CIRCUIT-FRONTIER REPRESENTATIONS FOR QUANTUM ARCHITECTURE SEARCH

Anonymous authors
Paper under double-blind review

ABSTRACT

Hardware-aware quantum architecture search must connect logical operation dependencies with the physical resources used to execute a circuit. We introduce a circuit-frontier representation that maintains a global circuit summary and an ordered embedding for each logical wire. An autoregressive builder uses these embeddings to select operations. We study two placement strategies within the same construction framework: a learned hardware-attention pointer that reuses the embeddings, and a compiled-cost heuristic that maps the constructed circuit directly. This separation lets us test learned placement without assuming that both decisions must be neural policies. A balanced study covers six H_2/LiH/H_2O Hamiltonians, single- and ten-calibration training, three paired training blocks, and density-matrix simulation with native-gate and readout noise from IBM Fez calibrations. The learned-placement model has favorable aggregate policy-only point estimates against the learned controls under single-snapshot training, but several structural directions reverse under ten-snapshot training and no registered aggregate advantage survives correction. Retaining acquired circuits reveals a different comparison: the heuristic-placement learner's archive improves eight of twelve task–exposure means relative to its own frozen policy and ties the other four at reported precision, with reductions of 10.455 and 8.876 mHa on the two water tasks under ten-snapshot training. Source-matched remapping shows that archive usefulness depends on deployment placement, while a fixed policy–archive mixture does not improve the full model's normalized aggregate. The results identify circuit acquisition, placement, and candidate retention as distinct requirements for evaluating a learned circuit representation.

1 INTRODUCTION

A circuit-search algorithm chooses logical operations, whereas a quantum processor executes native gates on particular physical qubits. The quality of a proposed architecture depends on both choices. An interaction that is useful for the target Hamiltonian may be expensive on one physical assignment and inexpensive on another. This coupling motivates noise-aware architecture and mapping search (Wang et al., 2022), but it does not prescribe how the two decisions should be represented or whether both should be learned.

The representation problem is asymmetric. During construction, a policy must summarize the operations already chosen and determine which logical wires a new operation will affect. During placement, it must distinguish the histories and roles of those wires while comparing them with calibrated physical sites. A single circuit vector compresses away the explicit wire correspondence. A serialization retains operations but introduces an ordering between independent events. We instead retain the *frontier* of a circuit's dependency graph: the current endpoint of every logical wire. Global and wire-indexed readouts then provide complementary views of the same constructed prefix.

We introduce a circuit-frontier graph encoder and use it in an autoregressive architecture-search policy. The encoder propagates information along port-labeled dependencies, preserving wire identity and the ordered operands of each operation. We instantiate the circuit builder with two placement strategies. In *learned placement*, an attention pointer reuses the terminal frontier embeddings together with a hardware graph. In *heuristic placement*, an additive compiled-cost rule maps the circuit selected by the frontier builder. The heuristic does not consume neural embeddings; it is a nonlearned placement

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

alternative within the same learned-construction framework. Both strategies train the builder using the noisy energy of the resulting physical circuit.

We ask *when learning placement adds value to a structured circuit builder*. Its effect on search can differ from its effect when redeploying acquired circuits. We therefore preserve both the final policy and the circuits evaluated during training. Each archive belongs to its own acquisition run. We compare policy, archive, and fixed-budget mixed deployment; remapping interventions hold archive membership fixed.

Learned placement has favorable aggregate policy-only point estimates against heuristic-placement PPO under both exposures, but that ordering does not extend to the archive-backed comparison. Heuristic-placement archives improve or match their own policy's task means, and remapping the full model's archive can reverse its ranking. Structural comparisons also depend on exposure. These findings separate representation, placement, and retention rather than establish general superiority.

Our contributions are: a dependency-sensitive, per-wire representation for online circuit construction and learned assignment; a common construction framework in which learned and heuristic placement can be compared without changing the physical objective; and a balanced molecular evaluation that separates representation effects from acquisition mapping, deployment mapping, and circuit retention.

2 Related Work

Quantum architecture search and placement. QAS optimizes circuit structure for variational algorithms (Du et al., 2022), including reinforcement-learning construction (Ostaszewski et al., 2021; Patel et al., 2024). Calibration-aware compilation accounts for physical-qubit and gate reliability (Murali et al., 2019); QuantumNAS jointly searches circuits and mappings through supercircuit parameter sharing and evolutionary search (Wang et al., 2022). We study the interface between sequential construction and assignment, not the general idea of noise-aware co-search. Our compiled-cost heuristic is a comparison rule, not a new routing algorithm or an energy oracle.

Representations for QAS. Lourens et al. (2023) organize circuit architectures as hierarchical directed-graph motifs. Sun et al. (2026) learn latent representations from circuit graphs for subsequent search. Particularly close, He et al. (2024) preserve qubit information-flow paths in DAGs and use asynchronous predecessor-dependent encoding in a meta-trained circuit generator. EAQAS combines representations of data-embedding operations and gate topology using hierarchical graph autoencoding and cross-attention (Dai et al., 2026). Its data-embedding problem differs from our Hamiltonian descriptors and hardware assignment. These works already establish circuit graphs, qubit-flow dependencies, and learned circuit generation. Our focus is an ordered readout for every logical wire of an observed construction prefix, reused with a global summary by the operation head and hardware-conditioned assignment head. We train separate molecular policies, rather than a cross-task meta-generator. Our experiments compare components within this construction-and-placement framework, not these approaches head-to-head.

Graph and attention components. Topological propagation follows the broader asynchronous DAG-encoding approach (Zhang et al., 2019). Quarl uses graph representations and decomposed RL decisions for circuit rewriting (Li et al., 2024). Our circuit encoder propagates along operand-port-labeled dependencies; hardware features use GATv2 (Brody et al., 2022), and assignment uses sequential attention (Kool et al., 2019). The proposed interface combines these ideas; neither graph encoding nor attention alone is the novelty claim.

Evaluating acquired solutions. A learned generator and its acquired circuit archive are different outputs. We compare their deployment with matched parent histories and candidate budgets; retention itself is not a new optimization principle. It provides a control for interpreting construction and placement. Uncertainty is measured across independent training blocks, not candidate draws (Agarwal et al., 2021).

3 CIRCUIT-FRONTIER SEARCH

3.1 PHYSICAL OBJECTIVE AND CONSTRUCTION STATE

We study circuit search for variational eigensolvers (Peruzzo et al., 2014). Let $H = h_0 I + \sum_\ell h_\ell P_\ell$ be a qubit Hamiltonian, c a sequence of logical operations, and m an injective logical-to-physical assignment. A hardware context G_d describes connectivity and calibration d. Compilation restricts a fixed candidate roster to feasible assignments $\mathcal{M}_d(c)$. For one selected pair $u = (c, m)$, a bounded optimizer starts at zero angles and returns $\hat{\theta}_B(u, d)$ after at most B objective calls. Write $E_{d,B}(u)$ for the resulting gate-and-readout-affected energy estimate. The reward is

$$R(u, d) = E_d^{\mathrm{bare}} - E_{d,B}(u), \tag{1}$$

using a common recorded reference-preparation convention. No exact ground-state energy enters training. The gate-only diagnostic is $\mathrm{Tr}[H\rho_d(u, \hat{\theta}_B)]$; readout distortion means that the full objective is not a variational state-energy certificate.

The builder observes only the prefix c_t already constructed, the Hamiltonian/reference descriptors, and a compiler-derived action mask. The original builder has no direct numerical calibration input. Hardware can restrict construction through feasibility, while the placement stage has access to numerical hardware descriptors. Each action appends one operation from the fixed conserving catalog or terminates construction.

3.2 A GLOBAL SUMMARY WITH EXPLICIT LOGICAL-WIRE FRONTIERS

Task-dependent inputs. A Hamiltonian-derived logical-qubit graph initializes INPUT embeddings. Its seven node channels summarize coefficient participation, reference occupation, and optional annotations; its seven directed-pair channels summarize co-support and overlap. For example, node i includes $\sum_{\ell: i \in S_\ell} |h_\ell|$ and pair (i, j) includes $\sum_{\ell: \{i,j\} \subseteq S_\ell} |h_\ell|$, where $S_\ell = \mathrm{supp}(P_\ell)$. One edge-conditioned mean-aggregation layer yields task embeddings d_i. This engineered context is lossy, not a complete encoding of H or a quantum state. Appendix A specifies the feature channels and the task-encoder update.

Operation graph. For n wires and L selected operations, the observed-prefix graph contains n INPUT nodes, L distinct OPERATION-occurrence nodes, and n FRONTIER nodes. Each operation receives the previous tail of each operand wire. Edges record the wire and both operand ports; each current wire tail connects to its frontier. Thus there are $2n + L$ nodes and $n + \sum_{\ell=1}^{L} \mathrm{arity}(a_\ell)$ edges. Repeated operations remain separate occurrences. Disjoint operations have no artificial total-order edge. Idle wires retain their INPUT-to-FRONTIER path. Operation nodes encode catalog actions, not native-gate expansions, and parameter slots are marked UNKNOWN rather than using later optimized angles.

Topological encoding. Information propagates in dependency order. In row-vector notation, an operation v with k_v parameter slots receives

$$\begin{aligned} \mu_v &= \sum_{(u,v) \in \mathcal{E}} h_u S_{\mathrm{src}(u,v)} T_{\mathrm{dst}(u,v)}, \\ h_v &= \tanh(F_{\mathrm{op}}(\mu_v + e_{a_v} + k_v e_?) + \mu_v)\,, \end{aligned} \tag{2}$$

where S, T are learned operand-port matrices, F_{op} is affine, and $e_?$ marks an unknown angle. A frontier transforms its tail message by $z_i = \tanh(F_{\mathrm{fr}}(\eta_i) + \eta_i)$, with $\eta_i = h_{\mathrm{tail}(i)} S_{\mathrm{src}(i)}$. Its readout therefore depends on the complete ancestor chain. The encoder returns

$$Z_t = (z_{1,t}, \ldots, z_{n,t}), \qquad g_t = \frac{1}{n} \sum_{i=1}^{n} z_{i,t}. \tag{3}$$

The ordered readouts preserve an explicit interface for individual wires; the mean supports circuit-level decisions. They are neural circuit descriptors, not reduced density matrices. Width is 32, with padding confined to readouts. The action head uses g_t, the frontiers of an action's ordered operands, action identity, and arity/parameter-slot/STOP features, normalized over legal actions.

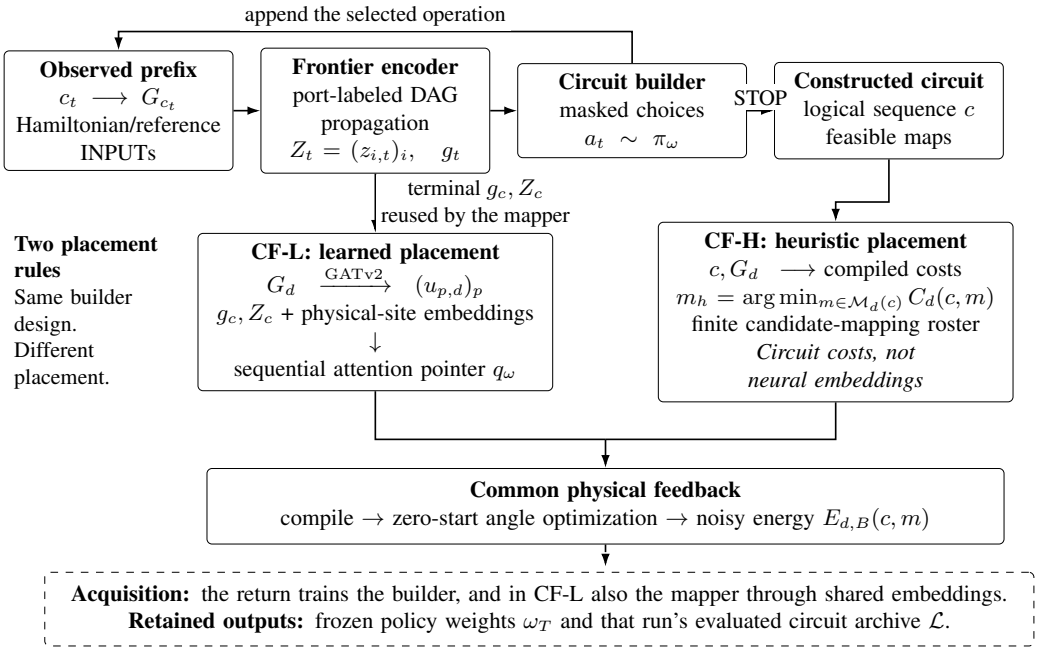


Figure 1: A common frontier-based builder architecture with two separately trained placement variants. CF-L reuses global and wire-indexed embeddings in a learned hardware-attention pointer. CF-H places the constructed circuit using compiled native/assignment costs; it does not consume frontier embeddings. Both use the selected physical pair's energy to train the builder. Each acquisition run retains its own policy and evaluated circuits. The placement alternatives are compared, not combined or selected using evaluation outcomes.

Structural rationale. A label-preserving reindexing of this DAG leaves its exact-arithmetic frontier readouts unchanged: corresponding INPUT states agree, and equality propagates through equal port-labeled predecessor multisets. Appendix A.6 gives the proof and a measured serialization probe. This removes dependence on irrelevant node storage or compatible ordering of disjoint events. It does not imply invariance to arbitrary commuting rewrites, changing masks, or logical-wire permutations. A canonicalized GRU can also remove serialization variation; the empirical encoder control, not invariance alone, determines comparative utility.

3.3 Learned and heuristic placement within the same framework

Learned placement (CF-L). The learned variant factorizes the decisions conditionally:

$$p_\omega(c, m \mid H, G_d) = \prod_t \pi_\omega(a_t \mid c_{<t}, H; \mathcal{A}_d) \prod_i q_\omega(m_i \mid g_c, Z_c, G_d, m_{<i}). \tag{4}$$

The mapper receives the completed circuit readouts; the two factors are not independent. A two-layer GATv2 encodes physical sites and directed couplings. Fifteen raw node/edge descriptors include calibrated errors, durations, readout information, and statistics of the installed noise channels. Concatenating raw, signed-log, and standardized features gives 45 channels. With physical embeddings $u_{p,d}$ and circuit query $q_c = \mathrm{MLP}(g_c)$, the pointer uses

$$\begin{aligned} s_{ip} &= v^\top \tanh(W_q q_c + W_k u_{p,d}) + f_\omega(u_{p,d}, g_c, z_{i,c}, o_{<i}, \xi_{ip,d}), \\ q_\omega(m_i = p) &= \mathrm{softmax}_{p \in \mathcal{F}_i} s_{ip}. \end{aligned} \tag{5}$$

Here $o_{<i}$ represents ordered previous assignments and $\xi_{ip,d}$ includes slot and compiled-plan features. Masks enforce injectivity and extension to a feasible complete map. The model assigns every wire in canonical order within the finite roster; it does not perform unrestricted routing. Appendix A.4 details the GATv2 and pointer computations.

Heuristic placement (CF-H). The alternative retains the same frontier-encoder and builder architecture, but chooses

$$m_h(c,d) = \arg\min_{m \in \mathcal{M}_d(c)} \left[\sum_{g \in \mathcal{P}_d(c,m)} \epsilon_g(d) + \sum_{i=1}^{n} \epsilon^{\mathrm{ro}}_{m_i}(d) \right] \tag{6}$$

with deterministic ties, where $\mathcal{P}_d$ is the compiled native block sequence. The rule depends on the circuit and hardware costs, not the GNN latent vectors or queried mapping energies. It is a deterministic conditional assignment, $q_h(m \mid c, G_d) = \mathbf{1}\{m = m_h(c,d)\}$. Circuit selection is still trained from physical feedback, so this alternative remains hardware-aware despite having only one learned decision policy.

Common training rule. Both variants collect one selected physical pair per episode and use Eq. (1). An initial-context critic supplies a detached, batch-standardized advantage. Four PPO epochs follow each complete 32-episode behavior batch (Schulman et al., 2017). CF-L updates builder and mapper through the shared encoder; CF-H updates only the learned construction path. There is no differentiation through simulation. The learned mapper can optimize against the energy return rather than the additive heuristic score, but that distinction motivates an experiment, not a guarantee of improvement. The exact surrogate, gradient ownership, and collection algorithm are in Appendix A.5.

4 Separating Acquisition from Deployment

An acquisition run a produces final parameters $\omega_{a,T}$ and evaluated trajectories $\mathcal{T}_a$. Retain both

$$\mathcal{O}_a = (\pi_{\omega_{a,T}}, \mathcal{L}_a), \qquad \mathcal{L}_a = \mathrm{Select}_{\mathrm{TRAIN}}(\mathcal{T}_a). \tag{7}$$

The archive contains eight distinct action sequences ranked by best observed TRAIN energy, with deterministic ties. Membership is specific to the parent algorithm, Hamiltonian, exposure, and training block. It is not selected from later-calibration energies. Best-observed ranking also does not estimate mean performance over every TRAIN date; we retain the stated selection rule without an extra reranking experiment.

We evaluate three eight-candidate deployment rules. P_8 generates one greedy and seven sampled circuit–mapping pairs from the frozen policy. Native A_8 maps all eight archived circuits using the parent's frozen, greedily decoded mapper; the native mapper for CF-H is the heuristic. X_8 uses the first four archived circuits, one greedy policy candidate, and three sampled candidates:

$$E(X_8) = \min\{E(\ell_0), \ldots, E(\ell_3), E(g), E(s_0), E(s_1), E(s_2)\}. \tag{8}$$

All pairs receive the same fresh angle-optimization allowance. The four-plus-four split is fixed, not selected per task. It need not outperform either pure mode: taking slots from a branch can omit its best candidate. Evaluating both full eight-candidate modes and returning the better result would instead cost up to sixteen candidates.

A fourth, attribution-only rule, A_8^{heur}, remaps the *same* archive with Eq. (6). This separates three questions. Comparing P_8 between acquisition algorithms measures their deployed policies. Comparing their archives under common heuristic placement holds the deployment rule fixed while varying acquired circuit sets. Comparing native and heuristic placement of one archive holds membership fixed while intervening on the mapper. An archive–policy difference is not a pure causal measure of forgetting: ranking, distinctness, and mapping decoding also differ.

No training history is borrowed across parents. Archive extraction adds no quantum acquisition, but memory, classical selection, and later physical optimization still have costs. The archive comparison uses the same acquired resources in two ways, not an independently hand-designed nonlearning baseline.

5 Experimental Protocol

Tasks and matched acquisition. Table 1 gives the six Hamiltonians. Every task has seven fresh acquisition configurations and three paired training blocks under S1 (May 27, 2026 only) and S10

Table 1: The same task set is used for both training exposures, all structural controls, and all deployment modes. Every instance uses STO-3G, Jordan–Wigner encoding, and no tapering. Water coordinates and inactive-orbital conventions are in Appendix B.

Molecule	Geometries	Active (e, o)	Logical qubits
H_2	$R = 0.7414,\ 1.5$ Å	$(2, 2)$	4
LiH	$R = 2.2,\ 3.4$ Å	$(2, 5)$	10
H_2O	Reference; symmetric $1.25\times$ O–H stretch	$(4, 4)$	8

(May 18–27, round robin). Each policy receives 1,024 acquisition episodes under either exposure, with width 32, a four-action horizon, and native depth/two-qubit caps of 128. The S10 episodes are distributed approximately evenly across its ten dates. A policy is trained separately for each Hamiltonian; this is not zero-shot cross-molecule transfer. Inputs and hyperparameters were fixed after earlier development.

Structural controls. CF-L is the shared-frontier builder with learned hardware-GNN placement; CF-H is its heuristic-placement counterpart. Five further configurations replace the circuit encoder with a GRU, replace hardware message passing with an incidence-feature MLP, separate the circuit encoders, block direct mapper gradients into the shared circuit readouts, or use the hardware MLP's previously selected lower rate. Policy rates are 10^{-3} except for the selected-rate MLP (3×10^{-4}); critic rate is 10^{-3}. Matched-rate MLP tests structure, while the lower-rate version checks a better-configured simpler alternative. Common modules are initialized consistently; independent encoders begin with equal values and distinct storage. No pretrained donor is used. Appendix B.6 specifies each intervention.

Physical evaluation and calibration exposure. Each task has sixteen complete candidate mappings before circuit-specific feasibility. Double-precision density-matrix simulation uses native-gate and assignment/readout errors derived from archived `ibm_fez` calibrations. Thermal relaxation, finite-shot sampling, and live-device execution are excluded. Each selected pair receives zero-start Powell optimization (Powell, 1964) capped at 32 calls. Hardware feature statistics are fitted once on May 27 and held fixed for every model, exposure, and evaluation date.

All frozen policies and archives are evaluated on July 28–30, 2026 as a common three-snapshot evaluation set. These calibrations are excluded from policy training and preprocessing. Evaluation history and the retrospective scope of the archive comparison are documented in Appendix B.5.

Repeated deployment and source matching. Each policy provides one greedy candidate and 32 ordered samples per date. Four P_8 minima use the same greedy candidate and disjoint seven-sample blocks. The X_8 repetitions use the first three samples of each corresponding block and four fixed archive members. Average the four minima within date, then dates within training block. Archive-only deployment supplies one eight-circuit minimum per date, not four independent repeats. All seven acquisition configurations have their own archive and the same four deployment modes; CF-H's native and heuristic archive modes coincide. There is no reference fallback or replacement of a failed selected candidate.

Statistical analysis. Paired task differences are divided by the nonidentity Hamiltonian coefficient L_1 norm, averaged equally over the two geometries in each species, then equally over species. Raw mHa differences remain available by task. Student-t intervals use three independent blocks ($df = 2$). The original 14 CF-L-minus-comparator and eight exposure contrasts retain one Holm22 correction (Holm, 1979); the 20 nominated archive/hybrid contrasts use a separate Holm20 correction. Further same-mapper and task-level views are descriptive. Neither dates nor candidate draws enlarge N (Agarwal et al., 2021).

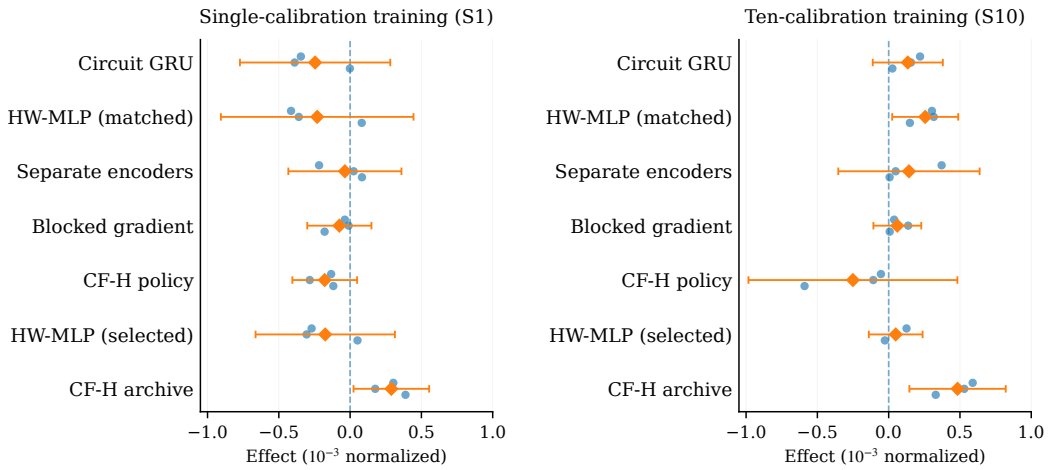


Figure 2: CF-L policy-only differences against every registered comparator. Small circles are the three independent paired blocks; diamonds and whiskers are means and pointwise 95% intervals. Negative favors CF-L. The two panels use the same normalized axis. The CF-H archive is a deployment comparator, not a structural ablation. None of the original contrasts passes Holm22.

6 Results

6.1 Representation effects depend on the training exposure

Under S1, CF-L's aggregate P_8 point estimate is lower than those of all six learned-control configurations. Its differences against GRU, blocked mapper gradients, and CF-H favor CF-L in each of the three observed training blocks. Under S10, the aggregate directions against GRU, hardware MLP, independent encoders, and blocked gradients reverse (Figure 2). The normalized CF-L–GRU means change from -0.245 to $+0.134$ in 10^{-3} units. Individual tasks also differ: the S1 water-reference comparison favors CF-L by 6.091 mHa, whereas S1 stretched LiH favors GRU by 1.881 mHa. Appendix E retains every task mean and contrast.

These are conditional empirical benefits, not a universal graph ranking. None of the 22 registered aggregate contrasts passes Holm correction (minimum adjusted $p = 0.559$). The CF-L exposure effect itself is positive, with an interval spanning zero. The sign changes do not establish an interaction or identify an optimization mechanism, and nonsignificance is not evidence of equivalence. The controlled matrix is nevertheless informative: it prevents a favorable encoder comparison under one exposure from being generalized to all uses of the representation.

6.2 Learned placement and heuristic placement serve different roles

CF-L has lower aggregate P_8 energy than CF-H in every observed block under both exposures. This is a comparison of complete learned policies, whose circuit distributions and mapping rules differ. It cannot attribute the difference to the mapper alone. Table 2 separates that comparison from two interventions available through retained circuits.

Under *common heuristic deployment*, the CF-L archive does not improve the aggregate mean over the CF-H archive. Thus the policy-level advantage is not evidence that learned placement during acquisition discovers better reusable circuit sets. On *the same CF-L archive*, changing only the placement rule favors the heuristic in the aggregate: native-minus-heuristic effects are $+1.611$ and $+1.807$ in 10^{-3} normalized units. The native-archive losses are particularly large on H_2. This observation identifies a deployment-mapping effect, not the cause of the learned pointer's behavior.

The learned strategy remains a principled way to use frontier embeddings for circuit-conditioned assignment, and it yields favorable complete-policy comparisons in specific conditions. The heuristic is nevertheless an effective alternative within the same learned-construction framework. The data do

Table 2: Three distinct placement questions. Effects are paired, species-balanced normalized means in 10^{-3} units, with pointwise 95% intervals. Negative favors the learned-placement choice in each row. The first row belongs to the original 22-test family; the two archive interventions are descriptive comparisons. No row establishes that two learned policies are necessary.

Comparison	S1	S10
CF-L P_8 − CF-H P_8	-0.178 $[-0.405, +0.049]$	-0.251 $[-0.983, +0.482]$
CF-L A_8^{heur} − CF-H A_8^{heur}	$+0.042$ $[-0.095, +0.179]$	$+0.091$ $[-0.176, +0.358]$
CF-L A_8 − CF-L A_8^{heur}	$+1.611$ $[+1.571, +1.652]$	$+1.807$ $[+1.025, +2.588]$

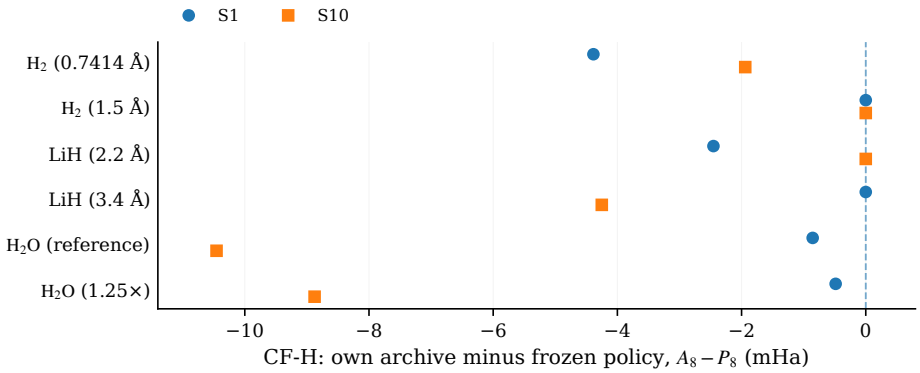


Figure 3: CF-H's own archive minus its frozen policy at the same eight-candidate budget. Points are task means over three dates and three training blocks; no task-level intervals are inferred from these rounded published means. The mapper and acquisition parent are held fixed. All means improve or tie at the displayed precision. Complete policy/control results and exact aggregate blocks are retained in the appendix.

not justify recommending the extra learned mapper for every deployment, or assuming that shared gradients alone improve acquisition.

6.3 Retaining circuits preserves value that a final policy can miss

The clearest same-source retention comparison is CF-H: its archive and its frozen policy share the acquisition history and deployment mapper. The archive lowers task-level mean returned energy in eight of twelve conditions and ties the other four at the published precision (Figure 3). Under S10, the two water reductions are 10.455 and 8.876 mHa. The three aggregate block effects favor the archive under each exposure. This is evidence for retaining useful outcomes of learned construction, not evidence that the acquisition run was unnecessary.

CF-L also acquires reusable circuits, but the deployment mapper changes their usefulness. Its common-heuristic archive means relative to its own P_8 are -0.247 and -0.392 in 10^{-3} normalized units, negative in every observed block. Native archive placement instead yields $+1.364$ and $+1.415$ (Figure 4). Across all seven acquisition configurations, common-heuristic archive means improve on their corresponding policy means under both exposures; native archives improve only three of seven means per exposure. These descriptive comparisons support retention and remapping as separate decisions, not unique frontier superiority.

Combination is not an automatic improvement. The fixed CF-L hybrid lowers the task mean in seven of twelve conditions, including stretched LiH and both water geometries under both exposures. H_2 regressions and the prescribed normalization yield positive $X_8 - P_8$ aggregate effects under both exposures, in all three blocks. No nominated contrast passes Holm20 (minimum adjusted $p = 0.053$). The complete equal-recipe baseline comparison is retained in Appendix D. Neither a favorable task count nor a retrospective better-of-two selector establishes a practical $K = 8$ advantage.

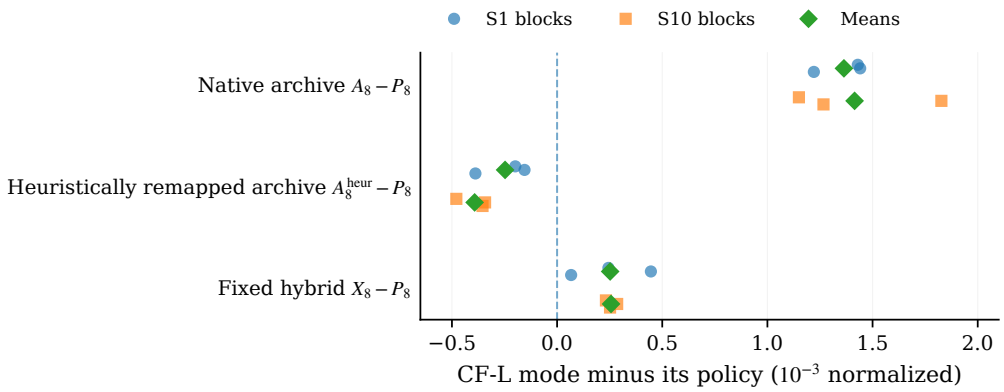


Figure 4: CF-L deployment relative to its own P_8: native archive, the same archive remapped heuristically, and fixed hybrid. Circles/squares are S1/S10 independent blocks; diamonds are means. All modes use eight candidate slots. Common-mapper remapping is an attribution control and is not substituted for the native/hybrid procedure. The change in sign demonstrates why archive membership alone is not a deployment result.

7 Discussion and Conclusion

The circuit frontier supplies a structured construction state and an explicit per-wire interface for learned assignment. This contribution does not require declaring every downstream decision a neural policy. Learned placement and compiled-cost placement are two implementations of a common construction framework, with different empirical strengths. In this study, learned placement has favorable complete-policy directions against heuristic-placement PPO, while heuristic placement gives lower aggregate energies for retained circuit sets. The common-mapper archive comparison does not establish a compensating general acquisition advantage for the learned mapper.

The practical lesson is to evaluate the circuit representation through the whole search process. A final policy, an archive, and a placement rule encode different decisions; none is a sufficient surrogate for the other two. Preserving acquired circuits is useful in the measured conditions, but combining them with policy samples at a fixed budget does not guarantee a better system. Our framework and controlled comparisons expose these distinctions without selecting a winning variant for each task.

The scope is six fixed active-space Hamiltonians, four-action circuits, sixteen candidate mappings per task, and three training blocks. Simulation omits thermal relaxation and finite shots, and the readout-affected objective does not certify state accuracy. The results therefore support a concrete representation and a conditional comparison of search strategies, not universal superiority, unrestricted routing scalability, or a causal explanation of optimization failures. Stronger claims require evidence beyond this bounded regime.

Reproducibility Statement

The anonymous repository[1] provides the experimental implementation, paper configurations, recorded observations, and analysis scripts. It also provides exact numerical exports of the thirteen IBM Fez calibration snapshots used here (May 18–27 and July 28–30, 2026), with the frozen May-27 preprocessing, checksums, and provenance. Its public calibration loader reconstructs the dated hardware graphs and gate/readout noise models without an IBM account or the original provider pickle files. Sections 3–5 and the appendices specify the model and evaluation; Appendix E.4 describes the released inputs. The supplementary numerical snapshot independently reconstructs the reported tables and both statistical families. Repository documentation distinguishes this reconstruction from simulator-based execution: full acquisition additionally requires the molecular task inputs, action schemas, and mapping rosters; re-evaluating the historical frozen policies also requires their checkpoints. These remaining external artifacts are not supplied by the calibration release.

AI Use Statement

Generative AI assisted literature search, hypothesis refinement, mathematical arguments, code implementation/review, and text/figure preparation.

AI agents helped schedule and execute experiments for resource efficiency.

The authors take responsibility for the final manuscript, its scientific claims and reported results, and the accompanying artifacts.

References

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron Courville, and Marc G. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2108.13264.

Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, et al. PennyLane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2018. URL https://arxiv.org/abs/1811.04968.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2105.14491.

Jiawen Dai, Jialiang Gu, Raymond Ho, Gary Man-Tat Man, Kwok Tai Chui, and Kevin Hung. EAQAS: Embedding-aware quantum architecture search via cross-attention fusion and hierarchical representation learning. *EPJ Quantum Technology*, 13:22, 2026. doi: 10.1140/epjqt/s40507-026-00478-y.

Yuxuan Du, Tao Huang, Shan You, Min-Hsiu Hsieh, and Dacheng Tao. Quantum circuit architecture search for variational quantum algorithms. *npj Quantum Information*, 8:62, 2022. doi: 10.1038/s41534-022-00570-y.

Zhimin He, Chuangtao Chen, Zhengjiang Li, Haozhen Situ, Fei Zhang, Shenggen Zheng, and Lvzhou Li. A meta-trained generator for quantum architecture search. *EPJ Quantum Technology*, 11:44, 2024. doi: 10.1140/epjqt/s40507-024-00255-9.

W. J. Hehre, R. F. Stewart, and J. A. Pople. Self-consistent molecular-orbital methods. I. use of Gaussian expansions of Slater-type atomic orbitals. *The Journal of Chemical Physics*, 51(6): 2657–2664, 1969. doi: 10.1063/1.1672392.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. URL https://www.jstor.org/stable/4615733.

IBM Quantum. *IBMBackend: Backend Properties and Calibration Metadata*, 2026. URL https://quantum.cloud.ibm.com/docs/en/api/qiskit-ibm-runtime/ibm-backend. API documentation, accessed September 26, 2026.

[1] https://anonymous.4open.science/r/CF-GNN-RL-QAS/README.md

P. Jordan and E. Wigner. Über das Paulische äquivalenzverbot. *Zeitschrift für Physik*, 47:631–651, 1928. doi: 10.1007/BF01331938.

Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1803.08475.

Zikun Li, Jinjun Peng, Yixuan Mei, Sina Lin, Yi Wu, Oded Padon, and Zhihao Jia. Quarl: A learning-based quantum circuit optimizer. *Proceedings of the ACM on Programming Languages*, 8 (OOPSLA1):114, 2024. doi: 10.1145/3649831.

Matt Lourens, Ilya Sinayskiy, Daniel K. Park, Carsten Blank, and Francesco Petruccione. Hierarchical quantum circuit representations for neural architecture search. *npj Quantum Information*, 9:79, 2023. doi: 10.1038/s41534-023-00747-z.

Prakash Murali, Jonathan M. Baker, Ali Javadi-Abhari, Frederic T. Chong, and Margaret Martonosi. Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. In *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 1015–1029, 2019. doi: 10.1145/3297858.3304075.

Mateusz Ostaszewski, Lea M. Trenkwalder, Wojciech Masarczyk, Eleanor Scerri, and Vedran Dunjko. Reinforcement learning for optimization of variational quantum circuit architectures. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2103.16089.

Yash J. Patel, Akash Kundu, Mateusz Ostaszewski, Xavier Bonet-Monroig, Vedran Dunjko, and Onur Danaci. Curriculum reinforcement learning for quantum architecture search under hardware errors. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2402.03500.

Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014. doi: 10.1038/ncomms5213.

M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2):155–162, 1964. doi: 10.1093/comjnl/7.2.155.

Qiskit Aer Developers. *Qiskit Aer Documentation*, 2026. URL https://qiskit.github.io/qiskit-aer/. Accessed September 26, 2026.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. URL https://arxiv.org/abs/1707.06347.

Yize Sun, Zixin Wu, Volker Tresp, and Yunpu Ma. Quantum architecture search with unsupervised representation learning. *Quantum*, 10:1994, 2026. doi: 10.22331/q-2026-02-03-1994.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17: 261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Hanrui Wang, Yongshan Ding, Jiaqi Gu, Zirui Li, Yujun Lin, David Z. Pan, Frederic T. Chong, and Song Han. QuantumNAS: Noise-adaptive search for robust quantum circuits. In *IEEE International Symposium on High-Performance Computer Architecture*, 2022. doi: 10.1109/HPCA53966.2022.00057. URL https://arxiv.org/abs/2107.10845.

Muhan Zhang, Shali Jiang, Zhicheng Cui, Roman Garnett, and Yixin Chen. D-VAE: A variational autoencoder for directed acyclic graphs. In *Advances in Neural Information Processing Systems*, 2019. URL https://arxiv.org/abs/1904.11088.

A Network Inputs and Computation

This appendix makes the three graph representations explicit. The Hamiltonian graph supplies fixed task descriptors, the operation graph changes as the circuit is constructed, and the hardware graph supplies physical connectivity and calibration. All learned hidden states have width 32. No exact ground-state energy or future optimized angle is an input to the policy.

A.1 Hamiltonian-derived logical-qubit graph

Write $H = h_0 I + \sum_\ell h_\ell P_\ell$, omitting identity from the following statistics, and let $S_\ell = \mathrm{supp}(P_\ell)$. There is one node per logical qubit and a directed edge for every pair $i \neq j$. Thus the stored graph has $n(n-1)$ edges, including pairs with zero attributes. Edges do not indicate only nonzero Hamiltonian interactions. Pair attributes are symmetric, but both orientations are stored.

Table 3: Seven Hamiltonian node channels. Optional annotations are supplied inputs, not inferred energy labels. In the reported molecular adapter the last two channels are zero.

Channel	Definition at logical qubit i	Meaning
1	$\sum_{\ell: i\in S_\ell} \lvert h_\ell\rvert$	Absolute coefficient mass
2	$\sum_{\ell: i\in S_\ell} \lvert h_\ell\rvert/\lvert S_\ell\rvert$	Support-distributed mass
3	$\sum_{\ell: i\in S_\ell} h_\ell$	Signed coefficient mass
4	$\sum_\ell \mathbf{1}\{i \in S_\ell\}$	Term-incidence count
5	b_i	Supplied reference descriptor
6	Sum of supplied symmetry-sector annotations incident on i	Optional symmetry context
7	Sum of supplied allowed-operation support weights incident on i	Optional operation context

For pair statistics, define

$$W_{ij} = \sum_{\ell:\{i,j\}\subseteq S_\ell} |h_\ell|, \qquad D_{ij} = \sum_{\ell:\{i,j\}\subseteq S_\ell} \frac{|h_\ell|}{\binom{|S_\ell|}{2}}, \tag{9}$$

$$N_{ij} = \sum_\ell \mathbf{1}\{\{i,j\} \subseteq S_\ell\}, \tag{10}$$

$$O_{ij} = \sum_{\ell<k:\{i,j\}\subseteq S_\ell\cap S_k} \min(|h_\ell|, |h_k|). \tag{11}$$

Let C_{ij} and A_{ij} restrict the last sum to commuting and anticommuting Pauli pairs, respectively. The seven stored channels are

$$e^H_{ij} = (W_{ij}, D_{ij}, N_{ij}, O_{ij}, C_{ij}, A_{ij}, O^{\mathrm{obs}}_{ij}). \tag{12}$$

All terms in the supported Pauli input are observed, so $O^{\mathrm{obs}}_{ij} = O_{ij} = C_{ij} + A_{ij}$. These redundant channels are retained from the implementation. Terms sharing only one wire do not contribute to these pair-overlap channels, even when they anticommute. Higher-body terms contribute to every pair in their support.

The encoder is one mean-aggregation layer, not a GAT. With learned affine maps F_N, F_E, F_M,

$$u_i = F_N(x^H_i), \qquad d_i = \mathrm{LayerNorm}\left[u_i + F_M\left(\mathrm{ReLU}\left[\frac{1}{n-1}\sum_{j\neq i}\big(u_j + F_E(e^H_{ji})\big)\right]\right)\right]. \tag{13}$$

Raw task features are supplied as floating-point tensors, without the three-channel hardware expansion described below. A zero-feature edge still transmits its neighbor embedding. Affine projection followed by averaging also compresses neighbor–edge associations; this is not a lossless Hamiltonian representation.

These features supply auxiliary task context; the experiments do not isolate their independent performance contribution.

A.2 Operation-graph metadata

The operation graph has $2n + L$ nodes and $n + \sum_{\ell=1}^{L} \mathrm{arity}(a_\ell)$ edges. Parallel dependencies between the same operations are retained when they follow different wires. No extra edge imposes a total order on disjoint operations.

Table 4: Stored circuit records and their use. Categorical IDs select embeddings or matrices rather than being treated as continuous measurements.

Record	Fields	Encoder use
Node	Type, wire, action, occurrence, parameter-slot count, reference occupation	Type selects the update. INPUT uses wire/occupation embeddings; OPERATION uses action and UNKNOWN-slot embeddings. Occurrence identifies the node but is not a numerical position.
Edge	Dependency type, wire, source operand port, destination operand port	Port labels select the source/target transformations; wire identity retains the correct dependency.
Parameter	Owner, slot index, UNKNOWN state	Optimized values are unavailable during construction and placement.
Frontier	One node per logical wire	Canonical wire order supplies the mapper; the live-wire mean supplies the global context. Idle wires keep an INPUT-to-FRONTIER path.

Exact circuit updates and action readout. INPUT state $h_i = \tanh(d_i + e_i^{\mathrm{wire}} + e_{b_i}^{\mathrm{ref}})$ incorporates the task embedding from Eq. (13). Operation and frontier updates are given in Eq. (2); incoming messages are summed in semantic-port order. Distinct source and destination matrices preserve the operand positions on both ends of a dependency. With ordered action support $(i_1, \ldots, i_r)$, the builder's unmasked score is

$$\ell(a) = F_{\mathrm{score}}\left(\tanh\left[F_g(g_t) + e_a + \sum_{p=1}^{r} z_{i_p,t} U_p + F_{\mathrm{static}}(r, k_a, \mathbf{1}_{a=\mathrm{STOP}})\right]\right). \tag{14}$$

$F_g, F_{\mathrm{static}}, F_{\mathrm{score}}$ are affine, and the final map outputs one scalar. The legal-action mask is applied before the categorical softmax. Neither this score nor its inputs include a numerical hardware embedding. The same builder mechanism is used with learned and heuristic placement.

A.3 Hardware descriptors and preprocessing

The molecular hardware graph contains the 156 physical sites in the archived `ibm_fez` context and directed coupling edges. Fifteen raw descriptors are computed for each node and edge. Channel statistics are averaged when several native-gate channels apply to a site or directed pair. Choi/PTM quantities describe the installed calibration-derived simulation channels, not measured process tomography.

Table 5: Raw hardware features in stored order. Durations are seconds. Choi eigenvalues are descending eigenvalues of the normalized Choi state; PTM singular values refer to its traceless block.

Index	Node	Directed edge
1	Mean one-qubit error	Mean two-qubit error
2	Mean one-qubit duration	Mean two-qubit duration
3–8	Choi rank/purity, Kraus rank, unitarity, nonunitality, trace-preservation error	Same six channel descriptors
9–10	First two Choi eigenvalues	First two Choi eigenvalues
11–12	First two PTM singular values	Third/fourth Choi eigenvalues
13	Third PTM singular value	First PTM singular value
14	Readout-assignment error	Second PTM singular value
15	Readout duration	Third PTM singular value

Missing raw values are imputed with TRAIN medians. Entirely missing columns have median zero. For imputed $u \in \mathbb{R}^{15}$, the 45-channel input is

$$x = [u \parallel \operatorname{sgn}(u)\log(1+|u|) \parallel (u-\mu)/\sigma]\,. \tag{15}$$

Statistics are separate for nodes and edges; zero population standard deviation is replaced by one. For the unified experiment, all medians, means and scales are fitted only on May 27, 2026. The same frozen transform is used by both S1 and S10, by every structural control, and by every later evaluation. No July data enter the fit. Compiled-plan tokens are additional features, not extra hardware-graph columns. Neither duration/coherence descriptors nor the token interface imply that thermal relaxation was simulated.

A.4 Hardware attention and the ordered pointer

For one GATv2 head, in column-vector notation,

$$e_{pq}^{(r)} = (a^{(r)})^{\top}\,\mathrm{LeakyReLU}\left(W_1^{(r)}h_p + W_2^{(r)}h_q + W_e^{(r)}x_{qp}\right), \tag{16}$$

$$\alpha_{pq}^{(r)} = \mathrm{softmax}_{q\in\mathcal{N}(p)\cup\{p\}}\, e_{pq}^{(r)}, \qquad \bar{h}_p = \tfrac{1}{4}\sum_{r=1}^{4}\sum_{q}\alpha_{pq}^{(r)}W_2^{(r)}h_q. \tag{17}$$

The first of two layers is followed by ELU; heads are averaged and the final site representation has width 32. This attention aggregates local hardware information. It is separate from logical-to-physical assignment.

Let $q_c = \mathrm{MLP}(g_c)$ and let u_p denote the resulting physical embedding. A circuit-conditioned starting score is $b_p = v^{\top}\tanh(W_q q_c + W_k u_p)$. For a previous assignment (j, m_j),

$$t_j = \mathrm{LayerNorm}\left(W_H u_{m_j} + W_L(z_{c,j} + e_j)\right), \qquad o_i = \sum_{j<i}\lambda_j t_j, \tag{18}$$

where λ_j is additive attention conditioned on q_c and $o_0 = 0$. A feed-forward scorer consumes $u_p, q_c, z_{c,i} + e_i, o_i$, the mean of previously assigned physical embeddings, and five scalar features describing selection/slot status and connectivity, error, and duration relative to assigned sites. A separate compiled-plan projection adds its scalar score. The sum in Eq. (5) is normalized over legal physical choices only. Ordered tokens retain which logical wire occupies each site; averaging only the set of selected physical sites would lose that correspondence.

A.5 Collection and the implemented PPO objective

The same terminal return trains all selected construction and assignment decisions. With initial values v_b, set $z_b = R_b - v_b$, $\mu = |\mathcal{B}|^{-1}\sum_b z_b$, and $s = \sqrt{|\mathcal{B}|^{-1}\sum_b (z_b-\mu)^2}$, lower-bounded numerically. The detached advantage is $\widehat{A}_b = (z_b - \mu)/s$. For each recorded decision j,

$$r_{bj} = \exp[\mathrm{clip}\left(\log p_\omega(a_{bj}) - \log p_{\mathrm{old}}(a_{bj}), -20, 20\right)], \tag{19}$$

$$\mathcal{L}_\pi = -\frac{1}{|\mathcal{B}|}\sum_b\sum_{j\in D_b}\left[\min\{r_{bj}\widehat{A}_b, \mathrm{clip}(r_{bj}, 0.8, 1.2)\widehat{A}_b\} + 0.01\mathcal{H}_{bj}\right], \tag{20}$$

$$\mathcal{L}_V = \frac{1}{2|\mathcal{B}|}\sum_b \big(V_\nu(H_b) - R_b\big)^2. \tag{21}$$

The surrogate sums decisions; it is neither a trajectory-probability ratio nor a length-normalized objective. Four full-batch epochs recompute both phases. One Adam optimizer owns all policy parameters and another owns the critic. Gradient norm clipping is 1.0. The principal policy rate is 10^{-3}; comparison-specific MLP rates are stated in the tables. Rewards are in Ha, without an additional depth penalty, and the terminal task is undiscounted. The initial critic has no completed circuit or current mapping input.

Algorithm 1 Circuit-frontier acquisition and retained outputs

Initialize shared circuit encoder, builder, hardware encoder, pointer, and critic.
for each batch of 32 acquisition episodes **do**
  Freeze behavior weights and assign the prescribed calibration to each episode.
  for each assigned (H, G_d) **do**
    Initialize the reference circuit and its Hamiltonian descriptors.
    while construction has not terminated **do**
      Encode the observed operation DAG; sample a legal next action.
      Append the selected block, or terminate at STOP/the prescribed horizon.
    end while
    if learned placement (CF-L) **then**
      Re-encode the completed circuit; sample assignments from the learned pointer.
    else
      Use the input-only heuristic $m_h(c, d)$ over feasible complete maps.
    end if
    Record (c, m) and the behavior scores of the learned decisions.
    Optimize only this pair's angles under allowance B; record R.
  end for
  Form detached advantages from the complete batch.
  for four epochs **do**
    Replay the recorded decisions; update the shared policy and critic separately.
  end for
end for
Freeze weights and retain the evaluated acquisition history $\mathcal{T}_a$.
Extract eight distinct TRAIN-ranked circuits to form the source-bound archive $\mathcal{L}_a$.
Generate only the candidates required by the declared deployment mode.
Charge fresh bounded angle optimization for every evaluated pair and return the declared statistic.

No optimizer update uses a resampled mapping after its energy is known. Physical jobs can run in parallel, but the learner waits for the complete behavior batch. Native compiler masks and calibration identities are retained during replay.

A.6 Proof and scope of the structural property

Proposition 1 (Conditional reindexing invariance)**.** *Consider two observed-prefix operation DAGs related by a node bijection preserving node types, action and slot labels, logical-wire identities, operand-port edge labels, and INPUT descriptors. With fixed encoder weights, the exact-arithmetic global and ordered frontier readouts are identical. The corresponding builder or mapper distributions also agree when their non-encoder inputs and legality masks agree.*

Proof of Proposition 1. Let φ be the label-preserving node bijection. Corresponding INPUT nodes have identical task, wire, and reference descriptors and therefore identical initialized embeddings. Assume equality at every predecessor of an OPERATION node v. The bijection preserves its incoming edge multiset, logical wires, and both operand ports, so its port-projected message sum equals that of $\varphi(v)$. Action identity and parameter-slot count also agree. Applying the same affine map, residual, and nonlinearity gives $h_v = h_{\varphi(v)}$. A topological induction covers every operation. Corresponding FRONTIER nodes transform equal tail messages through the same map; their canonical logical-wire ordering and mean readout are consequently identical. □

This argument concerns exact arithmetic. The implementation orders message accumulation by semantic ports, avoiding dependence on edge-storage order. Disjoint-operation reorderings preserving this labeled DAG satisfy the premise. General commuting rewrites, changes in compilation, changes in masks, and logical-wire permutations need not satisfy it.

The recorded probe contains 17 disjoint-operation alternatives and 128 storage reindexings. Graph discrepancies are zero in the measured cases. Canonicalizing the GRU serialization also gives zero sensitivity. Natural legality masks change in nine of the seventeen disjoint alternatives, so an invariant

representation alone does not guarantee an identical masked policy. Overlapping-support controls can change the readouts. These tests establish a conditional implementation property, not a universal injectivity or physical-equivalence result.

B Chemistry, Calibration, and Feasible Placement

B.1 Molecular task definitions

All six unified Hamiltonians are neutral singlets in STO-3G with Jordan–Wigner mapping, interleaved alpha/beta spin orbitals, and no tapering. H_2 retains its two spatial orbitals. The LiH study uses two active electrons in five spatial orbitals (ten qubits), with no substitution of smaller LiH active spaces. The two LiH bond lengths are 2.2 and 3.4 Å; they are not different species.

Table 6: Exact H_2/H_2O task definitions and diagnostic energies. Exact energy means the lowest eigenvalue in the specified active-particle/spin-projection sector, not an all-orbital molecular energy.

Molecule	Geometry	Active (e, o)	Qubits	HF (Ha)	Exact sector (Ha)
H_2	$R = 0.7414$ Å	$(2, 2)$	4	-1.116684387	-1.137270175
H_2	$R = 1.5$ Å	$(2, 2)$	4	-0.910873555	-0.998149354
H_2O	Reference	$(4, 4)$	8	-74.964529818	-74.972335608
H_2O	$r_{OH} \times 1.25$	$(4, 4)$	8	-74.889830331	-74.909317116

H_2 atoms are at $(0, 0, 0)$ and $(0, 0, R)$ in Å. The H_2O reference coordinates, in H–O–H order, are $(-0.021, -0.002, 0)$, $(0.835, 0.452, 0)$, and $(1.477, -0.273, 0)$. The stretched coordinates are $(-0.235, -0.1155, 0)$, the same oxygen position, and $(1.6375, -0.45425, 0)$. The O–H vectors are multiplied by 1.25 without changing their angle; these are not newly optimized equilibrium structures.

The water active spatial orbitals are canonical indices $3, 4, 5, 6$, with $0, 1, 2$ inactive (zero-based). Thus six electrons are inactive, not just the oxygen $1s$ pair. Reference occupations are 1100 for H_2 and 11110000 for active-space water. H_2 has 15 Pauli terms and water 193. The nonidentity coefficient norms are respectively 1.885050483, 1.216685783, 6.894991596, and 6.413604625 Ha in table order. For LiH at 2.2 and 3.4 Å, the corresponding norms are 8.872602198 and 8.984296932 Ha. All six full-precision normalization constants are included with the supplementary numerical data; the identity coefficient is excluded in each case.

Table 7: Constants retained in each breadth Hamiltonian. All entries are Ha.

Task	Nuclear repulsion	Inactive electronic	Sum
H_2, 0.7414 Å	0.713753994	0	0.713753994
H_2, 1.5 Å	0.352784807	0	0.352784807
H_2O, reference	9.088300262	-77.914970113	-68.826669851
H_2O, stretch	7.270640210	-76.430206247	-69.159566038

B.2 Action catalog and compilation

The molecular catalog uses reference-valid same-spin SingleExcitation and a globally number-conserving PairCorrelator, a single-excitation/CZ/single-excitation construction. PairCorrelator is not identified with DoubleExcitation away from the reference subspace. Each selected block has one continuous parameter. The four-block horizon, adjacent-identical-action exclusion, and native depth/two-qubit-gate caps of 128 are shared within the main molecular comparisons. Native counts include the actual compiled operations rather than block count alone.

Each H_2/H_2O task has sixteen complete mappings generated from the two retained hardware regions. The input-only rule enumerates compiler-valid pair anchors, completes to the task width by deterministic breadth-first traversal, and retains eight lexicographically ordered mappings per region. It does not truncate a ten-wire LiH mapping or use energies to select a mapping set. Initial legal-action

counts are three for H_2 and eighteen for water. Full-set nonforced assignment counts range from one to three and two to four, respectively; a particular circuit can reduce these further. A larger Hamiltonian therefore does not necessarily induce a larger effective assignment problem.

A prefix is legal only when some declared mapping compiles every block selected so far and remains within the resource constraints. Mapping is sequential but bounded by this complete set. The model is not performing arbitrary SWAP routing or selecting from all injections into the processor.

B.3 Physical conventions and software

The active-space Hamiltonians use the STO-3G basis (Hehre et al., 1969) and untapered Jordan–Wigner encoding (Jordan & Wigner, 1928). The implementation uses PennyLane for quantum-operator definitions (Bergholm et al., 2018), Qiskit Aer for density-matrix simulation (Qiskit Aer Developers, 2026), and SciPy for numerical optimization and statistical calculations (Virtanen et al., 2020). Compilation and circuit-specific feasibility follow the native-plan routines described above, rather than a newly introduced general-purpose routing algorithm. IBM backend-property documentation defines the reported calibration fields (IBM Quantum, 2026); it does not guarantee recovery of the exact historical snapshots.

The released implementation specifies Python 3.12 and pins, among other dependencies, PennyLane 0.44.0, Qiskit 2.3.0, Qiskit Aer 0.17.2, SciPy 1.17.1, PyTorch 2.10.0, and PyTorch Geometric 2.7.0. These are release-environment pins, not evidence of a new full-campaign rerun. The anonymous repository's lockfile and input manifests distinguish the implementation environment, retained observations, and external historical inputs.

B.4 Heuristic placement and libraries

For molecular circuit c, the heuristic chooses

$$m_H(c,d) = \arg\min_{m \in \mathcal{M}_d(c)} \left[\sum_{g \in \mathcal{P}_d(c,m)} \epsilon_g(d) + \sum_{i=1}^{n} \epsilon^{\mathrm{ro}}_{m_i}(d) \right], \tag{22}$$

with lexicographic tie-breaking. $\mathcal{P}_d$ is the compiled native block sequence. This additive score is a placement heuristic, not the energy objective. Optimizing it exactly over a finite set does not make it an energy oracle. The selected pair alone undergoes bounded angle optimization in heuristic-policy training.

A TRAIN library retains eight distinct feasible circuits ranked by permitted TRAIN energy records before the corresponding deployment evaluation. It then applies Eq. (22) with current calibration inputs. The measured unified libraries use only the complete matching frontier–heuristic acquisition run. Their parent task, exposure, training block, checkpoint and acquisition records are retained. No archive is borrowed across parents. The generalized native rule instead uses the frozen mapper belonging to that archive's own parent; all seven parents are evaluated with the native and common-heuristic rules in the completed archive comparison. Its acquired data are charged in standalone cost comparisons even when physically shared between an arm and its library. Candidate diversity is explicit: eight library circuits are distinct, whereas stochastic policies can repeat a circuit or mapping. Repetitions remain valid samples and are not replaced by additional evaluations.

B.5 Calibration split and evaluation history

Table 8: Calibration roles. July 28–30 form one evaluation set for every method; none enters policy training or preprocessing.

Role	Dates in 2026	Use
S1 acquisition	May 27	Single calibration; also the common preprocessing fit
S10 acquisition	May 18–27	One policy and optimizer, with round-robin exposure
Development	July 25–27	Development contexts, excluded from the reported evaluation mean
Evaluation	July 28–30	Frozen policies and fixed archive membership; current-calibration placement and angle optimization

S10 assigns 103 of the 1,024 episodes to each of the first four TRAIN dates and 102 to each remaining date. No evaluation calibration contributes to the feature-statistics fit.

July 28/29 results had been inspected during earlier development, whereas July 30 was reported uninspected before the unified evaluation; all three were observed before the retrospective archive analysis. The reported evaluation set is therefore held out from policy training, not uniformly uninspected during method development.

B.6 Structural controls

In the detailed tables, Full denotes CF-L and Frontier–heuristic (or Heuristic) denotes CF-H; these names refer to the same acquisition configurations throughout. The Full model shares one circuit encoder. The independent-encoder control initializes two equal copies with distinct parameter storage; construction uses one and mapping the other. The blocked-gradient control keeps the shared values but detaches floating circuit readouts only on mapper-loss paths. It does not eliminate the dependence of construction rewards on mapping. The circuit-GRU control supplies the same global and per-wire output interfaces from its sequential encoder. The hardware-incidence MLP retains physical features and incidence information but replaces hardware message passing. The two MLP learning rates are separate configurations: 10^{-3} is matched; 3×10^{-4} is the previously selected optimization-robustness setting. These controls do not inherit a heuristic-trained encoder or optimizer state.

Heuristic-placement PPO uses the same circuit-frontier builder with Eq. (22) at training and deployment. Its archive therefore compares two outputs of the same acquisition process under the same mapping rule. All seven configurations receive 1,024 selected-pair episodes, batch 32, four full-batch PPO epochs, width 32, and identical within-task circuit and mapping constraints. Distinct architectures do not necessarily have the same parameter count or classical compute cost; no matched-runtime claim is made.

C Exact Deployment Indexing and Failure Handling

For a parent configuration a, let g denote its greedy candidate and s_j its zero-indexed sampled candidates. Archive circuit ℓ_j is ranked by TRAIN observations, not destination score. Four policy repetitions use the sets

$$\mathcal{C}_r^P = \{g, s_{7r}, s_{7r+1}, \ldots, s_{7r+6}\}, \qquad r = 0, 1, 2, 3.$$

The primary policy value is the average of the four returned minima; all 32 samples also define a separate candidate mean. In the native archive mode, ℓ_j is assigned by the parent's frozen greedy mapper on the current context; for the heuristic parent this equals the common input-only heuristic. The fixed mixed sets are

$$\mathcal{C}_r^X = \{\ell_0, \ell_1, \ell_2, \ell_3, g, s_{7r}, s_{7r+1}, s_{7r+2}\}.$$

Archive membership is identical across repetitions. Repeating it does not create additional acquisition seeds or independent archive measurements. A common-heuristic archive provides an attribution control using the same eight circuits under Eq. (22). A circuit can acquire different mappings and optimized angles on different calibrations, while its logical sequence remains frozen.

Selection failures and costs. An infeasible candidate keeps its selected identity and cannot be resampled into a favorable case. An unresolved numerical failure does not become a valid returned minimum. The original protocol has no bare-reference fallback for the primary output. Physically reusing an exactly matching deterministic calculation can reduce new work, but its logical candidate/query cost remains charged. Candidate identity includes Hamiltonian, calibration, circuit, mapping, compilation, noise, optimizer initialization and budget, rather than circuit structure alone. All repeated policy samples, including duplicates, remain in their originally assigned positions.

D Complete Source-Matched Archive Comparison

All 252 acquisition parents have eight-member archives and four dated deployment modes: policy-only, native archive, common-heuristic archive, and fixed four-plus-four deployment. The resulting

3,024 dated mode observations are complete. Native and common archives coincide by definition for the heuristic-placement parent; these aliases are not independent physical evaluations. Policy and hybrid repetitions are averaged within date, then dates within each of the three independent training blocks.

D.1 Retention under the same parent and the same deployment recipe

Table 9 shows every acquisition parent. Its native-archive and common-map columns concern the same selected archive membership. The latter changes placement to the common input-only heuristic. For Full, common-map archive-minus-policy effects are -0.247 and -0.392 in 10^{-3} normalized units. This does not authorize replacing native A_8 or X_8 after seeing results; it is the declared attribution control.

Table 9: Every parent's own-mode minus its own-policy effect, in 10^{-3} dimensionless normalized units. Negative favors the mode over P_8. Values are descriptive means; no additional family of tests is created.

	S1			S10		
Parent	$A_8 - P_8$	$A_8^{\mathrm{heur}} - P_8$	$X_8 - P_8$	$A_8 - P_8$	$A_8^{\mathrm{heur}} - P_8$	$X_8 - P_8$
Full	+1.364	-0.247	+0.252	+1.415	-0.392	+0.257
GRU	+1.086	-0.528	-0.167	+1.453	-0.381	+0.212
HW-MLP (matched)	-0.050	-0.484	-0.118	-0.061	-0.164	-0.029
Separate encoders	+1.324	-0.284	+0.177	+1.657	-0.312	+0.294
Blocked gradient	+1.287	-0.323	-0.070	+1.435	-0.394	+0.334
Frontier–heuristic	-0.467	-0.467	-0.421	-0.734	-0.734	-0.467
HW-MLP (selected)	-0.209	-0.413	-0.242	-0.285	-0.442	-0.249

Table 10: Archive acquisition controls: Full archive minus comparator archive under the same deployment rule, in 10^{-3} normalized units. Native uses each parent's own mapper; common holds the heuristic rule fixed. These are descriptive comparisons, not new superiority tests.

	S1		S10	
Comparator	Native A_8	Common A_8^{heur}	Native A_8	Common A_8^{heur}
GRU	+0.032	+0.035	+0.095	+0.123
HW-MLP (matched)	+1.183	+0.006	+1.732	+0.028
Separate encoders	+0.003	-0.000	-0.101	+0.063
Blocked gradient	+0.001	-0.000	+0.040	+0.063
Frontier–heuristic	+1.653	+0.042	+1.898	+0.091
HW-MLP (selected)	+1.398	-0.009	+1.749	+0.099

D.2 All individual Full task–block effects

The paired block tables preserve the magnitude of individual improvements rather than treating a favorable task mean as three successful replications. In stretched LiH, for example, the large hybrid improvement occurs in block 3; blocks 1 and 2 tie or have small positive effects. The normalization and geometry/species weighting remain fixed throughout.

Table 11: Full S1: all paired task effects in mHa. Columns 1–3 refer to the same independent training blocks across modes. Negative favors the left-hand mode.

	$A_8 - P_8$			$X_8 - P_8$		
Task	1	2	3	1	2	3
H_2 (0.7414 Å)	+9.337	+8.458	+9.042	+1.768	+2.539	+1.253
H_2 (1.5 Å)	+5.726	+5.451	+5.207	+1.814	+1.937	+1.810
LiH (2.2 Å)	-1.064	-1.356	-1.512	+0.000	+0.692	-1.512
LiH (3.4 Å)	+0.000	+0.279	-5.642	+0.000	+0.054	-5.642
H_2O (base)	-6.647	+1.007	-2.661	-6.647	+0.000	-2.661
H_2O (1.25×)	+0.020	-2.211	-3.638	+0.000	-2.225	-3.638

Table 12: Full S10: all paired task effects under the same definitions and units.

	$A_8 - P_8$			$X_8 - P_8$		
Task	1	2	3	1	2	3
H_2 (0.7414 Å)	+7.243	+14.363	+7.890	+1.362	+2.405	+0.677
H_2 (1.5 Å)	+5.207	+4.287	+5.431	+1.068	+1.321	+1.870
LiH (2.2 Å)	+2.164	+2.759	-1.224	+0.454	+0.000	+1.013
LiH (3.4 Å)	+0.064	+0.246	-5.674	+0.016	+0.045	-5.674
H_2O (base)	-5.404	+0.001	-1.698	-3.672	+0.000	+1.156
H_2O (1.25×)	-4.430	-3.324	-0.212	+1.750	-4.188	-0.232

D.3 The nominated retrospective family

The 20 archive/hybrid contrasts are retrospective exploratory comparisons, with evaluation history described in Appendix B.5. They are corrected separately by Holm; the original 22-test family remains unchanged. Tables 13–14 report all nominated effects. The smallest adjusted p is 0.053; no corrected result passes 0.05.

Table 13: All ten S1 addendum contrasts. Means and pointwise 95% Student-t intervals are in 10^{-3} normalized units; $df = 2$. Corrected p values use all 20 S1/S10 nominations.

Contrast	Mean	Pointwise 95% interval	Holm20 p
$A_8^F - P_8^F$	+1.364	$[+1.056, +1.672]$	0.053
$X_8^F - P_8^F$	+0.252	$[-0.219, +0.724]$	0.776
$X_8^F - A_8^F$	-1.112	$[-1.365, -0.858]$	0.053
$X_8^F - A_8^h$	+0.541	$[+0.334, +0.749]$	0.118
$X_8^F - X_8^{\mathrm{GRU}}$	+0.174	$[-0.306, +0.654]$	0.956
$X_8^F - X_8^{\mathrm{MLP-m}}$	+0.139	$[-0.100, +0.378]$	0.776
$X_8^F - X_8^{\mathrm{Sep.}}$	+0.038	$[-0.351, +0.428]$	1.000
$X_8^F - X_8^{\mathrm{Block}}$	+0.247	$[+0.010, +0.483]$	0.323
$X_8^F - X_8^h$	+0.495	$[+0.287, +0.703]$	0.131
$X_8^F - X_8^{\mathrm{MLP-s}}$	+0.320	$[+0.249, +0.391]$	0.053

Table 14: All ten S10 addendum contrasts with the same inference and units.

Contrast	Mean	Pointwise 95% interval	Holm20 p
$A_8^F - P_8^F$	+1.415	$[+0.515, +2.314]$	0.233
$X_8^F - P_8^F$	+0.257	$[+0.189, +0.324]$	0.062
$X_8^F - A_8^F$	-1.158	$[-1.992, -0.324]$	0.269
$X_8^F - A_8^h$	+0.740	$[+0.400, +1.080]$	0.146
$X_8^F - X_8^{\mathrm{GRU}}$	+0.179	$[+0.078, +0.280]$	0.200
$X_8^F - X_8^{\mathrm{MLP-m}}$	+0.542	$[+0.400, +0.684]$	0.062
$X_8^F - X_8^{\mathrm{Sep.}}$	+0.105	$[-0.168, +0.378]$	0.956
$X_8^F - X_8^{\mathrm{Block}}$	-0.017	$[-0.910, +0.877]$	1.000
$X_8^F - X_8^h$	+0.473	$[+0.117, +0.828]$	0.269
$X_8^F - X_8^{\mathrm{MLP-s}}$	+0.555	$[+0.130, +0.979]$	0.269

Computation for the original cohort and archive evaluation is summarized together in Appendix E.4.

E Complete Unified Measurements

E.1 All policy and measured archive task means

Tables 15 and 16 report all 96 task–exposure–output means from the unified study. Energy entries are reported to 10^{-6} Ha. The Full column and all structural controls are policy-only P_8 scores. The final column is the archive of the frontier–heuristic parent. It is not an archive of Full. The corresponding source-matched comparison therefore uses the Heur. column, not Full. The policy/source-heuristic comparisons use this published precision. The completed Full archive and hybrid differences in Appendix D instead use full-precision paired task effects.

Table 15: Single-calibration training (S1): mean returned energy in Ha. Each policy mean includes three dates and three independent training blocks.

Task	Full	GRU	MLP-m	Sep.	Block	Heur.	MLP-s	A_8^h
H_2 (0.7414 Å)	-1.095007	-1.094917	-1.091649	-1.094804	-1.094790	-1.092104	-1.094425	-1.096490
H_2 (1.5 Å)	-0.968898	-0.968502	-0.968719	-0.968610	-0.968682	-0.968719	-0.968719	-0.968719
LiH (2.2 Å)	-7.799228	-7.799041	-7.797493	-7.799029	-7.799228	-7.797493	-7.799398	-7.799946
LiH (3.4 Å)	-7.703654	-7.705535	-7.705520	-7.703629	-7.703654	-7.705353	-7.702493	-7.705353
H_2O (base)	-74.902703	-74.896612	-74.904383	-74.902387	-74.901922	-74.905760	-74.900541	-74.906615
H_2O (1.25×)	-74.839700	-74.837106	-74.841542	-74.840956	-74.839392	-74.840869	-74.838624	-74.841354

Table 16: Ten-calibration training (S10), with the same acquisition and deployment allowances. Sep. means separate circuit encoders; Block means mapper-gradient blocking; MLP-m/s are matched/selected rates.

Task	Full	GRU	MLP-m	Sep.	Block	Heur.	MLP-s	A_8^h
H_2 (0.7414 Å)	-1.093650	-1.094517	-1.094849	-1.094387	-1.094003	-1.094549	-1.094233	-1.096490
H_2 (1.5 Å)	-0.968411	-0.968570	-0.968768	-0.968848	-0.968426	-0.968719	-0.968719	-0.968719
LiH (2.2 Å)	-7.798127	-7.798746	-7.797493	-7.798209	-7.798498	-7.797493	-7.798130	-7.797493
LiH (3.4 Å)	-7.703643	-7.705374	-7.705509	-7.703628	-7.703654	-7.701102	-7.702219	-7.705353
H_2O (base)	-74.903350	-74.905504	-74.906894	-74.903831	-74.903269	-74.896160	-74.900155	-74.906615
H_2O (1.25×)	-74.837265	-74.834951	-74.836992	-74.837435	-74.838123	-74.831899	-74.839547	-74.840775

E.2 Source-matched retention over independent blocks

Let D_b^{F-h} be the original Full-minus-heuristic-policy normalized effect in block b, and $D_b^{F-A_h}$ the Full-minus-heuristic-archive effect. Linearity of the original within-block weighting yields

$$D_b^{A_h-h} = D_b^{F-h} - D_b^{F-A_h}.$$

Table 17 reports this exact algebraic reanalysis of the full-precision aggregate records. It is a descriptive, post-hoc contrast, not an extra member retrospectively added to the original 22-test family. All three values have the same observed direction under each exposure; S1 and S10 are paired conditions, not six independent replications of one exposure.

Table 17: Archive-minus-its-source-policy effects, in 10^{-3} units of the dimensionless normalized energy difference. No additional significance claim is attached.

Exposure	Block 1	Block 2	Block 3	Mean
S1	-0.436115	-0.459276	-0.505985	-0.467126
S10	-0.643621	-0.638235	-0.919676	-0.733844

E.3 THE COMPLETE ORIGINAL INFERENCE FAMILY

For each task t with nonidentity coefficient scale $L_t = \sum_{\ell \neq I} |h_{t,\ell}|$, define $d_{t,b}$ as the date-averaged paired energy difference for block b. For species s with its two geometries $t \in s$, the aggregate block effect is

$$D_b = \frac{1}{3} \sum_{s \in \{\mathrm{H_2}, \mathrm{LiH}, \mathrm{H_2O}\}} \frac{1}{2} \sum_{t \in s} \frac{d_{t,b}}{L_t}. \tag{23}$$

The inferential unit is the three-element vector (D_1, D_2, D_3). Pointwise intervals use $\bar{D} \pm t_{2,.975} s_D / \sqrt{3}$. The original family tests seven Full-minus-comparator contrasts in each exposure and all eight within-method exposure contrasts. Ordered p values are adjusted by Holm's step-down rule over all 22. An eligible favorable effect additionally requires finite complete blocks, positive variance, and negative mean. All effects in this family are nonconstant; none has adjusted $p < .05$. Nonsignificance is not equivalence, and a descriptive adverse mean is not a corrected superiority result for the comparator.

Table 18: S1: original Full-minus-comparator aggregate effects, 10^{-3} normalized units. Intervals are pointwise, while decisions use the complete Holm22 family.

Comparator	Mean	Pointwise 95% interval	Holm p
GRU	-0.245	$[-0.772, +0.282]$	1.000
HW-MLP (matched)	-0.231	$[-0.906, +0.445]$	1.000
Separate encoders	-0.037	$[-0.433, +0.360]$	1.000
Blocked gradient	-0.076	$[-0.301, +0.150]$	1.000
Frontier–heuristic	-0.178	$[-0.405, +0.049]$	1.000
HW-MLP (selected)	-0.174	$[-0.663, +0.314]$	1.000
Heuristic archive	+0.289	$[+0.024, +0.554]$	0.828

Table 19: S10: the same aggregate contrasts. The heuristic-archive row is the original library comparator.

Comparator	Mean	Pointwise 95% interval	Holm p
GRU	+0.134	$[-0.112, +0.380]$	1.000
HW-MLP (matched)	+0.256	$[+0.025, +0.488]$	0.828
Separate encoders	+0.142	$[-0.353, +0.638]$	1.000
Blocked gradient	+0.061	$[-0.107, +0.229]$	1.000
Frontier–heuristic	-0.251	$[-0.983, +0.482]$	1.000
HW-MLP (selected)	+0.049	$[-0.139, +0.238]$	1.000
Heuristic archive	+0.483	$[+0.145, +0.821]$	0.559

Table 20: Original S10-minus-S1 effects for every evaluated output, in the same 10^{-3} normalized units. Negative favors broader exposure.

Acquisition / output	Mean	Pointwise 95% interval	Holm p
Full	+0.255	$[-0.337, +0.848]$	1.000
GRU	-0.124	$[-0.370, +0.121]$	1.000
HW-MLP (matched)	-0.232	$[-0.612, +0.148]$	1.000
Separate encoders	+0.076	$[-0.766, +0.919]$	1.000
Blocked gradient	+0.119	$[-0.566, +0.804]$	1.000
Frontier–heuristic	+0.328	$[+0.043, +0.612]$	0.806
HW-MLP (selected)	+0.031	$[-0.174, +0.236]$	1.000
Heuristic archive	+0.061	$[-0.004, +0.126]$	1.000

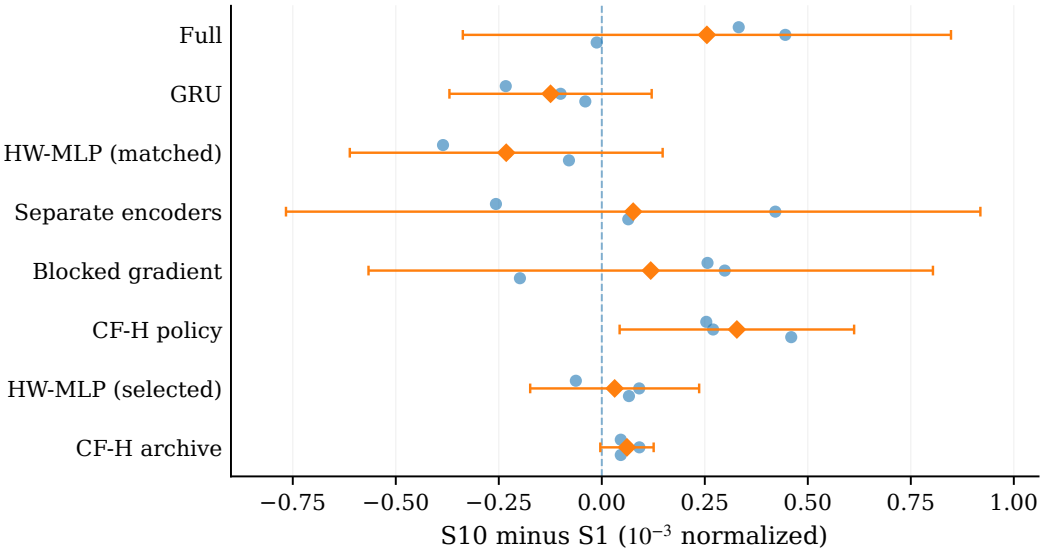


Figure 5: Exposure effects for all eight original outputs. Circles are independent-block differences, diamonds are means, and bars are pointwise 95% intervals. Calibration dates and candidate draws remain repeated measurements within each block.

E.4 Computation and reproducibility

The original cohort comprises 252 trained endpoints and 36 heuristic-parent archives, all evaluated on three dates. Recorded training and deployment use 7,134,578 and 673,825 objective requests, respectively: 7,808,403 requests and 7,835,831 simulator executions in total, with no failed or unfinished executions. "Physical attempts" in the records denote simulator executions, not live-device runs.

The source-matched addendum contains 12,096 logical archive candidate slots and 1,594 distinct physical identities. Of these, 985 reuse paid evaluations and 609 require new parameter optimization, using 18,749 objective requests and 20,129 simulator executions with no failures. Standalone archive deployment is charged its parent's acquisition once; shared campaign accounting does not duplicate it. Reuse does not allow more than eight candidate slots per deployment. Equal query allowances do not establish equal classical runtime, and summed worker time is not elapsed campaign time.

Numerical exports in `data/calibrations/` preserve the dated native gate facts, ordered graph features, installed channels, and May-27 preprocessing. Their manifest authenticates distributed bytes; raw-provider hashes record provenance only. The public loader needs no original provider files. Remaining molecular-input and historical-checkpoint requirements are documented in `docs/reproduction.md` and `data/README.md`. The numerical supplement reconstructs statistics without rerunning simulations.