, we see that swaps may be introduced to map wires to the correct locations. This is an artifact of the notation and does not appear in the compiled quantum circuits. In the second example we see a wire being deleted since there are no unused wires to assign it to, and in the process this creates an ungrammatical box for 'hates', since it is a transitive verb.

## B Proofs of Theorems

**Theorem 2.** QDISCO-CIRC-QA can be solved on a quantum computer in time

$$O\left(\frac{k \log\left(\frac{k}{\delta}\right) |V|_w (|T| + \max_i |Q_i|)}{\varepsilon^2}\right)$$

with failure probability  $\delta$ , where  $|V|_w$  is the maximum size of any word embedding in  $V$ .

*Proof.* To calculate each

$$p_j = \text{tr}(\rho_T(\rho_{Q_j} \otimes I))$$

we produce a circuit using the swap test where the probability of measuring a zero on the first qubit is  $\frac{1}{2} + \frac{p_j}{2}$ . Let  $\mu_{j,n}$  be the estimate of  $p_j$  after  $n$  shots. By Hoeffding's inequality,

$$P(|\mu_{j,n} - p_j| \geq \varepsilon') \leq 2 \exp\left(-\frac{n\varepsilon'^2}{2}\right)$$

thus if we take  $n = \frac{2}{\varepsilon'^2} \log\left(\frac{2k}{\delta}\right)$ , then we have

$$P\left(\bigwedge_{j=1}^k |\mu_{j,n} - p_j| \leq \varepsilon'\right) \geq \left(1 - \frac{\delta}{k}\right)^k \geq 1 - \delta$$

by Bernoulli's inequality. Note that if  $|\mu_{j,n} - p_j| \leq \varepsilon'$  for all  $j$ , then  $|\max_i \mu_{n,i} - \max_i p_i| \leq \varepsilon'$ , and hence if  $j = \arg \max_i \mu_{n,i}$  then  $|p_j - \max_i p_i| \leq 2\varepsilon'$ . Therefore, if  $\varepsilon' = \frac{\varepsilon}{2}$ , then this  $j$  correctly solves the QDISCO-CIRC-QA instance with probability  $1 - \delta$ . For each  $j$ , each shot of the swap test circuit takes  $O((|T| + |Q_j|)|V|_w)$  time since  $U_T$  and  $U_{Q_j}$  contain  $O(|T||V|_w)$  and  $O(|Q_j||V|_w)$  gates respectively, and the width of  $U_{Q_j}$  must be at most  $2|Q_j||V|_w$ . Therefore, we have an overall time complexity of:

$$O\left(\sum_i (|T| + |Q_i|)|V|_w n\right) = O\left(\frac{k \log\left(\frac{k}{\delta}\right) |V|_w (|T| + \max_i |Q_i|)}{\varepsilon^2}\right)$$

□

**Definition 3 (APPROX-QCIRCUIT-PROB).** Suppose we are given a quantum circuit  $C$  on  $n$  qubits with  $m$  2-local gates, where  $C$  is taken from a uniform family such that  $m$  is polynomial in  $n$ , along with the promise that either

- measuring the first qubit  $C|0\rangle$  has probability at least  $\frac{2}{3}$ , or
- measuring the first qubit  $C|0\rangle$  has probability at most  $\frac{1}{3}$ .

Then the APPROX-QCIRCUIT-PROB problem is to distinguish between these two cases. This is BQP-hard.

**Theorem 3.** Suppose that a set of word embeddings  $V$  satisfies the following:

1. The operations of  $V$  use one qubit for each input wire,
2.  $V$  contains arbitrarily many proper nouns,
3.  $V$  contains at least two adjectives that generate a dense subset of  $SU(2)$ ,

4.  $V$  contains at least one transitive verb that is entangling

then for any fixed  $\varepsilon < \frac{1}{7}$ , QDISCO-CIRC-QA is BQP-hard.

*Proof.* Let  $U_1$  and  $U_2$  be the adjectives defined by condition three, and  $U_3$  be the transitive verb identified by condition four. Then, by the inverse-free Solovay-Kitaev algorithm [8], any single-qubit unitary  $U$  can be approximated to precision  $\varepsilon'$  by a sequence of  $U_1$  and  $U_2$  operations of length  $O(\log^{8.62}(\frac{1}{\varepsilon'}))$ . From a result of Zhang et al [55], any entangling gate combined with arbitrary single-qubit gates can be used to implement any two-qubit gate exactly in a constant number of gates (where the constant depends only on the definition of the entangling gate). Hence, any two-qubit gate can also be approximated arbitrarily well by  $O(\log^{8.62}(\frac{1}{\varepsilon'}))$  applications of  $U_1$ ,  $U_2$ , and  $U_3$ .

Now we can reduce from the APPROX-QCIRCUIT-PROB problem to QDISCO-CIRC-QA, by constructing a QDisCoCirc circuit corresponding to the given circuit  $C$ . First, assign each qubit a proper noun  $N_i$  with corresponding unitary  $U_{N_i}$  in  $V$ . To each noun apply the unitary  $U_{N_i}^\dagger$ . Each  $U_{N_i}^\dagger$  is approximated as defined above. Then, for each gate in  $C$ , we approximate it as discussed above and apply it to the corresponding wires. This yields a QDisCoCirc circuit  $U_T$  which approximates the original circuit  $C$ .

From this, we can generate  $T$  as follows: for each gate in the circuit, we generate a sentence. If the gate acts on a single qubit  $i$ , we write “ $\{N_i\}$  is  $\{U_k\}$ .” where  $U_k$  is the adjective corresponding to the gate. If the gate acts on two qubits  $i$  and  $j$ , we write “ $\{N_i\}$   $\{U_3\}$   $\{N_j\}$ .” To produce  $Q_1$  and  $Q_2$ , we first assemble quantum circuits  $U_{Q_1} = U_{N_1}^\dagger U_{N_1}$  and  $U_{Q_2} = U_X U_{N_1}^\dagger U_{N_1}$  where  $U_X$  is an approximation of the X gate, and  $U_{N_1}^\dagger$  is also approximated as above. Then  $Q_1$  and  $Q_2$  can be generated in the same way as  $T$ .

Let  $\varepsilon' \geq \frac{1}{42}$  be such that  $\varepsilon' + \varepsilon < \frac{1}{6}$ , and let  $P_U(i)$  be the probability of measuring  $|i\rangle$  on the first qubit of  $U|0\rangle$ . Suppose we solve the QDISCO-CIRC-QA instance associated with  $V$ ,  $T$ ,  $\{Q_1, Q_2\}$  to precision  $\varepsilon$ . Note that since  $U_{Q_1}|0\rangle \approx |0\rangle$ ,  $\text{tr}(\rho_T(\rho_{Q_1} \otimes I)) \approx P_{U_T}(0)$ . Likewise, since  $U_{Q_2}|0\rangle \approx |1\rangle$ ,  $\text{tr}(\rho_T(\rho_{Q_2} \otimes I)) \approx P_{U_T}(1)$ . As  $U_T \approx C$ , if we find  $j = 1$ , then we can conclude that  $P_C(0) > P_C(1) = 1 - P_C(0) \implies P_C(0) > \frac{1}{2}$  up to some error  $\varepsilon + \varepsilon' < \frac{1}{6}$ . However, in the definition of APPROX-QCIRCUIT-PROB, we are guaranteed that  $P_C(0) > \frac{2}{3}$  or  $P_C(0) < \frac{1}{3}$ , which are separated from  $\frac{1}{2}$  by  $\frac{1}{6}$ , so by checking whether  $j = 1$  or  $j = 2$ , we can distinguish between these cases. Therefore, QDISCO-CIRC-QA is BQP-hard.  $\square$

**Theorem 4.** *Given a set of word embeddings  $V$ , suppose that operations in  $V$  are independent Haar-random unitaries. Then conditions three and four of Theorem 3 are almost surely satisfied for all word embeddings containing at least two adjectives and one transitive verb.*

*Proof.* Let  $U_1$  and  $U_2$  be the adjectives defined by condition three of Theorem 1, and  $U_3$  be the transitive verb identified by condition four. Consider  $U_3$  as a point in  $SU(4)$ . If  $U_3$  is not entangling it is either separable or locally equivalent to the SWAP gate. The spaces formed in both of these cases are isomorphic to  $SU(2) \times SU(2)$  which is a 9-dimensional subspace of the 15-dimensional  $SU(4)$ . Hence condition four is satisfied almost everywhere in  $SU(4)$ . Therefore, for Haar random  $U_3$  it is almost surely satisfied.

Consider arbitrary  $U_1$  and  $U_2$ . If  $U_1$  and  $U_2$  are non-commuting elements of infinite order in  $SU(2)$ , then they generate a dense subset of  $SU(2)$ . This is because the only closed positive-dimension Lie subgroup of  $SU(2)$  is itself, so it must be equal to the closure of the group generated by  $U_1$  and  $U_2$ . Consider  $U_1$  and  $U_2$  as a point in  $SU(2) \times SU(2)$ . Since elements of  $SU(2)$  can also be characterized as rotations of the Bloch sphere, then commuting elements correspond to rotations around the same axis. Thus the subspace of commuting  $U_1$  and  $U_2$  is isomorphic to  $SU(2) \times S$ . Likewise, elements of finite order in  $SU(2)$  correspond to rotations about any axis by a rational multiple of  $\pi$ , and the subspace is isomorphic to  $S^2 \times \mathbb{Q}$ , which has measure zero. Thus almost everywhere in  $SU(2) \times SU(2)$ ,  $U_1$  and  $U_2$  generate a dense subset of  $SU(2)$ . Hence, for Haar random  $U_1$  and  $U_2$ , this is almost surely the case.  $\square$

**Theorem 5.** *Given a specific set of word embeddings  $V$  it is possible to check numerically that the conditions of Theorem 3 are satisfied.*

*Proof.* Let  $U_1$  and  $U_2$  be the adjectives defined by condition three of Theorem 3, and  $U_3$  be the transitive verb identified by condition four. Consider the state  $|U_3\rangle$  obtained via the Choi-Jamiołkowski isomorphism. If we compute the reduced density matrix  $\hat{\rho} = \text{tr}_2(|U_3\rangle\langle U_3|)$  (where the partial trace is taken over both qubits corresponding to the second qubit of  $U_3$ ), then this is pure (i.e has rank one) if and only if  $U_3$  is separable. Therefore, we can check that  $U_3$  is entangling by performing this test on both  $U_3$  and  $U_3 \cdot \text{SWAP}$  (to check that it is not locally equivalent to a SWAP).

To check that  $U_1$  and  $U_2$  generate a dense subset of  $SU(2)$ , first check that they do not commute. We wish to check that the closure of the subgroup generated by  $U_1$  and  $U_2$  is  $SU(2)$  itself. The only infinite closed nonabelian proper subgroup of  $SU(2)$  is isomorphic to  $S \rtimes_{\phi} C_2$  (where  $\phi_0(x) = x$ ,  $\phi_1(x) = -x$ ). In this subgroup for all  $x$  and  $y$ ,  $x^2$  and  $y^2$  commute, so to rule out this case it suffices to find a non-commuting pair. This can be done numerically by generating elements of the group by taking products and checking randomly until an example is found.

The only finite closed nonabelian subgroups of  $SU(2)$  are the binary dicyclic groups, and three groups of order 24, 48 (the single-qubit Cliffords), and 120, by the ADE classification. The binary dicyclic groups are subgroups of  $S \rtimes_{\phi} C_2$ , so we have already ruled these out. Therefore, it remains to check that the order of the group generated by  $U_1$  and  $U_2$  is more than 120. This can be done numerically by testing products of group elements until 121 unique elements are generated.  $\square$

**Theorem 6.** *There exists a polynomial  $f(m, n)$ , such that given an oracle to solve QDISCO-CIRC-QA instances with arbitrary word embeddings and text circuits drawn from  $\mathcal{D}_{f(m, n)}$ , we can perform arbitrary quantum computations with  $m$  2-local gates on  $n$  qubits with high probability in polynomial time. That is, QDISCO-CIRC-QA for worst-case word embeddings is average-case BQP-hard over texts.*

*Proof.* We will reduce from the problem APPROX-QCIRCUIT-PROB. We are given an arbitrary quantum circuit on  $n$  qubits with  $m$  two-qubit gates given by arbitrary unitaries, and we assume that any single-qubit gates is combined into an adjacent two-qubit gate – the only time this is not possible is when a qubit is unentangled, so it can be disregarded (if it is the measurement qubit, we can calculate the outcome immediately, and if it is not, it doesn't affect the outcome).

Suppose that we have sampled a large random text circuit from  $\mathcal{D}_k$  for some  $k \gg n$ . We wish to calculate how many boxes are required so that there exists a subsequence which matches the given circuit. By match, we mean that each gate has a corresponding box in the circuit with a unique label, operating on the same wires as the gate. Let  $N_i$  be the gap between gates  $i$  and the previous gate (or that start of the circuit) – that is, the number of boxes between them in the circuit that are not part of the subsequence. Because all boxes in the circuit are chosen in an identical random manner,  $P(N_i \geq q) = (1 - p_i)^q$  where  $p_i$  is the probability of any particular box being correct for gate  $i$ . Since inputs to boxes are chosen uniformly, we have that all the  $p_i = p$  are identical.

Therefore, we can bound the probability that the total sequence length required is at least some constant  $q$ :

$$\begin{aligned} P\left(\sum_i N_i \geq q\right) &\leq 1 - P\left(\bigwedge_i N_i < \frac{q}{m}\right) = 1 - \left(1 - P\left(N_i \geq \frac{q}{m}\right)\right)^m = 1 - \left(1 - (1 - p)^{\frac{q}{m}}\right)^m \\ &\leq 1 - \left(1 - m(1 - p)^{\frac{q}{m}}\right) = m(1 - p)^{\frac{q}{m}} \leq me^{-\frac{pq}{m}} \end{aligned}$$

The second line follows from Bernoulli's inequality. Thus letting  $q = \frac{m}{p}(\log(m) + 1)$ , we have that  $P(\sum_i N_i \geq q) \leq \frac{1}{e}$ . We have

$$p = P(A = 2) \cdot P(\text{hapax}) \cdot \frac{2}{k(k-1)}$$

where  $P(\text{hapax})$  is the probability that any particular box has a unique label, since  $\frac{2}{k(k-1)}$  is the probability of selecting any particular pair of wires uniformly, assuming that  $k > n$ . According to Heap's law, the number of distinct words in a text grows like  $\sqrt{w}$  where  $w$  is the total number of words. Furthermore from Zipf's law, a constant proportion of distinct words occur only once (*hapax legomena*), usually about half for large corpora. By assumption, this law holds exactly for circuits drawn from  $\mathcal{D}_k$ , so this implies that

$$P(\text{hapax}) \leq \frac{c_1 \sqrt{k^\gamma}}{k^\gamma} = \frac{c_1}{\sqrt{k^\gamma}} \quad \text{and} \quad p \geq \frac{c_2}{k^{2+\frac{\gamma}{2}}}$$

for some constants  $c_1, c_2 > 0$ , since we assume  $w \geq c_3 k^\gamma$  for some  $c_3$ . Therefore, we can set  $q \leq \frac{1}{c_2} m(\log(m) + 1) k^{2+\frac{\gamma}{2}}$ , but we also require  $c_3 k^\gamma \geq q$  to ensure the number of boxes required is less than the total in the text. This can be enforced by setting:

$$c_3 k^\gamma \geq \frac{1}{c_2} m(\log(m) + 1) k^{2+\frac{\gamma}{2}} \implies k \geq \left( \frac{1}{c_2 c_3} m(\log(m) + 1) \right)^{\frac{2}{\gamma-4}}$$

Finally, in order to make sure  $k \geq n$ , we can set  $k = f(m, n) = \max(n, \lceil (cm(\log(m) + 1))^{\frac{2}{\gamma-4}} \rceil)$  for some  $c > 0$ .

The protocol to solve APPROX-QCIRCUIT-PROB is as follows: we are given a circuit  $C$  on  $n$  qubits with  $m$  two-qubit gates (single-qubit gates are eliminated as before). Sample a text circuit from  $\mathcal{D}_k$  with  $k$  as given above. With probability at least  $1 - e^{-1}$ , there exists a subsequence of boxes in the circuit that matches  $C'$ . We construct a set of word embeddings for the circuit by setting each word that does not appear in the subsequence to be the identity, and every word that is part of the subsequence to be the unitary of the corresponding gate in the circuit. Each state is assigned to the  $|0\rangle$  quantum state. This defines the text circuit and embeddings for the context text  $T$ . Similarly, we can obtain the questions  $Q_1$  and  $Q_2$  by drawing from  $\mathcal{D}_1$ , setting all boxes to the identity, and assigning the state to  $|0\rangle$  and  $|1\rangle$  respectively.

In the same way as in Theorem 3, solving this instance of QDISCO-CIRC-QA with  $\varepsilon < \frac{1}{7}$  is sufficient to solve the corresponding APPROX-QCIRCUIT-PROB instance.  $f(m, n)$  as given above grows polynomially if  $\gamma > 4$ , and the construction of the word embeddings is done in polynomial time. Since this procedure has fixed positive success probability, it can be amplified to any high success probability with logarithmically many iterations.  $\square$

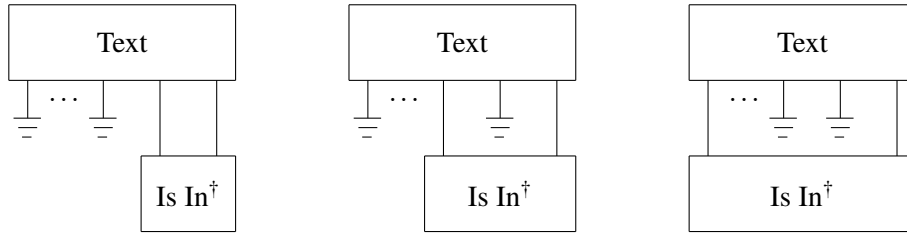
## C Oracle Construction

As a warmup, we will start by constructing oracles for a different task called bAbI1, first proposed in [50]. In the DisCoCirc model of the bAbI1 task, we have a context text from which we extract information by asking a series of yes/no questions. An example context text is:

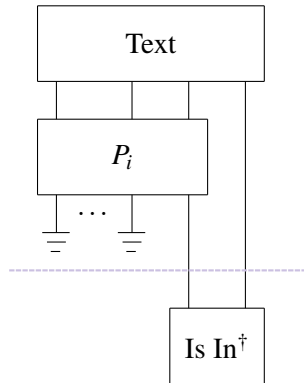
John goes to the bedroom.  
Mary walks to the hallway.  
John goes back to the bathroom.



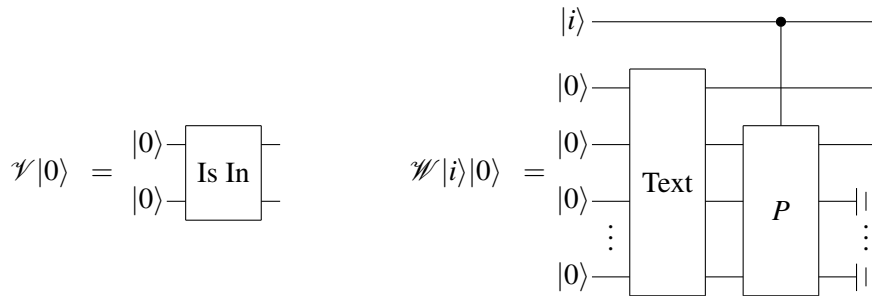
And the task then is to answer a question about the state of the world. For example: “Where is John?”. In DisCoCirc, we do this by stating a potential answer to the question as an affirmative: “John is in the bathroom.” Then we measure which potential answer (“John is in the bathroom”, “John is in the hallway”, etc) has the highest overlap with the context text, discarding the nouns that do not occur in the answer. Graphically, this means we wish to find the maximum of the following circuits:



We can phrase this as a instance of the closest vector problem by noting that this is equivalent to finding  $P_i$  which maximizes the following probability. The set  $\{P_1, P_2 \dots P_i \dots\}$  is any set of permutations such that  $P_j$  brings the  $j$ th input to the top.

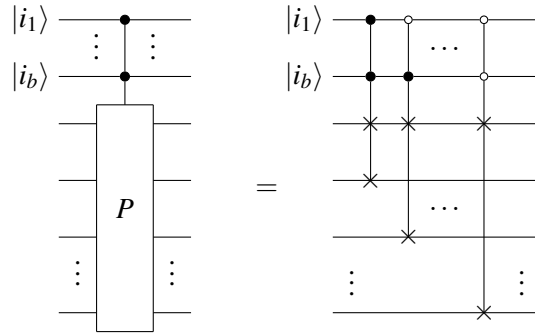


To view this as a closest vector problem, we divide the problem along the dotted line, letting the upper portion be the variable set of states  $\rho_i$  and the lower portion be  $\rho_0 = |v_0\rangle\langle v_0|$ . Since  $\rho_0$  is pure, we can apply the quantum closest vector algorithm of Basheer et al [5] to find  $i = \arg \max_{j \geq 1} F(\rho_j, \rho_0)$ . For this task the oracles  $\mathcal{V}$  and  $\mathcal{W}$  are given by the following circuits:

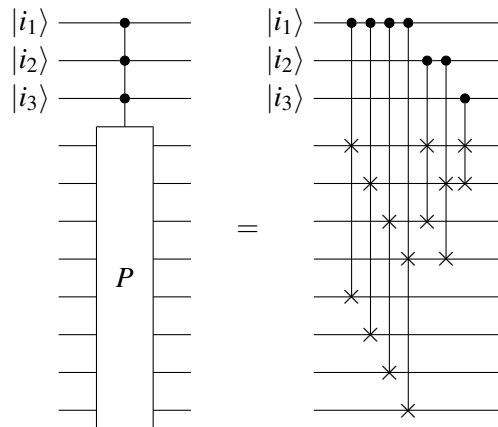


Here,  $P$  is a 1-to- $n$  multiplexer, it maps each wire to the top output (leaving the rest undefined) depending on the control register  $|i\rangle$ . There are several ways of constructing this operation. For instance, we can do a CSWAP controlled on every possible bitstring of the control register  $|i\rangle = |i_1 i_2 \dots i_b\rangle$ . Here,

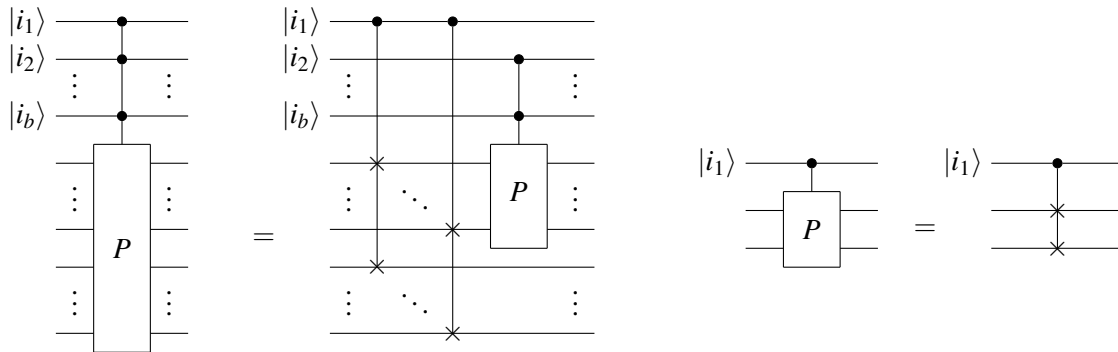
the CSWAPs are swapping nouns, so they can be realized with  $O(\log N)$  CSWAPs acting on qubits. This looks like the following circuit:



The gate complexity of this is  $O(n \log(N) \log(n))$  since we have to put a CX gate with  $O(\log(n))$  controls on each of  $O(n \log(N))$  qubits, and swaps the noun at the index with the top noun, leaving the rest unchanged. However, we can do better by leaving the other nouns in an unspecified order using a binary search method - first, we swap the first and second half of the nouns depending on the first bit of the index, then the first and second quarter depending on the second bit of the index, and so on, until the last bit controls whether we swap the top two nouns. This requires  $n \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots\right) O(\log(N)) = O(n \log(N))$  singly-controlled Toffolis, so has an overall gate complexity of  $O(n \log(N))$ . For example, the circuit on eight words looks like the following:



$P$  can be constructed recursively as:



Therefore, the oracles are given as follows in pseudocode:

```

fn V(n1, n2) {
  IsIn(n1, n2);
}

fn NounCSWAP(ctrl, n1, n2) {
  for q in 0..log(N) {
    if ctrl {
      SWAP(n1[q], n2[q]);
    }
  }
}

fn P(i, nouns) {
  if nouns.len() == 1 {
    return;
  } else if nouns.len() == 2 {
    NounCSWAP(i[0], nouns[0], nouns[1])
  } else {
    const m = nouns.len() / 2;

    for j in 0..m {
      NounCSWAP(i[0], nouns[j], nouns[m + j]);
    }

    P(i[1..], nouns[..m]);
  }
}

fn W(i, n1, n2) {
  let ancillas = InitAncillas(log(N) * (n - 2));
  let nouns = [n1, n2, .. ancillas];
  Text(nouns);
  P(i, nouns);
  Discard(ancillas);
}

```

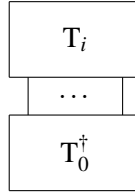
The gate complexity of the  $\mathcal{W}$  oracle is thus  $O(wA(N, d) + n \log(N))$ , and for the  $\mathcal{V}$  oracle it is  $O(A(N, 2)) = O(A(N, d))$ . It is reasonable to make the assumption that  $A(N, d) = O(d \log(N))$ . For example, this is the case if we are using standard ansatze like those from Sim et al [47]. In the closest vector algorithm for this task, we have  $M = O(n)$ . Therefore, for a fixed success probability and precision  $\epsilon$  for calculating fidelity, the overall gate complexity of the algorithm is:

$$O\left(\frac{\sqrt{n}}{\epsilon}(wd + n + \log(N))\log(N)\right) \leq O\left(\frac{\sqrt{nw}d\log^2(N)}{\epsilon}\right)$$

The inequality here follows from  $n = O(w)$ . Furthermore, the number of ancilla used by the oracles is at most  $O(n \log(N))$ .

### C.1 The Text-Text Similarity Task

The text-text similarity task involves finding a text that is most similar to a target text, by maximizing the overlap between the states representing the two texts. This is a versatile primitive that can be used both for full text-text similarity and for some variants of question answering. As discussed in Section 5, we can use the closest vector algorithm to solve this problem. Pictorially, we want to find  $i$  such that



is maximized. For this task the oracles  $\mathcal{V}$  and  $\mathcal{W}$  are given by:

- $\mathcal{V}|0\rangle = |T_0\rangle$
- $\mathcal{W}|i\rangle = |T_i\rangle$

Clearly, we have  $\mathcal{V} = U_{T_0}$  which is the direct translation of the text circuit into a quantum circuit via QDisCoCirc. However, for  $\mathcal{W}$ , we need a significantly more complicated circuit. Let us define the following parameters:

- Each noun wire has dimension  $N$ ,
- Each text circuit considered has at most  $w$  unitaries and  $n$  noun wires total,
- Each unitary in the text circuit acts on at most  $d$  wires,
- There are  $V$  words in the vocabulary,
- The number of gates in the ansatz for  $d$  wires of dimension  $N$  scales as  $A_g(N, d)$ ,
- The number of parameters in the ansatz for  $d$  wires of dimension  $N$  scales as  $A_p(N, d)$ ,
- Each parameter for the ansatz is stored with  $P$  bits of precision,
- We consider  $M$  possible texts, so that  $i \leq M$ .

We will construct the circuit for  $\mathcal{W}$  by interleaving layers of two operations. One operation is a parameterized partial permutation, which permutes some of its inputs according to information stored in a QRAM. The second is a parameterized ansatz, which applies an ansatz of a given size to a set of qubits with the parameterized angles of the ansatz stored in a QRAM. By combining these two operations, we can reconstruct a text circuit by loading from QRAM the parameters and argument locations for each unitary.

We assume access to the following QRAM oracles:

- $\text{Param}_{ij}|k\rangle|0\rangle = |k\rangle|\phi_{ijk}\rangle$  where  $\phi_{ijk}$  is a  $P$ -bit fixed-precision binary encoding of the  $j$ th parameter for the ansatz of the  $i$ th unitary in  $T_k$ ,
- $\text{Width}_i|k\rangle|0\rangle = |k\rangle|W_{ik}\rangle$  where  $W_{ik}$  is a binary encoding of the width of the  $i$ th unitary in  $T_k$ ,
- $\text{Index}_{ij}|k\rangle|0\rangle = |k\rangle|I_{ijk}\rangle$  where  $I_{ijk}$  is an encoding of the position of the  $j$ th argument to unitary  $i$  in  $T_k$ . Note that the encoding here is not a straightforward mapping from zero-indexed position to binary, and we will elaborate more later.

Each of these oracles can be implemented using a series of calls (one for each bit) to bucket-brigade QRAMS of size  $M$  [23]. In pseudocode, each of these lookups is as follows:

```
fn QRAMLookup(dict, qidx, cidx, out) {
  const ORACLES: [[Unitary]];
  let size = out.len();
  for i in 0..size {
    ORACLES[dict][cidx * size + i](qidx, out[i]);
  }
}
```

where `dict` specifies the chosen oracle, `qidx` the quantum index  $|k\rangle$  and `cidx` the classical index (i.e.  $ij$  or  $i$ ).

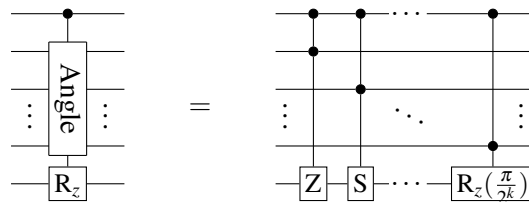
For the parameterized ansatz, we can construct a circuit out of some building blocks:

- Firstly, we have a parameterized controlled rotation gate, which implements a controlled-Z rotation based on an angle specified in auxiliary qubits.
- For a given ansatz of a specific size, we can implement a parameterized controlled circuit by replacing all gates except for parameterized rotations with their controlled versions. For the parameterized rotations, we then load the parameters from QRAM and implement each rotation using a parameterized controlled rotation circuit.
- This gives a parameterized ansatz of a specific width. To cover all possible widths, first we can load the width from QRAM, and perform the parameterized controlled ansatz of all possible widths, controlled on the width loaded from QRAM. This gives a general parameterized ansatz.

For the parameterized controlled rotation gate, we have:

```
fn PCRz(ctrl, angle, q) {
  const MAX_PREC;
  for i in 0..=MAX_PREC {
    CCRz(pi / (2^i), angle[i], ctrl, q);
  }
}
```

which is diagrammatically:



This has a gate complexity of  $O(P)$ .

Then for a specific realization of an ansatz of a given width, we can convert it to a parameterized controlled ansatz as follows. Here `angle_base` refers to the location in QRAM of the parameters for this ansatz, `text_idx` is the quantum register specifying the text index, and `angle` is an ancilla register for storing a parameter.

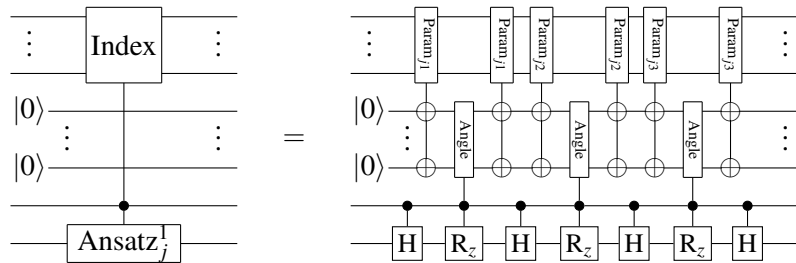
```
fn PCAnsatz(ansatz, ctrl, text_idx, angle_base, angle, qs) {
  for gate in ansatz {
```

```

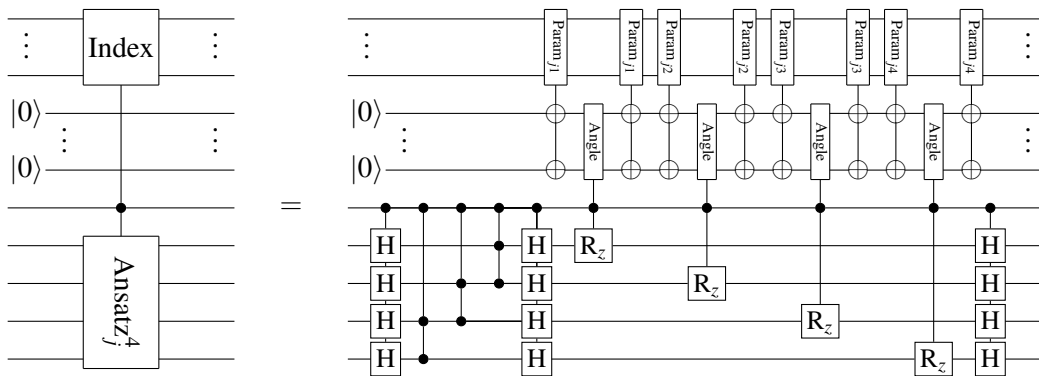
match gate {
  CX(a, b) => CCX(ctrl, qs[a], qs[b]),
  RzParam(idx, a) => {
    QRAMLookup(
      ANGLES, text_idx, angle_base + idx, angle
    );
    PCRz(ctrl, angle, qs[a]);
    QRAMLookup(
      ANGLES, text_idx, angle_base + idx, angle
    );
  },
  RzFixed(p, a) => CRz(p, ctrl, qs[a]),
  H(a) => CH(ctrl, qs[a])
}
}
}

```

For example, for an ansatz of width one (parameterized Euler rotations), corresponding to  $d = 1, N = 2$ , we have:



And for a one-layer of ansatz 9 from Sim et al. [47] of width four (i.e for  $d = 4$  and  $N = 2$  or  $d = 2$  and  $N = 4$ ) [47], we have the following:



This has a gate complexity of  $O(A_g(N, d) + P \log(M) A_p(N, d))$ . These ansatz of differing widths can be combined into one parameterized ansatz that controls all of these based on a width loaded from QRAM. We wish to do the following operation

```

fn PAnsatz(width, text_idx, angle_base, angle, qs) {
  const MAX_WIDTH;
  const ANSATZE = [

```

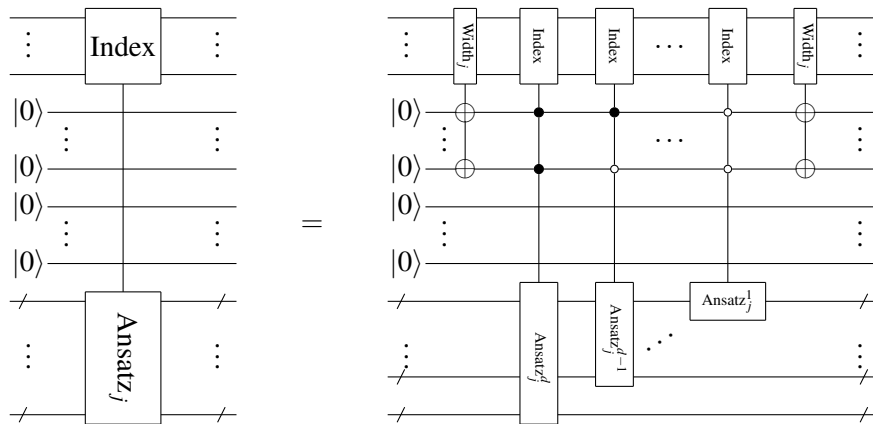
```

    AnsatzGates(w) for w in 1..=MAX_WIDTH
  ];

  QRAMLookup(WIDTHS, text_idx, i, width);
  for w in 1..=MAX_WIDTH {
    let ctrl = w == width;
    PCAnsatz(ANSATZE[w], ctrl, text_idx, angle_base, angle, qs)
    uncompute ctrl;
  }
  QRAMLookup(WIDTHS, text_idx, i, width);
}

```

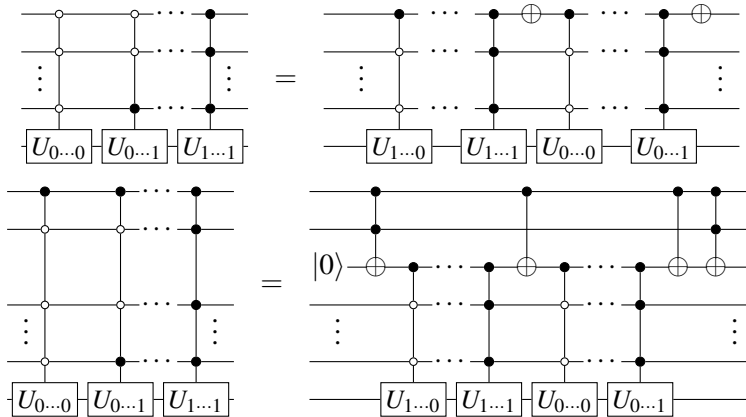
which is represented diagrammatically as:



However, compiling this circuit naively results in a gate complexity of

$$O(d \log(d) \cdot (A_g(N, d) + P \log(M) A_p(N, d)))$$

since each parameterized ansatz has  $O(\log(d))$  controls that must be combined. By making use of the unary iteration method, we can compile this more efficiently with gate complexity of  $O(d(A_g(N, d) + P \log(M) A_p(N, d)))$ . Given a series of unitaries  $U_{0..0}$  to  $U_{1..1}$  which we want to apply conditionally given an index register of size  $\log(d)$ , this technique synthesizes a circuit using singly-controlled versions of  $U_{i..j}$  with only  $O(d)$  gates. It is based on the following recurrences:



This second equation is applied until only single controls on each  $U_{i\dots j}$  remain. In pseudocode, this process is given by:

```

fn UnaryIteration(ctrls , ancillas , func) {
  X(ctrls [0]);
  CUnaryIteration(ctrls [1..], 0, ctrls [0], ancillas , func);
  X(ctrls [0]);
  CUnaryIteration(ctrls [1..], 1, ctrls [0], ancillas , func);
}

fn CUnaryIteration(ctrls , base , prev , ancillas , func) {
  if ctrls.len() > 0 {
    X(ctrls [0]);
    let fresh = ancillas [0];
    CCX(prev , ctrls [0], fresh);
    CUnaryIteration(
      ctrls [1..], 2*base + 0, fresh , ancillas [1..], func
    );
    CX(prev , fresh);
    CUnaryIteration(
      ctrls [1..], 2*base + 1, fresh , ancillas [1..], func
    );
    CCX(prev , ctrls [0], fresh);
    Discard(fresh);
  } else {
    func(base , prev);
  }
}

```

Applying this to create a parameterized ansatz, we have the following pseudocode:

```

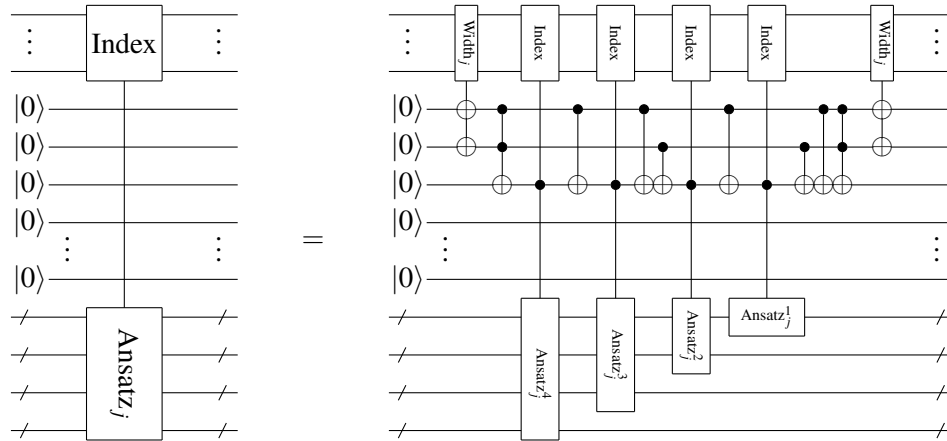
fn PAnsatz(width , text_idx , angle_base , angle , ancillas , qs) {
  const MAX_WIDTH;
  const ANSATZE = [
    AnsatzGates(w) for w in 1..=MAX_WIDTH
  ];

  QRAMLlookup(WIDTHS, text_idx , i , width);
  UnaryIteration(width , ancillas , |idx , ctrl| {
    if idx > 0 && idx <= MAX_WIDTH {
      PCAnsatz(
        ANSATZE[idx], ctrl , text_idx , angle_base , angle , qs
      );
    }
  });
  QRAMLlookup(WIDTHS, text_idx , i , width);
}

```



For the case that  $\text{MAX\_WIDTH} = 4$  and hence  $\log(d) = 2$ , this can be represented diagrammatically as follows (although note that we use a more optimized form of unary iteration on two index qubits).



Note that in this setup, the index  $0\dots 0$  is mapped to an ansatz with width one, and in general  $x$  in binary maps to the ansatz with width  $x + 1$ . Note that if we wish to apply the identity operator, we can apply an ansatz of width one with all angles set to zero. This has an overall gate complexity of  $O(d(A_g(N, d) + P \log(M) A_p(N, d)))$ .

For the parameterized partial permutation, we can reuse the multiplexer given before to implement each bit of the partial permutation. This puts a bit of our choice at the top of the multiplexer and leaves the rest permuted. By stacking  $d$  of these together, we can specify which qubits should form the top  $d$  qubits that the ansatz is applied too. We can calculate the required indices by propagating the desired index forward through all previous changes to calculate the actual index required at each step. Noting that each swap in the multiplexer is actually a swap of noun wires (which consist of multiple qubits), we can represent the multiplexer in pseudocode as:

```

fn NounCSWAP(ctrl , n1 , n2) {
    const N;

    for q in 0..log(N) {
        CSWAP(ctrl , n1[q] , n2[q]);
    }
}

fn P(idx , nouns) {
    if nouns.len() == 1 {
        return;
    } else if nouns.len() == 2 {
        NounCSWAP(idx[0] , nouns[0] , nouns[1])
    } else {
        const mf = floor(nouns.len() / 2);
        const mc = ceil(nouns.len() / 2);

        for j in 0..mf {
            NounCSWAP(i[0] , nouns[j] , nouns[mc + j]);
        }
    }
}

```

```

    }
    P(idx[1..], nouns[..mc]);
  }
}

```

This has a gate complexity of  $O(n \log(N))$ . From this the parameterized partial permutation can be constructed as

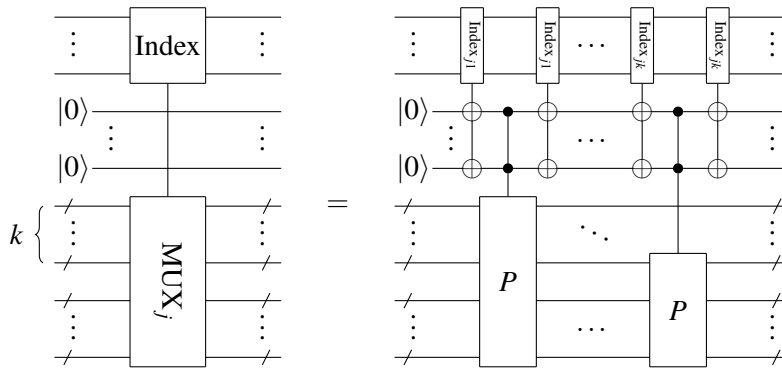
```

fn NounMux(i, idx, text_idx, nouns) {
  const MAX_ARITY;

  for j in 0..MAX_ARITY {
    QRAMLookup(INDICES, text_idx, i * MAX_ARITY + j, idx);
    P(idx, nouns[j..]);
    QRAMLookup(INDICES, text_idx, i * MAX_ARITY + j, idx);
  }
}

```

and diagrammatically, this looks as follows:



This has an overall gate complexity of  $O(d(n \log(N) + \log(n) \log(M)))$ .

Finally, by layering these two parts, we can construct the oracle. In this way, we first move the wires for each unitary to the top of the circuit, apply an ansatz (which applies the given unitary), and apply the inverse of the permutation. Then we repeat until no unitaries are remaining. In pseudocode, we have

```

fn Oracle(text_idx, nouns) {
  const MAX_PARAMS;
  const MAX_WORDS;
  const MAX_WIDTH;
  const MAX_NOUNS;
  const MAX_PREC;

  let idx = InitAncillas(MAX_NOUNS.log2());
  let angle = InitAncillas(MAX_PREC);
  let width = InitAncillas(MAX_WIDTH.log2());
  let ui_ancillas = InitAncillas(MAX_WIDTH.log2());
  for i in 0..MAX_WORDS {

```

```

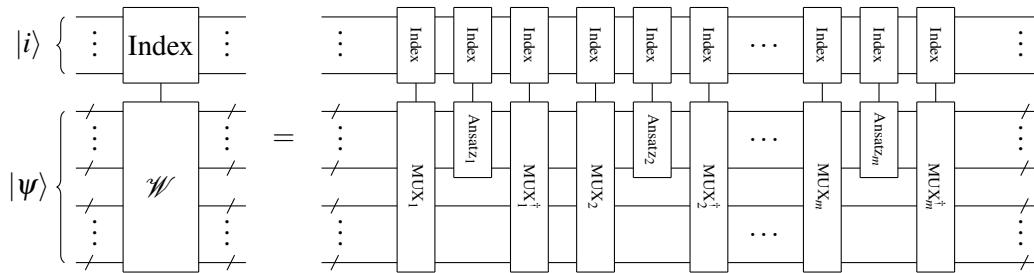
NounMux(i, idx, text_idx, nouns);

PAnsatz(
    width, text_idx, i * MAX_PARAMS,
    angle, ui_ancillas, nouns[..MAX_ARITY]
);

NounMux'(i, idx, text_idx, nouns);
}
Discard(idx);
Discard(angle);
Discard(width);
Discard(ui_ancillas);
}

```

and diagrammatically this looks like:



This has an overall gate complexity of

$$O(wd(n \log(N) + \log(n) \log(M) + A_g(N, d) + P \log(M) A_p(N, d)))$$

which can be simplified to:

$$O(wd \log(M)(n \log(N) + P A_g(N, d)))$$