Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

QBM: A Quantum Bayesian Machine for Relaxed Energy Prediction on Metal-Oxide Catalysts

Chien-Chai Chang

frank20040506@gapp.nthu.edu.tw

National Tsing Hua University https://orcid.org/0009-0001-5314-7976

Research Article

Keywords: Quantum Computing, Machine Learning Interatomic Potential

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11557640/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

QBM: A Quantum Bayesian Machine for Relaxed Energy Prediction on Metal-Oxide Electrocatalysts

Chien-Chai Chang [1]

Abstract

Whether quantum circuits add useful search or representation to catalyst-energy surrogates remains unclear under matched classical comparisons. We evaluate a Quantum Bayesian Machine (QBM) on the OC22 relaxed-total-energy task, holding structural descriptors and data partitions fixed while varying subset proposal, fidelity-kernel regression and variational residual learning. In the recorded CPU-simulated workflow, the circuit-assisted selector fitted 333 candidates versus 65,535 under enumeration, with mixed validation-error changes; original candidate lists and timing logs are unavailable. A selected fidelity-kernel setting lowered in-distribution mean absolute error by a median 0.00433 eV per atom against paired linear ridge across six fits, but showed no median out-of-distribution gain. Re-uploading had a median error only 0.00128 eV per atom below ridge across five seeds with overlapping ranges, and its residual-target construction is incompletely archived. These results delimit component behaviour under ideal simulation and identify the controls and provenance needed to test whether circuit-specific benefits persist.

1. Introduction

Relaxed energies of oxide catalysts are expensive to obtain by density-functional theory (DFT), yet they are needed across large and chemically diverse structure spaces. Learned atomistic surrogates amortise this cost, and equivariant graph neural network (GNN) potentials now provide strong classical models with locality and symmetry built into their representations (Hohenberg & Kohn, 1964; Kohn & Sham, 1965; Tran et al., 2023; Batzner et al., 2022; Batatia et al., 2022; Musaelian et al., 2023). Any quantum-learning contribution must therefore be assessed against a mature classical landscape rather than against DFT cost alone.

[1]Department of Engineering and System Science, National Tsing Hua University , 300044, Hsinchu City, Taiwan.

Quantum circuits can enter such a surrogate as optimisation routines, fixed feature maps or trainable predictors (Farhi et al., 2014; Schuld & Killoran, 2019; Jerbi et al., 2023). The unresolved question is whether these circuit-defined components contribute information or search behaviour that survives matched classical controls. A difference observed under classical circuit simulation may instead reflect candidate pruning, generic nonlinearity, model capacity or validation reuse. Circuit evaluation must also be separated from quantum-hardware execution: a simulated state and its finite-shot readout do not establish a hardware speed-up.

Figure 1a separates first-principles label generation from supervised surrogate prediction: DFT supplies the relaxed total energies, whereas the Quantum Bayesian Machine (QBM) predicts those labels from fixed structural descriptors. Figure 1b then locates QBM within the simulator landscape. QBM propagates its small ideal circuit states exactly on a CPU; the fidelity and re-uploading branches read overlaps, Pauli-Z expectations and gradients directly from those amplitudes, whereas the QAOA branch draws 1,024 seeded samples from a simulated state to propose subsets. None of these circuit quantities is an electronic Hamiltonian expectation or an OC22 energy.

Here we test circuit-defined search and representation within one controlled OC22 IS2RE-Total surrogate. The descriptor, partitions and classical preprocessing remain fixed, while simulated QAOA proposes feature subsets, a fidelity kernel supplies a Gram matrix to ridge regression and a variational re-uploading circuit learns residuals. Classical rescoring and paired ridge controls provide conditional component comparisons, subject to shared validation selection and the record boundary. The architecture and evaluation logic are summarised in Fig. 2. The study consequently provides a documented component test under ideal local simulation; it does not test quantum hardware, replace equivariant GNN potentials or presume quantum advantage. Reproducing the IS2RE-Total fits requires the original run configuration and outputs, which are not fully archived with the released code (Appendices F.1 and F.4).

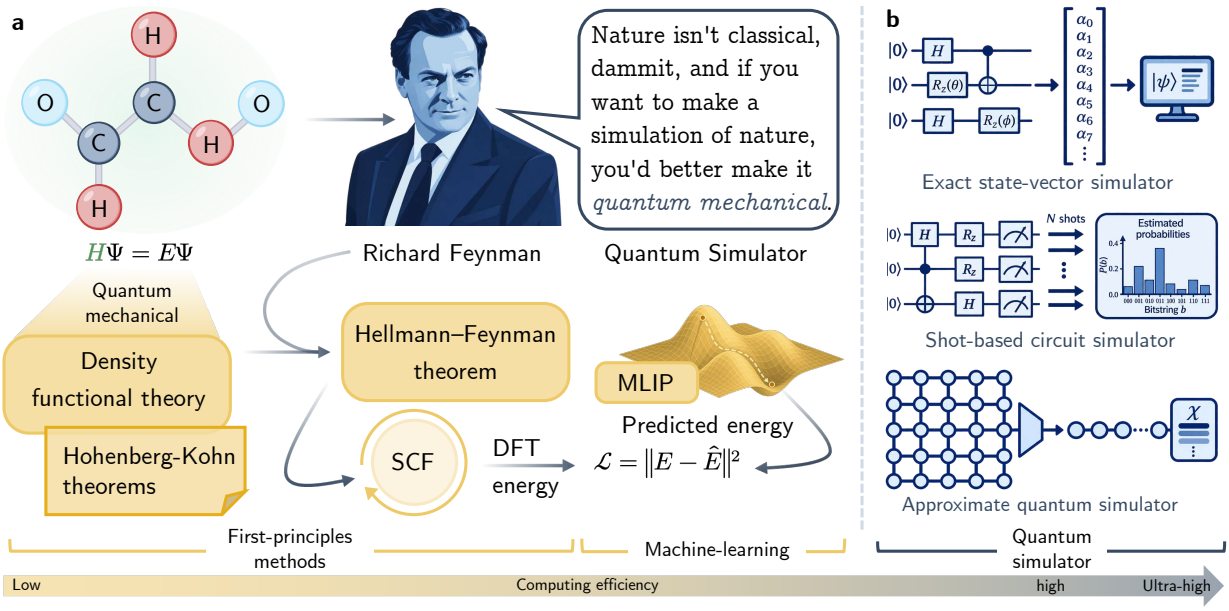


Figure 1. **First-principles energetics, learned energy surrogates and quantum-circuit simulation. a,** Density-functional theory (DFT) supplies relaxed total energies for supervised learning. QBM lies downstream of DFT and predicts these labels from fixed structural descriptors; it does not learn forces or propagate molecular dynamics. The quotation is a shortened excerpt from Feynman's discussion of simulating nature (Feynman, 1982); it omits the concluding clause of the source sentence and does not describe this surrogate's computation. **b,** Three classes of classical circuit simulation are shown: complete state-vector propagation, shot-based output estimation and approximate compressed-state simulation. QBM uses the first class; the other branches locate it within the broader simulator landscape. Arrows and the lower efficiency scale are conceptual rather than measured comparisons; the high-efficiency placement of the quantum-simulator branch does not describe QBM runtime or establish a quantum advantage. The energy landscape and force-related motifs in panel a are illustrative; QBM predicts relaxed energies only.

2. Background and Related Work

OC22 defines supervised total-energy tasks for oxide surfaces from periodic DFT relaxations and supplies fixed in-domain (ID) and out-of-domain (OOD) partitions, extending the earlier OC20 adsorption-energy setting (Tran et al., 2023; Chanussot et al., 2021). Classical atomistic models range from prescribed invariant descriptors to learned equivariant representations (Behler & Parrinello, 2007; Bartók et al., 2013; Artrith et al., 2017; Batzner et al., 2022; Batatia et al., 2022). Their physical priors, energy–force objectives and training corpora make modern GNN potentials the practical reference for atomistic accuracy and scaling. They are not matched controls for the present fixed-descriptor study, which instead asks whether changing one downstream computational component alters behaviour under otherwise shared inputs.

QBM adopts a modular sparse-surrogate structure. Composition regression removes a coarse baseline, projected descriptor subsets are ranked with a multiplicity-penalised criterion, held-out predictions determine non-negative combination weights and a final model learns the residual (Mamun et al., 2020; Chen & Chen, 2008). This decomposition permits component tests, but it is only Bayesian-motivated: the executed path returns point predictions and does not define a coherent posterior or calibrated predictive interval.

The quantum components occupy distinct statistical roles. QAOA proposes binary subsets for subsequent classical fitting (Farhi et al., 2014); fidelity kernels convert fixed circuit-state overlaps into a Gram matrix (Havlíček et al., 2019; Schuld & Killoran, 2019); and variational re-uploading trains data-dependent circuit observables (Pérez-Salinas et al., 2020; Jerbi et al., 2023). Expressive or noisy fidelity maps can concentrate and become costly to estimate, which motivates the spectral diagnostics used here (Thanasilp et al., 2024). These supervised representations differ from variational quantum eigensolvers, which estimate physical Hamiltonian expectations (Cervera-Lierta et al., 2021). Section 3 defines the quantities required for QBM, while Appendix G provides the fuller history of classical potentials, quantum surrogates, simulator–hardware studies and direct quantum energy estimation.

3. Preliminaries

We first define the physical target and then the statistical and circuit-defined quantities used to predict it. The target is a

relaxed DFT total energy. By contrast, the QAOA objective, quantum-state fidelity and Pauli-Z expectations are intermediate algorithmic quantities evaluated by local classical simulation; they are neither electronic Hamiltonian expectations nor relaxed energies. This separation is maintained throughout the model specification. Appendix H collects the notation.

3.1. Problem setup for IS2RE-Total on OC22

The OC22 initial-structure-to-relaxed-energy task predicts the relaxed DFT total energy from an unrelaxed periodic structure (Tran et al., 2023). A sample is $\mathcal{S} = (\{\mathbf{r}_i\}_{i=1}^N, \{Z_i\}_{i=1}^N, \mathbf{L})$, where $\mathbf{r}_i$, Z_i and $\mathbf{L}$ denote atomic positions, atomic numbers and the periodic lattice. The target is $E = E_{\mathrm{DFT}}(\mathrm{relax}(\mathcal{S}))$. Thus, prediction replaces the relaxation but not its relaxed-energy label. Dataset provenance and splits are given in Appendix F.1.

3.2. Atom-centred structure representation

We first represent each local environment by an atom-centred feature vector $\mathbf{f}_i$ that is invariant to translation, orthogonal transformations and permutations of like atoms within a cutoff R_{c}. These features are pooled into a structure-level vector before regression (Section 4.1). Consequently, the complete nonlinear predictor is not an atomically additive energy model, even though its global pooling block is additive. To prevent descriptor size from increasing with chemical diversity, we extend the constant-complexity radial and angular representation of Artrith et al. (2017) with a torsional distribution function (TDF). The radial (RDF), angular (ADF) and torsional terms encode neighbour distances R_{ij}, bond angles ϑ_{ijk} and unsigned dihedrals ϕ_{ijkl}, respectively (Appendix B.2).

Analytic projection onto first-kind Chebyshev polynomials gives finite neighbour sums,

$$c^{\mathrm{R}}_{i,\alpha} = \sum_{j\neq i} \tau(Z_j)\, T_\alpha\big(\bar{R}_{ij}\big)\, f_{\mathrm{c}}(R_{ij}), \tag{1}$$

$$c^{\mathrm{A}}_{i,\alpha} = \sum_{j\neq i} \sum_{k\neq i,j} \tau(Z_j)\,\tau(Z_k)\, T_\alpha\big(\cos\vartheta_{ijk}\big) \times f_{\mathrm{c}}(R_{ij})\, f_{\mathrm{c}}(R_{ik}), \tag{2}$$

$$c^{\mathrm{T}}_{i,\alpha} = \sum_{(j,k,l)\in\mathcal{T}_i} \tau(Z_j)\,\tau(Z_k)\,\tau(Z_l)\, T_\alpha\big(\cos\phi_{ijkl}\big) \times f_{\mathrm{c}}(R_{ij})\, f_{\mathrm{c}}(R_{jk})\, f_{\mathrm{c}}(R_{kl}), \tag{3}$$

where $\bar{R}_{ij} = 2R_{ij}/R_{\mathrm{c}} - 1$, $\tau(Z)$ is a species weight and f_{c} is a cosine cut-off (Behler & Parrinello, 2007). Uniform and pseudospin weights yield structural and compositional blocks. Their concatenation has dimension $D = 2(n_{\mathrm{R}} + n_{\mathrm{A}} + n_{\mathrm{T}})$, independent of the number of elements. Appendix B gives the derivation and invariance properties.

3.3. Kernel ridge regression

For feature matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$, kernel ridge regression (KRR) solves

$$(\mathbf{K} + \lambda\mathbf{I})\,\boldsymbol{\beta} \;=\; \mathbf{y} - \bar{y}\mathbf{1}, \qquad K_{mn} = \kappa(\mathbf{x}_m, \mathbf{x}_n), \tag{4}$$

and predicts $\hat{f}(\mathbf{x}) = \bar{y} + \sum_m \beta_m \kappa(\mathbf{x}, \mathbf{x}_m)$. We use either an RBF kernel or the fidelity kernel defined below. Nyström landmarks provide an explicit approximation when required. Ridge strength is selected by generalised cross-validation (GCV), which also defines

$$\hat{\sigma}^2 \;=\; \frac{\mathrm{SSE}}{n\,\big(1 - \mathrm{df}/n\big)^2}, \tag{5}$$

where df is the effective degrees of freedom.

3.4. Modified BIC and model combination

Candidate subsets are ranked by an implemented heuristic denoted mBIC,

$$\mathrm{mBIC}(\mathcal{M}) \;=\; \ln\frac{\hat{\sigma}^2_{\mathcal{M}}}{\hat{\sigma}^2_0} \;+\; \frac{k_{\mathrm{B}}\ln n}{n} \;+\; \frac{2\gamma}{n}\, k_{\mathrm{S}} \ln\frac{p}{k_{\mathrm{S}}}, \tag{6}$$

where $d = |\mathcal{M}|$, $k_{\mathrm{S}} = d + 1$ is the subset-complexity count and $\hat{\sigma}^2_0$ is the smallest scored variance in the candidate pool. For classical and fidelity KRR candidates, the implementation also sets the BIC-style count $k_{\mathrm{B}} = k_{\mathrm{S}}$, rather than counting the fitted Nyström coefficients. For a variational re-uploading finalist it instead sets $k_{\mathrm{B}} = N_\theta = 4qL + q + 1$, while the multiplicity term remains a function of k_{S}. The KRR branches use GCV variance; the variational finalist uses its recorded training-residual variance. The score is therefore a backend-aware ranking heuristic, not a likelihood-based BIC for either fitted model. Because k_{S} includes the intercept, the implemented multiplicity term becomes negative for the full subset $d = p$; it is retained only to reproduce the executed ranking and must not be interpreted as the exact combinatorial EBIC penalty. The last term is EBIC-style (Bogdan et al., 2004; Chen & Chen, 2008); we do not transfer the selection-consistency theorem of exact EBIC (Appendix C).

The K_{ens} lowest-scoring models are combined using conditional fold-excluding subset-refit predictions,

$$\mathbf{w} \;=\; \operatorname*{arg\,min}_{\mathbf{w}\geq 0,\ \mathbf{1}^\top\mathbf{w}=1} \big\|\mathbf{Y}^{\mathrm{oof}}\mathbf{w} - \tilde{\mathbf{y}}\big\|_2^2. \tag{7}$$

These predictions arise from fold-excluding subset-regressor refits after the supervised composition baseline, PLS projection and subset identities have been fitted or chosen on the full training set. They are conditional out-of-fold predictions, not predictions from a fully cross-fitted pipeline; the residual targets inherit that distinction (Appendices A and C.4).

3.5. Quantum primitives

A quadratic unconstrained binary optimisation (QUBO) represents a subset by $\mathbf{z} \in \{0,1\}^p$. The mapping $z_i = (1 - s_i)/2$ yields an Ising Hamiltonian $\hat{H}_C$. QAOA alternates this cost Hamiltonian with a transverse mixer and optimises the circuit angles classically (Farhi et al., 2014). Samples from the optimised state provide candidate subsets (Appendix D.4).

For $F \in \{Z, ZZ\}$, the implemented feature map is the unitary $U_F(\mathbf{x})$ shown in Fig. 2d,e and prepares $|\phi_F(\mathbf{x})\rangle = U_F(\mathbf{x})|0\rangle^{\otimes q}$. Using the figure's phase-gate notation $P(\varphi) = \mathrm{diag}(1, e^{i\varphi})$, its fidelity kernel is

$$\kappa_F(\mathbf{x}, \mathbf{x}') = \left|\langle \phi_F(\mathbf{x}') \mid \phi_F(\mathbf{x})\rangle\right|^2, \tag{8}$$

which is positive semidefinite (Havlíček et al., 2019; Schuld & Killoran, 2019). We evaluate an implemented single-body Z map and a ZZ map with data-dependent ring couplings; both reapply a Hadamard wall in every encoding layer (Appendix E.1).

Data re-uploading instead interleaves input-dependent and trainable gates (Pérez-Salinas et al., 2020). Because the resulting state depends jointly on data and parameters, we use it as a variational residual model, not as a fixed kernel (Jerbi et al., 2023). Circuit definitions are given in Appendix E.

4. Method

QBM combines a fixed atom-centred representation with sparse selection, model combination and residual correction. Five stages are shared across all variants.

1. Construct and pool pseudospin descriptors (Section 4.1).
2. Remove the composition baseline and learn a supervised projection (Section 4.2).
3. Propose sparse subsets by enumeration or QUBO/QAOA, then rank them by mBIC (Sections 4.3 and 4.4).
4. Fit retained subsets with classical or fidelity-kernel ridge regression (Section 4.5).
5. Combine held-out predictions and fit a residual correction (Section 4.6).

Figure 2a–c follows these shared stages from structural encoding to subset fitting and residual correction, whereas Fig. 2d–g isolates the circuit constructions tested within that pipeline. The four named QBM variants use different assignments of subset proposer, regression kernel and residual learner (Section 5.1). Because the complete variants are not a factorial design, component-level attribution is restricted to the *baseline–quantum* selector comparison and the paired residual ablations that hold the remaining fitted state fixed. Appendix A gives the complete configuration and pseudocode.

4.1. Pseudospin descriptor and locality-aware pooling

Each atom descriptor $\mathbf{f}_i$ is standardised and passed through a Nyström RBF map $\boldsymbol{\zeta}$. The global sum

$$\boldsymbol{\Phi}_{\mathrm{global}}(\mathcal{S}) \;=\; \sum_{i=1}^{N} \boldsymbol{\zeta}(\mathbf{f}_i) \tag{9}$$

is permutation invariant and additive. A linear readout of this block is size-extensive, although the subsequent projection and nonlinear stages are not.

To retain adsorption-site information, we concatenate a locality-weighted sum. Let r_i^{ads} be the minimum-image distance to the nearest adsorbate atom and ℓ_{ads} the decay length. Then

$$\boldsymbol{\Phi}(\mathcal{S}) \;=\; \Big[\, \sum_i \boldsymbol{\zeta}(\mathbf{f}_i) \;\Big\|\; \sum_i e^{-r_i^{\mathrm{ads}}/\ell_{\mathrm{ads}}}\, \boldsymbol{\zeta}(\mathbf{f}_i) \,\Big]. \tag{10}$$

If adsorbate tags are absent, the second block defaults to the global pool. Figure 2a shows where these structural and pseudospin-weighted distributions enter the shared representation, before any selector or circuit-specific regressor is introduced.

4.2. Composition baseline and dimensionality reduction

A cross-validated ridge model first removes the dominant composition dependence,

$$\hat{E}_{\mathrm{comp}}(\mathcal{S}) \;=\; \sum_Z \mu_Z\, n_Z(\mathcal{S}) \;+\; E_0.$$

Subsequent stages learn the residual target $\tilde{y} = E - \hat{E}_{\mathrm{comp}}$. The pooled descriptor is standardised and projected onto p partial least-squares (PLS) components (Wold et al., 2001). PLS is fitted only on the training split. The shared width p bounds the full subset space and the QUBO size; the encoding qubit count is set separately. As traced in Fig. 2b, this projection is the common input to both enumeration and QUBO/QAOA proposal, which prevents descriptor changes from confounding the selector comparison.

4.3. Heuristic subset selection via mBIC

The reference selector fits every non-empty subset of the p PLS components. Each Nyström RBF KRR chooses its ridge strength by GCV, after which all subsets are scored by equation (6). The exact sweep is feasible at small p and defines the control for the quantum proposer. The heuristic mBIC values are pool-relative because $\hat{\sigma}_0^2$ is defined within the scored candidate set.

4.4. Quantum subset selection (QUBO and QAOA)

The quantum selector replaces complete subset enumeration, not exact scoring. Stage 1 fits every singleton and

pair, defining the finite-difference coefficients a_i and b_{ij} of a quadratic log-variance surrogate. Stage 2 solves one cardinality-anchored QUBO for each target size d_t,

$$\mathcal{Q}(\mathbf{z}) = \sum_i a_i z_i + \sum_{i<j} b_{ij} z_i z_j + c_t \sum_i z_i + \nu\Big(\sum_i z_i - d_t\Big)^2, \tag{11}$$

where $c_t = \ln n/n + (2\gamma/n)\max[\ln(p/d_t), 0]$ and $\nu = \nu_0 \max\{\mathrm{median}_i\,|a_i|, 10^{-8}\}$, with $\nu_0 = 0.5$ in the canonical configuration. This scaling makes the anchor relative to the singleton-effect scale rather than an absolute QUBO coefficient. Depth-L_Q QAOA proposes low-energy bitstrings through a seeded local sampler. Stage 3 refits every distinct proposal and applies the exact mBIC. Constructing the quadratic surrogate requires $O(p^2)$ fitted singleton and pair models, compared with $O(2^p)$ fitted subsets for enumeration, but this count excludes circuit optimisation and sampling and the resulting candidate pool no longer covers the full power set (Appendix D).

Here, "energy" in the QAOA objective denotes the sampled expectation of the combinatorial Ising cost in equation (11); it is not the OC22 relaxed total energy. The code admits each distinct returned bitstring that satisfies the subset-size bounds without using its empirical weight for ranking. If the sampled pool exceeds its budget, the unanchored quadratic surrogate ranks the proposals before every retained subset enters the same classical fitting and mBIC rescoring pipeline.

4.5. Quantum-kernel ridge regression

To test the regression feature space independently of subset proposal, we replace the RBF kernel with κ_{ZZ} from equation (8). A selected subset is standardised and encoded on $q = \min(d, q_{\max})$ qubits. The layer count is $\max(L_0, \lceil d/q \rceil)$, ensuring that every coordinate is encoded. Direct batched state-vector calculation evaluates the Gram matrix before the classical ridge solve. The Z alternative removes all pair gates and is retained as an ablation; 200-landmark Nyström features approximate the configured large-sample kernel (Appendices E.1 to E.3).

This component changes the induced similarity, not the statistical target or the amount of training data. Figure 2d,e distinguishes the single-body Z encoding from the pair-coupled ZZ encoding that generates this alternative geometry. The direct amplitude evaluator produces state overlaps, the overlaps form the Gram matrix, and the Gram matrix is consumed by classical ridge regression; the circuit does not output an energy prediction directly. With $q \leq 6$, the reported kernels are evaluated exactly on the CPU and do not demonstrate a computational advantage.

4.6. Model combination and residual learning

The K_{ens} best subset identities, chosen using the full training set, are refitted across n_{fold} folds to form $\mathbf{Y}^{\mathrm{oof}}$. Non-negative least squares estimates simplex-constrained weights from those conditional fold-excluding predictions. The composition baseline and supervised PLS projection precede these folds and use all training labels, so the residual targets are not fully cross-fitted. The reduced-profile re-uploading ablations omitted even this conditional refit (Appendix F.6).

The residual learner uses the pooled descriptor or a residual-targeted PLS projection, following the conditional residual construction described alongside Fig. 2c. The *baseline* and *quantum* variants use a linear-kernel Gaussian process (Rasmussen & Williams, 2006); above 2,000 samples, its posterior mean is evaluated by the equivalent exact ridge solve. The *quantum_kernel* variant uses the fidelity kernel through a Nyström map. The *quantum_reupload* variant uses the variational residual model in Appendix E.5. Figure 2f,g contrasts the implicit/interleaved and explicit/separated circuit layouts; the executed model uses the implicit/interleaved construction. Experimental comparisons use the unambiguous labels *trained circuit* and *frozen-encoding kernel control*. Both models were evaluated by classical state-vector calculations, not quantum hardware. The final estimate is

$$\hat{E}(\mathcal{S}) = \hat{E}_{\mathrm{comp}}(\mathcal{S}) + \underbrace{\textstyle\sum_m w_m\, \hat{y}_m(\mathbf{x})}_{\text{ensemble}} + \underbrace{\hat{r}(\mathbf{u})}_{\text{residual}}, \tag{12}$$

where $\mathbf{u}$ denotes the residual learner input.

5. Experiments

5.1. Overview

QBM is instantiated as four named variants: *baseline*, *quantum*, *quantum_kernel* and *quantum_reupload*. They share the descriptor, locality-aware pooling, composition baseline, PLS projection and combination procedure, but use the component assignments in Table 1. The residual learner itself is one of the assigned components.

The reported full-profile variants run at the same projection width $p = 16$, which bounds the 2^p sweep of the baseline selector and keeps the QAOA instance simulable. The *baseline* and *quantum* variants differ only in subset proposal. The other named variants change multiple assignments relative to baseline and are not, by themselves, single-component controls; attribution instead uses the paired kernel-off, frozen-kernel and circuit-off ablations reported below. QAOA is sampled locally, whereas the fidelity and re-uploading circuits use direct state-vector updates; no quantum hardware or cloud account is required. Appendix A gives the software stack and hyperparameters, Appendix F.1 the dataset provenance and splits, and Appendix F.2 the training and evaluation protocol.

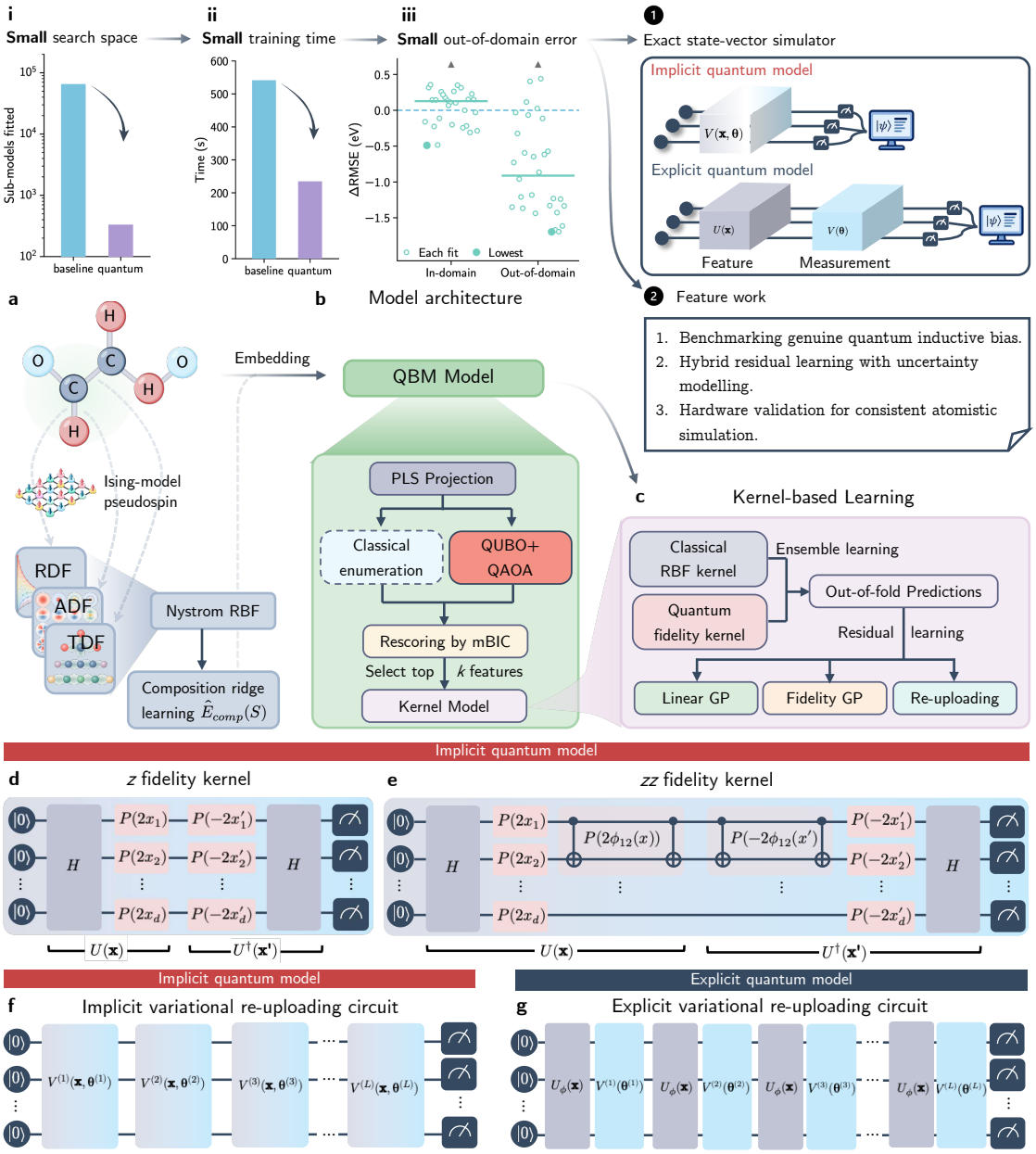


Figure 2. **QBM architecture and component-level evaluation. i,** Recorded fitted candidate counts for exhaustive enumeration and QAOA proposal; candidate identifiers are unavailable. **ii,** Recorded CPU fitting times for the matched selector comparison; primary timing logs are unavailable. **iii,** Changes in ID and OOD RMSE relative to the enumerated baseline across the recorded fits; filled points mark the lowest value in each split, not a consistent OOD benefit. **❶,** Exact local state-vector evaluation of the circuit models. **❷,** Prospective tests of quantum inductive bias, uncertainty-aware residual learning and hardware execution. **a,** Structural and pseudospin-weighted radial, angular and torsional distributions encode each atomic environment. **b,** A composition baseline and pooled Nyström features feed a supervised PLS projection; enumeration or QUBO/QAOA proposes coordinate subsets for mBIC rescoring and kernel fitting. **c,** The canonical pipeline uses conditional fold-excluding subset-refit predictions to construct ensemble residual targets. The reduced-profile re-uploading ablations omit this pass and their replacement target construction is unavailable. The artwork label "Fidelity GP" denotes the fidelity-kernel residual branch, implemented as fidelity KRR for the reported variant; "top k features" refers to retained subset models. **d,e,** The Z and ZZ circuits define the fidelity kernels; the ZZ map adds data-dependent pair phases. **f,g,** The implicit layout combines data-dependent and trainable operations within $V^{(\ell)}(\mathbf{x}, \boldsymbol{\theta}^{(\ell)})$, whereas the explicit layout displays $U_{\phi}^{(\ell)}(\mathbf{x})$ and $V^{(\ell)}(\boldsymbol{\theta}^{(\ell)})$ separately. The executed *quantum_reupload* model uses the implicit/interleaved construction in Appendix E.5; experimental comparisons are labelled trained circuit and frozen-encoding kernel control. All reported circuits use CPU state-vector simulation; the ❷ callout is prospective.

Table 1. Component assignments for the four pipeline variants evaluated on OC22 IS2RE-Total. The subset selector, regression kernel and residual learner distinguish the variants; the remaining stages follow Section 4.

Variant	Selector	Kernel	Residual
baseline	Enumeration	RBF	Linear GP
quantum	QAOA	RBF	Linear GP
quantum_kernel	QAOA	Fidelity kernel	Fidelity KRR
quantum_reupload	QAOA	RBF	Re-uploading circuit

Comparison discipline. Computational repeats used three to six seeds, depending on the comparison, and altered subset proposal, Nyström landmarks or circuit initialisation. The fixed structure partitions and shared validation selection make these computational repeats dependent for generalisation claims. We report medians and observed ranges as measures of computational variation, not confidence intervals. Configuration selection and performance reporting reuse the validation partitions. Component claims therefore rely on paired ablations that share the selected subset and reference variance seed by seed; the complete repeat structure and selection rules are given in Appendix F.3.

Table 2 places representative fits on a common metric scale but also shows why the variant names do not support an uncontrolled ranking. Only the *baseline–quantum* rows isolate subset proposal; the residual-family rows were shortlisted by ID RMSE under different residual widths and training profiles. The component-level evidence is therefore developed below from matched panels and paired fit records.

5.2. QAOA reduces fitted search without establishing predictive advantage

We first asked whether QAOA could reduce the number of fitted subsets while preserving the surrounding estimator. This is the only recorded end-to-end comparison in which the two named variants differ by a single assigned component; it does not constitute a budget-matched comparison against another heuristic. At $p = 16$, exhaustive enumeration fitted all 65,535 non-empty subsets, whereas the QAOA workflow fitted 333 candidates (0.51% of the search space). Figure 2i,ii displays the reported candidate counts and timing, which cannot be checked against primary candidate lists or logs: selection time decreased from 323 to 22 s and total fitting time from 541.67 to 234.84 s.

The selected QAOA subset lay one Hamming step from the enumerated optimum. Figure 2iii shows that the corresponding RMSE shifts also differed by split, whereas the matched rows of Table 2 quantify the MAE change: relative to enumeration, ID MAE increased by 0.00172 eV/atom and OOD MAE decreased by 0.00253 eV/atom. Only two QAOA runs are available, and no classical heuristic was evaluated under the same candidate or runtime budget. The result therefore supports a reduction in this CPU-simulated fitting workflow, not a quantum speed-up or a general predictive improvement. Candidate identities and raw logs are unavailable, so the 333-model pool is retained as a reported result rather than independently reconstructed (Appendix F.4).

5.3. Fidelity-kernel effects remain validation-specific

We next tested whether a circuit-defined fidelity geometry improved residual prediction relative to a paired linear-ridge control. Figure 3a locates the lowest recorded ZZ-map errors within a narrow low-scale region and against the paired ridge range. Figure 3c shows how the same scale sweep changes the offline Gram effective rank, whereas Fig. 3b relates Gaussian-probe effective rank to ID MAE across 23 dependent Z- and ZZ-map cells (ρ_s=0.946). Read together, panels a and c show that increasing the ZZ scale drove the diagnostic Gram matrix towards an identity-like high-rank regime while validation error rose after the rank plateau. The Gaussian probe inputs and validation-selected cells, however, do not establish a causal concentration mechanism.

At the selected ZZ scale, the fidelity residual had lower ID MAE than ridge in all six paired fits, with a median reduction of 0.00433 eV/atom (range, 0.00204–0.00472 eV/atom). The same comparison did not transfer to OOD validation: ridge had lower OOD MAE in five of six pairs, with a median difference of 0.00080 eV/atom, smaller than the observed OOD seed range. The Z and ZZ sweeps also used different selected subsets and reference variances, so their 0.00704 eV/atom median difference does not isolate the pair phase.

Thus, the selected ZZ kernel changed ID residual prediction, but the present evidence cannot distinguish a quantum-specific inductive bias from generic nonlinear capacity or validation selection. Table 6 supplies the fit-level errors and paired ridge rows behind this comparison, whereas Table 8 supplies the Gram-rank and off-diagonal-similarity values used to interpret Fig. 3b,c. The diagnostic protocol is detailed in Appendix F.5.

5.4. Re-uploading remains unresolved against ridge

The re-uploading experiments asked whether a trainable circuit improved residual learning beyond its frozen-encoding kernel and a linear ridge model. Figure 4a visualises the

Table 2. Representative validation outcomes for the four QBM variants on the official OC22 IS2RE-Total ID and OOD splits. The *baseline* and *quantum* variants are the matched selector pair from 3 September 2026. The remaining rows are the three distinct non-ridge fits with the lowest ID RMSE within each residual family in Tables 6 and 7; OOD metrics were not used for selection. Rows are individual validation fits, and differences in residual width and training profile preclude a controlled cross-family ranking. MAE is the mean of absolute total-energy errors divided by each structure's atom count (eV/atom); RMSE uses unnormalised total-energy errors (eV).

Model	ID MAE	ID RMSE	OOD MAE	OOD RMSE
baseline	0.15195	17.1657	0.22673	28.1378
quantum	0.15367	17.3149	0.22420	27.7175
quantum_kernel (ZZ, $\eta_{\mathrm{K}} = 0.03$)	0.15631	17.6336	0.22515	29.0394
quantum_kernel (ZZ, $\eta_{\mathrm{K}} = 0.03$)	0.15532	17.7073	0.22113	28.9042
quantum_kernel (ZZ, $\eta_{\mathrm{K}} = 0.02$)	0.15604	17.7206	0.21957	29.0160
quantum_reupload ($q = 4$, $L = 36$, $\eta_{\mathrm{R}} = 0.10$)	0.15232	16.6763	0.21478	27.4100
quantum_reupload ($q = 6$, $L = 24$, $\eta_{\mathrm{R}} = 0.15$)	0.15214	16.6819	0.21704	27.5431
quantum_reupload ($q = 6$, $L = 24$, $\eta_{\mathrm{R}} = 0.45$)	0.15369	16.8564	0.21301	26.7101

encoding-scale component of the ablation. Across the recorded scale, depth and register-width comparisons, median ID shifts were no larger than 0.00198 eV/atom and seed ranges overlapped. Figure 4b shows that initial tangent-kernel effective rank did not order the fitted errors across the scale sweep ($\rho_{\mathrm{s}} = 0.000$). Figure 6a,b broadens this diagnostic across register width and upload depth, respectively, but those configurations were not fitted end to end; effective rank therefore serves as a capacity descriptor rather than evidence of predictive improvement.

Figure 4c and the paired rows of Table 7 record lower validation error for the trainable circuit than the frozen-encoding kernel at every seed within the reduced profile. Because that profile omits the canonical out-of-fold residual pass and its replacement target construction is not archived, these rows do not establish a fold-safe residual-learning advantage. Reducing the effective-rank mismatch narrowed their median gap from 0.01911 to 0.01152 eV/atom, but effective rank alone cannot attribute the remaining difference. Against linear ridge, the circuit median was only 0.00128 eV/atom lower across five seeds. Figure 5a,b places this difference within the observed ID and OOD seed ranges, which overlap for the circuit and ridge controls. The recorded profile shows higher validation error for the frozen fidelity kernel; incomplete residual-target provenance precludes a controlled residual-learning inference and does not establish that re-uploading outperforms a classical residual learner.

These comparisons use a reduced ablation profile whose exact residual-target construction cannot be reconstructed from the released configuration. Full settings, seed ranges and this reproducibility boundary are given in Appendix F.6, Table 7, and Fig. 4.

6. Discussion

QBM separates subset search from residual representation, which prevents unlike end-to-end variants from being mistaken for evidence of one quantum effect. The selector panels in Fig. 2i–iii show that a sampled QAOA proposal can sharply restrict a small combinatorial search before classical rescoring, while the matched rows in Table 2 show no consistent accuracy direction across ID and OOD validation. The residual evidence is less favourable: Fig. 3a and Fig. 4c show that the selected ZZ kernel changed ID error without an OOD median gain, and that re-uploading was not resolved from ridge at the scale of seed variation. Taken together, these observations support component-specific behaviour under exact simulation, not a general accuracy or computational advantage.

The comparison also defines the role that circuit models can reasonably play beside equivariant GNN potentials. QBM applies circuits only after an invariant descriptor has been pooled and compressed; it does not learn locality, equivariance or conservative forces. Its circuit states therefore test alternative subset proposals and similarity geometries rather than a replacement for end-to-end atomistic learning. The association between near-identity fidelity Gram matrices and larger validation error is compatible with known quantum-kernel concentration, but the present offline probe cannot establish that mechanism (Thanasilp et al., 2024). Likewise, the trained–frozen comparison primarily reveals weakness in the frozen kernel, not a uniquely quantum benefit of the trainable circuit.

The evidence boundary is set by validation reuse, three to six computational seeds, unmatched nonlinear capacity, one OC22 task and the absence of hardware or noise experiments. Raw per-structure predictions, checkpoints and candidate identities are also unavailable, while the negative OOD mean errors remain unexplained. QBM is consequently a code-available scaffold for sharper tests, but the IS2RE-Total results are not fully reproducible from the archived configuration and aggregate metrics. It is not a calibrated Bayesian predictor, a force-conservative MLIP or evidence of quantum advantage. The matched controls, locked partitions

and hardware progression required to test stronger claims are specified in Section 7.

7. Future work

We should test whether circuit-defined search or representation provides an inductive bias that survives comparably resourced classical controls. Reproducibility first requires retaining per-structure predictions, candidate identities, checkpoints and seeded configurations. On locked independent partitions, QAOA should be compared with classical selectors under matched candidate evaluations and runtime, while fidelity and re-uploading models should be compared with kernels and compact neural residuals matched in input information, effective capacity and optimisation budget. Scaling over data size, projection width, qubit number, depth, encoding scale and shots should relate predictive error to kernel spectra, effective rank and target alignment, thereby distinguishing useful representation from concentration (Huang et al., 2021; Thanasilp et al., 2024).

Rather than replacing equivariant graph potentials, a small circuit could learn a residual correction to a frozen symmetry-preserving atomistic representation. Tests should target low-data and chemically shifted regimes and include matched classical heads (Batzner et al., 2022; Batatia et al., 2022; Wood et al., 2025). Projected or symmetry-aware quantum kernels offer alternatives when full-state fidelity becomes uninformative (Glick et al., 2024). A coherent predictive distribution or calibrated ensemble could further support active learning, evaluated by interval coverage and the number of additional DFT labels required to reach a target error.

Hardware experiments should follow these matched simulation studies. Exact state-vector, finite-shot, noise-injected and real-device results must remain separate, with compiled circuit cost, shots, mitigation and end-to-end resource use reported alongside accuracy (Lo Monaco et al., 2024; Belaloui et al., 2025). Longer-term work may extend the surrogate to differentiable energies and conservative forces, assessed through force error, energy conservation and molecular-dynamics stability (Kiss et al., 2022; Le et al., 2025; Couzinié et al., 2025). Quantum utility would require any improvement to persist across classical controls, distribution shift, uncertainty calibration and physical execution constraints.

8. Conclusion

QBM demonstrates an exploratory comparison of circuit-defined subset proposal and residual representation within an OC22 energy surrogate. Records report fewer fitted candidates and shorter CPU fitting time for the QAOA workflow than for enumeration, but the underlying candidate lists and timing logs are unavailable. Fidelity-kernel differences depend on validation selection, and the reduced-profile re-uploading targets cannot be reconstructed; neither establishes quantum-specific predictive benefit. The methodological contribution is a component-level evaluation framework that makes these attribution and provenance limits explicit. Establishing utility will require locked test partitions, budget- and capacity-matched classical controls, retained run-level artefacts and a staged comparison of ideal simulation, noise and hardware.

Source availability

The Open Catalyst 2022 (OC22) dataset used in this work is publicly available through the Open Catalyst Project at `https://facebookresearch.github.io/fairchem/oc22/`. The QBM source code and pinned environment are available at `https://github.com/FrancisPhillipsBuger/Quantum_Bayesian`. Aggregate metrics are retained in this manuscript repository, but the original run configurations, raw logs, candidate identities and per-structure predictions needed to independently regenerate all reported results are not archived. An interactive demonstration website is available at `https://francisphillipsbuger.github.io/pub-bayesian-energy`.

References

Artrith, N., Urban, A., and Ceder, G. Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. *Physical Review B*, 96 (1):014112, 2017. doi: 10.1103/PhysRevB.96.014112.

Bartók, A. P., Payne, M. C., Kondor, R., and Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical Review Letters*, 104(13):136403, 2010. doi: 10.1103/PhysRevLett.104.136403.

Bartók, A. P., Kondor, R., and Csányi, G. On representing chemical environments. *Physical Review B*, 87(18): 184115, 2013. doi: 10.1103/PhysRevB.87.184115.

Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C., and Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 11423–11436, 2022.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022. doi: 10.1038/s41467-022-29939-5.

Behler, J. and Parrinello, M. Generalized neural-network

representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98(14):146401, 2007. doi: 10.1103/PhysRevLett.98.146401.

Belaloui, N. E., Tounsi, A., Khamadja, A. R., Louamri, M. M., Benslama, A., Bernal Neira, D. E., and Rouabah, M. T. Ground-state energy estimation on current quantum hardware through the variational quantum eigensolver: A practical study. *Journal of Chemical Theory and Computation*, 21(14):6777–6792, 2025. doi: 10.1021/acs.jctc.4c01657.

Bogdan, M., Ghosh, J. K., and Doerge, R. W. Modifying the schwarz bayesian information criterion to locate multiple interacting quantitative trait loci. *Genetics*, 167(2):989–999, 2004. doi: 10.1534/genetics.103.021683.

Cervera-Lierta, A., Kottmann, J. S., and Aspuru-Guzik, A. Meta-variational quantum eigensolver: Learning energy profiles of parameterized hamiltonians for quantum simulation. *PRX Quantum*, 2(2):020329, 2021. doi: 10.1103/PRXQuantum.2.020329.

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., Palizhati, A., Sriram, A., Wood, B., Yoon, J., Parikh, D., Zitnick, C. L., and Ulissi, Z. Open catalyst 2020 (OC20) dataset and community challenges. *ACS Catalysis*, 11 (10):6059–6072, 2021. doi: 10.1021/acscatal.0c04525.

Chen, C. and Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022. doi: 10.1038/s43588-022-00349-3.

Chen, J. and Chen, Z. Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771, 2008. doi: 10.1093/biomet/asn034.

Couzinié, Y., Daimon, S., Nishi, H., Ito, N., Harazono, Y., and Matsushita, Y.-i. Towards improved quantum machine learning for molecular force fields. *Physical Review A*, 112(6):062442, 2025. doi: 10.1103/5s9m-h7yb.

Dai, J. and Krems, R. V. Quantum gaussian process model of potential energy surface for a polyatomic molecule. *The Journal of Chemical Physics*, 156(18):184802, 2022. doi: 10.1063/5.0088821.

Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J., and Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023. doi: 10.1038/s42256-023-00716-3.

Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B*, 99(1): 014104, 2019. doi: 10.1103/PhysRevB.99.014104.

Farhi, E., Goldstone, J., and Gutmann, S. A quantum approximate optimization algorithm, 2014.

Fey, M. and Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. arXiv:1903.02428.

Feynman, R. P. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6–7):467–488, 1982. doi: 10.1007/BF02650179.

Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR)*, 2020.

Gasteiger, J., Becker, F., and Günnemann, S. GemNet: Universal directional graph neural networks for molecules. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 6790–6802, 2021.

Glick, J. R., Gujarati, T. P., Córcoles, A. D., Kim, Y., Kandala, A., Gambetta, J. M., and Temme, K. Covariant quantum kernels for data with group structure. *Nature Physics*, 20(3):479–483, 2024. doi: 10.1038/s41567-023-02340-9.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.

Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., and Gambetta, J. M. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019. doi: 10.1038/s41586-019-0980-2.

Himanen, L., Jäger, M. O. J., Morooka, E. V., Federici Canova, F., Ranawat, Y. S., Gao, D. Z., Rinke, P., and Foster, A. S. DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications*, 247:106949, 2020. doi: 10.1016/j.cpc.2019.106949.

Hohenberg, P. and Kohn, W. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, 1964. doi: 10.1103/PhysRev.136.B864.

Huang, H.-Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., and McClean, J. R. Power of data in quantum machine learning. *Nature Communications*, 12(1):2631, 2021. doi: 10.1038/s41467-021-22539-9.

Jain, A. and Bligaard, T. Atomic-position independent descriptor for machine learning of material properties. *Physical Review B*, 98(21):214112, 2018. doi: 10.1103/PhysRevB.98.214112.

Javadi-Abhari, A., Treinish, M., Krsulich, K., Wood, C. J., Lishman, J., Gacon, J., Martiel, S., Nation, P. D., Bishop, L. S., Cross, A. W., Johnson, B. R., and Gambetta, J. M. Quantum computing with Qiskit, 2024.

Jerbi, S., Fiderer, L. J., Poulsen Nautrup, H., Kübler, J. M., Briegel, H. J., and Dunjko, V. Quantum machine learning beyond kernel methods. *Nature Communications*, 14(1): 517, 2023. doi: 10.1038/s41467-023-36159-y.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. arXiv:1412.6980.

Kiss, O., Tacchino, F., Vallecorsa, S., and Tavernelli, I. Quantum neural networks force fields generation. *Machine Learning: Science and Technology*, 3(3):035004, 2022. doi: 10.1088/2632-2153/ac7d3c.

Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140 (4A):A1133–A1138, 1965. doi: 10.1103/PhysRev.140.A1133.

Lam, S. K., Pitrou, A., and Seibert, S. Numba: A LLVM-based Python JIT compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC (LLVM '15)*, pp. 1–6, 2015. doi: 10.1145/2833157.2833162.

Le, I. N. M., Kiss, O., Schuhmacher, J., Tavernelli, I., and Tacchino, F. Symmetry-invariant quantum machine learning force fields. *New Journal of Physics*, 27(2): 023015, 2025. doi: 10.1088/1367-2630/adad0c.

Li, W., Zhang, S.-X., Sheng, Z., Gong, C., Chen, J., and Shuai, Z. Quantum machine learning of molecular energies with hybrid quantum–neural wavefunction. *Digital Discovery*, 4(10):2697–2710, 2025. doi: 10.1039/D5DD00222B.

Liao, Y.-L. and Smidt, T. Equiformer: Equivariant graph attention transformer for 3D atomistic graphs. In *International Conference on Learning Representations (ICLR)*, 2023.

Liao, Y.-L., Wood, B., Das, A., and Smidt, T. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations (ICLR)*, 2024.

Lloyd, S. Universal quantum simulators. *Science*, 273(5278): 1073–1078, 1996. doi: 10.1126/science.273.5278.1073.

Lo Monaco, G., Bertini, M., Lorenzo, S., and Palma, G. M. Quantum extreme learning of molecular potential energy surfaces and force fields. *Machine Learning: Science and Technology*, 5(3):035014, 2024. doi: 10.1088/2632-2153/ad6120.

Mamun, O., Winther, K. T., Boes, J. R., and Bligaard, T. A bayesian framework for adsorption energy prediction on bimetallic alloy catalysts. *npj Computational Materials*, 6(1):177, 2020. doi: 10.1038/s41524-020-00447-8.

Manin, Y. I. *Vychislimoe i nevychislimoe [Computable and Noncomputable]*. Sovetskoye Radio, Moscow, 1980.

Musaelian, A., Batzner, S., Johansson, A., Sun, L., Owen, C. J., Kornbluth, M., and Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023. doi: 10.1038/s41467-023-36329-y.

Paine, A. E., Elfving, V. E., and Kyriienko, O. Quantum kernel methods for solving regression problems and differential equations. *Physical Review A*, 107(3):032428, 2023. doi: 10.1103/PhysRevA.107.032428.

Park, Y., Kim, J., Hwang, S., and Han, S. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 20(11):4857–4868, 2024. doi: 10.1021/acs.jctc.4c00190.

Passaro, S. and Zitnick, C. L. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27420–27438, 2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., and Latorre, J. I. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020. doi: 10.22331/q-2020-02-06-226.

Powell, M. J. D. A direct search optimization method that models the objective and constraint functions by linear interpolation. In Gomez, S. and Hennart, J.-P. (eds.), *Advances in Optimization and Numerical Analysis*, Mathematics and Its Applications, pp. 51–67. Kluwer Academic, Dordrecht, 1994. doi: 10.1007/978-94-015-8330-5_4.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006.

Rhodes, B., Vandenhaute, S., Šimkus, V., Gin, J., Godwin, J., Duignan, T., and Neumann, M. Orb-v3: Atomistic simulation at scale, 2025.

Schuld, M. and Killoran, N. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4): 040504, 2019. doi: 10.1103/PhysRevLett.122.040504.

Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A., and Müller, K.-R. SchNet – a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018. doi: 10.1063/1.5019779.

Shang, H., Fan, Y., Shen, L., Guo, C., Liu, J., Duan, X., Li, F., and Li, Z. Towards practical and massively parallel quantum computing emulation for quantum chemistry. *npj Quantum Information*, 9(1):33, 2023. doi: 10.1038/s41534-023-00696-7.

Thanasilp, S., Wang, S., Cerezo, M., and Holmes, Z. Exponential concentration in quantum kernel methods. *Nature Communications*, 15(1):5200, 2024. doi: 10.1038/s41467-024-49287-w.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, 2018.

Tran, R., Lan, J., Shuaibi, M., Wood, B. M., Goyal, S., Das, A., Heras-Domingo, J., Kolluru, A., Rizvi, A., Shoghi, N., Sriram, A., Therrien, F., Abed, J., Voznyy, O., Sargent, E. H., Ulissi, Z., and Zitnick, C. L. The open catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13(5):3066–3084, 2023. doi: 10.1021/acscatal.2c05426.

Usuga, A. F., Praveen, C. S., and Comas-Vives, A. Local descriptors-based machine learning model refined by cluster analysis for accurately predicting adsorption energies on bimetallic alloys. *Journal of Materials Chemistry A*, 12(5):2708–2721, 2024. doi: 10.1039/D3TA06316J.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Weaving, T., Ralli, A., Love, P. J., Succi, S., and Coveney, P. V. Contextual subspace variational quantum eigensolver calculation of the dissociation curve of molecular nitrogen on a superconducting quantum computer. *npj Quantum Information*, 11(1):25, 2025. doi: 10.1038/s41534-024-00952-4.

Williams, C. K. I. and Seeger, M. Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems 13 (NIPS 2000)*, pp. 682–688. MIT Press, 2001.

Wold, S., Sjöström, M., and Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130, 2001. doi: 10.1016/S0169-7439(01)00155-1.

Wood, B. M., Dzamba, M., Fu, X., Gao, M., Shuaibi, M., Barroso-Luque, L., Abdelmaqsoud, K., Gharakhanyan, V., Kitchin, J. R., Levine, D. S., Michel, K., Sriram, A., Cohen, T., Das, A., Rizvi, A., Sahoo, S. J., Ulissi, Z. W., and Zitnick, C. L. UMA: A family of universal models for atoms. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, 2025. arXiv:2506.23971.

Zhang, L., Han, J., Wang, H., Car, R., and E, W. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Physical Review Letters*, 120(14):143001, 2018. doi: 10.1103/PhysRevLett.120.143001.

Zhang, Y. and Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Computational Materials*, 4(1):25, 2018. doi: 10.1038/s41524-018-0081-z.

A. Implementation Details

This section centralises the settings that determine the released QBM execution path. Table 3 separates shared settings from the component-specific circuit choices, while Algorithm 1 defines descriptor construction, Algorithm 2 the complete fit, Algorithms 3 and 4 subset proposal and scoring, and Algorithm 5 split-wise prediction and evaluation. The corresponding source is available in the repository given under Source availability; mathematical details are separated into Appendices B, D, E.1, E.2 and E.5.

Table 3. Shared and component-specific hyperparameters for the four OC22 IS2RE-Total pipeline variants. Ablation-specific departures from these settings are stated below.

Setting	Value
Cut-off R_c (Å)	6.0
Chebyshev terms n_R, n_A, n_T	11
Atom-level Nyström landmarks M	1024
Adsorbate decay length ℓ_{ads} (Å)	3
PLS width p	16
KRR Nyström landmarks	200
mBIC sparsity weight γ	1.0
Ensemble size K_{ens}	16
Out-of-fold splits n_{fold}	3
Cardinality anchor scale ν_0	0.5
QAOA depth L_Q	1
Maximum COBYLA iterations	25
Global random seed	0
Fidelity map / qubit cap / base layers	ZZ / 6 / 3
Fidelity angle scale η_K	0.1
Circuit-parameter learning rate	0.005
Readout learning rate	0.05

Computational environment and data partitions. All recorded experiments were executed on an Intel(R) Xeon(R) 6520P CPU without GPU or CUDA computation. The released pipeline can select CUDA when available, but the reported runs used its CPU branch. The reported partitions and the train-only fitting boundary for all transforms are described in Appendix F.1. The currently released variant YAML files select a different input directory, so they are not a complete record of the IS2RE-Total runs.

Numerical environment. NumPy and Numba evaluate descriptors (Harris et al., 2020; Lam et al., 2015), and PyTorch-Geometric reads sharded LMDB records (Fey & Lenssen, 2019). Scikit-learn and SciPy supply standardisation, PLS, ridge regression, Nyström features and non-negative least squares (Pedregosa et al., 2011; Virtanen et al., 2020; Williams & Seeger, 2001). Qiskit constructs and optimises the QUBO/QAOA problem (Javadi-Abhari et al., 2024; Powell, 1994); the re-uploading model uses mini-batch Adam (Kingma & Ba, 2015). Fidelity and re-uploading states are stored as complex128 arrays, with float64 inputs and regression algebra.

Role and reporting scope of the quantum simulator. QBM uses the exact local state-vector class shown in Fig. 1b: the ideal amplitudes of each small gate-based circuit are represented in full on a CPU, without tensor-network compression or quantum-hardware execution. The route used to read those simulated states depends on the component. QAOA uses Qiskit Aer `SamplerV2`, a level-one preset pass manager with `rx`, `ry`, `rz`, `rzz`, `cx` and `h` basis gates, serial shot execution and seed 0. Because the code neither sets `default_shots` nor passes `shots` to the primitive, the installed default of 1,024 shots applies; `StatevectorSampler` is the fallback. Fidelity-kernel and re-uploading branches instead update complex amplitudes directly, without shot sampling: the former supplies exact overlaps to a classical ridge solve, and the latter supplies Pauli-Z expectations and adjoint gradients to a classical readout. Thus, finite-shot sampling is a QAOA readout choice within the state-vector workflow, not evidence of quantum-hardware execution. No OC22 electronic Hamiltonian is encoded, and no VQE ground-state energy is computed.

The QAOA seed controls both sampler random numbers and the initial point drawn uniformly from $[0, \pi]^{2L_Q}$; COBYLA is capped at 25 iterations. Nyström landmark sampling, three-fold shuffled out-of-fold splitting and circuit initialisation also use seed 0 in the canonical configurations. Direct state-vector branches are deterministic conditional on their inputs and initial parameters, whereas QAOA retains finite-shot variability. No noise model, hardware coupling map, calibration data or quantum-device timing is used. The results are therefore noiseless CPU simulations with a sampled QAOA proposer, not quantum-processor measurements or evidence of hardware speed-up.

A.1. Subset proposal and exact rescoring

Shared fitting protocol. Each atom was represented by 66 RDF, ADF and TDF coefficients obtained with $R_c = 6.0$ Å and 11 Chebyshev terms per distribution in both unweighted and pseudospin-weighted blocks. The species-to-pseudospin map, atom standardiser and 1,024-landmark RBF Nyström transform were learned from training data. Global and adsorbate-local sums were concatenated with $\ell_{ads} = 3$ Å to give 2,048 structure features. A composition `RidgeCV` baseline was subtracted before the pooled representation was standardised and projected, under supervision by the composition residual, to $p = 16$ PLS components. This is PLS, not PCA. Subset KRR used 200 landmarks drawn without replacement with seed 0 and selected its ridge penalty by GCV over $\{10^{-3}, 10^{-2}, 0.05, 0.1, 0.5, 1, 5, 20, 100\}$. Candidate subsets were ranked with $\gamma = 1$; the 16 highest-ranked subset identities were held fixed while three-fold shuffled out-of-fold refits (seed 0) generated predictions for simplex-

Algorithm 1. Training-time descriptor construction and pooling

Input: Training structures $\mathcal{D}_{\mathrm{tr}}$; R_{c}; $n_{\mathrm{R}}, n_{\mathrm{A}}, n_{\mathrm{T}}$; ℓ_{ads}; landmark count M
Output: Pooled features $\{\mathbf{\Phi}_s\}$ and fitted atom transform

derive the species-specific pseudospin map from $\mathcal{D}_{\mathrm{tr}}$
foreach $\mathcal{S}_s \in \mathcal{D}_{\mathrm{tr}}$ **do**
  construct minimum-image pairwise displacement tensors and apply the R_{c} mask
  compute the structural and pseudospin-weighted RDF, ADF and TDF coefficients for each atom (Appendices B.3 and B.5)
fit the atom standardiser and the M-landmark RBF Nyström map on the training-atom descriptors
foreach $\mathcal{S}_s \in \mathcal{D}_{\mathrm{tr}}$ **do**
  map its standardised atom descriptors to $\boldsymbol{\zeta}_{s,i}$
  form $\mathbf{\Phi}_s^{\mathrm{glob}} = \sum_i \boldsymbol{\zeta}_{s,i}$ and the adsorbate-weighted pool in equation (10)
  use the global pool for the second block when adsorbate labels are unavailable
  $\mathbf{\Phi}_s \leftarrow \mathbf{\Phi}_s^{\mathrm{glob}} \| \mathbf{\Phi}_s^{\mathrm{loc}}$
return $\{\mathbf{\Phi}_s\}$ and the fitted atom transform

Algorithm 2. Full-profile model fitting

Input: Training structures $\{\mathcal{S}_s\}$; pooled features $\{\mathbf{\Phi}_s\}$; targets $\{y_s\}$; variant v; p; K_{ens}; n_{fold}
Output: Fitted state Θ

fit the composition ridge baseline and set $\tilde{y}_s = y_s - \hat{E}_{\mathrm{comp}}(\mathcal{S}_s)$
fit the structure standardiser and the p-component PLS projection to obtain $\mathbf{X}$
$\mathcal{C} \leftarrow$ `GenerateCandidates` $(\mathbf{X}, \tilde{\mathbf{y}}, v)$ ▷ *algorithm 3*
$\mathcal{P} \leftarrow$ `ExactRescore` $(\mathcal{C}, \mathbf{X}, \tilde{\mathbf{y}}, v)$ ▷ *algorithm 4*
retain the top K_{ens} subset identities and refit only those fixed subsets to obtain $\mathbf{Y}^{\mathrm{oof}}$ over n_{fold} splits
fit simplex-constrained weights $\mathbf{w}$ by non-negative least squares on $(\mathbf{Y}^{\mathrm{oof}}, \tilde{\mathbf{y}})$
set $r_s = \tilde{y}_s - \sum_m w_m Y_{sm}^{\mathrm{oof}}$ and prepare the variant-specific residual input $\mathbf{u}_s$
fit the residual learner specified for v in Table 1
return $\Theta = (\mathrm{transforms}, \hat{E}_{\mathrm{comp}}, \mathcal{P}_K, \mathbf{w}, \mathcal{R})$

constrained non-negative least squares. Algorithms 1 and 2 give the shared workflow.

Variant-specific fitting and model selection. The *baseline* variant enumerates all non-empty subsets. The other selectors use depth-one QAOA with the relative anchor $\nu = \nu_0 \max(\mathrm{median}_i\,|a_i|, 10^{-8})$, where $\nu_0 = 0.5$, and an automatically generated grid of at most eight cardinalities from 3 to 16. COBYLA runs for at most 25 iterations, sampled bitstrings and the reported optimum are deduplicated with all singleton and pair subsets, and at most 500 proposals are retained by quadratic-surrogate energy before exact refitting. The *quantum_kernel* variant uses a ZZ fidelity kernel, at most six qubits, three base layers and default angle scale 0.1 for both subset regression and a 16-dimensional residual-targeted projection. The *quantum_reupload* variant uses six qubits, a 48-dimensional residual-targeted projection and $3\lceil 48/6 \rceil = 24$ layers; it trains for at most 200 epochs with batch size 256 and patience 5. The *baseline* and *quantum* variants use a linear-kernel Gaussian process residual, evaluated by the equivalent exact ridge solve above 2,000 training structures. The first three manuscript identifiers match the released code keys.

The released source specifies canonical QAOA initialisation, sampling, cardinality-grid construction, proposal deduplication and the 500-candidate cap. Computational repeats, artefact gaps and reduced-profile overrides are documented in Appendices F.3, F.4 and F.6. Because $\hat{\sigma}_0^2$ is defined within each candidate pool, absolute mBIC values are not comparable across variants.

Algorithm 3. Candidate subset proposal

Input: $\mathbf{X} \in \mathbb{R}^{n \times p}$; residual target $\tilde{\mathbf{y}}$; QBM variant v; ν_0; L_{Q}; seed; proposal budget
Output: Candidate set $\mathcal{C}$

if $v =$ baseline **then**
  return all non-empty subsets of $\{1, \ldots, p\}$
fit every singleton and pair with the subset regressor assigned to v and store them in $\mathcal{C}_{\leq 2}$
compute v_0, a_i and b_{ij} from their fitted log-variances (equation (D.6))
choose at most eight target cardinalities in $[3, d_{\max}]$, initialise $\mathcal{C}_{\geq 3} = \varnothing$ and set $\nu = \nu_0 \max(\mathrm{median}_i\,|a_i|, 10^{-8})$
foreach *target cardinality* d_{t} **do**
  assemble the anchored QUBO and its Ising Hamiltonian (equations (D.8), (D.11) and (D.13))
  optimise the depth-L_{Q} QAOA state with seeded COBYLA initialisation
  sample 1,024 shots with the seeded local sampler and add admissible size-≥ 3 bitstrings, including the reported optimum, to $\mathcal{C}_{\geq 3}$
deduplicate $\mathcal{C}_{\geq 3}$ and, if necessary, retain its proposal-budget subsets with the lowest unanchored score in equation (D.7); singleton and pair fits do not consume this budget
return $\mathcal{C}_{\leq 2} \cup \mathcal{C}_{\geq 3}$

Algorithm 4. Candidate refitting and mBIC rescoring

Input: Candidate set $\mathcal{C}$; $\mathbf{X}$; residual target $\tilde{\mathbf{y}}$; variant v
Output: Ranked fitted pool $\mathcal{P}$

foreach $\mathcal{M} \in \mathcal{C}$ **do** ***(in parallel)***
  fit the subset regressor assigned to v on $\mathbf{X}_{:,\mathcal{M}}$
  store $\hat{\sigma}_{\mathcal{M}}^2$ and the fitted model
set $\hat{\sigma}_0^2 = \min_{\mathcal{M} \in \mathcal{C}} \hat{\sigma}_{\mathcal{M}}^2$
compute $\mathrm{mBIC}(\mathcal{M})$ using equation (6), with $k_{\mathrm{S}} = |\mathcal{M}| + 1$ and the backend-reported k_{B}
return the fitted candidates sorted by increasing mBIC

Algorithm 5. Prediction and split-wise evaluation

Input: Evaluation split $\mathcal{D}$; fitted state Θ
Output: Point predictions and regression metrics
replay the stored atom transform, pooling, structure standardiser and PLS projection for every $\mathcal{S} \in \mathcal{D}$
evaluate the stored composition baseline and retained subset models
set $\hat{y}_{\text{ens}} = \sum_m w_m \hat{y}_m$ and evaluate the residual learner with `return_std=False`
$\hat{E} \leftarrow \hat{E}_{\text{comp}} + \hat{y}_{\text{ens}} + \hat{r}$
compute R^2, RMSE and MAE in eV per structure, MAE in meV/atom, Pearson r, and error mean and standard deviation
compute MAPE only when every target is non-zero and has the same sign
return point predictions and metrics

A.2. Prediction and evaluation

Prediction replays every fitted training transform without refitting. The implemented evaluation path requests point predictions only and combines the composition baseline, weighted subset ensemble and residual learner as in equation (12). It evaluates the training split in sample and the official ID and OOD validation splits. RMSE and MAE are first calculated in eV per structure; per-atom MAE is the mean of $|E - \hat{E}|/N$ multiplied by 1,000 to give meV/atom. The current evaluation code does not emit a combined predictive variance or calibrated interval, so no such output is claimed in Algorithm 5.

B. Pseudospin Descriptor Details

We give the descriptor definitions, invariance properties and closed-form Chebyshev coefficients used by the implementation. The RDF and ADF follow Artrith et al. (2017); the TDF is the extension introduced here. Structural and compositional blocks differ only through the species weight $\tau(Z)$.

B.1. Cosine cut-off

The cosine cut-off makes the representation continuously differentiable as neighbours enter or leave the local environment (Bartók et al., 2013):

$$f_{\text{c}}(r) = \begin{cases} \frac{1}{2}\left[\cos\left(\frac{\pi r}{R_{\text{c}}}\right) + 1\right], & r < R_{\text{c}}, \\ 0, & r \geq R_{\text{c}}. \end{cases} \tag{B.1}$$

Proposition B.1 (Regularity of the cut-off)**.** *f_{c} is continuously differentiable on $[0, \infty)$, with $f_{\text{c}}(0) = 1$, $f_{\text{c}}(R_{\text{c}}) = 0$ and $f'_{\text{c}}(0) = f'_{\text{c}}(R_{\text{c}}) = 0$, but it is not twice differentiable at R_{c}.*

Proof. For $r < R_{\text{c}}$, $f'_{\text{c}}(r) = -\pi \sin(\pi r/R_{\text{c}})/(2R_{\text{c}})$. Both the function and its first derivative match the zero branch at R_{c}, establishing C^1 continuity. The second derivative approaches $\pi^2/(2R_{\text{c}}^2)$ from the left but is zero on the constant branch, so $f_{\text{c}} \notin C^2$. □

The C^1 property prevents the cut-off term itself from introducing a first-derivative discontinuity at R_{c}. It does not establish continuous forces for the complete predictor, which is neither trained nor evaluated as a force model; the second derivative of this cut-off is discontinuous.

B.2. Atom-centred distribution functions

For the central atom i, each neighbour (pair, triple) within R_{c} contributes a Dirac delta at its bond length $R_{ij} = \|\mathbf{r}_j - \mathbf{r}_i\|$, bond angle ϑ_{ijk}, or dihedral angle ϕ_{ijkl}, weighted by the cut-off factors of the participating bonds and by a species weight $\tau(Z)$.

$$\text{RDF}_i(r) = \sum_{j \neq i} \tau(Z_j)\, f_{\text{c}}(R_{ij})\, \delta(r - R_{ij}), \tag{B.2}$$

$$\text{ADF}_i(\vartheta) = \sum_{j \neq i} \sum_{k \neq i,j} \tau(Z_j)\, \tau(Z_k)\, f_{\text{c}}(R_{ij})\, f_{\text{c}}(R_{ik}) \times \delta(\vartheta - \vartheta_{ijk}), \tag{B.3}$$

$$\text{TDF}_i(\phi) = \sum_{(j,k,l) \in \mathcal{T}_i} \tau(Z_j)\, \tau(Z_k)\, \tau(Z_l)\, f_{\text{c}}(R_{ij})\, f_{\text{c}}(R_{jk}) \times f_{\text{c}}(R_{kl})\, \delta(\phi - \phi_{ijkl}), \tag{B.4}$$

Here $\mathcal{T}_i$ contains bonded chains i–j–k–l whose three consecutive bonds lie inside R_{c}. The available record does not specify whether reverse chains or symmetry-equivalent ordered tuples are deduplicated, so the exact torsional enumeration cannot be reconstructed without the implementation.

Scalar geometric arguments. The distributions depend only on minimum-image distances, angle cosines and unsigned dihedral cosines. Degenerate chains with an undefined plane normal are excluded. Using $\cos\phi_{ijkl}$ rather than a signed dihedral preserves reflection invariance.

Proposition B.2 (Descriptor symmetries)**.** RDF_i, ADF_i *and* TDF_i *are invariant under* (a) *a global translation* $\mathbf{r}_m \mapsto \mathbf{r}_m + \mathbf{t}$, (b) *a global orthogonal transformation* $\mathbf{r}_m \mapsto \mathbf{R}\mathbf{r}_m$ *with* $\mathbf{R}^\top\mathbf{R} = \mathbf{I}$ *(rotations are the case* $\det\mathbf{R} = +1$*, reflections* $\det\mathbf{R} = -1$*), and* (c) *a relabelling* π *of atoms that preserves species,* $Z_{\pi(m)} = Z_m$.

Proof. Translations leave every difference vector unchanged. Orthogonal transformations preserve distances and inner products; both plane normals acquire the same factor $\det\mathbf{R}$, which cancels in their cosine. Relabelling like atoms only reorders finite sums and preserves $\tau(Z)$. The atomic descriptor family is therefore permutation equivariant, and the pooled structure representation is invariant. □

B.3. Derivation of the expansion coefficients

Each distribution is represented by its first $n_\bullet$ projection coefficients in a finite basis $\{\varphi_\alpha\}_{\alpha=0}^{n_\bullet-1}$. A finite smooth basis does not exactly reconstruct the Dirac-delta sum; it defines a truncated projection. To keep the derivations readable in two columns, write $\tau_j = \tau(Z_j)$ and $f_{ab} = f_c(R_{ab})$ below. For the radial distribution,

$$\Pi_{n_R}\mathrm{RDF}_i(r) \;=\; \sum_{\alpha=0}^{n_R-1} c_{i,\alpha}^{R}\,\varphi_\alpha(r), \tag{B.5}$$

and we equip the basis with a dual $\{\bar\varphi_\alpha\}$ on $(0, R_c)$ through the biorthogonality relation

$$\int_0^{R_c} \bar\varphi_\beta(r)\,\varphi_\alpha(r)\,\mathrm{d}r \;=\; \delta_{\alpha\beta}. \tag{B.6}$$

Proposition B.3 (Closed-form truncated-projection coefficients)**.** *For the distribution equation (B.2), the projection coefficients of equation (B.5) are the finite neighbour sums*

$$c_{i,\alpha}^{R} \;=\; \sum_{j\neq i} \tau_j f_{ij}\,\bar\varphi_\alpha(R_{ij}), \tag{B.7}$$

requiring no integral or histogram. The angular and torsional coefficients follow identically from equations (B.3) and (B.4).

Proof. Define the truncated coefficient by pairing the distribution with $\bar\varphi_\alpha$. Substitution of the finite Dirac-delta sum gives

$$c_{i,\alpha}^{R} = \sum_{j\neq i} \tau_j f_{ij} \int_0^{R_c} \bar\varphi_\alpha(r)\delta(r - R_{ij})\,\mathrm{d}r.$$

The sifting property yields equation (B.7). The angular and torsional cases are identical after replacing the scalar argument and summation set. □

The descriptor consists of these exact finite moments of a truncated representation, evaluated directly from the neighbour list; no claim of exact distribution reconstruction is made.

B.4. Chebyshev polynomials of the first kind

We take the dual basis to be the Chebyshev polynomials of the first kind, T_α, in a rescaled argument. They obey the three-term recurrence

$$\begin{aligned} &T_0(x) = 1, \quad T_1(x) = x, \\ &T_{\alpha+1}(x) = 2x\,T_\alpha(x) - T_{\alpha-1}(x). \end{aligned} \tag{B.8}$$

Lemma B.4 (Trigonometric form and orthogonality)**.** *For $x = \cos\theta$, $\theta \in [0, \pi]$, one has $T_\alpha(\cos\theta) = \cos(\alpha\theta)$. Consequently the T_α are orthogonal on $[-1, 1]$ for the weight $(1 - x^2)^{-1/2}$,*

$$\int_{-1}^{1} \frac{T_\alpha(x)T_{\alpha'}(x)}{\sqrt{1-x^2}}\,\mathrm{d}x = \begin{cases} \pi, & \alpha = \alpha' = 0, \\ \pi/2, & \alpha = \alpha' \neq 0, \\ 0, & \alpha \neq \alpha'. \end{cases} \tag{B.9}$$

Proof. The identity follows by induction from equation (B.8) and $2\cos\theta\cos(\alpha\theta) = \cos[(\alpha+1)\theta] + \cos[(\alpha-1)\theta]$. With $x = \cos\theta$, the weighted integral in equation (B.9) becomes $\int_0^\pi \cos(\alpha\theta)\cos(\alpha'\theta)\,\mathrm{d}\theta$, which has the stated values by Fourier orthogonality. □

For the radial expansion we map $r \in [0, R_c]$ to $x \in [-1, 1]$ by

$$\begin{aligned} x(r) &= \frac{2r}{R_c} - 1, \\ r(x) &= \frac{R_c}{2}(x+1), \qquad \frac{\mathrm{d}r}{\mathrm{d}x} = \frac{R_c}{2}. \end{aligned}$$

Thus $r = 0$ maps to $x = -1$, $r = R_c$ maps to $x = 1$, and every neighbour inside the cut-off is evaluated at $\bar R_{ij} = x(R_{ij}) = 2R_{ij}/R_c - 1$. We take the dual basis

$$\bar\varphi_\alpha(r) = T_\alpha\Big(\tfrac{2r}{R_c} - 1\Big). \tag{B.10}$$

The matching reconstruction basis carries the Chebyshev weight in equation (B.9). With $x(r) = 2r/R_c - 1$, exact biorthogonality requires

$$\varphi_\alpha(r) = \frac{2\,(2 - \delta_{\alpha0})}{\pi R_c\,\sqrt{1 - x(r)^2}}\,T_\alpha\big(x(r)\big), \tag{B.11}$$

which follows from Lemma B.4 and $\mathrm{d}r/\mathrm{d}x = R_c/2$. Angular and torsional projections use the plain angle measure because $T_\alpha(\cos\theta) = \cos(\alpha\theta)$. Only the dual basis enters the coefficients; the primal basis is used for reconstruction.

Inserting equation (B.10) into equation (B.7) reproduces the radial coefficient of the main text,

$$c_{i,\alpha}^{R} = \sum_{j\neq i} \tau_j\,T_\alpha\Big(\tfrac{2R_{ij}}{R_c} - 1\Big)\,f_{ij}. \tag{B.12}$$

The angular and torsional arguments are already cosines in $[-1, 1]$, so by Lemma B.4 we expand directly in $\cos\vartheta_{ijk}$ and $\cos\phi_{ijkl}$,

$$\begin{aligned} c_{i,\alpha}^{A} &= \sum_{j\neq i}\sum_{k\neq i,j} \tau_j\tau_k\,T_\alpha(\cos\vartheta_{ijk}) \\ &\quad\times f_{ij}f_{ik}, \end{aligned} \tag{B.13}$$

$$\begin{aligned} c_{i,\alpha}^{T} &= \sum_{\mathcal{T}_i} \tau_j\tau_k\tau_l\,T_\alpha(\cos\phi_{ijkl}) \\ &\quad\times f_{ij}f_{jk}f_{kl}. \end{aligned} \tag{B.14}$$

The recurrence equation (B.8) is evaluated once per neighbour, pair, or chain, so the whole coefficient vector costs $O(n_\bullet)$ per term, for $n_\bullet$ basis functions.

B.5. Structural/compositional weighting and pseudospins

Each distribution is evaluated with uniform and species-dependent weights. Uniform $P \equiv 1$ produces the structural block; Ising-style pseudospins produce the compositional block. For S species, the weights are symmetric about zero, omitting zero when S is even. Concatenating both blocks gives $\mathbf{f}_i \in \mathbb{R}^{2(n_\mathrm{R}+n_\mathrm{A}+n_\mathrm{T})}$, independent of S. This constant dimension motivates the representation for chemically diverse oxides (Artrith et al., 2017).

The species table is reported as sorted over the training species, but the supplied materials do not list the resulting numerical pseudospins or the behaviour for a species absent from that table. These missing conventions prevent exact descriptor reproduction and must be fixed by releasing the fitted species map or implementation.

C. Heuristic mBIC Score

We motivate the implemented subset score and explain why model weights are fitted to held-out predictions. The score ranks input-coordinate subsets fitted by KRR and, for the final re-uploading shortlist, by a variational circuit. It is neither exact Schwarz BIC for either backend nor a posterior over all model and kernel hyperparameters.

C.1. Gaussian residual likelihood

For subset $\mathcal{M}$, let $e_{s,\mathcal{M}} = \tilde{y}_s - \hat{f}_\mathcal{M}(\mathbf{x}_{s,\mathcal{M}})$ and $\mathrm{SSE}_\mathcal{M} = \sum_s e^2_{s,\mathcal{M}}$. Under conditionally Gaussian residuals (Mamun et al., 2020),

$$\ell_\mathcal{M}(\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\mathrm{SSE}_\mathcal{M}}{2\sigma^2}. \tag{C.1}$$

Maximisation gives

$$\hat{\sigma}^2_{\mathrm{ML},\mathcal{M}} = \frac{\mathrm{SSE}_\mathcal{M}}{n}, \tag{C.2}$$

and therefore

$$-2\ell_\mathcal{M}(\hat{\sigma}^2_{\mathrm{ML},\mathcal{M}}) = n\ln\hat{\sigma}^2_{\mathrm{ML},\mathcal{M}} + n\{1+\ln(2\pi)\}. \tag{C.3}$$

The second term is common to subsets fitted to the same targets.

C.2. From BIC motivation to the implemented heuristic

Dropping model-independent constants, Schwarz BIC is

$$\mathrm{BIC}(\mathcal{M}) \equiv n\ln\hat{\sigma}^2_{\mathrm{ML},\mathcal{M}} + k_\mathrm{B}\ln n, \tag{C.4}$$

with a likelihood-model parameter count k_B. The implementation makes two backend-specific substitutions. For classical and fidelity KRR it sets $k_\mathrm{B} = k_\mathrm{S} = |\mathcal{M}| + 1$, which counts selected PLS coordinates rather than the up to 200 fitted Nyström coefficients, and replaces the maximum-likelihood variance by the GCV variance in equation (5). For a variational re-uploading finalist it records $k_\mathrm{B} = N_\theta = 4qL + q + 1$ and uses the residual variance of the fitted circuit; the canonical full-profile refit uses all training rows and therefore scores that variance in sample. These choices define a practical cross-backend ranking rule, not Schwarz BIC. Any Bayesian interpretation or selection-consistency statement would require a separate derivation.

Let $\hat{\sigma}^2_0 = \min_{\mathcal{M}\in\mathcal{C}}\hat{\sigma}^2_\mathcal{M}$ within candidate pool $\mathcal{C}$. Subtracting $n\ln\hat{\sigma}^2_0$ and dividing by n gives

$$\mathrm{mBIC}_0(\mathcal{M}) = \ln\frac{\hat{\sigma}^2_\mathcal{M}}{\hat{\sigma}^2_0} + \frac{k_\mathrm{B}\ln n}{n}.$$

Because the reference variance and, for re-uploading, the variance estimator are pool- and backend-specific, absolute mBIC values cannot be compared across variants.

C.3. Large-model-space penalty

Searching p features introduces $\binom{p}{d}$ subsets of size d. Exact EBIC adds $2\gamma\ln\binom{p}{d}$ to unnormalised BIC (Bogdan et al., 2004; Chen & Chen, 2008). We instead use

$$\psi(k_\mathrm{S}; p) = k_\mathrm{S}\ln\frac{p}{k_\mathrm{S}}, \tag{C.5}$$

with $k_\mathrm{S} = d+1$. This surrogate retains the leading $d\ln(p/d)$ behaviour while simplifying the QUBO cardinality term. It is not monotone in subset size. For KRR candidates, monotone size control comes from $k_\mathrm{B}\ln n/n$ because $k_\mathrm{B} = k_\mathrm{S}$; re-uploading instead uses its circuit parameter count. For the saturated subset $d = p$, ψ becomes slightly negative because $k_\mathrm{S} = p + 1$. At $p = 16$, $\gamma = 1$ and $n = 45{,}890$, the resulting advantage is approximately 4.5×10^{-5} in mBIC. The implemented score is

$$\mathrm{mBIC}(\mathcal{M}) = \ln\frac{\hat{\sigma}^2_\mathcal{M}}{\hat{\sigma}^2_0} + \frac{k_\mathrm{B}\ln n}{n} + \frac{2\gamma}{n}k_\mathrm{S}\ln\frac{p}{k_\mathrm{S}}. \tag{C.6}$$

We accordingly describe the third term as EBIC-style and do not transfer the consistency theorem of the exact EBIC to this surrogate.

C.4. Combination weights

BIC posterior approximations assign

$$\Pr(\mathcal{M}\mid\mathcal{D}) \propto \exp\left[-\frac{1}{2}\mathrm{IC}(\mathcal{M})\right]\Pr(\mathcal{M}). \tag{C.7}$$

Because mBIC is normalised per sample, the corresponding weights require the total scale,

$$w_m^{\mathrm{BIC}} = \frac{\exp(-n\,\Delta\mathrm{mBIC}_m/2)}{\sum_j \exp(-n\,\Delta\mathrm{mBIC}_j/2)}, \qquad \Delta\mathrm{mBIC}_m = \mathrm{mBIC}_m - \min_{m'} \mathrm{mBIC}_{m'}. \tag{C.8}$$

At $n = 45{,}890$, these weights concentrate on one model; applying the same formula to per-sample gaps instead gives an almost uniform average. Neither provides useful data-adaptive combination.

We therefore estimate weights from conditional fold-excluding subset-refit predictions; upstream target-supervised preprocessing and subset identities still use the complete training set. The simplex-constrained least-squares problem in equation (7) is solved by non-negative least squares with an augmented sum-to-one constraint. Both this solution and BIC-derived weights are evaluated on the same held-out prediction matrix. The ensemble prediction is

$$\hat{y}_{\mathrm{ens}}(\mathbf{x}) = \sum_m w_m \hat{f}_m(\mathbf{x}_{\mathcal{M}_m}),$$

which defines the residual target in Section 4.6.

D. Quantum Subset Selection Details

Subset selection occurs after composition removal and the training-only projection of the 2,048 pooled structure features to $p = 16$ PLS scores, and before ensemble weighting and residual learning (Algorithms 2 and 3). Figure 2b locates this hand-off within the full pipeline, while Fig. 2i,ii reports its observed candidate-count and CPU-time consequences. Its unit is therefore a *PLS component*, not an atom, structure, raw descriptor, principal component or training sample. The statistical problem remains prediction from all available training structures. QAOA changes only the computational candidate pool used to search the $2^{16} - 1$ non-empty component subsets; it is not the predictive mechanism and does not subsample the training set in the canonical configuration.

The selection stage is needed because exhaustive refitting scales with the power set. The *baseline* variant evaluates all 65,535 non-empty subsets, whereas the QAOA branches use fitted singleton and pair behaviour to construct a quadratic approximation and then propose a smaller pool for classical refitting. The matched predictive outcome is reported separately in Fig. 2iii and Table 2, preventing search reduction from being read as an accuracy claim. This primarily addresses subset-search cost under the simulated quantum-resource constraint; it neither reduces the original sample size nor establishes an advantage over budget-matched classical heuristics. We next derive the implemented QUBO surrogate, show how its coefficients are assembled and state the Ising mapping used by QAOA; the numerical settings are collected in Table 3.

D.1. Objective and binary encoding

Write a candidate feature subset $\mathcal{M} \subseteq \{1, \ldots, p\}$ as a binary indicator $\mathbf{z} \in \{0,1\}^p$ with $z_i = 1$ iff feature i is selected, and $|\mathbf{z}| = \sum_i z_i$. The selection objective is the exact mBIC of equation (6). Its complexity terms depend on $\mathbf{z}$ only through $|\mathbf{z}|$, but the data-fit term is the log-variance of the fitted model,

$$v(\mathbf{z}) \;=\; \ln \hat{\sigma}^2(\mathbf{z}),$$

which is neither separable nor quadratic in $\mathbf{z}$. We replace it with a quadratic surrogate $\hat{v}(\mathbf{z})$ that is *exact* on all subsets of size ≤ 2.

The exact pool score can be written directly in binary variables. Let $d(\mathbf{z}) = \sum_i z_i$ be the selected-feature count and let $\hat{\sigma}_0^2$ be the best residual variance in the candidate pool used for exact rescoring. In the QUBO proposal stages the assigned proxy is KRR, so $k_{\mathrm{B}}(\mathbf{z}) = k_{\mathrm{S}}(\mathbf{z}) = d(\mathbf{z}) + 1$ in equation (6). The variational parameter count is introduced only after the *quantum_reupload* finalists are refitted and does not enter the QUBO. Up to constants independent of $\mathbf{z}$,

$$\begin{aligned}\mathrm{mBIC}_{\mathrm{KRR}}(\mathbf{z}) &= \ln \hat{\sigma}^2(\mathbf{z}) - \ln \hat{\sigma}_0^2 + \frac{k_{\mathrm{S}}(\mathbf{z}) \ln n}{n} \\ &\quad + \frac{2\gamma}{n}\, k_{\mathrm{S}}(\mathbf{z}) \ln \frac{p}{k_{\mathrm{S}}(\mathbf{z})}, \qquad k_{\mathrm{S}}(\mathbf{z}) = d(\mathbf{z}) + 1, \\ &\equiv v(\mathbf{z}) + \Pi\big(d(\mathbf{z})\big) + \mathrm{const}.\end{aligned}$$

The constant $-\ln \hat{\sigma}_0^2$ is irrelevant for optimisation and for QAOA phases. The non-quadratic pieces are therefore isolated. The term $v(\mathbf{z})$ depends on a refit, and $\Pi(d)$ contains $\ln[p/k_{\mathrm{S}}(d)]$. The QUBO below is obtained by replacing the former with a second-order finite-difference surrogate and the latter with a target-cardinality linearisation.

D.2. Second-order log-variance surrogate

We approximate v by the unique quadratic that *interpolates* it on every subset of size ≤ 2. Anchor at the intercept-only model $v_0 = \ln \mathrm{Var}(\tilde{\mathbf{y}})$, let $\mathbf{e}_i$ be the i-th unit indicator, and posit the general quadratic

$$\hat{v}(\mathbf{z}) = c_0 + \sum_i c_i z_i + \sum_{i<j} c_{ij}\, z_i z_j. \tag{D.1}$$

Evaluating equation (D.1) on the empty, singleton, and pair indicators and matching v fixes the coefficients one order at a time.

$$\hat{v}(\mathbf{0}) = c_0 = v_0, \tag{D.2}$$

$$\hat{v}(\mathbf{e}_i) = c_0 + c_i = v(\mathbf{e}_i), \quad c_i = v(\mathbf{e}_i) - v_0, \tag{D.3}$$

$$\begin{aligned}\hat{v}(\mathbf{e}_i{+}\mathbf{e}_j) &= c_0 + c_i + c_j + c_{ij} \\ &= v(\mathbf{e}_i{+}\mathbf{e}_j).\end{aligned} \tag{D.4}$$

and hence

$$c_{ij} = v(\mathbf{e}_i{+}\mathbf{e}_j) - v(\mathbf{e}_i) - v(\mathbf{e}_j) + v_0.$$

Writing $a_i \equiv c_i$ and $b_{ij} \equiv c_{ij}$, these are the first- and second-order finite differences of v on the Boolean lattice,

$$a_i = v(\mathbf{e}_i) - v_0, \tag{D.5}$$

$$b_{ij} = v(\mathbf{e}_i+\mathbf{e}_j) - v(\mathbf{e}_i) - v(\mathbf{e}_j) + v_0, \tag{D.6}$$

so that a_i is the marginal log-variance change of feature i and b_{ij} is the non-additive interaction of i and j. Substituting back gives the truncated Möbius / inclusion–exclusion surrogate

$$\hat{v}(\mathbf{z}) = v_0 + \sum_i a_i z_i + \sum_{i<j} b_{ij}\, z_i z_j. \tag{D.7}$$

Proposition D.1 (Exactness and residual)**.** *$\hat{v}$ agrees with v on every subset of size ≤ 2. For general $\mathbf{z}$ the error is the sum of all order-≥ 3 interaction terms of v supported on $\mathbf{z}$, so $\hat{v} = v$ exactly iff v has no interactions of order ≥ 3.*

Proof. The equalities in equations (D.2) to (D.4) establish exact interpolation for subsets of size at most two. The Boolean-lattice Möbius expansion decomposes v into interaction terms indexed by subsets. Equation (D.7) retains orders zero to two, so its residual is precisely the sum of order-three and higher terms. □

The coefficients equations (D.5) and (D.6) are read off from the Stage-1 singleton and pair fits, so construction of the quadratic surrogate costs $O(p^2)$ exact KRR fits rather than the 2^p fits of a complete sweep. This comparison excludes the QAOA optimisation and sampling costs.

D.3. Complexity terms and the cardinality anchor

Complexity term. The mBIC complexity penalty in equation (6) depends on $\mathbf{z}$ only through the cardinality $d = |\mathbf{z}|$, since $k_{\mathrm{S}} = d + 1$. Splitting off the intercept,

$$\begin{aligned}\Pi(d) &= \frac{(d+1)\ln n}{n} + \frac{2\gamma}{n}(d+1)\ln\frac{p}{d+1} \\ &= \frac{\ln n}{n} + \frac{2\gamma}{n}\ln\frac{p}{d+1} + d\left[\frac{\ln n}{n} + \frac{2\gamma}{n}\ln\frac{p}{d+1}\right].\end{aligned}$$

To obtain a term *linear* in $\mathbf{z}$ we freeze the bracket at the target cardinality d_{t}, defining

$$c_{\mathrm{t}} = \frac{\ln n}{n} + \frac{2\gamma}{n}\,\max\left(\ln\frac{p}{d_{\mathrm{t}}},\, 0\right), \tag{D.8}$$

so that $c_{\mathrm{t}}\sum_i z_i = c_{\mathrm{t}} d$ reproduces the d-dependent part of $\Pi(d)$ up to the $O(1/n)$ shift between d_{t} and $d_{\mathrm{t}}+1$ inside the logarithm, and approximates it for nearby sizes. The $\max(\cdot, 0)$ guard keeps the per-bit cost non-negative for any target size; on the grid used here it never binds, because the grid stops at $d_{\mathrm{t}} = p$. Both the shift and the size-dependent constants are removed when the pool is rescored exactly in Stage 3.

Cardinality anchor. The configuration stores a dimensionless relative scale ν_0, not the absolute QUBO coefficient. After the singleton fits, the code computes

$$\nu = \nu_0 \max\left\{\mathrm{median}_i\,|a_i|, 10^{-8}\right\}, \qquad \nu_0 = 0.5, \tag{D.9}$$

once and reuses this value for every target cardinality. The floor prevents a vanishing anchor when singleton effects are numerically negligible. We then add ν times a quadratic anchor to keep proposals near d_{t}, so that the sampled pool spans several cardinalities. Because $z_i \in \{0, 1\}$ implies $z_i^2 = z_i$,

$$\begin{aligned}\mathcal{A}_{\mathrm{t}}(\mathbf{z}) &\equiv \Big(\sum_i z_i - d_{\mathrm{t}}\Big)^2 \\ &= \sum_i z_i^2 + 2\sum_{i<j} z_i z_j - 2d_{\mathrm{t}}\sum_i z_i + d_{\mathrm{t}}^2 \\ &= \sum_i (1 - 2d_{\mathrm{t}}) z_i + 2\sum_{i<j} z_i z_j + d_{\mathrm{t}}^2.\end{aligned} \tag{D.10}$$

Assembly. Summing the surrogate equation (D.7), the complexity term $c_{\mathrm{t}}\sum_i z_i$, and ν times equation (D.10), and discarding the additive constants v_0 and νd_{t}^2, the coefficient of z_i is $a_i + c_{\mathrm{t}} + \nu(1 - 2d_{\mathrm{t}})$ and that of $z_i z_j$ is $b_{ij} + 2\nu$. Thus $\mathcal{Q}(\mathbf{z}) = \sum_i Q_{ii} z_i + \sum_{i<j} Q_{ij} z_i z_j$ with

$$\begin{aligned}Q_{ii} &= a_i + c_{\mathrm{t}} + \nu(1 - 2d_{\mathrm{t}}), \\ Q_{ij} &= b_{ij} + 2\nu \qquad (i < j).\end{aligned} \tag{D.11}$$

which is the upper-triangular matrix used in the implementation.

Thus, the QUBO diagonal combines singleton fit, local cardinality cost and the linear anchor term. Off-diagonal entries combine pairwise non-additivity with the quadratic anchor.

D.4. Ising form and QAOA

The QUBO is mapped to a spin Hamiltonian by the usual change of variables.

Proposition D.2 (Ising coefficients)**.** *Under $z_i = (1 - s_i)/2$, $s_i \in \{-1, +1\}$, the QUBO energy equals the Ising energy*

$$\begin{aligned}\mathcal{Q}(\mathbf{z}) &= \sum_i Q_{ii} z_i + \sum_{i<j} Q_{ij} z_i z_j \\ &= \sum_i h_i s_i + \sum_{i<j} J_{ij} s_i s_j + \mathrm{const},\end{aligned}$$

with

$$h_i = -\tfrac{1}{2} Q_{ii} - \tfrac{1}{4}\sum_{j\neq i} \bar{Q}_{ij}, \qquad J_{ij} = \tfrac{1}{4}\, Q_{ij}. \tag{D.12}$$

where $\bar{Q}_{ij}$ is the symmetric off-diagonal (Q_{ij} if $i < j$, else Q_{ji}).

Proof. Substitution of $z_i = (1 - s_i)/2$ and $z_i z_j = (1 - s_i - s_j + s_i s_j)/4$ gives the stated linear and quadratic coefficients. The remaining spin-independent term is $\frac{1}{2}\sum_i Q_{ii} + \frac{1}{4}\sum_{i<j} Q_{ij}$. □

Promoting s_i to Pauli Z_i gives

$$\hat{H}_{\mathrm{C}} = \sum_i h_i \hat{Z}_i + \sum_{i<j} J_{ij} \hat{Z}_i \hat{Z}_j + \mathrm{const}, \tag{D.13}$$

whose computational-basis eigenvalues equal the QUBO energies. QAOA alternates this cost Hamiltonian with $\hat{H}_{\mathrm{B}} = \sum_i \hat{X}_i$,

$$|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle = \prod_{\ell=1}^{L_{\mathrm{Q}}} \mathrm{e}^{-\mathrm{i}\beta_\ell \hat{H}_{\mathrm{B}}}\, \mathrm{e}^{-\mathrm{i}\gamma_\ell \hat{H}_{\mathrm{C}}} |+\rangle^{\otimes p}. \tag{D.14}$$

Measurement in the computational basis assigns $p_{\boldsymbol{\gamma},\boldsymbol{\beta}}(\mathbf{z}) = |\langle \mathbf{z} \mid \boldsymbol{\gamma}, \boldsymbol{\beta}\rangle|^2$. COBYLA minimises the sampled QUBO expectation for at most 25 iterations. The depth-one initial point and sampler are both seeded with 0 in the canonical configuration. Aer `SamplerV2` is instantiated without an explicit shot argument, so its 1,024-shot default supplies the returned bitstrings; the state-vector sampler is a fallback with the same default number of samples. The code converts every returned vector by selecting indices with bit value greater than 0.5, filters to $0 < d \le d_{\max}$, and admits the resulting subset without probability weighting. It also admits the optimiser-reported solution. This makes proposal stochastic in principle but reproducible conditional on software versions and seed; deduplication occurs before exact rescoring.

D.5. Three-stage selection

Stage 1 fits every singleton and pair with the regression backend assigned to the QBM variant; these fits define the surrogate and remain eligible ensemble candidates. Stage 2 solves one QUBO for each of at most eight automatically spaced target cardinalities between 3 and p. It deduplicates admissible size-≥ 3 samples across targets and, only when this sampled pool exceeds 500 subsets, retains the 500 lowest values of the unanchored surrogate in equation (D.7). Singleton and pair fits remain outside that cap. Stage 3 refits the retained sampled proposals and recomputes their backend-specific variance and pool-relative mBIC together with the Stage-1 fits. The *quantum* variant uses RBF Nyström KRR; *quantum_kernel* uses the fidelity-kernel Nyström regressor. The *quantum_reupload* variant first ranks this pool with the configured ZZ fidelity proxy and then refits the top 64 proposals with the variational circuit, whose mBIC uses $k_{\mathrm{B}} = N_\theta$. All selectors receive the same $n \times p$ PLS matrix and target, but their fitted-model budgets differ: exhaustive enumeration covers every non-empty subset, whereas the quantum branches fit an $O(p^2)$ basis plus sampled proposals. This accounting excludes QAOA optimisation, finite-shot sampling and subsequent classical refits, and no budget-matched classical proposal heuristic was evaluated.

E. Quantum Kernel Details

The *quantum_kernel* variant fixes a circuit-defined feature map and uses its fidelity matrix in classical KRR; *quantum_reupload* instead trains circuit and readout parameters. Figure 2d,e shows the implemented Z and ZZ fidelity maps, while Fig. 2f,g uses implicit and explicit solely for the interleaved and separated re-uploading layouts. They share direct complex128 amplitude primitives but optimise different objects. The canonical *quantum_kernel* configuration uses ZZ; the Z map appears in the scale and geometry ablation of Fig. 3a–c. The trainable and frozen re-uploading controls are compared in Fig. 4c.

E.1. Implemented Z and ZZ encodings

Each candidate subset is standardised independently with statistics fitted on its training rows. For a standardised vector $\mathbf{x} \in \mathbb{R}^d$, the feature-map geometry is

$$q = \min(d, q_{\max}), \qquad L_{\mathrm{K}} = \max(L_0, \lceil d/q \rceil) \tag{E.1}$$

qubits and layers. Layer ℓ assigns qubit i to

$$\iota(\ell, i) = (\ell q + i) \bmod d, \tag{E.2}$$

so every coordinate is visited at least once and later slots wrap cyclically. The canonical cap is $q_{\max} = 6$ and $L_0 = 3$; qubit count is therefore determined by the selected feature dimension rather than by the 16-variable QUBO register. Define

$$\omega_{\ell i} = \eta_{\mathrm{K}} x_{\iota(\ell,i)}, \qquad P(\varphi) = \mathrm{diag}(1, \mathrm{e}^{\mathrm{i}\varphi}).$$

Starting from $|0\rangle^{\otimes q}$, *every* layer applies a Hadamard wall followed by the phase gates $P(2\omega_{\ell i})$ shown in Fig. 2d,e. The implemented Z state is

$$|\phi_Z(\mathbf{x})\rangle = \left[\prod_{\ell=0}^{L_{\mathrm{K}}-1} \left(\bigotimes_{i=0}^{q-1} P(2\omega_{\ell i})\right) H^{\otimes q}\right] |0\rangle^{\otimes q}. \tag{E.3}$$

Here $P(2\omega) = \mathrm{e}^{\mathrm{i}\omega} R_Z(2\omega)$, so this figure-primary notation is identical to the implemented R_Z update up to a global phase. No two-qubit gate is applied in the Z map.

For the ZZ map, define the pair phase

$$\phi_{\ell,ct}(\mathbf{x}) = (\pi - \omega_{\ell,c})(\pi - \omega_{\ell,t}) \tag{E.4}$$

and, following Fig. 2e,

$$W_{\mathcal{E}}(\mathbf{x}, \ell) = \prod_{(c,t)\in\mathcal{E}} \mathrm{CNOT}_{ct}\; P_t(2\phi_{\ell,ct}(\mathbf{x}))\, \mathrm{CNOT}_{ct}\,. \tag{E.5}$$

Each pair factor equals $e^{i\phi_{\ell,ct}} R_{ZZ}(2\phi_{\ell,ct})$, so it reproduces the implemented R_{ZZ} update up to a global phase. The edge set is empty for one qubit, contains $(0,1)$ for two qubits and is the circular ring $\{(i, i+1 \bmod q)\}$ for $q \geq 3$. Thus,

$$|\phi_{ZZ}(\mathbf{x})\rangle = \left[\prod_{\ell=0}^{L_K-1} W_{\mathcal{E}}(\mathbf{x},\ell)\left(\bigotimes_{i=0}^{q-1} P(2\omega_{\ell i})\right) H^{\otimes q}\right]|0\rangle^{\otimes q}. \tag{E.6}$$

This map follows the data-dependent second-order phase construction of Havlíček et al. (2019), restricted to the implemented ring topology. Its pair phases depend jointly on neighbouring encoded coordinates, so its induced similarity differs from the product Z map. The repeated Hadamard walls in equations (E.3) and (E.6) are essential implementation details; an initial-only Hadamard construction would be a different circuit.

E.2. Fidelity kernel and positive semidefiniteness

For $F \in \{Z, ZZ\}$, let $|\phi_F(\mathbf{x})\rangle$ be the state in equations (E.3) and (E.6) and $\rho_F(\mathbf{x}) = |\phi_F(\mathbf{x})\rangle\langle\phi_F(\mathbf{x})|$. The implemented fidelity kernel is equivalently a Hilbert–Schmidt inner product,

$$\begin{aligned}\kappa_F(\mathbf{x},\mathbf{x}') &= \left|\langle 0^{\otimes q}|U_F^\dagger(\mathbf{x}')U_F(\mathbf{x})|0^{\otimes q}\rangle\right|^2 \\ &= |\langle\phi_F(\mathbf{x}') \mid \phi_F(\mathbf{x})\rangle|^2 = \mathrm{Tr}[\rho_F(\mathbf{x})\rho_F(\mathbf{x}')]\,.\end{aligned} \tag{E.7}$$

Proposition E.1 (Validity of the fidelity kernel)**.** *For either implemented unitary feature encoding, κ_F is symmetric and positive semidefinite, satisfies $0 \leq \kappa_F \leq 1$, and has unit diagonal.*

Proof. For coefficients c_u,

$$\sum_{u,u'} \overline{c_u} c_{u'} \kappa_F(\mathbf{x}_u, \mathbf{x}_{u'}) = \left\| \sum_u c_u \rho_F(\mathbf{x}_u) \right\|_{\mathrm{HS}}^2 \geq 0.$$

Symmetry follows from conjugating the state overlap. Unitarity normalises each state, and Cauchy–Schwarz gives the unit upper bound. □

E.3. Exact state-vector evaluation

Let Ψ contain one simulated feature state per row. The full fidelity Gram matrix is

$$\mathbf{K} = (\Psi\Psi^\dagger) \odot \overline{(\Psi\Psi^\dagger)}, \qquad K_{ab} = \kappa_F(\mathbf{x}_a, \mathbf{x}_b). \tag{E.8}$$

The implementation applies the Hadamard, R_Z and diagonal R_{ZZ} operations directly to batched amplitudes in Qiskit's little-endian basis. These updates are algebraically identical to per-circuit state-vector evolution. Equation (E.8) therefore differs only by floating-point roundoff while avoiding one simulation per pair.

No overlap is estimated from finite shots in this branch: the 1,024-shot setting applies only to QAOA. For the configured large-sample path, 200 standardised training rows are drawn without replacement with seed 0 as landmarks. Write the symmetrised landmark Gram matrix as $\mathbf{K}_{MM,F} = \mathbf{V}\,\mathrm{diag}(\boldsymbol{\lambda})\mathbf{V}^\top$. The code retains r eigenvalues satisfying $\lambda_j > 10^{-8}\max_k \lambda_k$ and forms

$$\mathbf{W}_F = \mathbf{V}_r\,\mathrm{diag}(\lambda_1^{-1/2}, \ldots, \lambda_r^{-1/2}), \qquad \mathbf{F}_F = \mathbf{K}_{XM,F}\mathbf{W}_F. \tag{E.9}$$

Thus, $\boldsymbol{\zeta}_F(\mathbf{x}) = \boldsymbol{\kappa}_{M,F}(\mathbf{x})\mathbf{W}_F$, and the classical ridge stage solves

$$(\mathbf{F}_F^\top\mathbf{F}_F + \lambda_F\mathbf{I})\mathbf{a}_F = \mathbf{F}_F^\top(\mathbf{y} - \bar{y}\mathbf{1}). \tag{E.10}$$

Ridge strength is selected by the same GCV grid as the RBF subsets. No Gram-matrix centring or additional kernel normalisation is applied; unit diagonal follows from state normalisation, while eigentruncation and ridge regularisation provide numerical stabilisation. At validation time, the stored subset scaler is applied first, a cross-kernel to the stored landmarks is evaluated, and the stored $\mathbf{W}_F$, ridge coefficients and intercept produce the point prediction. The subsequent ridge fit costs $O(nM^2)$ and approximates the full dual kernel solution.

E.4. The kernel as a linear quantum model

Following Jerbi et al. (2023), the density operator acts as a feature map and the fitted observable is

$$O_{\boldsymbol{\beta},F} = \sum_m \beta_m \rho_F(\mathbf{x}_m).$$

The KRR predictor then satisfies

$$\hat{f}(\mathbf{x}) = \sum_m \beta_m \kappa_F(\mathbf{x}, \mathbf{x}_m) = \mathrm{Tr}[\rho_F(\mathbf{x})O_{\boldsymbol{\beta},F}].$$

Thus, fidelity-kernel ridge is a fixed-feature linear quantum model. The representer theorem makes this expansion exact within the kernel hypothesis space, but does not imply superior generalisation. This distinction motivates the matched classical comparisons.

E.5. Variational data-re-uploading residual model

In the canonical full-profile pipeline, *quantum_reupload* trains on residuals from conditional fold-excluding subset refits

$$r_s = \tilde{y}_s - \sum_m w_m Y_{sm}^{\mathrm{oof}}, \qquad \mathbf{u}_s = \mathrm{std}(\Pi_{\mathrm{R}}\boldsymbol{\Phi}_s).$$

Here the configured Π_{R} is a residual-targeted 48-component PLS projection of the pooled descriptor. Its standardiser and PLS weights are fitted only on training residual inputs.

With $q = 6$ and three uploads per feature block, the code sets $L = 3\lceil 48/6 \rceil = 24$.

To match the symbols in Fig. 2f,g, write the standardised residual input locally as $\mathbf{x} \equiv \mathbf{u}$ and define the data block $U_\phi^{(\ell)}(\mathbf{x}) \equiv D_\ell(\mathbf{u})$. Qubit i receives $\omega_{\ell i}(\mathbf{x}) = \eta_{\ell i} x_{\iota(\ell,i)} + \chi_{\ell i}$. The executed implicit/interleaved layout is the combined block

$$|\psi_{\ell+1}\rangle = V^{(\ell)}(\mathbf{x}, \boldsymbol{\theta}^{(\ell)})|\psi_\ell\rangle,$$
$$V^{(\ell)}(\mathbf{x}, \boldsymbol{\theta}^{(\ell)}) \equiv W_{\mathcal{E}}^{\mathrm{CZ}} V^{(\ell)}(\boldsymbol{\theta}^{(\ell)}) U_\phi^{(\ell)}(\mathbf{x}). \tag{E.11}$$

The fixed ring is not inert here because interleaved R_Y rotations do not commute with it. Figure 2g uses *explicit* only for the alternative drawing in which $U_\phi^{(\ell)}$ and $V^{(\ell)}$ are shown as separate blocks; these layout terms are not model-performance labels.

An affine Pauli-Z readout maps the bounded expectations to the residual scale,

$$\hat{r}_{\boldsymbol{\theta},\mathbf{g},b}(\mathbf{x}) = \sum_{i=0}^{q-1} g_i \,\mathrm{Tr}[\rho_{\boldsymbol{\theta}}(\mathbf{x}) Z_i] + b. \tag{E.12}$$

With $E = qL$ gate slots, the flat trainable vector contains E encoding scales, E encoding offsets, E variational R_Y angles, E variational R_Z angles, q readout weights and one bias, giving $N_\theta = 4qL + q + 1$. This is the parameter count passed to the BIC-style term when re-uploading finalists are rescored.

The frozen-encoding control fixes this same circuit at its seeded initial parameter vector $\boldsymbol{\theta}_0$ and forms

$$\kappa_{\mathrm{RU}}^{\mathrm{frz}}(\mathbf{x}, \mathbf{x}') = \left|\langle \psi(\mathbf{x}; \boldsymbol{\theta}_0) \mid \psi(\mathbf{x}'; \boldsymbol{\theta}_0)\rangle\right|^2, \tag{E.13}$$

which then enters the Nyström ridge path in equations (E.9) and (E.10). By contrast, the trained-circuit model jointly optimises the circuit parameters and Pauli-Z readout. Mini-batch Adam minimises mean-squared residual error with optional L_2 penalty; the canonical setting uses no L_2 penalty. The direct amplitude simulator evaluates adjoint gradients, with learning rates 0.005 for circuit parameters and 0.05 for the readout. Early stopping does not use a validation set: an epoch stalls when its training MSE improves by less than a relative 10^{-4}, and optimisation stops after five consecutive stalled epochs.

At prediction, the residual mean is unstandardised. The regressor can form a linearised-Laplace covariance when requested, but the training pipeline calls it with `return_std=False`; the reported evaluation therefore uses only its point mean. Unlike the fixed kernel, re-uploading jointly optimises the encoding and readout, so the two variants test different hypotheses.

F. Experiment Details

F.1. Dataset and partitions

OC22 contains periodic PBE+U DFT relaxations for oxide electrocatalysts and defines total-energy prediction tasks beyond adsorption-energy learning (Tran et al., 2023). The aggregate record, without an archived sample manifest, reports IS2RE-Total using the training, in-domain validation (val_id) and out-of-domain validation (val_ood) partitions, without a further random split. Table 4 records their distinct roles and structure counts, making explicit that the two validation sets are retained as released rather than derived from a new random split. The dataset release defines the electronic-structure settings and the adsorbate–slab and clean-slab composition of each partition.

Table 4. Structure counts and roles of the official OC22 IS2RE-Total training and validation splits. ID and OOD denote in-domain and out-of-domain validation, respectively.

Split	Role	Structures
train	Training	45,890
val_id	Validation (ID)	2,624
val_ood	Validation (OOD)	2,780

Species weights, standardisers, Nyström landmarks, the composition baseline and PLS projections were fitted on the training partition and replayed without refitting on both validation partitions.

F.2. Metrics and configuration selection

For reference energy y_s and prediction $\hat{y}_s$, the implementation reports structure-level RMSE in eV and a separately normalised per-atom MAE,

$$\mathrm{MAE}_{\mathrm{atom}} = \frac{1}{N_{\mathrm{eval}}} \sum_s \frac{|\hat{y}_s - y_s|}{N_s},$$
$$\mathrm{RMSE} = \Big(\frac{1}{N_{\mathrm{eval}}} \sum_s (\hat{y}_s - y_s)^2\Big)^{1/2},$$

where N_s is the number of atoms in structure s. Per-atom MAE is stored in meV/atom and divided by 1,000 for tables in eV/atom. ID validation determined hyperparameters and retrospective shortlists; OOD values were not used for selection. Because the same ID partition supports selection and reporting, all selected-setting differences are descriptive validation results rather than untouched test estimates.

F.3. Computational repeats and paired comparisons

Repeated settings used three to six computational seeds. The seeds changed subset sampling, Nyström landmarks and, for trainable circuits, parameter initialisation. We report the

median and observed range. These ranges are not confidence intervals and do not represent uncertainty over structures, training sets or hyperparameter choice.

ID MAE spanned 0.00381 eV/atom for the repeated Z-map setting and 0.00509 eV/atom for the repeated re-uploading setting at $\eta_{\mathrm{R}} = 0.10$. The corresponding linear-ridge ranges were 0.00164 and 0.00128 eV/atom. Re-uploading OOD MAE spanned 0.00844 eV/atom. These scales motivated paired comparisons: the residual learner and its control used the same selected subset and candidate-pool reference variance within each seed. The fidelity-kernel and re-uploading records contain 50 and 49 displayed rows, respectively, with six re-uploading rows reused across comparison blocks; the aggregate's stated 103-run count cannot be reconciled to unique runs without primary identifiers (Tables 6 and 7).

F.4. Selector timing and record boundary

The matched selector experiment used cached descriptors and the same downstream estimator. Enumeration fitted 65,535 subsets and required 323 s for selection and 541.67 s for the complete fit. QAOA proposal followed by classical rescoring fitted 333 subsets and required 22 s for selection and 234.84 s in total. The two retained QAOA runs differed by 0.1045 eV in ID RMSE and 0.3327 eV in OOD RMSE; the enumeration result reproduced to the reported precision.

The aggregate record contains mean-dispersion fields of 16.55 and 16.59 eV, but the released evaluation code does not emit this quantity. We therefore do not interpret it as predictive uncertainty. Candidate identities, per-structure predictions, checkpoints and complete logs are also absent. The 333-model pool and its timing are retained results, but cannot be regenerated exactly from the current source.

F.5. Fidelity-kernel diagnostic protocol

Each scale–map cell was accompanied by an offline 300×300 fidelity Gram matrix on standardised Gaussian inputs rather than OC22 structures. Its participation ratio defines the reported effective rank. Across 23 cells, effective rank correlated with ID MAE ($\rho_{\mathrm{s}} = 0.946$) and candidate-pool reference variance ($\rho_{\mathrm{s}} = 0.949$); the within-map correlations were 0.972 for ZZ and 0.936 for Z. These cells share data and selection procedures and are not independent replicates.

For the ZZ map, increasing η_{K} from 0.05 to 0.30 reduced the mean off-diagonal Gram entry from 0.26 to 0.016 and increased effective rank from 10.2 to 257.7. The selected subset contracted from 14 to 4 coordinates while $\hat{\sigma}_0^2$ increased from 434 to 652. At $\eta_{\mathrm{K}} \geq 0.30$, rank remained near 258–260 and validation error remained high. The near-constant $\eta_{\mathrm{K}} = 0.001$ cell had rank 1.00 and ID MAE 0.16335 eV/atom, but used a different subset and only one run.

At the selected $\eta_{\mathrm{K}} = 0.03$, six paired ZZ fits had a median ID MAE of 0.15671 eV/atom, compared with 0.16042 for ridge. The Z-map median was 0.16375 eV/atom but did not share the same selected subset and reference variance. The matched ZZ–ridge OOD comparison favoured ridge in five of six seeds by a median 0.00080 eV/atom. Complete scale cells and diagnostic values appear in Tables 6 and 8 and Fig. 3.

F.6. Re-uploading ablation protocol

The reduced profile has incompletely archived target construction. The recorded scale, depth and register-width changes shifted median ID MAE by 0.00153, 0.00198 and 0.00057 eV/atom, respectively. Excluding the highest-error $\eta_{\mathrm{R}} = 1.00$ cell, the scale sweep spanned 0.00685 eV/atom. The fitted-error register comparison covered $q = 4$ and $q = 6$; a separate diagnostic grid used $q = 3$–12. Because $L = n_{\mathrm{layers}}\lceil 48/q \rceil$, the diagnostic grid retained 580–611 trainable parameters while shortening the circuit as q increased. Prediction error was not recorded for every diagnostic-only setting.

Initial tangent-kernel effective rank did not correlate monotonically with fitted error. Across nine scale cells $\rho_{\mathrm{s}} = 0.000$, and excluding the highest-error cell gave $\rho_{\mathrm{s}} = -0.357$. In the paired trained–frozen comparison, the trainable circuit reduced median MAE by 0.01911 eV/atom when the frozen kernel had 5.7-fold greater rank and by 0.01152 eV/atom after rank matching. Against linear ridge, the trainable circuit median was 0.00128 eV/atom lower across five seeds, with overlapping observed ranges.

These ablations used a reduced profile that rescored 16 rather than 64 candidates, trained selection circuits for 40 rather than 200 epochs, used 4,000 selection structures and omitted the canonical out-of-fold residual pass. The released configurations do not preserve these overrides or identify how residual targets were formed after that omission. Comparisons are therefore valid only within the recorded profile, and the canonical pipeline must not be described as having regenerated them. Full rows and visual summaries are provided in Tables 7 and 8 and Figs. 4 and 6.

G. Full Related Work

The literature spans two distinct uses of quantum mechanics in atomistic prediction. First-principles electronic-structure methods generate physical energies and forces, which descriptor models and GNN MLIPs learn as classical surrogates. Quantum-learning studies instead use circuit states, overlaps or observables as computational representations of those classically generated labels, unless they explicitly measure an electronic Hamiltonian through methods such as VQE. The survey below preserves this distinction

and identifies simulator and hardware evidence separately; model-specific statements are supported by the corresponding primary publications.

Fixed representations. Early machine-learning interatomic potentials combine symmetry-preserving descriptors with neural or kernel regressors. Representative constructions include atom-centred symmetry functions (Behler & Parrinello, 2007), DeepMD local coordinates (Zhang et al., 2018), SOAP kernels (Bartók et al., 2010; 2013) and the atomic cluster expansion (Drautz, 2019). Their species scaling motivated constant-complexity descriptors (Artrith et al., 2017; Jain & Bligaard, 2018) and shared descriptor libraries (Himanen et al., 2020). Our representation extends the constant-complexity formulation of Artrith et al. (2017).

Learned geometric representations. Invariant graph networks learn local features from distances and angles, including SchNet, DimeNet and GemNet (Schütt et al., 2018; Gasteiger et al., 2020; 2021). Equivariant architectures propagate tensor features that transform with the input. Tensor-field networks established this construction, followed by NequIP, MACE and Allegro for atomistic modelling (Thomas et al., 2018; Batzner et al., 2022; Batatia et al., 2022; Musaelian et al., 2023). Equiformer and $SO(2)$-reduced variants combine equivariance with attention (Liao & Smidt, 2023; Passaro & Zitnick, 2023; Liao et al., 2024).

Pretrained and scalable potentials. Pretraining extends one potential across broader structural and chemical domains. M3GNet learns three-body graph interactions from Materials Project relaxations, whereas CHGNet incorporates magnetic-moment targets to represent charge-sensitive environments (Chen & Ong, 2022; Deng et al., 2023). SevenNet instead focuses on parallel deployment of an equivariant potential in molecular dynamics (Park et al., 2024). More recent universal families, including UMA and Orb-v3, enlarge the training domain while making different choices about equivariance, conservatism and model scale (Wood et al., 2025; Rhodes et al., 2025). Because their training corpora and objectives differ, these systems contextualize the design space but are not directly comparable controls for the present OC22 study.

Sparse and modular materials surrogates. When data or fitting budgets are limited, materials models commonly combine prescribed descriptors, regularised regression and explicit feature or model selection. Small-data strategies emphasise validation discipline and low-complexity representations (Zhang & Ling, 2018); Bayesian model averaging has likewise been used to combine adsorption-energy models rather than select a single specification (Mamun et al., 2020). Local-descriptor models refined by clustering provide a related route to adsorption-energy prediction on alloy surfaces (Usuga et al., 2024). QBM inherits the modular logic of these approaches but uses held-out simplex weights and a heuristic, pool-relative mBIC score. It therefore does not claim a coherent Bayesian posterior.

Circuit algorithms and learning primitives. Early proposals framed quantum computers as simulators of quantum dynamics (Manin, 1980; Feynman, 1982; Lloyd, 1996). The present use is different: QAOA treats feature-subset proposal as a combinatorial optimisation problem (Farhi et al., 2014), while fidelity kernels and re-uploading circuits define supervised representations of classical data (Havlíček et al., 2019; Schuld & Killoran, 2019; Pérez-Salinas et al., 2020). Trainable variational models and fixed-kernel models form distinct hypothesis classes, so their empirical difference cannot by itself be assigned to one circuit property (Jerbi et al., 2023). Quantum-kernel advantage additionally depends on data access, target alignment and classical comparison class, while concentrated kernels can require prohibitive measurement precision (Huang et al., 2021; Thanasilp et al., 2024). These conditions motivate the matched controls and spectral diagnostics used here.

Quantum representations for atomistic regression. Quantum kernels and variational circuits have previously been applied to molecular potential-energy surfaces and force prediction. Kiss *et al.* used re-uploading quantum neural networks to learn molecular total energies and derivative forces from ab initio labels, explicitly distinguishing supervised force-field learning from direct Hamiltonian solution (Kiss et al., 2022). Dai and Krems combined a quantum kernel with Gaussian-process regression for a global six-dimensional molecular potential-energy surface (Dai & Krems, 2022), while Paine *et al.* developed quantum-kernel regression methods for general regression and differential equations (Paine et al., 2023). Symmetry-invariant variational models later incorporated molecular rotations and permutations into the circuit architecture (Le et al., 2025). Couzinié *et al.* evaluated revised equivariant quantum force models on rMD17 and found improved force prediction but persistent difficulty in energy learning and data scaling (Couzinié et al., 2025). These studies support the use of circuit outputs as learned representations; none makes the circuit expectation itself a relaxed oxide energy.

Simulator–hardware separation in quantum surrogates. Lo Monaco *et al.* provide an unusually explicit comparison: their quantum extreme learner predicts LiH, water and formamide potential-energy surfaces using state-vector and finite-shot QASM simulation and also runs the same feature-generation idea on IBM Brisbane (Lo Monaco et al., 2024). Glick *et al.* similarly estimated a covariant kernel on a 27-qubit superconducting processor, while noting that their shallow structured circuit remained classically simulable

(Glick et al., 2024). These experiments show why a hardware result must report shots, noise handling, compilation and device identity separately from an ideal simulation. The present QBM study uses seeded finite-shot QAOA together with exact direct-amplitude kernel and re-uploading calculations, all on a CPU. It does not establish that the same accuracy, runtime or scaling would persist on hardware.

Direct quantum energy estimation. Variational electronic-structure algorithms address a different task. Meta-VQE learns an energy profile of a parameterised Hamiltonian by repeatedly minimising Hamiltonian expectations (Cervera-Lierta et al., 2021); hybrid quantum–neural wavefunctions likewise use circuit samples and a neural correction to estimate molecular electronic energies and have been tested on superconducting hardware (Li et al., 2025). Contextual-subspace VQE has produced an experimental nitrogen dissociation curve on superconducting hardware, and practical VQE studies compare noiseless, device-noise and real-hardware estimates of molecular ground-state energy (Weaving et al., 2025; Belaloui et al., 2025). Large matrix-product-state emulation of VQE further shows that a "quantum simulator" can itself be an HPC calculation (Shang et al., 2023). These works compute expectations of physical Hamiltonians. The present method instead learns an OC22 PBE+U DFT label from classical structural descriptors; its QAOA cost, fidelities and Pauli-Z observables are intermediate algorithmic quantities rather than total, formation, potential or ground-state energies.

The present work instead evaluates QAOA subset proposal, a fidelity residual kernel and a variational residual circuit after classical descriptor construction and compression on OC22. Its contribution is the modular separation of selector, fixed-kernel and trainable-residual evidence, not the invention of these components or a capability shown to be unavailable to classical graph or kernel models. The results therefore support a bounded component study rather than replacement of graph-based potentials.

H. Notation

Table 5: Definitions of the principal symbols used in the model formulation, evaluation and appendices.

Symbol	Description
$\mathcal{S},\ \mathbf{r}_i$	Periodic structure of the IS2RE-Total task, and the unrelaxed position of atom i; $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$.
$E,\ \hat{E}$	DFT total energy of the relaxed structure and its prediction, in eV.
$\hat{E}_{\mathrm{comp}},\ \tilde{y}_s$	Composition baseline, and the target of structure s after subtracting it.
$R_{\mathrm{c}},\ f_{\mathrm{c}}$	Cut-off radius of a local environment, and the cosine cut-off applied inside it.
$R_{ij},\ \vartheta_{ijk},\ \phi_{ijkl}$	Interatomic distance, bond angle, and unsigned dihedral angle.
$\tau(Z)$	Species weight of element Z, written $\tau_j = \tau(Z_j)$; $\tau \equiv 1$ gives the structural block.
$\mathcal{T}_i$	Bonded chains $i{-}j{-}k{-}l$ rooted at atom i.
$T_\alpha,\ c^{\mathrm{R}}_{i,\alpha}, c^{\mathrm{A}}_{i,\alpha}, c^{\mathrm{T}}_{i,\alpha}$	Chebyshev polynomial at order α, and the radial, angular and torsional expansion coefficients of atom i.
$\mathbf{f}_i,\ D$	Per-atom descriptor and its dimension, $D = 2(n_{\mathrm{R}} + n_{\mathrm{A}} + n_{\mathrm{T}})$.
$\boldsymbol{\zeta}(\cdot),\ M$	Nyström RBF map and its landmark count: $M = 1024$ at atom level, $M = 200$ in the per-subset KRR.
$r^{\mathrm{ads}}_i,\ \ell_{\mathrm{ads}}$	Distance from atom i to the nearest adsorbate atom, and the pooling decay length.
$\boldsymbol{\Phi}$	Structure-level descriptor: global pool concatenated with the adsorbate-local pool.
$\mathbf{X},\ \mathbf{x}_s$	PLS latent feature matrix and the feature vector of sample s.
$n,\ p$	Training samples, and candidate features after the projection.
$\kappa(\cdot,\cdot),\ \mathbf{K}$	Kernel and its Gram matrix, $K_{mn} = \kappa(\mathbf{x}_m, \mathbf{x}_n)$.
$\lambda,\ \boldsymbol{\beta}$	Ridge strength and dual coefficient vector of KRR.
$\hat{f},\ \mathrm{df}$	Fitted predictor, and the effective degrees of freedom entering equation (5).
$\mathcal{M},\ d,\ k_{\mathrm{S}},\ k_{\mathrm{B}}$	Candidate subset, its size $d = \lvert\mathcal{M}\rvert$, the subset-complexity count $k_{\mathrm{S}} = d + 1$, and the backend-specific BIC-style count. For KRR, $k_{\mathrm{B}} = k_{\mathrm{S}}$; for a re-uploading finalist, $k_{\mathrm{B}} = N_\theta$.
$\hat{\sigma}^2_{\mathcal{M}},\ \hat{\sigma}^2_0$	Backend-reported residual variance of sub-model $\mathcal{M}$, and the smallest one in its candidate pool; KRR uses equation (5), whereas variational refits use their recorded residual variance.
$\gamma,\ \mathrm{mBIC}$	EBIC sparsity weight $0 \le \gamma \le 1$, and the criterion of equation (6).
$K_{\mathrm{ens}},\ n_{\mathrm{fold}}$	Sub-models combined, and conditional fold-excluding subset-refit splits used to weight them.
$w_m,\ \mathbf{Y}^{\mathrm{oof}}$	Combination weights and the conditional fold-excluding subset-refit prediction matrix they are fitted on; upstream supervised transforms are fitted on all training rows.
$r_s,\ \hat{r},\ \mathbf{u}$	Conditional fold-excluding subset-refit residual, the learnt correction, and the residual learner's input.
$\mathbf{z},\ \mathbf{Q}$	Binary subset indicator and the upper-triangular QUBO cost matrix.
$a_i,\ b_{ij}$	First- and second-order finite differences of the log-variance surrogate.
$d_{\mathrm{t}},\ c_{\mathrm{t}},\ \nu_0,\ \nu$	QUBO target cardinality, its frozen per-feature cost, the relative anchor scale and the resulting coefficient in equation (D.9).
$\hat{H}_{\mathrm{C}},\ \hat{H}_{\mathrm{B}},\ L_{\mathrm{Q}}$	Ising cost Hamiltonian, QAOA mixer, and circuit depth.
$q,\ q_{\max},\ L_{\mathrm{K}},\ L$	Active qubits, the qubit budget $q = \min(d, q_{\max})$, the fidelity-map layer count, and the re-uploading layer count.
$n_{\mathrm{seed}},\ N_\theta$	Number of independently seeded pipeline fits in a repeated configuration, and number of trainable circuit parameters.
$\iota(\ell, i),\ \omega_{\ell i}$	Cyclic layer-to-feature map, and the encoding angle it feeds.
$\eta_{\mathrm{K}},\ \eta_{\mathrm{R}}$	Angle scale of the fidelity kernel, and the initial re-uploading scale.
$F,\ \kappa_F,\ U_F,\ \lvert\phi_F(\mathbf{x})\rangle$	Feature-map label $F \in \{Z, ZZ\}$, its fidelity kernel, unitary encoding and encoded state.
$P(\varphi),\ \phi_{\ell,ct}$	Phase gate used in Fig. 2d,e, and the data-dependent pair phase of the ZZ map.
$\boldsymbol{\gamma},\ \boldsymbol{\beta}$	QAOA cost and mixer angles.
$\boldsymbol{\theta}^{(\ell)},\ \mathbf{g},\ b$	Layerwise variational angles, and the readout coefficients and offset of the re-uploading model.

I. Supplementary Figures and Tables

Figures 3 and 4 report the complete fidelity-kernel and re-uploading ablations. Figures 5 and 6 then summarize seed variation and initial tangent-kernel capacity. The corresponding fit-level records and offline spectral diagnostics are given in Tables 6 to 8.

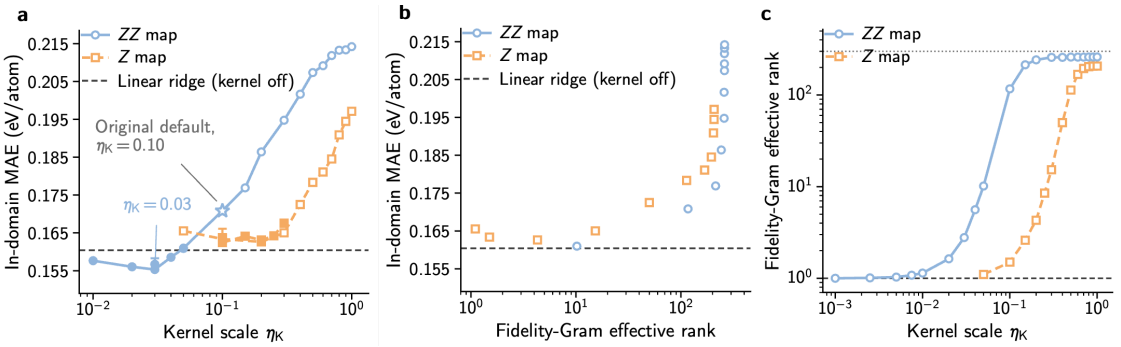


Figure 3. **Fidelity-Gram geometry and validation error across the kernel-scale sweep. a**, In-domain validation MAE as a function of kernel scale η_K for the entangled ZZ map (blue circles) and single-body Z map (orange squares). Open symbols show the initial sweep and filled symbols the lower-scale extension. The two ZZ grids overlap at $\eta_K = 0.05$, where their MAEs differ by 0.00008 eV/atom. At each validation-selected scale, the filled symbol and error bar denote the median and observed range over six computational seeds. The grey band and dashed line denote the observed range and median of the seed-matched linear-ridge control. The star marks the original default $\eta_K = 0.10$. **b**, In-domain MAE against effective rank for the 23 dependent scale–map cells in the initial sweep. The grey ridge-control range is repeated from **a**; ρ_s is the Spearman rank correlation across these cells. **c**, Effective rank of offline fidelity Gram matrices evaluated on 300 standardised Gaussian inputs, not OC22 structures. Horizontal guides mark the 300-sample rank ceiling and unit-rank constant-kernel limit; blue shading marks $0.02 \leq \eta_K \leq 0.05$, which contains the lowest recorded ZZ-map MAE. Error bars and bands are observed seed ranges, not confidence intervals. Inferential p values are omitted because hyperparameter cells and computational seeds are not independent experimental replicates.

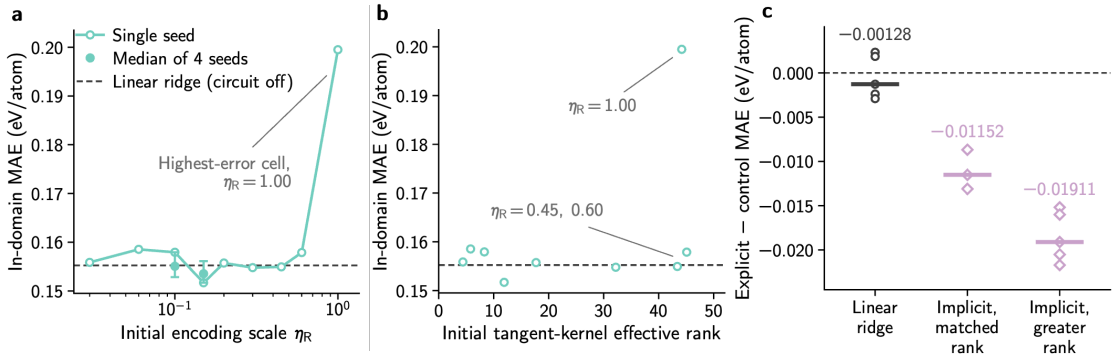


Figure 4. **Variational re-uploading residuals and matched residual controls. a**, In-domain validation MAE across the initial encoding scale η_R at $q = 6$ and $L = 24$. Open symbols show the single-seed sweep; filled symbols and error bars denote the median and observed range over four seeds at $\eta_R = 0.10$ and 0.15. The grey band and dashed line denote the observed range and median of the five-seed circuit-off linear-ridge control. **b**, In-domain MAE for the same nine sweep cells against tangent-kernel effective rank at initialisation, evaluated on the 300-input Gaussian probe. The reported ρ_s values include all nine cells or exclude the highest-error cell at $\eta_R = 1.00$. **c**, Seed-paired difference in MAE between the trained circuit and each residual control; negative values favour the trained circuit. Open symbols show individual seeds and horizontal bars their medians. The reduced profile omitted the canonical out-of-fold residual pass and does not identify its replacement residual targets or early-stopping partition. Implicit controls are Gaussian processes using the fidelity kernel of the frozen encoding, with NTK-to-fidelity effective-rank ratios of $1.0\times$ ($\eta_R = 0.06$, $n = 3$) or $5.7\times$ ($\eta_R = 0.45$, $n = 5$); the linear-ridge comparison uses $n = 5$. Paired arms share the selected subset and reference variance within the recorded profile; the target provenance and leakage boundary cannot be independently checked. All ranges are descriptive spreads over computational seeds and do not define inferential intervals.

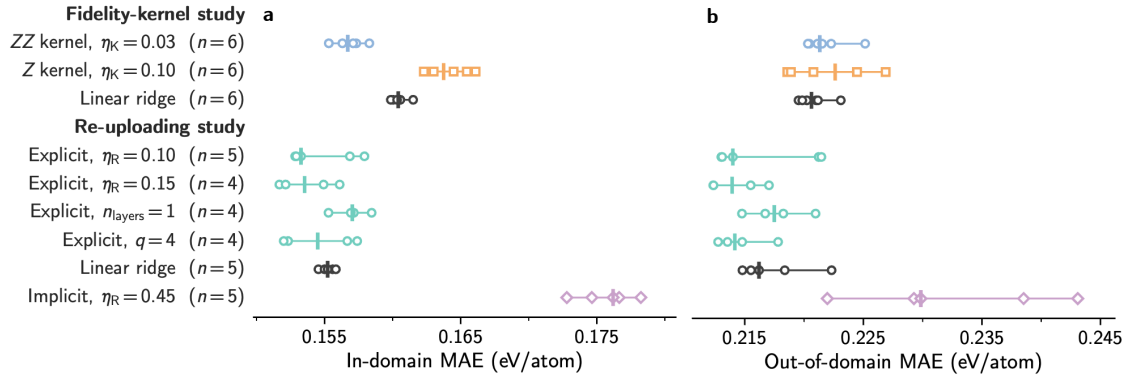


Figure 5. **Seed-to-seed variation in repeated pipeline fits.** Each row contains all recorded validation fits for one configuration; the row label gives the number of seeds, n. Open symbols denote individual fits, the vertical bar denotes the median and the horizontal segment spans the observed minimum and maximum. Blue identifies fidelity-kernel configurations, orange identifies re-uploading configurations and grey identifies matched linear-ridge controls. The value at right is the observed range, defined as maximum minus minimum, in eV/atom. **a**, In-domain MAE. **b**, Out-of-domain MAE. The displayed ranges quantify computational-seed variation only; they are neither confidence intervals nor estimates of a noise floor.

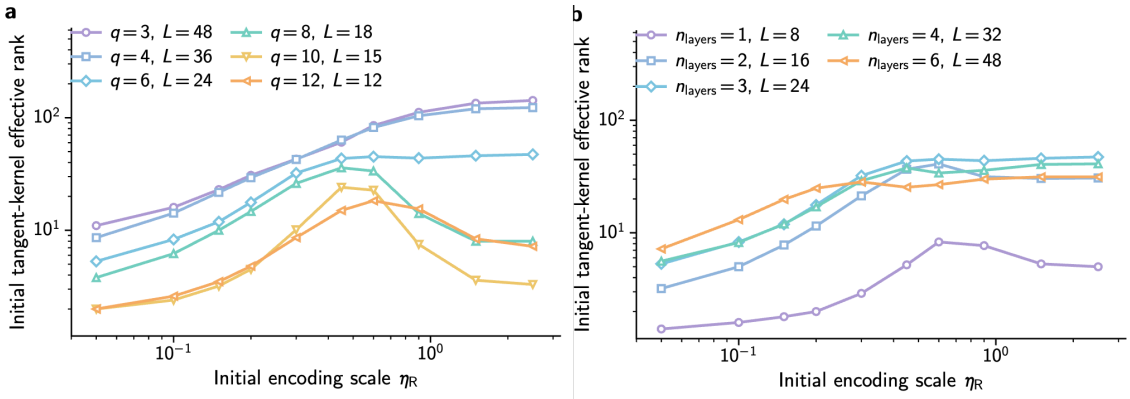


Figure 6. **Initial tangent-kernel effective rank across re-uploading circuit configurations.** Effective rank is evaluated on 300 standardised Gaussian inputs and plotted against the initial encoding scale η_R. **a**, Register width q at an approximately matched trainable-parameter budget. For $q = 3, 4, 6, 8, 10$ and 12, the corresponding parameter counts N_θ are 580, 581, 583, 585, 611 and 589. **b**, Number of upload layers at $q = 6$. For $n_{layers} = 1, 2, 3, 4$ and 6, N_θ is 199, 391, 583, 775 and 1,159, respectively. The $q = 6$, $L = 24$ series appears in both panels and matches the NTK entries in Table 8. These diagnostic configurations were evaluated without fitting the full pipeline; effective rank alone did not order the fitted errors in Fig. 4b.

Table 6. **Fidelity-kernel ablations and paired linear-ridge controls.** The 50 fit-level records are ordered as the initial scale–map sweep, lower-scale extension, seed study and kernel-off controls. All residual models use width 16 and the official OC22 IS2RE-Total training, in-domain (ID) validation and out-of-domain (OOD) validation partitions. Training and validation MAE are reported in eV/atom; OOF residual-target error and RMSE are in eV. $\hat{\sigma}_0^2$, $|\mathcal{M}|$ and rank denote the reference variance, selected-subset size and fidelity-Gram effective rank, respectively. Seed-paired kernel and ridge rows share $\hat{\sigma}_0^2$ and the selected subset. Bold and underlined values mark the lowest and second-lowest entries in the performance columns across the complete table; they do not imply statistical significance. A dash denotes an unrecorded diagnostic.

Map	η_K	seed	rank	$\hat{\sigma}_0^2$	$\lvert\mathcal{M}\rvert$	OOF	train MAE	ID MAE	ID RMSE	OOD MAE	OOD RMSE
ZZ	0.05	0	10.2	433.67	14	19.9987	0.15045	0.16098	18.4337	0.22337	<u>28.2610</u>
ZZ	0.10	0	116.8	563.21	7	22.3706	0.16316	0.17083	20.7492	0.22860	29.6325
ZZ	0.15	0	213.9	597.19	4	23.3518	0.17223	0.17690	21.4487	0.23873	31.6693
ZZ	0.20	0	242.2	622.09	4	24.1124	0.18074	0.18640	22.4477	0.24720	32.9123
ZZ	0.30	0	257.7	652.03	4	25.1316	0.19081	0.19477	23.4429	0.25485	34.1076
ZZ	0.40	0	257.8	682.40	4	25.7410	0.19643	0.20165	24.2053	0.25415	34.4825
ZZ	0.50	0	259.6	697.32	4	26.1543	0.20190	0.20736	24.5132	0.25623	34.7247
ZZ	0.60	0	259.4	702.53	4	26.3520	0.20584	0.20917	24.7816	0.25755	34.8367
ZZ	0.70	0	259.5	705.45	2	26.4462	0.20789	0.21186	24.8568	0.25879	34.7981
ZZ	0.80	0	260.0	707.64	2	26.5010	0.20997	0.21327	24.9358	0.25844	34.7647
ZZ	0.90	0	259.9	709.54	2	26.5591	0.21067	0.21337	25.0248	0.25846	34.7904
ZZ	1.00	0	260.3	711.78	2	26.6210	0.21148	0.21421	24.9992	0.25840	34.9112
Z	0.05	0	1.1	448.11	10	20.3383	0.15844	0.16553	18.9154	0.22850	30.4879
Z	0.10	0	1.5	429.65	14	19.8526	0.15417	0.16337	18.9710	0.22486	29.8752
Z	0.20	0	4.3	433.65	13	19.9959	0.15276	0.16263	18.5333	0.22374	29.3407
Z	0.30	0	15.3	477.59	14	20.8001	0.15605	0.16502	19.1696	0.22485	28.7278
Z	0.40	0	50.0	527.79	14	21.9570	0.16284	0.17250	20.5106	0.22934	29.4808
Z	0.50	0	113.3	569.72	14	22.9862	0.17079	0.17835	21.7323	0.23397	30.5174
Z	0.60	0	168.1	607.73	14	23.7666	0.17728	0.18109	22.3561	0.24268	32.0543
Z	0.70	0	194.8	636.17	14	24.3517	0.17989	0.18453	22.7182	0.24545	32.2215
Z	0.80	0	204.0	654.98	13	24.8842	0.18512	0.19090	23.1643	0.25110	33.6048
Z	0.90	0	206.6	666.56	9	25.2983	0.18823	0.19447	23.5448	0.25115	33.6056
Z	1.00	0	207.0	678.67	13	25.6188	0.19227	0.19712	23.8526	0.25338	34.2255
ZZ	0.01	0	—	370.74	14	19.1238	0.14819	0.15765	17.8856	0.22061	29.5337
ZZ	0.02	0	—	366.02	14	**18.8795**	**0.14484**	0.15604	17.7206	0.21957	29.0160
ZZ	0.03	0	—	376.12	14	<u>19.0664</u>	0.14537	<u>0.15532</u>	<u>17.7073</u>	0.22113	28.9042
ZZ	0.04	0	—	405.02	14	19.5031	0.14815	0.15855	17.8396	0.22047	**28.0682**
ZZ	0.05	0	—	437.52	14	20.0444	0.15097	0.16090	18.2686	0.22596	28.5953
Z	0.10	0	—	429.79	14	19.9121	0.15369	0.16245	18.5202	**0.21859**	29.0111
Z	0.15	0	—	424.93	14	19.7418	0.15426	0.16413	18.6436	0.23007	30.2939
Z	0.20	0	—	433.40	14	19.9942	0.15226	0.16311	18.3767	0.22557	29.8416
Z	0.25	0	—	454.71	14	20.3366	0.15687	0.16423	18.8313	0.22136	28.6798
Z	0.30	0	—	477.59	14	20.7476	0.15830	0.16756	19.2300	0.22418	28.7133
ZZ	0.03	1	—	382.23	14	19.1520	<u>0.14491</u>	0.15733	17.8004	0.22152	28.9492
ZZ	0.03	2	—	383.36	14	19.1335	0.14620	**0.15531**	17.7272	0.22045	28.6345
ZZ	0.03	3	—	378.04	14	19.0696	0.14523	0.15631	**17.6336**	0.22515	29.0394
ZZ	0.03	4	—	385.91	14	19.1607	0.14606	0.15829	18.0985	0.22229	28.8190
ZZ	0.03	5	—	385.16	14	19.1095	0.14537	0.15710	17.8119	0.22032	28.5717
Z	0.10	1	—	429.65	14	19.8009	0.15273	0.16229	18.6142	0.22688	30.1857
Z	0.10	2	—	429.65	14	19.9392	0.15390	0.16448	19.0147	0.22445	29.7608
Z	0.10	3	—	429.65	14	19.9720	0.15692	0.16610	18.9291	0.22079	29.7428
Z	0.10	4	—	429.65	14	19.9129	0.15275	0.16302	18.6456	0.22447	29.8672
Z	0.10	5	—	429.65	14	19.7749	0.15648	0.16545	18.9122	<u>0.21889</u>	29.1282
ridge	0.03	0	—	376.12	14	<u>19.0664</u>	0.15001	0.16004	17.8959	0.21953	28.7018
ridge	0.03	1	—	382.23	14	19.1520	0.14969	0.16050	18.0207	0.22310	29.0906
ridge	0.03	2	—	383.36	14	19.1335	0.15042	0.15987	18.0173	0.22023	28.6672
ridge	0.03	3	—	378.04	14	19.0696	0.14980	0.16057	18.0163	0.22102	28.8355
ridge	0.03	4	—	385.91	14	19.1607	0.15020	0.16033	18.0648	0.22117	28.6963
ridge	0.03	5	—	385.16	14	19.1095	0.15029	0.16151	18.0637	0.21985	28.6291
ZZ GP	0.001	0	—	400.27	16	20.0268	0.15578	0.16335	18.5164	0.22136	29.2661

Table 7. **Re-uploading ablations, frozen-encoding kernels and paired ridge controls.** The table contains 49 recorded rows, corresponding to 43 distinct fits because six fits are reused across comparison blocks. All models use residual width 48, the reduced training profile in Appendix A and the official OC22 IS2RE-Total training, ID-validation and OOD-validation partitions. Here, q is the number of qubits, $L = n_{\text{layers}}\lceil 48/q \rceil$ is the encoding-layer count, N_θ is the number of trainable parameters and NTK is the tangent-kernel effective rank at initialisation. MAE is in eV/atom and RMSE in eV. Paired arms share the selected subset and reference variance. Bold and underlined values mark the lowest and second-lowest entries in each error column, including ties and reused fits; this emphasis is descriptive, not inferential. A dash denotes a non-applicable diagnostic.

Cell	η_{R}	q	L	N_θ	NTK	seed	$\hat{\sigma}_0^2$	$\lvert\mathcal{M}\rvert$	ID MAE	ID RMSE	OOD MAE	OOD RMSE
trained	0.03	6	24	583	4.4	0	437.97	11	0.15586	17.2777	0.21870	28.1625
trained	0.06	6	24	583	5.8	0	437.16	11	0.15853	17.1659	0.21284	**26.4422**
trained	0.10	6	24	583	8.3	0	431.63	11	0.15793	17.2336	0.21399	26.7002
trained	0.15	6	24	583	11.9	0	449.39	10	**0.15168**	16.9645	0.21550	27.2736
trained	0.20	6	24	583	17.7	0	450.56	11	0.15573	17.4292	0.22317	28.5783
trained	0.30	6	24	583	32.2	0	447.20	11	0.15478	17.4800	0.22076	27.5661
trained	0.45	6	24	583	43.4	0	460.01	11	0.15494	17.4824	0.22136	28.1069
trained	0.60	6	24	583	45.1	0	458.73	11	0.15788	17.5229	0.22194	27.8144
trained	1.00	6	24	583	44.2	0	572.42	10	0.19949	23.3946	0.24697	30.5972
trained	0.15	6	24	583	11.9	1	421.32	11	0.15214	<u>16.6819</u>	0.21704	27.5431
trained	0.15	6	24	583	11.9	2	432.60	11	0.15493	16.8795	<u>0.21234</u>	26.8040
trained	0.15	6	24	583	11.9	3	429.16	12	0.15611	17.3091	0.21236	26.7708
trained	0.10	6	24	583	8.3	1	426.15	12	0.15327	17.0532	0.22123	28.2530
trained	0.10	6	24	583	8.3	2	434.60	12	0.15686	17.1358	0.21302	26.5245
trained	0.10	6	24	583	8.3	3	425.89	12	0.15284	16.9407	0.21309	26.9290
trained	0.10	6	8	199	1.6	0	475.45	11	0.15846	17.5163	0.21825	27.5216
trained	0.10	6	8	199	1.6	1	457.86	11	0.15529	17.3099	0.21476	27.1788
trained	0.10	6	8	199	1.6	2	448.23	12	0.15715	17.3146	0.21672	26.8992
trained	0.10	6	8	199	1.6	3	450.68	12	0.15693	17.4255	0.22097	28.0706
trained	0.10	4	36	581	14.2	0	436.73	11	0.15667	17.3105	0.21781	27.4908
trained	0.10	4	36	581	14.2	1	415.71	12	0.15740	17.3178	0.21354	26.4683
trained	0.10	4	36	581	14.2	2	426.78	12	0.15232	**16.6763**	0.21478	27.4100
trained	0.10	4	36	581	14.2	3	421.23	12	<u>0.15200</u>	16.9212	0.21275	26.8080
trained	0.45	6	24	583	43.4	0	460.01	11	0.15494	17.4824	0.22136	28.1069
frozen	0.45	6	24	583	43.4	0	460.01	11	0.17665	20.3390	0.22926	29.0032
trained	0.45	6	24	583	43.4	1	425.27	12	0.15369	16.8564	0.21301	26.7101
frozen	0.45	6	24	583	43.4	1	425.27	12	0.17280	20.1967	0.22985	29.7800
trained	0.45	6	24	583	43.4	2	430.18	12	0.15945	17.3465	0.21440	26.9555
frozen	0.45	6	24	583	43.4	2	430.18	12	0.17464	19.9810	0.23852	31.0524
trained	0.45	6	24	583	43.4	3	443.18	12	0.15773	17.3833	0.22539	28.5405
frozen	0.45	6	24	583	43.4	3	443.18	12	0.17824	20.4471	0.24308	30.3761
trained	0.45	6	24	583	43.4	4	439.66	11	0.16020	17.3933	0.21482	<u>26.4443</u>
frozen	0.45	6	24	583	43.4	4	439.66	11	0.17620	20.7754	0.22196	28.8730
trained	0.06	6	24	583	5.8	0	437.16	11	0.15853	17.1659	0.21284	**26.4422**
frozen	0.06	6	24	583	5.8	0	437.16	11	0.16721	19.0929	0.22062	28.2370
trained	0.06	6	24	583	5.8	1	423.66	11	0.15295	17.0058	**0.21209**	26.7874
frozen	0.06	6	24	583	5.8	1	423.66	11	0.16606	19.1900	0.22695	28.5883
trained	0.06	6	24	583	5.8	2	437.31	12	0.15584	16.9482	0.21607	26.9068
frozen	0.06	6	24	583	5.8	2	437.31	12	0.16736	18.6710	0.22298	28.9642
trained	0.10	6	24	583	8.3	0	431.63	11	0.15793	17.2336	0.21399	26.7002
ridge	0.10	6	24	583	—	0	431.63	11	0.15560	17.2751	0.21480	27.1386
trained	0.10	6	24	583	8.3	1	426.15	12	0.15327	17.0532	0.22123	28.2530
ridge	0.10	6	24	583	—	1	426.15	12	0.15455	17.1292	0.21837	27.7866
trained	0.10	6	24	583	8.3	2	434.60	12	0.15686	17.1358	0.21302	26.5245
ridge	0.10	6	24	583	—	2	434.60	12	0.15500	17.0722	0.21552	27.2217
trained	0.10	6	24	583	8.3	3	425.89	12	0.15284	16.9407	0.21309	26.9290
ridge	0.10	6	24	583	—	3	425.89	12	0.15523	17.1779	0.21620	27.2814
trained	0.10	6	24	583	8.3	4	435.63	11	0.15292	17.2706	0.22146	28.0239
ridge	0.10	6	24	583	—	4	435.63	11	0.15583	17.4004	0.22232	28.0577

Table 8. **Offline spectral diagnostics for fidelity and re-uploading feature maps.** All dimensionless diagnostics are evaluated on 300 standardised Gaussian inputs without fitting the full prediction pipeline; effective rank is the participation ratio of the Gram spectrum. The upper block reports fidelity-kernel diagnostics at $d = 16$. Mean off-diagonal similarity records the transition from a near-constant to an increasingly identity-like Gram matrix. The lower block compares the tangent-kernel rank at initialisation with the fidelity rank of the same frozen re-uploading encoding at $d = 48$, $q = 6$ and $L = 24$. Their ratio defines the effective-rank-matched and effective-rank-imbalanced controls in Table 7. A dash denotes an uncomputed cell.

Fidelity kernel of the selection features ($d = 16$)			
	ZZ map		Z map
η_{K}	eff. rank	mean off-diag.	eff. rank
0.001	1.00	0.99934	—
0.0025	1.01	0.99591	—
0.005	1.03	0.98375	—
0.0075	1.08	0.96389	—
0.01	1.14	0.93690	—
0.02	1.63	0.77521	—
0.03	2.78	0.57710	—
0.05	10.2	0.26241	1.1
0.10	116.8	0.044	1.5
0.15	213.9	—	—
0.20	242.2	—	4.3
0.30	257.7	0.016	15.3
0.40	257.8	—	50.0
0.50	259.6	—	113.3
0.60	259.4	—	168.1
0.70	259.5	—	194.8
0.80	260.0	—	204.0
0.90	259.9	—	206.6
1.00	260.3	—	207.0
Re-uploading encoding ($d = 48$, $q = 6$, $L = 24$)			
η_{R}	NTK rank	fidelity rank	ratio
0.03	4.4	3.6	0.8×
0.06	5.8	5.8	1.0×
0.10	8.3	11.1	1.3×
0.15	11.9	22.2	1.9×
0.20	17.7	43.2	2.4×
0.30	32.2	146.0	4.5×
0.45	43.4	247.3	5.7×
0.60	45.1	260.7	5.8×
1.00	44.2	262.3	5.9×