On connectivity-dependent resource requirements for digital quantum simulation of $d$-level particles

NICOLAS P. D. SAWAYA$^1$, GIAN GIACOMO GUERRESCHI$^1$, AND ADAM HOLMES$^{2,3}$

$^1$Intel Labs, Santa Clara, CA 95054, USA
$^2$Intel Labs, Hillsboro, OR 97124, USA
$^3$University of Chicago, Chicago, IL 60615, USA

Corresponding author: Nicolas P. D. Sawaya (e-mail: nicolas.sawaya@intel.com).

ABSTRACT A primary objective of quantum computation is to efficiently simulate quantum physics. Scientifically and technologically important quantum Hamiltonians include those with spin-$s$, vibrational, photonic, and other bosonic degrees of freedom, i.e., problems composed of, or approximated by, $d$-level particles (qudits). Recently, several methods for encoding these systems into a set of qubits have been introduced, where each encoding’s efficiency was studied in terms of qubit and gate counts. Here, we build on previous results by including effects of hardware connectivity. To study the number of SW AP gates required to Trotterize commonly used quantum operators, we use both analytical arguments and automatic tools that optimize the schedule in multiple stages. We study the unary (or one-hot), Gray, standard binary, and block unary encodings, with three connectivities: linear array, ladder array, and square grid. Among other trends, we find that while the ladder array leads to substantial efficiencies over the linear array, the advantage of the square over the ladder array is less pronounced. Additionally, analytical and numerical results show that the Gray code is less advantageous when connectivity constraints are considered. These results are applicable in hardware co-design and in choosing efficient encodings for a given set of near-term quantum hardware when simulating Hamiltonians with $d$-level degrees of freedom.

INDEX TERMS quantum computation, vibrational, spin-$s$, bosonic, quantum simulation, connectivity, qudit.

I. INTRODUCTION

Hamiltonian simulation—the simulation of quantum physics using a quantum computer—is likely to be a primary early application of quantum computation. Much theoretical work has been done on the general problem of Hamiltonian simulation as well as its application to problems in chemistry and materials science [1, 2], condensed matter theory [3], and nuclear physics [4]. Most work in this area has focused on fermionic and spin-$\frac{1}{2}$ particles, although there is a large set of relevant problems of scientific interest involving ensembles of $d$-level systems (i.e., qudits), including photonic [5, 6], vibrational [7–13], and spin-$s$ [14, 15] degrees of freedom.

In contrast to fermionic particles which require the use of Jordan-Wigner [16] or related [17–19] transformations, here we are instead encoding $d$-level particles with bosonic commutation relations. This task consists primarily of mapping a series of local $d$-by-$d$ matrix operators to a set of qubits. For recent theoretical work on Hamiltonian simulation of photonic, vibrational, and bosonic degrees of freedom, the unary and standard binary (SB) encodings in second quantization have been considered [5], [11], [12], [20–24], and a first quantization approach was also studied [25], [26]. A systematic study of $d$-level encoding approaches, using both previously used and novel encodings, was recently published [27], and the purpose of the current work is to build on these results.

In this work we study the effects of hardware connectivity on two-qubit operation counts, when simulating $d$-level systems on a digital quantum computer. Most actively studied classes of quantum hardware [28, 29] (excluding ion trap systems [30]) do not inherentially allow for operations between arbitrary pairs of qubits. Hence a series of SWAP gates must be performed to make relevant qubit pairs adjacent. Here we study which encodings are superior for a given set of problem parameters when hardware connectivity is taken into account, and determine the added utility of changing connectivity patterns. We use a greedy scheduler of quantum circuits,
A small subset of results are shown in Figure 1 for $d$-level two-particle operators $\hat{n}_i \hat{n}_j$ and $\hat{q}_i \hat{q}_j$. The radius from the center of each radar chart equals the number of two-qubit gates and includes both the CNOTs needed by algorithm and the SWAPs required to overcome the limited connectivity. The blue (outermost) polygon of the radar chart represents linear hardware, with the inner polygons representing increased connectivity density. The plots demonstrate one of the main results of this work—namely, that the increase in connectivity from linear to ladder yields much more benefit than the subsequent transition to a 2D grid.

It is useful to think of encoding choices in terms of a “hardware budget” [27]—the optimal encoding depends on both the coherence time and the number of available qubits. As some encodings may require more gates with fewer qubits while others require fewer gates but more qubits, the choice of encoding will often depend on the available quantum hardware.

In Section II we summarize theory relevant to this work including encoding $d$-level particles to qubits. In Section III we derive upper bounds for the number of SWAP gates required to approximate a matrix exponential of local operators on hardware with linear connectivity. In Sections IV and V we respectively give our numerical methods and results. We end with discussion and outlook in section VI.

II. THEORY

A. TROTTERIZATION

Our goal is to implement the exponential of a Hamiltonian operator $\hat{H}$, a subroutine used both for dynamics and for eigenvalue estimation. Begin with a Hamiltonian $\hat{H}$ acting on $N_p$ particles:

$$\hat{H} = \sum_i c_i \hat{g}_i \tag{1}$$

where each term $c_i \hat{g}_i$ is a tensor product of $d$-level single-particle operators. Throughout this work, we assume that operators on different particles commute, as is the case for bosonic degrees of freedom.

In order to simulate the Hamiltonian on a qubit-based digital quantum computer, one must decompose the operator into a sum of Pauli strings such that

$$\hat{H} \mapsto \sum_j \hat{h}_j = \sum_j w_j \bigotimes_{k=1}^{N_q} \hat{\sigma}_{jk} \tag{2}$$

where $N_q$ is the number of qubits and each $\hat{\sigma}_{jk} \in \{\hat{X}_k, \hat{Y}_k, \hat{Z}_k, \hat{I}_k\}$ is either a Pauli matrix or the identity on qubit $k$. A well-known example of this step is in the simulation of fermions, where the second quantized Hamiltonian is converted to a sum of Pauli strings using the Jordan-Wigner [16], Bravyi-Kitaev [17], [18], or related transformations [19]. In simulation of $d$-level particles (qubits, truncated bosons, spin-$s$ particles), other approaches are instead required, as summarized in the next subsection. Several previous works have implemented mappings for qubit-based quantum simulation of bosons [5], [11], [12], [20]–[27].

To approximate the exponential, one may implement the Suzuki-Trotter formula [33], [34]

$$\exp(-i\hat{H} \tau) \approx \left( \prod_j \exp(-i\hat{h}_j \tau/\eta) \right)^\eta \tag{3}$$

which is exact in the limit of large $\eta$. Each Pauli string may be exponentiated using the well known “CNOT staircase” circuit [35], examples of which are shown in Figures 2 and 3.

B. ENCODING $D$-LEVEL SYSTEMS

We study the unary, standard binary, Gray, and block unary encodings. We refer to the Gray and standard binary encodings as compact encodings. Examples for these encodings are shown in Tables 1 and 2. Because the Gray code’s defining feature is a unity Hamming distance between consecutive integers [36], it was shown (using all-to-all connectivity) to often require fewer entangling gates [27] when Trotterizing common $d$-level operators.

Note that the qubit counts are not constant across encodings. Compact codes require $N_q = \lceil \log_2 d \rceil$ qubits,
qubits must be included in the mapping of four possible terms since \( x \) corresponds to a one-qubit operator that can be expressed as down the mapping

unary encoding includes just one bit. One may then write encoding includes all bits in the encoding, while that of a integer the bitmask subset encoding is that, to determine whether the system is in state

Table 1: SB, Gray, and unary encodings for decimal values 0 through 8.

| Decimal | Std. Binary | Gray Code | Unary |
|---------|-------------|-----------|-------|
| 0       | 0000        | 00000000  | 0     |
| 1       | 0001        | 00000001  | 1     |
| 2       | 0010        | 00000010  | 2     |
| 3       | 0011        | 00000100  | 3     |
| 4       | 0100        | 00001000  | 4     |
| 5       | 0101        | 00010000  | 5     |
| 6       | 0110        | 00100000  | 6     |
| 7       | 0111        | 01000000  | 7     |
| 8       | 1000        | 10000000  | 8     |

Table 2: Block unary encodings for decimal values 0 through 8.

| Decimal | BU_{SB} | BU_{Gray} |
|---------|---------|-----------|
| 0       | 0 0 0 0 0 0 0 0 | 0 0 0 0 0 0 0 0 |
| 1       | 0 0 0 0 0 1 0 0 | 0 0 0 0 0 1 0 0 |
| 2       | 0 0 0 0 1 0 0 0 | 0 0 0 0 1 0 0 0 |
| 3       | 0 0 0 1 0 0 0 0 | 0 0 1 0 0 0 0 0 |
| 4       | 0 0 1 0 0 0 0 0 | 0 1 0 0 0 0 0 0 |
| 5       | 0 1 0 0 0 0 0 0 | 1 0 0 0 0 0 0 0 |
| 6       | 1 0 0 0 0 0 0 0 | 0 0 0 0 0 0 0 0 |
| 7       | 1 1 0 0 0 0 0 0 | 1 1 0 0 0 0 0 0 |
| 8       | 1 1 0 0 0 0 0 0 | 1 1 0 0 0 0 0 0 |

unary requires \( N_q = d \), and block unary requires \( N_q = \lceil \frac{d}{l} \rceil \log_2 (g + 1) \).

In the preceding expressions, \( d \) is the number of levels in the particle, \( g \) is the number of bits in one “block,” and \( \lceil \cdot \rceil \) is the ceiling function.

Because near-term hardware will be limited both in operations counts (due to decoherence times) and total qubits, the choice of encoding may be hardware-dependent. For instance, if one wants to simulate a Hamiltonian using a quantum computer with many qubits but shorter coherence time, an analysis might show that the unary code fits the hardware budget while a compact code does not.

Given a matrix element \(|l\rangle \langle l'|\), one first converts the integers \( l \) and \( l' \) to bitstrings denoted \( R_{enc}(l; d) \) and \( R_{enc}(l'; d) \) for some encoding. A relevant property of every non-compact encoding is that, to determine whether the system is in state \(|l\rangle\), one may inspect only a subset of the qubits. Specifically, the bitmask subset \( C(l) \equiv C_{enc}(l; d) \) determines which qubits must be included in the mapping of \(|l\rangle\). For any integer \( l \), the bitmask subset of a compact (SB or Gray) encoding includes all bits in the encoding, while that of a unary encoding includes just one bit. One may then write down the mapping

\[
|l\rangle \langle l'| \mapsto \bigotimes_{i \in C(l) \cup C(l')} |x_i\rangle_i \langle x'_i|_i. \tag{4}
\]

For each factor in the right hand side of eq. \( 4 \) there are four possible terms since \( x_i, x'_i \in \{0, 1\} \). Each possibility corresponds to a one-qubit operator that can be expressed as the linear combination of the identity and at most two Pauli operators \( \{X, Y, Z\} \) using the formulas

\[
|0\rangle\langle 1| = \frac{1}{2} (\hat{X} + i\hat{Y}) \equiv \hat{\sigma}^+, \tag{5}
\]

\[
|1\rangle\langle 0| = \frac{1}{2} (\hat{X} - i\hat{Y}) \equiv \hat{\sigma}^-, \tag{6}
\]

\[
|0\rangle\langle 0| = \frac{1}{2} (\hat{I} + \hat{Z}), \tag{7}
\]

\[
|1\rangle\langle 1| = \frac{1}{2} (\hat{I} - \hat{Z}). \tag{8}
\]

It has been shown previously that it is often advantageous to convert between encodings within the same Trotter step \([27]\), though we do not consider encoding conversions in this work.

C. OPERATORS

We study one- and two-particle operators for this work, for both bosonic and spin-\( s \) degrees of freedom. Here we enumerate these operators and summarize properties that are relevant to resource requirements. We work in second quantization and define \( \hat{a} \) and \( \hat{a}^\dagger \) respectively as the bosonic creation and annihilation operators. The bosonic position \( \hat{q} = (\hat{a}^\dagger + \hat{a})/\sqrt{2} \) and momentum \( \hat{p} = i(\hat{a}^\dagger - \hat{a})/\sqrt{2} \) operators are tridiagonal matrices when represented in the number basis, as are spin-\( s \) operators \( \hat{S}_x \) and \( \hat{S}_y \) when represented in the \( Z \) basis. The triadiagonality is relevant partly because the Gray code yields unity Hamming distance between integers \( l \) and \( l \pm 1 \), which often leads to efficiency improvements over SB.

We explicitly show the sparsity patterns for \( \hat{q} \) and \( \hat{q}^2 \) to aid in the interpretation of our numerical results:

\[
\hat{q} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots \\ 1 & 0 & \sqrt{2} & 0 & \ldots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \ldots \\ 0 & 0 & \sqrt{3} & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{9}
\]

\[
\hat{q}^2 = \frac{1}{2} \begin{pmatrix} 1 & 0 & \sqrt{1 - \frac{1}{2}} & 0 & \ldots \\ 0 & 3 & 0 & \sqrt{2 - \frac{3}{2}} & \ldots \\ \sqrt{1 - \frac{1}{2}} & 0 & 5 & 0 & \ldots \\ 0 & \sqrt{2 - \frac{3}{2}} & 0 & 7 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{10}
\]

Using the nomenclature introduced in \([27]\), both the bosonic number operator \( \hat{n} = \hat{a}^\dagger \hat{a} \) and the spin-\( s \) \( \hat{S}_z \) operator are diagonal binary-decomposable (DBD) and diagonal evenly spaced (DES), with the result that the SB encoding allows one to exactly implement their exponential form with only one-qubit gates when \( d \) is a power of 2.

We include bosonic interaction operators \( \hat{n}_i \hat{n}_j, \hat{q}_i \hat{q}_j \), and \( \hat{q}_i \hat{q}_j \). We choose these three two-particle operators because they constitute all tensor product combinations of a diagonal and a banded tridiagonal matrix. Hence the results are applicable to other \( d \)-level systems with these operator characteristics—for example, results for \( \hat{n}_i \hat{n}_j \) are similar to
results for an operator $\hat{S}_x^{(i)} \hat{S}_z^{(j)}$ and $\hat{q}_i \hat{n}_j$, similar to results for $\hat{S}_x \hat{S}_z$. It is also notable that, though we do not show results in the paper, we confirmed that resource counts for $\hat{q}_i \hat{n}_j$ and the bosonic interaction operator $\hat{a}^\dagger_i \hat{a}_{i+1} + \text{h.c.}$ are nearly identical.

Finally, we note that, although for many common problems $d < 10$ tends to be sufficient, there are indeed cases where one requires $d$ to be 70 or greater [11].

### III. UPPER BOUNDS FOR SWAP GATES

In this section we study upper bounds for the number of SWAP gates required. Our analytical results consider scheduling only on a 1D line. Numerical results extend to denser connectivity and are provided in Section [11].

#### A. COMPACT OPERATORS

Here we consider analytical upper bounds for compact codes, i.e. those requiring $K = \lceil \log_2 d \rceil$ qubits per particle. For this analytical study, we assume that $d$ is a power of 2. $N_{\text{UB}}$ denotes the upper bound for the number of SWAP gates. We define $h$ as the Hamming distance between two bit strings.

**Arbitrary Pauli string.** A single arbitrary Pauli string on $K$ qubits has length $p$, where the length is the number of non-identity Pauli operators. In order to implement the Trotter circuit, relevant qubits must be adjacent at some point in the calculation. In the worst case, the relevant qubits are split in two equally sized groups and separated as far as possible along the line. In calculating upper bounds, we consider two routes for determining a SWAP pattern for this worst case. As shown in Figure [2], one may either consider a “cluster move” or a “shuttle move.” The former moves all the qubits next to each other, while the latter moves a qubit back and forth to complete the calculation. Asymptotically, this worst case requires SWAP counts of either

$$N_{\text{UB}}^{\text{Cl}}(p; K) = \frac{(K - p)^{p}}{2}$$

or

$$N_{\text{UB}}^{\text{Sh}}(p; K) = 2(K - p).$$

In our first analysis, we assign $N_{\text{UB}}^{\text{Sh}}(p; K)$ to all Pauli strings of length $K$ and $p$ non-identity terms. This upper bound may be tightened in future work by considering how the Pauli operators are distributed among $K$ qubits.

**Single term.** Consider a real Hermitian operator with only two non-zero entries, $|l\rangle \langle l'| + |l'\rangle \langle l|$. The mapping of one such term to a set of qubits leads to a sum of multiple Pauli strings. The distribution of Pauli string lengths is [27]

$$f(p; h, K) = \frac{1}{2^h} \binom{K - h}{p - h},$$

where $f(p; h, K)$ is the number of length-$p$ Pauli strings and $h = h(R(l), R(l'))$ is the Hamming distance between the bit-string representations $R$ of $l$ and $l'$. Note that $h \leq p, K$.

The upper bound for the number of SWAP gates for one $|l\rangle \langle l'| + |l'\rangle \langle l|$ term is thus

$$N_{\text{UB}}^{l,l'}(p; K) = \sum_{p=2}^{K} f(p; h, K) N_{\text{UB}}(p; K)$$

$$= \sum_{p=2}^{K} 2^h \frac{(K - h)(K - p)}{(p - h)}$$

$$= \frac{1}{2} K (K - h)$$

$$= \frac{1}{2} d(\log_2 d - h)$$

where $p = 2$ is the first value of $p$ for which SWAP gates may be required.

A key result is that $N_{\text{UB}}^{l,l'}$ decreases as $h$ increases. This is the opposite of the analytical trend in CNOT counts [27]. This is an intuitive result, since a higher $h$ leads to a higher density of Pauli terms, meaning that there are fewer “gaps” caused by local identity operators.

**All Pauli strings for $K$ qubits.** It is instructive to consider the highest possible bounds for $K$ qubits, for example in the worst case of encoding a fully dense matrix operator. At most, on $K$ qubits one may have all possible combinations of the four single-qubit operators. Though there are at most $4^K$ unique Pauli strings, there are only $2^K$ combinations of $I$ and non-identity $\sigma \in \{X, Y, Z\}$. If one considers the latter number of strings, then the cluster movement is a more appropriate upper bound, as one may implement many Pauli matrix exponentials with one particular qubit ordering.

We consider all length-$m$ combinations of $\{1, \sigma\}$ leading up to $K$, $m \in \{3, 4, ..., K\}$. For a given $m$, the number of
Pauli strings terminated with a non-identity is $2^m - 2^{m-1} = 2^{m-1}$. This leads to bounds of

$$N_{\text{All},K}^{\text{UB}} = \sum_{m=3}^{K} \sum_{i} N_{\text{UB}}^{\text{Cl}}(p_i; m)$$

$$= \sum_{m=3}^{K} \sum_{i} (m - p)p/2$$

$$< \sum_{m=3}^{K} 2^{m-1}m^2/8$$

$$< \frac{1}{8}(d \log_2 d - 2d \log_2 d + 3d - 12)$$

where using $p \leq m$ allowed us to write $(m-p)p \leq m^2/4$. This upper bound has lower complexity than $d$ times $N_{\text{UB}}^{\text{Cl},l,h}$ ($d \times O(d \log_2 d$), indicating that our original approach produces upper bounds that are too loose for full typical quantum operators of $O(d)$ non-zero entries. However, note that a greedy scheduling algorithm might not be able to “find” the upper bound derived from cluster and/or shuttle movements, unless there are additional pre-compilation steps to order the Pauli strings in a favorable fashion; hence our approach may be closer to what a greedy algorithm can achieve.

For this particular case (all strings on $K$ qubits) we also derive lower bounds for SWAP counts. We consider only the number of SWAP gates required for the case of $p = K/2$, the $p$ with the highest number of distinct Pauli strings. Since any SWAP sequence used for this value of $p$ will result in many legitimate groupings for other values of $p$, limiting analysis to $p = K/2$ yields a lower bound.

There are $(\frac{K}{2})^2 \approx \frac{K^2}{4}$ unique strings of non-identity terms (i.e. combinations using the two-member set $\{I, \sigma\}$, where in the last step we have used Stirling’s approximation which is accurate even for very small $K$.

In any given placement on a line, there are exactly $K/2 + 1$ consecutive groups of length-$K/2$ strings (those starting from position 0, from 1, . . . , and from $K/2$). A single SWAP yields at most 2 new groups (neglecting the qubit order inside the group) and therefore even in the best case the SWAP counts needed are at least

$$N_{\text{LB}}^{\text{All},K} \sim O(2^K/\sqrt{K}).$$

Hence we have relatively close lower ($O(2^K/\sqrt{K}) = O(d/\sqrt{\log_2 d})$) and upper ($O(2^K K^2) = O(d \log_2^2 d)$) bounds.

**Two-particle operators.** Two-particle upper bounds for SWAP gate counts will display a similar trend with respect to the Hamming distance, because the same argument with respect to density of “gaps” applies. Here we wish to highlight the bounds for any coupled pair of one-particle operators, by considering the case of multiplying two $K$-qubit operators together where every possible Pauli string is present. Plugging $2K$ into equations (15) and (16), the worst case for complete $2K$ qubits is

$$N_{\text{UB}}^{\text{All},2K} = \frac{1}{2}(2d \log_2^2 d - d^2 \log_2 d + \frac{3}{4}d^2 - 3),$$

and the lower bound is

$$N_{\text{LB}}^{\text{All},2K} = O(d^2/\sqrt{\log_2 d}).$$.  

**B. UNARY**

**One-particle operators.** Unlike the compact case, single-particle unary operators are linear combinations of Pauli strings with at most 2 non-identity factors. Exponentiating a single-particle unary operator consists of at most two-qubit operators. As a result, it is appropriate to take a different route than the compact case for estimating upper bounds.

Consider an operator of $O(d)$ terms, where terms occupy only a particular off-diagonal pair of bands, with the parameter $w = |l-l'|$ indicating the band including $|l\rangle \langle l'|$. Common matrix operators tend to be “banded” in this way (e.g. $w = 1$ for $q$) or are sums of a small number of matrices with this characteristic (e.g. $q^2$ is a sum of two matrices: one with only...
shown in Figure 3(a). If gates are required when the qubits are in default ordering, as calculation, and one to return to the original position.

Thus, given the ordering, the number of inversions required for the number of SW AP gates. The adjacent CNOT motifs without SW AP gates. The adjacent CNOT indices naturally partition the qubits into distinct groups, each of which can be internally ordered so as to create all required adjacencies. One such ordering is the sequential placement of each group onto the linear connectivity graph (Table 3).

To analyze the SWAP overhead of this configuration we use a well-known result of computer science: the number of required adjacent SWAP gates to convert one ordering to another is equal to the number of inversions in the array \( w \). Thus, giving the ordering, the number of inversions required to re-order the circuit can be calculated directly.

Table 3 counts inversions for different values of \( w \). From inspection, the upper bound for the number of SWAP gates will be

\[
N_{\text{SWAP}, \text{small } d} = \sum_{g=1}^{w-1} \left( \sum_{j=1}^{[d/w]-1} \right) g
\]

which leads to

\[
N_{\text{SWAP}, \text{small } d} = d^2 \frac{w - 1}{4w} - d \frac{w - 1}{4}
\]

It is notable, especially when considering low \( d \) values, that as \( w \) increases the second (linear) coefficient increases faster than the first (quadratic) coefficient.

A simpler route leads to linear scaling. One may instead consider every relevant pair of qubits, SWAP to make them adjacent, and then SWAP back to the original ordering once the two-qubit exponentiation has been performed pair, before moving on to the next. This leads us to

\[
N_{\text{SWAP}, \text{large } d} = 2(d - w)w
\]

Expression (21) does not lead to lower SWAP counts than (20) until at least \( d \geq 30 \), so expression (21) is a more accurate upper bound for many applications.

**Dense one-particle operator.** We consider the case of a fully dense matrix in the unary encoding, in which all pairs of \( d \) qubits must be adjacent at some point in the calculation, in order to exponentiate all terms \( X_i X_j + Y_i Y_j \). In such a case, one can implement a linear-depth SWAP network based on previous work in \( [38] \) that considered SWAP gates required for the quantum Fourier transform (QFT). The SWAP pattern for QFT leads to a SWAP count upper bound of

\[
N_{\text{SWAP}, \text{Unary}} = \frac{d^2}{2} + \frac{3}{2} + 1.
\]

These SWAP gates alone may be implemented in depth \( 2d - 3 \) \( [38] \).

**Two-particle operators.** Because a one-particle operator will include at most two-qubit terms, a two-particle operator (as it is built from products of one-particle operators) will include at most four-qubit terms. A few cases are instructive to consider. A product of two diagonal operators has structure

\[
a_0 Z_0 Z_d + a_1 Z_0 Z_{d + 1} + \cdots
\]

which yields at most \( d^2 \) two-qubit terms. The product of a diagonal and a banded off-diagonal \( (w > 0) \) operator has structure

\[
a_0 Z_0 (X_d X_{d + w} + Y_d Y_{d + w}) + a_1 Z_0 (X_{d + 1} X_{d + w + 1} + Y_{d + 1} Y_{d + w + 1}) + \cdots,
\]

leaving at most \( 2(d - w)d = 2(d^2 - dw) \) three-qubit terms. And the product of two banded operators (with equal \( w > 0 \)) has structure \((X_0 X_w + Y_0 Y_w)(X_d X_{d + w} + Y_d Y_{d + w}) + \cdots\), leading to at most \( 2(d(w - w)) = 4(d^2 - dw - w^2) \) four-qubit terms. Examples of Trotterized two-particle operators are given in Figures 3(c) and 3(d). Because of the \( X_i X_j + Y_i Y_j \) motif, multiple Pauli strings can be exponentiated with the same qubit placements, a fact we take advantage of in our upper bound calculations.

Unlike in the single-banded one-particle unary case, one cannot execute all of the exponentials of the two-particle case with a single ordering, as there are \( O(d^2) \) terms. For the diagonal–diagonal interaction operator \( (e.g. \hat{n}_i \hat{n}_j) \), we note that the interactions form a complete bipartite graph with \( d \) nodes in each partition. Beginning with ordering \( \{0, 1, \cdots, d, d + 1, \cdots, 2d - 1\} \), one may move the qubits of the bottom particle up until one reaches ordering \( \{d, d + 1, \cdots, 2d - 1, 0, 1, \cdots\} \). On the way to the final ordering, all relevant pairs of qubits are adjacent at some point in the procedure. This requires \( d^2 \) SWAP operations.

Next we consider the product of two banded one-particle operators \( (e.g. \hat{q}_i \hat{q}_j) \), both with \( w = 1 \). In this case, the same SWAP procedure from the diagonal-diagonal case may be used. Note that both terms like \( (X_1 X_2 + Y_1 Y_2)(X_{10} X_{11} + \cdots) \)
Y_{10}Y_{11}) and \((X_1X_2 + Y_1Y_2)(X_{11}X_{12} + Y_{11}Y_{12})\) need to be considered. As the second particle’s qubits are moved up, all relevant combinations of four qubits can be made adjacent at some point, though in an order that may be different from how the Hamiltonian term was originally written down. For instance, if \{1, 2, 10, 11\} is an ordering at one point in the swap network, then a few SWAPs later an ordering \{1, 11, 2, 12\} will naturally appear. This does not affect the calculation, as the only requirement for exponentiating a Pauli string is that the relevant qubits are adjacent.

We note this upper bound for SWAP operations required:

\[
N_{\text{UB, Unary}}^{2p_c l, \text{diag-diag}} = N_{\text{UB, Unary}}^{2p_c l, w=1} = d^2. \tag{23}
\]

Methods used in previous work \[38\] can be used to show that this sequence of SWAPS can be performed in depth \(2d - 1\).

In the case of a product of two banded operators with \(w > 1\), one may calculate an upper bound for the SWAP count by considering the reordering of each particle’s qubits to the order \(\{0, w, 2w, \ldots, 1, w + 1, 2w + 1, \ldots\\}\). This leaves us with the expression

\[
N_{\text{UB, Unary}}^{2p_c l, w>0} = d^2 + 2N_{\text{UB, Unary}}^{w}, \tag{24}
\]

where either equation \[20\] or \[21\] may be used for \(N_{\text{UB, Unary}}^{w}\) and the factor of 2 in the second expression is due to the presence of two particles.

**C. COMPARISONS BETWEEN UNARY AND COMPACT**

Though the upper bounds for compact and unary were calculated using different routes, both of which led to relatively loose bounds, it may be somewhat useful to make tentative comparisons between them.

Equation \[20\] scales worse than the absolute upper bound for the compact case calculated in \[15\]. This is partly because more qubits are present in the unary case, often leading to more distance that must be travelled. However, one often is not interested in increasing \(d\) asymptotically, as most physics applications involved building up a system of particles with bounded \(d\). The fact that \(\frac{1}{2}d \log_2^2 d\) (equation \[15\]) is larger than \(d^2 - 2d^2 - d^2\) (equation \[20\]) for small \(d\) suggests that unary may often have fewer SWAP gates.

Note that it was previously shown that unary usually requires fewer CNOT gates than compact codes for Trotterization. Hence this comparison of SWAP gates is relevant because it suggests that unary, at least for low-\(w\) operators that are common in quantum simulation, will often have superior overall gate counts for both all-to-all connectivity (which consider only CNOT gates) and linear connectivity (which consider both CNOT and SWAP counts). As we will show, the numerical results of this work bear out some of these trends.

Another relevant fact is that Pauli strings of the unary encoding always have length \(p \leq 4\), assuming one deals with two-particle operators at the most. One would expect this to affect circuit depth, as it means that more Pauli exponentials can be implemented in parallel in the unary as compared to the compact case.

**IV. NUMERICAL METHODS**

We consider three connectivities in this work: linear, ladder, and square grid. These are shown in Figure 4 together with the all-to-all case.

To determine the series of SWAP gates required in the routing process, we adopt the scheduler of quantum circuits described in \[31\], \[32\]. The inputs to the scheduling algorithm are the quantum circuit, the hardware’s connectivity, and the initial placement of program qubits on the physical qubits. When a two-qubit gate between not-connected qubits needs to be executed, the scheduler chooses one of them according to a greedy, stochastic policy. Due to this stochasticity, for each distinct input of the scheduler, we generate > 1000 stochastic schedules and report the one with the minimum number of SWAP gates.

Our initial qubit placement for the linear connectivity was the standard ordering \(0, 1, 2, \ldots\). For the ladder, we considered the “vertical snake” and “horizontal snake” placements explicitly visualized in Figure 4 and reported the lowest SWAP count that the scheduler was able to find. When the square grid is implemented, the length of the sides are \(\sqrt{N_q}\), we also use the snake placement. These initial positions are arbitrary, since an operation will most often be run after previous operations that have already reordered the qubits. However, choosing a consistent starting placement allows for direct comparisons between different encodings, and trends with respect to operator characteristics (such as sparsity structure) will often be applicable regardless of starting placement.

In our numerical results, we do not consider the depth of the circuits. This is because optimizing depth requires an entirely new set of considerations, namely the order in which the single exponentials \(\exp(-i\hat{h}_j^i \tau/\eta)\) are imple-
mented. This leads to a rich set of additional optimization considerations that we leave for future work. We note that one would expect the SWAP counts to be correlated with the circuit depth and its minimization a good proxy for overall circuit fidelity [41].

V. NUMERICAL RESULTS

The scheduling process took between a few seconds to almost one hour on a dual-socket Xeon Platinum 8280 (28 cores Cascade Lake per socket at 1.8GHz AVX base frequency, turbo disabled, 192 GB, dual-rail OPA) depending on the input.

A. ONE PARTICLE WITH LINEAR CONNECTIVITY

Figure 5 shows gate counts for one Suzuki-Trotter step of single-particle bosonic operators. The left column shows SWAP gate counts while the right column counts total two-qubit gates (SWAP and CNOT). Even though the Gray code requires more SWAP gates than SB (consistent with analytical arguments), for the tridiagonal \( \hat{q} \) it usually requires slightly fewer total two-qubit gates. For \( \hat{q}^2 \), the larger number of SWAP gates leaves Gray inferior for all simulated \( d \) values, when using the greedy scheduler.

For \( \hat{q} \), the unary code is improved relative to the all-to-all connectivity case, as it does not require any SWAP gates. Even in the case of \( \hat{q}^2 \), fewer SWAP gates provides it with a larger advantage over compact codes than before. For diagonal operators \( \hat{n} \) and \( \hat{n}^2 \), the unary always requires only single-qubit gates, and hence no SWAP gates are required either. Gray and SB yield similar results, because the extra SWAP gates required for Gray mostly cancel out its previous advantage in all-to-all connectivity.

The visibly favorable decrease in operations counts when \( \log_2 d \) is an integer has been discussed previously [27], including the fact that the bosonic \( \hat{n} \) is always most efficient with SB, requiring only one-qubit gates.

When dealing specifically with bosonic degrees of freedom, this implies that increasing the truncation often is helpful. For instance, if a problem requires a truncation of \( d = 5 \) or greater, then it is appropriate to compare unary’s \( d = 5 \) gate counts to the lowest compact (SB or Gray) gates counts in the range of \( d = 5−8 \), as all of the latter require the same number of qubits. In all plotted results, we show the values that result for truncating the given matrix at \( d \) (e.g. if Gray’s gate counts for \( d = 8 \) are less than for \( d = 5 \), we still plot the original \( d = 5 \) value). This choice was made in order to show behavior for general matrix-to-qubit mappings with these sparsity structures, as some problems types (like spin-s or classical combinatorial problems) do not allow one to increase the truncation.

Figure 6 shows results for a single-particle Suzuki-Trotter step of \( \hat{S}_x \) and \( \hat{S}_z \). In this case, one may not increase the truncation, as \( d \) is determined by the inherent properties of the particle (its spin \( s \)). There are no particularly strong trends, except that the Gray code tends to be slightly favorable despite its higher SWAP gate counts. We omit plots for

![Figure 5: SWAP gate counts and total two-qubit gate counts for a Suzuki-Trotter step of single-particle bosonic operators with increasing truncation \( d \), using linear hardware connectivity. Some data points are cut off in order to show low-\( d \) orderings more clearly.](image)

![Figure 6: SWAP gate counts and total two-qubit gate counts for a Suzuki-Trotter step of single-particle spin-\( s \) particles, using linear hardware connectivity.](image)
two-particle spin-$s$ operators, since the trends are similar to bosonic interaction operators (e.g. $S_z^i S_z^j \sim \hat{q}_i \hat{q}_j$).

Note that there is a limited set of instances where block unary may prove useful for specific hardware. These are the cases in which a BU code requires fewer qubits than the unary and fewer operations than the compact codes (e.g. $d = 9$ through 12 for $\hat{q}$). These limited use cases will be hardware-dependent, and are best thought of in terms of the hardware budgets as discussed previously [27].
B. TWO PARTICLES WITH LINEAR CONNECTIVITY

Figure 7 shows SWAP gate counts and total two-qubit gate counts for Suzuki-Trotter circuits of $\hat{n}_i \hat{n}_j$, $\hat{q}_i \hat{n}_j$, and $\hat{q}_i \hat{q}_j$. Though it is not shown, we note that numerical results for $\hat{a}_i^\dagger \hat{a}_{i+1} + \text{h.c.}$ were nearly identical to those for $\hat{q}_i \hat{q}_j$. The total number of operations increases as the number of non-diagonal operators (i.e. $\hat{q}$) increases from 0 to 2, because off-diagonal terms lead to more complex Pauli Hamiltonians.

In the unary code, the empirical trend for the SWAP gate counts is now super-linear, consistent with the analytics, and consistent with the fact that there are $O(N^2)$ Pauli strings instead of $O(N)$ in the single-particle unary case.

When one allows for increasing the bosonic truncation for compact codes (as discussed previously), unary-encoded $\hat{n}_i \hat{n}_j$ and $\hat{q}_i \hat{n}_j$ do not appear to be advantageous for any $d$. Considering $\hat{q}_i \hat{n}_j$, for example, unary with $d = 9$ requires more gates than SB at $d = 16$.

On the other hand, for $\hat{q}_i \hat{q}_j$, the optimal encoding is $d$-dependent, with unary often requiring the fewest total two-qubit gates. Finally, as in the one-particle case, there are very limited cases (such as $d = 9$ for all shown two-particle operators) where BU may be occasionally advantageous for near-term hardware.

C. VARYING HARDWARE CONNECTIVITY

We now compare the scheduler’s results across the three hardware connectivities. Each plot in this section studies one encoding-operator pair while varying $d$. Only SWAP gate counts are compared. We do not include square grid results for compact encodings, as the ladder and square grid layouts are either identical or very similar for small qubit counts.

Figure 8 shows results for compact representations of $\hat{n}$ and $\hat{n}_j^2$, for which all data points use at most $N_q = 5$ qubits. Unary is excluded because it yields zero SWAP gates. A substantial improvement is shown due to the ladder connectivity.

Figure 9 shows results for a selection of encodings of $\hat{q}$ and $\hat{q}_j^2$. Ladder grid improvements for the compact codes are substantial and similar in magnitude to the diagonal cases. The unary case for $\hat{q}_j^2$ shows substantial improvements when switching from linear to ladder, though the switch to square grid is either not advantageous or is detrimental. The square grid results become especially poor around the point that a 5-by-5 qubit grid is needed. This is intriguing, because a linear schedule can always be mapped to a snaked qubit placement on a grid, implying that linear should never be better than a higher connectivity. We conclude that the near-term decisions of the scheduler are made to the detriment of future gates in the quantum circuit since no look-ahead mechanism is present.

Figure 10 gives SWAP gate counts for a selection of two-particle operators. $\hat{q}_i \hat{n}_j$ is excluded because results are very similar to $\hat{n}_i \hat{n}_j$ and $\hat{q}_i \hat{n}_j$. Again, we see diminishing returns when increasing connectivity.

VI. DISCUSSION & OUTLOOK

Using analytical upper bound calculations and a greedy quantum circuit scheduler, we have studied the connectivity-dependent two-qubit operation counts required to approximate the exponentials of common bosonic and spin-$s$ operators.

Most scaling trends derived in the upper bounds calculations were observed in the numerical data. Importantly, the Gray code indeed required more SWAP gates than the standard binary code, due to the former’s lower-length Pauli strings. The unary code requires more SWAP gates the further its matrix elements are from the diagonal. The optimal encoding for two-particle operators is closely dependent on $d$, though compact encodings tends to be superior.

Compared to the previously studied all-to-all connectivity, the low-$d$ advantage of the unary code is slightly increased. As was the case before, block unary code is optimal only for a narrow set of operator-$d$ pairs. Interestingly, the Gray code is not as advantageous as previously predicted, often in fact showing a higher total two-qubit gate counts than standard binary. This is surprising, because the Gray code usually produces shorter Pauli string lengths. It may be that a scheduler that looks beyond the local optimum would find a more beneficial schedule for Gray-encoded circuits. We also note that these simpler Pauli Hamiltonians still suggest that the Gray code will be advantageous for algorithms like the variational quantum eigensolver, as simpler Hamiltonians require fewer measurements.

We tested three connectivities: linear, ladder, and square grid. In most cases, an increase in dimensionality lowers the number of required SWAP gates. Notably, changing from linear to ladder leads yielded a much larger percent improvement than changing from ladder to grid. Additionally, in some cases the scheduler’s results for the ladder are superior to the square grid. This is partially attributable to the scheduler considering a limited number of upcoming gates.

Fruitful future work may include tightening upper bounds by considering the fine-grained distribution of single-qubit Pauli operators. Additionally, the scheduler may be improved by taking more future gates into account, or by actively constructing the Trotterized quantum circuit using the qubit Hamiltonian as a starting point. Further, it will be important to study circuit depth for different hardware connectivities, which would require more detailed consideration of the order of the Suzuki-Trotter product.

Notably, there are instances for which each one of the studied encodings is the preferable choice. This fact highlights the need to consider multiple encoding types when preparing for Hamiltonian simulation of $d$-level particles, and multiple hardware connectivities when considering hardware design.

References

[1] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. KiefertovÁg, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, “Quantum chemistry in the age of quantum computing.” Chemical Reviews, vol. 2019, p. DOI: 10.1021/acs.chemrev.8b00803, 2019. arXiv:1812.09976.
