Term Grouping and Travelling Salesperson for Digital Quantum Simulation

Kaiwen Gui[1],* Teague Tomesh[2], Pranav Gokhale[3], Yunong Shi[4],
Frederic T. Chong[3], Margaret Martonosi[2], and Martin Suchara[1,5]
[1] *Pritzker School of Molecular Engineering, University of Chicago*
[2] *Department of Computer Science, Princeton University*
[3] *Department of Computer Science, University of Chicago*
[4] *Department of Physics, University of Chicago and*
[5] *Mathematics and Computer Science Division, Argonne National Laboratory*

(Dated: January 17, 2020)

Digital simulation of quantum dynamics by constructing the time evolution of a Hamiltonian is the initially proposed application of quantum computing. The large number of quantum gates required for emulating the complete second quantization form of the Hamiltonian, however, makes such an approach unsuitable for near-term devices with limited gate fidelities. In addition, Trotter error caused by noncommuting terms can accumulate and harm the overall circuit fidelity. In this paper, we propose two integrated techniques that address these problems. First, we improve the Trotter fidelity compared with previously proposed optimization by reordering Pauli terms and partitioning them into commuting families. We demonstrate the practicality of this method by constructing and evaluating quantum circuits that simulate different molecular Hamiltonians. We also provide theoretical explanations for the fidelity improvements provided by our term grouping method. Second, we describe a new gate cancellation technique that reduces the high gate counts by formulating the gate cancellation problem as a travelling salesperson problem.

I. INTRODUCTION

The idea of simulating unknown quantum systems by using controllable quantum systems was originally proposed by Richard Feynman in the 1980s [1] and has remained one of the major motivations for building quantum computers. Similar to some other quantum algorithms, it can provide exponential speed-ups compared with classical simulation of quantum systems [2].

Digital quantum simulation (DQS) aims to map a quantum evolution, defined by a time-dependent Hamiltonian, to a digital quantum circuit. The digital quantum circuit is executed with respect to some evolution time in order to "emulate" the real Hamiltonian evolution. Typically, one will initialize the Hamiltonian using the second quantization form and convert to the qubit representation using mapping methods such as Jordan-Wigner [3] or Bravyi-Kitaev [4]. Then the resulting Pauli terms, shown in the "Qubit Representation" in Figure 1, are mapped to the digital quantum circuit (more details are given in Section II).

Unlike near-term quantum simulation methods such as VQE [5], the naïve presentations of the quantum dynamics algorithm have extremely high depth (e.g., 10^7 gates [6]). This is due to the fact that Pauli terms need to be concatenated in the same quantum circuit. Details about DQS circuit concatenations are provided in Section IV. This high circuit depth is unsuitable for near-term quantum devices with limited coherence times of hardware qubits[7] as well as future devices that will consists of qubits with improved but non-negligible error properties.

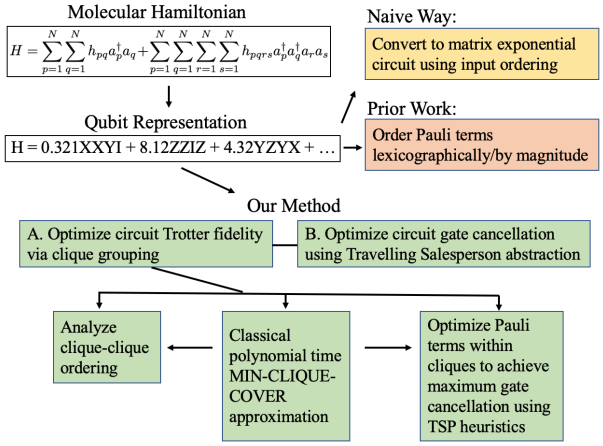

Figure 1. Summary of our method

In addition to the large gate depth, DQS circuits suffers from large Trotter errors. In general, the Pauli terms do not commute with each other. Therefore, we can only map the Hamiltonian into an approximated circuit by using the Trotter-Suzuki decomposition [8]. This will lead to Trotter errors that can devastate the total circuit fidelity even under perfect quantum gates [2]. Details about Trotter errors are provided in Section IV D. We denote the fidelity of DQS under perfect physical gates as "Trotter fidelity."

In this work, we propose a new term-ordering strategy. Assuming that all quantum gates are perfect (no dissipation or crosstalk), this strategy minimizes the Trotter errors. Figure 1 illustrates the key steps of our method.

- We first group the Pauli terms into commuting groups that aim to provide better Trotter fidelity.

* Corresponding author: kgui@uchicago.edu

Next, we optimize the ordering of Pauli terms inside each commuting group to achieve maximum gate cancellation. We can reorder the terms inside each commuting group because reordering these Pauli terms will not harm the Trotter fidelity (Corollary 2 in Section VIII).

The contributions of this work are as follows:

- We describe a new term-grouping method that improves circuit fidelity (Section V). We also describe a theoretical explanation in Sections VI, and provide an experimental demonstration, VII. This contribution partially overlaps with some concurrent work [9, 10].
- We analyze the impact of clique-clique ordering on the total circuit fidelity for two and more commuting groups, and perform experimental demonstrations (Section VII). This contribution differs from previous and concurrent work which only considered ordering of two commuting groups.
- We use the traveling salesperson problem (TSP) for gate cancellation abstraction, with analysis and initial implementations (Section VIII and IX). We are the first to propose this TSP strategy on DQS circuits.

Although we are using the same term-finding technique that was described in our earlier work that optimizes VQE circuits [11], here it serves a different purpose. In our earlier work, grouping the Pauli terms into cliques was done in order to reduce the total number of required measurements. In this work, we utilize the clique grouping to provide better quantum dynamics circuit fidelity and a more suitable platform for gate cancellation using TSP.

The role of noise will be considered in follow-up work. In particular, different levels or different types of noise models would impact the actual optimization strategy. For instance, a larger physical error would likely imply that we should perform the TSP gate cancellation on the entire quantum circuit instead of inside each commuting group, since gate cancellation will be prioritized to minimize physical errors on top of minimizing Trotter errors.

II. BACKGROUND

The simulation of quantum systems by using programmable quantum computers is one of the most promising applications of quantum computing. Understanding the evolution of a quantum state under some Hamiltonian, H,

$$|\psi(t)\rangle = e^{-iHt/\hbar}\,|\psi(0)\rangle \tag{1}$$

is a problem of great interest in chemistry and physics. Classically, the molecular Hamiltonian for electronic structure problems can be exactly simulated by using full configuration interaction with a runtime that scales as a factorial of the number of basis functions [12]. The same problem can be solved in polynomial time on a quantum computer by leveraging quantum phenomena [13].

In general, simulation of a molecular system consists of three integrated steps: state preparation, time evolution, and measurement of observables [13].

In this paper, we focus on the second step: proposing techniques for improving the fidelity and reducing the depth of the quantum circuits used to implement the evolution unitary (Eq. (1)). We demonstrate the improved fidelity by examining the simulated quantum operation matrix, explained in Section VII A.

A. Quantum Gate Basics

In DQS circuits, we use the following quantum gates, represented by their unitary matrices:

Gate	Unitary
Hadamard	$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$
Z-Rotation	$Rz(\theta) = \begin{pmatrix}e^{-i\frac{\theta}{2}} & 0\\ 0 & e^{i\frac{\theta}{2}}\end{pmatrix}$
Phase	$S = \begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix}$
Controlled-Not	$CNOT = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}$

B. Quantum Dynamics Circuit

In general, the quantum evolution unitary $e^{-iHt/\hbar}$ is complicated and hard to implement by using quantum circuits. Fortunately, in electronic structure problems, there is a relatively easy way [14–16] to implement the summation of the Pauli terms that come after the Jordan-Wigner or Bravyi Kitaev transformations.

For general electronic structure problems, the molecular Hamiltonians are initialized with the form

$$H = \sum_{p=1}^{N}\sum_{q=1}^{N} h_{pq} a_p^\dagger a_q + \sum_{p=1}^{N}\sum_{q=1}^{N}\sum_{r=1}^{N}\sum_{s=1}^{N} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s. \tag{2}$$

We first prepare the initial imperfect coefficients h_{pq} and h_{pqrs} using classical algorithms such as Hartree-Fock. We use OpenFermion [17] to obtain these coefficients.

Next, we map the fermionic operators $a_p^\dagger$ and a_q to the qubit representation. This process can be done via transformation methods such as Jordan-Wigner or Bravyi-Kitaev [18]. We denote the resulting strings of the form

$\{X, Y, Z, I\}^{\otimes n}$ as "Pauli terms." We also used OpenFermion to obtain these Pauli terms [17].

The Pauli terms are then mapped to the quantum circuits. Consider a simple simulation of the unitary evolution $U = e^{-iZt}$ (Z is the single Pauli Z matrix). It can be mapped to the single Rz(2t) gate (t is the time parameter) defined in Section II A.

Accordingly, the unitary evolution $U = e^{-iZZt}$ and $U = e^{-iZZZt}$ can be evaluated with the circuits in Figs. 2 and 3.

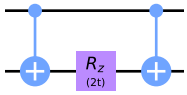

Figure 2. $U = e^{-iZZt}$ circuit

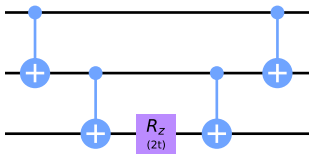

Figure 3. $U = e^{-iZZZt}$ circuit

Similarly, the Pauli X matrix can be simulated by adding Hadamard gates to the front and back of the circuits. Figure 4 provides an example of the $U = e^{-i(XX)t}$ evolution. The Pauli Y matrix can be simulated by

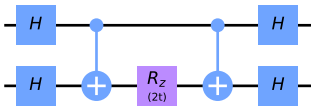

Figure 4. $U = e^{-iXXt}$ circuit

adding the Hadamard gate and the S gate in the front and then Hadamard and the $S^\dagger$ gate at the end. Figure 5 provides an example of the circuit representing $U = e^{-i(YY)t}$. The coefficients in front of each Pauli

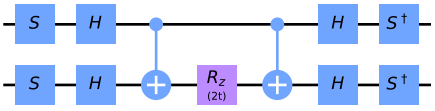

Figure 5. $U = e^{-iYYt}$ circuit

term are multiplied with 2t in the Rz gate's parameter.

C. Travelling Salesperson Problem

The travelling salesperson problem (TSP) aims to solve the following question:

- Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city and returns to the origin city?

TSP can be modeled as an undirected weighted graph, such that cities are the graph's vertices, paths are the graph's edges, and a path's distance is the edge's weight. It is a minimization problem starting and finishing at a specified vertex after having visited each other vertex exactly once. In another words, it is trying to find a "Hamiltonian cycle" with the lowest total weight [19].

In the optimization problems we study, we focus on the related problem of finding "Hamiltonian paths" instead. Similar to the TSP problem, we seek a path in an undirected weighted graph that visits each vertex exactly once but does not need to start and finish at the same vertex.

In Section IX we describe how the vertices, edges, and weights in the original Hamiltonian path problem map to our problem of interest: optimization for gate cancellation to get the minimal total number of gates.

The exact solution of TSP is is NP-hard and we cannot find a solution in polynomial time (unless P = NP) [19]. In Section IX A we discuss TSP heuristics that can approximate the solution in polynomial time.

III. PREVIOUS WORK

Three major term-ordering techniques have been proposed by previous work: *lexicographic* ordering, *interleave* ordering, and *magnitude* ordering.

In particular, *lexicographic* ordering aims to provide a better solution for quantum gate cancellations. *Interleave* ordering and *magnitude* ordering, on the other hand, are more concerned with reducing Trotter errors, and the two share some similarities. In later sections, we will demonstrate that our grouping method can match the Trotter fidelities provided by *magnitude* ordering and can achieve a greater gate cancellation factor than *lexicographic* ordering does.

A. Lexicographic Term Ordering

Prior work has proposed the *lexicographic* ordering to achieve better gate cancellation. The idea is to order the Pauli terms in alphabetic order. In particular, Hastings et al. [20] propose lexicographic order at the level of fermionic operator strings (which is specific to molecular simulation use cases). Tranter et al. [6] broaden the lexicographic treatment to operate at the level of Pauli strings, thus broadening the applicability beyond molecular chemistry tasks.

For instance, assuming X < Y < Z < I, the ordering for the following unordered string collection {YYYYYXXZ, XXXXXYYX, YYYYYZII, XXXXXIXZ} would be XXXXXYYX < XXXXXIXZ < YYYYYXXZ <

YYYYYZII. Then the corresponding quantum circuit will be constructed by concatenation of the Pauli terms in this order. We see that the higher-order characters are grouped next to each other. When applied to the time evolution quantum circuits, this ordering can provide many CNOT gates cancellations. In contrast, the original ordering would have no possible neighboring-term gate cancellation since all neighboring characters are different from each other. Using *lexicographic* ordering, we can group the X's next to each other and Y's next to each other, resulting in many gate cancellation.

In general, *lexicographic* ordering can provide a good level of gate cancellation since similar Pauli characters are grouped together, which results in some similarities in the control and target qubits of CNOTs to cancel. Moreover, *lexicographic* ordering can provide some improvement of Trotter fidelity, since similar terms imply similar physical properties.

B. Interleave Term Ordering

Hastings et al. [20] also propose the *interleave* algorithm, which orders the Pauli terms in the following way, based on intuition from physics:

Algorithm 1: *Interleave* Ordering

Input: Molecular Hamiltonian in the Fermionic Representation
Result: Mapped Quantum Circuit in the Optimized Fermionic Ordering
1. Execute all H_{pp} and H_{pqqp} terms
2. **for** p, q **do**
   a. Execute H_{pq} and all H_{prrq} terms
end
3. Execute all H_{pqrs} terms

The *interleave* ordering improves Trotter fidelity because each step groups the terms with similar physical properties (this similarity is greater than what is provided by *lexicographic* ordering) and all terms in lines 1 and 2a commute with each other, under the following condition [20]:

$$t_{pq} + \frac{1}{2}\sum V_{prrq} n_r = 0. \tag{3}$$

C. Magnitude Term Ordering

Prior work has also has considered sorting the Pauli terms according to the magnitude of their coefficients [6, 20], namely, the absolute value of h_{pq} and h_{pqrs} defined in Eq. 2.

The idea is that sorting the Pauli terms with similar coefficients magnitude will imply that Pauli terms with similar physical properties will be grouped together, thereby leading to fewer Trotter errors.

For instance, in [20], Hastings et al. observe that for the particular molecules they studied, the coefficient magnitude has the following ordering:

$$h_{pp} > h_{pqqp} > h_{pq} > h_{pqqr} > h_{pqrs}. \tag{4}$$

Therefore, the *magnitude* ordering likely can produce orderings similar to those of the *interleave* ordering method.

IV. MULTITERM CIRCUIT CONSTRUCTION

In this section, we make several observations that justify the validity of using TSP gate cancellation (Section IX) after the term-grouping preprocessing. This also provides some insight into the *group-commutation* ordering strategy we discuss in more detail in Section V.

A. Quantum Evolution Primitive

Observation: For unitary evolution of the form $U(t) = U_1(t)U_2(t)$, suppose we have a quantum circuit A that efficiently simulates $U_1(t)$ and another quantum circuit B for $U_2(t)$. We can efficiently simulate $U(t)$ by concatenating circuit B and circuit A. For example, the evolution $U = e^{-i(aXX)t}e^{-i(bZZ)t}$ can be simulated by concatenating the $e^{-i(aXX)t}$ and $e^{-i(bZZ)t}$ circuits.

B. Pairwise Commutation Circuit

Theorem 1: For Hamiltonians of the form $H = H_1 + H_2$, if $[H_1, H_2] = 0$, we have

$$e^{-i(H_1+H_2)t} = e^{-iH_1 t}e^{-iH_2 t}.$$

Using Theorem 1, we can effectively construct the quantum circuit that simulates any two equal-sized Pauli terms that commute with each other. Figure 6 shows the circuit simulating the $U = e^{-i(aXX+bZZ)t}$ evolution.

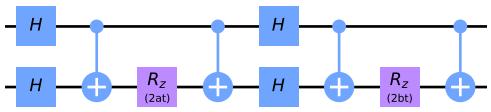

Figure 6. $U = e^{-i(aXX+bZZ)t}$ circuit

C. Group Commutation Circuit

Theorem 2: For an arbitrary Hamiltonian of the form

$$H = \sum_{n=1}^{M} H_n,$$

if

$$[H_j, H_k] = 0, \forall j, k \in [M],$$

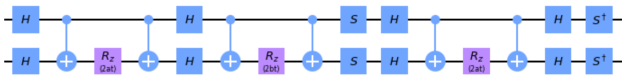

Figure 7. $U = e^{-i(aXX+bYY+cZZ)t}$ circuit

then

$$e^{-iHt} = e^{-iH_1 t} e^{-iH_2 t} ... e^{-iH_M t}, \forall t. \quad (5)$$

Corollary 2: For any $k, l \in \{1, ..., n\}$, if $[H_i, H_j] = 0$ $\forall i, j \in \{1, ..., n\}$, we have

$$e^{-iH_1 t} ... e^{-iH_k t} ... e^{-iH_l t} ... e^{-iH_n t} = e^{-iH_1 t} ... e^{-iH_l t} ... e^{-iH_k t} ... e^{-iH_n t}. \quad (6)$$

Theorem 2 is a powerful tool. If we can find a group of Pauli terms that mutually commute with all other terms in the group, we can construct the exact quantum circuit without any loss in Trotter fidelity. Therefore, it allows us to fully utilize the result we get from the classical MIN-COMMUTING-PARTITION preprocessing (detailed explanation in the next section).

As a simple example, we know that [XX, YY] = 0, [YY, ZZ] = 0, and [XX, ZZ] = 0. Therefore, we can construct the exact quantum circuit mapping of $U = e^{-i(aXX+bYY+cZZ)t}$ by concatenating them one after another, as shown in Fig. 7. Corollary 2 then allows us to freely order the Pauli terms with a particular commuting partition without losing the overall Trotter fidelity. We can then formulate the problem of finding the optimal ordering inside each group as a TSP problem.

D. Trotterization Circuit

If the Pauli terms do not commute, we need to apply Trotterization/Trotter decomposition, justified by Lloyd [2]:

$$e^{-iHt} \approx (e^{-iH_1 t/n} e^{-iH_2 t/n} ... e^{-iH_m t/n})^n,$$

where

$$H = \sum_{i=1}^{m} H_i.$$

More precisely [13],

$$e^{-iHt} = (e^{-iH_1 \Delta t} e^{-iH_2 \Delta t} ... e^{-iH_m \Delta t})^{t/\Delta t} + O(t\Delta t). \quad (7)$$

We denote $\frac{t}{\Delta t} = r$, also called the "Trotter number."

The approximation can be made tight by limiting the time step and Trotter number. The reason is that we have

$$O(t\Delta t) = O(\frac{1}{r} t^2).$$

As $r \to \infty$ and $t \to 0$, the error in the approximation vanishes.

Higher-order approximations, such as the second-order Suzuki-Trotter approximation, also exist that further reduce the error. For demonstration purposes, we only implemented the first-order Trotter decomposition in our experimental simulation and theoretical examination. However, our techniques also apply for higher-order decompositions.

Trotterization Circuit Example

Consider the following simple Hamiltonian for the deuteron [21].

$$H = 5.907II + 0.2183ZI - 6.125IZ - 2.143XX - 2.143YY$$

Naïvely, we would apply Trotterization to all the Pauli terms. Following Eq. 7, we set the Trotter number to 4. We then have

$$e^{-iHt} \approx (e^{\frac{-0.2183iZIt}{4}} e^{\frac{6.125iIZt}{4}} e^{\frac{2.143iXXt}{4}} e^{\frac{2.143iYYt}{4}})^4.$$

This corresponds to the quantum circuit in Fig. 8.

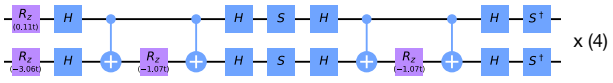

Figure 8. Quantum circuit simulating the quantum dynamics of deuteron, naïvely partitioned. The entire circuit consists of the shown portion repeated 4 times.

V. GROUP-COMMUTATION TERM ORDERING

In this section, we describe our *group-commutation* ordering strategy that aims to achieve better total circuit fidelity and thus provide a desirable precondition for the TSP gate cancellation described in Section IX.

Intuitively, we know that if the commuting Pauli term pairs are grouped together in mutually commuting partitions, the Trotter errors can potentially be reduced. In other words, we are trying to first create the kind of concatenated sub-circuits that contain as much of the pairwise commuting Pauli terms as possible, shown in Section IV C. Then we concatenate those sub-circuits one after another to construct the entire quantum circuit.

This intuition is explained more rigorously in Section VI. In Section VII, we will experimentally demonstrate that the *group-commutation* ordering is able to match the Trotter fidelity with that of *magnitude* ordering, and provide much better Trotter fidelity than *lexicographic* ordering does.

More importantly, not only can this *group-commutation* method provide good fidelity, but it

also has another benefit: after we arrange the Pauli terms into groups, we can freely arrange the term orderings in each group without affecting the fidelity (with the help of Corollary 2 in Section IV). Then we can consider the optimal ordering that maximizes gate cancellation inside each group.

A. Problem Abstraction

With this idea in mind, we thus abstract the problem as a graph problem. All the Pauli terms are represented as graph vertices, and an edge between a pair of vertices is introduced for commuting Pauli terms [11].

Since the pairwise commutativity is not affected by the coefficients (while computing the term commutators, the coefficients are cancelled) and since the coefficients affect only the parameter of the Rz gates in the circuit mapping, we drop the coefficients of the Pauli terms in the initial grouping step (but will consider them later in the clique-clique ordering).

Consider the Hamiltonian for the H_2 molecule as an example. We obtain the following molecular Hamiltonian after applying the Jordan-Wigner transformation, and ignoring the IIII term that corresponds to a trivial circuit.

$$
\begin{aligned}
H = \; & 0.0871IIIZ - 0.0243IIZI + 0.0871IZII \\
& - 0.0243ZIII + 0.0785IIZZ + 0.135IZIZ \\
& + 0.0590XXYY + 0.0590YYYY + 0.0590XXXX \\
& + 0.0590YYXX + 0.138ZIIZ + 0.138IZZI \\
& + 0.143ZIZI + 0.0785ZZII
\end{aligned}
$$

This Hamiltonian is represented by the graph abstraction in Figure 9. Note that the coefficients are not captured by the graph representation and they are not needed to determine commutativity relations. In this specific example, the 14 Pauli terms are separated into two fully connected subgraphs, or cliques. Then, when mapping the problem to a quantum dynamics circuit, we can first concatenate all the IIIZ, IZIZ, ... terms in the blue colored clique, and then concatenate all the XXYY,YXXY terms in the red colored clique.

B. Clique-Finding Algorithm

We use the clique finding algorithm that we described in our earlier work [11, 22]. Given a Hamiltonian whose terms are a set of Pauli strings, we wish to partition the terms into commuting families where the number of partitions is minimized.

After representing the Hamiltonian as the graph described above, we then use the recursive Bron-Kerbosch [23] algorithm to find MIN-CLIQUE-COVERs on this graph. As presented in our previous work [11], algorithms such as Boppana-Halldórsson [24] are able to provide polynomial time MIN-CLIQUE-COVER heuristics.

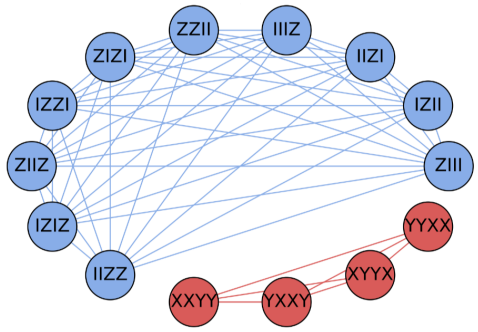

Figure 9. H_2 graph abstraction

C. Possible Grouping Algorithm Modifications or Enhancements

Since the term-grouping method so far does not consider any physical properties of the Hamiltonian, we can consider the following modifications and enhancements that contain physics intuitions.

1. *Considering Magnitude Factors*

We can consider putting the Pauli terms with similar scale of the coefficients magnitude in the same clique. For instance, we can make the original unweighted graph into a weighted one (e.g., the edge weight is the product of the coefficients magnitude of the two Pauli terms it connects) and group together the pairwise commuting Pauli terms with similar coefficients.

2. *Considering Physical Property Factors*

We can also apply the *group-commutation* technique in the *interleave* ordering. In the original *inteleave* method (Section III B), step 3 treats all H_{pqrs} terms with *lexicographic* ordering. In our case, we order the H_{pqrs} terms into cliques. One can also investigate whether the groups can be expanded to larger cliques and better clique orderings in step 2.

VI. THEORETICAL ANALYSIS ON GROUP-COMMUTATION ORDERING

In this section, we will justify the intuition that *group commutation* ordering can potentially provide better

Trotter fidelity than naive ordering for general Hamiltonians. In the next section, we will experimentally verify this intuition.

We analyze the effectiveness of our group commutation ordering strategy through the example of Hamiltonians with two commutation groups. The analysis can be generalized to more commutation groups by the reader. Consider a Hamiltonian $H = \sum_{i=1}^{k} \alpha_i H_i$, where a_i are real numbers and H_i are simple Hamiltonians that can be mapped to quantum circuits (or diagonalized) directly. Suppose H can be divided into two commuting groups (cliques) $H_1^c = \sum_{m=1}^{p} H_m$ and $H_2^c = \sum_{m=p+1}^{k} H_m$, *i.e.*, $[H_m, H_n] = 0$ if $0 < m, n \leq p$ or $p < m, n \leq k$.

We compare the Trotter error of group commutation ordering and other orderings.

With group commutation ordering

It has been shown that [25–27], for H, the approximation error (in the additive form) of the Lie-Trotter formula is given by the variation-of-parameters formula,

$$\begin{aligned}\delta_{gc} &= e^{-it(H_1^c)}e^{-it(H_2^c)} - e^{-itH} \\ &= \int_0^t d\tau e^{-i(t-\tau)H}[e^{-i\tau H_1^c}, H_2^c]e^{\tau H_2^c}. \end{aligned} \tag{8}$$

Note that the error form also applies to general H^g where $H^g = H_1^g + H_2^g$. For simplicity, we denote the integral in Eq. 8 as $I(H_1^g, H_2^g)$ for general H_1^g, H_2^g. Thus, we can simply write the approximation error as $\delta_{gc} = I(H_1^c, H_2^c)$.

We are interested primarily in the operator norm (*i.e.*, spectral norm) of $||\delta_{gc}|| = ||I(H_1^c, H_2^c)||$, which gives the worst-case analysis of the error.

Without group commutation ordering

We can recursively apply the error formula in Eq. 8 to the Lie-Trotter formula of arbitrary ordering for H. Let π be a permutation of the set $\{1, .., p\}$ that defines the ordering. First, we can approximate e^{-itH} by separating $H_{\pi(1)}$ from other terms:

$$\begin{aligned}\delta_1 &= e^{-itH_{\pi(1)}}e^{-it\sum_{m=2}^{k} H_{\pi(m)}} - e^{-itH} \\ &= I(H_{\pi(1)}, \sum_{m=2}^{k} H_{\pi(m)}).\end{aligned}$$

Then we can recursively repeat the process for the rest of the Hamiltonian and arrive at the following expression for the approximation error δ of the Lie-Trotter formula.

$$\begin{aligned}\delta_{ngc} &= e^{-itH_{\pi(1)}}e^{-itH_{\pi(2)}}...e^{-itH_{\pi(2)}} - e^{-itH} \\ &= I(H_{\pi(1)}, \sum_{m=2}^{p} H_{\pi(m)}) + e^{-itH_{\pi(1)}} I(H_{\pi(2)}, \sum_{m=3}^{p} H_{\pi(m)}) \\ &+ e^{-itH_{\pi(1)}}e^{-itH_{\pi(2)}} I(H_{\pi(3)}, \sum_{m=4}^{p} H_{\pi(m)}) + ... \\ &+ e^{-itH_{\pi(1)}}...e^{-itH_{\pi(j)}} I(H_{\pi(j+1)}, \sum_{m=j+2}^{p} H_{\pi(m)}) + ...\end{aligned}$$

Using the triangle inequality and the submultiplicativity of the operator norm, together with the fact that the operator norm of a unitary is 1, we have

$$\begin{aligned}||\delta_{ngc}|| \approx ||I(H_{\pi(1)}, \sum_{m=2}^{p} H_{\pi(m)}) + I(H_{\pi(2)}, \sum_{m=3}^{p} H_{\pi(m)}) \\ + ... + I(H_{\pi(j+1)}, \sum_{m=j+2}^{p} H_{\pi(m)}) + ...||.\end{aligned}$$

Also we know that $H_{\pi(k)}$ is either in H_1^c or H_2^c. Thus, we have

$$\begin{aligned}||\delta_{ngc}|| \approx || \sum_{\pi(j)\in[1,p]} I(H_{\pi(j)}, H_2^c)|| + \\ || \sum_{\pi(j)\in[p+1,k]} I(H_{\pi(j)}, H_1^c)||.\end{aligned}$$

Although we have no proof that $||\delta_{gc}|| < ||\delta_{ngc}||$ (because we do not have information about the full commutation relation and magnitude information in H), we can make several observations why group commutation ordering is advantageous. First, $||\delta_{gc}||$ in general has a much lower upper bound than does the first term in $||\delta_{ngc}||$. In fact, the upper bound of $||\delta_{gc}||$ does not scale with the number of terms p and that of $||I(H_{\pi(1)}, \sum_{m=2}^{p} H_{\pi(m)})||$ is of $O(p)$. Second, $||\delta_{gc}||$ does not include the second term in $||\delta_{ngc}||$. Thus, there is strong evidence that *group commutation* ordering has an advantage over naive ordering in terms of eliminating Trotter errors.

In the next section, we will support the above theoretical intuition by implementing real molecular Hamiltonian and test their Trotter fidelities versus other ordering methods such as *lexicographic* and *magnitude* orderings.

VII. EXPERIMENTAL DEMONSTRATION OF GROUP-COMMUTATION ORDERING

In this section, we experimentally demonstrate the fact that *group-commutation* ordering can provide total circuit fidelities similar to the fidelities provided by *magnitude* ordering, thus preparing for the in-clique TSP gate cancellation described in Section IX. Additionally, we show that different permutations of the cliques will lead to different total circuit fidelities.

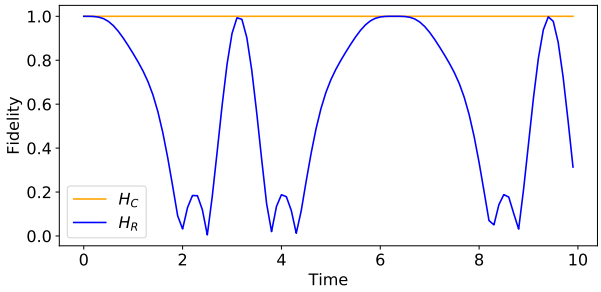

Figure 10. The same Hamiltonian is simulated using two different term orderings H_C and H_R.

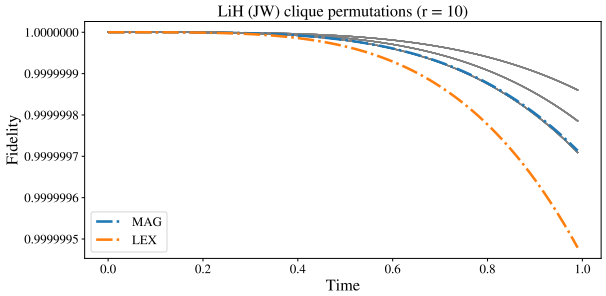

Figure 11. LiH Hamiltonian partitioned into 4 commuting cliques. The resulting 24 permutations for the Hamiltonian are shown as solid lines while the *magnitude* and *lexicographic* orderings are dotted. The "Time" in the x-axis is a unitless parameter. Fidelity = 1 means the result we get from the quantum dynamics circuit matches perfectly the exact matrix exponential it is trying to simulate. A smaller Fidelity value means that the quantum circuit results show more differences from the exact matrix exponential.

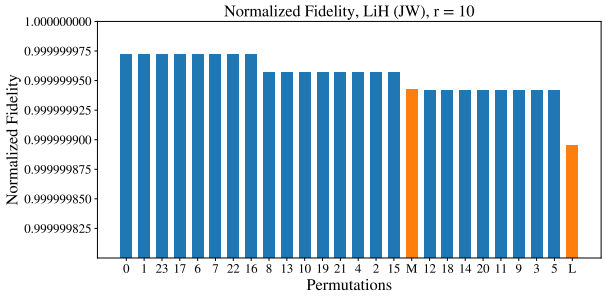

Figure 12. Normalized fidelity as defined by Eq. 10 for various permutations of the Hamiltonian. Fidelities of random permutation orderings are shown in blue and Fidelities of the *magnitude* and *lexicographic* orderings are shown in orange.

A. Experimental Demonstration Method

To show that the group commutation circuit can provide better fidelity, we construct an example DQS circuit to simulate a simple 2-qubit Hamiltonian with its terms sorted according to group commutation,

$$H_C = IZ + ZI + ZZ + XX + YY,$$

and with terms sorted in some random order,

$$H_R = XX + ZI + YY + IZ + ZZ.$$

Our fidelity metric measures how closely the DQS circuit's unitary matrix, U_{approx}, matches the exact unitary evolution matrix, $U_{exact} = e^{-iHt}$,

$$\mathcal{F} = \frac{|\mathrm{Tr}(U_{exact} U_{approx}^{\dagger})|}{\dim(U_{exact})} \tag{9}$$

The result is plotted in Fig. 10, with different time iteration t values. The orange curve (group ordering) is a straight line that achieves perfect fidelity matching the exact unitary evolution. On the other hand, the random ordering fidelity represented by the blue curve oscillates between 0 and 1.

The group ordering places the two all-pairwise commuting groups of Pauli terms $\{XX, YY, ZZ\}$ and $\{ZI, IZ\}$ together. We also observe that $[(XX + YY + ZZ), (ZI + IZ)] = 0$. Therefore, applying Theorem 2 twice, we can explain why the resulting circuit achieves perfect fidelity.

B. Benchmarking of Molecular Hamiltonian

For a typical Hamiltonian, the groups of Pauli terms do not usually commute with each other with coefficients added. Thus, in the general case we are unlikely to achieve the perfect fidelity as depicted in Fig. 10.

We examined three molecules—H_2, LiH, and H_2O—with Hamiltonian sizes ranging from 14 to 61 terms. For each of these Hamiltonians, we first partitioned the terms into n maximally commuting cliques and then enumerated all $n!$ permutations of the cliques. For each permutation we constructed a DQS circuit to simulate the Hamiltonian evolution and computed this circuit's fidelity according to Equation (9), with time iteration step t = 0.01 and Trotter number r = 10. Additionally, we found the fidelity given by *magnitude* and *lexicographic* orderings of the Hamiltonian. Figures 11 and 12 show the results of this benchmark for the LiH molecule whose Hamiltonian contains 26 terms that can be grouped into 4 commuting cliques. The normalized fidelity values in Figure 12 were computed (using the fidelities $\mathcal{F}(t)$ in Figure 11) according to

$$\frac{\int_0^{t'} \mathcal{F}(t) dt}{\int_0^{t'} dt}. \tag{10}$$

Figure 12 shows the fidelities of different permutations (blue) as well as the fidelities of *magnitude* and *lexicographic* orderings (orange). Note that the fidelity differences in Figure 12 are small. We use these benchmark

Table I. Fidelity comparison of *magnitude* ordering vs. *group-commutation* ordering, across different molecules and active spaces (t = 0.01, r = 10). The performance of group-commutation vs *magnitude* orderings varies but is close to each other. Both are much better than *lexicographic* ordering in providing better Trotter fidelity.

Molecule	Encoding	Active Spaces	# of Terms	# of Cliques	# of Perms	$\frac{1-\mathcal{F}(bestperm)}{1-\mathcal{F}(MAG)}$	$\frac{1-\mathcal{F}(bestperm)}{1-\mathcal{F}(worstperm)}$	$\frac{1-\mathcal{F}(bestperm)}{1-\mathcal{F}(avgperm)}$
LiH	JW	4	26	4	24	0.485	0.478	0.648
LiH	BK	4	26	4	24	0.286	0.379	0.555
H2	JW	4	14	2	2	1.00	1.00	1.00
H2	JW	6	61	6	720	1.01	0.240	0.413
H2O	BK	6	61	6	720	1.30	0.456	0.725

results to show that the best clique-clique permutations are able to achieve fidelities on a par with the *magnitude* ordering. Table I contains similar information for all the molecules and Hamiltonians considered in this benchmark.

C. Result Discussion

Although both the *group-commutation* and *magnitude* orderings achieve similar fidelities, the number of possible gate cancellations differs greatly between the two ordering schemes. The *group-commutation* ordering is especially well suited for gate optimizations because of the freedom it provides for term orderings within each clique. In Section IX we describe how we take advantage of this fact to order terms within cliques, which is optimal for gate cancellations. On the other hand, a *magnitude* ordering is determined by term coefficients that may have no bearing whatsoever on the number of gate cancellations between terms, and therefore we do not expect this ordering to benefit from many gate cancellation techniques.

For other physical pictures (i.e., Hamiltonians that describe solid-state structures, high energy physics, etc.), the fidelity comparison would still require investigation. For instance, if we have a Hamiltonian that has Pauli terms all with similar magnitude coefficients, *group-commutation* ordering might offer a much better total circuit fidelity. Moreover, the *magnitude* and *interleave* ordering are based on the Hamiltonians that evolve only the single and double electron integrals defined in Eq. 2. Other Hamiltonians of physical systems with different quantization structures are likely to have different results.

The next question is how to arrange the clique sequences and whether there is a way we can optimize their order. Follow-up work will continue to explore this question.

VIII. PAIRWISE GATE CANCELLATION TRICKS AND RESTRICTIONS

In this section we first provide some of the important gate cancellation techniques and restrictions. Then we show a useful "*star + ancilla*" trick that fully utilizes the gate cancellation techniques and at the same time is able to bypass the restrictions. The next section (Section IX will describes how we actually formulate these tricks and restrictions, together with the *star + ancilla* trick, as a TSP problem, after we group the Pauli terms into cliques using the *group-commutation* ordering described in Section V.

A. Gate Cancellation Techniques

We observe that many cancellations of the gates specified in Section II A are possible in the quantum dynamics circuit.

1. Cancellation of Single-Qubit Gates

First, we observe that two Hadamard gates acting on the same qubit consecutively can cancel each other since $HH = I$. This will happen when two Pauli X characters or an X and a Y character are next to each other.

We can also cancel the following gate sequence when two Pauli Y characters are next to each other, since $S^\dagger HHS = S^\dagger S = I$.

2. Cancellation of Multiqubit Gates

Next, we observe that two ideal CNOT gates acting on the same control and target qubits can cancel each other. We can prove this cancellation by multiplying two CNOT matrices.

$$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix} * \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix} = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} = I$$

In addition, when other gates are in between the CNOT gate pairs, we can do a "swap & cancel" if the CNOT gate matrix commutes with the other gate matrices. The commutation relation does not always hold, however, as we explain next.

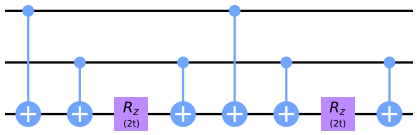

Figure 13. *Star* implementation original circuit

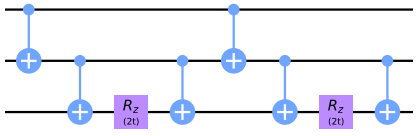

Figure 14. *Ladder* implementation original circuit

B. Gate Cancellation Restrictions

Certain gate orderings can prevent the gate cancellations describe above. We identify several kinds of these restrictions below.

1. Restriction I

Two ways of implementing the CNOT entangler exist. The first is to put the target qubits of CNOT gates on all the qubits, as used in the quantum dynamics circuit construction in Section II and IV. We denote this the "*ladder*" implementation. The second way is to put the target qubits of CNOT gates on one qubit. We denote this the "*star*" implementation. The two implementations will yield identical outputs.

Consider the following Hamiltonian $H = ZZZ + ZZI$ as an example. Its two implementations are shown in Fig. 13 and Fig. 14

We can cancel the third and fifth CNOT gates in the *star* implementation (Fig. 15) but not in the *ladder* implementation. Canceling the third and fifth CNOT gates in the *ladder* implementation would result in a different quantum circuit output.

2. Restriction II

Sometimes when Hadamard gates exist between the CNOT sequences (resulting from X or Y characters in the Pauli terms), we cannot cancel certain pairs of CNOT gates. The reason is that H and the NOT part of CNOT

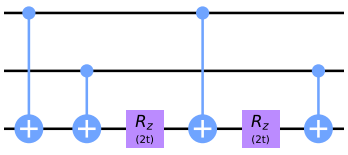

Figure 15. *Star* implementation optimized circuit

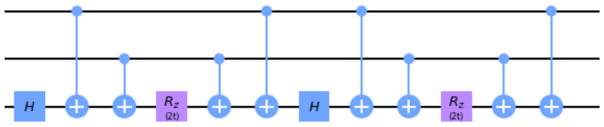

Figure 16. Hadamard restriction

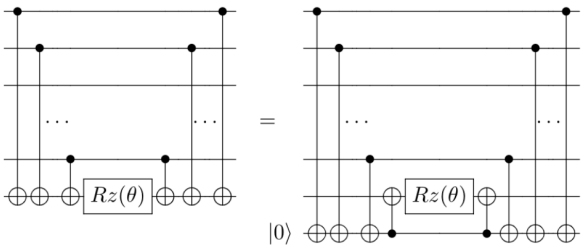

Figure 17. Trick I basic idea

do not commute with each other.

$$[H, X] = HX - XH = \sqrt{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \neq 0$$

The fact that H and X do not commute implies that we cannot simply swap the H and CNOT gates. Therefore, we cannot perform the "swap and cancel" technique in the previous example for the CNOT pairs that have the same control and target qubits.

Consider the following example of the Hamiltonian $H = XZZ + ZZZ$ (Fig. 16). We cannot cancel the third & sixth CNOT and the fourth & fifth CNOT pairs because there is a Hadamard gate in between.

Similarly, if the Hadamard gate is operating on the same qubit as one of CNOT's control qubits, we also cannot swap the Hadamard with the CNOT, because

$$[H \otimes I, CNOT] \neq 0.$$

3. Restriction III

In some cases CNOT target qubits are not the same for different Pauli terms.

Consider the Hamiltonian $H = ZZZZIIII + IIIIZZZZ$ as an example. The target qubit of the $ZZZZIIII$ term can never overlap with the $IIIIZZZZ$ term, and hence no possible CNOT gate cancellation can occur.

C. Star + Ancilla Trick for CNOT Cancellation

One interesting trick presented in [28] suggested that we can move the target qubit of the CNOT entanglers to an additional zero-initialized qubit by adding two CNOT gates (Fig. 17). This trick allows us to bypass all the restrictions described above.

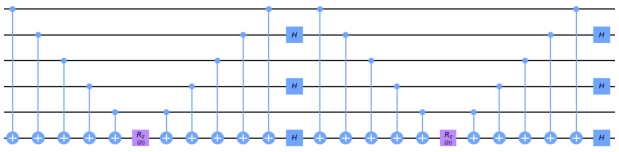

Figure 18. ZZZZZZ & ZXZXZX original *star* implementation

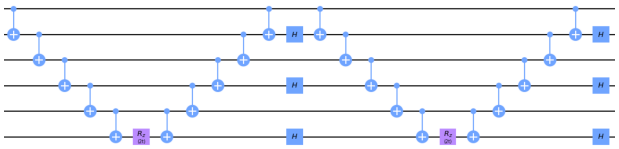

Figure 19. ZZZZZZ & ZXZXZX original *ladder* implementation

1. *Bypassing the Restrictions*

Consider the example of the two Pauli terms ZZZZZZ and ZXZXZX. The original *star* implementation (Fig. 18) is unable to have any CNOT cancellation because there is one Hadamard acting on the sixth/last qubit that blocks any swaps and cancellation of the CNOTs. The CNOT pairs in the original *ladder* implementation (Fig. 19) are also unable to cancel with each other, since there are always Hadamard gates standing on either the target or the control qubit of the CNOTs.

With the help of the trick, we can first modify the *star* implementation to the circuit shown as Fig. 20. Then the middle six CNOT gates are canceled (Fig. 21), thus solving Restriction II. Having the trick also allows us to move all target qubits to the ancilla qubit and thus allows us to do cancellation of CNOT gates. Thus, Restriction III is also solved. Since the trick is a *star* implementation by definition, Restriction I is inherently bypassed.

Although we add four more CNOT gates and one more ancilla qubit in Fig. 20, this eventually leads to six CNOT gate cancellations. In the more general cases where the Pauli terms are very long, this trick can provide greater cancellation numbers, since the additional cost in Fig. 20 is constant (2 additional CNOT gates for each Pauli term and only one ancilla qubit that works for the entire circuit)

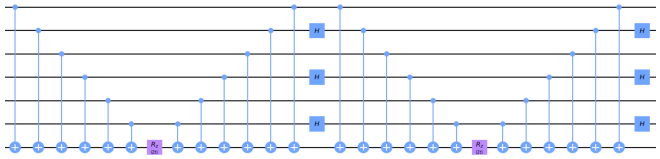

Figure 20. ZZZZZZ & ZXZXZX modified *star* implementation

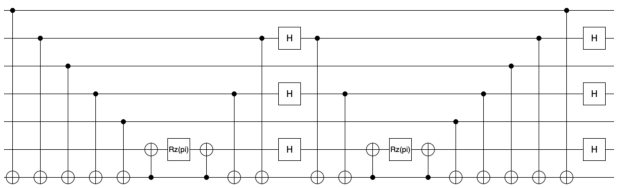

Figure 21. ZZZZZZ & ZXZXZX modified *star* implementation with gate cancellation

IX. GATE CANCELLATION VIA TSP

In this section, we describe a strategy for maximizing gate cancellation in a quantum dynamics circuit, in order to reduce its hardware implementation cost. We focus primarily on the cancellation of two-qubit CNOT gates, which are the dominant source of errors and latency in the prevalent quantum platforms. Typically, CNOT gates have at least 10x lower fidelity and 2–5x longer duration than single qubit gates [29–31]. We also show that TSP provides better gate cancellation results than *lexicographic* ordering under the same circuit Trotter fidelities. Since *magnitude* ordering does not concern the gate cancellation regime, TSP and *lexicographic* perform much better than *magnitude* ordering.

The task we perform is ordering a set of k Pauli strings, $\{P_1, P_2, ..., P_k\}$, to maximize the CNOT gate cancellation in the corresponding quantum dynamics circuit. To simplify our notation in this discussion, we have ignored the coefficients of the Pauli strings, because they do not affect the CNOT cancellation properties. We assume that any permutation of the Pauli strings, $\pi : [1, 2, ..., k] \rightarrow [1, 2, ..., k]$, is valid. This is indeed the case when starting from the *group-commuting* ordering method described in Section V, since the resulting Trotter error is zero, invariant of term ordering (Corollary 2). Alternatively, one may decide that the gate cancellation gains outweigh the Trotter error variances. In that case each permutation could be treated as valid.

Each of the $k!$ permutations corresponds to a quantum circuit that implements

$$\prod_{j=1}^{k} \exp(-iP_{\pi(j)}t).$$

We focus on the "*star + ancilla*" circuit architecture, described in Section VIII C and illustrated in Figure 17), because this circuit architecture simplifies the gate cancellation logic. Each non-I Pauli character originates two identical CNOTs, controlled on the non-I index and targeting the ancilla qubit. These two CNOTs are symmetric around the middle R_z on the ancilla. Without any gate cancellation, we know that the number of CNOT gates required for $\exp(-iP_{\pi(j)}t)$ is

$$2\sum_{j=1}^{k} |P_j|_{\text{Ham}} = 2\sum_{j=1}^{k}\sum_{i=1}^{N} \mathbb{1}_{P_j[i]\neq I}, \tag{11}$$

where N denotes the width of the Pauli strings and Ham refers to Hamming weight.

However, a good permutation π can substantially reduce the CNOT count against Eq. 11's upper bound because the CNOT gates between neighboring $\exp(-iP_{\pi(j)}t)$ subcircuits can cancel each other. Concretely, let us consider two consecutive Pauli strings, $[P_1, P_2]$. We focus on the "transition zone" from $\exp(-iP_1t)$ to $\exp(-iP_2t)$. This spans the CNOT gates between the R_z gates for P_1 and for P_2, and this span is where we can orchestrate gate cancellation. From the previous analysis, prior to gate cancellation, the number of CNOT gates required for the transition zone is

$$|P_1|_{\text{Ham}} + |P_2|_{\text{Ham}}. \tag{12}$$

For two neighboring CNOTs to cancel each other, they must be identical, in other words have the same control and target. Moreover, as described in Section VIII, the single-qubit gates on the controls must also cancel out—this requires the Pauli characters of the controls to be identical. There are no other restrictions, since the shared-target CNOT blocks are order insensitive. Intuitively, this analysis suggests examining the Hamming distance, which reports the number of disagreeing indices:

$$|P_1 - P_2|_{\text{Ham}} = \sum_{i\in[N]} \mathbb{1}_{P_1[i]\neq P_2[i]}. \tag{13}$$

However, this undercounts the number of CNOTs needed. On indices where $P_1[i] \neq P_2[i]$ but neither is I, we see that $P_1[i]$ and $P_2[i]$ will each originate (control) a CNOT, but Eq. 13 only increments by 1 on index i. To adjust this undercounting, we define the corrected CNOT distance as

$$\begin{aligned} |P_1 - P_2|_{\text{CNOT}} &:= |P_1 - P_2|_{\text{Ham}} + \sum_{i\in[N]} \mathbb{1}_{I\ \neq P_1[i]\neq P_2[i]\neq I} \\ &= \sum_{i\in[N]} \mathbb{1}_{P_1[i]\neq P_2[i]}(1 + \mathbb{1}_{I\notin\{P_1[i],P_2[i]\}}). \end{aligned} \tag{14}$$

This distance, which is symmetric in its arguments, reports the number of CNOTs needed to implement the transition zone from $\exp(-iP_1t)$ to $\exp(-iP_2t)$, after performing gate cancellation.

Given this distance, we could in principle find the ordering with minimum CNOT cost by summing the CNOT cost along each of the possible π permutation functions. However, this approach would be impractical, taking $O(k!)$ factorial time. Instead, we can formulate our objective as a TSP problem, in which we are given a weighted graph and asked to find the shortest cycle visiting each vertex once.

This bijects clearly to our setting—we simply define a complete graph G on k vertices and set the distance between vertices i and j as the CNOT cost $|P_i - P_j|_{\text{CNOT}}$. Thus, TSP gives a permutation that cycles through $\{P_1, ..., P_k\}$ with minimal CNOT cost. To be precise, we actually seek the shortest Hamiltonian path; in other words, we want to visit each Pauli string once, without returning to the start. We can do so by creating a virtual node connected to every other node with cost 0, as described by the *computer wiring* bijection in [32]. Algorithm 2 shows pseudocode for the full gate cancellation subroutine.

However, TSP is NP-hard. While it can be solved in $O(k^2 2^k)$ time by the Bellman-Held-Karp dynamic algorithm [33, 34]—an improvement over the $O(k!)$ brute-force solution—it is still exponentially expensive to solve in general. It would be impractical to invest exponential classical resources to optimize a polynomially sized quantum circuit. Fortunately, there are good heuristics and approximation algorithms for TSP, which we discuss next.

Algorithm 2: CNOT Gate Cancellation

Input: Weighted Pauli strings, unordered: $\{P_1, ..., P_k\}$
Result: Permutation approximately maximizing CNOT cancellation: $[P_{\pi(1)}, ..., P_{\pi(k)}]$
finalOrdering $\leftarrow$ [ ];
$G \leftarrow$ undirected complete graph with $k+1$ nodes;
Set $G[k+1, *] = 0$;
for $i, j \in [k] \times [k]$ **do**
  Dist$[i, j] = |P_i - P_j|_{\text{CNOT}}$;
end
Solve TSP on G;
Delete $(k+1)$th vertex in TSP route to break cycle;
Append Hamiltonian path to finalOrdering;
return finalOrdering

A. TSP Approximation

Solving TSP in the most general setting is NP-hard. Moreover, no polynomial time algorithms exists that is guaranteed to approximate it to a constant ratio (unless P = NP) [19]. For the case of *metric graphs*, however, TSP can be efficiently 1.5-approximated via Christofides' algorithm [35], meaning that the approximation will return an ordering that requires at most 1.5x as many CNOTs as the optimal ordering does. Moreover, Christofides' algorithm is fast, running in $O(k^3)$ time as originally proposed and later improved to $\tilde{O}(k^2)$ [36–38]. In practice, Christofides' algorithm, as well as other heuristics that do not have rigorous approximation ratio guarantees, can perform well in practice, often attaining near-optimal solutions [39, 40].

Indeed, the graph defined by the $|P_1 - P_2|_{\text{CNOT}}$ distance function is a metric graph. We have already seen that it is symmetric in its arguments (i.e., the graph is undirected), so we need only show now that it satisfies the triangle inequality:

$$|P_1 - P_2|_{\text{CNOT}} + |P_2 - P_3|_{\text{CNOT}} - |P_1 - P_3|_{\text{CNOT}} \geq 0.$$

We prove this below.

$$\begin{aligned} |P_1 - P_2|_{\text{CNOT}} + |P_2 - P_3|_{\text{CNOT}} - |P_1 - P_3|_{\text{CNOT}} \\ = \sum_{i \in [N]} \mathbb{1}_{P_1[i] \neq P_2[i]}(1 + \mathbb{1}_{I \notin \{P_1[i], P_2[i]\}}) \\ + \mathbb{1}_{P_2[i] \neq P_3[i]}(1 + \mathbb{1}_{I \notin \{P_2[i], P_3[i]\}}) \\ - \mathbb{1}_{P_1[i] \neq P_3[i]}(1 + \mathbb{1}_{I \notin \{P_1[i], P_3[i]\}}) \end{aligned} \tag{15}$$

We will prove that each three-term expression in the sum is non-negative for each i, so that the full sum must also be non-negative. The third term is 0, -1, or -2.

- If it is 0, the three-term expression is already non-negative since the first two terms are non-negative.
- If it is -1, then $P_1[i] \neq P_3[i]$, and one of the two is I. We must also have $P_1[i] \neq P_2[i]$ or $P_2[i] \neq P_3[i]$, so the first two terms must sum to at least 1. Thus, the three-term expression is non-negative.
- If it is -2, then $P_1[i] \neq P_3[i]$, and neither is I. Suppose that $P_1[i] = P_2[i]$; then the first term is +2, and the three-term expression is non-negative. Similarly, if $P_2[i] = P_3[i]$, then the second term is +2, and the three-term expression is non-negative. And if $P_1[i] \neq P_2[i] \neq P_3[i]$, then the first two terms are both at least +1, so the three-term expression is non-negative.

Thus, we conclude that Eq. 15 is a sum over non-negative numbers, so the final result is non-negative. This proves that the triangle inequality holds for the CNOT distance, and thus our graph is metric. Therefore, Christofides' algorithm can be used to efficiently attain a 1.5-approximation to the optimal TSP.

However, this approach is not directly compatible with the shortest Hamiltonian path bijection described previously. This bijection required adding a virtual node connected to every other node with zero cost, which would break the triangle inequality. We instead propose to approximate the shortest Hamiltonian path by first approximating TSP and then deleting the most expensive edge to break the cycle. Note that this patch requires two levels of approximation: first approximating the TSP cycle and then approximating the shortest Hamiltonian path based on this cycle. This approach amounts to modifying Algorithm 2 by (1) initializing G without the $(k+1)$th node, (2) approximating TSP instead of solving it exactly, and (3) deleting the most expensive edge in TSP instead of deleting the $(k+1)$th vertex to realize a Hamiltonian path. There may be more direct approaches to approximating the shortest Hamiltonian path, but TSP is better studied and has more well-known approximation algorithm results.

B. Comparison with Lexicographic Ordering

Lexicographic ordering does often lead to good gate cancellation properties and has the advantage of being extremely fast to compute, for example by a linear-time string sort. However, it does not achieve the optimality of TSP. As an example, consider the 4-qubit Hamiltonian with lexicographically ordered strings [$XXXX$, $XXYY$, $XYXY$, $XYYX$, $YXXY$, $YXYX$, $YYXX$, $YYYY$]. These eight commuting strings arise in the Jordan-Wigner encoding of H_{pqrs} for molecules, so this example is ubiquitous [11, 13]. Applying Eq. 11, we see that $8 \times 2 \times 4 = 64$ CNOTs are needed prior to gate cancellation. After CNOT cancellation, summing Eq. 14 along the lexicographic order gives a total of 40 CNOTs. Now consider the TSP order [$XXXX$, $XXYY$, $XYXY$, $XYYX$, $YXYX$, $YXXY$, $YYXX$, $YYYY$], which flips the third and fourth terms from the lexicographic order. Under TSP, we are able to execute the dynamics with just 36 CNOTs. In summary, unoptimized to *lexicographic* to TSP have CNOT costs of $64 \to 40 \to 36$.

In certain cases, TSP can have an even greater factor of improvement over *lexicographic* ordering. For example, consider the nine Pauli strings below, specially annotated and monospaced for pattern readability.

1. `XX|XXXXXXXX`
2. `XY|XXXZZZZZ`
3. `XZ|XXXXXXXZ`
4. `YX|XXZZZZZZ`
5. `YY|XXXXXXZZ`
6. `YZ|XZZZZZZZ`
7. `ZX|XXXXXZZZ`
8. `ZY|ZZZZZZZZ`
9. `ZZ|XXXXZZZZ`

Without gate cancellation, $9 \times 10 \times 2 = 180$ CNOTs are required. Under the given lexicographic order above, gate cancellation yields 112 CNOTs. However, with reordering into the TSP route, only 62 CNOTs are needed. The TSP route is as follows.

(i) `XX|XXXXXXXX`

(ii) `XZ|XXXXXXXZ`

(iii) `YY|XXXXXXZZ`

(iv) `ZX|XXXXXZZZ`

(v) `ZZ|XXXXZZZZ`

(vi) `XY|XXXZZZZZ`

(vii) `YX|XXZZZZZZ`

(viii) `YZ|XZZZZZZZ`

(ix) `ZY|ZZZZZZZZ`

In fact, this is a special case of an *asymptotic advantage* in gate cancellation that TSP can achieve over *lexicographic* ordering. Consider, a set of $N+1$ Pauli strings, where each Pauli string is the concatenation of two of Pauli substrings, $P_{\text{index}}|P_{XZ(i)}$. $P_{XZ(i)}$, defined for $i \in [N]$, has width of N and consists of $(N-i)$ X's followed by i Z's. Note that $P_{XZ(i)}$ and $P_{XZ(i+\lfloor N/2 \rfloor \mod N)}$ differ by a Hamming distance of $\sim N/2$ and therefore have a CNOT distance of $O(N)$. However, $P_{XZ(i)}$ and $P_{XZ(i)+1}$ have a CNOT distance of $O(1)$ since only one index changes. We denote these two sequences as *unfavorable* and *favorable*, respectively.

An adversarial strategy follows from these sequences. We prefix the $P_{XZ(i)}$'s with P_{index}, of width $O(\log N)$. We assign these indices so that lexicographically sorting the Pauli strings arranges the $P_{XZ(i)}$ in the *unfavorable* sequence where each adjacent pair has $O(N)$ CNOT distance. In this sequence, the total CNOT distance is $O(N^2)$, which has no asymptotic improvement over the no-gate-cancellation baseline. The first list above is lexicographically ordered into the unfavorable sequence—notice that each adjacent $P_{XZ(i)}$ pair has a Hamming distance of 4 or 5.

By contrast, TSP would disregard the P_{index} qubits and instead optimizes the CNOT distances by choosing the favorable sequence. In this ordering, there are a total of $O(N \log N)$ CNOTs along the P_{index} indices and $O(N \times 1) = O(N)$ CNOTs along the $P_{XZ(i)}$ indices, amounting to $O(N \log N)$ CNOTs via TSP. The second list above (with Roman numeral headings) shows the TSP route—note that each adjacent $P_{XZ(i)}$ pair has a Hamming distance of just 1. The bottleneck in CNOT cost therefore originates from the P_{index} substrings, which have only logarithmic width.

While this particular example is pathological, it demonstrates scenarios where *lexicographic* ordering is asymptotically identical to no-gate cancellation, but TSP achieves an asymptotic advantage. This advantage is from $O(N^2)$ to $O(N \log N)$, which is a significant improvement. For future work, we will evaluate empirical improvements in terms of the gains both from TSP over lexicographic and from approximate/heuristic TSP over *lexicographic*.

C. Extension for Future Cost Functions and Optimizations

The TSP bijection in Algorithm 2 is flexible in that it can be extended to other cost functions, i.e, new Dist$[i, j]$ assignments. One natural extension would be to account for single-qubit gate cancellations in addition to CNOT cancellations. For example, the cancellation between two Pauli X characters may be preferable to the cancellation between two Pauli Z characters, since the latter cancels out Hadamard gates in addition to a CNOT pair. The cost function could also account for hardware topological constraints or cancellations between non-nearest-neighbor Pauli strings. However, these cost functions might not be metric, which could hamper the approximability of TSP.

We could also explore different circuit architectures for the $\exp -iP_{\pi(j)}t$ terms. For example, the ancilla could be removed, and instead the R_z could target the last non-I qubit, as in Figures 13 (*star*) and 14 (*ladder*). In the corresponding *star* architecture, we would need to prepartition the Pauli strings into subsets that share common last non-I characters, to ensure that the CNOTs all cancel out. This strategy is analogous to a lexicographic sort on the last non-I character (rather than a lexicographic sort across all indices). We propose exploring this direction in future work.

D. Immediate Additional Work on TSP

We have also identified several directions for further work in applying TSP for gate cancellation. An immediate next step is to benchmark TSP gate cancellation results for actual Hamiltonians. In particular, we wish to study the improvement due both to TSP (over the no-gate-cancellation baseline and over lexicographic cancellation) and to an approximated TSP. The approximated TSP could be either via Christofides' algorithm or a heuristic. We also propose adding additional cost functions and optimizations, as suggested in Section IX C. Beyond the TSP bijection that we proposed for metric graphs, better approximations may be available since Pauli strings are in a restricted Hamming-like space. We propose to further study TSP in this space, following directions from [41, 42].

Besides optimizing gate cancellation within the clique, we can optimize the gate cancellation for the entire quantum dynamics circuit. Ideally this would give us the minimum quantum gate depth. However, the Trotter fidelity might be affected. Future work can explore the relationship between total circuit fidelity and maximized total circuit gate cancellation, with the help of physical noise models.

X. APPENDIX

Lemma 1: Binomial theorem for matrices: For arbitrary Hermitian matrices A and B of the same size, if $[A, B] = 0$, then we can binomially expand $(A+B)^n$

Proof. In general, expanding out $(A+B)^n$ would result in a summation of the permutation of the string with A and B, or namely, $\{A, B\}^{\bigotimes n}$. Since A and B commute with each other, we can freely arrange the orders of A and B in each AB strings, and group them together just like the binomial series for numbers.

$$(A+B)^n = \sum_{l=0}^{n} \binom{n}{l} A^{n-l} B^l$$

□

Corollary 1: In a multiplication series of matrix $A_1A_2...A_n$, if $[A_j,\ A_k] = 0\ \forall j,k \in \{1,2,...,n\}$, we can binomially expand the series as $(A_1A_2...A_{n-1} + A_n)^k$

Proof. By induction. Take Theorem 1 as base case, and induct on the newly added A_n term. □

1. Theorem 1 Proof

Proof. Using Taylor expansion and Lemma 1, we have

$$\begin{aligned} e^{A+B} &= \sum_{k=0}^{\infty} \frac{1}{k!}(A+B)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{l=0}^{k} \binom{k}{l} A^{k-l}B^l \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{k} \frac{1}{k!} \frac{k!}{l!(k-l)!} A^{k-l}B^l. \end{aligned}$$

Since l is bounded by k, as $k \to \infty$, l will sum up to ∞; therefore we can swap k and l as

$$e^{A+B} = \sum_{k=l}^{\infty} \sum_{l=0}^{\infty} \frac{1}{l!(k-l)!} A^{k-l}B^l.$$

Let $k - l = m$, we have

$$\begin{aligned} e^{A+B} &= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{l!m!} A^m B^l \\ &= e^A e^B. \end{aligned}$$

□

2. Theorem 2 Proof

Proof. By induction. For simplicity, we substitute $-iH_1t = h_1,\ -iH_2t = h_2, ..., -iH_Mt = h_M$. The commutation relations still hold since i and t are just coefficients.

Base case: application of Theorem 1.

Induction hypothesis: Suppose the theorem holds for all values of M, up to some number P, $P \geq 3$.

Induction steps: Let $M = P + 1$.

Denote $\sum_{n=1}^{P} h_n$ as H_{BEG}. By induction hypothesis, we have

$$e^{H_{BEG}} = e^{h_1}e^{h_2}...e^{h_P}.$$

Since h_{P+1} commutes with all the other h_n terms, from Corollary 1 we have

$$e^{H_{BEG}+h_{P+1}} = \sum_{k=0}^{\infty} \frac{1}{k!}(H_{BEG} + h_{P+1})^k.$$

Using the same technique from Theorem 2's proof by letting H_{BEG} = A and h_{P+1} = B, we have

$$e^{H_{BEG}+h_{P+1}} = e^{H_{BEG}}e^{h_{P+1}}.$$

Therefore,

$$e^{\sum_n^{P+1} h_n} = e^{h_1}e^{h_2}...e^{h_P}e^{h_{P+1}}.$$

□

ACKNOWLEDGMENTS

We thank Nathan Wiebe for the introduction to digital quantum simulation and for helpful discussions on commuting Pauli terms. We thank Pavel Lougovski for initial thoughts and helpful discussions. We also thank Zain Saleem for offering helpful comments.

Funding acknowledgments: P. G. is supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. P.G., Y. S., and F. C. are funded in part by EPiQC, an NSF Expedition in Computing, under grant CCF-1730449, by the NSF STAQ project under grant NSF Phy-1818914, and by DOE grants DE-SC0020289 and DE-SC0020331. Y. S. is also funded in part by the NSF QISE-NET fellowship under grant number 1747426. T. T. and M. M. are also funded by EPiQC, under grant CCF-1730082. The work of K. G. and M. S. is supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357.

[1] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, Jun 1982.

[2] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

[3] Pascual Jordan and Eugene P Wigner. About the Pauli exclusion principle. *Z. Phys.*, 47:631–651, 1928.

[4] Jacob T Seeley, Martin J Richard, and Peter J Love. The Bravyi-Kitaev transformation for quantum computation of electronic structure. *The Journal of Chemical Physics*, 137(22):224109, 2012.

[5] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014.

[6] Andrew Tranter, Peter J Love, Florian Mintert, and Peter V Coveney. A comparison of the Bravyi–Kitaev and Jordan–Wigner transformations for the quantum simulation of quantum chemistry. *Journal of Chemical Theory and cCmputation*, 14(11):5617–5630, 2018.

[7] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, Aug. 2018.

[8] Masuo Suzuki. Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations. *Physics Letters A*, 146(6):319 – 323, 1990.

[9] Andrew Tranter, Peter J. Love, Florian Mintert, Nathan Wiebe, and Peter V. Coveney. Ordering of Trotterization: Impact on errors in quantum simulation of electronic structure. 2019.

[10] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu. A theory of Trotter error, 2019.

[11] Pranav Gokhale, Olivia Angiuli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, Margaret Martonosi, and Frederic T Chong. Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families. *arXiv preprint arXiv:1907.13623*, 2019.

[12] A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Books on Chemistry. Dover Publications, 2012.

[13] James D Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. Simulation of electronic structure Hamiltonians using quantum computers. *Molecular Physics*, 109(5):735–750, 2011.

[14] G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme. Quantum algorithms for fermionic simulations. *Phys. Rev. A*, 64:022319, July 2001.

[15] R. Somma, G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme. Simulating physical phenomena by quantum networks. *Phys. Rev. A*, 65:042323, Apr 2002.

[16] M. Hjorth-Jensen E. Ovrum. Quantum computation algorithm for many-body studies. *arXiv preprint arXiv:0705.1928*, 2007.

[17] Jarrod R McClean, Ian D Kivlichan, Kevin J Sung, Damian S Steiger, Yudong Cao, Chengyu Dai, E Schuyler Fried, Craig Gidney, Brendan Gimby, Pranav Gokhale, et al. OpenFermion: the electronic structure package for quantum computers. *arXiv preprint arXiv:1710.07629*, 2017.

[18] Sam McArdle, Suguru Endo, Alan Aspuru-Guzik, Simon Benjamin, and Xiao Yuan. Quantum computational chemistry. *arXiv preprint arXiv:1808.10402*, 2018.

[19] Sartaj Sahni and Teofilo Gonzalez. P-complete approximation problems. *Journal of the ACM (JACM)*, 23(3):555–565, 1976.

[20] Matthew B Hastings, Dave Wecker, Bela Bauer, and Matthias Troyer. Improving quantum algorithms for quantum chemistry. *arXiv preprint arXiv:1403.1539*, 2014.

[21] Eugene F Dumitrescu, Alex J McCaskey, Gaute Hagen, Gustav R Jansen, Titus D Morris, T Papenbrock, Raphael C Pooser, David Jarvis Dean, and Pavel Lougovski. Cloud quantum computing of an atomic nucleus. *Physical Review Letters*, 120(21):210501, 2018.

[22] Pranav Gokhale and Frederic T. Chong. $o(n^3)$ measurement cost for variational quantum eigensolver on molecular Hamiltonians, 2019.

[23] Coen Bron and Joep Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

[24] Ravi Boppana and Magnús M Halldórsson. Approximating maximum independent sets by excluding subgraphs. *BIT Numerical Mathematics*, 32(2):180–196, 1992.

[25] Anthony W Knapp. *Basic real analysis*. Springer Science & Business Media, 2005.

[26] John Day Dollard and Charles N Friedman. *Product integration*. 1979.

[27] Andrew M Childs, Yuan Su, Minh C Tran, Nathan Wiebe, and Shuchen Zhu. A theory of Trotter error. *arXiv preprint arXiv:1912.08854*, 2019.

[28] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, USA, 10th edition, 2011.

[29] Norbert M Linke, Dmitri Maslov, Martin Roetteler, Shantanu Debnath, Caroline Figgatt, Kevin A Landsman, Kenneth Wright, and Christopher Monroe. Experimental comparison of two quantum computing architectures. *Proceedings of the National Academy of Sciences*, 114(13):3305–3310, 2017.

[30] Shantanu Debnath, Norbert M Linke, Caroline Figgatt, Kevin A Landsman, Kevin Wright, and Christopher Monroe. Demonstration of a small programmable quantum computer with atomic qubits. *Nature*, 536(7614):63, 2016.

[31] J Gambetta and S Sheldon. Cramming more power into a quantum device. *IBM Research Blog*, 2019.

[32] Jan Karel Lenstra and AHG Rinnooy Kan. Some simple applications of the travelling salesman problem. *Journal of the Operational Research Society*, 26(4):717–733, 1975.

[33] Richard Bellman. Dynamic programming treatment of the travelling salesman problem. *J. ACM*, 9(1):61–63, January 1962.

[34] Michael Held and Richard M Karp. A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics*, 10(1):196–210, 1962.

[35] Nicos Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical report,

Carnegie-Mellon Univ Pittsburgh PA Management Sciences Research Group, 1976.

[36] Chandra Chekuri and Kent Quanrud. Approximating the Held-Karp bound for metric TSP in nearly-linear time. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 789–800. IEEE, 2017.

[37] Chandra Chekuri and Kent Quanrud. Fast approximations for Metric-TSP via linear programming. *arXiv preprint arXiv:1802.01242*, 2018.

[38] Google Developers. C++ reference: Christofides. *OR-Tools Reference*, 2018.

[39] Albert Haque, Jay Shah, Faisal Ejaz, and Jia Xing Xu. An empirical evaluation of approximation algorithms for the metric traveling salesman problem. 2013.

[40] Kyle Genova and David P Williamson. An experimental evaluation of the best-of-many Christofides' algorithm for the traveling salesman problem. *Algorithmica*, 78(4):1109–1130, 2017.

[41] Luca Trevisan. When Hamming meets Euclid: the approximability of geometric TSP and MST. In *STOC*, volume 97, pages 21–29. Citeseer, 1997.

[42] Gérard Cohen, Simon Litsyn, and Gilles Zemor. On the traveling salesman problem in binary Hamming spaces. *IEEE Transactions on Information Theory*, 42(4):1274–1276, 1996.