METHOD STUDY

The Configuration Atlas of Quantum Kernels

Which settings of a quantum kernel classifier (QSVM) are measurable at all - and which are dead on arrival. Five feature maps, four repeat counts, five encoding bandwidths, four register sizes and three label families: 7,200 kernel-concentration measurements, 150 refereed validation runs, five pre-registered hypotheses with a public design deposit before the first run. Two settings the product offered turned out to encode nothing or to be classical in disguise; both are fixed. A method study behind the QML RQP.

September 2026

qubit-lab.ch

qubit-lab.ch
contact@qubit-lab.ch

Contents

The study on one page

A quantum kernel classifier does one thing: it compares data points through a quantum similarity score and lets a standard classical algorithm draw the decision boundary. The product lets you choose how that score is computed - five encodings, how often to repeat them, how strongly to scale the inputs - and that choice matters more than most users know. At larger register sizes many settings produce scores that cannot be measured with a realistic number of repeated measurements (shots); other settings produce scores that are perfectly measurable but carry no information about the data; one setting the product offered produced no information at all. This study measured every combination the product offers - 1,200 configurations across six replicate datasets, producing 7,200 kernel matrices - verified 150 configurations against a full classical referee (the best classical models we can mount, which any quantum claim must beat), and turned the result into guardrails in the product.

Everything was registered before the first run: the design was locked, publicly tagged and deposited with an independent registry (DOI 10.5281/zenodo.22498143). Five hypotheses were pre-registered; two of our own bets lost and are printed with the same prominence as the findings. Two product knobs the study needed were added before the lock, so the study measures the product's real configuration space.

Throughout, "advantage" means a statistically significant predictive advantage against the specified classical referee under the stated conditions; never a speed-up, a hardware result, or advantage over all classical algorithms. This is not a new theory of quantum kernels - the mechanisms are known [1-3], and systematic hyperparameter studies exist [4]. It is a measured operating envelope for configuring one, in a single chain: configuration → kernel diagnostic → refereed verdict → software behaviour.

The study produced five main findings:

1. **Above 8 qubits, most settings cannot be measured.** At 512 shots, 100% of configurations are measurable at 4 qubits, 86% at 8, 33% at 12 and 25% at 16; even 32,768 shots reach only 36% of the 16-qubit settings. More repeats and entangling encodings make it worse exactly as the theory predicts (hypothesis supported in 48 of 48 groups).
2. **Input scaling is the lever where it matters - but our registered bet lost.** We bet that the scaling knob (the bandwidth) would explain more of the outcome than the choice of encoding, at every size. It did not: the encoding factor dominates (0.57 to 0.62 of the variance versus 0.14 to 0.16), because one encoding is dead (finding 3) and swamps the comparison. Among live encodings - a disclosed post-hoc reading - bandwidth's share rises with size (0.25 → 0.46) and leads from 12 qubits.
3. **One offered setting encodes nothing.** The angle encoding with a Z rotation gives every row the same quantum state: the score is 1 for every pair, in all 1440 cells, and the diagnostic read it as "trivially measurable". Dead on arrival, and an instrument blind spot. The product now rejects the setting and flags constant scores.
4. **Measurable is not sufficient.** Of 150 refereed settings, 4 beat the referee at first sight - and all four use the simplest encoding, whose score we show has an exact classical formula (identical to 10^{-15}). With that formula in the referee, genuine wins are 0. Most other settings memorize the training data instead of learning: training accuracy near 0.99, new data near chance.
5. **Two more registered bets are printed as lost.** The diagnostic did not predict which settings lose their win under shot noise (only one setting flipped, so the test is inconclusive), and the winners did not sit near the shot-budget line (0 of 4).

Scope, once and meant everywhere: exact and shot-sampled simulation, no hardware; synthetic label families from the island studies - best-case envelopes, not real-world claims; one kernel family (fidelity); 4-16 qubits.

1. Management summary

Written for decision-makers; the measured basis follows in sections 2-7. Terms: a "setting" is one combination of encoding, repeats and input scaling; "measurable" means the similarity scores can be resolved with a realistic number of shots; the "referee" is the set of classical models a quantum result must beat.

Q1 - Which settings can we safely use, and which should we never run? Far fewer than the configuration screen suggests, and it depends on size. At 4 qubits every setting is measurable. At 12 qubits two thirds are not, at any realistic shot budget; at 16 qubits three quarters. The only 16-qubit settings measurable at 512 shots use the narrowest input scaling with one or two repeats. One offered setting - the angle encoding with a Z rotation - computed the same score for every pair of rows, encoding nothing; it is now rejected. The full map is in section 3 and ships in the product as configuration warnings.

Q2 - Does what the theory predicts actually show up in our product? Yes for the mechanisms; our own bet about which knob matters most lost. More repeats and entangling encodings make the scores collapse exactly as predicted, at every tested size and across all three label families. We had bet that the input-scaling knob would matter more than the choice of encoding at every size. On the pre-registered grid it did not - one dead encoding dominates the comparison. Among live encodings, scaling is the strongest lever at 12 and 16 qubits, which is where anything is decided; that reading is printed as post-hoc, not as a result.

Q3 - If the scores can be measured, does the classifier win? No. Of 150 tested settings, none genuinely beat the classical referee. Four looked like wins at first. All four used the simplest encoding, whose score has an exact classical formula; once that formula is in the referee, the wins vanish. Most other settings memorize the training data and fail on new data, even though their scores are perfectly measurable. Measurable is necessary, not sufficient. The one winning region known from the first island study is small, and this study's sample of 150 settings did not land on it - a limitation we print, not hide.

Q4 - What does the product do differently now? It stops users from running the settings the study found dead, and it refuses to call a classical formula a quantum win. Four changes shipped: the setting that encodes nothing is rejected with an explanation; a constant score matrix is flagged; the exact classical twin of the simplest encoding sits in the referee, so that encoding can at best reach Parity - the truth, since it uses no quantum resource; and the new scaling knob is bounded where it would wrap around. The guide now says: use an entangling encoding if you want to claim anything; keep scaling narrow at 12-16 qubits or nothing is measurable; and check both faces of the score matrix - can it be measured, and does it still tell rows apart - before spending shots.

The implications in four lines: **Do not run** settings that encode nothing or cannot be measured economically. **Do not claim quantum value** where an exact classical twin exists. **Screen before training** - screening rejects bad settings; it does not prove good ones. **Guardrails over freedom** - a configuration screen should expose evidence-based choices, not every mathematically legal circuit.

Q5 - What does it cost to find this out before training? About 17 seconds per setting, against about a minute for a refereed verdict - and the verdict needs a whole workflow that screening skips. The atlas: 7,200 screens in 35 processor-hours across four parallel shards. The verdicts: 150 refereed runs in 2.6 processor-hours, each requiring the quantum evaluation at three shot levels, ten classical models and three bootstrap inferences. Screening removes the dead settings before that workflow starts; it does not replace it.

2. What was measured

The study has two layers. The atlas layer measures the kernel-concentration diagnostic on every configuration the product offers, on every replicate dataset, without training anything; the validation layer takes a pre-registered sample of those configurations through the full refereed run. Both are described here, together with the registration that preceded them.

2.1 Factors and instrument

Factor	Levels
Feature map	angle-X, angle-Y, angle-Z, iqp, zz_like (amplitude and basis maps excluded: a different encoding paradigm; bandwidth undefined)
Feature-map repeats	1, 2, 3, 4
Encoding bandwidth	0.25, 0.5, 1, 2, 4 (a multiplier on the product's scaled angles - a knob added to the product for this study)
Register size	4, 8, 12, 16 qubits
Label family	quantum-native prototypes, 50% mixture, classical low-order (the island studies' generators, unchanged)
Replicates	6 datasets per family and size, seeds fixed before the run

1,200 configurations × 6 replicates = 7,200 kernel matrices, each on the 600-row training split. The instrument is the product's own kernel-concentration indicator, unmodified: bulk statistics of the off-diagonal fidelities, the per-row maximum fidelity as the tail that classification actually rides on, and **shots-to-resolve** $S = (3/v)^2$ with v the tail median - the number of measurement shots at which the typical nearest-neighbour fidelity stands three standard errors above noise. A configuration is resolvable at a budget when S is at or below it; the study considers 512, 4,096 and 32,768 shots.

2.2 The referee and the validation layer

150 configurations were drawn from the finished atlas by a pre-registered stratified sampler (strata: S* relative to the 4,096-shot line × label family × size; seed fixed and committed before the atlas finished) and run through the island studies' refereed cell: the QSVM trained at exact simulation and re-evaluated with shot-estimated kernels at 512 and 4,096 shots, scored against ten classical variants - logistic regression, random forest, gradient boosting, the order-3 periodic twin (two modes each), a tuned RBF kernel SVM and a neural network - each threshold-tuned on validation. A win requires the whole 95% confidence interval of the paired-bootstrap Quantum Advantage Factor to clear 1.0; a selection-aware panel-max bootstrap is reported alongside.

2.3 Related work, and what is new here

Kernel concentration and its sources are established theory [1]; bandwidth as the decisive hyperparameter is established [2, 3], and a systematic hyperparameter study of quantum kernels exists [4]. That bandwidth-tuned quantum kernels come to resemble classical radial-basis kernels [5], that near-identity kernels overfit [6], and that even entangling maps admit classical surrogates in some regimes [7] are all recent results. This study does not contest any of them. Its contribution is the chain: the product's real configuration screen, measured with its own diagnostic, checked against a full classical referee, and turned into software behaviour - including the discovery that its own simplest encoding had an exact classical twin the referee lacked.

2.4 Registration

Design, hypotheses, decision rules and the sampler were locked, tagged in the public repository (**kernel-config-atlas-design-v1**) and deposited with Zenodo (DOI 10.5281/zenodo.22498143) before the first atlas cell ran. This is the series' first study under the deposit-first commitment of the methodology paper. Appendix A lists every deviation and every analysis that was not pre-registered.

3. The atlas: where configurations are resolvable

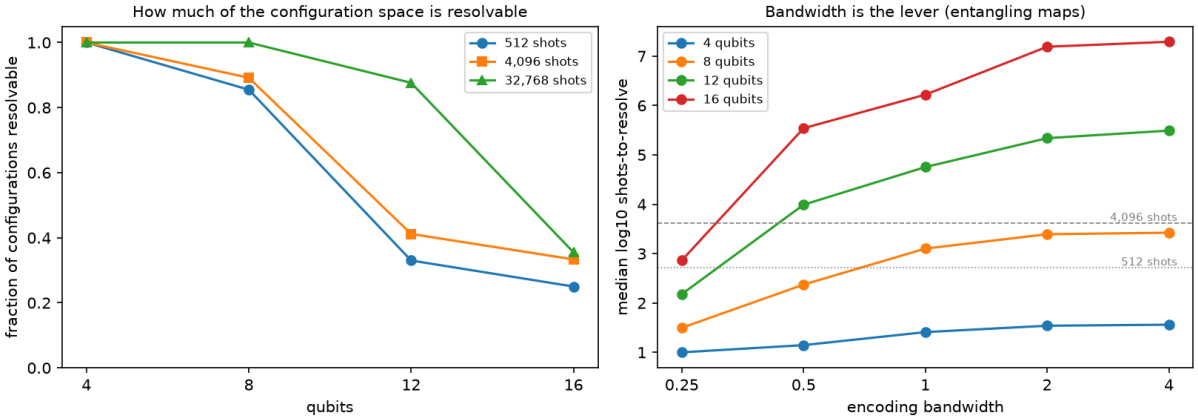

Figure 1 - Left: the fraction of the 1,200 configurations (× replicates) resolvable at three shot budgets, by register size. Right: for the entangling maps, median log10 shots-to-resolve against encoding bandwidth; dashed lines mark the 512- and 4,096-shot budgets.

The regime is size-gated and steep. At 4 qubits every configuration is resolvable at 512 shots; at 8 qubits 86%; at 12 qubits 33% (41% at 4,096 shots); at 16 qubits 25%, and raising the budget 64-fold to 32,768 shots only lifts this to 36%. 144 cells - all iqp at 16 qubits - are unresolvable at any budget the study considers.

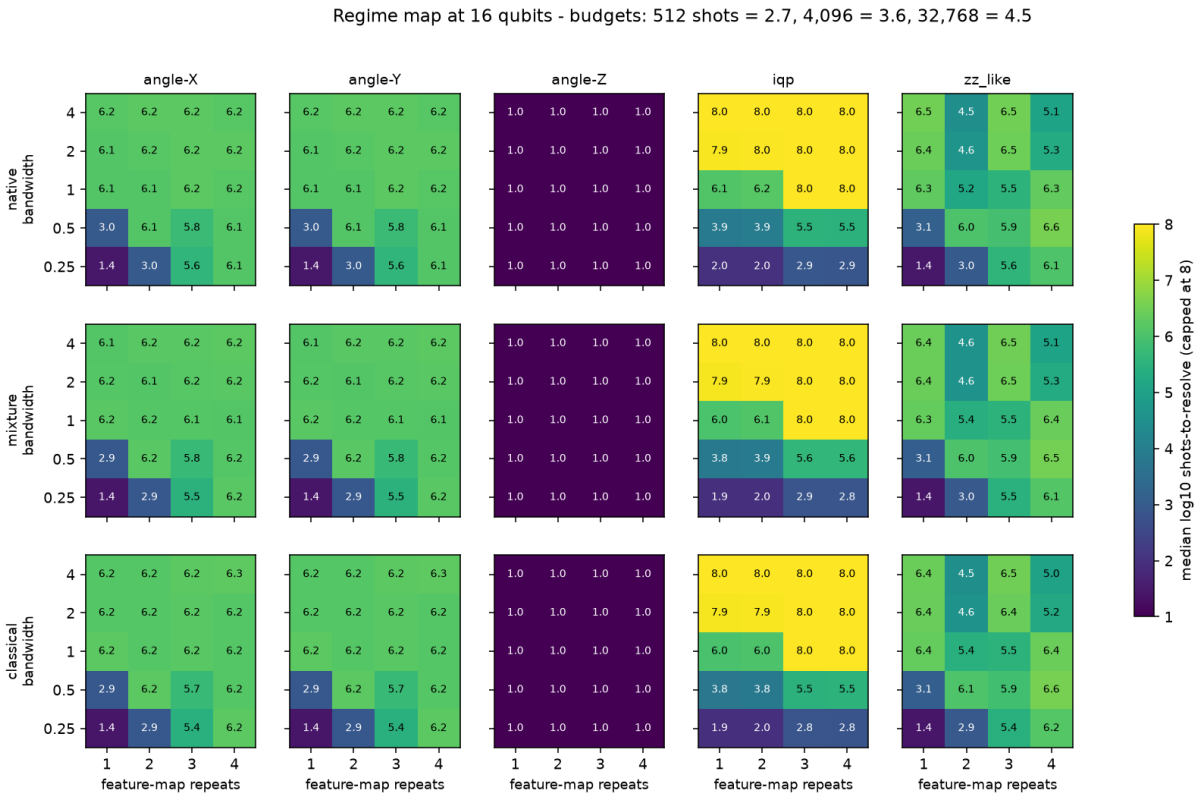

Figure 2 - The 16-qubit regime map: median log10 shots-to-resolve per configuration cell (bandwidth × repeats), one panel per feature map (columns) and label family (rows). Budgets: 512 shots = 2.7, 4,096 = 3.6, 32,768 = 4.5. Angle-Z is constant by degeneracy (section 4); the non-monotone pattern in zz_like at bandwidths 2-4 is angle aliasing.

S5-H1 - repeats and entanglement concentrate (supported). Within every family × size × map group with a defined statistic, shots-to-resolve rises with repeats (Spearman with cluster-bootstrap CI above zero in 48 of 48 groups; the 12 undefined groups are all angle-Z, where the statistic is constant - section 4). Entangling maps concentrate faster than product maps: +0.061 decades of S* per repeat, 95% CI [0.06, 0.062].

S5-H2 - the registered bet that bandwidth beats structure (refuted as registered). In a variance decomposition of log10 S over the balanced grid, the feature-map factor explains 0.57, 0.63, 0.62 and 0.62 of the variance at 4, 8, 12 and 16 qubits; bandwidth 0.16, 0.15, 0.15, 0.14; repeats 0.01 to 0.03. Bandwidth beats repeats everywhere (CI above zero at

every size) but not the map factor; the bet is lost. The cause is the degenerate angle-Z map, which sits five decades away from every live encoding at 16 qubits. Excluding it - a post-hoc sensitivity analysis, not a registered test - bandwidth's share is 0.25, 0.30, 0.40, 0.46 and the map factor's 0.48, 0.41, 0.19, 0.02: bandwidth overtakes from 12 qubits (CI of the difference [0.204, 0.212] at 12, [0.435, 0.444] at 16). The honest reading: which map you pick decides most at small registers; at the sizes where concentration decides anything, the bandwidth knob is the one that moves the boundary. For the entangling maps, median log10 S at 16 qubits runs 2.86 → 5.54 → 6.22 → 7.19 → 7.29 across bandwidths 0.25 to 4; only the narrowest setting crosses under the 512-shot line.

Two structural facts the atlas made visible, both disclosed rather than modelled: for product-state (angle) maps, repeats and bandwidth are exactly degenerate - the encoding angle is repeats × bandwidth × the scaled feature, so two repeats at bandwidth 0.5 is one repeat at bandwidth 1 - while entangling maps interleave entanglers and are not; and bandwidths above about 2 wrap the encoding angles beyond 2π, which is why the entangling maps show non-monotone repeat patterns at wide bandwidths. The product knob is bounded accordingly.

S5-H5 - the rotation axis is irrelevant (null-shaped; exact for X vs Y, refuted for Z). Angle-X and angle-Y produce identical kernels in every cell (mean difference 0.0) - the fidelity of product states does not depend on the rotation axis. Angle-Z differs by 2.1 decades, for the reason in section 4.

4. Dead on arrival: the degenerate encoding

A Z rotation applied to the $|0\ldots0\rangle$ start state changes only a global phase. Every row is therefore encoded to the same quantum state, and the kernel is identically 1: in all 1440 angle-Z cells the off-diagonal fidelity is exactly 1.0000 with zero spread. The setting was selectable in the product.

The instrument did not catch it. Shots-to-resolve keys on the magnitude of the typical nearest-neighbour fidelity, and a fidelity of 1 reads as "resolvable in 9 shots" - the blind spot is concentration toward one, the opposite direction from the concentration the theory warns about. In the validation layer the same configurations score at chance, as they must.

Both are fixed in the product: workbook validation rejects the angle map with rotation Z at inspection and at every run route, with an explanation; the kernel-concentration indicator carries an additive flag for degenerate constant kernels and the report prints a warning row. The study's registered instrument is unchanged; the flag is additive so that every number here reproduces from the committed raw results.

5. Resolvable is not sufficient: the validation layer

Of 150 refereed configurations, 4 met the registered win rule at exact simulation and 3 at 512 shots. The four winners share a property: all use the angle map, and all sit far below the 4,096-shot line (S* around 100).

S5-H3 - shots-to-resolve predicts the shot-noise verdict flip (inconclusive). Only 1 configuration lost its win between exact simulation and 512 shots, so there is nothing to predict: the registered AUC (0.45, CI [0.369, 0.527]) is computed on one positive case and carries no information. The decision rule returns "not supported"; the honest label is inconclusive.

S5-H4 - winners sit near the boundary (refuted). 0 of 4 winners lie within one decade below the 4,096-shot line (Wilson 95% upper bound 0.49). The winners were not near any boundary; they were in the most resolvable corner of the atlas.

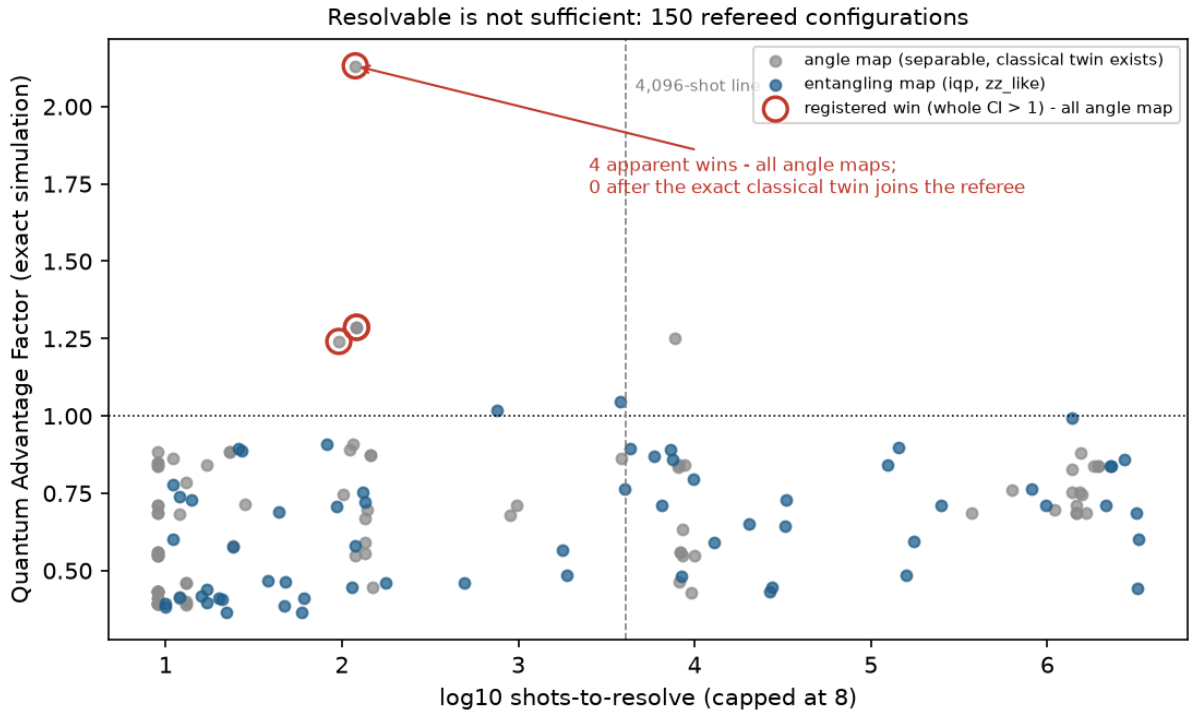

Figure 3 - The 150 refereed configurations: shots-to-resolve against the Quantum Advantage Factor at exact simulation. Grey: angle (product-state) maps; blue: entangling maps. Circled: the four configurations that met the registered win rule - all angle maps, and zero after their exact classical twin joins the referee.

5.1 Why almost nothing wins: memorization

The losing configurations are not concentrated in the theory's sense - their tails resolve - but their bulk is near zero: off-diagonal fidelities of 0.004 to 0.01 make the kernel nearly the identity matrix. A kernel SVM on a near-identity kernel is a nearest-neighbour memorizer: training accuracy 0.98-0.99 in the cells we dissected, test performance near chance. The instrument's tail statistic says "resolvable"; the bulk says "carries no generalization". Both must be read, and the product now prints both.

5.2 Why the four wins are not quantum (post-hoc, disclosed)

The angle map applies one single-qubit rotation per feature, so its state is a product state and its fidelity kernel has a closed classical form - the product over qubits of $\cos^2$ of half the angle difference, on the product's own scaled angles. We verified this identity against the product's statevector kernel to a maximum difference of 7e-15 on every winner and across sizes, repeats and bandwidths. A classical kernel SVM on the closed form reproduces the angle-map QSVM exactly: on the four winners its macro-F1 equals the quantum model's to four decimals (0.805, 0.620, 0.779, 0.779). The registered referee did not contain this twin; with it, wins among the 150 configurations are 0. The product's referee now contains it whenever a QSVM run uses the angle map - an angle-map QSVM can at best reach Parity, which is what a product-state kernel deserves.

5.3 Two disclosures about the procedure

Threshold tuning. The registered scoring rule tunes a positive-class F1 threshold on the validation split (the product's rule; the island studies used it too). For weak models it collapses predictions to a single class - 101 of 150 quantum models here score at the one-class floor of macro-F1 0.33. Re-scoring the four winners and three losers with plain 0.5 thresholds on both sides changes no verdict direction (winners' plain ratios 1.564, 1.205, 1.2, 1.2; losers' 0.714, 0.786, 0.812); it does change absolute scores, and a weak quantum model is printed harsher than it is. A guard in the product's threshold optimizer is the open improvement.

The sampler missed the known winner. The stratified sampler drew one replicate per configuration cell by S* bands. The configuration the first island study showed to win in ≥94% of arms - zz_like at two repeats and bandwidth about 1 on

native labels - was not among the 150 drawn. This study therefore cannot re-confirm that region; it shows that the region is small and that its neighbours in configuration space do not share it. Study 3's map stands; this study bounds it.

6. What the product ships now

- **Angle-Z rejected** at inspection and at every run route, with the reason printed.
- **Degenerate-kernel flag** in the concentration indicator and a warning row in the report.
- **Separable-kernel twin** in the classical referee for angle-map QSVM runs: an exact classical reproduction, so the verdict is Parity at best.
- **Bandwidth knob** (**input_scaling_bandwidth**, 0.05-4.0) with the guide stating that values above about 2 alias the encoding.
- **Guidance printed in the user guide:** advantage claims for kernel methods require an entangling map; at 12-16 qubits only narrow bandwidths with one or two repeats stay resolvable at realistic shot budgets; read the bulk mean as well as the tail before spending shots - a resolvable kernel that is nearly the identity will memorize.

7. What it cost

The atlas layer: 7,200 kernel matrices in 34.8 processor-hours across four parallel shards (about 17 seconds per configuration on average; 16-qubit IQP cells dominate at one to seven minutes each). The validation layer: 150 refereed configurations in 2.6 processor-hours (median 11 seconds per arm at 4-12 qubits, 100-700 seconds at 16). In processor time the two layers differ by a factor of about 3.6 per configuration (17 against 63 seconds), not by orders of magnitude - the 16-qubit atlas cells are expensive in their own right. What differs by orders of magnitude is the workflow a verdict needs: quantum evaluation at three shot levels, ten classical models, three bootstrap inferences, and a human reading the result. Screening removes the dead configurations before any of that starts. With the caveat of section 5, which is the point of the study: screening tells you what is dead, not what is alive.

8. Limitations

- **Synthetic label families**, inherited from the island studies: best-case envelopes for the native family, not real-world data. The atlas maps configurations, not datasets.
- **One kernel family** (fidelity); projected kernels out of scope. **Exact and shot-sampled simulation only**; no hardware, no noise model.
- **Replicate variance is tiny** at these sizes, so the cluster-bootstrap intervals are narrow: they reflect the sampling of datasets, not model uncertainty about the configuration effect.
- **The validation sampler** stratified by resolvability and drew one replicate per cell; it did not land on the known winning configuration (section 5.3).
- **Two registered bets lost and one test is inconclusive.** They are printed with the same prominence as the supported findings.

Appendix A - Registration, deviations and post-hoc analyses

A.1 Registration. Design tag **kernel-config-atlas-design-v1** (commit ecd0284) pushed to the public repository and Zenodo deposit DOI 10.5281/zenodo.22498143 published on 2026-09-06, both before the first atlas run (started the same day after both existed). The deposited package is byte-identical to the tagged files (SHA-256 manifest inside).

A.2 Implementation choices not pinned by the design, fixed before the run and disclosed: order-3 low-order labels for the classical family; one stratified split per dataset (seed 42); the full 600-row training split as the gram sample; the product's effective-rank statistic used as the secondary instrument.

A.3 Deviations. None from the factor grid, the instrument, the referee, the sampler or the decision rules. The validation layer finished with two processes working the arm list from both ends; 15 arms were computed twice and verified identical before deduplication.

A.4 Post-hoc analyses (not pre-registered, labelled as such where used): the H2 sensitivity excluding the degenerate angle-Z map (section 3); the separable-twin identity check and re-scoring (section 5.2); the threshold-rule robustness re-scoring on seven arms (section 5.3); the memorization dissection (section 5.1).

A.5 Product changes made during the study, each after the corresponding measurement and without touching the study's code path: angle-Z rejection, degenerate-kernel flag, separable-kernel twin in the referee, bandwidth guidance. All are additive or validation-level, so every number here reproduces from the committed raw results with the committed scripts.

References

1. S. Thanasilp, S. Wang, M. Cerezo, Z. Holmes, Exponential concentration in quantum kernel methods, Nature Communications 15 (2024); arXiv:2208.11060 - the concentration mechanism and its named sources (expressivity, entanglement, global measurement, noise).
2. R. Shaydulin, S. M. Wild, Importance of Kernel Bandwidth in Quantum Machine Learning, Phys. Rev. A 106, 042407 (2022); arXiv:2111.05451 - bandwidth as the lever on kernel concentration.
3. A. Canatar, E. Peters, C. Pehlevan, S. M. Wild, R. Shaydulin, Bandwidth Enables Generalization in Quantum Kernel Models, arXiv:2206.06686 - the generalization side of the same lever.
4. S. Egginger, A. Sakhnenko, J. M. Lorenz, A hyperparameter study for quantum kernel methods, Quantum Machine Intelligence (2024); arXiv:2310.11891 - prior systematic hyperparameter work on quantum kernels.
5. R. Flórez-Ablan, M. Roth, J. Schnabel, On the similarity of bandwidth-tuned quantum kernels and classical kernels, arXiv:2503.05602 - bandwidth-tuned quantum kernels converge toward classical radial-basis kernels.
6. J. Tomasi, S. Anthoine, H. Kadri, Benign Overfitting with Quantum Kernels, UAI 2026; arXiv:2503.17020 - near-identity fidelity kernels and overfitting.
7. E. Tekeli, The ZZ feature map induces a signless Laplacian metric: a closed-form classical surrogate for quantum kernel regression, arXiv:2608.29422 - a classical surrogate for an entangling map in the small-bandwidth regime.
8. qubit-lab.ch, The Advantage Island of Quantum Kernels (August 2026) - the island map, the label families, the referee and the concentration instrument this study reuses.
9. qubit-lab.ch, The Advantage Island of Variational Classifiers (August 2026) - the companion map; the registration chronology this study extends.
10. qubit-lab.ch, The RQP Evidence Engine (August 2026) - the deposit-first commitment kept here.

Disclaimer

This document reports a method study on synthetic benchmark data with a rapid quantum prototyping tool by qubit-lab.ch. Its figures are unedited outputs of the described runs, intended for technical review, education and structured discussion. The configurations found resolvable or unresolvable are properties of the tested feature maps on the tested label families; they are not financial advice, not a performance claim for real-world datasets, and not a hardware result. qubit-lab.ch is not affiliated with the authors of the cited references.