

**From Molecules to Qubits: A Complete Guide to  
Fermion-to-Qubit  
Encoding for Quantum Chemistry Simulation**

John Azariah<sup>\*</sup>

*University of Technology Sydney*

(Dated: February 14, 2026)

# Abstract

Simulating molecules on quantum computers requires translating the fermionic operators of quantum chemistry into the qubit operators of quantum hardware. This translation, the fermion-to-qubit encoding, is a critical step that determines the structure, cost, and feasibility of the resulting quantum algorithm. Despite its importance, no single pedagogical reference covers the complete pipeline from molecular integrals to qubit Hamiltonian with every step made explicit.

This article provides that reference. Using the hydrogen molecule ( $\text{H}_2$ ) in the STO-3G basis as a running example, we develop each stage in sequence: the Born–Oppenheimer approximation, basis-set discretization, second quantization, notation conversion, spin-orbital expansion, and fermion-to-qubit encoding. We derive every integral, operator product, and Pauli coefficient explicitly, obtaining the 15-term qubit Hamiltonian measured in practical quantum algorithms. We then validate the construction by exact diagonalization and recover the known ground-state energy.

We cover the Jordan–Wigner transform in full detail and survey four alternatives (Bravyi–Kitaev, Parity, balanced binary tree, balanced ternary tree), explaining how each trades off locality, weight, and complexity. A companion open-source implementation in F# reproduces every calculation in this paper.

## I. INTRODUCTION

Determining the ground-state energy of the hydrogen molecule is a canonical benchmark problem in molecular quantum mechanics. For  $\text{H}_2$ , the electronic Schrödinger equation can be solved to high accuracy on classical hardware; at an equilibrium bond length of approximately  $0.74 \text{ \AA}$ , the total energy is about  $-1.17$  Hartree (relative to separated nuclei and electrons).

The problem becomes difficult, and eventually intractable, as molecular size increases. For caffeine ( $\text{C}_8\text{H}_{10}\text{N}_4\text{O}_2$ , 102 electrons), an exact solution of the Schrödinger equation requires manipulating a wavefunction in a Hilbert space whose dimension grows combinatorially with the number of orbitals. Classical hardware cannot store or diagonalize matrices of that size except for very small systems.

---

\* [john.azariah@student.uts.edu.au](mailto:john.azariah@student.uts.edu.au)

In 1982, Richard Feynman observed that quantum systems might be efficiently simulated by other quantum systems [1]. This observation launched the field of quantum simulation, and quantum chemistry became its most promising application: if we could encode the electronic Hamiltonian of a molecule into the native operations of a quantum computer, we could in principle determine molecular energies, reaction rates, and material properties that are inaccessible to classical computation.

Today, quantum simulation of chemistry is a reality, at least for small molecules. Experiments have computed the ground-state energy of  $\text{H}_2$  [2],  $\text{LiH}$  [3], and  $\text{BeH}_2$  [4] on quantum hardware, using variational algorithms that require tens to hundreds of measurements of Pauli operators.

But between “the Schrödinger equation” and “measure Pauli operators on qubits,” there is a pipeline of transformations that every quantum chemistry simulation must pass through:

1. **Choose a molecule and a basis set** to reduce the continuous electronic wavefunction to a finite-dimensional problem.
2. **Compute molecular integrals**, namely the matrix elements of the electronic Hamiltonian in the chosen basis.
3. **Write the Hamiltonian in second quantization** to express the physics in terms of creation and annihilation operators.
4. **Convert from spatial to spin-orbital integrals** to account for electron spin, doubling the index space.
5. **Encode fermionic operators as qubit operators** through the fermion-to-qubit transform, which is the central topic of this paper.
6. **Run a quantum algorithm** via VQE, QPE, or related methods.

Each stage involves notation choices, sign conventions, and index manipulations that are often compressed to a few lines in the research literature. A reader with standard quantum mechanics preparation (Schrödinger equation and hydrogen atom) still encounters a substantial barrier: quantum chemistry texts [5, 6] typically assume prior familiarity with second quantization; quantum computing texts [7] rarely treat chemistry in detail; and

research papers on quantum simulation [2, 8] often present the full pipeline in only a few pages of dense notation.

This paper is designed as a single, self-contained reference for that transition. We execute each stage of the pipeline for  $\text{H}_2$ , with explicit matrix elements, index mappings, and sign conventions. Where conventions differ between chemistry and physics, we provide the corresponding translation rules explicitly and identify common failure modes.

By the end, the reader will:

- Understand why fermions and qubits are algebraically different and why encoding is necessary.
- Be able to construct the Jordan–Wigner encoding by hand for any number of modes.
- Have computed the complete 15-term qubit Hamiltonian for  $\text{H}_2$  and verified it by diagonalization.
- Know what alternatives to Jordan–Wigner exist and why they matter for larger molecules.

We assume a background roughly equivalent to a third-year undergraduate in physics or chemistry: linear algebra, introductory quantum mechanics (wavefunctions, the Schrödinger equation, the hydrogen atom), and basic chemistry (orbitals, bonds). No prior knowledge of second quantization, Fock space, Pauli algebra, or quantum computing is assumed.

A companion open-source library in F#, available at <https://github.com/johnazariah/encodings>, reproduces every numerical result in this paper.

## II. THE ELECTRONIC STRUCTURE PROBLEM

### A. The Schrödinger equation for molecules

The full molecular Hamiltonian for a molecule with  $M$  nuclei (charges  $Z_A$ , masses  $M_A$ , positions  $\mathbf{R}_A$ ) and  $N$  electrons (mass  $m_e$ , positions  $\mathbf{r}_i$ ) is:

$$\hat{H} = -\sum_{A=1}^M \frac{\hbar^2}{2M_A} \nabla_A^2 - \sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_i^2 + \sum_{A<B} \frac{Z_A Z_B e^2}{|\mathbf{R}_A - \mathbf{R}_B|} - \sum_{i,A} \frac{Z_A e^2}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i<j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1)$$

For  $\text{H}_2$ , this means two protons ( $A$  and  $B$ , separated by distance  $R$ ) and two electrons (1 and 2). The Hamiltonian is a function of six electronic coordinates (three per electron) plus the internuclear distance  $R$ .

For the hydrogen atom (one electron, one proton), this Schrödinger equation can be solved analytically; the result is the familiar  $1s$ ,  $2s$ ,  $2p$ ,  $\dots$  orbitals. For two electrons, exact analytical solution is already impossible. The electron–electron repulsion term  $e^2/|\mathbf{r}_1 - \mathbf{r}_2|$  couples the two electrons, making the equation non-separable.

## B. The Born–Oppenheimer approximation

Protons are roughly 1836 times heavier than electrons. On the timescale of electronic motion, the nuclei are nearly stationary. The Born–Oppenheimer approximation exploits this mass ratio by treating the nuclear positions  $\{\mathbf{R}_A\}$  as fixed parameters rather than dynamical variables.

The result is the *electronic Hamiltonian*:

$$\hat{H}_{\text{el}} = -\sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_{i,A} \frac{Z_A e^2}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i<j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2)$$

which depends on the nuclear positions only through the electron–nucleus attraction term. The nuclear–nuclear repulsion  $V_{nn} = Z_A Z_B e^2 / R$  is just a constant for fixed  $R$ ; it shifts every energy eigenvalue by the same amount.

For  $\text{H}_2$  at the equilibrium bond length  $R = 0.7414 \text{ \AA}$  ( $= 1.401 \text{ Bohr}$ ):

$$V_{nn} = \frac{e^2}{R} = 0.7151 \text{ Ha} \quad (3)$$

From here on, we fix  $R$  and solve the electronic problem.

## C. Basis sets: turning continuous into discrete

The electronic Hamiltonian  $\hat{H}_{\text{el}}$  acts on wavefunctions of  $3N$  continuous variables. To make the problem finite-dimensional, we expand the molecular orbitals in a finite set of known functions, the *basis set*.

The idea is analogous to Fourier series: approximate a function by keeping finitely many terms. Here, the “terms” are atomic orbitals, and the approximation improves as we add more of them.

*a. Atomic orbitals.* The student already knows the hydrogen atom eigenstates:  $1s$ ,  $2s$ ,  $2p$ , etc. These are characterised by exponential (Slater-type) radial dependence  $e^{-\zeta r}$ , but integrals involving products of exponentials on different centres are analytically intractable. The practical solution is to approximate each Slater-type orbital by a sum of Gaussians  $e^{-\alpha r^2}$ , which have the wonderful property that the product of two Gaussians is another Gaussian.

*b. STO-3G.* The “Slater-Type Orbital, 3 Gaussians” basis set approximates each atomic orbital by 3 Gaussian functions. It is the smallest meaningful basis set, the absolute minimum needed to describe molecular bonding. For hydrogen, STO-3G provides one basis function per atom: the  $1s$  orbital.

*c. Molecular orbitals for  $H_2$ .* With one  $1s$  orbital on each hydrogen atom, the Linear Combination of Atomic Orbitals (LCAO) procedure gives two molecular orbitals:

$$\sigma_g = \frac{1s_A + 1s_B}{\sqrt{2(1+S)}} \quad (\text{bonding}) \quad (4)$$

$$\sigma_u = \frac{1s_A - 1s_B}{\sqrt{2(1-S)}} \quad (\text{antibonding}) \quad (5)$$

where  $S = \langle 1s_A | 1s_B \rangle$  is the overlap integral. The bonding orbital  $\sigma_g$  has lower energy because the electron density is concentrated between the nuclei, while the antibonding orbital  $\sigma_u$  has a node at the midpoint.

With 2 molecular orbitals and 2 spin states ( $\alpha$  = spin-up,  $\beta$  = spin-down), we have  $2 \times 2 = 4$  *spin-orbitals*.

#### D. How many states? The configuration space

Two electrons distributed among 4 spin-orbitals can be arranged in  $\binom{4}{2} = 6$  ways. Using occupation-number notation  $|n_0 n_1 n_2 n_3\rangle$  where  $n_j \in \{0, 1\}$  indicates whether spin-orbital  $j$  is occupied:

The *exact* ground state of  $H_2$  is a superposition of these six configurations. The Hartree–Fock approximation uses only the first ( $|1100\rangle$ , the single-determinant ground state), which captures about 99% of the energy. The remaining 1%, the *correlation energy*, is what makes quantum simulation valuable.

Configuration Notation		Description
$ 1100\rangle$	$\sigma_{g\alpha} \sigma_{g\beta}$	Both in bonding (ground state)
$ 1010\rangle$	$\sigma_{g\alpha} \sigma_{u\alpha}$	One in each, same spin
$ 1001\rangle$	$\sigma_{g\alpha} \sigma_{u\beta}$	One in each, opposite spin
$ 0110\rangle$	$\sigma_{g\beta} \sigma_{u\alpha}$	One in each, opposite spin
$ 0101\rangle$	$\sigma_{g\beta} \sigma_{u\beta}$	One in each, same spin
$ 0011\rangle$	$\sigma_{u\alpha} \sigma_{u\beta}$	Both in antibonding

TABLE I. The six two-electron configurations for  $\text{H}_2$  in the STO-3G basis.

*Key observation:* These occupation vectors  $|n_0 n_1 n_2 n_3\rangle$  look exactly like qubit computational basis states  $|q_0 q_1 q_2 q_3\rangle$ . This is not a coincidence. It is why quantum simulation of chemistry works. But as we will see in Sec. VI, the correspondence is not as simple as setting qubit  $j$  = occupation of orbital  $j$ , because fermions and qubits obey different algebraic rules.

### III. SECOND QUANTIZATION

#### A. Limitations of first-quantized wavefunction methods

Electrons are fermions: the wavefunction must be antisymmetric under exchange of any two electrons. For two electrons,

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = -\Psi(\mathbf{r}_2, \mathbf{r}_1) \quad (6)$$

The standard way to enforce this is the Slater determinant, a determinant of single-particle orbitals:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_a(\mathbf{r}_1) & \phi_b(\mathbf{r}_1) \\ \phi_a(\mathbf{r}_2) & \phi_b(\mathbf{r}_2) \end{vmatrix} \quad (7)$$

For two electrons this determinant contains only two terms. For  $N$  electrons, however, antisymmetrization scales factorially, and the exact many-electron state is generally a superposition of many Slater determinants. This rapid combinatorial growth motivates a formulation in which antisymmetry is handled algebraically rather than by explicit expansion in coordinate space.

Second quantization solves this bookkeeping problem by encoding the antisymmetry into the *operators* rather than the *wavefunction*. The wavefunction becomes a simple binary string (which orbitals are occupied), and all the sign complexity lives in the definition of the operators.

## B. Occupation numbers and Fock space

Instead of tracking which electron is at which position, we track which *orbitals* are occupied. The state of the system is specified by the occupation number vector:

$$|n_0 n_1 n_2 \dots n_{K-1}\rangle \quad n_j \in \{0, 1\} \quad (8)$$

where  $K$  is the number of spin-orbitals and  $n_j = 1$  means orbital  $j$  is occupied. (Fermions can have at most one particle per orbital, by the Pauli exclusion principle.)

The *Fock space* is the vector space spanned by all  $2^K$  such occupation vectors. For  $\text{H}_2$  with  $K = 4$ , Fock space has dimension  $2^4 = 16$ , but only the  $\binom{4}{2} = 6$  states with exactly 2 occupied orbitals are physically relevant.

The *vacuum state*  $|0000\rangle$  has all orbitals empty.

## C. Creation and annihilation operators

The creation operator  $a_j^\dagger$  creates an electron in orbital  $j$ :

$$a_j^\dagger |\dots 0_j \dots\rangle = (-1)^{\sum_{k<j} n_k} |\dots 1_j \dots\rangle \quad (9)$$

$$a_j^\dagger |\dots 1_j \dots\rangle = 0 \quad (\text{Pauli exclusion}) \quad (10)$$

The annihilation operator  $a_j$  destroys an electron in orbital  $j$ :

$$a_j |\dots 1_j \dots\rangle = (-1)^{\sum_{k<j} n_k} |\dots 0_j \dots\rangle \quad (11)$$

$$a_j |\dots 0_j \dots\rangle = 0 \quad (\text{nothing to destroy}) \quad (12)$$

The sign factor  $(-1)^{\sum_{k<j} n_k}$  counts the number of occupied orbitals with index less than  $j$ . This factor is the source of *all* the complexity in fermion-to-qubit encoding.



a. *Worked examples* (4 spin-orbitals):

- $a_0^\dagger |0000\rangle = |1000\rangle$ : no occupied orbitals before index 0, so the sign is  $(-1)^0 = +1$ .
- $a_1^\dagger |1000\rangle = -|1100\rangle$ : one occupied orbital (index 0) before index 1, so the sign is  $(-1)^1 = -1$ .
- $a_0^\dagger |1000\rangle = 0$ : orbital 0 is already occupied.
- $a_1 |1100\rangle = -|1000\rangle$ : remove orbital 1; one occupied orbital before it gives the minus sign.

The *number operator*  $\hat{n}_j = a_j^\dagger a_j$  counts the occupation of orbital  $j$ :

$$\hat{n}_j |\dots 1_j \dots\rangle = |\dots 1_j \dots\rangle \quad (13)$$

$$\hat{n}_j |\dots 0_j \dots\rangle = 0 \quad (14)$$

Its eigenvalue is  $n_j$ , the occupation number itself.

#### D. The canonical anti-commutation relations

The creation and annihilation operators satisfy the *canonical anti-commutation relations* (CAR):

$$\{a_i, a_j^\dagger\} \equiv a_i a_j^\dagger + a_j^\dagger a_i = \delta_{ij} \quad (15)$$

$$\{a_i, a_j\} = 0 \quad (16)$$

$$\{a_i^\dagger, a_j^\dagger\} = 0 \quad (17)$$

where  $\{A, B\} = AB + BA$  is the anti-commutator.

The physical content:

- $\{a_i^\dagger, a_j^\dagger\} = 0$  says you cannot create two electrons in the same orbital ( $i = j$  gives  $2(a_i^\dagger)^2 = 0$ ), and creating in orbitals  $i$  then  $j$  is the *negative* of creating in  $j$  then  $i$  (antisymmetry).
- The cross-anticommutator  $\{a_i, a_j^\dagger\} = \delta_{ij}$  says creating then destroying in the same orbital recovers the original state, but in different orbitals the operations anti-commute.

*The encoding imperative:* These anti-commutation relations are the *definition* of fermionic algebra. Any mapping from fermions to qubits must preserve them exactly. If the qubit operators don't anti-commute in the right way, the encoded Hamiltonian has the wrong eigenvalues and the quantum simulation gives incorrect results.

### E. The Hamiltonian in second quantization

The electronic Hamiltonian can be written entirely in terms of creation and annihilation operators:

$$\hat{H} = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} \langle pq|rs \rangle a_p^\dagger a_q^\dagger a_s a_r \quad (18)$$

The one-body integrals  $h_{pq}$  encode kinetic energy and electron–nucleus attraction:

$$h_{pq} = \int \phi_p^*(\mathbf{r}) \left[ -\frac{\hbar^2}{2m_e} \nabla^2 - \sum_A \frac{Z_A e^2}{|\mathbf{r} - \mathbf{R}_A|} \right] \phi_q(\mathbf{r}) d\mathbf{r} \quad (19)$$

The two-body integrals  $\langle pq|rs \rangle$  encode electron–electron repulsion:

$$\langle pq|rs \rangle = \iint \frac{\phi_p^*(\mathbf{r}_1) \phi_q^*(\mathbf{r}_2) \phi_r(\mathbf{r}_1) \phi_s(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (20)$$

(This is physicist's notation; more on this in Sec. IV.)

*Warning:* The operator ordering in the two-body term is  $a_p^\dagger a_q^\dagger a_s a_r$ ; note that  $a_s$  comes before  $a_r$ . This “reversed” order relative to the integral indices comes from normal ordering (all creation operators to the left of all annihilation operators). Getting this wrong flips signs.

For  $\text{H}_2$  in the STO-3G basis, the non-zero one-body integrals are:

Integral Value (Ha) Physical meaning		
$h_{00}$	−1.2563	$\sigma_g$ orbital energy
$h_{11}$	−0.4719	$\sigma_u$ orbital energy

TABLE II. Non-zero one-body integrals for  $\text{H}_2/\text{STO-3G}$ .

The off-diagonal elements  $h_{01} = h_{10} = 0$  because  $\sigma_g$  and  $\sigma_u$  have different symmetry.

## IV. NOTATION CONVENTIONS AND CONVERSION

Multiple index-ordering conventions for two-electron integrals are in common use. Substituting one convention into formulas written in another can produce numerically plausible but incorrect Hamiltonians. This section makes the required conversions explicit.

### A. Chemist's notation

Chemist's notation (also called Mulliken notation or charge-density notation) groups indices by *spatial coordinate*:

$$[pq|rs] = \iint \phi_p^*(\mathbf{r}_1) \phi_q(\mathbf{r}_1) \frac{1}{r_{12}} \phi_r^*(\mathbf{r}_2) \phi_s(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (21)$$

The bracket  $[pq]$  refers to electron 1 (at  $\mathbf{r}_1$ ), and  $[rs]$  refers to electron 2. Within each bracket, the first index is the complex conjugate (bra) and the second is the ket.

### B. Physicist's notation

Physicist's notation (also called Dirac notation; confusingly, not the same as bra-ket notation for states) groups indices by *particle*:

$$\langle pq|rs \rangle = \iint \phi_p^*(\mathbf{r}_1) \phi_q^*(\mathbf{r}_2) \frac{1}{r_{12}} \phi_r(\mathbf{r}_1) \phi_s(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (22)$$

Here  $p$  and  $r$  belong to electron 1, while  $q$  and  $s$  belong to electron 2. The convention is: bra indices on the left ( $p, q$ ), ket indices on the right ( $r, s$ ).

### C. The conversion

Comparing the two definitions:

$$\boxed{\langle pq|rs \rangle_{\text{physicist}} = [pr|qs]_{\text{chemist}}} \quad (23)$$

The indices get *shuffled*: the physicist's bra-ket pairs  $(p, r)$  and  $(q, s)$  become the chemist's coordinate pairs, but the *positions within each bracket* change.

### D. Which notation for the Hamiltonian?

The second-quantized Hamiltonian (Eq. 18) uses *physicist’s* notation. If you have integrals in chemist’s notation (which most quantum chemistry codes output), you must convert before plugging into this formula.

*Frequent implementation errors:*

1. Using chemist’s integrals  $[pq|rs]$  directly in the physicist’s formula (or vice versa). This permutes the indices and gives wrong coefficients.
2. Forgetting the  $\frac{1}{2}$  prefactor on the two-body term. This double-counts electron–electron interactions.
3. Writing the operator ordering as  $a_p^\dagger a_q^\dagger a_r a_s$  instead of  $a_p^\dagger a_q^\dagger a_s a_r$ . The  $r$  and  $s$  are reversed.

## V. FROM SPATIAL TO SPIN-ORBITAL INTEGRALS

The molecular integrals computed by quantum chemistry codes are in the *spatial orbital* basis (e.g., 2 orbitals for  $\text{H}_2$ ). But the fermionic operators act on *spin-orbitals* (4 for  $\text{H}_2$ ), because each spatial orbital can hold one electron of each spin.

### A. Spin-orbital indexing

Each spatial orbital  $p$  gives rise to two spin-orbitals:  $p\alpha$  (spin up) and  $p\beta$  (spin down).

We use *interleaved* indexing:

Spin-orbital index	Spatial orbital	Spin
0	0 ( $\sigma_g$ )	$\alpha$
1	0 ( $\sigma_g$ )	$\beta$
2	1 ( $\sigma_u$ )	$\alpha$
3	1 ( $\sigma_u$ )	$\beta$

TABLE III. Spin-orbital indexing for  $\text{H}_2/\text{STO-3G}$ .

The conversion rules are: spatial orbital index =  $\lfloor j/2 \rfloor$  (integer division); spin =  $j \bmod 2$  ( $0 = \alpha$ ,  $1 = \beta$ ).

## B. One-body expansion

The spin-orbital one-body integral is:

$$h_{pq}^{\text{spin}} = h_{p/2, q/2}^{\text{spatial}} \times \delta(\sigma_p, \sigma_q) \quad (24)$$

In words: the integral equals the spatial integral if the spins match, and zero otherwise. An electron cannot change its spin through one-body interactions (in the non-relativistic limit).

For  $\text{H}_2$ , this gives 4 non-zero entries, all diagonal:

$p$	$q$	$h_{pq}^{\text{spin}}$ (Ha)	Origin
$0\alpha$	$0\alpha$	$-1.2563$	$h_{00}^{\text{spatial}}$ , same spin
$0\beta$	$0\beta$	$-1.2563$	$h_{00}^{\text{spatial}}$ , same spin
$1\alpha$	$1\alpha$	$-0.4719$	$h_{11}^{\text{spatial}}$ , same spin
$1\beta$	$1\beta$	$-0.4719$	$h_{11}^{\text{spatial}}$ , same spin

TABLE IV. Non-zero spin-orbital one-body integrals for  $\text{H}_2$ .

## C. Two-body expansion

The spin-orbital two-body integral in physicist's notation is:

$$\langle pq|rs\rangle_{\text{spin}} = \left[ \frac{p}{2} \frac{r}{2} \middle| \frac{q}{2} \frac{s}{2} \right]_{\text{spatial}} \times \delta(\sigma_p, \sigma_r) \times \delta(\sigma_q, \sigma_s) \quad (25)$$

Both electrons must independently conserve spin. This generates more non-zero integrals than one might expect, because *cross-spin* terms are allowed.

*Common error:* If you include only same-spin blocks ( $\alpha\alpha$  and  $\beta\beta$ ) and omit the cross-spin blocks ( $\alpha\beta$  and  $\beta\alpha$ ), your Hamiltonian will contain only Z-type (diagonal) Pauli terms and no XX/YY excitation terms. The eigenvalues will be wrong.

For  $\text{H}_2$ , there are 32 non-zero spin-orbital two-body integrals (before symmetry reduction). They are tabulated in full in Appendix A.

## D. The complete spin-orbital Hamiltonian

Combining one-body (4 terms) and two-body (32 terms, with  $\frac{1}{2}$  prefactor), plus the nuclear repulsion constant:

$$\hat{H} = V_{nn} \cdot \hat{I} + \sum_{pq} h_{pq}^{\text{spin}} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} \langle pq|rs \rangle_{\text{spin}} a_p^\dagger a_q^\dagger a_s a_r \quad (26)$$

with  $V_{nn} = 0.7151$  Ha.

## VI. THE ENCODING PROBLEM

### A. Fermions vs. qubits

Both fermionic Fock space and multi-qubit Hilbert space have dimension  $2^n$ , where  $n$  is the number of modes (spin-orbitals) or qubits. The computational basis states even look the same:  $|0110\rangle$  could be an occupation vector or a qubit state.

But the *algebras* are different. Fermionic operators anti-commute across all modes:

$$\{a_i, a_j^\dagger\} = \delta_{ij} \quad \text{for all } i, j \quad (27)$$

Qubit operators (Paulis) anti-commute *on the same qubit* but *commute on different qubits*:

$$\{X_i, Y_i\} = 0 \quad (\text{same qubit: anti-commute}) \quad (28)$$

$$[X_i, Z_j] = 0 \quad (\text{different qubits: commute}) \quad (29)$$

This algebraic mismatch defines the encoding problem.

### B. Naive mapping and its failure

The qubit raising and lowering operators  $\sigma_j^\pm = (X_j \mp iY_j)/2$  satisfy  $\{\sigma_j^-, \sigma_j^+\} = I$  on qubit  $j$ , exactly like  $\{a_j, a_j^\dagger\} = 1$ . So the tempting mapping is:

$$a_j^\dagger \stackrel{?}{\mapsto} \sigma_j^+ = \frac{X_j - iY_j}{2} \quad (30)$$

Consider the cross-mode anticommutator:

$$\{a_0, a_1^\dagger\} = 0 \quad (\text{fermions: must vanish}) \quad (31)$$

Since  $\sigma_0^-$  and  $\sigma_1^+$  act on *different* qubits, they commute rather than anti-commute. Consequently,  $\{\sigma_0^-, \sigma_1^+\} = 2\sigma_0^- \sigma_1^+ \neq 0$ , so this mapping does not preserve the fermionic algebra.

### C. The Jordan–Wigner transform

Jordan and Wigner (1928) [9] found the fix: insert a chain of  $Z$  operators on all lower-index qubits. Since  $Z$  has eigenvalues  $\pm 1$  depending on the qubit’s state, this chain tracks the *parity* of all preceding occupations, exactly the sign factor  $(-1)^{\sum_{k<j} n_k}$  that appears in the creation operator’s definition.

The Majorana decomposition makes this cleanest. Define:

$$c_j = a_j^\dagger + a_j \quad d_j = i(a_j^\dagger - a_j) \quad (32)$$

These are Hermitian operators satisfying  $\{c_j, c_k\} = 2\delta_{jk}$  and  $\{d_j, d_k\} = 2\delta_{jk}$ .

The Jordan–Wigner encoding maps them to:

$$c_j \mapsto X_j \otimes Z_{j-1} \otimes \cdots \otimes Z_0 \quad (33)$$

$$d_j \mapsto Y_j \otimes Z_{j-1} \otimes \cdots \otimes Z_0 \quad (34)$$

Or in Pauli string notation (reading left to right = qubit 0, 1, ...):

Mode $j$	$c_j$	$d_j$
0	$XIII$	$YIII$
1	$ZXII$	$ZYII$
2	$ZZXI$	$ZZYI$
3	$ZZZX$	$ZZZY$

TABLE V. Majorana operators under Jordan–Wigner for  $n = 4$ .

The ladder operators follow from  $a_j^\dagger = \frac{1}{2}(c_j - id_j)$  and  $a_j = \frac{1}{2}(c_j + id_j)$ .

*a. Algebraic mechanism.* Consider  $\{c_0, c_1\}$ :

$$c_0 c_1 = (XIII)(ZXII) = -YXII \quad (35)$$

$$c_1 c_0 = (ZXII)(XIII) = +YXII \quad (36)$$

The anti-commutator  $c_0 c_1 + c_1 c_0 = 0$ . The key is that  $X_0$  and  $Z_0$  anti-commute (they are different non-identity Paulis on the same qubit), generating the crucial minus sign.

*b. Asymptotic cost.* The  $Z$ -chain grows with  $j$ . Operator  $c_{n-1}$  acts on *all*  $n$  qubits; its Pauli weight is  $n$ . This  $O(n)$  scaling increases circuit depth for larger active spaces, motivating alternative encodings with logarithmic Pauli-weight growth.

#### D. Beyond Jordan–Wigner

The  $O(n)$  Pauli weight of Jordan–Wigner comes from its linear chain structure. Can we do better?

*a. Bravyi–Kitaev (2002).* Replaces the linear chain with a *Fenwick tree* (binary indexed tree) [10, 11]. Each qubit stores the parity of a logarithmically bounded subset of modes, resulting in  $O(\log_2 n)$  Pauli weight.

*b. Parity encoding.* The “dual” of Jordan–Wigner: each qubit stores the cumulative parity  $n_0 \oplus n_1 \oplus \dots \oplus n_j$  instead of the individual occupation  $n_j$ .

*c. Tree encodings.* Every labelled rooted tree defines a valid fermion-to-qubit encoding [12]. Jordan–Wigner corresponds to a linear chain. Bravyi–Kitaev corresponds to a Fenwick tree. A balanced ternary tree achieves the *provably optimal* worst-case Pauli weight of  $O(\log_3 n)$ .

The following table summarises the maximum single-operator Pauli weight for each encoding at various system sizes:

$n$	JW	BK	Parity	Balanced Binary	Balanced Ternary
4	4	3	4	3	<b>2</b>
8	8	4	8	4	<b>3</b>
16	16	5	16	5	<b>4</b>
24	24	5	24	5	<b>5</b>

TABLE VI. Maximum single-operator Pauli weight for different encodings.



## VII. BUILDING THE H<sub>2</sub> QUBIT HAMILTONIAN

### A. The recipe

1. For each non-zero one-body integral  $h_{pq}$ : encode  $a_p^\dagger$  and  $a_q$  as Pauli strings, multiply them, multiply by  $h_{pq}$ .
2. For each non-zero two-body integral  $\langle pq|rs\rangle$ : encode all four ladder operators, multiply the four Pauli strings, multiply by  $\frac{1}{2}\langle pq|rs\rangle$ .
3. Sum all terms, collecting Pauli strings with the same signature and adding their coefficients.
4. Add  $V_{nn} \cdot IIII$  (nuclear repulsion as a constant offset).

### B. One-body terms

The non-zero one-body integrals for H<sub>2</sub> are all diagonal:  $h_{00}, h_{11}, h_{22}, h_{33}$  (in the spin-orbital basis). These are number operators  $\hat{n}_j = a_j^\dagger a_j$ .

Under Jordan–Wigner:

$$\hat{n}_j = a_j^\dagger a_j = \frac{1}{2}(I - Z_j) \quad (37)$$

(The  $Z$ -chains cancel because both  $c_j$  and  $d_j$  have the same chain.)

So the one-body contribution is:

$$\begin{aligned} \hat{H}_1 = \frac{1}{2}(h_{00} + h_{11} + h_{22} + h_{33}) \cdot IIII \\ - \frac{h_{00}}{2} IIII - \frac{h_{11}}{2} IIZI - \frac{h_{22}}{2} IZII - \frac{h_{33}}{2} ZIII \end{aligned} \quad (38)$$

Substituting  $h_{00} = h_{11} = -1.2563$  and  $h_{22} = h_{33} = -0.4719$ :

### C. Two-body terms

After processing all 32 non-zero two-body integrals and combining like terms, the complete electronic Hamiltonian under Jordan–Wigner encoding has **15 Pauli terms**:

The  $Z$ -only terms (rows 2–11) represent classical electrostatic interactions: Coulomb repulsion and orbital energies. The  $XXYY$ -type terms (rows 12–15) represent *quantum*

Term Coefficient (Ha)	
<i>IIII</i>	-1.7282
<i>IIIZ</i>	+0.6282
<i>IIZI</i>	+0.6282
<i>IZII</i>	+0.2359
<i>ZIII</i>	+0.2359

TABLE VII. One-body Pauli terms for H<sub>2</sub>/JW.

*exchange*, a fundamentally non-classical effect arising from the indistinguishability of electrons.

#### D. Cross-encoding comparison

The same Hamiltonian encoded under all five transforms produces:

All encodings produce the same number of terms with the same identity coefficient, as expected, since the identity coefficient equals  $\text{Tr}(\hat{H})/2^n$ , which is invariant under any unitary change of basis.

### VIII. CHECKING OUR ANSWER

#### A. Exact diagonalization

To verify the qubit Hamiltonian, we construct its  $16 \times 16$  matrix representation. Each Pauli string  $\sigma_\alpha$  corresponds to a known matrix (the tensor product of its single-qubit Pauli matrices). The full Hamiltonian matrix is:

$$H = \sum_{\alpha} c_{\alpha} \cdot \sigma_{\alpha} \quad (39)$$

where  $c_{\alpha}$  are the 15 coefficients from Table VIII.

Diagonalising this matrix gives 16 eigenvalues. These can be grouped by the particle-number sector:

The ground state of the  $N_e = 2$  sector is  $E_0^{\text{el}} = -1.8573$  Ha. Adding nuclear repulsion:

$$E_0^{\text{total}} = E_0^{\text{el}} + V_{nn} = -1.8573 + 0.7151 = -1.1422 \text{ Ha} \quad (40)$$

<hr/> # Pauli String Coefficient (Ha) <hr/>	
1 <i>IIII</i>	−1.0704
2 <i>IIIZ</i>	−0.0958
3 <i>IIZI</i>	−0.0958
4 <i>IZII</i>	+0.3021
5 <i>ZIII</i>	+0.3021
6 <i>IIZZ</i>	+0.1743
7 <i>IZIZ</i>	−0.0085
8 <i>IZZI</i>	+0.1659
9 <i>ZIIZ</i>	+0.1659
10 <i>ZIZI</i>	−0.0085
11 <i>ZZII</i>	+0.1686
12 <i>XXYY</i>	−0.1744
13 <i>XYXX</i>	+0.1744
14 <i>YXXY</i>	+0.1744
15 <i>YYXX</i>	−0.1744

---

TABLE VIII. Complete 15-term qubit Hamiltonian for H<sub>2</sub>/STO-3G under Jordan–Wigner encoding.

## B. Comparison with known results

The Hartree–Fock energy (single determinant  $|1100\rangle$ ) is:

$$E_{\text{HF}} = 2h_{00} + [00|00] = 2(-1.2563) + 0.6745 = -1.8382 \text{ Ha (electronic)} \quad (41)$$

The Full CI correlation energy is:

$$E_{\text{corr}} = E_{\text{FCI}} - E_{\text{HF}} = -1.8573 - (-1.8382) = -0.0191 \text{ Ha} \approx -12.0 \text{ kcal/mol} \quad (42)$$

This correlation energy, about 1% of the total energy but  $\sim 12$  kcal/mol, is precisely what makes quantum simulation valuable. It captures the effect of electron–electron correlation that the single-determinant Hartree–Fock approximation misses.

All five encodings produce the same eigenspectrum to machine precision ( $|\Delta\lambda| < 5 \times$

Encoding	Terms	Max Weight	Avg Weight	Identity (Ha)
Jordan–Wigner	15	4	2.13	−1.0704
Bravyi–Kitaev	15	4	2.40	−1.0704
Parity	15	4	2.27	−1.0704
Balanced Binary	15	4	2.27	−1.0704
Balanced Ternary	15	4	2.40	−1.0704

TABLE IX. Cross-encoding comparison for  $\text{H}_2/\text{STO-3G}$ . All encodings produce the same eigen-spectrum.

Sector ( $N_e$ )	Dim.	Eigenvalues (Ha, electronic)
0	1	0
1	4	−1.2563, −1.2563, −0.4719, −0.4719
2	6	−1.8573, −1.3390, −0.9032, −0.9032, −0.6753, 0.0
3	4	−1.7282, −1.7282, −0.9438, −0.9438
4	1	−2.2001

TABLE X. Eigenspectrum by particle-number sector.

$10^{-16}$ ), confirming that the encoding is a unitary change of basis that preserves the physics exactly.

## IX. ALGORITHMIC IMPLICATIONS

The qubit Hamiltonian from Sec. VII is the input to quantum algorithms. Two families of algorithms can extract the ground-state energy:

*a. Variational Quantum Eigensolver (VQE).* Prepares a parameterised quantum state  $|\psi(\boldsymbol{\theta})\rangle$ , measures  $\langle\psi|\hat{H}|\psi\rangle$  by separately measuring each Pauli term, and uses a classical optimizer to minimize the energy over  $\boldsymbol{\theta}$  [2, 13]. VQE is designed for near-term noisy quantum hardware.

*b. Quantum Phase Estimation (QPE).* Applies the time-evolution operator  $e^{-i\hat{H}t}$  controlled on an ancilla register to extract eigenvalues directly [7]. QPE requires fault-tolerant hardware but provides exponential speedup for large systems.

The encoding choice directly affects algorithmic resource scaling:

- Each Pauli term must be measured separately, so more terms = more shots.
- Higher Pauli weight = deeper CNOT ladders = more gate errors.
- The ternary tree encoding’s  $O(\log_3 n)$  weight scaling means that for 100 modes, the deepest circuits are roughly 5 CNOTs instead of JW’s 100, a difference that may determine whether the simulation is feasible on early fault-tolerant hardware [14].

## X. CONCLUSION

We have traced the complete pipeline from the molecular Schrödinger equation to a qubit Hamiltonian, using  $\text{H}_2$  as a worked example with every step made explicit:

1. The Born–Oppenheimer approximation reduces the problem to the electronic Hamiltonian.
2. The STO-3G basis set turns it into a finite-dimensional matrix problem (2 spatial orbitals  $\rightarrow$  4 spin-orbitals  $\rightarrow$  6 configurations for 2 electrons).
3. Second quantization encodes antisymmetry into operators, giving a compact representation as creation and annihilation operators.
4. The spatial-to-spin-orbital expansion doubles the index space and introduces spin conservation constraints.
5. The Jordan–Wigner (or other) encoding maps fermionic operators to Pauli strings, producing a qubit Hamiltonian that a quantum computer can measure.
6. Exact diagonalization of the resulting 15-term Hamiltonian recovers the known ground-state energy, confirming the encoding’s correctness.

In addition, we have documented key notation transitions (chemist’s vs. physicist’s integrals and operator ordering), highlighted recurrent implementation errors (including omitted cross-spin contributions and incorrect index conversion), and provided a companion codebase that reproduces all numerical results.

For those interested in *why* each encoding has the structure it does, why the tree shape determines everything, and what this reveals about the relationship between fermionic and qubit descriptions, we refer to our companion paper.

## ACKNOWLEDGMENTS

This work is dedicated to Dr. Guang Hao Low, whose encouragement to explore Bravyi–Kitaev encodings seven years ago led to the development of this symbolic algebra processing library.

## Appendix A: H<sub>2</sub> STO-3G Integral Tables

### 1. Molecular parameters

Parameter	Value
Bond length $R$	0.7414 Å = 1.401 Bohr
Nuclear repulsion $V_{nn}$	0.7151043391 Ha
Spatial orbitals	2 ( $\sigma_g, \sigma_u$ )
Spin-orbitals	4
Electrons	2

### 2. Spatial one-body integrals $h_{pq}$ (Ha)

### 3. Spatial two-body integrals $[pq|rs]$ (Ha)

All other elements are zero by symmetry.

		$q = 0 \ (\sigma_g)$	$q = 1 \ (\sigma_u)$
$p = 0$	$-1.2563390730$	$0$	
$p = 1$	$0$	$-0.4718960244$	

  

Integral	Value
$[00 00]$	$0.6744887663$
$[11 11]$	$0.6973979495$
$[00 11] = [11 00]$	$0.6636340479$
$[01 10] = [10 01] = [01 01] = [10 10]$	$0.6975782469$

## Appendix B: Pauli Algebra Reference

### 1. Single-qubit Pauli matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (\text{B1})$$

### 2. Multiplication table

$$X \cdot Y = iZ \qquad Y \cdot Z = iX \qquad Z \cdot X = iY \quad (\text{B2})$$

$$Y \cdot X = -iZ \qquad Z \cdot Y = -iX \qquad X \cdot Z = -iY \quad (\text{B3})$$

Two Paulis on the same qubit either commute ( $[A, B] = 0$  when  $A = B$  or either is  $I$ ) or anti-commute ( $\{A, B\} = 0$  when  $A \neq B$  and neither is  $I$ ).

### 3. Multi-qubit Pauli strings

A Pauli string on  $n$  qubits is a tensor product:  $\sigma = P_0 \otimes P_1 \otimes \cdots \otimes P_{n-1}$  where each  $P_j \in \{I, X, Y, Z\}$ .

The product of two Pauli strings is another Pauli string (times a phase  $\pm 1$  or  $\pm i$ ). The Pauli weight of a string is the number of non-identity entries:  $w(\sigma) = |\{j : P_j \neq I\}|$ .

---

- [1] R. P. Feynman, Simulating physics with computers, [International Journal of Theoretical Physics](#) **21**, 467 (1982).
- [2] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, *et al.*, Scalable quantum simulation of molecular energies on a superconducting qubit processor, [Physical Review X](#) **6**, 031007 (2016).
- [3] A. Kandala, A. Mezzacapo, K. Temme, *et al.*, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, [Nature](#) **549**, 242 (2017).
- [4] A. Kandala, K. Temme, A. D. Córcoles, *et al.*, Error mitigation extends the computational reach of a noisy quantum processor, [Nature](#) **567**, 491 (2019).
- [5] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, 1996).
- [6] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley & Sons, 2000).
- [7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, 2010).
- [8] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Simulation of electronic structure Hamiltonians using quantum computers, [Molecular Physics](#) **109**, 735 (2011).
- [9] P. Jordan and E. Wigner, Über das Paulische äquivalenzverbot, [Zeitschrift für Physik](#) **47**, 631 (1928).
- [10] S. B. Bravyi and A. Y. Kitaev, Fermionic quantum computation, [Annals of Physics](#) **298**, 210 (2002).
- [11] J. T. Seeley, M. J. Richard, and P. J. Love, The Bravyi-Kitaev transformation for quantum computation of electronic structure, [The Journal of Chemical Physics](#) **137**, 224109 (2012).
- [12] Z. Jiang, A. Kaley, W. Mruczkiewicz, and H. Neven, Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states of chemistry, [PRX Quantum](#) **1**, 010306 (2020).
- [13] A. Peruzzo, J. McClean, P. Shadbolt, *et al.*, A variational eigenvalue solver on a photonic quantum processor, [Nature Communications](#) **5**, 4213 (2014).



- [14] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, [Proceedings of the National Academy of Sciences](#) **114**, 7555 (2017).