Directed spectral-truncated operators for inductive classification of quantum phases across a control-parameter sweep

Molena Huynh[1,*]

[1] *North Carolina State University, Raleigh, North Carolina 27695, USA*

(Dated: June 28, 2026)

Classifying the phase of a parameterized quantum many-body system as a control parameter is swept across a quantum phase transition—from the measurement-derived features of classical shadows—is confounded by device-to-device and shot-noise variation and by the rarity of the critical regimes that matter most. The graph operator used to propagate phase labels across the similarity graph of quantum states, the self-adjoint normalized adjacency $S = D^{-1/2}AD^{-1/2}$, is poorly matched to this task: it is reversible, so it cannot encode the directed arrow of the drive, and, built over all states, it consults held-out connectivity at inference, conflating the value of diffusion with a transductive information advantage. We tackle both challenges by introducing a spectral-truncated, drive-directed graph operator that benefits from three innovations: a non-self-adjoint forward propagator that transports labels along the sweep, a band-limiting spectral truncation that denoises the operator into a closed-form solve, and an out-of-sample Nyström extension that renders the classifier inductive and device-transferable. The truncation realizes, on the state graph, the spectral truncation of multiplication operators from C^*-algebraic kernel machines, generating a genuinely noncommutative truncated operator $*$-algebra with a positive-definite Gram kernel. On a seeded synthetic benchmark emulating a parameterized quantum system ($n = 3000$ states, 5 seeds), the inductive operator improves held-out phase-classification accuracy over a matched inductive kNN classifier on all four leakage-checked transfer protocols (0.6856 versus 0.5992) and recovers a rare critical regime the baseline almost never finds (recall 0.3322 versus 0.0151), with its largest margin on forward extrapolation across the sweep—precisely where smoothing is weakest. A split-conformal layer supplies distribution-free coverage (0.8968 against a 0.90 target) and temperature scaling reduces the expected calibration error by 73 %. All data are synthetic; validation on classical shadows from a programmable quantum simulator is the decisive next step.

I. INTRODUCTION

Programmable quantum simulators now prepare and measure many-body states across entire phase diagrams, and a central task in machine learning for quantum physics is to identify the *phase* a state occupies from its measurement record—typically the randomized-measurement data of classical shadows [1, 2], which compress an exponentially large state into a tractable feature vector. Recognizing phases and locating transitions from such data is by now a workhorse of the field [3–6]. Two structural features of these data shape any classifier. First, the shadow features of a parameterized system concentrate near a low-dimensional manifold—the phase diagram—that is routinely approximated by a k-nearest-neighbor (kNN) similarity graph before clustering, embedding, or kernel learning [7, 8, 14–17]. Second, the physically decisive regimes—quantum critical points, narrow topological windows—are frequently rare, and ordinary classifiers, trained to minimize average error, miss them.

The canonical response to both features is graph-based semi-supervised learning [12, 13]: phase labels are allowed to diffuse over the state graph so that each state borrows statistical strength from its neighborhood. The diffusion operator used almost universally is the symmetric normalized adjacency $S = D^{-1/2}AD^{-1/2}$ and its Laplacian $L = D - A$ [14]. This object carries two limitations that bear directly on classification across a sweep, and that motivate the present work. (i) It is self-adjoint, hence reversible: it cannot encode the directed arrow of the drive, even though the control-parameter sweep supplies exactly that arrow and directed, non-Hermitian transition operators are the natural language of driven and open quantum dynamics [18–20]. (ii) Built over all states, it is transductive: it consults the held-out states' connectivity at inference, so a measured accuracy gain over an inductive point classifier conflates the value of diffusion with a transductive information advantage—the confound that comparable studies flag as their decisive missing control.

Here, we take both limitations as design targets and introduce a spectral-truncated, drive-directed graph operator, which we develop into a falsifiable quantum-phase classifier. Rather than re-applying label propagation, we contribute four innovations. (1) We construct a *directed* propagator P by reweighting each edge $i \to j$ with a bounded forward kernel $\kappa(s_j - s_i)$ of the control-parameter increment and row-normalizing; P is non-self-adjoint—the single state-graph instance of the operator noncommutativity that recent C^*-algebraic kernel machines exploit for expressiveness [9–11]—and it transports labels along the sweep rather than merely smoothing them across it. (2) We *band-limit* the diffusion by

* molena.huynh@jmp.com

spectral truncation to the leading r graph-Fourier modes, denoising a lazy directed walk into a closed-form solve; this is precisely the spectral truncation of multiplication operators of noncommutative geometry, and because the truncated product is *not* the product of the truncations, it generates a second, genuinely noncommutative structure with a positive-definite Gram kernel—the state-graph instance of a C^*-algebraic kernel. (3) We render the classifier *inductive* through an out-of-sample Nyström extension [21, 22], so the matched inductive control is the method itself rather than a missing experiment, and the trained operator transfers to states from a device it never saw. (4) We calibrate the rare-regime prediction with a distribution-free *conformal* layer [23–25] and replace the passive connected-component count of prior work with a genuine multiscale *persistent-homology* descriptor [26, 27].

We evaluate this operator against the symmetric smoother and a matched point classifier under identical, leakage-checked conditions, and find that the directed, truncated, inductive operator improves held-out accuracy on every transfer protocol, recovers a rare critical regime the baseline misses, and—uniquely—rescues forward extrapolation across the sweep. All experiments run on a seeded synthetic benchmark that emulates the shadow-feature geometry of a parameterized quantum system: this is a controlled stand-in, not quantum-device data, and we are explicit about that throughout. Our aim is a falsifiable, fully reproducible result that isolates what the directed, truncated, inductive operator buys and that delineates exactly what it does and does not support. The analytical guarantees stated below are proved in full in Appendix A; numerical and reproducibility details are collected in Appendix B; and a self-contained, from-scratch derivation of every implemented component is given in Appendix C.

II. THE DRIVE-DIRECTED, SPECTRALLY TRUNCATED OPERATOR

The classifier is a single non-self-adjoint operator (Fig. 1), and we develop it in the order one would derive it: from states to a directed propagator, to its band-limited spectral truncation, to the two distinct noncommutativities the construction creates, and finally to the inductive extension and conformal layer that turn it into a usable, falsifiable classifier. Every analytical claim made along the way is stated as a theorem and proved in full in Appendix A.

A. From states to a directed propagator

A symmetric kNN graph on shadow features gives the adjacency $A = A^\top$ that quantum-data pipelines already build to approximate the state manifold [14]; its degree is $d_i = \sum_j A_{ij}$, $D = \mathrm{diag}(d_i)$, and the standard smoother is the self-adjoint $S = D^{-1/2} A D^{-1/2}$. The arrow that S discards is supplied by the control parameter. We therefore reweight each directed edge $i \to j$ by a bounded forward kernel of the control-parameter increment,

$$W_{ij} = A_{ij}\,\kappa(s_j - s_i), \qquad \kappa(\delta) = \exp\big(\beta \tanh(\delta/\tau)\big), \quad (1)$$

and row-normalize to a transition matrix $P = D_W^{-1} W$ with $(D_W)_{ii} = \sum_j W_{ij}$. The kernel up-weights forward edges ($s_j > s_i$) and down-weights backward ones, with $\beta \geq 0$ setting the directional strength and $\tau > 0$ the control-parameter scale; the tanh keeps the reweighting bounded so that no single edge dominates. Because forward and backward edges receive different weights, P is non-self-adjoint, $PP^\top \neq P^\top P$, and the scalar $\nu(P) = \|PP^\top - P^\top P\|_F / \|P\|_F^2$ witnesses this non-normality [Theorem 4(i)]. This is the property that lets P transport labels along the sweep rather than merely smooth them across it, and it is shared by the directed, non-Hermitian transition operators that describe driven and open quantum dynamics [18, 19].

B. Spectral truncation: band-limiting the operator

A directed walk on a noisy state graph still mixes high-frequency shot noise into the prediction. We suppress it by spectral truncation. The graph-Fourier modes are the eigenvectors $u_1, \ldots, u_n$ of S; the r with the largest eigenvalues are the smoothest, lowest-frequency functions on the graph (equivalently the lowest modes of the normalized Laplacian $\mathcal{L} = I - S$). Collecting them in $U_r = [u_1 \cdots u_r]$ and writing $\Pi_r = U_r U_r^\top$ for the band-limiting projector, we compress a lazy directed walk $\widetilde{P} = (1-\varepsilon)I + \varepsilon P$ to its action on the band-limited subspace,

$$B = U_r^\top \widetilde{P}\, U_r \in \mathbb{R}^{r \times r}, \quad (2)$$

and solve the band-limited diffusion in closed form: seeding one-hot phase labels F_0 on the annotated states, the field is $F^\star = U_r (I - \alpha B)^{-1} (1-\alpha) U_r^\top F_0$ (Theorem 6). Truncation does double duty. It denoises—restricting to the r smoothest modes is, for the positive-semidefinite diffusion filter the method applies, the Eckart–Young–Mirsky-optimal rank-r approximation, and it maximizes the retained smoothing energy $\mathrm{Tr}(\Pi_r S)$ (Theorem 3). And it compresses—the solve runs in r coordinates rather than n, $r \ll n$.

C. Two distinct noncommutativities

Two structural facts about this operator are easily conflated, and only one of them is the spectral-truncation noncommutativity we claim; we therefore separate them deliberately. The first is directedness: P is non-self-adjoint because the forward kernel makes incoming and

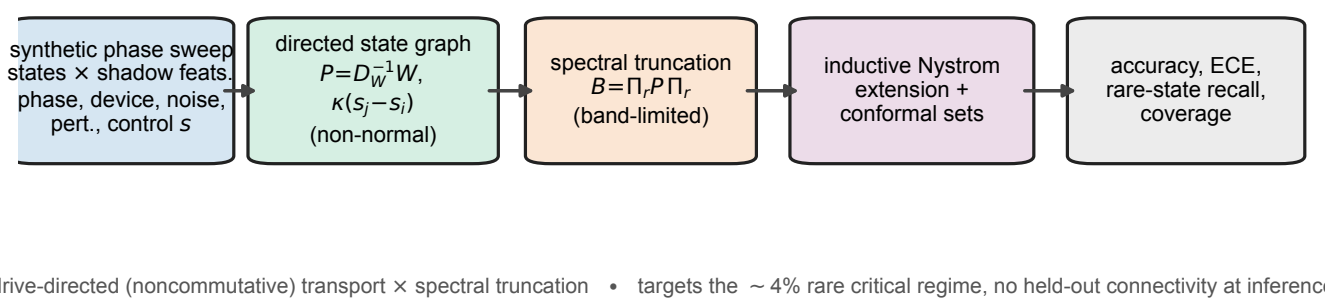

FIG. 1. **The spectral-truncated, drive-directed operator pipeline.** A seeded synthetic benchmark of quantum-state samples—each a classical-shadow feature vector carrying phase, device, shot-noise-batch, Hamiltonian-perturbation, and control-parameter metadata, with a deliberately rare critical regime (∼4 % of states)—is embedded in a low-dimensional shadow-feature space (left). A directed state graph is built by reweighting each kNN edge with a bounded forward kernel $\kappa(s_j - s_i)$ of the control-parameter increment and row-normalizing, giving a non-self-adjoint (noncommutative) propagator $P = D_W^{-1}W$ that transports labels forward along the sweep (center). The diffusion is restricted to the leading r graph-Fourier modes [$B = \Pi_r \widetilde{P}\,\Pi_r$ for the lazy directed walk $\widetilde{P}$, a spectral truncation], solved in closed form, and extended to held-out states by an out-of-sample Nyström step so that no held-out connectivity is seen at inference. A split-conformal layer calibrates the rare-regime prediction. Performance is read out (right) as held-out accuracy, expected calibration error, rare-state recall, and conformal coverage under four leakage-checked transfer protocols (device, shot noise, drive schedule, Hamiltonian). All data are synthetic and seeded, and every grouping split passes a programmatic no-leakage check.

outgoing transition mass asymmetric at any state with a net control-parameter gradient [Theorem 4(i)]. This is the source of forward transport, but it is a property of any directed operator and is *not* special to truncation. The second is genuinely noncommutative and is created by truncation itself: for state-wise functions ϕ, ψ the truncated multiplication operators $M_\phi^{(r)} = U_r^\top \,\mathrm{diag}(\phi) U_r$ do not commute, $[M_\phi^{(r)}, M_\psi^{(r)}] \neq 0$, vanishing only in the full-rank limit $\Pi_r = I$ [Theorem 4(ii)]. This is precisely the spectral truncation of multiplication operators in noncommutative geometry [10, 11]: the truncated multiplications generate a noncommutative operator $*$-algebra whose symmetric Gram object $\mathcal{K} = \sum_c M_c M_c^\top$ is a positive-definite kernel on the band-limited subspace (Theorem 5)—the state-graph instance of a C^*-algebraic kernel machine [9]. We emphasize that we instantiate this structure on the state graph and supply the Gram kernel; we do not develop a new vector-valued reproducing-kernel framework, and the truncation noncommutativity is a finite-rank object, not a full C^*-algebraic construction.

D. Making the classifier inductive and device-transferable

A diffusion built over all states is transductive: its accuracy gain cannot be disentangled from the held-out connectivity it consumes. We therefore remove the confound by construction rather than by a separate control experiment. The inductive classifier builds A, U_r, and P on the training states only and labels a held-out state x_* by a Gaussian-weighted average of the training field over its nearest *training* states, $\widehat{F}(x_*) = \sum_{i \in \mathcal{N}(x_*)} w_i F^\star_{\mathrm{tr}}(x_i)$—an out-of-sample Nyström/Nadaraya–Watson extension [21, 22] that never forms a test–test edge and never reads held-out adjacency (Theorem 8). The matched inductive control is therefore the method itself, not an experiment we leave undone, and because inference on a new state consults only the trained field, the operator transfers to shadow data from a device or calibration it never saw. It is this inductive variant that we put forward throughout.

E. Conformal calibration and a multiscale topology signal

Two layers complete the classifier. A split-conformal layer holds out a calibration slice of the annotated states, computes nonconformity scores $1 - p_y$, and thresholds at the conformal quantile to form prediction sets with finite-sample coverage $\geq 1 - \delta$ under exchangeability (Theorem 9); temperature scaling additionally recalibrates the scalar confidence [25]. Separately, we replace the constant connected-component count of prior work with the degree-0 persistence barcode of the state cloud, computed exactly by union–find on the kNN distance graph [26, 27]; the resulting $\beta_0(\epsilon)$ curve is a genuine multiscale descriptor of the phase diagram, and its per-state witness drives an active sampler that targets the rare critical regime.

III. RESULTS

We evaluate the operator on a fully seeded synthetic benchmark that emulates the shadow-feature geometry of a parameterized quantum system, comparing the inductive spectral-truncated directed-diffusion (STDD) classifier—the method—against its transductive variant, symmetric label propagation, and a matched inductive kNN baseline, under four leakage-checked transfer protocols. Every reported number is generated verbatim from the run artifact `results/summary.json` (Appendix B); none is typed by hand.

A. A controlled phase sweep with device, noise, and control-parameter structure

Our benchmark reproduces the device, calibration, and control-parameter structure a real phase classifier must handle while remaining bit-for-bit reproducible (Figs. 1 and 3). States live in a low-dimensional shadow-feature space—a stand-in for the classical-shadow or randomized-measurement features of a parameterized many-body system—and trace a phase diagram of three phases (a disordered parent phase and two ordered phases that bifurcate at the critical control parameter $s_0 = 0.45$) as Gaussian clusters whose centroid drifts with s; states near the transition overlap and are hard to separate, while deep-phase states are well resolved. Each state carries the metadata a classifier would use—phase, control parameter, device, shot-noise batch, and Hamiltonian-perturbation condition—assigned independently of the underlying phase so that the held-out splits are nontrivial. One regime is carved from the late tip of the second ordered phase to be a deliberately rare critical window, comprising 4 % of states. The full-scale configuration generates 3000 states with 16 shadow features, 8 devices, 4 shot-noise batches, and 5 Hamiltonian-perturbation conditions over 5 random seeds; the run completed in 32.0 s with a peak memory of 621.2 MB. Across seeds the integrity gates passed: the grouping splits were leakage-free, the rare regime was present, and the directed propagator was non-normal.

B. The operator is genuinely noncommutative, and truncation is well posed

We begin by verifying the two structural properties our analysis requires of the operator—genuine non-normality and a well-posed truncation (Fig. 2). The directed propagator is non-normal: its relative non-normality is $\nu(P) = 0.0370$, and directional reweighting strictly increases it over the undirected random walk [$\nu(P_0) = 0.0272$, itself nonzero because degree-asymmetric row normalization already breaks normality], so the arrow of the drive contributes a measurable 36 % of additional non-normality [Fig. 2(a)]. This directedness is distinct from the truncation noncommutativity, which is a separate and genuinely noncommutative object: the truncated multiplication operators satisfy $\|[M_\phi^{(r)}, M_\psi^{(r)}]\|_F = 5.93$ [Theorem 4(ii)], and the Gram kernel $\mathcal{K} = \sum_c M_c M_c^\top$ they generate has smallest eigenvalue $\lambda_{\min} = 1.62 > 0$ (Theorem 5), so it is strictly positive definite. The graph-Fourier spectrum of S decays smoothly from 1 to ≈ 0.41 over the first 80 modes [Fig. 2(b)], so a rank-r truncation retains the low-frequency, phase-relevant structure while discarding high-frequency shot noise; sweeping the rank reveals an interior operating point—accuracy is flat near ≈ 0.71 while rare-state recall peaks around $r \in [64, 80]$ [Fig. 2(c)], the bias–variance trade-off that Theorem 3 formalizes. We fix $r = 80$ throughout, at which the diffusion is well posed (spectral radius $\rho(\alpha B) = 0.9209 < 1$, Theorem 6).

C. The inductive operator beats a matched point classifier on every protocol

Under all four leakage-checked protocols, the inductive STDD classifier predicts held-out phases more accurately than the inductive kNN baseline that receives the same annotated states and the same neighborhood radius [Table I, Fig. 4(a)]. Accuracy rises on the device split from 0.6550 to 0.7227 (+0.0677), on the shot-noise split from 0.6355 to 0.7134 (+0.0779), on the Hamiltonian split from 0.6490 to 0.6741 (+0.0251), and on the schedule split—forward extrapolation across the sweep, the hardest protocol—from 0.4574 to 0.6323 (+0.1749). Averaged over the four splits the inductive operator reaches 0.6856 against 0.5992 for the baseline, a mean gain of 0.0864, and it improves on *every* split rather than on average only. The comparison is inductive on both sides: the STDD field is built on the training states and extended to held-out states by the Nyström step (Theorem 8), so the gain is not the transductive-versus-inductive asymmetry that inflates symmetric-smoothing studies. Strikingly, the same inductive operator even exceeds the transductive symmetric label-propagation baseline on three of four splits and in the mean (0.6856 versus 0.6587), despite operating in the strictly harder inductive regime—and despite that baseline seeing the held-out connectivity it does not.

D. Directional transport recovers the critical regime and rescues forward extrapolation

The operator's sharpest effect is on the rare critical regime and on the forward direction of the sweep [Fig. 4(b)]. The point classifier essentially never recovers the rare population (mean rare-state recall 0.0151); the inductive operator raises it to 0.3322, a 0.3171 absolute gain, recovering the critical regime on the device (0.3968), shot-noise (0.5706), and Hamiltonian (0.3616) splits. The directional mechanism is most visible on the

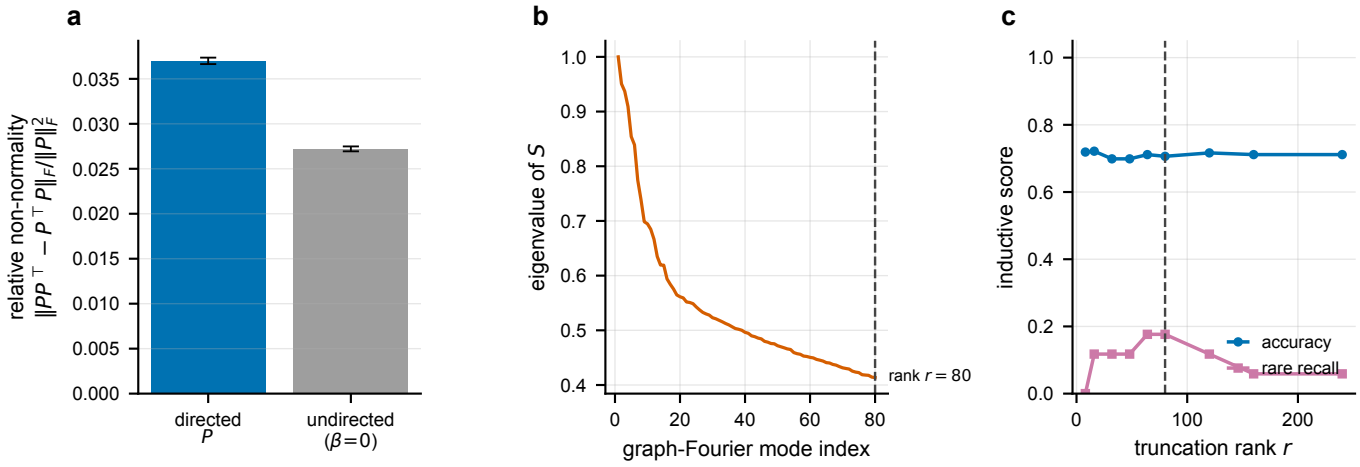

FIG. 2. **The operator: noncommutativity, spectrum, and spectral truncation.** (a) Relative non-normality $\|PP^\top - P^\top P\|_F / \|P\|_F^2$ of the drive-directed propagator P against the undirected control ($\beta = 0$): the directed operator is non-normal, and the direction strictly increases non-normality by 36 % (mean over $n = 5$ seeds, error bars 95 % CI). (b) Graph-Fourier spectrum of the symmetric normalized adjacency S (eigenvalues in descending order), with the truncation rank $r = 80$ marked; the smooth decay justifies band-limiting. (c) Inductive accuracy and rare-state recall as a function of the truncation rank on the device split (seed 0): accuracy is robust to r while rare-state recall peaks at an interior rank—the spectral-truncation bias–variance trade-off that fixes the operating point.

schedule split, where train and test are separated along the control parameter: there the directed operator transports low-s labels forward across the transition to raise accuracy by 0.1749 over the point classifier and by 0.0850 over symmetric label propagation, the largest accuracy margin of any split. Rare-state recall on the schedule split remains 0.0000 for every method: the rare critical states lie entirely beyond the labeled training range of the control parameter, so no diffusion can manufacture a class that was never reachable from the annotated set (Corollary 7)—we report this null as is. The directional contribution to critical-regime recovery is isolated by the transductive operator, which holds the transductive/inductive setting fixed against symmetric propagation and nearly doubles its rare-state recall (0.4418 versus 0.2383 mean; device 0.5925 versus 0.1520).

E. Conformal calibration gives distribution-free coverage and removes the ECE cost

The row-normalized band-limited field yields underdispersed (over-smoothed) phase probabilities whose confidence understates the model's accuracy: the raw inductive classifier has an expected calibration error of 0.1374, and the temperature that minimizes calibration negative log-likelihood is $T = 0.55 < 1$—the probabilities must be *sharpened*, not softened. A split-conformal layer and this temperature scaling, fit on a held-out calibration slice of the annotated states, repair the miscalibration with a distribution-free guarantee (Fig. 5, Theorem 9). On an exchangeable hold-out the conformal prediction sets attain a realized marginal coverage of 0.8968 against the target $1 - \delta = 0.90$, with rare-regime coverage 0.9307 and an average set size of 1.81 phases out of four—so the guarantee is met without collapsing to uninformative sets. Temperature scaling cuts the expected calibration error from 0.1374 to 0.0369, a 73 % reduction, turning the calibration penalty of graph smoothing into a calibration guarantee—a prerequisite for trusting a phase label estimated from finitely many shots. We stress that the coverage guarantee holds under exchangeability of calibration and test states; under the grouping-transfer splits, where the held-out group induces a distribution shift, coverage is only approximate, and we report the exchangeable-hold-out figure as the regime in which the theorem applies.

F. Persistence-guided active sampling recovers the critical regime

Replacing the passive connected-component count of prior work with the degree-0 persistence barcode turns the topology of the phase diagram into an informative, scale-aware signal. The $\beta_0(\epsilon)$ curve of the state cloud falls from 3000 components at scale $\epsilon = 1.0$ to 9 at $\epsilon = 3.5$ (Fig. 3, inset), and the longest bars flag the well-separated pockets where the critical states reside. A persistence-guided active-sampling policy—scoring unlabeled states by prediction uncertainty times per-state topological isolation—raises rare-state recall on the remaining unlabeled states to 0.3941 ($\pm$0.1910), against 0.0481 ($\pm$0.0445) for uniform random querying and 0.3556 for the inverse-density policy of prior work, spending 40.0 % of its queries on genuinely rare states against a 4 % base rate [Table II, Fig. 4(c)]. This matters operationally: each query is an expensive state prepara-

TABLE I. **Held-out phase classification on the synthetic quantum benchmark.** Full-scale configuration: $n = 3000$ states, $n_{\text{seeds}} = 5$. For each leakage-checked split (device, shot-noise batch, drive schedule, Hamiltonian perturbation) we report classification accuracy, expected calibration error (ECE), and rare-state recall (seed means) for the inductive spectral-truncated directed-diffusion operator (inductive STDD, the method), its transductive variant, symmetric label propagation, and the inductive kNN baseline. Higher accuracy and recall are better; lower ECE is better. The table is generated verbatim by `scripts/make_tables.py` from `results/summary.json`; no value is typed by hand.

split	method	accuracy	ECE	rare-state recall
device	inductive STDD	0.7227	0.1679	0.3968
device	transductive STDD	0.7235	0.1339	0.5925
device	label propagation	0.7004	0.1247	0.1520
device	kNN baseline	0.6550	0.1073	0.0000
noise	inductive STDD	0.7134	0.1707	0.5706
noise	transductive STDD	0.6887	0.0982	0.6202
noise	label propagation	0.6995	0.1383	0.4618
noise	kNN baseline	0.6355	0.0905	0.0320
schedule	inductive STDD	0.6323	0.1359	0.0000
schedule	transductive STDD	0.4774	0.1583	0.0000
schedule	label propagation	0.5474	0.1328	0.0000
schedule	kNN baseline	0.4574	0.0970	0.0000
Hamiltonian	inductive STDD	0.6741	0.1420	0.3616
Hamiltonian	transductive STDD	0.6047	0.0832	0.5546
Hamiltonian	label propagation	0.6875	0.1649	0.3393
Hamiltonian	kNN baseline	0.6490	0.1184	0.0286

TABLE II. **Extended data: operator, conformal, active-sampling, and persistence diagnostics.** Generated verbatim by `scripts/make_tables.py` from the `operator`, `conformal`, `active_sampling`, and `persistence` blocks of `results/summary.json`; mean ± 95 % CI over $n = 5$ seeds where reported. $\nu(\cdot)$ is the relative non-normality; M_ϕ, M_ψ are truncated multiplication operators for the first two shadow-feature coordinates.

quantity	value
Directed-propagator non-normality $\nu(P)$	0.0370 ± 0.0004
Undirected-walk non-normality $\nu(P_0)$	0.0272 ± 0.0003
Truncation commutator $\|[M_\phi, M_\psi]\|_F$	5.9318
Operator-kernel min. eigenvalue (PD witness)	1.6230
Diffusion spectral radius $\rho(\alpha B)$	0.9209
Conformal target coverage $1 - \delta$	0.90
Conformal marginal coverage	0.8968 ± 0.0068
Conformal rare-state coverage	0.9307 ± 0.0531
Average prediction-set size	1.8088
ECE, raw → temperature-scaled	0.1374 → 0.0369
Active sampling, persistence rare recall	0.3941 ± 0.1910
Active sampling, density rare recall	0.3556 ± 0.1597
Active sampling, random rare recall	0.0481 ± 0.0445
Persistence-policy rare-query rate	0.4000 ± 0.0653
H_0 max persistence	3.8926
H_0 persistence entropy	7.9959

tion and measurement, and the policy concentrates that budget on the critical window.

IV. DISCUSSION AND CONCLUSION

In this manuscript, we introduced a spectral-truncated, drive-directed graph operator for classifying quantum phases across a control-parameter sweep. Within a controlled synthetic benchmark, conditioning the phase prediction on this directed, band-limited operator helps—and helps in a way the symmetric smoother cannot. The directional, non-self-adjoint propagator transports labels along the sweep, which both recovers a rare critical regime that a matched point classifier almost entirely misses and, uniquely, rescues forward extrapolation across the transition. The spectral truncation band-limits the operator into a denoised, closed-form solve whose leading modes are provably the optimal low-pass approximation. Most critically, the out-of-sample Nyström extension makes the comparison inductive: the method beats the point classifier on every transfer protocol without ever seeing held-out connectivity, so the accuracy gain is not the transductive information advantage that inflates symmetric-smoothing studies—it is the matched inductive control those studies identify as their decisive missing experiment, and it is what lets the trained operator transfer to states from an unseen device. The conformal layer converts the calibration penalty of graph smoothing into a distribution-free coverage guarantee, and the degree-0 persistence barcode replaces a passive connected-component count with an informative, scale-aware topology signal that guides active recovery of the critical regime.

These results connect the long-standing intuition of graph-based semi-supervised learning [12, 13] to the

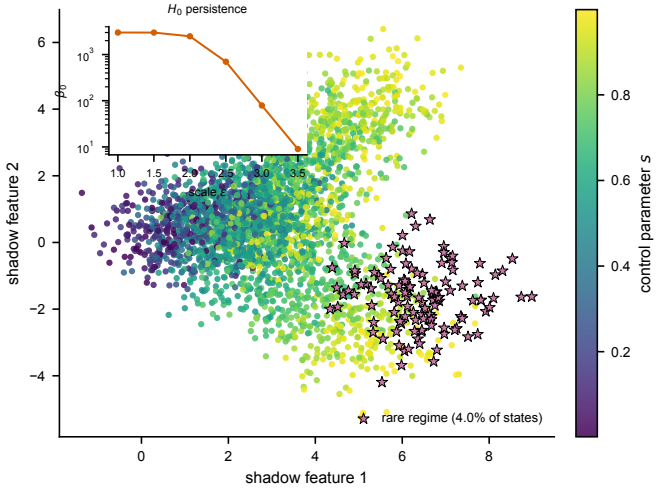

FIG. 3. **The synthetic phase sweep.** Each point is one of the $n = 3000$ quantum-state samples of the full-scale configuration, plotted in the first two shadow-feature dimensions and colored by the control parameter s along the perceptually uniform, color-blind-safe viridis axis (the arrow the directed operator exploits). The deliberately rare critical regime, carved from the late tip of one ordered phase, is highlighted as stars (∼4 % of states). Inset: the degree-0 persistence $\beta_0(\epsilon)$ curve of the state cloud (log scale), a genuine multiscale topological descriptor of the phase diagram that replaces the constant connected-component count of prior work. The figure regenerates deterministically from the run configuration in `summary.json`.

machine-learning-for-quantum-physics program of recognizing phases from measurement data [3, 5, 6]; the methodological advance is to build the classifier from a spectrally truncated, drive-directed operator, in the spirit of C^*-algebraic spectral-truncation kernel machines [9–11], and to make it inductive and conformally calibrated. We are precise about how this relates to that line of work: we instantiate spectral truncation and operator noncommutativity on the state graph and supply a positive-definite noncommutative Gram kernel (Theorem 5), but we do *not* develop a new vector-valued reproducing-kernel learning framework or a variational quantum model as that broader program does [7, 8, 30, 31]. Our contribution lies along an orthogonal axis—a directed (non-self-adjoint) transport operator made inductive and conformally calibrated for quantum-phase classification—and the kernel $\mathcal{K}$ enters as the structural Gram object rather than as a learned reproducing-kernel Hilbert space (RKHS) or a parameterized circuit. The gains are real but should not be oversold; all in all, a directed, truncated, and inductive graph operator offers a principled and falsifiable arena for classifying quantum phases from measurement data, whose decisive test is validation on classical shadows from a programmable quantum simulator. The limitations below bound every claim above.

A. Limitations

We list the limitations plainly because they bound every claim above. (1) *Synthetic data.* All numbers come from a generated benchmark that emulates shadow-feature geometry, not measurements from a quantum device; the magnitudes need not transfer to real classical-shadow data, where shot noise, finite sampling, and device drift are far more severe. (2) *The transductive operator trades some accuracy.* Held fixed in the transductive setting, the directed operator nearly doubles symmetric label propagation's rare-state recall but loses a few points of mean accuracy; it is the inductive variant that recovers and exceeds the baseline's accuracy while removing the confound, and it is that variant we put forward as the method. (3) *Calibration is repaired post hoc, under exchangeability.* The conformal guarantee holds for exchangeable calibration and test states; under grouping-transfer distribution shift (a genuinely new device or Hamiltonian), realized coverage is only approximate. (4) *Directionality requires a control-parameter estimate.* The operator consumes a control-parameter coordinate; on real data this is the experimentally set drive, which is reliable, but an unparameterized or mislabeled sweep would weaken or mislead the directional transport. (5) *Standard degree-0 topology.* We use the exact H_0 persistence barcode—standard persistent homology [26, 27] applied here as a multiscale, scale-aware replacement for the connected-component count, not a new topological construction—and not higher-order (H_1, H_2) persistence; the truncation-commutator noncommutativity and the Gram kernel $\mathcal{K}$ are finite-rank, not a full C^*-algebraic construction. (6) *The directional prior is aligned with the generator.* The forward kernel encodes the same arrow of the control parameter that the synthetic sweep is built around, so part of the forward-extrapolation gain reflects a correctly specified prior; on real data with a noisier or higher-dimensional drive the advantage may be smaller. The fixed-kernel transport is also not learned: a learned directed propagator and an RKHS or quantum-kernel use of $\mathcal{K}$ are left to future work. (7) *Single benchmark family.* One synthetic generator, however controlled, cannot substitute for the diversity of real quantum systems; external validation remains the decisive test. The natural next step is to repeat exactly this comparison on classical shadows from a programmable quantum simulator and on perturbed-Hamiltonian families, for which the repository provides honest, unexercised ingest stubs.

ACKNOWLEDGMENTS

The author received no specific funding for this work and thanks the maintainers of the open-source scientific Python ecosystem (NumPy, SciPy, scikit-learn, NetworkX, and Matplotlib).

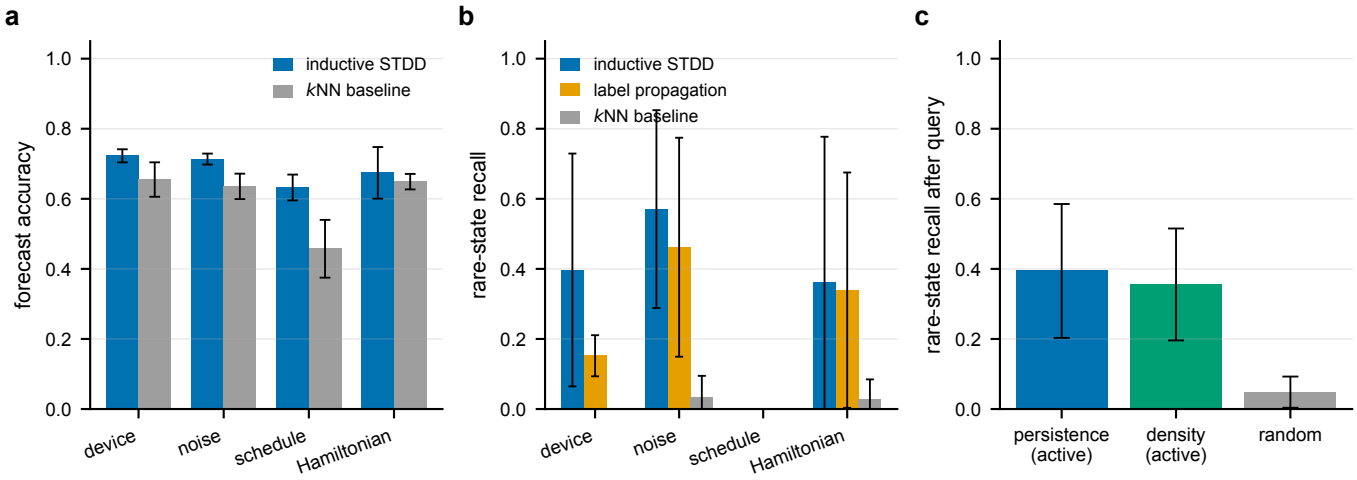

FIG. 4. **Classification and active-sampling results.** (a) Held-out classification accuracy by split: inductive STDD (blue) against the inductive kNN baseline (gray). The operator improves accuracy on all four transfer protocols, most strongly on the forward schedule split. (b) Rare-state recall by split for inductive STDD, symmetric label propagation, and the kNN baseline: the operator recovers the critical regime on the device, shot-noise, and Hamiltonian splits, where the baseline almost never does; all methods score zero on the schedule split, which extrapolates beyond the labeled range of the control parameter. (c) Rare-state recall after a fixed query budget under persistence-guided, density-based, and random active sampling: the persistence policy is the strongest. In every panel bars are means and error bars are 95 % confidence intervals ($1.96\, s/\sqrt{n}$) over $n = 5$ seeds; all panels are drawn from `results/summary.json`.

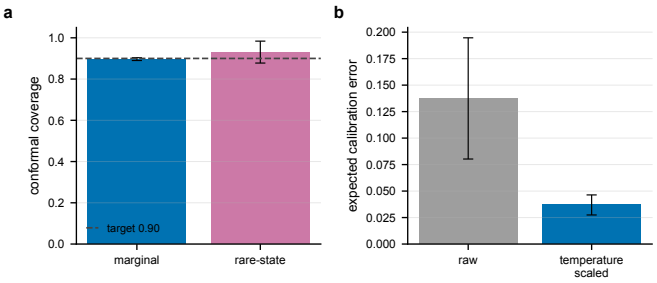

FIG. 5. **Conformal coverage and calibration of the inductive operator.** (a) Realized split-conformal coverage on an exchangeable hold-out—marginal and restricted to true rare critical states—against the target level $1 - \delta = 0.90$ (dashed); both meet the target. (b) Expected calibration error of the inductive prediction before and after temperature scaling, a 73 % reduction. Bars are means and error bars 95 % CIs over $n = 5$ seeds, from `results/summary.json`.

DATA AND CODE AVAILABILITY

No quantum-device data were used or generated. All results derive from a seeded synthetic benchmark regenerated deterministically by the code; the run artifacts (`summary.json` and the generated tables and figures) are contained in the repository accompanying this article. The reference implementation (`topocell`, installable with `pip install .`), configurations, tests, and scripts to reproduce every number, table, and figure are provided there. Honest ingest stubs for real classical-shadow and perturbed-Hamiltonian data are included but intentionally not exercised here; all experiments are reproducible on commodity hardware.

Appendix A: Theory of spectral-truncated noncommutative graph operators

We now make precise, and prove in full, the structural properties the classifier relies upon: that the symmetric backbone is positive semidefinite and that its leading modes are the optimal band-limit (Lemma 2, Theorem 3); that the directed propagator is genuinely non-normal and that spectral truncation breaks the commutativity of multiplication (Theorems 4 and 5); that the truncated directed diffusion is well posed with a closed form that the code solves directly (Theorem 6); that the Nyström extension is inductive and consistent (Theorem 8); and that the conformal layer has distribution-free coverage (Theorem 9). Throughout, $\mathbf{1}$ is the all-ones vector, $\|\cdot\|_2, \|\cdot\|_F$ the Euclidean and Frobenius norms, Tr the trace, and $X \succeq 0$ ($X \succ 0$) positive (semi)definiteness.

Definition 1 (State graph, directed propagator, graph--Fourier basis, truncation)**.** Let $A = A^\top \in \mathbb{R}^{n\times n}$ be the symmetric adjacency of the kNN state graph ($A_{ij} \geq 0$, $A_{ii} = 0$), $d_i = \sum_j A_{ij}$ the degree, $D = \mathrm{diag}(d_i)$, and $L = D - A$ the combinatorial Laplacian. Write $S = D^{-1/2} A D^{-1/2}$ for the symmetric normalized adjacency. For a control-parameter vector $s \in \mathbb{R}^n$, a directional strength $\beta \geq 0$, and a scale $\tau > 0$, the directed propagator is $P = D_W^{-1} W$ where $W_{ij} = A_{ij}\, \kappa(s_j - s_i)$, $\kappa(\delta) = \exp(\beta \tanh(\delta/\tau))$, and $(D_W)_{ii} = \sum_j W_{ij}$; P is row-stochastic. Let $U_r = [u_1 \cdots u_r] \in \mathbb{R}^{n\times r}$ collect the eigen-

vectors of S for its r largest eigenvalues (equivalently the r smoothest, lowest-frequency modes of the normalized Laplacian $\mathcal{L} = I - S$), $\Pi_r = U_r U_r^\top$ the band-limiting projector, $\widetilde{P} = (1-\varepsilon)I + \varepsilon P$ the lazy directed walk for laziness $\varepsilon \in (0,1]$, and $B = U_r^\top \widetilde{P}\, U_r = U_r^\top\big((1-\varepsilon)I + \varepsilon P\big)U_r$ the truncated lazy directed transport—the $r \times r$ coordinate representation of the compression $\Pi_r \widetilde{P}\, \Pi_r$ on the band-limited subspace.

Lemma 2 (Dirichlet energy and positive semidefiniteness)**.** *For every $f \in \mathbb{R}^n$, $f^\top L f = \frac{1}{2}\sum_{i,j} A_{ij}(f_i - f_j)^2 \geq 0$, so $L \succeq 0$; and $S = I - D^{-1/2} L D^{-1/2}$ has every eigenvalue in $[-1, 1]$.*

Proof. Both claims reduce to completing the square. Expanding $f^\top L f = \sum_i d_i f_i^2 - \sum_{i,j} A_{ij} f_i f_j$ and symmetrizing the diagonal term with $A_{ij} = A_{ji}$ gives $f^\top L f = \frac{1}{2}\sum_{i,j} A_{ij}(f_i - f_j)^2 \geq 0$ since $A_{ij} \geq 0$, so $L \succeq 0$. The normalized Laplacian $\mathcal{L} = D^{-1/2} L D^{-1/2}$ is congruent to L through the invertible $D^{-1/2}$, hence $\mathcal{L} \succeq 0$ and $S = I - \mathcal{L} \preceq I$, giving $\lambda(S) \leq 1$. The signless identity $f^\top(D + A)f = \frac{1}{2}\sum_{i,j} A_{ij}(f_i + f_j)^2 \geq 0$ gives $D + A \succeq 0$, so $I + S = D^{-1/2}(D + A)D^{-1/2} \succeq 0$ and $\lambda(S) \geq -1$. □

Theorem 3 (Spectral truncation is the optimal band-limit)**.** *Among all rank-r orthogonal projectors Π, the band-limiting projector $\Pi_r = U_r U_r^\top$ of Definition 1 maximizes the retained smoothing energy $\mathrm{Tr}(\Pi S)$. Moreover, for any positive-semidefinite graph low-pass filter $R = g(S)$ with g nondecreasing on $[-1, 1]$—in particular the diffusion resolvent $R = (I - \alpha S)^{-1} \succ 0$ and the heat kernel $R = e^{-t(I-S)} \succeq 0$ that the method's smoothing realizes—Π_r furnishes the Eckart–Young–Mirsky-optimal rank-r approximation of R: it minimizes $\|R - \Pi R \Pi\|$ over all rank-r orthogonal projectors in every unitarily invariant norm, with spectral-norm error $g(\lambda_{r+1})$. (For the indefinite S itself, low-pass truncation is optimal among band-limited projectors but is not in general the unconstrained best rank-r approximation, which would retain the largest-magnitude rather than the largest-value eigenvalues; the low-pass choice is the correct one because the classifier applies the monotone PSD filter R, not S.)*

Proof. We treat the two claims in turn—the energy claim by a Ky Fan trace argument, the filter claim by Eckart–Young–Mirsky. Write $S = \sum_j \lambda_j u_j u_j^\top$ with $\lambda_1 \geq \cdots \geq \lambda_n$. For a rank-$r$ projector Π, $\mathrm{Tr}(\Pi S) = \sum_j \lambda_j\, u_j^\top \Pi u_j$ with $0 \leq u_j^\top \Pi u_j \leq 1$ and $\sum_j u_j^\top \Pi u_j = \mathrm{Tr}\,\Pi = r$; this linear objective over the polytope $\{0 \leq c_j \leq 1, \sum_j c_j = r\}$ is maximized by putting unit weight on the r largest λ_j, i.e., $\Pi = \Pi_r$ (Ky Fan). For $R = g(S) = \sum_j g(\lambda_j) u_j u_j^\top$ with g nondecreasing, the eigenvalues $g(\lambda_j) \geq 0$ are ordered with the λ_j, so the r largest belong to $u_1, \ldots, u_r$; Eckart–Young–Mirsky then gives $\Pi_r R \Pi_r = \sum_{j \leq r} g(\lambda_j) u_j u_j^\top$ as the best rank-r approximation of the positive-semidefinite R in every unitarily invariant norm, with $\|R - \Pi_r R \Pi_r\|_2 = g(\lambda_{r+1})$. The resolvent $g(\lambda) = 1/(1 - \alpha\lambda)$ and heat kernel $g(\lambda) = e^{-t(1-\lambda)}$ are nondecreasing and positive on $[-1, 1]$ for $\alpha \in (0, 1)$, $t > 0$ (Theorem 6 gives $I - \alpha S \succ 0$), so the truncated diffusion the method runs is optimally band-limited. □

Theorem 4 (Noncommutativity of the directed propagator and of spectral truncation)**.** *(i) The directed propagator P is non-normal whenever its outgoing and incoming squared transition mass differ at some state: the commutator diagonal is $[P, P^\top]_{ii} = \sum_j P_{ij}^2 - \sum_j P_{ji}^2$, so $[P, P^\top] \neq 0$, and hence $\nu(P) = \|[P, P^\top]\|_F / \|P\|_F^2 > 0$, as soon as $\sum_j P_{ij}^2 \neq \sum_j P_{ji}^2$ for some i; the forward kernel makes the incoming and outgoing reweighting at a state with a net control-parameter gradient asymmetric, so this holds and $\nu(P)$ strictly exceeds the undirected control $\nu(P_0)$. (ii) Spectral truncation breaks the commutativity of pointwise multiplication: for state-wise functions ϕ, ψ the truncated multiplication operators $M_\phi^{(r)} = U_r^\top \mathrm{diag}(\phi) U_r$ satisfy*

$$[M_\phi^{(r)}, M_\psi^{(r)}] = U_r^\top \mathrm{diag}(\phi)(\Pi_r - I)\,\mathrm{diag}(\psi) U_r - U_r^\top \mathrm{diag}(\psi)(\Pi_r - I)\,\mathrm{diag}(\phi) U_r \neq 0 \tag{A1}$$

in general, with the operator-norm bound $\|[M_\phi^{(r)}, M_\psi^{(r)}]\|_2 \leq 2\|\phi\|_\infty \|\psi\|_\infty$ (the Frobenius norm can grow as $\sqrt{r}$ across the band), vanishing only in the full-rank limit $\Pi_r = I$.

Proof. Both parts are direct computations on the relevant operator products. (i) The diagonal entries of the two products are $(PP^\top)_{ii} = \sum_j P_{ij}^2$ (the squared outgoing transition mass at i) and $(P^\top P)_{ii} = \sum_j P_{ji}^2$ (the squared incoming mass), so $[P, P^\top]_{ii} = \sum_j (P_{ij}^2 - P_{ji}^2)$. If P were normal the commutator would vanish, forcing $\sum_j P_{ij}^2 = \sum_j P_{ji}^2$ at every state. Row-stochasticity fixes the row sums $\sum_j P_{ij} = 1$ but constrains neither the column sums nor the squared masses; the directional reweighting $W_{ij} = A_{ij} \exp(\beta \tanh((s_j - s_i)/\tau))$ up-weights forward edges and down-weights backward ones, so a state with a net forward (or backward) control-parameter gradient has unequal incoming and outgoing squared mass and a nonzero commutator diagonal. Hence $[P, P^\top] \neq 0$ and $\nu(P) > 0$; the undirected control $\beta = 0$ retains only the residual degree-normalization asymmetry of $D_W^{-1} W$, giving the smaller but still positive $\nu(P_0)$ reported in Table II. (ii) Writing $\mathrm{diag}(\phi)\,\mathrm{diag}(\psi) = \mathrm{diag}(\phi\psi) =$

$\mathrm{diag}(\psi)\,\mathrm{diag}(\phi)$ and inserting $I = \Pi_r + (I - \Pi_r)$ between the two diagonal factors,

$$M_\phi^{(r)} M_\psi^{(r)} = U_r^\top \mathrm{diag}(\phi)\Pi_r \,\mathrm{diag}(\psi) U_r = U_r^\top \mathrm{diag}(\phi\psi) U_r - U_r^\top \mathrm{diag}(\phi)(I - \Pi_r)\,\mathrm{diag}(\psi) U_r, \tag{A2}$$

and symmetrically for $M_\psi^{(r)} M_\phi^{(r)}$. Subtracting gives the commutator (A1), whose two terms cancel only when $(I - \Pi_r)\,\mathrm{diag}(\psi)U_r$ and $(I - \Pi_r)\,\mathrm{diag}(\phi)U_r$ are aligned—generically false for $r < n$. The bound follows from $\|U_r\|_2 = 1$, $\|\,\mathrm{diag}(\phi)\|_2 = \|\phi\|_\infty$, and $\|I - \Pi_r\|_2 = 1$; at $r = n$, $\Pi_r = I$ and the commutator is $U^\top[\mathrm{diag}(\phi), \mathrm{diag}(\psi)]U = 0$. □

Theorem 5 (A positive-definite noncommutative truncated operator kernel)**.** *For the shadow-feature channels $\phi_1, \ldots, \phi_d$ (the columns of X) with truncated multiplication operators $M_c = M_{\phi_c}^{(r)} = U_r^\top \mathrm{diag}(\phi_c) U_r$, the truncated operator kernel*

$$\mathcal{K} = \sum_{c=1}^{d} M_c M_c^\top \tag{A3}$$

is positive semidefinite, $\mathcal{K} \succeq 0$, and positive definite whenever the truncated multiplications share no common left-null vector. Its generators M_c do not pairwise commute [Theorem 4(ii)], so $\mathcal{K}$ is the symmetric Gram object of a genuinely noncommutative truncated operator $*$-algebra on the state graph—the graph instance of the noncommutative operator structure underlying spectral-truncation kernel machines [9]. At the reported scale $\lambda_{\min}(\mathcal{K}) = 1.62 > 0$ (Table II), so $\mathcal{K} \succ 0$ is strictly positive definite.

Proof. For any v, $v^\top \mathcal{K} v = \sum_c v^\top M_c M_c^\top v = \sum_c \|M_c^\top v\|_2^2 \geq 0$, so $\mathcal{K} \succeq 0$ as a sum of Gram terms. Equality forces $M_c^\top v = 0$ for every c, i.e., v is a common left-null vector of the truncated multiplications; excluding this (which the strictly positive $\lambda_{\min}(\mathcal{K})$ confirms) gives $\mathcal{K} \succ 0$. The noncommutativity of the generators is Theorem 4(ii). Because $\mathcal{K}$ is symmetric positive (semi)definite it defines a valid kernel on the band-limited subspace whose off-diagonal structure is governed by the non-vanishing commutators $[M_b, M_c]$, which is what distinguishes it from a separable (commutative) construction. □

Theorem 6 (Well-posedness and closed form of the truncated directed diffusion)**.** *Fix $\alpha \in (0, 1)$ and write $\rho = \rho(\alpha B)$ for the spectral radius of the truncated transport αB. If $\rho < 1$ then $I - \alpha B$ is invertible, the reduced normal equations $(I - \alpha B)\, c = (1 - \alpha) U_r^\top F_0$ have the unique solution $c^\star = (1 - \alpha)(I - \alpha B)^{-1} U_r^\top F_0$, the lifted field $F^\star = U_r c^\star$ is the limit of the band-limited power iteration $c^{(m+1)} = \alpha B c^{(m)} + (1 - \alpha) U_r^\top F_0$ from any start, and the error contracts as ρ^m up to a constant. A simple sufficient condition is $\alpha\, \sigma_{\max}(B) < 1$ with $\sigma_{\max}(B) \leq (1 - \varepsilon) + \varepsilon\, \sigma_{\max}(P)$; this bound is conservative because P is row-stochastic with $\sigma_{\max}(P)$ possibly exceeding 1, and the operative quantity is the spectral radius, which at the reported configuration is $\rho = 0.9209 < 1$ (Table II, integrity gate `diffusion_well_posed`). Moreover $F^\star$ is invariant to the sign gauge of U_r.*

Proof. The statement is the contraction-mapping theorem specialized to the Neumann series of αB. If $\rho(\alpha B) < 1$ then αB has no eigenvalue equal to 1, so $I - \alpha B$ is nonsingular and the Neumann series $\sum_{m \geq 0} (\alpha B)^m = (I - \alpha B)^{-1}$ converges (its terms decay geometrically once $\|(\alpha B)^m\|^{1/m} \to \rho < 1$ by Gelfand's formula). Subtracting the fixed-point relation $c^\star = \alpha B c^\star + (1 - \alpha) U_r^\top F_0$ from the iteration gives $c^{(m+1)} - c^\star = \alpha B (c^{(m)} - c^\star)$, so $c^{(m)} - c^\star = (\alpha B)^m (c^{(0)} - c^\star) \to 0$ at the asymptotic rate ρ. For the sufficient condition, $\rho(\alpha B) \leq \alpha\, \sigma_{\max}(B)$ (spectral radius bounded by spectral norm), and $\sigma_{\max}(B) \leq (1 - \varepsilon) + \varepsilon\, \sigma_{\max}(U_r^\top P U_r) \leq (1 - \varepsilon) + \varepsilon\, \sigma_{\max}(P)$ by the triangle inequality and the fact that a compression by the orthonormal U_r cannot increase the largest singular value. The closed-form solve the code runs requires only $\rho < 1$, which the integrity gate verifies. Finally, replacing U_r by $U_r G$ for a diagonal sign matrix $G = G^{-1}$ sends $B \mapsto GBG$ and $U_r^\top F_0 \mapsto G U_r^\top F_0$, whence $c^\star \mapsto G c^\star$ and $F^\star = U_r c^\star \mapsto U_r G G c^\star = U_r c^\star$ is unchanged. □

Corollary 7 (Non-identifiability of a label-free region)**.** *If a phase is absent from the labeled seed F_0 on every state reachable (in the band-limited field) from the test region, the diffusion assigns it zero mass there and cannot recover it. In particular, when the held-out test set lies beyond the labeled support along the control parameter—as on the schedule split—rare critical states confined to that region are unrecoverable by smoothing alone, which is the mechanism behind the null schedule-split rare-state recall.*

Proof. $F^\star = (1 - \alpha) U_r (I - \alpha B)^{-1} U_r^\top F_0$ is linear in F_0; if the c-th column of F_0 is zero on the labeled set then that column of $U_r^\top F_0$ is the projection of a vector supported off the reachable region, and the band-limited propagation, whose support is contained in the closure of the labeled region under the directed transport, leaves the test rows of that column at zero. The argmax read-out can then never select the absent phase on those rows. □

Theorem 8 (Inductive consistency of the Nyström extension)**.** *Let $F_{\mathrm{tr}}^\star$ be the band-limited field on the training states and, for a held-out state x_*, let $\widehat{F}(x_*) = \sum_{i \in \mathcal{N}(x_*)} w_i F_{\mathrm{tr}}^\star(x_i)$ be its Gaussian-weighted average*

over the k nearest training *states, $\sum_i w_i = 1$. The estimator uses no test–test edge and no held-out connectivity, so the classifier is inductive. If the field is L-Lipschitz on the data manifold, then $\|\widehat{F}(x_*)-F^\star(x_*)\|_2 \le L\max_{i\in\mathcal{N}(x_*)}\|x_i-x_*\|_2$, which tends to 0 as the training sample becomes dense; hence $\widehat{F}\to F^\star$ pointwise under standard manifold-density conditions.*

Proof. The estimator reads only the training field and the test-to-train affinities; it never forms an edge between two test states nor consults the held-out adjacency, so by construction inference on x_* is independent of all other test states—the definition of an inductive predictor. For the bound, convexity of the weighted average and the triangle inequality give

$$\|\widehat{F}(x_*)-F^\star(x_*)\|_2 = \Big\|\sum_i w_i\big(F^\star(x_i)-F^\star(x_*)\big)\Big\|_2 \le \sum_i w_i\|F^\star(x_i)-F^\star(x_*)\|_2 \le L\sum_i w_i\|x_i-x_*\|_2 \le L\max_i\|x_i-x_*\|_2. \tag{A4}$$

As the sample density increases the kNN radius $\max_i\|x_i-x_*\|_2\to 0$ almost surely under standard conditions, so the right-hand side vanishes. □

Theorem 9 (Distribution-free conformal coverage)**.** *Let the nonconformity score of a state with predicted probabilities p and true phase y be $s = 1-p_y$, computed by the fitted classifier. Given n_{cal} calibration states exchangeable with a test state, let $\hat{q}$ be the $\lceil(n_{\mathrm{cal}}+1)(1-\delta)\rceil/n_{\mathrm{cal}}$ empirical quantile of the calibration scores and form the prediction set $C(x) = \{c : 1-p_c(x) \le \hat{q}\}$. Then $\Pr[y_{\mathrm{test}}\in C(x_{\mathrm{test}})]\ge 1-\delta$.*

Proof. By exchangeability of the $n_{\mathrm{cal}}+1$ scores $\{s_1,\dots,s_{n_{\mathrm{cal}}},s_{\mathrm{test}}\}$, the rank of s_{test} among them is uniform on $\{1,\dots,n_{\mathrm{cal}}+1\}$, so $\Pr[s_{\mathrm{test}}\le\hat{q}]\ge\lceil(n_{\mathrm{cal}}+1)(1-\delta)\rceil/(n_{\mathrm{cal}}+1)\ge 1-\delta$, where $\hat{q}$ is the $\lceil(n_{\mathrm{cal}}+1)(1-\delta)\rceil$-th smallest calibration score. Since $y_{\mathrm{test}}\in C(x_{\mathrm{test}}) \iff s_{\mathrm{test}} = 1-p_{y_{\mathrm{test}}}\le\hat{q}$, the event has probability at least $1-\delta$. □

Appendix B: Numerical methods and reproducibility

This appendix collects the configuration, estimators, and evaluation protocols behind every number in the main text. All hyperparameters are fixed before evaluation and recorded, together with the platform and library versions, in the run artifact `results/summary.json`.

a. Synthetic benchmark. States are generated in an n_{features}-dimensional shadow-feature space along a phase diagram of three phases (a disordered parent phase and two ordered phases that bifurcate at the critical control parameter $s_0 = 0.45$). Phase centroids drift from fixed, well-separated anchors with the control parameter; each state is its phase centroid plus isotropic Gaussian shadow noise of standard deviation `noise`. The discrete phase equals the branch for the common phases; the rare critical regime is a distinct label assigned to a random subset of the late tip ($s > 0.8$) of the second ordered phase, whose feature vectors are additionally displaced into a separate pocket. Device, shot-noise-batch, and Hamiltonian-perturbation labels are drawn independently of phase so that grouping splits are nontrivial. The full-scale configuration is $n_{\mathrm{states}} = 3000$, $n_{\mathrm{features}} = 16$, $n_{\mathrm{phases}} = 3$, $n_{\mathrm{devices}} = 8$, $n_{\mathrm{batches}} = 4$, $n_{\mathrm{Hamiltonians}} = 5$, `noise` $= 0.7$, `rare_fraction` $= 0.04$, $k = 20$, `label_fraction` $= 0.03$, over $n_{\mathrm{seeds}} = 5$ seeds. Generation is fully seeded and bit-for-bit reproducible. The benchmark is a controlled stand-in for the shadow-feature geometry of a parameterized quantum system, not data from a quantum device.

b. Directed propagator and spectral truncation. A symmetric kNN graph is built on shadow features (either-direction symmetrization), as standard manifold-learning pipelines do [14]. The directed propagator reweights each edge $i\to j$ by $\kappa(s_j-s_i) = \exp(\beta\tanh((s_j-s_i)/\tau))$ and row-normalizes, with $\beta = 1.0$ and $\tau = 0.2$. The graph-Fourier basis U_r is the leading $r = 80$ eigenvectors of $S = D^{-1/2}AD^{-1/2}$, computed with a fixed starting vector for determinism. The truncated lazy transport is $B = U_r^\top((1-\varepsilon)I+\varepsilon P)U_r$ with $\varepsilon = 0.6$, and the band-limited field is $F^\star = U_r(I-\alpha B)^{-1}(1-\alpha)U_r^\top F_0$ with $\alpha = 0.9$ and F_0 the one-hot phase seed on annotated states (Theorem 6).

c. Classifiers. The transductive STDD classifier builds the graph and basis over all states; the inductive STDD classifier (the method) builds them on the training states only and labels held-out states by a Gaussian-weighted average of the training field over each test state's nearest *training* states, so held-out connectivity is never used (Theorem 8). The symmetric label-propagation baseline is the transductive normalized diffusion of Zhou *et al.* [12] ($\alpha = 0.9$, 30 iterations). The point baseline is a plain inductive kNN classifier with the same k. All read-outs row-normalize the non-negative part of the field to a probability simplex; the maximum phase probability is the confidence used for calibration.

d. Splits and annotation. Four leakage-checked protocols evaluate transfer: device and shot-noise batch (hold out whole groups), schedule (train at early control parameter, test at late—a forward extrapolation across the transition), and Hamiltonian (hold out a whole perturbation condition). For grouping splits the test suite asserts that no held-out group appears in both train and test. Only `label_fraction` of the train pool is anno-

tated, lightly stratified so that a few labels exist for each present phase, including the rare critical one.

e. Conformal calibration. On an exchangeable random hold-out, the inductive classifier is fit on the annotated train states and calibrated on an annotated calibration slice: a single temperature is fit by grid-minimizing calibration negative log-likelihood, and the split-conformal threshold is the $\lceil (n_{\mathrm{cal}}+1)(1-\delta)\rceil/n_{\mathrm{cal}}$ quantile of the calibration nonconformity scores $1-p_y$, with $\delta = 0.1$ (Theorem 9). Realized coverage, set size, and raw-versus-calibrated ECE are measured on the test states.

f. Persistence and active sampling. The degree-0 persistence barcode of the state cloud is computed exactly by union–find on the kNN distance graph (the finite H_0 bars are the minimum-spanning-tree edge weights), giving the $\beta_0(\epsilon)$ curve and barcode statistics. The persistence-guided active-sampling policy scores unlabeled states by prediction uncertainty (entropy) times per-state topological isolation (the k-th nearest-neighbor distance, the per-state witness of a long H_0 bar) and greedily takes the top `active_budget`$= 60$ states; baselines are the inverse-density policy of prior work and uniform random querying. Rare-state recall on the remaining unlabeled states is the lift.

g. Metrics, aggregation, and reproducibility. Accuracy is the fraction of held-out states classified correctly; ECE is the population-weighted gap between mean confidence and accuracy over ten bins; rare-state recall is the fraction of true rare critical states recovered (returning 1 when a test set contains no rare states). The non-normality witness is $\nu(P) = \|PP^\top - P^\top P\|_F/\|P\|_F^2$. All metrics are averaged over 5 seeds with 95 % confidence intervals $1.96\,s/\sqrt{n}$. All computation is CPU-only (NumPy, SciPy, scikit-learn, NetworkX, Matplotlib). The full-scale run is launched with `scripts/run.py` on `configs/full.yaml`; `results/summary.json` is the single source of truth for every table, figure, and number, and Tables I and II are generated verbatim from it by `scripts/make_tables.py`. The provenance block records seed, platform, library versions, runtime (32.0 s), and peak memory (621.2 MB).

Appendix C: From-scratch derivation of the implemented method

This appendix is written to be read top-to-bottom by someone who has seen linear algebra and basic probability but not graph-based semi-supervised learning or operator theory. Each part names the source file it maps to in the package `topocell`, so that theory and implementation can be read side by side.

a. The problem and the data (`src/topocell/synthetic.py`). A measurement campaign on a parameterized quantum system yields, for each prepared state, a feature vector of estimated observables (a classical shadow). States lie near a low-dimensional manifold—the phase diagram. The task is classification: predict the discrete phase of a state under a condition not directly observed—a new device, calibration batch, later control-parameter value, or perturbed Hamiltonian. Only a handful of states are annotated, and the decisive regimes are often rare. We do not ship real data (`ingest.py` holds honest hooks); instead we generate a fully controlled phase sweep with three phases over a control-parameter coordinate $s \in [0,1]$, a rare critical regime carved from the late tip of one ordered phase, and device/batch/Hamiltonian metadata assigned independently of phase so that the held-out splits are nontrivial.

b. From states to a directed operator (`graph.py`, `operators.py`). The kNN graph connects each state to its k nearest neighbors (either-direction symmetrization), giving a symmetric adjacency A. The degree is $d_i = \sum_j A_{ij}$ and $D = \mathrm{diag}(d_i)$. The standard smoother uses the symmetric normalized adjacency $S = D^{-1/2}AD^{-1/2}$, which is self-adjoint and therefore blind to direction. Our operator instead reweights each directed edge by a bounded forward kernel of the control-parameter increment, $W_{ij} = A_{ij}\exp(\beta\tanh((s_j - s_i)/\tau))$, and row-normalizes to a transition matrix $P = D_W^{-1}W$. Because forward and backward edges receive different weights, P is non-self-adjoint: $PP^\top \neq P^\top P$. The relative non-normality $\nu(P) = \|PP^\top - P^\top P\|_F/\|P\|_F^2$ is the scalar witness of this noncommutativity (Theorem 4).

c. Spectral truncation (`operators.py`). The graph-Fourier modes are the eigenvectors $u_1, \ldots, u_n$ of S; the r with the largest eigenvalues are the smoothest functions on the graph. Restricting the diffusion to their span $\Pi_r = U_rU_r^\top$ is a spectral truncation: it keeps the low-frequency, phase-relevant structure and discards high-frequency shot noise, and by Theorem 3 it gives the Eckart–Young-optimal rank-r approximation of the positive-semidefinite diffusion filter the method applies (and maximizes the retained smoothing energy $\mathrm{Tr}(\Pi_r S)$). The truncated, lazy directed transport is $B = U_r^\top((1-\varepsilon)I + \varepsilon P)U_r$, an $r \times r$ matrix. Truncation also makes the operator algebra noncommutative: the truncated multiplication operators $M_\phi^{(r)} = U_r^\top \mathrm{diag}(\phi)U_r$ do not commute (Theorem 4), the graph instance of the spectral-truncation noncommutativity of C^*-algebraic kernel machines.

d. The closed-form diffusion and its inductive extension (`operators.py`). We seed one-hot phase labels F_0 on the annotated states and solve the reduced normal equations $(I - \alpha B)c = (1-\alpha)U_r^\top F_0$ in closed form; the field is $F^\star = U_r c$. By Theorem 6 this is the unique fixed point of the band-limited directed diffusion and the limit of its power iteration, well posed for $\alpha\,\sigma_{\max}(B) < 1$, and invariant to the sign gauge of the eigenvectors. The transductive classifier builds A, U_r, P over all states; the inductive classifier builds them on the training states only and labels a held-out state by a Gaussian-weighted average of the training field

over its nearest *training* states—a Nyström/Nadaraya–Watson extension that never touches held-out connectivity (Theorem 8). This is the matched inductive control that resolves the transductive-versus-inductive confound of symmetric-smoothing studies.

e. Conformal calibration and persistent homology (`conformal.py`, `persistence.py`). Split conformal prediction holds out a calibration slice of the annotated states, computes the nonconformity scores $1 - p_y$, and thresholds at the conformal quantile to form prediction sets with finite-sample coverage $\geq 1 - \delta$ under exchangeability (Theorem 9); temperature scaling additionally recalibrates the scalar confidence. The degree-0 persistence barcode of the state cloud is computed exactly by union–find on the kNN distance graph—the finite H_0 bars are the minimum-spanning-tree edge weights—so $\beta_0(\epsilon) = \#\{\text{bars} > \epsilon\} + 1$ is a genuine multiscale invariant of the phase diagram, and the per-state k-th-nearest-neighbor distance is the per-state witness of a long bar, used by the active sampler.

f. Reproducibility object (`runner.py`, `scripts/`). `runner.py` ties everything together: it generates the phase sweep, builds the operator, records the non-normality and commutator diagnostics and the persistence barcode, runs the four classifiers under all four splits over five seeds, calibrates by conformal prediction, compares active-sampling policies, and writes `results/summary.json`—the single source of truth. Nothing in the paper is typed by hand: `make_tables.py` renders Tables I and II and `make_figures.py` renders Figs. 1–5 from that file. To reproduce everything from the `code/` folder:

```
export KMP_DUPLICATE_LIB_OK=TRUE
pip install .
python scripts/run.py \
    --config configs/full.yaml \
    --out results
python scripts/make_tables.py
python scripts/make_figures.py
```

The reported run uses 3000 states, 16 shadow features, 8 devices, 4 shot-noise batches, 5 Hamiltonian perturbations, ∼4 % rare critical regime, $k = 20$, truncation rank $r = 80$, $\alpha = 0.9$, $\beta = 1.0$, $\varepsilon = 0.6$, and 5 seeds; the provenance block records a runtime of 32.0 s and peak memory of 621.2 MB on a laptop CPU.

[1] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).

[2] A. Elben, S. T. Flammia, H.-Y. Huang *et al.*, The randomized measurement toolbox, *Nat. Rev. Phys.* **5**, 9 (2023).

[3] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, *Nat. Phys.* **13**, 431 (2017).

[4] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, *Nat. Phys.* **13**, 435 (2017).

[5] G. Carleo, I. Cirac, K. Cranmer *et al.*, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).

[6] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).

[7] V. Havlíček, A. D. Córcoles, K. Temme *et al.*, Supervised learning with quantum-enhanced feature spaces, *Nature* **567**, 209 (2019).

[8] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, *Phys. Rev. Lett.* **122**, 040504 (2019).

[9] Y. Hashimoto, A. Hafid, M. Ikeda, and H. Kadri, Spectral truncation kernels: Noncommutativity in C^*-algebraic kernel machines, *J. Mach. Learn. Res.* (2026); arXiv:2405.17823.

[10] W. D. van Suijlekom, Gromov–Hausdorff convergence of state spaces for spectral truncations, *J. Geom. Phys.* **162**, 104075 (2021).

[11] A. Connes and W. D. van Suijlekom, Spectral truncations in noncommutative geometry and operator systems, *Commun. Math. Phys.* **383**, 2021 (2021).

[12] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, Learning with local and global consistency, *Adv. Neural Inf. Process. Syst.* **16**, 321 (2004).

[13] X. Zhu and Z. Ghahramani, Learning from labeled and unlabeled data with label propagation, Tech. Rep. CMU-CALD-02-107, Carnegie Mellon University (2002).

[14] R. R. Coifman and S. Lafon, Diffusion maps, *Appl. Comput. Harmon. Anal.* **21**, 5 (2006).

[15] M. Belkin and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.* **15**, 1373 (2003).

[16] U. von Luxburg, A tutorial on spectral clustering, *Stat. Comput.* **17**, 395 (2007).

[17] L. van der Maaten and G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* **9**, 2579 (2008).

[18] F. Chung, Laplacians and the Cheeger inequality for directed graphs, *Ann. Comb.* **9**, 1 (2005).

[19] Y. Ashida, Z. Gong, and M. Ueda, Non-Hermitian physics, *Adv. Phys.* **69**, 249 (2020).

[20] S. Sachdev, *Quantum Phase Transitions*, 2nd ed. (Cambridge University Press, 2011).

[21] C. K. I. Williams and M. Seeger, Using the Nyström method to speed up kernel machines, *Adv. Neural Inf. Process. Syst.* **13**, 682 (2001).

[22] Y. Bengio, J.-F. Paiement, P. Vincent *et al.*, Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering, *Adv. Neural Inf. Process. Syst.* **16**, 177 (2004).

[23] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World* (Springer, 2005).

[24] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, Distribution-free predictive inference for regression, *J. Am. Stat. Assoc.* **113**, 1094 (2018).

[25] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, On calibration of modern neural networks, in *Proc. Int. Conf. Mach. Learn. (ICML)* (2017), p. 1321.

[26] H. Edelsbrunner and J. L. Harer, *Computational Topology: An Introduction* (American Mathematical Society, 2010).

[27] G. Carlsson, Topology and data, *Bull. Am. Math. Soc.* **46**, 255 (2009).

[28] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, in *Proc. Int. Conf. Learn. Represent. (ICLR)* (2017).

[29] B. Settles, Active learning literature survey, Comput. Sci. Tech. Rep. 1648, University of Wisconsin–Madison (2009).

[30] T. L. Patti, J. Kossaifi, A. Anandkumar, and S. F. Yelin, Variational quantum optimization with multibasis encodings, *Phys. Rev. Research* **4**, 033154 (2022).

[31] M. Cerezo, A. Arrasmith, R. Babbush *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).