

# Quantum Circuit Optimization in the ZX-Calculus



Ryan Krueger  
Christ Church College  
University of Oxford

A thesis submitted for the degree of  
*Master of Science in Mathematical Sciences*

Trinity 2021

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod  
tempor incididunt ut labore et dolore magna aliqua. Ut enim ad  
minim veniam, quis nostrud exercitation ullamco laboris nisi ut  
aliquip ex ea commodo consequat.

## Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

## Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background</b>	<b>3</b>
2.1	Qubits . . . . .	3
2.2	Quantum Circuits . . . . .	4
2.3	ZX-Calculus . . . . .	6
2.4	Circuit Optimization in the ZX-Calculus . . . . .	8
<b>3</b>	<b>Methods</b>	<b>12</b>
3.1	Congruences . . . . .	13
3.2	Search Procedures . . . . .	16
3.2.1	Simulated Annealing . . . . .	16
3.2.2	Genetic Algorithms . . . . .	17
3.2.3	Objective Functions . . . . .	19
3.3	Circuit Optimization Strategies . . . . .	20
3.3.1	Primary: Post-Simplification Search . . . . .	20
3.3.2	Alternative: Congruences as First-Class Citizens . . . . .	21
<b>4</b>	<b>Results</b>	<b>23</b>
<b>5</b>	<b>Application: Quantum Chemistry</b>	<b>24</b>
<b>6</b>	<b>Conclusion</b>	<b>25</b>
	<b>Bibliography</b>	<b>26</b>

# Chapter 1

## Introduction

Quantum computing arose from the idea that the “bugs” of quantum mechanics (e.g., superposition) may serve as useful features for information processing. Since this framing was posed in the late 1970’s, a rich theory of quantum information has developed that treats a *qubit*, a superposition of two classical bits, as the basic unit of information. Quantum information science is an exciting field of work as many quantum algorithms have been discovered that offer significant speedups from their classical counterparts (e.g., Shor’s algorithm for integer factorization). However, like the formulation of the Turing machine decades before the invention of the transistor, efforts to engineer a real-world quantum computer that can implement these theories are in their infancy.

The current state-of-the-art are so-called noisy intermediate-scale quantum (NISQ) computers [7]. These are devices without enough qubits to spare for error correction (i.e., noisy) and with a relatively small qubit number (i.e., intermediate-scale). With  $\sim 50$  qubits, NISQ devices cannot run particularly resource-intensive quantum algorithms (e.g., Shor’s algorithm at a meaningful scale) but do afford the occasional speedup over classical computers. To stress these capabilities, there is significant effort towards minimizing *quantum circuits*, the standard model for a quantum computation, to mitigate noise and decoherence. More formally, this field is known as *quantum circuit optimization*.

The ZX-calculus, a graphical formalism for quantum circuits rooted in category theory, provides a unique setting for circuit optimization. In this language, quantum circuits are represented as *ZX-diagrams* that are manipulated via a set of rewrite rules. ZX-diagrams provide a lower-level representation for quantum computations and can be deformed arbitrarily, enriching the space of possible simplifications. While there is no known procedure for recovering a quantum circuit from a general ZX-diagram, subsets of transformations have been identified that preserve circuit extractability.

These transformations have enabled a suite of circuit optimization procedures in the ZX-calculus (e.g., for reducing T-count) with various tradeoffs.

Still, circuit extraction from a ZX-diagram remains a bottleneck for optimization; the best known extraction procedures can introduce an unwieldy degree of complexity in the resulting circuit. For example, circuit extraction can introduce many CNOT gates which are particularly problematic on NISQ devices. Interestingly, however, closely related ZX-diagrams can induce drastically different circuits. Current optimization procedures employing the ZX-calculus do not exploit this fact (e.g., by searching local variants) but rather return a ZX-diagram (or its associated circuit) once no more graph simplifications can be applied.

In this thesis, we explore the role that such a search over local variants of ZX-diagrams could play in circuit optimization. More specifically, we apply various search procedures (i.e., simulated annealing and genetic algorithms) over the application of rewrite rules that do not necessarily simplify the graph itself but may reduce the complexity of the associated circuit. By default, we apply this search over an already simplified diagram (via off-the-shelf methods) but also explore more general strategies for combining these search procedures and/or rewrite rules with existing optimization procedures. We find that (FIXME: this combination is effective for reducing the 2-qubit count on random circuits when the number of qubits is less than 10). FIXME: Also present application.

# Chapter 2

## Background

To understand quantum circuit optimization in the ZX-calculus, one first must understand quantum circuits and the ZX-calculus. In this section we provide the requisite background in both of these, preceded by a brief discussion on qubits. We conclude with an overview of quantum circuit optimization in the ZX-calculus.

### 2.1 Qubits

A classical bit has a value of 0 or 1. A quantum bit, or *qubit*, encodes a quantum superposition of these two values and therefore more information than a classical bit. In the traditional bra-ket notation of quantum mechanics, a qubit is represented by a vector in  $\mathbb{C}^2$ ,

$$|\psi\rangle = \lambda_1|0\rangle + \lambda_2|1\rangle = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

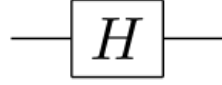
where

$$\begin{aligned} |0\rangle &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ |1\rangle &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned}$$

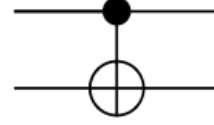
and  $\lambda_1$  and  $\lambda_2$  are the probability amplitudes of observing  $|0\rangle$  and  $|1\rangle$  upon measurement in this basis, respectively. Notably, qubits can be written with respect to any orthonormal basis. The set  $\{|0\rangle, |1\rangle\}$  is known as the *computational* basis. Another common basis is the  $+/-$  basis, where

$$\begin{aligned} |+\rangle &= \frac{|0\rangle + |1\rangle}{\sqrt{2}} \\ |-\rangle &= \frac{|0\rangle - |1\rangle}{\sqrt{2}} \end{aligned}$$





(a) The Hadamard gate.



(b) The CNOT gate. The top and bottom qubits are the control and target, respectively.

Figure 2.1: The circuit representation of two common quantum logic gates. Wires represent qubits and boxes represent quantum logic gates. Quantum circuits are read from left to right.

The computational and  $+/-$  bases are also known as the Z and X bases, respectively.

More generally, a system of  $n$  qubits is represented by the tensor product of the individual states. So, an  $n$ -qubit state corresponds to a  $2^n$ -dimensional vector. For example, the two-qubit state with both qubits in  $|0\rangle$  is

$$|0\rangle \otimes |0\rangle = |00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

In the ZX-calculus and this work, qubits are abstracted as wires in string diagrams. However, this definition is included both for completeness and to aid understanding of quantum circuits.

## 2.2 Quantum Circuits

A quantum circuit models a quantum computation as a sequence of discrete gates. In the traditional notation with qubits as vectors in  $\mathbb{C}^2$ , quantum gates correspond to unitary matrices. For example, the Hadamard gate, a quantum gate that acts on a single qubit and maps

$$\begin{aligned} |0\rangle &\mapsto \frac{|0\rangle + |1\rangle}{\sqrt{2}} \\ |1\rangle &\mapsto \frac{|0\rangle - |1\rangle}{\sqrt{2}} \end{aligned}$$

has the matrix form

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Figure 2.1a depicts the circuit representation of the Hadamard gate. Note that  $|+\rangle = H|0\rangle$  and  $|-\rangle = H|1\rangle$ . An important family of single-qubit gates are the *phase shift*

gates that modify the phase of the quantum state by some  $\alpha$ . A phase shift gate maps

$$\begin{aligned} |0\rangle &\mapsto |0\rangle \\ |1\rangle &\mapsto e^{i\alpha}|1\rangle \end{aligned}$$

and has the matrix form

$$R_\alpha = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{bmatrix}$$

Examples of phase gates are the T-gate ( $\alpha = \frac{\pi}{4}$ ) and the S-gate ( $\alpha = \frac{\pi}{2}$ ). Phase gates where  $\alpha$  is a multiple of  $\pi/2$  are called *Clifford* gates.

Quantum gates are not restricted to acting on single qubits. For example, the controlled-NOT (CNOT) gate acts on two qubits and flips the second (target) qubit only when the first (control) qubit is  $|1\rangle$ . The CNOT gate is typically depicted as shown in Figure 2.1b and has the following matrix form:

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

In general, an  $n$ -qubit gate is represented by a  $2^n$ -dimensional unitary. However, attention is typically restricted to single- and 2-qubit gates as it is well known that the Clifford + T gate set, consisting of the H, T, S, and CNOT gates, is *universal* – any other operation can be represented by a finite sequence from this set. The *Clifford circuits* are those that can be generated by  $\{H, S, CNOT\}$ . As the T and S gates are instances of phase shift gates,  $\{H, R_\alpha, CNOT\}$  is also universal though in practice any non-Clifford phase gate is implemented with T gates.

Quantum gates can be composed sequentially or in parallel. Consider two gates  $A$  and  $B$ . The effect of  $B$  applied in series after  $A$  can be described by a single gate (i.e., linear map) via matrix multiplication (i.e.,  $B \cdot A$ ). Alternatively, the tensor product is used to describe  $A$  and  $B$  in parallel (i.e.,  $B \otimes A$ ). For example, the simple circuit shown in Figure 2.3a can be described by the following linear map:

$$(\mathbb{I} \otimes CNOT) \cdot (\mathbb{I} \otimes H \otimes \mathbb{I}) \cdot (CNOT \otimes \mathbb{I}) \cdot (S \otimes \mathbb{I} \otimes \mathbb{I})$$

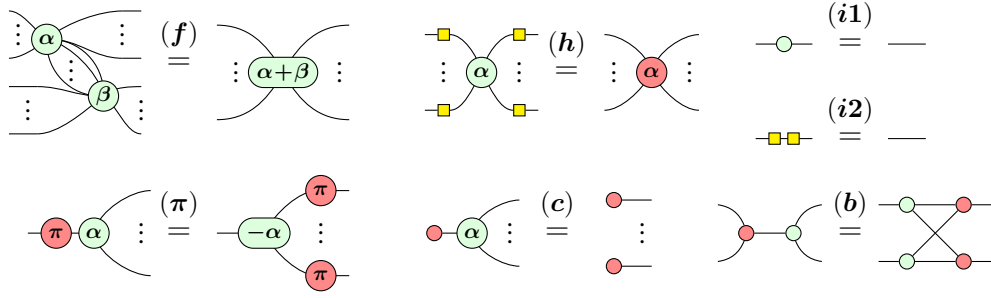


Figure 2.2: The rules of the ZX-calculus. All rules hold with the colors interchanged and for  $\alpha, \beta \in [0, 2\pi)$ .

## 2.3 ZX-Calculus

The ZX-calculus is a graphical language for representing and reasoning about quantum processes. Quantum processes are represented by *ZX-diagrams* which consist of *wires* and *spiders*. Spiders can have an arbitrary number of inputs and outputs and come in two flavors: Z spiders (shown in green) and X spiders (shown in red). As with quantum circuits, a ZX-diagram is interpreted from left to right.

ZX-diagrams abstract away the complexities of linear algebra and tensor products from traditional formalisms of quantum processes. Wires still represent qubits while spiders provide a general form for linear maps, where

$$\begin{aligned} \text{Z spider } \alpha &:= |0\dots 0\rangle\langle 0\dots 0| + e^{i\alpha} |1\dots 1\rangle\langle 1\dots 1| \\ \text{X spider } \alpha &:= |+\dots +\rangle\langle +\dots +| + e^{i\alpha} |-\dots -\rangle\langle -\dots -| \end{aligned}$$

Two diagrams can be composed in serial by joining the outputs of the first to the inputs of the second, or in parallel by stacking the two diagrams.

From this definition, we can identify the diagrammatic representations for several common components of quantum computation:

$$\begin{aligned} \text{green circle} &= |0\rangle + |1\rangle \propto |+\rangle & \text{red circle} &= |+\rangle + |-\rangle \propto |0\rangle \\ \text{green circle with } \pi &= |0\rangle\langle 0| - |1\rangle\langle 1| = Z & \text{red circle with } \pi &= |+\rangle\langle +| - |-\rangle\langle -| = X \end{aligned}$$

where  $Z$  and  $X$  are the corresponding Pauli matrices. Note that spiders without any inputs can be regarded as qubit state preparations. The Hadamard gate is so pervasive that it merits shorthand notation; we use a yellow square to represent the Hadamard gate as shown below

$$\text{yellow square} = \text{green circle with } \frac{\pi}{2} \text{ --- red circle with } \frac{\pi}{2} \text{ --- green circle with } \frac{\pi}{2} \quad (2.1)$$

and replace a Hadamard gate between two spiders with a blue dashed edge:

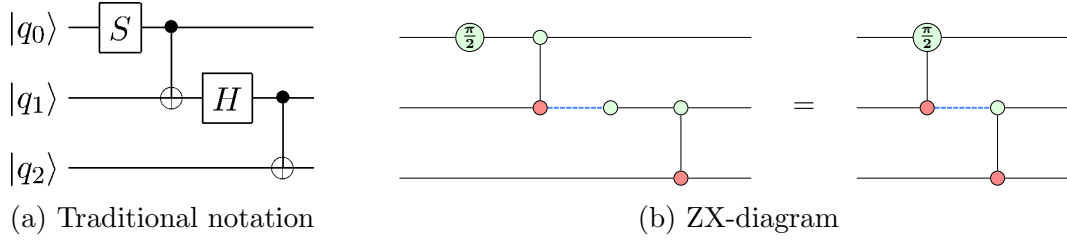


Figure 2.3: A simple 3-qubit quantum circuit shown in both the traditional notation and as a ZX-diagram.

$$\begin{array}{c} \diagup \quad \diagdown \\ \vdots \text{---} \text{green spider} \text{---} \text{blue dashed line} \text{---} \text{green spider} \text{---} \vdots \\ \diagdown \quad \diagup \end{array} := \begin{array}{c} \diagup \quad \diagdown \\ \vdots \text{---} \text{green spider} \text{---} \text{yellow box} \text{---} \text{green spider} \text{---} \vdots \\ \diagdown \quad \diagup \end{array}$$

We can also reason about ZX-diagrams. The first rule for determining equality is that *only connectivity matters* (OCM). In other words, two ZX-diagrams are equal when one can be deformed into the other (e.g., via moving vertices around in the plane or bending wires) while maintaining connectivity and the order of the inputs and outputs. In addition to this OCM principle, the ZX-calculus includes a primary set of rewrite rules as shown in Figure 2.2. These rules only hold up to non-zero scalar factors, though these scalars are typically ignored as they correspond to negligible differences in global phase. Additional rules can be derived from this set, such as the antipode rule

$$\begin{array}{c} \diagup \quad \diagdown \\ \vdots \text{---} \text{green spider} \text{---} \text{red spider} \text{---} \vdots \\ \diagdown \quad \diagup \end{array} \stackrel{(a)}{=} \begin{array}{c} \vdots \text{---} \text{green spider} \text{---} \vdots \quad \vdots \text{---} \text{red spider} \text{---} \vdots \end{array}$$

and the  $\pi$ -copy rule

$$\begin{array}{c} \diagup \quad \diagdown \\ \vdots \text{---} \text{red spider} \text{---} \text{green spider} \text{---} \vdots \\ \diagdown \quad \diagup \end{array} \stackrel{(\pi c)}{=} \begin{array}{c} \text{red spider} \text{---} \text{green spider} \text{---} \text{red spider} \\ \vdots \quad \vdots \quad \vdots \end{array}$$

All quantum circuits can be translated to ZX-diagrams. This is a consequence of the fact that the following universal gate set can be easily represented in the ZX-calculus:

$$CNOT = \begin{array}{c} \vdots \text{---} \text{green spider} \text{---} \vdots \\ \vdots \text{---} \text{red spider} \text{---} \vdots \end{array} \quad R_\alpha = \text{green spider with phase } \alpha \quad H = \text{yellow box}$$

In the CNOT diagram, the green and red spiders are the control and target qubits, respectively. As an example, Figure 2.3 depicts both the traditional and diagrammatic representations of a simple quantum circuit.

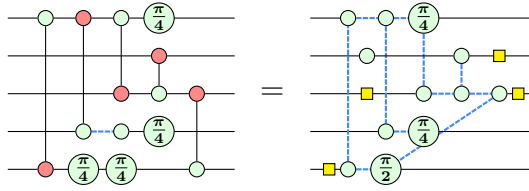


Figure 2.4: An example of a ZX-diagram and an equivalent, graph-like ZX-diagram.

## 2.4 Circuit Optimization in the ZX-Calculus

The goal of quantum circuit optimization is to simplify quantum circuits (e.g., via reducing the number of gates) to reduce noise and decoherence when run on modern quantum computers. More specifically, the T-count and 2-qubit-count are typically targeted; fault tolerant implementations of T gates generally require an order of magnitude more resources than Clifford gates [2], and the fidelity of single-qubit gates is demonstrably higher than that of 2-qubit gates [1]. It is standard to perform optimizations at the circuit-level. Examples of circuit-level optimizations are 3-qubit gate decomposition and cancelling back-to-back CNOT gates.

The ZX-calculus enables circuit optimization at the graph-level, a more flexible setting that is well-studied in its own right. Such an optimization involves converting a circuit to a ZX-diagram, simplifying the diagram, and converting the diagram back to a circuit. However, there is no known general-purpose procedure for recovering a quantum circuit from a generic ZX-diagram. This restricts the space of permissible simplifications to those that preserve whichever diagrammatic properties are required by the extraction procedure.

In 2020, Duncan et al. reported an extraction procedure that enabled current best practices [3]. Firstly, this procedure operates on *graph-like* ZX-diagrams. A ZX-diagram is graph-like when:

1. All spiders are Z-spiders.
2. Z-spiders are only connected via Hadamard edges
3. There are no parallel Hadamard edges or self-loops
4. Every input or output is connected to a Z-spider and every Z-spider is connected to at most one input or output

Every ZX-diagram is equal to a graph-like diagram (see Figure 2.4 for an example), and this form admits an underlying graph structure that permits graph-theoretic

analyses. Secondly, this procedure requires that the underlying graph of the input ZX-diagram satisfies a graph-theoretic invariant called *focused generalized flow* (focused gFlow). Given such an input ZX-diagram, extraction proceeds in such a way that each non-zero phase corresponds to one phase gate in the resulting circuit but each edge can correspond to multiple CNOTs. For this reason, local changes in a ZX-diagram (e.g., connectivity) can have significant effects on the complexity of the associated circuit. Importantly, this extraction procedure is not optimal; if a given circuit is converted to a ZX-diagram and the extraction procedure is immediately applied, the extracted circuit is often more complex than the original circuit. In particular, this procedure can drastically increase the number of 2-qubit gates. For more details on the extraction procedure or focused gFlow, see [3].

Alongside this extraction procedure, Duncan et al. also introduced an optimization procedure that relies on focused gFlow-preserving graph simplifications. More specifically, this procedure relies on two graph-theoretic transformations that each correspond to rewrite rules in the ZX-calculus: *local complementation* and *pivoting*. Given a graph  $G$  and some vertex  $u$  of  $G$ , the *local complementation* of  $G$  according to  $u$ , denoted  $G \star u$ , is the graph in which the connectivity of all pairs of neighbors of  $u$  is inverted. For example,



Given a ZX-diagram, local complementation can be induced in the underlying graph structure while maintaining equality by applying an  $X_{-\pi/2}$  gate on (the spider corresponding to)  $u$  and a  $Z_{\pi/2}$  gate to its neighbors [4]:

$$(2.2)$$

Relatedly, given two connected vertices  $u$  and  $v$  in  $G$ , the *pivot* of  $G$  along the edge  $uv$  is the graph  $G \wedge uv := G \star u \star v \star u$ . In practice, this consists of complementing the edges between three subsets of vertices: (1) the common neighborhood of  $u$  and  $v$ , (2) the exclusive neighborhood of  $u$ , and (3) the exclusive neighborhood of  $v$ :



Again, given a ZX-diagram, the rewrite rule reported in [5] introduces a pivot by applying Hadamard gates on  $u$  and  $v$ , and  $Z_\pi$  gates on their common neighborhood:

$$\begin{array}{c}
 \begin{array}{c}
 \text{Diagram 1: A vertex } u \text{ (yellow square) and } v \text{ (yellow square) are connected to a common neighborhood } N(u) \cap N(v) \text{ (green circles). The neighborhood is connected to } u \text{ and } v \text{ via blue dashed lines. The neighborhood is also connected to other vertices in } N(u) \setminus (N(v) \cup \{v\}) \text{ and } N(v) \setminus (N(u) \cup \{u\}) \text{ via blue dashed lines.} \\
 \end{array}
 =
 \begin{array}{c}
 \text{Diagram 2: A simplified version of the diagram where the neighborhood } N(u) \cap N(v) \text{ is replaced by a single vertex (green circle) connected to } u \text{ and } v \text{ via blue dashed lines. The neighborhood is also connected to other vertices in } N(u) \setminus (N(v) \cup \{v\}) \text{ and } N(v) \setminus (N(u) \cup \{u\}) \text{ via blue dashed lines.} \\
 \end{array}
 \end{array}
 \quad (2.3)$$

Each rewrite rule can be extended to a simplification by performing local complementation (resp. pivoting), removing the vertex (resp. the pair of vertices), and updating the phases:

$$\begin{array}{c}
 \begin{array}{c}
 \text{Diagram 1: A vertex } \alpha_1 \text{ (green circle) is connected to } \alpha_2 \text{ (green circle) and } \alpha_n \text{ (green circle). The vertex } \alpha_1 \text{ is also connected to } \alpha_{n-1} \text{ (green circle). The vertex } \alpha_1 \text{ is marked with a red asterisk.} \\
 \end{array}
 =
 \begin{array}{c}
 \text{Diagram 2: A simplified version of the diagram where the vertex } \alpha_1 \text{ is replaced by a single vertex (green circle) connected to } \alpha_2 \text{ and } \alpha_n \text{ via blue dashed lines. The vertex } \alpha_1 \text{ is also connected to } \alpha_{n-1} \text{ via blue dashed lines.} \\
 \end{array}
 \\
 \begin{array}{c}
 \text{Diagram 3: A pair of vertices } \alpha_1 \text{ (green circle) and } \alpha_n \text{ (green circle) are connected to } \alpha_2 \text{ (green circle) and } \alpha_{n-1} \text{ (green circle). The vertices } \alpha_1 \text{ and } \alpha_n \text{ are marked with red asterisks.} \\
 \end{array}
 =
 \begin{array}{c}
 \text{Diagram 4: A simplified version of the diagram where the pair of vertices } \alpha_1 \text{ and } \alpha_n \text{ is replaced by a single vertex (green circle) connected to } \alpha_2 \text{ and } \alpha_{n-1} \text{ via blue dashed lines.} \\
 \end{array}
 \end{array}$$

Importantly, the existence of a focused gFlow is preserved in both cases. At a high-level, optimization can then be performed by applying these simplifications to fixpoint and extracting a circuit from the resulting ZX-diagram. Details on this optimization procedure and proofs of the simplification rules or their focused gFlow-preservation can be found in [3]. A variant of this procedure is implemented in the `full_reduce` method of the PyZX library.

One issue with `full_reduce` is its reliance on circuit extraction; it is not uncommon that the circuit extracted from the final, simplified ZX-diagram has more gates than the input circuit. Kissinger and van de Wetering introduced an alternative optimization method that uses *phase teleportation* to sidestep the issue of circuit extraction altogether [6]. Crucially, simplification of the ZX-diagram is performed symbolically. This symbolic simplification leverages the fact that several rewrite rules involving *phase-gadgets*, a particular ZX-diagrammatic motif, correspond to changes in the original phases. Therefore, non-Clifford phases can potentially cancel or combine with each other while leaving the original graphical structure of the ZX-diagram

intact. So, as simplification proceeds, application of these rules is simply tabulated to enable reconstruction of a final circuit with the same connectivity but a possible reduction in non-Clifford gates (and therefore T-count). Importantly, using phase teleportation rather than circuit extraction ensures that the number of 2-qubit and Hadamard gates is unchanged. By changing the angles of many non-Clifford phase gates to either 0 or multiples of  $\pi/2$ , this procedure can also make a circuit-level optimization procedure much more effective. This procedure is implemented in the `teleport_reduce` method of PyZX.

In both cases, it is standard to apply a set of circuit-level optimizations post-simplification. This set of optimizations is implemented in the `basic_optimization` routine in PyZX. Note that the circuit must first be extracted after `full_reduce` whereas these optimizations can be applied directly after `teleport_reduce`. In summary, the two standard circuit optimization pipelines in the ZX-calculus are `full_reduce + extract_circuit + basic_optimization` and `teleport_reduce + basic_optimization`.



# Chapter 3

## Methods

At a high level, the goal of any quantum circuit optimization is to reduce the complexity of the input circuit. A core assumption of circuit optimization in the ZX-calculus is that ZX-diagrams with fewer spiders typically correspond to simpler circuits. Earlier, however, we noted that local changes (e.g., connectivity) in a ZX-diagram can drastically affect the complexity of the associated circuit obtained via extraction. However, no existing optimization procedure over ZX-diagrams addresses this and searches this local space. The `full_reduce` method described in Section 2.4 is the quintessential example of this lost opportunity. After the two simplification rules can no longer be applied, the final circuit is taken to be that which is extracted from the resulting simplified ZX-diagram; however, there may be an equivalent ZX-diagram lurking nearby whose associated circuit is markedly less complex.

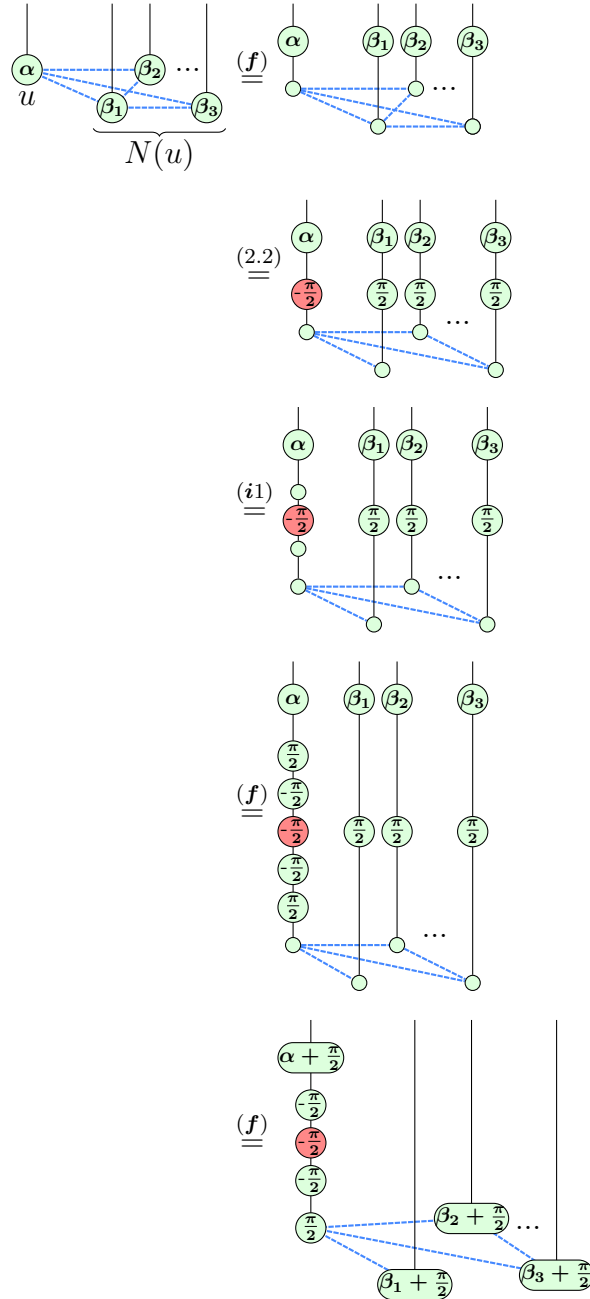
In this thesis, we explore the utility of searching this local space of equivalent ZX-diagrams. To do so, we first need rewrite rules that modify a ZX-diagram in a manner other than spider removal that may reduce circuit complexity. Indeed, a rule that introduces several additional spiders alongside changes in connectivity may in turn yield a less complex circuit. We refer to these rewrite rules as *congruences*. Given a set of congruences, we then need a procedure to search the space of equivalent ZX-diagrams generated by an input ZX-diagram and these rules. We can then devise strategies for incorporating this local search into existing methods for circuit optimization in the ZX-calculus.

We present the methods of our work in this order. First, we generalize the original (non-simplification) variants of local complementation and pivoting to ZX-diagrams with arbitrary phases. We then describe two search procedures, simulated annealing (SA) and genetic algorithms (GA). We also define a measure of circuit complexity and discuss candidate objective functions to guide search. Lastly, we discuss how this search can be incorporated into existing optimization pipelines.

### 3.1 Congruences

The non-simplification versions of local complementation and pivoting presented in Section 2.4 embody the desired properties of congruences. They change the connectivity of the ZX-diagram without introducing an unwieldy number of additional gates, presenting an opportunity for a potential reduction in circuit complexity. However, Equations 2.2 and 2.3 only apply to spiders with zero phase and a single wire.

We can apply the rules of the ZX-calculus and Equation 2.2 to generalize local complementation to arbitrary phases ( $\alpha_i, \beta_i \in [0, 2\pi)$ ):



$$\begin{array}{c}
 \stackrel{(2.1)}{=} \begin{array}{c} \text{---} \\ | \\ \alpha + \frac{\pi}{2} \\ | \\ \frac{\pi}{2} \end{array} \begin{array}{c} \text{---} \\ | \\ \beta_2 + \frac{\pi}{2} \end{array} \dots \begin{array}{c} \text{---} \\ | \\ \beta_3 + \frac{\pi}{2} \end{array} \\
 \text{---} \quad \text{---} \quad \text{---} \\
 \beta_1 + \frac{\pi}{2} \quad \beta_3 + \frac{\pi}{2}
 \end{array}
 \tag{3.1}$$

Equation 3.1 can be easily extended to apply for an arbitrary number of wires connected to each spider:

$$\begin{array}{c}
 \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ | \quad | \quad | \quad | \\ \alpha \quad \beta_2 \quad \dots \quad \beta_3 \\ | \quad | \quad | \quad | \\ \beta_1 \quad \beta_1 \quad \beta_1 \quad \beta_1 \\ \underbrace{\hspace{10em}}_{N(u)} \end{array} \stackrel{(f)}{=} \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ | \quad | \quad | \quad | \\ \alpha \quad \beta_2 \quad \dots \quad \beta_3 \\ | \quad | \quad | \quad | \\ \beta_1 \quad \beta_1 \quad \beta_1 \quad \beta_1 \end{array} \\
 \stackrel{(3.1)}{=} \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ | \quad | \quad | \quad | \\ \alpha + \frac{\pi}{2} \quad \beta_2 + \frac{\pi}{2} \quad \dots \quad \beta_3 + \frac{\pi}{2} \\ | \quad | \quad | \quad | \\ \frac{\pi}{2} \quad \beta_1 + \frac{\pi}{2} \quad \beta_3 + \frac{\pi}{2} \end{array} \\
 \stackrel{(f)}{=} \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ | \quad | \quad | \quad | \\ \alpha + \frac{\pi}{2} \quad \beta_2 + \frac{\pi}{2} \quad \dots \quad \beta_3 + \frac{\pi}{2} \\ | \quad | \quad | \quad | \\ \frac{\pi}{2} \quad \beta_1 + \frac{\pi}{2} \quad \beta_3 + \frac{\pi}{2} \end{array}
 \end{array}
 \tag{C1}$$

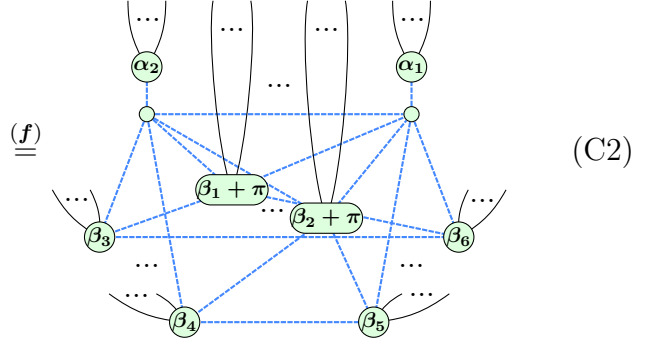
Similarly, we can generalize pivoting to arbitrary phases:

$$\begin{array}{c}
 \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ | \quad | \quad | \quad | \\ \alpha_1 \quad \beta_1 \quad \alpha_2 \quad \beta_2 \\ | \quad | \quad | \quad | \\ \beta_3 \quad \beta_4 \quad \beta_5 \quad \beta_6 \\ \underbrace{\hspace{10em}}_{N(u) \setminus (N(v) \cup \{v\})} \quad \underbrace{\hspace{10em}}_{N(v) \setminus (N(u) \cup \{u\})} \end{array} \stackrel{(f)}{=} \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ | \quad | \quad | \quad | \\ \alpha_2 \quad \beta_1 \quad \beta_2 \quad \alpha_1 \\ | \quad | \quad | \quad | \\ \beta_3 \quad \beta_4 \quad \beta_5 \quad \beta_6 \end{array}
 \end{array}$$

$$\begin{array}{c}
\begin{array}{c}
\text{---} \alpha_2 \text{---} \beta_1 \beta_2 \text{---} \alpha_1 \text{---} \\
\text{---} \pi \pi \text{---} \\
\text{---} \text{---} \text{---} \\
\text{---} \beta_3 \text{---} \beta_6 \text{---} \\
\text{---} \beta_4 \text{---} \beta_5 \text{---}
\end{array} \\
(2.3) \\
= \\
\begin{array}{c}
\text{---} \alpha_2 \text{---} \alpha_1 \text{---} \\
\text{---} \beta_1 + \pi \text{---} \\
\text{---} \beta_2 + \pi \text{---} \\
\text{---} \beta_3 \text{---} \beta_6 \text{---} \\
\text{---} \beta_4 \text{---} \beta_5 \text{---}
\end{array}
\end{array} \tag{3.3}$$

Again, we can extend this rewrite rule for arbitrary wiring:

$$\begin{array}{c}
\begin{array}{c}
\overbrace{N(u) \cap N(v)} \\
\text{---} \alpha_1 \text{---} \alpha_2 \text{---} \\
\text{---} \beta_1 \text{---} \beta_2 \text{---} \\
\text{---} \beta_3 \text{---} \beta_6 \text{---} \\
\text{---} \beta_4 \text{---} \beta_5 \text{---}
\end{array} \\
N(u) \setminus (N(v) \cup \{v\}) \quad N(v) \setminus (N(u) \cup \{u\}) \\
(f) \\
= \\
\begin{array}{c}
\text{---} \alpha_1 \text{---} \alpha_2 \text{---} \\
\text{---} \beta_1 \text{---} \beta_2 \text{---} \\
\text{---} \beta_3 \text{---} \beta_6 \text{---} \\
\text{---} \beta_4 \text{---} \beta_5 \text{---}
\end{array} \\
(3.3) \\
= \\
\begin{array}{c}
\text{---} \alpha_2 \text{---} \alpha_1 \text{---} \\
\text{---} \beta_1 + \pi \text{---} \beta_2 + \pi \text{---} \\
\text{---} \beta_3 \text{---} \beta_6 \text{---} \\
\text{---} \beta_4 \text{---} \beta_5 \text{---}
\end{array}
\end{array}$$



Equations C1 and C2 will serve as our primary congruences because they maintain a similar graph complexity (e.g., number of spiders and wires) but, due to their effects on connectivity, can produce circuits of varying complexities. Equation C1 can be applied to any spider with a degree of more than 1. Equation C2 can be applied to any pair of connected spiders.

## 3.2 Search Procedures

The space of ZX-diagrams generated by an input ZX-diagram and these congruences is combinatorial and therefore cannot be searched exhaustively. We can instead formulate this search as an optimization problem where the set of isomorphic ZX-diagrams are the states (reachable by congruence applications) that are evaluated by some quantitative measure of circuit complexity. Here we describe two search procedures as well as several candidate objective functions that either measure circuit complexity directly or use ZX-diagrammatic properties as a proxy.

### 3.2.1 Simulated Annealing

Simulated annealing is an optimization technique named after the annealing process in metallurgy in which a molten hot metal is cooled in a slow, controlled fashion in order for it to reach its most stable form. In SA, this notion of slow cooling is interpreted as a slow decrease in the probability of accepting worse solutions. In this way, the algorithm initially explores a broad region of the the search space and progressively narrows its scope.

More formally, the SA algorithm maintains a current state  $s$  and at each step randomly considers some neighboring state  $s^*$ . The algorithm then chooses whether or not to replace  $s$  with  $s^*$  via a Bernoulli random variable parameterized by the *acceptance probability*, denoted  $P(s, s^*, T)$ . The acceptance probability depends on (1) the *energy* (i.e., objective) *function*, and (2) the *temperature*. The energy function

---

**Algorithm 1** Simulated annealing

---

```
1: function SIMANNEAL( $s_0, T, c, k_{max}$ )
2:    $s \leftarrow s_0$ 
3:   for  $i \leftarrow 0$  to  $k_{max}$  do
4:      $s^* \leftarrow$  randomly sampled neighbor of  $s$ 
5:     if  $E(s^*) < E(s)$  or  $\text{random}(0, 1) < \exp(-\frac{E(s^*)-E(s)}{T})$  then
6:        $s \leftarrow s^*$ 
7:        $T \leftarrow T * c$ 
8:   return  $s$ 
```

---

$E(s)$  assigns a score to each state (lower is better). If  $E(s') < E(s)$ , then  $s$  is always replaced with  $s^*$  (i.e.,  $P(s, s^*, T) = 1$ ). Otherwise,  $P(s, s^*, T) = \exp(-\frac{E(s^*)-E(s)}{T})$  where the temperature  $T$  controls the likelihood of moving to a higher energy state. Note that  $P(s, s^*, T)$  is inversely proportional to  $E(s^*) - E(s)$  and directly proportional to  $T$ . Informally, this means that  $s^*$  is more likely to be accepted if it is closer in energy to  $s$  and with a higher temperature.  $T$  is initialized to some positive value and progressively decreases to zero. At each step,  $T$  is updated as  $T = T * c$  where  $c \in [0, 1)$  is the *cooling* parameter. In this way, SA converges towards a greedy algorithm and is more likely to accept  $s^*$  when  $E(s^*) > E(s)$  early in search when  $T$  is high. The algorithm terminates when a maximum number of steps  $k_{max}$  is reached. Algorithm 1 provides an overview of this procedure.

In this work, the state is a ZX-diagram and neighbors are sampled via the probabilistic application of rewrite rules. For example, to anneal using the rewrite rules described in Section 3.1,  $s^*$  would be sampled by first choosing one of Equation C1 or Equation C2 and then choosing a spider or pair of connected spiders to which the rule will be applied. Note that we use Equations C1 and C2 by default but in theory any set of rewrite rules can be used.

### 3.2.2 Genetic Algorithms

Genetic algorithms are a class of optimization procedures inspired by the process of natural selection. Rather than maintaining a single state (e.g., as in SA), a *population* of candidate solutions is *evolved*. This evolution is guided by a *fitness function* and operates via biologically inspired notions of *mutation*, *crossover*, and *selection*. In this work, we define fitness such that lower fitness is better to remain consistent with the energy function described in Section 3.2.1.

A genetic algorithm operates in the following general fashion. First, a population of candidate solutions (i.e., *mutants*) is randomly generated. The population size is

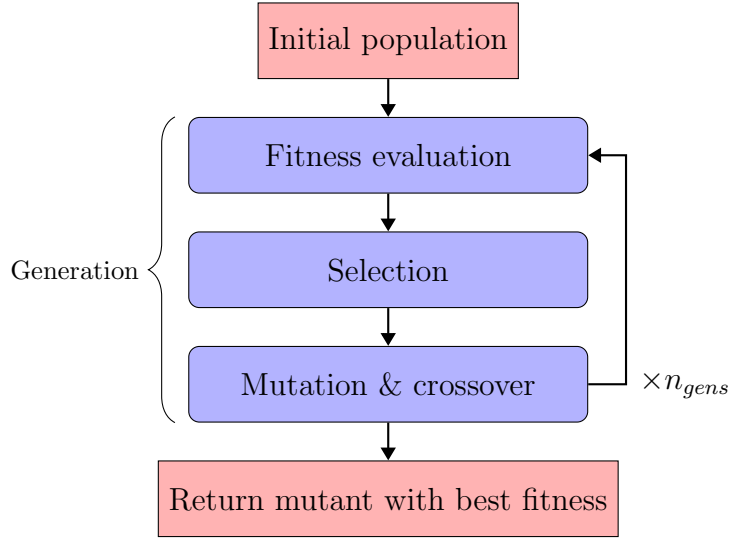


Figure 3.1: An overview of genetic algorithms.

typically fixed to some  $n_{mutants}$ . This initial population then evolves in a series of *generations*. Each generation consists of two parts: (1) selection and (2) application of genetic operators (e.g., mutation and crossover). In the selection step, a score is assigned to each mutant via the fitness function and a new population is chosen given these scores. For example, a naive selection method would set the population to  $n_{mutants}$  copies of the best-scoring mutant. Given this surviving population, the second step involves applying genetic operators to obtain a revised, diverse population. The two most common genetic operators are mutation and crossover. Analogous to mistakes being made during replication of a DNA sequence, a mutation operator involves randomly tweaking a single mutant. In the case where mutants are represented as bit strings, the canonical example of a mutation operator is a bit-flip. Alternatively, a crossover operator produces a new mutant given two or more existing mutants; this is analogous to reproduction and biological crossover. For example, given two mutants represented as strings, a new mutant can be generated by swapping the two segments defined by a random position. Evolution proceeds for  $n_{gens}$  generations and the algorithm returns the mutant with the highest fitness throughout search. This process is summarized in Figure 3.1.

The default in this work is to represent mutants as ZX-diagrams and to use rewrite rules as mutation operators. As is common in practical applications of genetic algorithms, we do not use any crossover operators though this is an attractive area for future exploration given the graphical nature of ZX-diagrams. Lastly, we use *tournament selection* to determine the surviving mutants at each generation. In this

selection method,  $k_{\text{tourn}}$  mutants are selected to compete in a “tournament” from which a winner is chosen based on fitness. This is repeated  $n_{\text{mutants}}$  times with replacement to obtain a surviving population. We use a simple variant of tournament selection in which  $k_{\text{tourn}} = 2$  and the winner is simply the mutant with better (i.e., lower) fitness.

### 3.2.3 Objective Functions

Both the energy function (used in SA) and the fitness function (used in GA) are instances of an *objective function*. An objective function measures the “goodness” of a particular solution to an optimization problem. For our purposes, we require an objective function that, given a ZX-diagram, provides a reasonable measure of the complexity of the circuit obtained via extraction. For consistency, all objective functions are defined in such a way that a low score is better (i.e., a low score indicates low complexity).

The most straightforward objective function involves extracting a circuit and measuring its complexity directly. A quantitative measure of circuit complexity is also useful for evaluating our optimization techniques. Given a circuit  $C$ , we define its complexity  $Comp(C)$  to be a weighted average of its T-count and 2-qubit count:

$$Comp(C) = 4 * (\# \text{ 2-qubit gates}) + 1 * (\# \text{ T gates})$$

**FIXME: Why?** Note that `basic_optimization` is always applied to the extracted circuit.

However, extracting a circuit from the ZX-diagram every time we want to measure its complexity is costly. For example, evolving a population of 50 mutants for 100 generations would require 5000 extractions. Instead, we could measure some property of the ZX-diagram that is a reasonable proxy for  $Comp(C)$  (where  $C$  is the circuit obtained via extraction). The following are candidate properties to serve as such a proxy:

- The number of edges
- The density of the underlying graph
- The centrality of the underlying graph

The ZX-diagram is assumed to be graph-like.

Measuring the ZX-diagram is clearly preferable but relies on the existence of a property that is sufficiently correlated to  $Comp$ . Note that none of the above objective



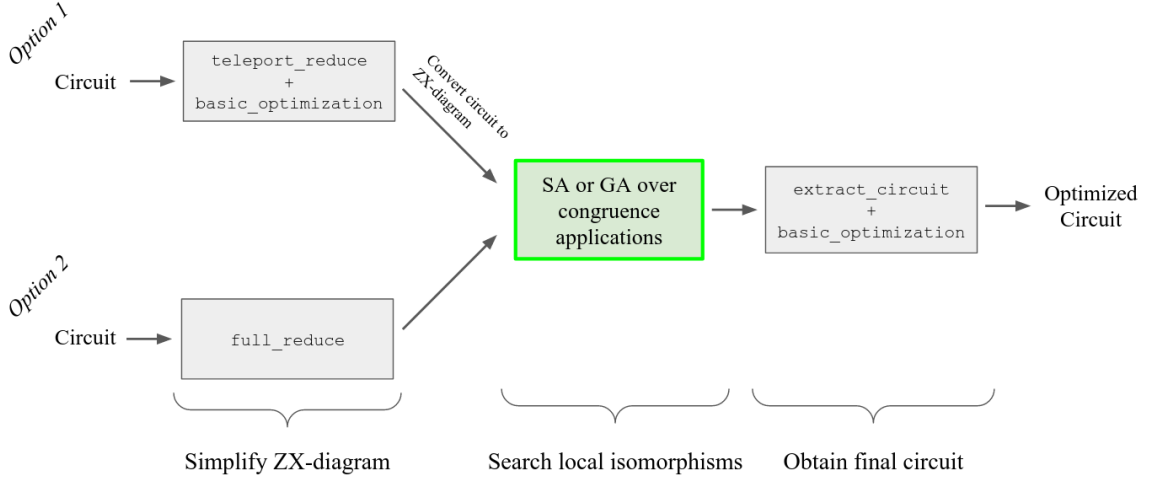


Figure 3.2: An overview of our primary optimization strategy. Given an input circuit, a simplified ZX-diagram is obtained via one of two off-the-shelf methods. Then, we search for an equivalent ZX-diagram with similar graph complexity but lower circuit complexity (highlighted in green). A circuit is extracted from the final ZX-diagram and standard circuit-level optimizations are performed.

functions, including *Comp*, are normalized and therefore scores can only be compared between isomorphic ZX-diagrams.

### 3.3 Circuit Optimization Strategies

Here we describe how we use these congruences and search procedures for quantum circuit optimization. We first outline our primary method which involves first applying off-the-shelf methods to obtain a simplified ZX-diagram and subsequently searching the space of equivalent ZX-diagrams generated by Equations C1 and C2. We then describe an alternative, more general approach in which Equations C1 and C2 are combined with simplification rules to form a single action set over which search is applied.

#### 3.3.1 Primary: Post-Simplification Search

Our primary optimization strategy improves upon the two pipelines described in Section 2.4 by searching local variants of the simplified ZX-diagrams. Given an input circuit, we can obtain a simplified ZX-diagram to seed search in one of two ways. Firstly, we can apply `full_reduce`. Importantly, we do not apply the entire `full_reduce + extract_circuit + basic_optimization` pipeline and convert the simplified circuit to a ZX-diagram as the complexity cost incurred by extraction

will likely outweigh any simplifications achieved at the circuit-level. Importantly, Alternatively, since `teleport_reduce` short-circuits circuit extraction, we can apply `teleport_reduce + basic_optimization` and convert the simplified circuit to a ZX-diagram. Given a simplified ZX-diagram, we can then apply SA or GA to search the space of ZX-diagrams generated by Equations C1 and C2 for an equivalent ZX-diagram whose extracted circuit is less complex. An overview of this strategy is depicted in Figure 3.2.

There are several opportunities to refine this strategy. One example is applying Equations C1 and C2 with unequal probabilities. This could be beneficial if one congruence more effectively navigates the search space than the other. Another possible refinement is non-uniform sampling of spiders (resp. pairs of spiders) to which Equation C1 (resp. Equation C2) will be applied. The following are possible spider metrics to weight sampling for application of Equation C1:

- Degree
- Centrality (e.g., betweenness, Katz)
- Total degree of neighbors

Similarly, the following are possible metrics to weight sampling of pairs of nodes for Equation C2:

- Sum of spider-wise metrics (e.g., total degree of the union of all neighbors)
- Edge centrality (e.g., load, betweenness)
- Edge dispersion

### 3.3.2 Alternative: Congruences as First-Class Citizens

An underlying assumption of `full_reduce` is that simplifying the ZX-diagram as much as possible is optimal. However, removing spiders to fixpoint may come at a cost. For example, later-stage spider removals may induce connectivity properties in the ZX-diagram (e.g., high-degree spiders, high density) that translate to a high 2-qubit count upon extraction.

To test this, we employ an alternative simplification strategy in which congruences are applied alongside simplifications rather than after the ZX-diagram is fully simplified. More specifically, we treat individual simplification rules as actions and

combine these with Equations C1 and C2 to form a master action set. These individual simplifications are the following: **FIXME itemize, and then a sentence on what they can be applied to.** Given an input circuit, we convert it to a ZX-diagram and search over this master action set. For search, we only use GA as maintaining a population of mutants rather than a single state (as in SA) is well-suited for the significantly higher branching factor.

**FIXME:** If have time and space, can extend to the case where the state of GA is both circuit + ZX-diagram so that we can apply non-PyZX simplifications.

# Chapter 4

## Results

Before reporting on the results of our optimization strategies, we first conduct preliminary analyses to refine our strategy. These opportunities for refinement are the following:

- Optimization function
- Probability of applying Equation C1 vs. Equation C2.
- Sampling of subjects (i.e., spiders or pairs of connected spiders) for congruence application

FIXME: For all, we use SA because we expect behavior to be same in GA and SA is more efficient.

### Optimization Function

# Chapter 5

## Application: Quantum Chemistry

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

# Chapter 6

## Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

# Bibliography

- [1] CJ Ballance, TP Harty, NM Linke, MA Sepiol, and DM Lucas. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. *Physical review letters*, 117(6):060504, 2016.
- [2] Earl T Campbell, Barbara M Terhal, and Christophe Vuillot. Roads towards fault-tolerant universal quantum computation. *Nature*, 549(7671):172–179, 2017.
- [3] Ross Duncan, Aleks Kissinger, Simon Perdrix, and John Van De Wetering. Graph-theoretic simplification of quantum circuits with the zx-calculus. *Quantum*, 4:279, 2020.
- [4] Ross Duncan and Simon Perdrix. Graph states and the necessity of euler decomposition. In *Conference on Computability in Europe*, pages 167–177. Springer, 2009.
- [5] Ross Duncan and Simon Perdrix. Pivoting makes the zx-calculus complete for real stabilizers. *arXiv preprint arXiv:1307.7048*, 2013.
- [6] Aleks Kissinger and John van de Wetering. Reducing t-count with the zx-calculus. *arXiv preprint arXiv:1903.10477*, 2019.
- [7] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.