

Optimal number-conserved linear encoding for practical fermionic simulation

M. H. Cheng,^{1,2,*} Yu-Cheng Chen,^{3,4,*} Qian Wang⁵,³ V. Bartsch⁶,⁵ A. C. Medina,² M. S. Kim,^{1,†}
Alice Hu,^{3,6,‡} and Min-Hsiu Hsieh^{4,§}

¹*Department of Physics Blackett Laboratory, Imperial College London, SW7 2AZ London, United Kingdom*

²*Fraunhofer Institute for Industrial Mathematics, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany*

³*Department of Mechanical Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR 999077, China*

⁴*Hon Hai (Foxconn) Research Institute, Taipei, Taiwan*

⁵*Fraunhofer Center for Maritime Logistics, Blohmstrasse 32, 21079 Hamburg, Germany*

⁶*Department of Materials Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR 999077, China*



(Received 14 October 2023; revised 13 August 2024; accepted 15 August 2025; published 24 September 2025)

Number-conserved subspace encoding reduces resources needed for quantum simulations, but scalable complexity trade-off bounds for M modes and N particles with $\mathcal{O}(N \log M)$ qubits have remained unknown. We study qubit-gate-measurement trade-offs through the lens of classical/quantum error correction complexity and develop a framework of fermionic gate and measurement complexity based on classical encoder/decoder appearing in the error correction framework. We demonstrate optimal encoding with random classical parity check code and propose the Fermionic Expectation Decoder for scalable probability decoding in $\mathcal{O}(M^4)$ bases. The protocol is tested with variational quantum eigensolver on LiH in the STO-3G and 6-31G bases, and H_2 potential energy curve in the 6-311G* basis.

DOI: [10.1103/lzf5-x6hc](https://doi.org/10.1103/lzf5-x6hc)

I. INTRODUCTION

Simulating many-body fermionic systems is an essential tool in various fields like quantum chemistry, condensed matter, and high-energy physics [1–3], promising better designs for batteries and drugs, new quantum technologies, and discoveries in the exotic nature of matter [4–6]. However, accurate fermionic simulation on classical computers requires exponential computational resources [7]. While a quantum computer is expected to overcome this limit, it remains a central challenge to demonstrate quantum advantages in this domain.

Electronic structure Hamiltonian encoding maps fermionic operators from the Fock space to qubit operators in the qubit space. The standard approaches such as the Jordan-Wigner (JW), Parity, and Bravyi-Kitaev transformations [8,14] encode M fermionic modes to M qubits, regardless of the number of electrons N . However, since all fermionic Hamiltonians must conserve fermions, interesting physics depends only on the number-conserved subspace and takes only $\lceil N \log M \rceil$ qubits to encode in the precision limit $M \gg N$ [15,16]. Logarithmic scaling in M enormously relieves the qubit resource

requirements for simulation in the high precision limit, and it could serve as a powerful tool in the practical implementation of quantum simulation (QS) on near-term and fault-tolerant quantum computing (FTQC) devices [17].

Another important aspect of fermionic encoding is the measurement scalability of the two-electron reduced density matrix (2-RDM). Unlike classical encoding, quantum encoding changes the measurement complexity in two ways. First, nonlinear methods such as the projector approach [18] and the qubit-efficient encoding (QEE) [12] have attained the $\lceil N \log M \rceil$ qubit scaling via a direct fermionic state to qubit state encoding. However, it is plagued with unscalable encoding and measurement basis preparation cost of $\mathcal{O}(M^N)$, limiting its applications to fermionic systems with few electrons. Even if a specific grouping is found to mitigate the basis preparation cost, it will still suffer from an unbounded number of commuting Pauli strings.

Linear compression resolves the measurement and decoding cost. All linear compression methods, such as the segment code [9] and graph-based encoding [10], exhibit polynomial measurement costs with respect to M . However, they fail to efficiently encode fermionic modes and large number of electrons. The current best linear encoding demonstrates polylogarithmic $\mathcal{O}(N^2(\log M)^4)$ qubit scaling [11] that overcomes the scaling in the previous works, but still is limited to molecules requiring large basis set not accessible to near-term devices. The optimal compression rate in linear compression remains unknown.

We identify a natural complexity duality between linear compression for quantum simulation and quantum error correction (QEC) framework across qubit, gate, and measurement encoding-decoding process. Subsequently, we develop a framework of fermionic gate complexity based on

*These authors contributed equally to this work.

†Contact author: m.kim@imperial.ac.uk

‡Contact author: alicehu@cityu.edu.hk

§Contact author: min-hsiu.hsieh@foxconn.com

TABLE I. Comparison of the qubit-to-measurement trade-off of the existing encoding schemes at the regime $M \gg N$. We also provide the scaling of measurement bases for the listed linear encoders contributed from the FED algorithm.

	Jordan-Wigner [8]	Segment [9]	Graph-based [10]	Polylog [11]	QEE [12]	Our work
Qubit cost	M	$M - \frac{M}{2N}$	$M - \frac{M}{N}$	$\mathcal{O}(N^2(\log M)^4)$	$[N \log M]$	$[2N \log M]$
Measurements	$\mathcal{O}(M^3)$ [13]	$\mathcal{O}(M^4)$	$\mathcal{O}(M^4)$	$\mathcal{O}(M^4)$	$\mathcal{O}(M^N)$	$\mathcal{O}(M^4)$

quantum decoder and fermionic measurement complexity based on classical decoder. The latter result leads to the fermionic expectation decoding (FED) algorithm, which classically “error” decodes expectation value from compressed fermionic probability distribution (encoded Hamiltonian) for arbitrary linear encoding schemes. By avoiding calculation in the encoded space, FED allows for polynomial time classical post-processing with $\mathcal{O}(M^4)$ measurement basis, compared to JW $\mathcal{O}(M^3)$ measurement basis scaling [2] Table I. Meanwhile, using parity check code of error correction code (CEC) that obeys the Gilbert-Varshamov (GV) bound [10,19,20], we show that optimal linear encoder has a qubit resource upper bound of $2N \log M$, demonstrating a encoder-decoder protocol with the optimal qubit-to-measurement trade-off. Polynomial time encoders and decoders can be found in examples such as the geometric Goppa code [21,22] and Gallager’s low density parity check (LDPC) code [23,24]. Our protocol resolves scalability issue in nonlinear encodings in both the encoding and decoding processes via parity check code, polynomial measurement bases, and parity check decoder for fermionic observable postprocessing. Compatibility with other linear encoding schemes allows us to leverage the complexity of classical decoder. As a result, this scheme finds board applications for encoding fermionic systems in Noisy Intermediate Scale Quantum (NISQ) or FTQC quantum algorithms such as qubit Adaptive Derivative-Assembled Pseudo-Trotter ansatz Variational Quantum Eigensolver (qubit ADAPT-VQE) [25] and quantum signal processing [26].

We benchmark Gilbert bounded linear encodings with a randomized linear encoder (RLE) and a classical lookup table decoder to encode large-mode problems with a bounded number of electrons. RLE-encoded circuits adhere to the $2N \log M$ qubit bound, surpassing the qubit compression rates of the segment code and graph-based encoding for all N and M . The encoding is tested using the variational quantum eigensolver (VQE) on LiH in the STO-3G and 6-31G bases at a bond length of 2.5 Å with the hardware-efficient ansatz (HEA) [27,28], and on H₂ in the 6-311G* basis. The encoded LiH noiseless simulation achieves chemical accuracy using 30 CNOT gates with five layers, compared to the circuit in the JW basis, which does not converge even after six layers. We also compare both cases with the IBM Sherbrooke noise model. H₂ simulation achieves chemical accuracy across all bond lengths using only 66 local CNOT gates.

II. OPERATOR FRAMEWORK

Given a second-quantized electronic structure Hamiltonian

$$\hat{H} = \sum_{ij} h_{ij} \hat{a}_i^\dagger \hat{a}_j + \sum_{ijkl} g_{ijkl} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l, \quad (1)$$

qubit operator representation of the excitation operators $\hat{a}_i^\dagger \hat{a}_j$ and $\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l$, collectively denoted as $\hat{O} := \hat{a}_i^\dagger \hat{a}_j, \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l$, determines qubit, gate, and measurement complexities of quantum simulation. In what follows, we review the framework and complexity of fermionic-to-qubit mapping, drawing connections to linear compression of number-conserved particle basis.

Number-conserved bitstrings have fixed Hamming weight that can be compressed via a binary linear map $\mathbf{G} : \mathbb{F}_2^M \rightarrow \mathbb{F}_2^Q$. A fermionic/spin operator $\hat{O}$ can be represented with three types of binary vectors: $\vec{a}, \vec{b}, \vec{c} \in \mathbb{F}_2^M$. In terms of qubit operator, these vectors, respectively, correspond to transition, states undergoing transitions as projectors, and the parity picked up during transitions:

$$X^{\vec{a}} = \prod_{m=1}^M \sigma_{x,m}^{\vec{a}[m]}, \quad P^{\vec{b}} = \prod_{m=1}^M \frac{1 + (-1)^{\vec{b}[m]} \sigma_{z,m}}{2},$$

$$Z^{\vec{c}} = \prod_{m=1}^M \sigma_{z,m}^{\vec{c}[m]}, \quad (2)$$

dubbed as the X , P , and Z strings, respectively. Given an $M \times M$ encoder, such as the Bravyi-Kitaev (BK) encoding [14], the operators are encoded as [9]

$$\mathcal{E}(X^{\vec{a}}) = X^{\mathbf{G}\vec{a}}, \quad \mathcal{E}(P^{\vec{b}}) = P^{\mathbf{G}\vec{b}}, \quad \mathcal{E}(Z^{\vec{c}}) = Z^{(\mathbf{G}^{-1})^T \vec{c}}. \quad (3)$$

However, when $\mathbf{G}$ is $M \times Q$ with $M > Q$, the parity information cannot be represented as a Z -string operator due to the lack of a left inverse. Instead, Z -string representation is discarded via $Z^{\vec{c}} P^{\vec{b}} = (-1)^{\vec{c} \cdot \vec{b}} P^{\vec{b}}$, leading to instead the following decomposition of $\hat{O}$:

$$\text{Re/Im} \hat{O} = X^{\vec{a}} \sum_{\vec{b} \in S_{\hat{O}}} (-1)^{\vec{c} \cdot \vec{b}} (P^{\vec{b}} \pm P^{(\vec{a} \oplus \vec{b})}), \quad (4)$$

$$\text{Re/Im} \mathcal{E}(\hat{O}) = X^{\mathbf{G}\vec{a}} \sum_{\vec{b} \in S_{\hat{O}}} (-1)^{\vec{c} \cdot \vec{b}} (P^{\mathbf{G}\vec{b}} \pm P^{\mathbf{G}(\vec{a} \oplus \vec{b})}), \quad (5)$$

where we denote $S_{\hat{O}} = \{\vec{b} \in \mathbb{F}_2^M | H(\vec{b}) = H(\vec{a} \oplus \vec{b}) = N\}$ as the set of physical state basis where $\vec{a} \oplus \vec{b}$ is also a physical state. For a more thorough discussion of the mappings between classical bitstrings and quantum operators, see Appendix B.

III. QUBIT AND MEASUREMENT COMPLEXITY

The connection between qubit operator and linear code allows us to incorporate classical coding theory to compute fermionic expectation values. Qubit operator encoding-decoding is translated to finding $\mathbf{G}$, which preserves $S_{\hat{O}}$, and the decoder $F : \mathbf{G}\vec{b} \rightarrow \vec{b}$. Parity check codes turn out to be the

ALGORITHM 1. Fermionic expectation decoder.

Input: Probability distribution of projector variable $P(\vec{d} \in \mathbb{F}_2^Q)$,
 X -string vector $\vec{a}$ of $\hat{O}$, Z -string vector $\vec{c}$ of $\hat{O}$, Parity Check
Matrix $\mathbf{G}$, Minimal Weight Decoder $F(\cdot)$

Output: Fermionic Expectation Value $\langle \hat{O} \rangle = \text{val}$

$\vec{e} \leftarrow \vec{0} \in \mathbb{F}_2^Q$;
 $k \leftarrow \min \{k, \vec{a}[k] = 1\}$;
 $\vec{e}[k] \leftarrow 1$;
Find $\mathbf{A}$ s.t. $\mathbf{A}\vec{e} = \mathbf{G}\vec{a}$;
 $\text{val} = 0$;
for $P(\vec{d}) \neq 0$
 $\vec{m} \leftarrow \mathbf{A}\vec{d}$;
 $\vec{b} \leftarrow F(\vec{m})$;
Post-Select $\vec{b}$ which obeys the number conservation constraint.
 $s \leftarrow (-1)^{\vec{b} \cdot \vec{c}} * (-1)^{\vec{d} \cdot \vec{c}}$;
 $\text{val} \leftarrow \text{val} + P(\vec{d}) * s$;
end for
return val

right candidate for this task. In our context, it distinctively encodes the conserved particle basis as fixed Hamming weight $\vec{b}$ bit-flip errors and decodes them in polynomial time. CEC with distance $2N + 1$ has a parity check matrix that distinctively encodes a number-conserved fermionic state. By applying the GV bound to the parity check code, we obtain an upper bound on the minimal size of the fermionic encoder, leading to our first result.

Theorem 1. The minimal qubit cost required to encode an M fermionic modes N electrons problem has an upper bound of $Q \leq 2N \log M$.

Proof. See Appendix C. ■

This theorem asserts the existence of a logarithmic mode scaling linear encoder. Scalable construction of such encoders can be outsourced. Existing LDPC code packages such as AFF3CT [29] and OPENFEC [30] can generate not only scalable GV bound-converging encoders but also polynomial time decoders. To study the impact of the optimal encoder on quantum chemistry problems, we introduce RLE linear code in Appendix I. The compression rates of RLE Q/M are presented in Fig. 9 and Table III (Appendix H).

Meanwhile, error decoding in CEC can be translated to decoding number-conserved states for quantum simulation. However, unlike classical decoding, the expectation values of $\hat{O}$ are decoded in multiple measurement bases, which contributes to the measurement complexity. To calculate the total measurement complexity, we note that $\text{Re}/\text{Im} \hat{O}$ in Eq. (5) can be expanded as commuting Pauli strings measurable in one basis. It leads to the following measurement cost.

Theorem 2. The number of Clifford bases for measuring the expectation value of a quantum chemistry Hamiltonian is upper bounded by $\mathcal{O}(M^4)$.

Proof. See Appendix D. ■

Upon sampling in the compressed measurement basis, FED [31] is introduced in Algorithm 1 to classical postprocess the probability distributions of the projectors $X^{\mathbf{G}\vec{a}}(P^{\mathbf{G}\vec{b}} \pm P^{\mathbf{G}(\vec{a} \oplus \vec{b})}) \rightarrow X^{\vec{a}}(P^{\vec{b}} \pm P^{\vec{a} \oplus \vec{b}})$, where the parity $(-1)^{\vec{c} \cdot \vec{b}}$ is subsequently computed.

Compared to Ref. [12], FED avoids computing expectation value in the encoded subspace. It is compatible with all linear encoders, allowing resource leverages between the qubit and gate complexity while keeping the measurement efficient.

IV. GATE COMPLEXITY

The RLE encoder and its associated classical decoder have a computational complexity of $\mathcal{O}(M^N)$. In practice, they are replaced with polynomial encoder-decoder constructions that attain the GV bound, such as the Goppa code and Gallager's LDPC code. FED is fully scalable when combined with optimal CEC.

Such scalability is extended to the encoding of fermionic gate. In other words, the encoder-decoder formalism from error correction code also contributes directly to the gate complexity. We observe that the complexity growth is directly dual to the construction of QEC decoder, leading to the duality between QEC and QS. In what follows, we will formulate fermionic gate construction in the language of QEC decoder and demonstrate its implementation in the language of decoders.

A. QS-QEC duality

QEC involves error encoding via controlled gates, syndrome measurement, and error decoding [see Fig. 1(a)]. In QS, the roles of the first and second registers are reversed. Number-conserved configurations are coherently stored as “errors” in the second register. Rather than encoding errors, the encoder now maps quantum gate operations onto the parity check qubits. Similarly, the decoder not only removes “error” information from the first register but also decodes these configurations in the JW basis. This duality leads to the fermionic gate operation picture shown in Fig. 1(b).

In both cases, decoder complexity determines the scalability of the encodings. There are key differences between these protocols. In QS, the ancillary qubits $|s\rangle$ cannot always be the $|0\rangle$ state. Measurement and reset is forbidden to preserve the coherence, the decoder does not need to decode in the uncompressed number-conserved basis, and the decoder only needs to partially decode the necessary bit to encode gate operation. However, as we shall show, classical decoder from QEC [Fig. 1(a)] can be directly quantized and mapped [Fig. 1(b)], ultimately allowing us to characterize fermionic gate complexity with classical decoder complexities.

B. Partial decoding

Although QEC protocol is designed to eliminate a complete set of quantum errors, such as single-qubit bit-flip errors, syndrome measurements often reveal only *local* errors that require local error corrections via partial classical decoders. Similarly, decoding local errors corresponds to local fermionic gate operations in QS. These processes do not require full knowledge of the number-conserved basis. The information needed depends on the particular gate one aims to implement. In fermionic simulation, the relevant operations

are single- and double-excitation operators, which admit the following decomposition in the JW basis:

$$\hat{a}_i^\dagger \hat{a}_j \pm \hat{a}_j^\dagger \hat{a}_i = X_i X_j \prod_{i < m < j} Z_m (P_i^0 P_j^1 \pm P_i^1 P_j^0), \quad (6)$$

$$\begin{aligned} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l \pm \hat{a}_l^\dagger \hat{a}_k^\dagger \hat{a}_j \hat{a}_i &= X_i X_j X_k X_l \prod_{i < m < j} Z_m \prod_{k < n < l} Z_n \\ &\times (P_i^0 P_j^0 P_k^1 P_l^1 \pm P_i^1 P_j^1 P_k^0 P_l^0). \end{aligned} \quad (7)$$

In the postprocessing of FED, X strings in the first row of Eqs. (6) and (7) require no decoding. Meanwhile, Z strings in the second rows and the P operators in the third rows require decodings before expectation values are computed. Utilizing the QS-QEC duality, we recognize that logical gate operations share the same complexity composition as FED—decode and compute. The P (third rows) operators are first decoded to select the fermionic modes $[i, j]$ in Eq. (6) and i, j, k, l in Eq. (7) for transitions induced by the X and Z strings. These decoded modes, denoted as $\vec{b}[i]\vec{b}[j]$ and $\vec{b}[i]\vec{b}[j]\vec{b}[k]\vec{b}[l]$, then compute fermionic transitions $|\mathcal{E}(\vec{b})\rangle \rightarrow |\mathcal{E}(\vec{b} \oplus \vec{a})\rangle$. Meanwhile, the decoded Z strings (second rows), by summing over $\mathcal{O}(M)$ fermionic modes in the JW basis, compute the fermionic signs $|\mathcal{E}(\vec{b})\rangle \rightarrow \pm |\mathcal{E}(\vec{b})\rangle$.

For linearly encoded state, the summation and selection of Z and P operators requires distinctive decoder for fermionic mode a , and this necessitates M single-mode fermionic decoders to implement fermionic gate acting on arbitrary fermionic modes:

$$\begin{array}{c} \rho \text{ --- } \boxed{D_a} \text{ --- } \times M \\ |\mathcal{E}(\psi)\rangle \text{ --- } \bullet \text{ --- } \end{array}$$

where ρ is a single-qubit density matrix and a denotes the fermionic mode decoded from the linearly encoded multi-qubit state $|\mathcal{E}(\psi)\rangle$ on the second register. We note that these sets of decoders are dual to QEC decoders where each error syndrome too corresponds to a unique decoding. Up next, we will demonstrate the constructions and the corresponding complexity of these single-mode decoders, and subsequently, fermionic gate construction.

C. Decoder construction

Decoder from Fig. 1(a) can be directly quantized to quantum decoder in Fig. 1(b), summarized as follows:

Proposition 1. Classical linear decoder algorithms find direct quantum circuit representation.

Proof. By the virtue of Bennett's Theorem, any classical algorithm can be made reversible with at most polynomial overhead in space and time using ancillary bits [32].

Next, due to deferred measurement principle [33], measurement probability distribution remains the same if one defers the measurement after a reversible classical logical gate operation (see Fig. 2). This property corresponds to an isomorphism between reversible classical logic and coherent quantum gate operations when processing unentangled classical bits. That is, for a general reversible classical logic gate $f: \mathbb{F}_2^A \rightarrow \mathbb{F}_2^A$, provided that it can be decomposed into universal reversible classical gate set, such as classical multicontrol gates, it finds a direct quantum gate representation.

Since classical single-mode decoders $\mathcal{D}_i(\cdot)$ may be decomposed as classical reversible logic (which is further decomposed into multicontrol gates), for a classical single bit decoder

$$\mathcal{D}_i(\mathcal{E}(\vec{b})) = \mathcal{E}(\vec{b}) \otimes \vec{b}[i], \quad (8)$$

the same classical circuit $\hat{\mathcal{D}}_i(\cdot)$ can be directly implemented on quantum bits:

$$\hat{\mathcal{D}}_i(|\mathcal{E}(\vec{b})\rangle) = |\mathcal{E}(\vec{b})\rangle \otimes |\vec{b}[i]\rangle. \quad (9)$$

It immediately follows that

$$\hat{\mathcal{D}}_i\left(\sum_j c_j |\mathcal{E}(\vec{b}_j)\rangle\right) = \sum_j c_j |\mathcal{E}(\vec{b}_j)\rangle \otimes |\vec{b}_j[i]\rangle \quad (10)$$

due to linearity of quantum circuit. ■

Therefore, we have proved the existence of polynomial time quantum decoder for implementing single-mode fermionic decoders. In fact, this result allows us to analyze any encoding (beyond linear) quantum gate complexity in terms of classical decoder complexity. The optimization of fermionic gate complexity boils down to the optimization of classical decoder complexity. This result unifies previous frameworks of second-quantized fermionic encodings and provides a standardized means to study qubit-gate-measurement complexity trade-offs in the language of decoders.

D. Decoding subroutines

Based on single-mode fermionic decoders, we demonstrate subroutines that implement fermionic operators in the compressed subspace. We focus on projector P decoding, which enables selection of specific fermionic configurations required for the gate transitions, and Z -string decoding, which determines the sign structure in the JW basis. The latter can be implemented straightforwardly via the subroutine illustrated in Fig. 3. The signs of mode from a to z are decoded and summed together coherently from the encoded state. After fermionic gate operation, the sign-decoded fermionic modes remain the same, such that a second round decoding, as illustrated in Fig. 1(b), uncomputes the coherence on the ancillary qubit.

Meanwhile, the decoder construction of the P projector is less straightforward. In Eqs. (6) and (7), decoded fermionic information of P projectors is accompanied with transitions on fermionic modes i, j and i, j, k, l via $X_i X_j$ and $X_i X_j X_k X_l$, representing a pair of states that we wish to decode. These pairs of states undergoing transition must therefore be decoded to the same quantum state (indistinguishable to controlled state), and the controlled state contains enough information to induce the right transition. This is achieved by preparing $|s\rangle$ as the Greenberger-Horne-Zeilinger (GHZ) state on the fermionic modes that required decoding. Suppose we have a pair of encoded fermionic states $|\mathcal{E}(\vec{b})\rangle, |\mathcal{E}(\vec{b} \oplus \vec{a})\rangle$ such that in our notation

$$X^{\vec{a}} = X_i X_j, \quad (11)$$

acting on modes i, j , and a GHZ state on the same modes $(|0_i 0_j\rangle + |1_i 1_j\rangle)/2$. With decoders $\hat{\mathcal{D}}_i$ and $\hat{\mathcal{D}}_j$, if $|\mathcal{E}(\vec{b})\rangle$ is

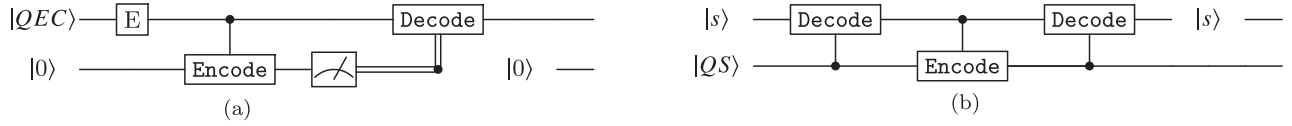


FIG. 1. QS-QEC duality. In QEC code, error channel on the $|QEC\rangle$ register is followed by error encoding and decoding after measurement, and the ancillary is reset to $|0\rangle$ state to be reused. The structure of error encoding and decoding finds its analogy in QS for gate operation, except that all operations are coherently implemented and two rounds of decoding are required to decouple $|s\rangle$, known as uncomputation in quantum reversible logic. Uncomputation allows one to reset the ancillary qubits, contrasting the nature of QS and QEC. (a) Circuit illustration of single round QEC and (b) General subroutine for implementing fermionic operators on linearly compressed fermionic state.

decoded to the value $m_i n_j$, then $|\mathcal{E}(\vec{b} + \vec{a})\rangle$ is decoded to $(m \oplus 1)_i (n \oplus 1)_j$ in the computational basis due to Eq. (11). However, when decoding the information on the GHZ state, we obtain the following:

$$m_i n_j \rightarrow \frac{|(m \oplus 0)_i (n \oplus 0)_j\rangle + |(m \oplus 1)_i (n \oplus 1)_j\rangle}{2}, \quad (12)$$

$$(m \oplus 1)_i (n \oplus 1)_j \rightarrow \frac{|(m \oplus 1)_i (n \oplus 1)_j\rangle + |(m \oplus 0)_i (n \oplus 0)_j\rangle}{2}. \quad (13)$$

That is, the resulting pair of GHZ states are indistinguishable despite the outputs on the left-hand side of the equation. Similar argument can be made with four modes decoding for double excitations. In that case, we just prepare the state $(|0_i 0_j 0_k 0_l\rangle + |1_i 1_j 1_k 1_l\rangle)/2$ on the ancillary qubits. Figure 4 illustrates double-excitation P decoding subroutine, where

$$X^{\vec{a}} = X_i X_j X_k X_l. \quad (14)$$

In this subroutine, $|s\rangle = |\text{GHZ}\rangle$, $\text{Decode} = \hat{D}_{i,j,k,l}$, and circuit in between $\hat{D}_{i,j,k,l}$ is the Encode subroutine in Fig. 1(b). The Encode subroutine implements controlled rotation if the decoded state is $(|1_i 1_j 0_k 0_l\rangle + |0_i 0_j 1_k 1_l\rangle)/2$, corresponding to the P projector in Eq. (7). The efficiency of one-to-one mapping of the X strings [Eq. (3)] is reflected by the single controlled gate when implementing transition.

In Appendix E, we will demonstrate time evolution encodings via P and Z decoding subroutines, and additionally discuss the worst-case overhead when implementing gate complexity without fermionic decoders.

E. Near-term complexity

Although QS-QEC duality demonstrates thoroughly complementary qubit-gate-measurement complexity contribution in the language of classical/quantum decoders, a complete fermionic simulation encoding only finds application in early FTQC regime. For near term, our protocol can implement quantum algorithms without fermionic gates, as it is done in our VQE benchmarks. Besides HEA ansatz, our protocol is compatible with the operator selection techniques in qubit ADAPT-VQE, which encodes the Hamiltonian information on the qubit operators rather than fermionic operators. In Appendix G, we introduce an algorithm that performs operator selection in the encoded subspace. The numerics illustrate the advantages of reducing operator actions in the compressed qubit space, suggesting potential benefits in minimizing the operator pools for implementing qubit ADAPT-VQE. Our results provide more effective ground state preparation strategies, expanding the scope of Variational Quantum Algorithms.

V. NUMERICAL RESULTS

We benchmark the RLE-FED protocol on VQE trained with the L-BFGS-B optimizer, HEA ansatz, and Hartree-Fock initialization. The impact of compression on gate resources and convergence without noise is investigated. In Fig. 5, LiH molecule [34] in STO-3G compressed from $M = 8, N = 2$ to $Q = 6$ qubits is compared against the JW-encoded LiH; in Fig. 6, the dissociation curve of H_2 molecule [35] in the 6-311G* basis compressed from $M = 12, N = 2$ to $Q = 7$ qubits is simulated to study its gate resource requirement. All data points are averaged over 30 instances of gradient descent to account for the randomness of our encoder and the quantum noise.

Noiseless RLE-simulated LiH energy converges to chemical accuracy (< 1 kcal/mol) with 30 local CNOT gates, with the noisy part converged to lower energy than the HF state. On the other hand, the JW-encoded noiseless/noisy LiH simulation requires more entangling gates but fails to show any signs of convergence. Meanwhile, using as few as 66 local CNOT gates, all H_2 energy points converge to chemical accuracy, compared to the JW-encoded H_2 that requires 11 CNOT gates at least per layer and double excitations that require at least 13 CNOT gates. Training the JW-encoded LiH and H_2 molecules with HEA or the UCCSD ansatz often requires significantly more CNOT gates to achieve the same accuracy. Subspace encoding allows state preparation with substantially less coherence time and consequently less circuit depth [36]. This is further supported with the benchmarks of 16-mode 6-31G LiH in Tables III and IV in Appendix H.

Overall, successful ansatz training on the RLE-encoded circuit demonstrates significantly fewer CNOT gates, and consequently, the spatial and temporal aspects of the quantum circuits are less prone to noise than the uncompressed case. Noisy energy variance shows that the compressed problem

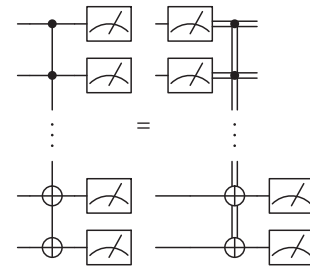


FIG. 2. Deferred measurement principle for multicontrol gate. It demonstrates direct correspondence between classical reversible circuit and quantum circuit.

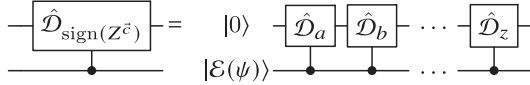


FIG. 3. Fermionic sign decoder circuit for the binary vector of the Z string Z^z built from single-mode fermionic decoders. The indices $a, b, \dots, z$ denote the nonzero entries of binary vectors $\vec{c}$.

also produces a stable result. While the compressed HEA simulation results do not directly imply scalability in VQE, our protocol is compatible to scalable VQE algorithm such as qubit ADAPT-VQE [25]. See Appendix G for implementation. Exponential improvement on the scaling of fermionic modes relieves the problem of barren plateau contributed from qubit scaling and enhances the performance of qubit ADAPT-VQE.

VI. CONCLUSION

In this manuscript, we have proposed the use of parity check matrix and FED for encoding and decoding number-conserved problems in quantum simulation. We established a connection between number-conserved linear encoding and the parity check matrix of a dual CEC code, leading to insights for constructing number-conserved encoding. Leveraging this duality, we showed that the optimal number-conserved linear encoding is bounded above by $2N \log M$. Our proposal also prepared efficient measurement bases for number-conserved observables, including the 2-RDM for quantum chemistry, with a measurement complexity upper bound of $1 + \binom{M}{2} + \binom{M}{4}$. These results can be applied to construct encoding with optimal qubit-to-measurement trade-off. Parity check code with polynomial time decoder is available in the geometric Goppa and Gallager's LDPC codes. Meanwhile, FED offers a universal protocol to decode fermionic information for all

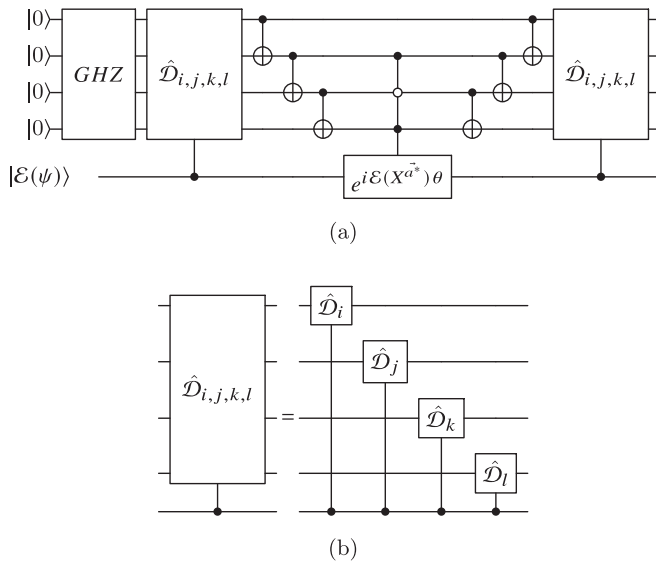


FIG. 4. (a) Building block of fermionic double excitations built from single-mode fermionic decoders. (b) Projector decoder based on Eq. (7).

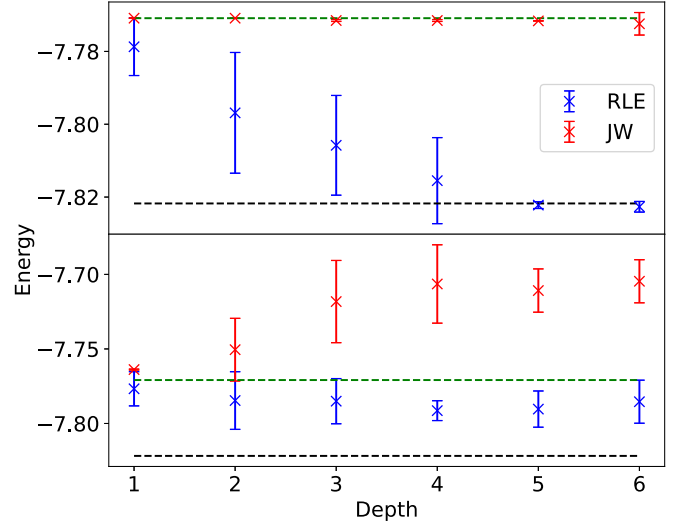


FIG. 5. VQE energy of STO-3G LiH at 2.5 Å without compression using HEA at varying depth. The JW-encoded noiseless energy (top red) cannot find energy better than the HF energy (green), whereas the RLE energy (blue) converges to chemical accuracy (black) at the fifth layer (25 CNOT). In the *ibm-sherbrooke* noisy simulation, the RLE energy shift is significantly less than the JW energy, which diverges from the HF energy.

linear encodings in second-quantized simulation, enabling trade-offs between the qubit and gate complexity.

To benchmark optimal codes, we constructed the RLE encoding and FED decoding algorithms. We compared VQE simulations of LiH in the RLE- and JW-encoded STO-3G/6-31G basis, and H₂ in the RLE-encoded 6-311G* basis using HEA at different circuit depths. We conducted a VQE experiment on H₂ in the 6-311G* basis against full configuration interaction (FCI), demonstrating chemical accuracy across all bond lengths. All results favored the optimal encoding in terms of CNOT cost and convergence, which showcased the potential benefits of implementing VQE with our protocol.

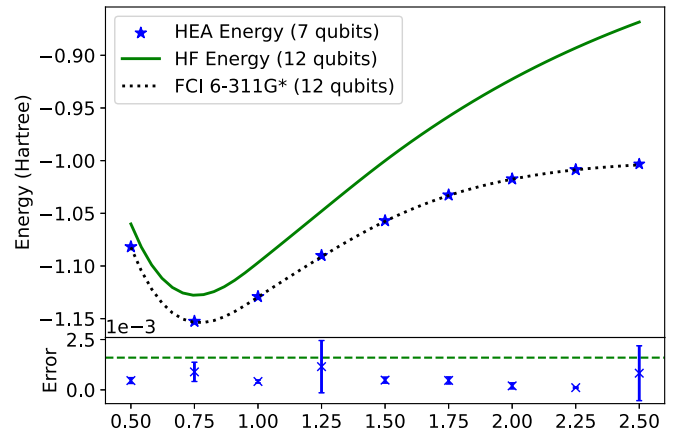


FIG. 6. Noiseless VQE potential energy curve of 6-311G* (12 modes) H₂ potential energy curves encoded with seven qubits using HEA with 11 layers (66 CNOT). The energy uncertainty is displayed at the bottom, which lies within the chemical accuracy threshold (green dot).

TABLE II. The numerical result of RLE compression. This table shows the maximum number of modes M that we can encode given the number of electrons N (rows) and qubits Q (columns) with RLE. All results were computed on an eight-core laptop.

$N \backslash M \backslash Q$	10	12	14	16	18	20	22	24	26	28	30	
2		22	36	48	64	90	118	158	226	316	420	580
3		13	20	25	31	38	46	58	72	88	105	140
4		11	14	18	23	27	31	36	42	50	60	71
5		11	14	16	19	22	26	31	34	39	44	49
6		11	13	16	18	21	23	27	31	34	37	41

We propose a subroutine based on our protocol to implement qubit ADAPT-VQE in the encoded subspace, further relieving the resource requirement for variational quantum simulation of number-conserved quantum systems.

The duality between number conservation and error correction extends beyond FED. We show that QEC decoding is dual to simulating quantum gates in the compressed number-conserved subspace. We proposed a generalized framework on scalable fermionic gate encoding, echoing the results in Ref. [37]. This result opens the doors for designing sparse access oracles with technology in QEC decoders, expanding the current tool set of quantum simulation in FTQC regime and allowing flexible resource trade-offs when encoding number-conserved systems, while ensuring scalability.

Overall, the connection between CEC and number conservation has implications for resource trade-offs in quantum simulation, in both the NISQ and FTQC regimes. Our example demonstrates that this connection can be leveraged to reduce quantum resource requirements for variational quantum algorithms. We leave as an open question how these resource trade-offs can be further exploited in the encoding of more general quantum algorithms.

ACKNOWLEDGMENTS

A.H. gratefully acknowledges the sponsorship from Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11200120), City University of Hong Kong (Projects No. 7005615 and No. 7006103), CityU Seed Fund in Microelectronics (Project No. 9229135), and Hong Kong Institute for Advanced Study, City University of Hong Kong (Project No. 9360157). M.S.K. thanks the Samsung GRC project and the UK EPSRC (EP/@032643/1 and EP/Y004752/1). This work was supported by the project AnQuC-3 of the Competence Center Quantum Computing Rhineland Palatinate (Germany).

DATA AVAILABILITY

All data and code supporting the findings of this study are openly available [31].

TABLE III. The noiseless simulation of (8, 2) LiH using HEA- and UCC-type ansätze on the state vector simulator. ΔE is the difference between the VQE energy and energy from exact diagonalization. The HEA results are chosen from the lowest energy results using the VQE with L-BFGS-B optimizer and 100 different initial states. The UCCSD and SUCCD ansätze are constructed from qiskit package.

System	LiH (8, 2)			
	HEA		UCC	
	8 qubits	6 qubits	SUCCD	UCCSD
Number of parameters	48	36	6	15
Number of CNOT	35	25	480	768
Circuit depth	41	31	668	1043
ΔE (kcal/mol)	0.4126	0.0002	0.5216	0.00003

APPENDIX A: FRAMEWORK FOR NUMBER-CONSERVED LINEAR ENCODING

We provide a detailed discussion of the dual code correspondence that connects number-conserved linear encoding with classical error-correcting codes, inspired by Ref. [10].

In this work, we make use of a $Q \times M$ parity check code $\mathbf{G}$ of an $(M - Q) \times M$ CEC $\mathbf{C}$ to compress the number-conserved configurations. $\mathbf{G}$ is defined as the dual code of $\mathbf{C}$:

$$\mathbf{G}\mathbf{C}^T = \mathbf{0}. \quad (\text{A1})$$

Due to this duality, constraints on $\mathbf{P}$ can be directly translated to constraints in $\mathbf{C}$. With respect to $\mathbf{G}$, we can think of $\mathbf{C}$ as the generator of equivalence class C on the M -dimensional binary vector space $\mathbb{F}_2^M$:

$$C = \{\vec{v} \in \mathbb{F}_2^M \mid \forall \vec{c} \in \mathbb{F}_2^{M-Q}, \vec{v} = \mathbf{C}^T \vec{c}\} = \ker \mathbf{G}. \quad (\text{A2})$$

This class C is the kernel $\ker \mathbf{G}$ of the encoded space. $\ker \mathbf{G}$ also defines the equivalence class of all encoded bitstrings $\vec{b} \in \mathbb{F}_2^M$:

$$\vec{b} \sim \vec{b} \oplus \vec{v} \quad \text{for } \vec{v} \in C. \quad (\text{A3})$$

TABLE IV. The noiseless simulation of (16, 2) LiH using HEA- and UCC-type ansatz on the state vector simulator. ΔE is the difference between the VQE energy and energy from exact diagonalization. The HEA results are chosen from the lowest energy results using the VQE with L-BFGS-B optimizer and 100 different initial states. The UCCSD and SUCCD ansätze are constructed from the qiskit package.

System	LiH (16, 2)			
	HEA		UCC	
	16 qubits	8 qubits	SUCCD	UCCSD
Number of parameters	96	48	28	63
Number of CNOT	75	35	4032	7280
Circuit depth	81	41	4874	8675
ΔE (kcal/mol)	2.2099	0.7186	0.5749	0.00001

Let us define $\vec{b} \oplus C$ to be the equivalence class of $\vec{b}$. To ensure that all configurations are distinctly encoded, number conservation demands that each bitstring with Hamming weight N shall only appear once in $\vec{b} \oplus C$, which constrains the Hamming weight of C and thus its generator $\mathbf{C}$, the error-correcting code. If $\mathbf{C}$ generates a subspace of even Hamming weights with code distance $2N + 2$, then for the subspace with Hamming weight less than or equal to N , $\mathbf{G}$ will distinctively encode them into the encoded space. Proof of this statement is provided in Lemma C1. On the other hand, we can generate arbitrary odd equivalence classes because they do not contribute to mapping bitstrings with the same parity.

As we will show in the proof of Theorem 1, this is equivalent to saying that we would like our $\mathbf{C}$ to have one unconstrained row element with odd Hamming weight. As for the rest of the rows, they form a code with even Hamming weight and distance $2N + 2$. Thus, $\mathbf{C}$ contains an error-correcting subcode of dimension $(M - Q - 1) \times M$ with even Hamming weight and distance $2N + 2$ for us to construct a valid $Q \times M$ number-conserved linear encoding with code parameters $[M, M - Q - 1, 2N + 2]$.

APPENDIX B: FRAMEWORK FOR QUBIT OPERATOR ENCODING

The basics of operator encoding-decoding are reviewed here in the language of binary linear vector space. Literature such as Refs. [12,38] employs projectors for approaching the $[N \log M]$ lower bound. However, both methods suffer from unscalable measurement costs. One reason behind this broken scaling is the disruption of the linear structures of the projectors. In our framework, we connect the projector to linear binary vector space, which allows us to assess the measurement complexity with the linear structure.

1. Notations and properties

We define the notations of the X , P , and Z strings used in this paper.

Definition B1. Projectors in the computational basis are defined as

$$P^{0/1} = |0\rangle\langle 0| = \frac{1 + (-1)^{0/1}Z}{2}. \quad (\text{B1})$$

Definition B2. Let $\vec{b} \in \mathbb{F}_2^M$ be a binary vector; P string is defined as follows:

$$P^{\vec{a}} = \prod_{m=1}^M P_m^{\vec{b}[m]}, \quad (\text{B2})$$

with $\vec{b}[m]$ denoting the m th entry of the vector and the following property:

$$P^{\vec{b}} |\vec{a}\rangle = \delta_{ab} |\vec{a}\rangle. \quad (\text{B3})$$

Definition B3. Let $\vec{b} \in \mathbb{F}_2^M$ be a binary vector; X string is defined as follows:

$$X^{\vec{b}} = \prod_{m=1}^M X_m^{\vec{b}[m]}, \quad (\text{B4})$$

with $\vec{b}[m]$ denoting the m th entry of the vector and the following property:

$$X^{\vec{b}} |a\rangle = |\vec{a} \oplus \vec{b}\rangle. \quad (\text{B5})$$

Definition B4. Let $\vec{b} \in \mathbb{F}_2^M$ be a binary vector; we define a Z string as follows:

$$Z^{\vec{b}} = \prod_{m=1}^M Z_m^{\vec{b}[m]}, \quad (\text{B6})$$

with $\vec{b}[m]$ denoting the m th entry of the vector and the following property:

$$Z^{\vec{b}} |\vec{a}\rangle = (-1)^{\vec{b} \cdot \vec{a}} |\vec{a}\rangle. \quad (\text{B7})$$

Next, we will derive the transformation of the X , P , and Z strings with an encoder $\mathcal{E} : |\vec{a}\rangle \rightarrow |\mathbf{G}\vec{a}\rangle$. Immediately, we can infer from the state encoding of operators in Eqs. (B3), (B5), and (B7) the transformations as

$$\mathcal{E}(P^{\vec{b}}) = P^{\mathbf{G}\vec{b}}, \quad (\text{B8})$$

$$\mathcal{E}(X^{\vec{b}}) = X^{\mathbf{G}\vec{b}}, \quad (\text{B9})$$

$$\mathcal{E}(Z^{\vec{b}}) = Z^{(\mathbf{G}^{-1})^T \vec{b}}. \quad (\text{B10})$$

Linear compression preserves Eqs. (B8) and (B9), but breaks Eq. (B10) due to the lack of a left inverse for $\mathbf{G}$. Instead, the individual Pauli- Z operator can be encoded as

$$\mathcal{E}(Z^{\vec{c}}) = \sum_{\vec{b} \in S} (-1)^{\vec{c} \cdot \vec{b}} P^{\mathbf{G}\vec{b}}, \quad (\text{B11})$$

where $P^{\mathbf{G}\vec{b}}$ is a Q -qubit projector and $S = \{\vec{b} \mid H(\vec{b}) = N\}$ is the set of all number-conserved states. The exact gate implementation of Eq. (B11) can be performed with multiconrolled gates or a linear combination of an exponential number of Z strings. However, both methods suffer from scalability problems, as discussed in the gate complexity section in Appendix E. This problem can be tackled using Quantum Signal Processing (QSP) techniques, provided that the encoded $\mathcal{E}(Z^{\vec{c}})$ is also sparse.

2. XP decomposition

We can now derive a general representation of the qubit operator. In essence, all qubit operators consist of a selected set of states, state transitions, and phase evolution. In the computational basis, X strings capture transitions, Z strings capture phase, and the sum over projectors captures the states.

Given an X string $X^{\vec{a}}$ and a projector $P^{\vec{b}}$, we can observe the following properties:

$$X^{\vec{a}} P^{\vec{b}} = P^{\vec{a} \oplus \vec{b}} X^{\vec{a}}. \quad (\text{B12})$$

We can represent the Hermitian or anti-Hermitian part of the operator as

$$X^{\vec{a}} (P^{\vec{b}} \pm P^{\vec{a} \oplus \vec{b}}) = |\vec{a} \oplus \vec{b}\rangle \langle \vec{b}| \pm |\vec{b}\rangle \langle \vec{a} \oplus \vec{b}|, \quad (\text{B13})$$

$$(X^{\vec{a}} (P^{\vec{b}} \pm P^{\vec{a} \oplus \vec{b}}))^{\dagger} = \pm X^{\vec{a}} (P^{\vec{b}} \pm P^{\vec{a} \oplus \vec{b}}). \quad (\text{B14})$$

Summation over this basis operator with respect to the Boolean vectors $\vec{a}$ and $\vec{b}$ allows us to construct arbitrary

(anti-)Hermitian qubit operators. As we show in Theorem 2, arbitrary sum over projector $P^{\vec{b}}$ and its conjugate $P^{\vec{a} \oplus \vec{b}}$ and $X^{\vec{a}}$ generates commuting Pauli strings as we expand the projector into Z strings. The number of X strings determines the measurement cost of a given qubit operator. Suppose now we exponentiate the Hermitian operator. It also determines the Trotterization steps required to approximate the unitary evolution.

In the JW representation, the product of creation and annihilation operators always generates one X string under our decomposition. Our linear encoder then surjectively encodes the JW X strings. This property allows us to exactly implement fermionic gates, though at the risk of an unscalable number of gates, and efficiently prepare measurement bases for fermionic observables.

APPENDIX C: QUBIT COMPLEXITY PROOF

Lemma C1. Given bitwise linear map $\mathcal{E}$ that is number-conserved, for all bitstrings k that are an element of $\ker \mathcal{E}$, $D_H(k, 0) \geq 2N + 2$ if $D_H(k, 0) \equiv 0 \pmod{2}$.

Proof. This theorem claims that the even Hamming weight code words, generated from the kernel space, must be lower bounded by $2N + 2$.

Suppose there is a bitstring $\vec{k}$ with Hamming weight $2K < 2N + 2$. We can construct two distinct bitstrings with Hamming weight N such that $\vec{d} = \vec{c} \oplus \vec{k}$, and $\vec{c}, \vec{d}$ have $N - K$ overlapping bits. We observe the following relation in the encoded space:

$$\mathcal{E}(\vec{d}) = \mathcal{E}(\vec{c}) \oplus \mathcal{E}(\vec{k}) = \mathcal{E}(\vec{c}). \quad (\text{C1})$$

■

Proof of Theorem 1. From Theorem 1, we observe that all even kernel elements k , generated by the dual code $\mathbf{C}^T$, have even Hamming weight lower bounded as $D_H(k, 0) \geq 2N + 2$. We can interpret the row of $\mathbf{C}$ as an $M - Q$ set of M -bit bitstrings, whose arbitrary bitwise addition $\oplus$ among the bitstrings is encoded into the $\mathbf{0}$ vector in the encoded space. Since $\mathbf{C}$ is invariant under elementary bitwise row addition, it is possible to eliminate the odd Hamming weight M -bit bitstrings with odd Hamming weight M -bit bitstrings. $\mathbf{C}$ can be reorganized as a set of M -bit bitstrings with *only one* element with odd Hamming weight. In this form, we can impose Theorem 1 on the $M - Q - 1$ numbers of M -bit bitstrings.

Next, we reformulate the constraint as finding the maximal $M - Q - 1$ of $\mathbf{C}$ given qubits M , and minimal hamming distance $2N + 2$, and all Hamming weight of elements in $\mathbf{C}$ are even. Borrowing from CEC, the maximum is upper bounded by the Hamming bound [39] and lower bounded by the Gilbert-Varshamov bound [19,20]:

$$\begin{aligned} \frac{2^{M-1}}{\sum_{j=0}^N \binom{M}{2j}} &\leq 2^{M-Q-1} < \frac{2^{M-1}}{\sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{M}{2j}}, \quad (\text{C2}) \\ M - 1 - \log \sum_{j=0}^N \binom{M}{2j} &\leq M - Q - 1 \\ &< M - 1 - \log \sum_{j=0}^{\lfloor N/2 \rfloor} \binom{M}{2j}, \end{aligned}$$

$$2N \log M \geq Q > 2 \left\lfloor \frac{N}{2} \right\rfloor \log M. \quad (\text{C3})$$

We modify the Hamming and Gilbert-Varshamov bounds in Eq. (C2) such that we are only considering the bound on the even Hamming weight subspace, which has size 2^{M-1} , and the sum $\sum_{j=0}^N \binom{M}{2j}$ is only over the even binomial coefficients. However, the argument of the Hamming cube as the maximal nonoverlapping covering remained unchanged [19]. We also note that the Hamming bound for even subspace becomes a strict nonequality because the covering space does not cover the whole even code subspace.

We have proved that optimal bitwise linear encoding has $\mathcal{O}(N \log M)$ scaling. In particular, the lower bound coincides with the scaling of physical states.

APPENDIX D: PROOF FOR MEASUREMENT COMPLEXITY

Lemma D1. All fermionic operators that can be written as a product of creation and annihilation operators can be expressed in terms of the product of one X string and sum over P strings under linear compression.

Proof. Let us consider the Jordan-Wigner basis. The one-particle creation and annihilation operators are written as the product of the Pauli-Z operator Z_m , Pauli-X operator X_i , and projector $P_i^{0/1}$ defined in Definition B1:

$$\hat{a}_i^\dagger = \prod_{m=1}^{i-1} Z_m X_i P_i^0, \quad (\text{D1})$$

$$\hat{a}_j = \prod_{m=1}^{j-1} Z_m X_j P_j^1. \quad (\text{D2})$$

Using the completeness relation of the identity operator, it is possible to expand $P_i^{0/1}$ as a sum over projector strings $P^{a_1 a_2 \dots a_{M-1}}$ defined in Definition B2:

$$P_i^{0/1} = P_i^{0/1} \otimes \sum_{\vec{a} \in \{0,1\}^{M-1}} P^{\vec{a}}. \quad (\text{D3})$$

Thus, the Z strings $\prod_{m=1}^{i-1} Z_m$ from the above example can be absorbed into the projectors via the relationship

$$Z^{\vec{c}} P^{\vec{b}} = (-1)^{\vec{c} \cdot \vec{b}} P^{\vec{b}}. \quad (\text{D4})$$

After this procedure, fermionic observables can be represented with only the X strings, projectors, and signs. Each projector represents a fermionic state, and the unphysical states are discarded such that the encoded projectors would not coincide with the number-conserved projectors due to the equivalent classes. Let us denote $\hat{O}$ as the product of creation and annihilation operators. Given a linear encoder $\mathcal{E}(\cdot)$, we can succinctly express the unencoded and encoded gates as follows:

$$\hat{O} = X^{\vec{a}} \sum_{\vec{b} \in S_{\hat{O}}} (-1)^{\vec{c} \cdot \vec{b}} P^{\vec{b}}, \quad (\text{D5})$$

$$\mathcal{E}(\hat{O}) = X^{G\vec{a}} \sum_{\vec{b} \in S_{\hat{O}}} (-1)^{\vec{c} \cdot \vec{b}} P^{G\vec{b}}, \quad (\text{D6})$$

where $S_{\hat{O}}$ represents the set of number-conserved states on which $\hat{O}$ acts and $\vec{a}, \vec{c} \in \mathbb{F}_2^M$ are binary vectors uniquely associated with each operator $\hat{O}$: $\vec{a}$ induces physical state transition and $\vec{c}$ encodes the sign information for each vector $\vec{b}$.

Lemma D2. The real and imaginary parts of an observable $\text{Re}/\text{Im}\hat{O}$ of the form in Eq. (D6) can be measured with one measurement basis under linear encoding.

Proof. From Eqs. (B14) and (D6), we can write the (anti-)Hermitian part of the operators as Eq. (5):

$$\text{Re}/\text{Im}\mathcal{E}(\hat{O}) = X^{\vec{G}\vec{a}} \sum_{\vec{b} \in S_{\hat{O}}} (-1)^{\vec{c}\vec{b}} (P^{\vec{G}\vec{b}} \pm P^{\vec{G}(\vec{a} \oplus \vec{b})}). \quad (\text{D7})$$

We make use of the following CNOT transformation to eliminate the extra Pauli- X operations:

$$\text{CNOT}_{1,2} X_1 X_2 \text{CNOT}_{1,2} = X_1, \quad (\text{D8})$$

$$\text{CNOT}_{1,2} P^{a_1 a_2} \text{CNOT}_{1,2} = P^{a_1(a_1 \oplus a_2)}, \quad (\text{D9})$$

where we define i as control qubit and j as target qubit for $\text{CNOT}_{i,j}$. Applying Eq. (D9) on a pair of conjugate projectors in the form of Eq. (B14), we observe that the projector at the target qubit will have the same parity. The full set of CNOT gates $\hat{C}$ results in the following transformation:

$$\hat{C} X^{\vec{G}\vec{a}} \hat{C}^\dagger = X_B, \quad (\text{D10})$$

$$\hat{C} P^{\vec{G}\vec{b}} \hat{C}^\dagger = P^{\vec{M}\vec{G}\vec{b}} = P_B^0 P^{\vec{m}}, \quad (\text{D11})$$

$$\hat{C} P^{\vec{G}(\vec{a} \oplus \vec{b})} \hat{C}^\dagger = P^{\vec{M}\vec{G}(\vec{a} \oplus \vec{b})} = P_B^1 P^{\vec{m}}, \quad (\text{D12})$$

where the adjoint action of $\hat{C}$ on P is equal to the $Q \times Q$ matrix $\mathbf{M}$ action on $\vec{G}\vec{b}$, such that $\mathbf{M}(n\vec{a}) \oplus \vec{b} = ((n \bmod 2)_B, \vec{m}(\vec{b}))$, with B denoting the bit location on which X_B acts and n an integer. We write $\vec{m} = \vec{m}(\vec{b})$ as a function of $\vec{b}$ to simplify the notation in the sum. The operator becomes

$$\hat{C} \text{Re}/\text{Im}\mathcal{E}(\hat{O}) \hat{C}^\dagger = \sum_{\vec{b} \in S_{\hat{O}}} (-1)^{\vec{c}\vec{b}} X_B (P_B^0 P^{\vec{m}(\vec{b})} \pm P_B^1 P^{\vec{m}(\vec{b})}). \quad (\text{D13})$$

On qubit B , the sum of P^0 and P^1 generates local I or difference generates Z_N . We can rewrite them as X_N or $-iY_N$, as shown in Eq. (E1).

Since all operators, except the qubit Q that is either X_i when Hermitian or iY_Q otherwise, are projectors made of I, Z operators, the real/imaginary part of the bitwise linear encoded operators must generate commuting Pauli string in the Clifford transformed basis. Finally, since Clifford transforms one-to-one encoded Pauli strings, the operator $\text{Re}/\text{Im}\mathcal{E}(\hat{O})$ also generates commuting Pauli strings. ■

Proof of Theorem 2. We only need to determine the measurement complexity of 2-RDM.

Next, from Lemmas D1 and D2, we see that the (anti-)Hermitian part of the 2-RDM, defined as $\langle \psi | \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l \pm \hat{a}_l^\dagger \hat{a}_k^\dagger \hat{a}_j \hat{a}_i | \psi \rangle$, can be measured in one basis for each set of fermionic modes i, j, k, l . In quantum chemistry, we are only interested in the real part of the observables. Therefore, each i, j, k, l demands one measurement basis.

Due to the result in Lemma D2, each distinct X string corresponds to one unique measurement basis. Upon encoding

the 2-RDM operator via the JW transformation, we see that the encoded operator is of the form of the real part of Eq. (D5). The X string in this JW representation is invariant under any permutation among i, j, k, l in the JW representation. Thus, the measurement scaling is equal to the possible combinations of i, j, k, l , where we allow the indices to be equal to each other. Next, we count the number of distinct X strings derived from the possible permutation. In the JW representation, when all indices are not equal to each other, the number of distinct X strings corresponds to a complete set of M -dimensional binary vectors with Hamming weight 4. Meanwhile, when one pair of indices share the same digit, it corresponds to the set of M -dimensional binary vectors with Hamming weight 2. When two pairs are equal, then it corresponds to M -dimensional binary vectors with Hamming weight 0. Under our formalism, each X string can be associated with a binary vector. Thus, the total measurement cost, without parallelization [13], is equal to $\binom{M}{0} + \binom{M}{2} + \binom{M}{4}$. Parallelization is the simultaneous measurement of two or more sets of RDM elements with nonoverlapping indices; maximal linear encoding breaks parallelization.

Finally, since a linear compression surjectively encodes the X string guaranteed by the $2N + 2$ code distance of the CEC $\mathbf{C}$, $\binom{M}{0} + \binom{M}{2} + \binom{M}{4}$ is the upper bound of the measurement scaling for the encoded space.

Corollary D1. The scaling of measurement basis can be completely characterized by the number of distinct X strings that appear in the X and P decomposition.

Proof. From the previous proof, each unique X string, obtained from the X and P decomposition of an operator, generates commuting Pauli strings. Meanwhile, Pauli strings generated by two distinct X strings, and two pairs of respective projectors, contain mutually noncommuting terms. ■

This result can also be employed to characterize nonlinear encoding, which *does not* surjectively encode X strings to X strings. Instead, Eq. (D7) will further split into a sum over X strings. Scalable measurement is still possible if the distinct X strings for each single/double-excitation operator scale polynomially.

APPENDIX E: FERMIONIC GATE COMPLEXITY

1. Gate decompositions without decoder

We examine fermionic gate complexity without quantum decoder.

Proposition E1. The fermionic gate of single and double excitations can be decomposed as a series of multicontrolled gates and Clifford gates.

Proof. From Lemma D2, we can transform operators in Eq. (D7) into X string with the Clifford transformation $\hat{C}$ and operator of the form

$$\hat{C} \text{Re}/\text{Im}\mathcal{E}(\hat{O}) \hat{C}^\dagger = \sum_{\vec{m} \in S_{\hat{O}}} X_B / -iY_B P^{\vec{m}(\vec{b})}, \quad (\text{E1})$$

where $X_B / -iY_B$ is the last Pauli- X/Y operator at qubit B . For the implementation of fermionic gates, it is necessary to describe the exponentiation of this operator representation. $\hat{C}$, described in Eqs. (D10), (D11), and (D12), survives under exponentiation. Meanwhile, each operator $X_N / -iY_N P^{\vec{m}(\vec{b})}$ is

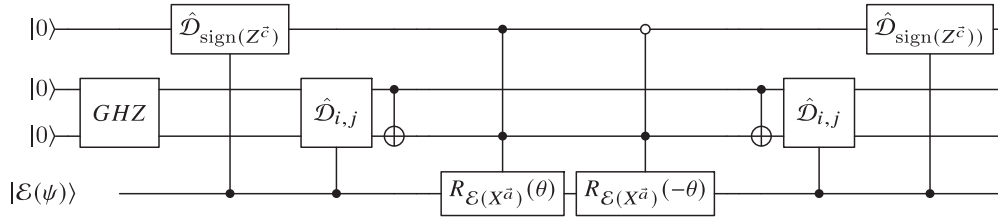


FIG. 7. Single-excitation quantum gate constructions based on sign and projector decoder in Figs. 3 and 4. The D operators represent the decoding subroutine and in between is the encoding subroutine in Fig. 1(b). Definitions of $\bar{a}$ and $\bar{c}$ follow the convention in Eq. (3).

mutually commutative, and its exponentiation corresponds to a multicontrolled X/Y rotation, with the projectors acting as the control bits. Consequently, the exponentiation of the diagonalized fermionic operators can be represented as a product of multicontrolled Pauli- X/Y gates.

For molecular simulation, where only the real part of fermionic operators matters, the multicontrolled gates with a Clifford transformation can exactly implement fermionic gates defined in Lemma D1. However, the number of multicontrolled gates given by the size of S_{δ} indicates that the worst-case scenario gate complexity has combinatorial scaling $\mathcal{O}(M^N)$ [10].

The segment code in Ref. [9] and polylog code in Ref. [11] circumvent the problem of decoding by imposing constraints on the encoder. Segment code employs block structure on the encoder to ensure that the compressed N -particle states can be locally distinguished upon encoding, partially preserving locality in fermionic gate operations. Although the corresponding explicit gate complexity is not specified, it is polynomial in both N and M based on the construction. Meanwhile, the polylog code further exploits encoding with block structure and Bravyi-Kitaev transformation for a more efficient fermionic information representation, and QSP to approximately implement single-qubit Pauli- Z operators, with a gate complexity lower scaling as $\mathcal{O}(N^2 \log M^4)$.

These approaches demonstrate example of scalable fermionic gate implementation on the compressed state, but they are fundamentally limited to their encodings. The QS-QEC duality provides a framework that generalizes these approaches.

2. Time evolution

a. Trotterization

Building on Sec. IV, we demonstrate two examples of time evolution algorithm encoding based on subroutines introduced in Figs. 3 and 4. First, we show that individual fermionic gates can be implemented with single-mode decoders. Unitary evolution of the single and double excitations is implemented in Figs. 7 and 8, respectively.

For each query of the single and double excitations, the single-mode decoders are queried at most $\mathcal{O}(M)$ times since the calculation of signs requires summing mode information from between fermionic modes $i < m < j$ for single excitations and additionally $k < n < l$ for double excitations. If the single and double excitations are queried G times in Trotterization, the query complexity of the single-mode decoders D_i

is given by

$$\mathcal{O}(GM). \quad (\text{E2})$$

We demonstrate two examples of time evolution algorithm encoding based on the subroutines introduced in Figs. 3 and 4. First, we show that individual fermionic gates can be implemented using single-mode decoders. The unitary evolution of single and double excitations is illustrated in Figs. 7 and 8, respectively.

However, the discussion centers around the JW basis. In principle, decoding can be done in the BK basis, leading instead to the query complexity of

$$\mathcal{O}(G \log M), \quad (\text{E3})$$

since only $\log M$ BK modes require decoding.

b. Block encoding

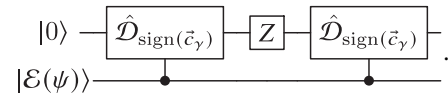
Meanwhile, the decoder is not limited to encoding fermionic single and double excitations; it is possible to encode second-quantized subroutines, such as block encoding for time evolution [40]. Given a second-quantized Hamiltonian of the form in Eq. (1) with a Pauli decomposition:

$$\hat{H} = \sum_i c_i P_i = \sum_i d_i Z^{\bar{c}_i} X^{\bar{a}_i}, \quad (\text{E4})$$

one can construct the $\mathcal{E}$ -SEL (SEL stands for select) primitive in the compressed space using just the sign decoder in Fig. 3:

$$\begin{aligned} \mathcal{E}\text{-SEL } |\gamma\rangle |0\rangle |\mathcal{E}(\psi)\rangle &= |\gamma\rangle \hat{O}_{\text{sign}(\bar{c}_\gamma)} |0\rangle \mathcal{E}(X^{\bar{a}_\gamma}) |\mathcal{E}(\psi)\rangle \\ &= |\gamma\rangle |0\rangle \mathcal{E}(Z^{\bar{c}_i}) \mathcal{E}(X^{\bar{a}_\gamma}) |\mathcal{E}(\psi)\rangle \\ &= |\gamma\rangle |0\rangle \mathcal{E}(P_\gamma) |\mathcal{E}(\psi)\rangle, \end{aligned} \quad (\text{E5})$$

where the operator $\hat{O}_{\text{sign}(\bar{c}_i)}$ is given by the following sign decoder construction from Fig. 3:



When combined with the prepared primitive

$$\text{PREP } |\gamma\rangle = \sum_i \sqrt{\frac{d_i}{N}} |i\rangle, \quad (\text{E6})$$

we obtain the following block encoding:

$$\begin{aligned} &|0\rangle |0\rangle \langle \mathcal{E}(\psi) | \text{PREP}^\dagger \mathcal{E}\text{-SEL } \text{PREP } |0\rangle |0\rangle |\mathcal{E}(\psi)\rangle \\ &= \langle \mathcal{E}(\psi) | \mathcal{E}(\hat{H}) | \mathcal{E}(\psi) \rangle / N. \end{aligned} \quad (\text{E7})$$

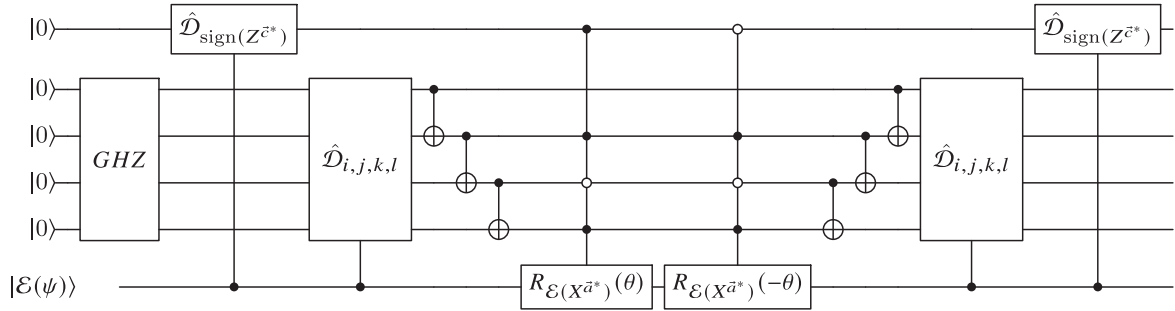


FIG. 8. Double-excitation quantum gate constructions based on sign and projector decoder in Figs. 3 and 4. The $\hat{D}$ operators represent the decoding subroutine and in between is the encoding subroutine in Fig. 1(b). $\vec{a}^*$ and $\vec{c}^*$ denote binary vectors for double excitations.

$|\gamma\rangle$ takes up $4 \log M$ qubits, corresponding to the $\mathcal{O}(M^4)$ Pauli terms in $\hat{H}$. Thus, G query of the $\mathcal{E}$ -SEL primitive corresponds to querying $\mathcal{O}(GM^4)$ $\hat{D}_{\text{sign}(\vec{c})}$ decoder and hence

$$\mathcal{O}(GM^5) \quad (\text{E8})$$

single-mode fermionic decoders.

3. Qubit-gate complexity trade-off

Via the QS-QEC duality, we have demonstrated two examples of second-quantized subroutines in the linearly compressed basis using $\mathcal{O}(M)$ single-mode fermionic decoders. In exchange for the qubit reduction, the single-mode fermionic decoders result in an increase in gate complexity. The duality provides a natural framework for studying the qubit-to-gate complexity trade-off. In what follows, we will provide both lower and upper bounds on decoder complexity and give examples of efficient constructions.

Since the ancillary qubits are reusable, fermionic excitation evolution can be queried sequentially without increasing the number of ancillary qubits. In the absence of compression, the single-mode decoders are simply implemented using CNOT gates, as shown by the following relation:

$$\text{CNOT}_{i,a} |m_i 0_a\rangle = |m_i m_a\rangle, \quad (\text{E9})$$

where m_i denotes the occupation number at mode i and a represents the ancillary qubit. Consequently, the $\hat{D}_{i,j}/\hat{D}_{i,j,k,l}$ decoders have $\mathcal{O}(1)$ complexity, while the $\hat{D}_{\text{sign}}$ decoder has $\mathcal{O}(M)$ complexity.

In the worst-case scenario, the $\hat{D}_{i,j}/\hat{D}_{i,j,k,l}$ and $\hat{D}_{\text{sign}}$ decoders can be trivially constructed using $\mathcal{O}(M^N)$ multi-controlled gates, similar to the construction in Eq. (D13). Therefore, the single-mode fermionic decoder has complexity lower bound $C_{\hat{D}_i}$ given by

$$\mathcal{O}(1) < C_{\hat{D}_i}. \quad (\text{E10})$$

Using QS-QEC duality, the question of qubit-gate-measurement trade-off can be studied in terms of the decoder space-time complexity scaling as one compresses the qubit resources for the time resource.

APPENDIX F: FERMIONIC EXPECTATION DECODER

1. Compatibility with classical decoder

FED, as depicted in Algorithm 1, avoids direct computation of fermionic/spin operator in the encoded space through

a decoding process. This approach circumvents the issue of measuring an unscalable number of Pauli operators and maintains polynomial scaling in measurement complexity. FED computes the expectation of the operator in Eq. (D13), from which a probability distribution $P(\vec{d})$ is sampled in the basis $\vec{d} = (1/0_B, \vec{m}(\vec{b}))$. To decode the bitstrings from $\vec{G}\vec{b}$ to $\vec{b}$, reversing the Clifford transformation is necessary to retrieve the projector $\vec{G}\vec{b}$ from $\vec{d}$ —a task that is infeasible for classical postprocessing.

Instead, when computing the fermionic sign $(-1)^{\vec{b} \cdot \vec{c}}$, one can focus solely on recovering either $P^{\vec{G}\vec{b}}$ or $P^{\vec{G}(\vec{a} \oplus \vec{b})}$ as shown in Eq. (5) via partial recovery transformation. Collectively, this is equivalent to retrieving the projector information $X^{\vec{G}\vec{a}}(P^{\vec{G}\vec{b}} + P^{\vec{G}(\vec{a} \oplus \vec{b})})$. This can be achieved through the CNOT transformation $\hat{C}$ in Eqs. (D11) and (D12), which can be efficiently implemented on a classical computer directly on the bitstrings of the histogram as **A**.

Under our operator decomposition, we avoid two problems that arise in nonlinear encoding: First, we avoid preparing an exponential number of measurement bases; and second, we avoid calculating expectation values in the encoded space, which would decompose into an exponential number of Pauli strings. The projector representation in Eq. (5) avoids the Pauli representation of the encoded operator under conserved subspace compression. This representation establishes a direct link to the decoding protocol of the CEC code and classifies the complexity of measurement grouping within the conserved subspace. Overall, this leads to a scalable encoding and decoding process for simulating systems with conserved particle numbers.

2. The decoding algorithm

Results in Appendix D can be directly employed for measurement basis preparations. Combined with the FED decoder, we have a quantum-classical observable decoder of compressed number-conserved fermionic data (Algorithm 2).

The Clifford gate preparation algorithm CLIFFORD($\cdot$) is based on the Clifford decomposition of Eqs. (D13) and (E1) with a single-qubit Clifford gate to diagonalize the $X_B/ -iY_B$ Pauli operator. Meanwhile, GETVEC($\cdot$) is the extraction of the binary vector representation of $\hat{O}_i$ from the XP decomposition in the JW basis. MEAS($\cdot$) is measurement of the encoded quantum state in the quantum computer.

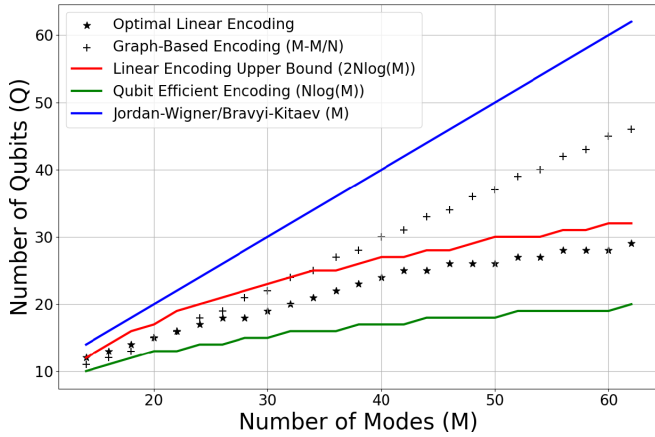


FIG. 9. Comparison of the qubit cost of RLE (star) to the JW basis (blue), graph-based encoding (cross), QEE (green) for a four-electron problem, and the $2N \log M$ bound (red) against the number of orbitals.

3. Applications

The FED algorithm, leveraging the duality between CEC and fermionic observables, is applicable to general linearly compressed quantum states. Combining with the QS-QEC duality introduced in Appendix E, the FED algorithm finds applications across NISQ and FTQC regimes depending on the qubit-to-gate trade-offs in the compression, although the numerical results place no emphasis on the encoding of quantum gates.

For a moderate/Gilbert bounded qubit compression rate such as segment code and QNN decoders, although gate complexity increases, the FED algorithm finds application in decoding such compressed NISQ/early FTQC algorithms such as Trotterization. Meanwhile, as we will show in the following, the encoding/decoding protocol can be useful for encoding qubit ADAPT-VQE, providing a promising way of minimizing qubits for NISQ algorithms.

APPENDIX G: QUBIT ADAPT-VQE

In our numerical results, we have demonstrated the benefits of training VQE in a number-conserved subspace while ensuring that the encoding/decoding process remains scalable. However, the demonstration is limited to nonphysically motivated ansätze, and it remains unexplored whether the RLE-FED protocol can facilitate the scalable implementation of VQE in the compressed subspace.

We present an algorithm that employs the RLE-FED scheme for encoding NISQ algorithms beyond nonphysically motivated ansätze. While compression disrupts the structure of fermionic gates, it is possible to encode Hamiltonian information directly into qubit operators using qubit ADAPT-VQE [25]. The algorithm selects the appropriate qubit gates by identifying the qubit operator that maximizes the gradient within an operator pool:

$$\frac{\partial}{\partial \theta_i} \langle \hat{H} \rangle = \langle \mathcal{E}(\psi) | [\mathcal{E}(\hat{H}), \hat{\tau}_i] | \mathcal{E}(\psi) \rangle, \quad (\text{G1})$$

where $\hat{H}$ is the fermionic Hamiltonian, $\hat{\tau}_i$ is a qubit operator belonging to the operator pool, and $|\mathcal{E}(\psi)\rangle$ represents the encoded fermionic state in the qubit space. Specifically, to implement the operator $[\mathcal{E}(\hat{H}), \hat{\tau}_i]$, we make use of the parameter shift rule. $\mathcal{E}(\hat{H})$ is the Hamiltonian in the encoded space that

ALGORITHM 2. Decode.

Input: Encoder $\mathcal{E}(\cdot)$, Decoder $F(\cdot)$, Encoded Quantum State $|\mathcal{E}(\psi)\rangle$, Hamiltonian $\hat{H} = \sum_i c_i \hat{O}_i$, $\hat{O}_i^\dagger = \hat{O}_i$

Output: $\langle \psi | \hat{H} | \psi \rangle$

```

sum = 0
for  $c_i, \hat{O}_i$  in  $\hat{H}$ 
   $\hat{C} \leftarrow \text{CLIFFORD}(\mathcal{E}(\hat{O}_i))$ 
   $\vec{a}, \vec{c} \leftarrow \text{GETVEC}(\hat{O}_i)$ 
   $P \leftarrow \text{MEAS}(\hat{C}^\dagger | \mathcal{E}(\psi))$ 
  val  $\leftarrow \text{FED}(P, \vec{a}, \vec{c}, \mathcal{E}, F)$ 
  sum  $\leftarrow \text{sum} + c_i * \text{val}$ 
end for
return sum

```

requires decoding, and $\hat{\tau}_i$ is a qubit operator that acts directly on the qubit space. One can act $R_{\hat{\tau}_i}(\pm\pi/2)$ directly on $|\mathcal{E}(\psi)\rangle$ and compute the gradient via decoding the following term:

$$\frac{\partial}{\partial \theta_i} \langle \hat{H} \rangle = \frac{\langle R_{\hat{\tau}_i}(\pi/2)^\dagger \mathcal{E}(\hat{H}) R_{\hat{\tau}_i}(\pi/2) \rangle - \langle R_{\hat{\tau}_i}(-\pi/2)^\dagger \mathcal{E}(\hat{H}) R_{\hat{\tau}_i}(-\pi/2) \rangle}{2}. \quad (\text{G2})$$

In particular, the terms $\langle R_{\hat{\tau}_i}(\pi/2)^\dagger \mathcal{E}(\hat{H}) R_{\hat{\tau}_i}(\pi/2) \rangle$ and $\langle R_{\hat{\tau}_i}(-\pi/2)^\dagger \mathcal{E}(\hat{H}) R_{\hat{\tau}_i}(-\pi/2) \rangle$ can be classically decoded. This leads to the subroutine for qubit operator selection in qubit compressed ADAPT-VQE in Algorithm 3. Qubit ADAPT-VQE shows evidence that avoids barren plateau and local minima, and it demonstrates CNOT gate reduction compared to the implementing fermionic ADAPT-VQE.

When combining RLE-FED with ADAPT-VQE, Eq. (G1) can be efficiently sampled. Building on the results from Ref. [25], the qubit operator pool $\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_n$ scales linearly with the number of qubits Q . Under our scheme's qubit scaling, the size of this operator pool is effectively reduced from M to $2N \log M$, potentially enabling VQE simulations involving a large number of fermionic modes. This reduction in operator pool size yields practical benefits, as demonstrated in our LiH simulation, where chemical accuracy is achieved with significantly fewer CNOT gates compared to the JW-encoded

ALGORITHM 3. Qubit operator optimizer.

Input: Encoder $\mathcal{E}(\cdot)$, Decoder $F(\cdot)$, Qubit operator pool $\text{Pool} = \{\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_n\}$, Hamiltonian $\hat{H}$, compressed quantum state $|\mathcal{E}(\psi)\rangle$

Output: $\text{Max}_{\hat{\tau}_i \in \text{Pool}} \langle \mathcal{E}(\psi) | [\mathcal{E}(\hat{H}), \hat{\tau}_i] | \mathcal{E}(\psi) \rangle$

```

L = [0]
g_max = 0
for  $\hat{\tau}_m \in \text{Pool}$  do
   $|\psi^+\rangle = e^{i\hat{\tau}_m\pi/2} |\mathcal{E}(\psi)\rangle$ 
   $|\psi^-\rangle = e^{-i\hat{\tau}_m\pi/2} |\mathcal{E}(\psi)\rangle$ 
   $g = \left| \frac{\text{DECODE}(\mathcal{E}, F, |\psi^+\rangle, \hat{H}) - \text{DECODE}(\mathcal{E}, F, |\psi^-\rangle, \hat{H})}{2} \right|$ 
  if  $g > g_{\text{max}}$  then
    L[0]  $\leftarrow \hat{\tau}_m$ 
    g_max  $\leftarrow g$ 
  end
  APPEND(L, ( $g_m, \hat{\tau}_m$ ))
end for
return L[0], g_max

```

basis. Furthermore, fermionic observable decoding plays a key role in reducing quantum gate overhead, enhancing the feasibility of near-term applications.

APPENDIX H: NUMERICAL DETAILS

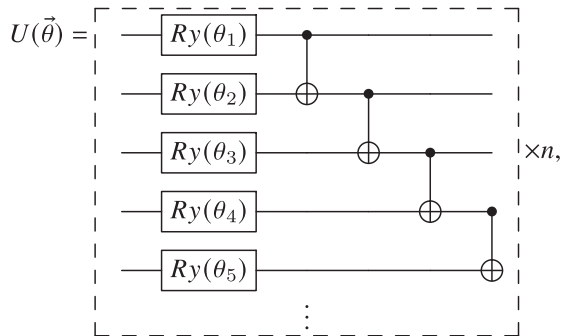
1. The compression rate

Table II shows the numerical study of the compression rate Q/M given Q . This table directly compares to the Graph-Based encoding table in Ref. [10], which shows qubit advantages in all settings. Meanwhile, the polylogarithmic encoding cannot be compared with because their encoding advantage appears only in the 10^5 qubit regime. Logarithmic scaling in M means that it is possible to compress a large number of fermionic modes with a qubit cost that is manageable in the FTQC regime. For instance, a two-electron problem with 100 modes can be encoded with fewer than 20 qubits and a four-electron problem with 400 modes in 50 qubits.

To analyze the scaling with respect to the number of modes, we evaluated the compression rate Q/M and presented the results in Fig. 9 for a four-electron problem. We benchmarked our results against the JW basis, graph-based encoding, and the theoretical upper and lower bounds. Our compression rate for linear encoding surpasses that of all previous works [10] across all scenarios. Compared to the nonlinear QEE scheme, we maintain $\mathcal{O}(N \log M)$ scaling while achieving polynomial measurement scaling.

2. Hardware-efficient ansatz

The hardware-efficient ansatz is commonly used in variational quantum eigensolver due to its resource efficiency given the current noisy constraints of quantum devices. Different hardware setups have their own specific HEA configurations, which include the selection of gates, circuit connectivity, and noise models. To meet these criteria, we design our HEA with two constraints. First, the gates used in our VQE simulations are restricted to the two-qubit controlled-NOT gate and the single-qubit Y gate for rotations. Second, the connectivity of the two-qubit gates is constrained to a one-dimensional chain topology. The design of the ansatz is therefore as follows:



where n is the number of repetitions of the circuit block.

3. Comparison with unitary couple cluster type of ansatz

We compared the performance of the hardware-efficient ansatz in the compressed space to the UCC-type ansatz [41], which is the gold standard for computational chemistry sim-

ulations, specifically for the ground-state simulation of LiH in the STO-3G and 6-31G basis sets. For the active space chosen as $M = 8$ and $M = 16$ fermionic modes, respectively, we restricted the Hilbert space generated from both bases to the subspace of two electrons, denoted as $(8, 2)$ and $(16, 2)$.

In the STO-3G case, both the 6-qubit compressed and 8-qubit uncompressed Hamiltonians achieved chemical accuracy of 1 Kcal/mol using HEA with a reasonable circuit depth. However, the 6-qubit Hamiltonian achieved accuracy four orders of magnitude higher compared to the 8-qubit Hamiltonian, with fewer parameters and shallower circuit depth.

In contrast, the UCC-type ansatz converges faster and more easily compared to HEA, owing to its smaller number of parameters. However, achieving approximately one order of magnitude higher accuracy than HEA requires a circuit depth approximately two orders of magnitude deeper and roughly ten times more two-qubit gates. However, when scaling up to the $(16, 2)$ LiH system, the hardware-efficient ansatz applied directly to the uncompressed Hilbert space starts exhibiting barren plateaus. Even with L-BFGS-B optimization using 100 initial states, convergence to the required chemical accuracy is not achieved.

In contrast, using the RLE approach, which encodes the original 16-qubit Hamiltonian into an 8-qubit Hamiltonian, HEA can successfully converge to chemical accuracy.

In conclusion, the LiH system scaling result shows that the qubit compression can improve VQE convergence while reducing the qubit cost and circuit depth.

APPENDIX I: RANDOMIZED LINEAR ENCODER

1. The algorithm

RLE (Algorithm 4) begins its random search by choosing a Q within the bounded region $N \log_2 M < Q < 2N \log_2 M$. Then, the algorithm initializes the parity check matrix $\mathbf{G} = [I_Q | D]$ in the *standard basis* [42], where I_Q is a $Q \times Q$ identity

ALGORITHM 4. Randomized linear encoder.

Input: Target States $\mathbf{S}$; Testing subset T_s ;
Output: Code word $\mathbf{C}$;

while True **do**
 Initialized Code word $\mathbf{W} = []$;
 Random generate D with constraints;
 $\mathbf{G} \leftarrow [I_Q | D]$;
 Create distance $2N + 2$ error correction matrix $\mathbf{C} \leftarrow [-D^T | I_{M-Q}]$;
 Create a checking list $w = []$
 for t_s in T_s **do**
 $w \leftarrow [w | D_H(\sum C^T(t_s) \bmod 2, \mathbf{0})]$;
 end for
 if $\min(w) \geq 2N + 2$ **then**
 for s in $\mathbf{S}$ **do**
 $\mathbf{c} = \mathbf{C}s$;
 $\mathbf{W} = [\mathbf{W} | \mathbf{c}]$;
 end For
 if no repeat element in $\mathbf{W}$ **then**
 return $\mathbf{W}$
 end If
 end if
end while

matrix. The standard basis helps reduce the space of randomized search. We randomly generate the $Q \times (M - Q)$ matrix D such that each column has even Hamming weight equal to the value $Q/2$, denoted as $\text{even}(Q/2)$.

The second part of the algorithm consists of checks. Initially, the algorithm censors a large set of flawed generators by randomly verifying the Hamming weight of the element $\vec{k}$ generated by the $(M - Q)$ -dimensional vector $\vec{v}$ and the CEC C :

$$\vec{k} = C\vec{v}, \vec{v} \in S, \quad (11)$$

where S is defined as the set of all vectors with Hamming weight 2 and $2K$ for $2K < N$, where K is a randomly generated integer. The random check serves to maximize the probability of obtaining the correct encoder. Finally, the algorithm explicitly computes the states to check for repetitions. This process is repeated until the minimum Q is found.

2. Scaling of RLE

We examine the scalability of RLE. The algorithm has a computation cost proportional to $\mathcal{O}(M^N)$ for each physical state check, multiplied by the number of checks performed. This complexity scaling is comparable to the Brouwer-Zimmermann algorithm, which is recognized as the most efficient method for checking the minimal distance of C [43]. RLE also shares the same computational cost as nonlinear encoding [12], but it is much simpler to implement. We can further enhance RLE scalability by incorporating auxiliary qubits into the optimal encoding. Figure 10 summarizes the

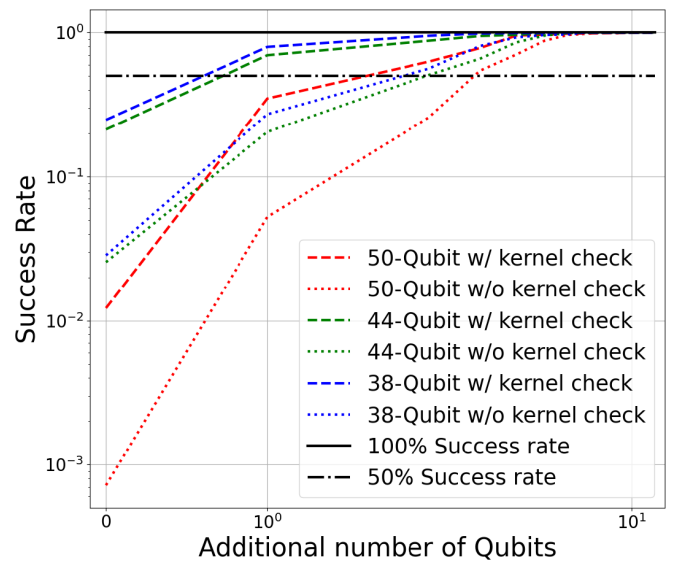


FIG. 10. Impact on the success probability of generating G when adding extra qubits to the generator. The x axis shows the number of auxiliary qubits and the y axis is the success rate. The dashed line, which resulted from a code-word check generated in Eq. (11), shows a systematic improvement in success rate compared to the unchecked (dotted line) case.

success probability of generating the correct parity check matrix G for four-electron systems with auxiliary qubits. The algorithm's run-time exponentially converges to $\mathcal{O}(M^N)$ —the complexity for one check—when auxiliary qubits are successively added.

- [1] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, *Chem. Rev.* **120**, 12685 (2020).
- [2] A. Kitaev and C. Laumann, Topological phases and quantum computation, *arXiv:0904.2771*.
- [3] C. W. Bauer, Z. Davoudi, A. B. Balantekin, T. Bhattacharya *et al.*, Quantum simulation for high-energy physics, *PRX Quantum* **4**, 027001 (2023).
- [4] A. Ho, J. McClean, and S. P. Ong, The promise and challenges of quantum computing for energy storage, *Joule* **2**, 810 (2018).
- [5] R. Santagati, A. Aspuru-Guzik, R. Babbush, M. Degroote *et al.*, Drug design on quantum computers, *Nat. Phys.* **20**, 549 (2024).
- [6] J. Zhang, P. W. Hess, A. Kyprianidis, P. Becker *et al.*, Observation of a discrete time crystal, *Nature (London)* **543**, 217 (2017).
- [7] R. P. Feynman, Simulating physics with computers, *Int. J. Theor. Phys.* **21**, 467 (1982).
- [8] P. Jordan and E. P. Wigner, About the Pauli exclusion principle, *Z. Phys.* **47**, 631 (1928).
- [9] M. Steudtner and S. Wehner, Fermion-to-qubit mappings with varying resource requirements for quantum simulation, *New J. Phys.* **20**, 063010 (2018).
- [10] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic Hamiltonians, *arXiv:1701.08213*.
- [11] W. Kirby, B. Fuller, C. Hadfield, and A. Mezzacapo, Second-quantized fermionic operators with polylogarithmic qubit and gate complexity, *PRX Quantum* **3**, 020351 (2022).
- [12] Y. Shee, P. K. Tsai, C. L. Hong, H. C. Cheng, and H. S. Goan, Qubit-efficient encoding scheme for quantum simulations of electronic structure, *Phys. Rev. Res.* **4**, 023154 (2022).
- [13] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang *et al.*, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, *Npj Quantum Inf.* **7**, 23 (2021).
- [14] J. T. Seeley, M. J. Richard, and P. J. Love, The Bravyi-Kitaev transformation for quantum computation of electronic structure, *J. Chem. Phys.* **137**, 224109 (2012).
- [15] J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi, Convergence of many-body wave-function expansions using a plane-wave basis: From homogeneous electron gas to solid state systems, *Phys. Rev. B* **86**, 035111 (2012).
- [16] A. Grüneis, J. J. Shepherd, A. Alavi, D. P. Tew, and G. H. Booth, Explicitly correlated plane waves: Accelerating convergence in periodic wavefunction expansions, *J. Chem. Phys.* **139**, 084112 (2013).
- [17] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug *et al.*, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).

- [18] T.-C. Yen, R. A. Lang, and A. F. Izmaylov, Exact and approximate symmetry projectors for the electronic structure problem on a quantum computer, *J. Chem. Phys.* **151**, 164111 (2019).
- [19] E. N. Gilbert, A comparison of signalling alphabets, *Bell Syst. Tech. J.* **31**, 504 (1952).
- [20] R. R. Varshamov, Estimate of the number of signals in error correcting codes, *Dokl. Akad. Nauk SSSR* **117**, 739 (1957).
- [21] G. Katsman, M. Tsfasman, and S. Vladut, Modular curves and codes with a polynomial construction, *IEEE Trans. Inf. Theory* **30**, 353 (1984).
- [22] S. Porter, B.-Z. Shen, and R. Pellikaan, Decoding geometric goppa codes using an extra place, *IEEE Trans. Inf. Theory* **38**, 1663 (1992).
- [23] R. Gallager, Low-density parity-check codes, *IEEE Trans. Inf. Theory* **8**, 21 (1962).
- [24] J. Mosheiff, N. Resch, N. Ron-Zewi, S. Silas, and M. Wootters, Low-density parity-check codes achieve list-decoding capacity, *SIAM J. Comput.* **53**, FOCS20-38 (2024).
- [25] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley *et al.*, Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor, *PRX Quantum* **2**, 020310 (2021).
- [26] G. H. Low and I. L. Chuang, Optimal Hamiltonian simulation by quantum signal processing, *Phys. Rev. Lett.* **118**, 010501 (2017).
- [27] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature (London)* **549**, 242 (2017).
- [28] X. Zeng, Y. Fan, J. Liu, Z. Li, and J. Yang, Quantum neural network inspired hardware adaptable ansatz for efficient quantum simulation of chemical systems, *J. Chem. Theory Comput.* **19**, 8587 (2023).
- [29] A. Cassagne, O. Hartmann, M. Léonardon, K. He *et al.*, AFF3CT: A fast forward error correction toolbox! *SoftwareX* **10**, 100345 (2019).
- [30] J. R. McClean, K. J. Sung, I. D. Kivlichan, Y. Cao *et al.*, Openfermion: The electronic structure package for quantum computers, *Quantum Sci. Technol.* **5**, 034014 (2020).
- [31] M. H. Cheng and A. C. Medina, Demonstration of Fermionic Expectation Decoder, 2024, <https://github.com/Tcsg22/RLE-FED-protocol-demo>.
- [32] C. H. Bennett, Time/space trade-offs for reversible computation, *SIAM J. Comput.* **18**, 766 (1989).
- [33] Y. Gurevich and A. Blass, Quantum circuits with classical channels and the principle of deferred measurements, [arXiv:2107.08324](https://arxiv.org/abs/2107.08324).
- [34] J. E. Rice, T. P. Gujarati, M. Motta, T. Y. Takeshita *et al.*, Quantum computation of dominant products in lithium–sulfur batteries, *J. Chem. Phys.* **154**, 134115 (2021).
- [35] J. Du, N. Xu, X. Peng, P. Wang *et al.*, Nmr implementation of a molecular hydrogen quantum simulation with adiabatic state preparation, *Phys. Rev. Lett.* **104**, 030502 (2010).
- [36] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [37] D. F. Locher, L. Cardarelli, and M. Müller, Quantum error correction with quantum autoencoders, *Quantum* **7**, 942 (2023).
- [38] C.-K. Lee, C.-Y. Hsieh, S. Zhang, and L. Shi, Variational Quantum Simulation of Chemical Dynamics with Quantum Computers, *J. Chem. Theory Comput.* **18**, 2105 (2022).
- [39] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes* (North-Holland Pub. Co., Amsterdam, 1977).
- [40] R. Babbush, D. W. Berry, Y. R. Sanders, I. D. Kivlichan *et al.*, Exponentially more precise quantum simulation of fermions in the configuration interaction representation, *Quantum Sci. Technol.* **3**, 015006 (2018).
- [41] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg *et al.*, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents? *J. Chem. Phys.* **152**, 124107 (2020).
- [42] V. Pless, *Introduction to the Theory of Error-Correcting Codes* (John Wiley & Sons, Ltd., Bognor Regis, UK, 1998), Chap. 1–2, pp. 1–38.
- [43] F. Hernando, F. D. Igual, and G. Quintana-Ortí, Algorithm 994: Fast implementations of the Brouwer-Zimmermann algorithm for the computation of the minimum distance of a random linear code, *ACM Trans. Math. Softw.* **45**, 1 (2019).