Parameter-performance scaling of hybrid quantum-classical policies for demand-charge-aware battery control

Eunsung Oh[1],*

[1] Department of Electrical Engineering, Gachon University, Gyeonggi-do 13120, South Korea
* Corresponding author: esoh@gachon.ac.kr

Supplementary Information

Supplementary Note 1. Cohort filtering and building-level split

This note documents the construction of the 549-building cohort and the locked building-level split. The project-local manifest starts from the previously screened cohort; therefore intermediate exclusions before the 549-building source manifest are recorded as inherited screening decisions rather than recalculated counts. The screening ledger (Table S1), the split balance summary (Table S2), the split manifest schema (Table S3) and the cohort distribution figure (Figure S1) together characterize the locked cohort.

Table S1. Cohort filtering ledger. Counts before the final screened cohort are inherited from the source screening manifest; the project-local audit verifies the 549-building manifest, split and file resolution.

Filtering step	Criterion	Remaining buildings	Excluded buildings	Notes/source field
Dataset release	ComStock 2025 Release 2, AMY2018	549 in source-screened manifest	Inherited	Source manifest SHA256 prefix 8ec2b4b757ae.
Geography	Los Angeles County	549	Inherited	County field retained in split manifest.
Climate zone	ASHRAE climate zone 3B	549	Inherited	ASHRAE/CEC climate fields retained.
Building type	Office: MediumOffice, SmallOffice and LargeOffice	549	Inherited	Type counts: 334/170/45.
Vintage	Vintage 1980 or later	549	Inherited	Vintage categories retained in manifest.
Tariff screening	Standardized PG&E B-10 scenario screen; native summer peak 75.61–497.95 kW	549	0 within project-local manifest	Source screening manifest defines the screen.
Time-series quality	Resolved June–September 15-minute load profiles under read-only bldg_data	549	0	All paths resolved; timestamp and accounting audits passed.
Final cohort	All filters passed	549	–	Building IDs stored in artifacts/data/split_manifest.csv.

Table S2. Split balance summary. Distribution cells are median [IQR]. Battery energy follows the 2-hour sizing rule and is included because floor area is not available in the frozen project-local split manifest.

Split	Buildings	Months/building	Episodes	Native summer peak kW	Summer kWh	Battery energy kWh
Train	329	4	1,316	162.9 [113.6, 239.5]	484,622.2 [309,975.0, 775,466.7]	81.4 [56.8, 119.7]
Validation	110	4	440	167.0 [119.9, 246.2]	470,538.9 [304,409.0, 754,507.6]	83.5 [60.0, 123.1]
Locked test	110	4	440	169.4 [113.9, 239.8]	487,916.7 [332,204.9, 740,356.2]	84.7 [56.9, 119.9]

Total	549	4	2,196	165.4 [114.4, 241.1]	484,622.2 [310,966.7, 762,488.9]	82.7 [57.2, 120.6]

Table S3. Split manifest schema and realized project-local columns.

Column	Type	Description	Status
bldg_id	integer	Unique ComStock building identifier.	Present
split	categorical	train, validation or test.	Present
in.county_name	string	County label used in inherited screen.	Present
in.comstock_building_type	categorical	MediumOffice, SmallOffice or LargeOffice.	Present
in.vintage	categorical	Vintage category retained from ComStock metadata.	Present
peak_kw	numeric	Maximum native summer load used for stratification and sizing.	Present
out.electricity.total.energy_consumption..kwh	numeric	Native annual electrical energy field used as the energy-scale balance variable.	Present
battery_power_kw, battery_energy_kwh	numeric	Battery power and energy from the frozen sizing rule.	Present
timeseries_path_resolved	path	Resolved read-only time-series path under bldg_data.	Present
split_seed	integer	Stratified split seed.	Present

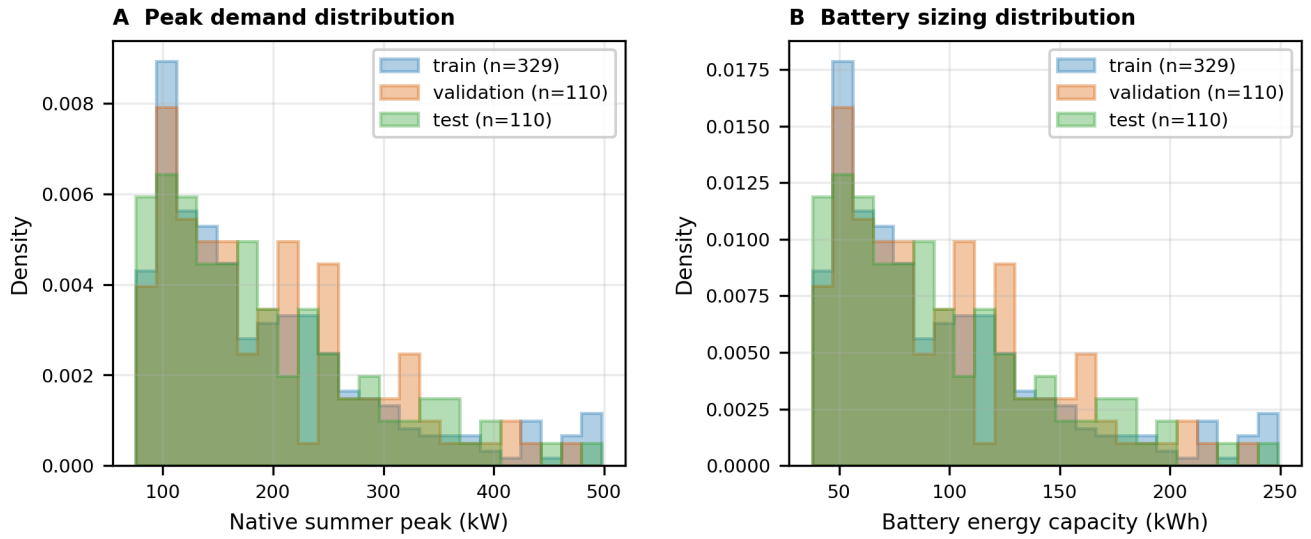

Figure S1. Cohort and split balance diagnostics. **A.** Native summer peak (kW) distribution per split. **B.** Battery energy capacity (kWh) distribution per split (size set by the 0.25-peak / 2-hour sizing rule). The three splits overlap closely across both axes, consistent with the stratified sampling described in Table S2. Source data: artifacts/data/split_manifest.csv (SHA256 prefix bb86e5e10df4).

Supplementary Note 2. Tariff and battery accounting

The standardized PG&E B-10 tariff constants (Table S4), the battery sizing and projection rules (Table S5) and the accounting audit (Table S6) together define the deterministic environment used throughout the locked test.

Table S4. Standardized PG&E B-10 tariff constants used in the benchmark.

Tariff component	Definition	Value	Notes
Summer season	Months included in primary study	June–September	Summer-only benchmark.
Peak period	Daily summer peak window	16:00–21:00 PST	All days in the scenario.
Part-peak period	Daily summer part-peak windows	14:00–16:00; 21:00–23:00 PST	All days in the scenario.
Off-peak period	All other intervals	All non-peak and non-part-peak hours	–

Energy price, peak	Summer peak energy price	0.33947 USD/kWh	Secondary bundled rate, effective 2026-03-01.
Energy price, part-peak	Summer part-peak price	0.27778 USD/kWh	Same snapshot.
Energy price, off-peak	Summer off-peak price	0.24522 USD/kWh	Same snapshot.
Demand charge	Monthly maximum-demand charge	20.50 USD/kW-month	Highest 15-minute average import demand.
Demand reset	Running peak initialization	$m_0 = 0$ each month	Primary accounting rule.
Voltage class	Selected B-10 voltage class	Secondary bundled	Fixed scenario, not native utility assignment.

Table S5. Battery sizing and control-constraint configuration.

Configuration item	Definition or formula	Value	Notes
Power sizing	Battery power rating as function of building scale	$P_b = 0.25\max_{t\in\text{summer}}\ell_{b,t}$	Native summer peak.
Energy sizing	Battery energy rating or duration	$E_b = 2P_b$	2-hour duration.
Charge efficiency	Interval charging efficiency	0.96	One-way efficiency.
Discharge efficiency	Interval discharging efficiency	0.96	One-way efficiency.
SoC lower/upper bounds	Feasible usable state of charge	0 and 1	Fraction of usable capacity.
Initial SoC	Start-of-month SoC	0.5	Same for all policies.
Terminal treatment	Month-end SoC correction	Economic terminal settlement	Same evaluator path for all policies.
No-export projection	Constraint on grid import	$g_t \geq 0$	Applied after action clipping and SoC projection.
Degradation proxy	Charge plus efficiency-adjusted positive discharge energy	0.0596 USD/kWh	Applied after projection.

Table S6. Tariff and battery accounting audit.

Audit item	Test performed	Result
15-minute alignment	Timestamp boundary and synthetic boundary checks for June–September mapping.	Passed: 12/12 manifest-month checks and 27/27 synthetic checks.
Monthly reset	Confirm demand running peak is reset at each month boundary.	Passed through no-storage telescoping audit across 2,196 episodes.
Demand charge telescoping	Check $\sum_t D\,(m_t - m_{t-1}) = Dm_T$ numerically.	Passed; maximum absolute difference 0.0 USD.
No-export projection	Verify post-projection import remains non-negative in projection/degradation tests.	Passed: 6/6 edge cases and 72/72 manifest samples.
SoC bounds	Verify projected trajectories respect SoC bounds in projection/degradation tests.	Passed in battery projection audit; oracle/environment smoke failures 0/9.
Terminal correction	Verify terminal SoC treatment is common across evaluator paths.	Economic terminal settlement; oracle/environment objective max absolute difference 1.82×10^{-11} USD.
Rate constants	Check tariff and demand constants against frozen config.	Snapshot effective 2026-03-01; constants in artifacts/configs/tariff_constants.json.

Supplementary Note 3. Objective, regret and frontier source data

The objective decomposition (Table S7), the regret-denominator handling (Table S8) and the dense-frontier implementation details (Table S9) document how the headline regret, bill-saving and degradation axes are constructed.

Table S7. Objective components and units.

Component	Formula/source	Unit	Reporting convention
Energy cost	$\sum_t \pi_t \, g_t \Delta t$	USD/month	Positive cost, lower better.
Demand cost	Dm_T	USD/month	Positive cost, lower better.
Bill component	Energy plus demand cost	USD/month	Raw monthly cost and bill saving from no storage.
Degradation component	$0.0596 \sum_t \left(q_t^{\mathrm{chg}} + q_t^{\mathrm{dis}}/\eta_{\mathrm{dis}}\right)$	USD/month	Positive cycling cost; plotted as negative degradation cost.
Scalar objective	$(1-w)C_{\mathrm{bill}} + wC_{\mathrm{deg}}$	USD/month	Lower better; no component standardization before scalarization.
Bill-saving axis	No-storage bill minus policy bill	USD/month	Higher better.
Negative-degradation axis	$-C_{\mathrm{deg}}$	USD/month	Higher better, equivalent to lower degradation cost.

Table S8. Normalized-regret denominator handling. The same validation-metric configuration is used for validation and locked-test replay.

Split	Primary preference grid	Denominator threshold	Valid primary-grid episodes	Handling
Train	$w \in \{0.0,0.3,0.4,0.8,0.9\}$ for labels	1 USD	Used for BC labels, not for selection aggregation	Event-weighted BC; no normalized-regret exclusion needed for fitting.
Validation	$w = 0.0,0.1,\ldots,0.9$	1 USD	Full validation replay uses all valid opportunity episodes	Episodes with $J^0 - J^\star \leq 1$ USD are excluded from normalized-regret means.
Locked test	$w = 0.0,0.1,\ldots,0.9$	1 USD	4,400/4,400 primary-grid building-month-weight cases per cell	No locked-test model changes; exclusion rule fixed before replay.

Table S9. Dense-grid frontier implementation details.

Item	Definition	Frozen implementation
Preference grid	Weights used to generate frontier points	$\{0.0,0.1,\ldots,0.9\}$ for validation and locked test.
Axes	Bill-saving and degradation dimensions	$x =$ bill saving vs no storage (USD/month); $y =$ negative degradation cost (USD/month); both higher better.
Reference point	Dominated point for qualitative frontier comparison	The no-storage/degradation-free origin and oracle frontier are shown; locked inference focuses on macro regret and paired contrasts rather than a single HV pass/fail.
Normalization	Oracle-relative regret denominator	$J^0 - J^\star$, with $\epsilon = 1$ USD opportunity threshold.
Filtering	Dominated and monotonicity diagnostics	Dominated-point counts and monotonicity-bad segments are reported in paper/source_data/s8_locked_test_frontier_shape.csv.
Source data	CSV columns for frontier calculation	paper/source_data/s8_locked_test_dense_frontier.csv; columns include model series, budget, preference weight, bill saving, negative degradation, regret and seed count.

Supplementary Note 4. Model family definitions and architecture cards

This note collects the family definitions (Table S10), the full locked-test parameter grid (Table S11), the hybrid VQC architecture card (Table S12) and the inference-time isotonic wrapper card (Table S13).

Table S10. Model-family definitions.

Family	Architecture definition	Main purpose	Placement
MLP-Cap	Feed-forward neural policy with a learned demand-cap action head and common action projection.	Strong active classical frontier and high-capacity threshold reference.	Main.

VQC-Cap	Skip-connected hybrid VQC policy with learned demand-cap action head and high-weight sparsity regularization.	Tests compact hybrid parameter-performance scaling.	Main.
VQC-RegimeCap	Hybrid VQC policy with smooth preference-regime gates over demand-cap heads and high-weight sparsity regularization.	Tests explicit preference-regime conditioning.	Main robustness family.
Screened MLP variants	Direct-action and residual-action MLP policies without the final demand-cap wrapper.	Shows why the demand-cap head is the active comparator.	SI/screening.
Screened structured classical variants	Fourier-feature and RBF-prototype policies with matched observation and projection.	Tests non-quantum structured-feature explanations.	SI/screening.
Screened VQC variants	Earlier direct, residual and one-hot regime VQC policies.	Negative evidence motivating VQC-Cap and VQC-RegimeCap.	SI/screening.
No storage and oracle	No-action policy and perfect-foresight optimization reference.	Normalization floor and opportunity ceiling.	Reference.

Table S11. Locked-test model grid. Actor-parameter counts are deployable actor parameters from the frozen S8 manifest.

Family	Budget label	Target params	Exact params	Seeds	Status/placement
MLP-Cap	Tiny	0.5k	495	10	Locked S8 classical comparator.
MLP-Cap	Compact	1.4k	1,393	10	Locked S8 classical comparator.
MLP-Cap	Medium	4k	3,985	10	Locked S8 classical scaling diagnostic.
MLP-Cap	Large	12k	11,929	10	Locked S8 high-capacity reference.
VQC-Cap	Tiny	0.5k	499	10	Locked S8 hybrid finalist.
VQC-Cap	Compact	1.4k	1,393	10	Locked S8 hybrid finalist.
VQC-Cap	Medium	4k	3,993	10	Locked S8 hybrid finalist.
VQC-RegimeCap	Tiny	0.5k	498	10	Locked S8 hybrid robustness family.
VQC-RegimeCap	Compact	1.4k	1,393	10	Locked S8 hybrid robustness family.
VQC-RegimeCap	Medium	4k	3,999	10	Locked S8 hybrid robustness family.
Screened classical families	Multiple	0.5k–12k	476–11,929 depending on family/head	3–10 depending on gate	Validation screening only.
Screened VQC families	Multiple	0.5k–4k	498–3,999 depending on family/head	3–10 depending on gate	Validation screening only.

Table S12. Hybrid VQC architecture card for the retained VQC-Cap and VQC-RegimeCap families.

Architecture item	Definition	Retained value
Input features	Observation features entering the VQC branch	Only the train-only scaled z16 (sixteen features) is encoded into the quantum circuit via data reuploading; the six-dimensional preference vector is not encoded into the quantum circuit and instead enters the classical skip path that feeds the policy head.
Feature scaling	Normalization before encoding	Robust train-split median/IQR scaling for continuous features, no scaling for binary/cyclic features, clipping to $[-8,8]$.
Qubits	Number of qubits	4.
Encoding	Data map	RY/RZ angle encoding with $\pi\tanh(x)$ and two data-reuploading layers.
Ansatz depth	Variational layers	2 reuploading/variational layers.

Single-qubit gates	Trainable gates per layer	RX, RY and RZ on each qubit; 24 trainable quantum parameters total.
Entanglement pattern	Multi-qubit gates	Ring CNOT entanglement in each layer.
Measurements	Observables	Pauli-Z and Pauli-X expectation values for each qubit; 8 readout values.
Quantum resource estimate	Per-action circuit resources	16 data rotations, 24 trainable rotations, 8 CNOTs, 48 estimated gates, depth 14, one analytic circuit evaluation per action.
Classical heads	Post-VQC readout	Demand-cap head for VQC-Cap; three smooth preference-regime demand-cap heads for VQC-RegimeCap.
Shots/simulator	Quantum backend	Analytic expectation on PennyLane default.qubit; finite-shot hardware execution is not claimed.
Gradient method	Differentiation	Torch-interface backpropagation through the statevector simulator.

Table S13. Implied-cap isotonic preference-path wrapper card. The wrapper is applied identically to all three retained finalist families (MLP-Cap, VQC-Cap and VQC-RegimeCap) and is the same operator that was frozen in the S7-C wrapper development pass before the S8 locked-test freeze.

Wrapper item	Definition	Frozen value or status
Wrapper name	Inference-time preference-path post-processor	implied_cap_isotonic.
Operator	Pool-adjacent-violators (PAV) projection on per-step cap sequence	Computed fresh every action step from the length-ten cap sequence.
Preference grid	Weights queried per action	{0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9}.
Fitting data	Data used to fit wrapper parameters	None; the wrapper is parameter-free.
Trainable parameters	Wrapper-added trainable parameters	0.
Frozen parameters	Wrapper-added frozen parameters	0.
Stored knots or values	Persistent wrapper state across steps	0; PAV is recomputed each step from the policy queries.
Inference cost multiplier	Policy forward passes per action	10x relative to a single query; reflected in the resource audit in main text Table 6.
Application scope	Families using the wrapper	MLP-Cap, VQC-Cap and VQC-RegimeCap (all retained finalists).
Validation use	Whether validation data is used during inference	No; the wrapper does not access validation, test or oracle data at inference.
Audit reference	Implementation audit	artifacts/protocol/isotonic_wrapper_accounting_audit_v1.md.

Supplementary Note 5. Training protocol and validation selection

The training-protocol parameters (Table S14) and the training-run ledger schema (Table S16) record the fixed-dose behavioral-cloning settings, while the epoch-dose sensitivity (Table S15) verifies that the 12k MLP-Cap anomaly is not under-training.

Table S14. Training protocol parameters for the active fixed-dose path.

Training item	Definition	Active value or status
BC target source	Perfect-foresight oracle action labels on the training split	Oracle targets only; learned policies remain causal at inference.
Train set	Building-month files	Full training split: 1,316 June–September building-month files.
BC sample construction	Row selection from label files	event_aware_stratified; max 5,760 rows per file; weights $w \in \{0.00,0.30,0.40,0.80,0.90\}$.

BC loss	Supervised action loss	Event-weighted mean squared error on clipped normalized oracle action targets; event weights have base 1, max 12 and per-episode mean normalization.
Event emphasis	Sparse demand-charge windows	Native peak window, top 97.5% load windows, running-peak increments, oracle nonzero action and oracle discharge windows.
Epochs and batches	Fixed supervised dose	Two epochs; MLP batch size 256; hybrid VQC batch size 128 with max 45 batches/file.
Learning rate	Optimizer learning rate	AdamW with learning rate 10^{-3} for MLP-Cap and 3×10^{-4} for hybrid VQC; weight decay 10^{-4}.
Checkpoint selection	Whether validation chooses best checkpoint during training	Disabled; final fixed-dose BC checkpoint is replayed.
Vectorized validation	Batched replay	Excluded because scalar parity was not acceptable for demand-charge ESS replay.
Validation preference grid	Replay weights	0.0,0.1, … ,0.9.
Hardware	Execution environment	CPU scalar replay; VQC simulated with PennyLane default.qubit; GPU not used in locked evidence path.

Table S15. Training-dose sensitivity for the 4k and 12k MLP-Cap rows. Ten seeds per budget were re-trained at four epochs (double the main-path two-epoch dose) under the same architecture, head, wrapper and validation protocol. Doubling the dose worsens validation regret at both budgets, so the 12k anomaly is not an under-training artifact.

Budget	Macro regret (epoch 2, S7 validation)	Macro regret (epoch 4, diagnostic)	Δ (e4 - e2)	Interpretation
4,000	0.597 [0.567, 0.633]	0.640 (seed SD 0.098)	+0.043	More training worsens regret; capacity-data mismatch rather than underdose.
12,000	0.659 [0.614, 0.708]	0.737 (seed SD 0.118)	+0.077	Anomaly is preserved or worsened with longer training; 1.4k sweet spot is genuine.

Table S16. Training-run ledger schema used by the S7-C/S8 manifests.

Column	Type	Example/status	Required	Notes
finalist_id or cell_id	string	s7c_hybrid_vqc_cap_isotonic_b1400_seed20260430	Yes	Unique family-budget-seed identifier.
family, action_head	string	h5_fullskip_peak_cap, peak_cap	Yes	Architecture lookup keys.
budget_target	integer	1400	Yes	Nominal actor-parameter budget.
deployable_parameters	integer	1,393	Yes	Actual actor parameter count.
seed	integer	20260430	Yes	Model seed; final locked test uses 10 seeds per retained cell.
checkpoint	path	policy_checkpoint.pt	Yes	Final fixed-dose checkpoint path.
checkpoint_sha256	string	recorded in S8 finalist manifest	Yes	Checkpoint immutability audit.
status	categorical	frozen_for_s8_locked_test or completed	Yes	Failed runs are not silently dropped.
failure_reason	String	empty for locked completed cells	If failed	Instability or code issue.

Supplementary Note 6. Validation threshold and threshold sensitivity

The validation-threshold construction (Table S17) and the threshold sensitivity mirror of main-text Table 2 (Table S18) together support the sensitivity ladder reported in the body.

Table S17. Validation-based threshold construction. Thresholds are calibrated from full-validation replay only; locked-test outcomes are used only for final crossing checks.

Step	Definition	Frozen value
Reference model	Validation-frozen 12k MLP-Cap with implied-cap isotonic wrapper	10 seeds; 11,929 deployable actor parameters.
Validation mean regret	Dense-grid mean regret on validation buildings	Unrounded 0.6594, rounded up to 0.660.
Bootstrap interval	Building-level 95% CI for validation regret	[0.6141,0.7077] (unrounded).
Primary threshold	Upper 95% confidence bound rounded upward to the nearest 0.005	Unrounded 0.7077, rounded up to $\tau_R^{\mathrm{val}} = 0.710$.
Rounding rule	Nearest 0.005 upward (rounding direction registered before locked-test access)	Frozen in paper/source_data/s7_validation_threshold_definitions.csv.
Tail threshold	High-weight tail constraint	Binary high-weight tail rate is non-discriminative for retained rows (0.20 for all); continuous high-weight burden is reported instead.
High-weight definition	Preference weights included in tail analysis	$w \in \{0.8,0.9\}$.
Locked-test rule	Family reaches threshold if test regret $\leq \tau_R^{\mathrm{val}}$	Applied once after S8 replay.

Table S18. Validation-threshold sensitivity (Supplementary Information mirror of main-text Table 2). Thresholds are derived from S7 full-validation replay only and applied to the S8 locked-test mean regret without retuning. This table is retained in the Supplementary Information for standalone use when the SI is split from the main text at journal submission; numerical content is identical to the main-text table.

Threshold	VQC-Cap min params	VQC-RegimeCap min params	MLP-Cap min params	Interpretation
Best MLP validation mean ($0.5758 \rightarrow \tau_R = 0.580$)	Not reached	Not reached	Not reached	Strict sensitivity; too strict for any retained family on locked test.
MLP 1.4k validation mean ($0.5863 \rightarrow \tau_R = 0.590$)	1,393	Not reached	Not reached	Compact-reference sensitivity; only VQC-Cap 1.4k crosses by mean rule.
MLP 12k validation mean ($0.6594 \rightarrow \tau_R = 0.660$)	1,393	1,393	495	High-capacity reference mean; all active families reach, with MLP crossing at 0.5k.
MLP 12k validation U95 ($0.7077 \rightarrow \tau_R^{\mathrm{val}} = 0.710$)	1,393	1,393	495	Primary U95 threshold; broad threshold, so frontier and paired contrasts remain the main interpretation.

Supplementary Note 7. Locked-test metrics and paired contrasts

The full locked-test metric ledger (Table S19), the primary paired contrasts including bill-saving and degradation effects (Table S20) and the seed-completion ledger (Table S21) expand the main-text Table 5 and Table 4.

Table S19. Full locked-test metric ledger. Values are building-level bootstrap means [95% CI] except the deterministic threshold-crossing flag.

Family/budget	Params	Regret	Bill saving	Negative degradation	High-w tail rate	Reached τ_R^{val}
MLP-Cap 0.5k	495	0.597 [0.552, 0.651]	433.2 [392.5, 475.2]	-182.7 [-200.2, -166.2]	0.200 [0.200, 0.200]	Yes

MLP-Cap 1.4k	1,393	0.596 [0.547, 0.654]	440.5 [400.4, 482.4]	-193.0 [-211.6, -175.5]	0.200 [0.200, 0.200]	Yes
MLP-Cap 4k	3,985	0.596 [0.566, 0.632]	397.6 [361.8, 434.1]	-187.2 [-204.3, -170.6]	0.200 [0.200, 0.200]	Yes
MLP-Cap 12k	11,929	0.659 [0.615, 0.711]	416.6 [376.0, 459.2]	-186.5 [-204.4, -169.9]	0.200 [0.200, 0.200]	Yes
VQC-Cap 0.5k	499	0.819 [0.750, 0.900]	402.8 [363.6, 443.9]	-214.9 [-235.5, -195.1]	0.200 [0.200, 0.200]	No
VQC-Cap 1.4k	1,393	0.581 [0.538, 0.632]	449.5 [407.7, 494.1]	-193.9 [-212.4, -177.1]	0.200 [0.200, 0.200]	Yes
VQC-Cap 4k	3,993	0.682 [0.632, 0.741]	435.0 [392.9, 477.9]	-212.7 [-232.7, -193.7]	0.200 [0.200, 0.200]	Yes
VQC-RegimeCap 0.5k	498	0.769 [0.711, 0.834]	403.0 [362.6, 446.1]	-206.5 [-226.1, -187.7]	0.200 [0.200, 0.200]	No
VQC-RegimeCap 1.4k	1,393	0.634 [0.596, 0.679]	411.0 [372.8, 451.8]	-193.5 [-211.8, -175.8]	0.200 [0.200, 0.200]	Yes
VQC-RegimeCap 4k	3,999	0.623 [0.588, 0.664]	392.1 [354.4, 431.2]	-185.7 [-203.6, -168.6]	0.200 [0.200, 0.200]	Yes
Oracle	–	0	Reference ceiling	Reference ceiling	0	Perfect-foresight reference

Table S20. Primary paired locked-test contrasts. Differences are hybrid minus matched MLP at the same nominal budget; negative regret is better, positive bill-saving difference is better and positive negative-degradation difference is less degradation cost.

Contrast	Regret difference	Bill-saving difference	Negative-degradation difference	Interpretation
VQC-Cap 0.5k – MLP-Cap 0.5k	0.222 [0.194, 0.253]	-30.5 [-40.0, -21.4]	-32.1 [-36.3, -28.3]	MLP better
VQC-Cap 1.4k – MLP-Cap 1.4k	-0.016 [-0.032, -0.001]	9.0 [1.9, 16.1]	-0.9 [-2.0, 0.4]	Hybrid (marginal)
VQC-Cap 4k – MLP-Cap 4k	0.086 [0.061, 0.113]	37.4 [26.9, 48.1]	-25.5 [-28.5, -22.8]	MLP better
VQC-RegimeCap 0.5k – MLP-Cap 0.5k	0.172 [0.153, 0.193]	-30.3 [-39.6, -21.1]	-23.7 [-27.1, -20.5]	MLP better
VQC-RegimeCap 1.4k – MLP-Cap 1.4k	0.038 [0.019, 0.054]	-29.5 [-36.1, -22.9]	-0.5 [-1.8, 0.9]	MLP better
VQC-RegimeCap 4k – MLP-Cap 4k	0.027 [0.013, 0.041]	-5.5 [-14.6, 3.5]	1.5 [0.2, 3.0]	MLP better

Table S21. Seed completion and replay-cost ledger for S8 locked test.

Family/budget group	Seeds complete	Validation regret range	Test regret range	Replay wall time/cell	Failed runs
MLP-Cap budgets	10/10 each	0.576–0.659 across budget means	0.596–0.659 across budget means	1.18–1.26 h mean	0
VQC-Cap budgets	10/10 each	0.561–0.788 across budget means	0.581–0.819 across budget means	32.09–42.20 h mean	0
VQC-RegimeCap budgets	10/10 each	0.615–0.742 across budget means	0.623–0.769 across budget means	20.42–52.84 h mean	0
S8 total	100/100 cells	Validation-only source used for freeze	Locked-test one-shot replay completed	Run time 97.9 h with 30 workers	0

Supplementary Note 8. Dense preference-grid replay

The dense-grid source-table schema (Table S22), the trained-vs-untrained preference-weight audit (Table S23) and the full dense-grid figure across every retained main family-budget cell (Figure S2) provide the full evidence behind main-text Figure 3.

Table S22. Dense-grid replay source table. S8 source rows aggregate over locked-test buildings, months and seeds for each model, budget and preference weight.

Column	Type	Description	Unit	Example	Required	Notes
series	string	Model-family display identifier	–	Hybrid demand-cap + isotonic	Yes	Paper-facing series.
budget_target	integer	Nominal actor-parameter budget	parameters	1400	Yes	Exact params stored separately.
preference_weight	numeric	Preference weight	–	0.8	Yes	Dense grid 0.0–0.9.
bill_saving_mean_usd	numeric	No-storage bill minus policy bill	USD/month	521.9 for oracle at $w = 0.8$	Yes	Higher better.
negative_degradation_mean_usd	numeric	Negative policy degradation cost	USD/month	-30.8 for oracle at $w = 0.8$	Yes	Higher better.
normalized_regret_mean	numeric	Oracle-relative normalized regret	–	0 for oracle	Yes	Eq. (6).
high_w_tail_failure_rate	numeric	Binary tail rate for high-w rows	–	0.2 in retained aggregate rows	Recommended	Non-discriminative for final retained rows.

Table S23. Trained-vs-untrained preference-weight audit. BC is trained on the sparse grid {0.0,0.3,0.4,0.8,0.9} and evaluated on the dense grid {0.0,0.1, … ,0.9}. Mean regret on the five untrained weights {0.1,0.2,0.5,0.6,0.7} is lower than on the trained set for every family-budget because the untrained set is mid-grid while the trained set contains the hardest weight $w = 0.9$; the dense per-w profile in Fig. 3B shows no localized spike at untrained weights.

Family	Budget	Trained-w mean regret	Untrained-w mean regret	Δ (untr. - tr.)	Notes
MLP-Cap	500	0.756	0.438	-0.318	Untrained mid-grid easier; smooth across grid.
MLP-Cap	1,400	0.760	0.433	-0.328	Same pattern; no generalization gap visible.
MLP-Cap	4,000	0.713	0.479	-0.235	Same pattern.
MLP-Cap	12,000	0.835	0.484	-0.351	Larger trained-vs-untrained gap consistent with the elevated 12k trained-set mean.
VQC-Cap	500	1.045	0.593	-0.453	Compact hybrid; underfit at trained endpoint $w = 0.9$.
VQC-Cap	1,400	0.743	0.418	-0.325	Best macro-regret point; gap matches MLP-Cap pattern.
VQC-Cap	4,000	0.855	0.509	-0.346	Higher trained-set mean reflects the worsening at $w = 0.9$.
VQC-RegimeCap	500	0.973	0.566	-0.407	Compact regime hybrid.
VQC-RegimeCap	1,400	0.776	0.492	-0.284	Modest gap consistent with smooth regime gates.
VQC-RegimeCap	4,000	0.740	0.505	-0.235	Smallest absolute gap of the retained rows.

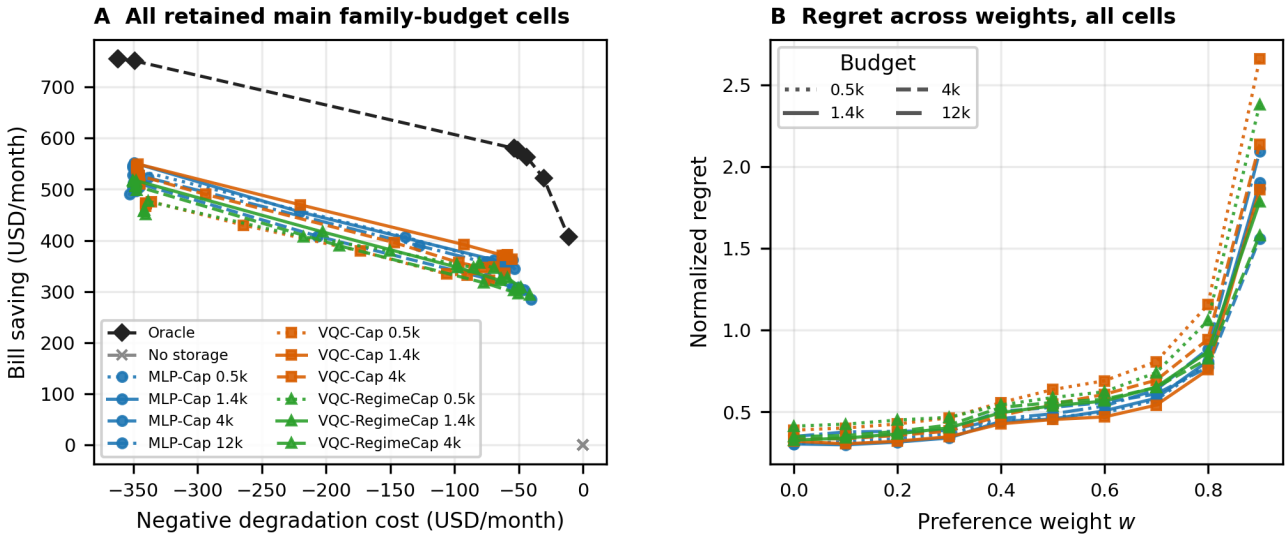

Figure S2. Full dense preference-grid replay for every retained main family-budget cell. **A.** Bill-saving versus negative-degradation-cost frontier with the oracle (dashed black) and no-storage (gray) references. **B.** Oracle-relative normalized regret across w; color encodes family, linestyle encodes budget. All 1.4k cells track the oracle frontier most closely at low and mid w; the spread across budgets is small relative to the oracle gap.

Supplementary Note 9. High-weight timing-tail taxonomy

The operational-burden categories (Table S24), the locked-test row-level high-w burden summary (Table S26) and the representative monthly trace (Figure S3) document the mechanistic content behind main-text Figure 4A.

Table S24. High-weight operational-burden categories used for trace interpretation.

Category	Operational definition	Diagnostic variables
Early depletion	Battery reserve is depleted before the native or oracle monthly-peak event.	SoC 1, 2, 4 and 8 hours before peak; pre-peak discharge.
Late response	Energy is available, but discharge occurs after the demand peak or too slowly to reduce it.	Battery power around peak; timing difference to oracle action.
Over-conservative	Policy avoids cycling despite high demand-charge opportunity.	Low throughput, high SoC at peak, degradation component.
No-export/projection clipping	Intended action is reduced by no-export or SoC projection.	Pre-projection action, projected action, grid import.
Forecast-blind event	Perfect-foresight oracle captures a future event that is not predictable from causal observations.	Load ramp statistics, time-to-peak, previous-day pattern.
Terminal artifact	End-of-month SoC correction materially changes regret or action sequence.	Final-day SoC and terminal correction cost.

Table S25. Per-weight locked-test mean normalized regret at the two high-w preference weights, every retained family-budget cell, ranked within each weight from lowest (best) to highest (worst). At $w = 0.9$ the MLP-Cap 4k row is the best high-weight point (1.561) and no hybrid cell beats it; at $w = 0.8$ VQC-Cap 1.4k is the best (0.758) but MLP-Cap rows are clustered between 0.77 and 0.88. The contrast between these two weights motivates the high-weight-burden discussion in the main text.

w	Family/budget	Regret	Bill saving (USD/mo)	Negative deg. (USD/mo)
0.8	VQC-Cap 1.4k	0.758	371.6	-58.8
0.8	MLP-Cap 4k	0.774	303.3	-46.2
0.8	MLP-Cap 0.5k	0.797	363.9	-59.3
0.8	MLP-Cap 1.4k	0.811	360.7	-59.2

0.8	VQC-RegimeCap 4k	0.824	305.4	-50.7
0.8	VQC-RegimeCap 1.4k	0.867	326.5	-58.9
0.8	MLP-Cap 12k	0.885	352.7	-66.9
0.8	VQC-Cap 4k	0.944	348.6	-69.4
0.8	VQC-RegimeCap 0.5k	1.058	355.4	-80.8
0.8	VQC-Cap 0.5k	1.156	343.6	-84.6
0.9	MLP-Cap 4k	1.561	284.0	-40.4
0.9	VQC-RegimeCap 4k	1.581	293.8	-41.7
0.9	VQC-RegimeCap 1.4k	1.786	308.3	-48.7
0.9	VQC-Cap 1.4k	1.859	362.9	-55.6
0.9	MLP-Cap 0.5k	1.867	362.4	-55.0
0.9	MLP-Cap 1.4k	1.902	345.0	-53.7
0.9	MLP-Cap 12k	2.093	336.7	-61.9
0.9	VQC-Cap 4k	2.137	335.6	-60.9
0.9	VQC-RegimeCap 0.5k	2.381	345.8	-69.4
0.9	VQC-Cap 0.5k	2.660	322.3	-72.8

Table S26. High-weight burden summary for retained locked-test rows. The binary tail rate is saturated at 0.20 for retained aggregate rows, so continuous bill-saving and degradation metrics are used for interpretation.

Family/budget	Binary high-*w* rate	Bill-saving mean	Negative degradation mean	Interpretation
MLP-Cap 1.4k	0.200 [0.200, 0.200]	440.5 USD/month	-193.0 USD/month	Strong comparator; tail-rate not discriminative.
VQC-Cap 1.4k	0.200 [0.200, 0.200]	449.5 USD/month	-193.9 USD/month	Best macro-regret locked row; similar degradation to MLP 1.4k.
VQC-RegimeCap 1.4k	0.200 [0.200, 0.200]	411.0 USD/month	-193.5 USD/month	More conservative bill-saving frontier at compact budget.
MLP-Cap 12k	0.200 [0.200, 0.200]	416.6 USD/month	-186.5 USD/month	High-capacity validation threshold reference.

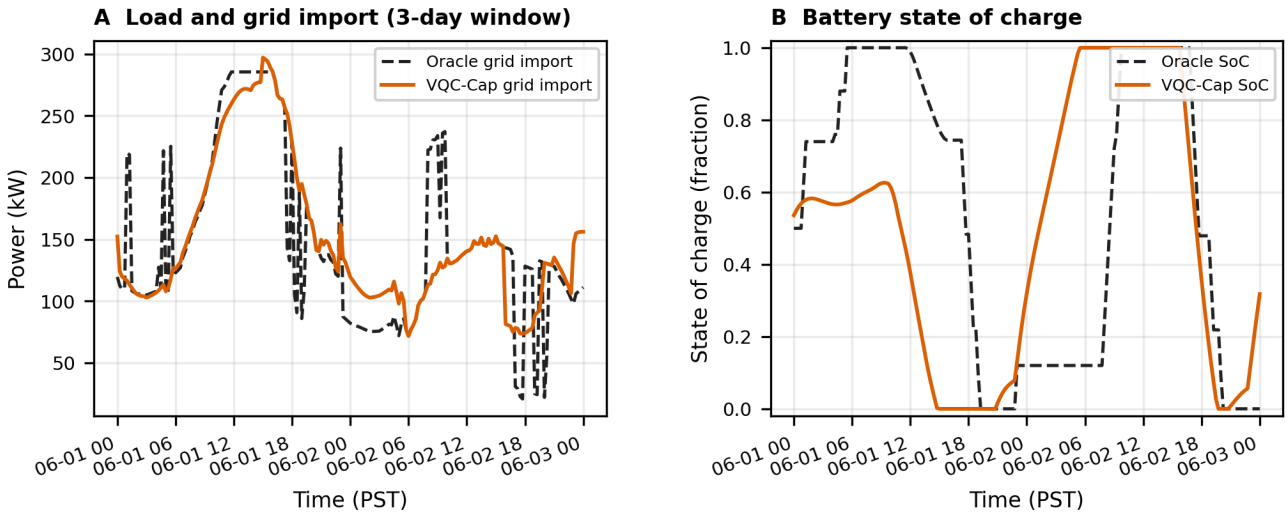

Figure S3. Representative monthly trace for a single building-month at the bill-priority weight (w=0). **A.** Native load (gray), oracle grid import (dashed black) and VQC-Cap 1.4k grid import (orange) over a three-day window centred on the monthly peak; the VQC-Cap policy reduces peak grid import relative to no storage but allows higher off-peak import than the oracle. **B.** State-of-charge trajectories for the oracle and the VQC-Cap policy; the policy discharges aggressively into the peak event and recovers afterwards, in contrast to the oracle's pre-planned conservative cycling.

Supplementary Note 10. Quantum, feature and resource ablations

The feature-block and VQC screening summary (Table S27), the cap-head isolation ablation (Table S28), the latency and resource audit (Table S29) and the shot-noise sensitivity (Table S30) constitute the resource and architectural ablations.

Table S27. Feature-block and VQC screening summary. Active rows use full-validation means [95% CI]; screened rows are summarized from validation-screening decisions and remain supplementary.

Ablation	Params	Validation regret	High-*w* rate	Interpretation
VQC-Cap 1.4k	1,393	0.561 [0.525, 0.600]	0.200	Active skip-connected hybrid finalist.
VQC-RegimeCap 1.4k	1,393	0.620 [0.590, 0.651]	0.200	Active preference-regime hybrid robustness family.
Earlier direct/residual VQC	0.5k–4k	Screening-only	Not locked	Did not form a reliable dense preference-conditioned frontier before demand-cap rescue.
Earlier one-hot-gated VQC	1.4k screening	Screening-only	Not locked	Monotonicity and frontier-shape risk motivated smooth gates instead.
Fourier feature block	476–3,976	Screening-only	Not locked	Produced weaker frontiers than MLP-Cap and was excluded from S8.
RBF feature block	Screening budgets	Screening-only	Not locked	Did not provide a stronger final comparator than MLP-Cap.
MLP direct/residual block	495–11,929	Screening-only	Not locked	Demand-cap head and isotonic wrapper were retained as the final classical comparator.

Table S28. Cap-head isolation ablation at the matched 1.4k budget (validation only). Replacing the demand-cap action head with a direct head—while keeping the same observation, BC dose, sparse-train grid and isotonic wrapper—worsens regret for both blocks. Internal VQC variant codes: *h5* is the skip-connected hybrid VQC used as the locked-test VQC-Cap finalist (parallel classical skip path with the VQC block); *h1* is the screened direct-readout hybrid VQC without a skip path. The cap-head baseline rows use the same 10-seed cells as the locked-test main path; the direct-head ablation rows are 2-seed validation cells, so the across-block contrast at the direct head (+0.063) carries wide seed-level uncertainty. The ablation also swaps the VQC variant (h5 for the cap-head baseline, h1 for the direct-head ablation), so the +0.063 direct-head contrast confounds the head change with a VQC architecture change; this is acknowledged in the body. The within-block contrast at the cap-head baseline is −0.026 (0.5606 vs 0.5863, 10 vs 10 seeds). All deltas are computed from the unrounded mean values; the rows in the table below display the conventional three-decimal rounding for readability and the seed standard deviations for the 2-seed rows are reported alongside.

Cell	Seeds	Macro regret	Bill saving	Neg. degradation
MLP + cap head (baseline)	10	0.586	443.9	-195.7
VQC (h5 skip) + cap head (baseline)	10	0.561	454.9	-197.0
MLP + direct head (ablation)	2	0.702 (sd 0.000)	384.5	-202.0
VQC (h1 direct) + direct head (ablation)	2	0.765 (sd 0.109)	210.6	-133.7
Δ from removing cap head, same family		MLP: +0.116; VQC: +0.204	MLP: -59.4; VQC: -244.4	MLP: -6.3; VQC: +63.3
VQC − MLP at cap head (10 vs 10 seeds, h5 vs MLP)	10 vs 10	-0.026	+11.0	-1.3
VQC − MLP at direct head (2 vs 2 seeds, h1 vs MLP; head + arch. swap)	2 vs 2	+0.063	-174.0	+68.3

Table S29. Resource and latency audit from scalar S8 locked-test replay. Per-action latency is computed from scalar replay wall time divided by replayed interval count and is reported as an implementation-level CPU replay cost, not a hardware-quantum claim.

Model family	Actor params	Mean replay latency	Extra resource metric	Notes
VQC-Cap 0.5k	499	6.90 ms/action	1 circuit eval/action; 48 gates; analytic shots	Mean cell replay 32.09 h.
VQC-Cap 1.4k	1,393	9.07 ms/action	1 circuit eval/action; 48 gates; analytic shots	Mean cell replay 42.18 h.
VQC-Cap 4k	3,993	9.07 ms/action	1 circuit eval/action; 48 gates; analytic shots	Classical head width increases parameters.
VQC-RegimeCap 0.5k–4k	498–3,999	4.39–11.36 ms/action	1 circuit eval/action; 48 gates; analytic shots; gate-head overhead	Smooth-gate head overhead included.
MLP-Cap 1.4k	1,393	0.253 ms/action	Matrix operations only	Mean cell replay 1.18 h.
MLP-Cap 12k	11,929	0.270 ms/action	Matrix operations only	High-capacity reference; mean cell replay 1.26 h.

Table S30. Shot-noise sensitivity for VQC-Cap 1.4k (validation subset, single seed). Each Pauli expectation $E \in [-1,1]$ from the analytic VQC readout is replaced by the sample mean of n_{shots} binomial single-shot outcomes drawn from $P(+1) = (1+E)/2$, matching the statistics of an ideal noiseless finite-shot device. The diagnostic uses VQC-Cap 1.4k seed 20260430 on the alphabetically deterministic first 30 of the 110 validation buildings (1,560 primary-grid episodes across 39 building-month-grid weight points). The macro-regret difference between analytic, 8,192-shot and 1,024-shot replays is at most 0.001, far below the locked-test bootstrap uncertainty in Table 5, indicating that the analytic-expectation result is robust to ideal finite-shot sampling in the 1,024–8,192 range. Coherent hardware-noise effects (depolarizing channel, gate errors) are not modelled in this sensitivity and would require a separate noise-aware validation pass.

Shot level	Macro regret	Bill saving (USD/month)	Neg. degradation (USD/month)	Notes
Analytic (state vector)	0.5575	484.77	-217.82	Reference; same backend as locked-test S8 replay.
$n_{\text{shots}} = 8{,}192$	0.5573	484.76	-217.82	$\Delta = -0.0002$ vs analytic; within sampling noise.
$n_{\text{shots}} = 1{,}024$	0.5568	484.75	-217.82	$\Delta = -0.0007$ vs analytic; binomial sampling variance dominates the gap.

Supplementary Note 11. Sensitivity analyses and reproducibility package

The sensitivity analyses (Table S31), the reproducibility manifest (Table S32) and the software/hardware environment (Table S33) record what was held fixed, where the source artifacts live and which versions of the analysis stack were used.

Table S31. Sensitivity analyses and status.

Sensitivity	Design	Result/status
Threshold sensitivity	Apply validation-derived thresholds $\tau_R \in \{0.580, 0.590, 0.660, 0.710\}$ to locked-test replay without retuning.	Minimum parameter count by family in main-text Table 2 (Supplementary mirror in Table S18).
High-weight burden sensitivity	Compare binary high-weight tail rate with continuous bill/degradation burden.	Binary rate is non-discriminative for retained rows; continuous frontier metrics are primary.
Tariff accounting	Monthly reset primary; carried-peak not used in the main scenario.	No-storage telescoping and oracle/environment parity passed.
Battery sizing	2-hour, 0.25 peak-power sizing primary.	Alternative sizing is not part of the locked evidence path.
Degradation coefficient	Base coefficient 0.0596 USD/kWh.	Sensitivity not required for the locked claim; coefficient is frozen in config.
Screened-family placement	Keep Fourier/RBF and earlier VQC variants in SI unless validation evidence changes the frozen candidate set.	S8 retains only MLP-Cap, VQC-Cap and VQC-RegimeCap.

Table S32. Reproducibility archive manifest.

Artifact	Contents	Status/identifier
split_manifest.csv	Building IDs, splits and stratification variables.	artifacts/data/split_manifest.csv; SHA256 prefix bb86e5e10df4.
tariff_constants.json	PG&E B-10 scenario constants and snapshot date.	artifacts/configs/tariff_constants.json.
battery_sizing_config.json	Battery sizing, efficiency, SoC and degradation settings.	artifacts/configs/battery_sizing_config.json.
feature_card_v1.yaml	Feature definitions, train-only scaling and clipping rules.	artifacts/configs/feature_card_v1.yaml.
s8_locked_test_finalist_manifest_v1.csv	Frozen finalist checkpoints, SHA256 hashes and replay roles.	artifacts/configs/s8_locked_test_finalist_manifest_v1.csv.
oracle_labels/	Oracle and no-storage results by episode and preference.	artifacts/oracle_labels/.
locked_test_replay_results/	Final scalar replay outputs.	runs/s8_locked_test_replay_20260515_090125/outputs/.
figure_source_data/	CSV data for main figures and statistical tables.	paper/source_data/.
analysis_scripts/	Bootstrap, threshold and plotting scripts.	scripts/generate_s9_statistical_tables.py; scripts/generate_s9_threshold_sensitivity.py.
manuscript_figures/	300 dpi PNG manuscript figures.	paper/figures/.

Table S33. Software and hardware environment used for the locked analysis.

Component	Version or hardware	Notes
Python	3.10.20	Conda environment pqmorl.
PyTorch	2.10.0	Neural policy framework.
PennyLane	0.38.0	Analytic default.qubit simulator for VQC training/replay.
Oracle solver	Gurobi 13.0.1	Perfect-foresight optimization backend.
NumPy/SciPy/pandas/matplotlib	NumPy 1.26.4; SciPy 1.15.3; pandas 2.3.3; matplotlib 3.10.8	Analysis and figure stack.
CPU	AMD Ryzen Threadripper 7970X, 32 cores/64 threads	Scalar replay used CPU workers.
GPU	Not used in locked evidence path	GPU acceleration is not part of reported comparison.
Operating system	Ubuntu 24.04 family, Linux 6.17.0-23-generic	Remote server execution under user-level systemd services.
Random seeds	Split seed 20260429; model seeds 20260430–20260509; bootstrap seed 20260519	Seed handling is recorded in manifests and source-data CSVs.