

Particle-Number-Conserving Quantum Autoencoders for Certifiable Molecular Compression

Jian Feng Kong, Zi Cheng Wong, Jiabin You, Adrian Matthew Mak

Institute of Advanced Intelligence and Computing (IAIC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore

Abstract

Quantum autoencoders (QAEs) compress families of quantum states onto a few latent qubits, but their latent space is a black box: one can neither certify what it contains nor read it in physical terms. We show that one physically motivated constraint, restricting the encoder to conserve particle number, resolves both problems at once. The latent becomes operationally certifiable (measuring the trash register and accepting the all-zeros outcome projects it exactly into the correct N -electron sector) and interpretable (the trained one-body encoder reproduces the state-averaged natural-orbital active space, an identification proved from the optimality condition and verified numerically, so the latent is a learned, state-averaged active space in the literal SA-CASSCF sense). We note that this interpretability comes at a cost: the constrained encoder compresses to at least as many qubits as a hardware-efficient ansatz (HEA). We map where the latent’s physical meaning holds and report actual numbers, including the negative results, on an exactly-solvable Hubbard control and on $\text{H}_2/\text{H}_4/\text{H}_6$ chains.

Background

Quantum machine-learning models are routinely criticised for the opacity of their learned representations. A QAE [1] trains a parameterised encoder $U(\theta)$ to compress a family of states, such as ground states along a potential energy surface (PES) or low-lying excited states, onto $k < n$ latent qubits, driving the remaining $n_T = n - k$ trash qubits to a fixed reference $|0\rangle^{\otimes n_T}$; the decoder is $U^\dagger(\theta)$. QAE compression of molecular PES families is by now established: UCCSD ground states have been compressed across geometries to chemical accuracy [2] and a QAE-plus-generator pipeline compresses and regenerates H_2/LiH states [3]. Every such encoder is hardware-efficient, so its latent space carries no physical meaning; symmetry-informed QAE ansätze are explicitly named an open problem [2]. Particle-number conservation has separately served as a compression resource through a fixed linear encoding [4], and equivariant quantum machine learning supplies number-conserving gate sets and pooling layers [5, 6], but no autoencoder has been built from them. Molecular state families are structured by symmetry and by a small set of chemically active orbitals. A latent space that does not respect this structure is difficult to certify and difficult to interpret physically. In this work we elucidate the advantages of imposing particle-number conservation.

Methods

We first benchmark on a Fermi–Hubbard model ($L = 4$, half filling, the reaction coordinate U/t) and on $\text{H}_2/\text{H}_4/\text{H}_6$ chains under symmetric stretch plus an H_4 automerization rectangle, all in STO-3G. The reference states are obtained from exact sector-restricted diagonalisation. For the H chains we express every geometry in a symmetrically orthogonalised (Löwdin) common frame, so the qubit-to-(atom, spin) labelling is geometry-independent and the family stays low-rank.

The number-conserving encoders are built from native particle-conserving fSim/Givens gates [7] with $[U, \hat{N}] = 0$ for all parameters (building in number-conservation by structure): an orbital-rotation Clements mesh, which maps Slater determinants to Slater determinants and is interpretable as a one-body active space, and a non-Gaussian symmetric brickwork that can represent correlations beyond orbital rotations. An HEA baseline (R_y, R_z rotations plus a CZ brickwork) deliberately breaks $\hat{N}$. Encoders are trained by minimising the local trash cost [8] with exact adjoint gradients. The latent size follows the sector-aware rule $\binom{k}{N} \geq d_{\text{eff}}$, with d_{eff} the effective Gram dimension of the family. For multi-state compression one encoder is trained jointly on the lowest few sector eigenstates at every geometry, tracked across crossings by maximum overlap and labelled by $\langle S^2 \rangle$. We report reconstruction fidelity F , sector validity P_N , acceptance probability p_{acc} , and the non-parallelity error (NPE) of the reconstructed PES.

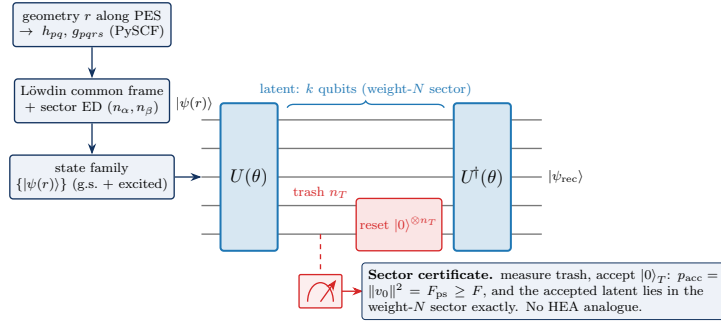


Figure 1: Workflow. A common-frame state family along the PES is compressed by a number-conserving encoder $U(\theta)$ onto k latent qubits (the weight- N sector). Measuring the trash register and accepting $|0\rangle_T$ gives the sector certificate: the acceptance probability equals the post-selected fidelity, upper-bounds the unconditional fidelity, and the accepted latent lies exactly in the weight- N sector.

Results

Parameter efficiency. The number-conserving brickwork gives a better fidelity–parameter tradeoff than the HEA. On the Hubbard control it approaches the best-rank- d compression ceiling [9, 10] with fewer parameters (median fidelity 0.982 at 168 parameters, versus 0.926 at 208 for the HEA), while the one-body orbital rotation saturates far below (0.64) because the family is not well described by a single Slater-determinant manifold. The same ordering holds for every molecular system, with the gap widening at H_6 (0.916 vs 0.833). Since the achievable compression can be limited by the ansatz parameter manifold as well as by the sector bound [11], we use the measured d_{eff} to choose the latent size.

Trainability. Hamming-weight preservation restricts the relevant Hilbert space from dimension 2^n to the fixed- N sector, for which barren-plateau scaling is correspondingly softened [12, 13]. Our gradient-variance scan supports this picture: the symmetric encoder has larger first-parameter gradient variance than the HEA at every size tested, by $\sim 48\times$ on H_2 and $\sim 238\times$ on H_6 . We therefore find evidence for improved trainability, while noting that per-component averages at unmatched depth are less uniform, with one H_6 case reversing the ordering.

A free sector certificate. Measuring the trash register and accepting $|0\rangle_T$ projects the latent exactly into the weight- N sector. We prove $p_{\text{acc}} = \|v_0\|^2 = F_{\text{ps}} \geq F$: the certificate is free to report and post-selection never hurts. The HEA admits no analogue; its accepted latent retains as little as 4–18% in-sector weight. Median acceptance is 0.97–0.99 on H_4 and 0.93–0.96 on H_6 , with the accepted latent in-sector to machine precision at every geometry. The certificate is also a realistic measurement: 10^3 trash shots give $\hat{p} = 0.997 \pm 0.002$ (H_4); under per-gate depolarizing and amplitude-damping noise, acceptance and accepted-sector weight decay approximately linearly and post-selection never hurts at any strength tested, while the HEA’s accepted in-sector weight never exceeds the maximally-mixed floor. Since the post-selected states are pure, PES-quality metrics are tractable at 12 qubits, where the density-matrix reconstruction is not: NPE 30.4 vs 139 mHa (symmetric vs HEA) on H_4 , 97 vs 109 mHa on H_6 , and 6.2 vs 320.1 mHa on the automerization family (Table 1).

The latent is a learned active space. We prove that a one-body encoder at its cost minimum discards exactly the lowest-occupation SA-NO subspace, and verify attainment to $\leq 2 \times 10^{-4}$ on every family: a gradient-trained latent that provably is the active space, connecting QAEs to automated active-space selection [14–16]. Degenerate α/β occupation pairs make individual SA-NOs non-unique, so the invariant test is membership in the degenerate occupation block rather than the raw principal-angle cosine to one chosen basis. This block-membership score is 1.000 for every family; the low H_2 raw cosine (0.366) is therefore a degeneracy artifact, not a failure of active-space recovery. Where a shared active space exists the raw cosine itself is high (H_6 , 0.985; H_4 , 0.982). The correlated brickwork then compresses families beyond any one-body active space (the fixed SA-NO active space reaches fidelity 0.753 on H_4 , the brickwork 0.974), so the latent can mean more than a choice of orbitals.

Multi-state compression (a state-averaged active space). One encoder trained jointly on the lowest three (H_4) or two (H_6) sector eigenstates reconstructs the ground and excited PES from post-selected pure states at median per-state fidelity 0.938 (H_4) and 0.912 (H_6), with median excitation-gap errors of 4.2 and 10.1 mHa. At the state-averaged level the two encoders are essentially tied on H_4 (0.919 vs 0.919), but the HEA’s worst case collapses (per-state fidelity minimum 0.303 vs 0.747; gap-error maximum 324 vs 119 mHa); on H_6 the symmetric encoder’s median gap error is also lower (10.1 vs 36.5 mHa). With

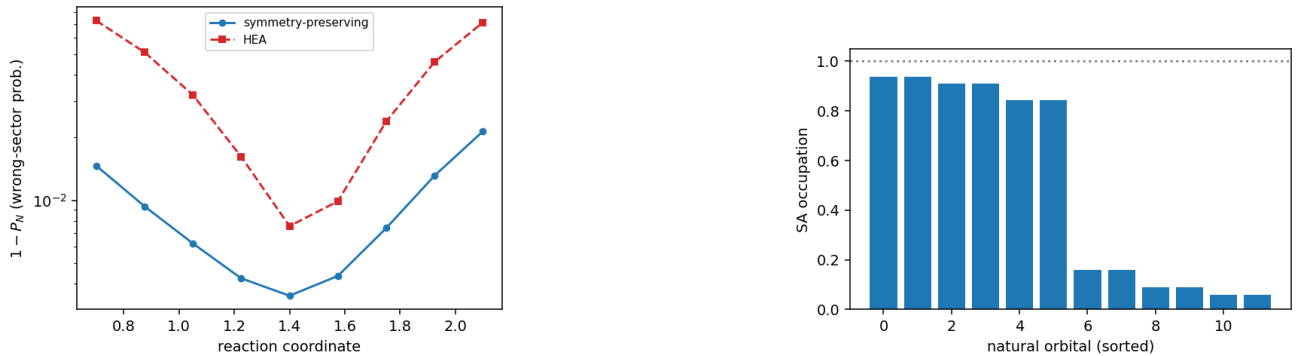


Figure 2: Left (H_4): wrong-sector probability $1 - P_N$ of the unconditional reconstruction, symmetry-preserving vs HEA. Right (H_6): the learned discarded subspace aligns with the lowest-occupation natural orbitals.

system	enc	k	params	fid	p_{acc}	NPE (mHa)
Hubbard-4	sym	5	168	0.982	—	—
	HEA	5	208	0.926	—	—
H_4	sym	5	168	0.974	0.993	30.4
	HEA	5	208	0.915	0.971	138.5
H_6	sym	7	440	0.916	0.962	97.1
	HEA	7	312	0.833	0.927	108.9
H_4 -auto	sym	5	168	0.989	0.996	6.2
	HEA	5	208	0.666	0.879	320.1

Table 1: Table showing the different metrics computed for the different systems. Values are obtained as the medians over seeds.

genuine state-averaging the discarded subspace aligns with the lowest-occupation SA-NOs at principal-angle cosine equal to 1.0 on both systems, justifying the SA-CASSCF interpretation.

Diagnostic validity. On monotone stretch paths, reconstruction infidelity tracks the natural-occupation entropy [17] (Spearman 1.0), consistent with compression error increasing with multireference character. The H_4 automerization path defines the limit of this interpretation: at the square geometry the bonding pattern changes, no single shared active space spans the full family, and the common-encoder infidelity reflects this change of active space as well as local correlation, reducing the rank correlation to 0.52. Even in this case, the correlated symmetric encoder compresses the full two-bonding-pattern family at median fidelity 0.989, while the matched-depth HEA reaches 0.666 (best over tested depths; 0.535 for a shallower variant), the largest separation observed.

Generalization and orbital frame. The learned compression generalizes across the PES rather than memorizing the training grid: on H_4 , alternate-point training gives 0.978 train fidelity and 0.983 test fidelity, while endpoint extrapolation gives $0.987 \rightarrow 0.960$ for the symmetric encoder compared with $0.946 \rightarrow 0.863$ for the HEA. The common orbital frame is essential at H_6 : using per-geometry canonical orbitals [1] lowers the median fidelity from 0.917 to 0.721 and doubles the NPE where orbitals cross.

Conclusion and outlook

We have shown that particle-number conservation turns a QAE latent register into a sector-certified and physically interpretable representation of molecular state families. The trash measurement gives an operational certificate, the one-body limit recovers the state-averaged natural-orbital active space, and the correlated number-conserving ansatz compresses beyond this one-body limit while preserving the sector structure. The tradeoff is latent dimension: the sector constraint $\binom{k}{N} \geq d$ is never more compact than the unconstrained HEA scaling $\log_2 d$, with the qubit-counting argument supported by [4, 7, 18]. The systems studied here are intentionally classically simulable ($n \leq 12$, exact diagonalisation) to validate the structural claims. In state preparation, symmetry-preserving ansätze reduce the resource cost of excited-state VQE [19]; the multi-state gap errors above suggest the compression analogue, and larger systems and excitation-gap calculations are a natural next step. These results identify symmetry preservation as a route to QAEs whose latent spaces are not only compressed, but also constrained, certifiable, and chemically interpretable.

References

- [1] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, 2017. doi:10.1088/2058-9565/aa8072.
- [2] Koen Mesman, Yinglu Tang, Matthias Möller, Boyang Chen, and Sebastian Feld. Nn-ae-vqe: Neural network parameter prediction on autoencoded variational quantum eigensolvers. *arXiv:2411.15667*, 2024.
- [3] Naipunnaya Raj, Rajiv Sangle, Avinash Singh, and Krishna Kumar Sabapathy. Quantum generative adversarial autoencoders: Learning latent representations for quantum data generation. *arXiv:2509.16186*, 2025.
- [4] Yu-Cheng Chen, Ronin Wu, M. H. Cheng, and Min-Hsiu Hsieh. Neural network assisted fermionic compression encoding: A lossy quantum-selected configuration interaction framework for resource-efficient ground-state simulations. *Physical Review Research*, 8:013128, 2026. doi:10.1103/9x8b-4j41.
- [5] Johannes Jakob Meyer, Marian Mularski, Elies Gil-Fuster, Antonio Anna Mele, Francesco Arzani, Alissa Wilms, and Jens Eisert. Exploiting symmetry in variational quantum machine learning. *PRX Quantum*, 4(1):010328, 2023. doi:10.1103/PRXQuantum.4.010328.
- [6] Quynh T Nguyen, Louis Schatzki, Paolo Braccia, Michael Ragone, Patrick J Coles, Frédéric Sauvage, Martín Larocca, and M Cerezo. Theory for equivariant quantum neural networks. *PRX Quantum*, 5(2):020328, 2024. doi:10.1103/PRXQuantum.5.020328.
- [7] Bryan T Gard, Linghua Zhu, George S Barron, Nicholas J Mayhall, Sophia E Economou, and Edwin Barnes. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. *npj Quantum Information*, 6(1):10, 2020. doi:10.1038/s41534-019-0240-1.
- [8] M Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1791, 2021. doi:10.1038/s41467-021-21728-w.
- [9] Chang-Jiang Huang, Hailan Ma, Qi Yin, Jun-Feng Tang, Daoyi Dong, Chunlin Chen, Guo-Yong Xiang, Chuan-Feng Li, and Guang-Can Guo. Realization of a quantum autoencoder for lossless compression of quantum data. *Physical Review A*, 102(3):032412, 2020. doi:10.1103/PhysRevA.102.032412.
- [10] Hailan Ma, Gary J. Mooney, Ian R. Petersen, Lloyd C. L. Hollenberg, and Daoyi Dong. Quantum autoencoders using mixed reference states. *npj Quantum Information*, 10(1):86, 2024. doi:10.1038/s41534-024-00872-3.
- [11] Felix Frohnert, Emiel Koridon, and Stefano Polla. Learning minimal representations of fermionic ground states. *arXiv:2512.11767*, 2025.
- [12] Michael Ragone, Bojko N. Bakalov, Frédéric Sauvage, Alexander F. Kemper, Carlos Ortiz Marrero, Martin Larocca, and M. Cerezo. A lie algebraic theory of barren plateaus for deep parameterized quantum circuits. *Nature Communications*, 15:7172, 2024. doi:10.1038/s41467-024-49909-3.
- [13] Léo Monbroussou, Eliott Z. Mamon, Jonas Landman, Alex B. Grilo, Romain Kukla, and Elham Kashefi. Trainability and expressivity of hamming-weight preserving quantum circuits for machine learning. *Quantum*, 9:1745, 2025. doi:10.22331/q-2025-05-15-1745.
- [14] Christopher J Stein and Markus Reiher. Automated selection of active orbital spaces. *Journal of Chemical Theory and Computation*, 12(4):1760, 2016. doi:10.1021/acs.jctc.6b00156.
- [15] Lexin Ding, Stefan Knecht, and Christian Schilling. Quantum information-assisted complete active space optimization (QICAS). *Journal of Physical Chemistry Letters*, 14(49):11022, 2023. doi:10.1021/acs.jpclett.3c02536.
- [16] Moritz Bensberg and Markus Reiher. Corresponding active orbital spaces along chemical reaction paths. *Journal of Physical Chemistry Letters*, 14(8):2112, 2023. doi:10.1021/acs.jpclett.2c03905.
- [17] Katharina Boguslawski, Pawel Tecmer, Ors Legeza, and Markus Reiher. Entanglement measures for single- and multireference correlation effects. *Journal of Physical Chemistry Letters*, 3(21):3129, 2012. doi:10.1021/jz301319v.
- [18] Yu Shee, Pei-Kai Tsai, Cheng-Lin Hong, Hao-Chung Cheng, and Hsi-Sheng Goan. Qubit-efficient encoding scheme for quantum simulations of electronic structure. *Physical Review Research*, 4(2):023154, 2022. doi:10.1103/PhysRevResearch.4.023154.
- [19] Chufan Lyu, Xusheng Xu, Man-Hong Yung, and Abolfazl Bayat. Symmetry enhanced variational quantum spin eigensolver. *Quantum*, 7:899, 2023. doi:10.22331/q-2023-01-19-899.