Methods of quantum machine learning in detecting the causality of stochastic processes

Based on Szymon Szurnicki *BSc thesis* under supervision of Adam Bednorz

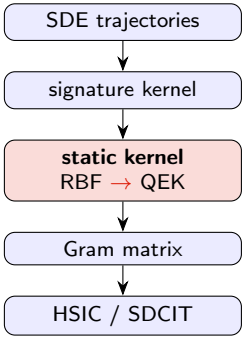
Szymon Szurnicki[1], Piotr Gawron[2], Adam Bednorz[1]

[1]Department of Physics, University of Warsaw
[2]Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences

3rd International Workshop on Machine Learning
and Quantum Computing Applications in Medicine and Physics
7–11 September 2026, Warsaw, Poland

Outline[1]

1. Motivation and otlook
2. Computational pipeline: from time series to test verdict
3. Signature kernel for time series
4. Quantum embedding kernel (QEK)
5. Circuit optimization and independence tests
6. Experiments E1–E4, including measurements on IBM Heron



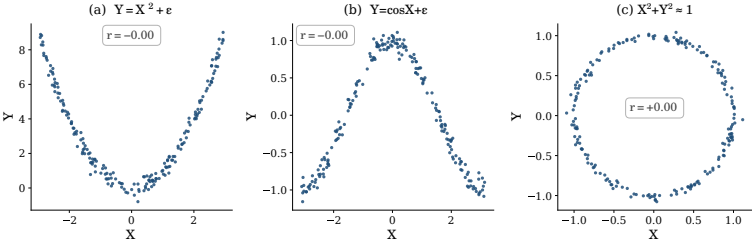
[1]Slides by Szymon Szurnicki (adapted)

Limitations of classical dependence measures

Pearson correlation coefficient

$$r_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \, \sigma_Y}$$

- captures only linear dependence;
- $r \approx 0$ does not imply independence;
- for time series, the order of events matters, not accounted for by pointwise measures.



Configurations with strong dependence, for which $r \approx 0$.

nature reviews earth & environment

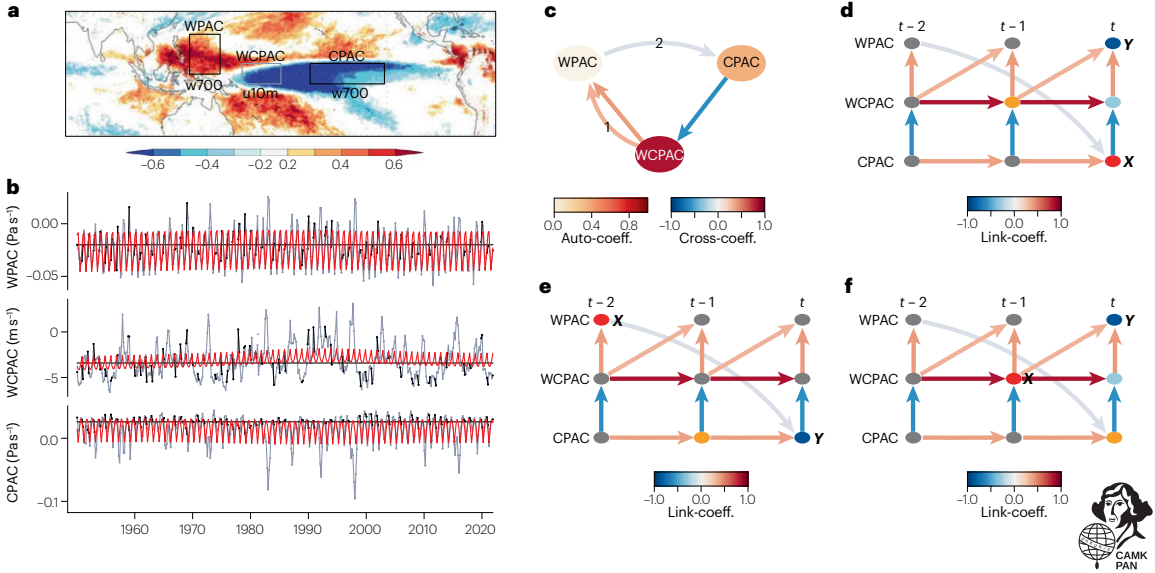
https://doi.org/10.1038/s43017-023-00431-y

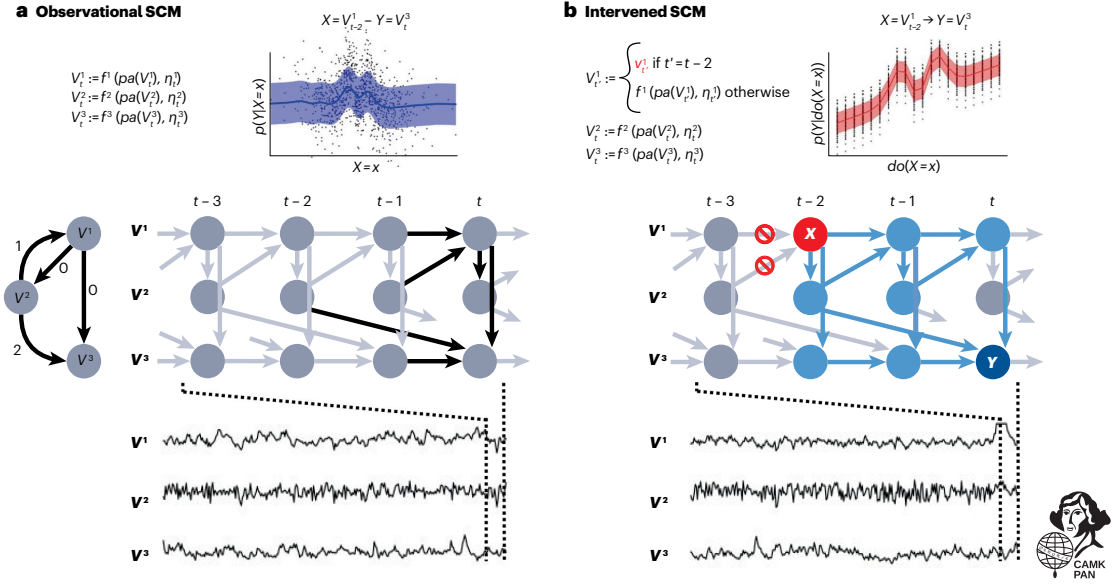
Technical review

Check for updates

Causal inference for time series

Jakob Runge [1,2] ✉, Andreas Gerhardus[1], Gherardo Varando [3], Veronika Eyring[4,5] & Gustau Camps-Valls[3]


a
WPAC
WCPAC
CPAC
w700
u10m
w700
−0.6 −0.4 −0.2 0.2 0.4 0.6
b
WPAC (Pa s⁻¹)
0.00
−0.05
WCPAC (m s⁻¹)
0
−5
CPAC (Pa s⁻¹)
0.0
−0.1
1960 1970 1980 1990 2000 2010 2020
c
WPAC
CPAC
WCPAC
2
1
0.0 0.4 0.8
Auto-coeff.
−1.0 0.0 1.0
Cross-coeff.
d
t − 2
t − 1
t
WPAC
WCPAC
CPAC
Y
X
−1.0 0.0 1.0
Link-coeff.
e
t − 2
t − 1
t
WPAC
X
WCPAC
CPAC
Y
−1.0 0.0 1.0
Link-coeff.
f
t − 2
t − 1
t
WPAC
Y
WCPAC
X
CPAC
−1.0 0.0 1.0
Link-coeff.
CAMK
PAN


a Observational SCM
$V_t^1 := f^1(pa(V_t^1), \eta_t^1)$
$V_t^2 := f^2(pa(V_t^2), \eta_t^2)$
$V_t^3 := f^3(pa(V_t^3), \eta_t^3)$
$X = V_{t-2}^1 \to Y = V_t^3$
$p(Y|X = x)$
$X = x$
b Intervened SCM
$V_{t'}^1$ if $t' = t - 2$
$f^1(pa(V_{t'}^1), \eta_{t'}^1)$ otherwise
$V_t^2 := f^2(pa(V_t^2), \eta_t^2)$
$V_t^3 := f^3(pa(V_t^3), \eta_t^3)$
$X = V_{t-2}^1 \to Y = V_t^3$
$p(Y|do(X = x))$
$do(X = x)$
V^1
V^2
V^3
1
0
0
2
$t-3$
$t-2$
$t-1$
t
X
Y

Aim of the work

Reference work — Manten and Casolo (2025)

Combining the signature kernel (RBF) with statistical (HSIC, SDCIT) tests allows recovering causal structures in complex time-dependant systems.

Scope of the work

1. Replacement of the RBF static kernel with a quantum embedding kernel (QEK) implemented by a variational circuit.
2. Circuit optimization (SPSA with Adam optimizer) to maximizing test power.
3. Influence of the number of qubits ($W = 4, 6, 8$) on test power.
4. Execution of computations on an IBM Heron quantum processor.

Research question: is there a quantum advantage in the considered application?

Synthetic data generation

Coupled linear SDEs

A system with controlled causal structure:

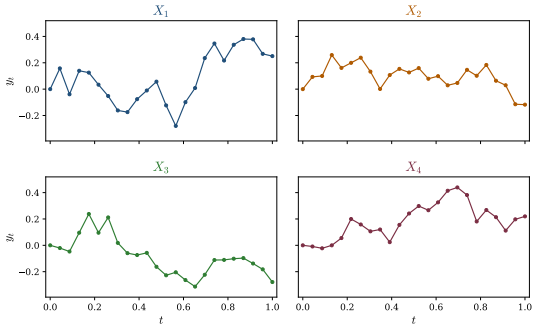
$$\mathrm{d}X_t = AX_t\mathrm{d}t + \mathrm{d}W_t, \qquad X_t \in \mathbb{R}^p$$

Two-variable case:

$$\mathrm{d}X = -0.5\,X\,\mathrm{d}t + \mathrm{d}W_1$$
$$\mathrm{d}Y = (a_{21}X - 0.5\,Y)\,\mathrm{d}t + \mathrm{d}W_2$$

- $a_{21} \neq 0 \;\Rightarrow\; X \to Y$; causal structure known a priori;
- $a_{ii} = -0.5$: self-drift (Ornstein–Uhlenbeck process);
- n trajectories observed at m time points on $[0, 1]$.

$$\mathrm{d}X_t = -\alpha X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$$



Independent realizations of a single process.

Kernel methods

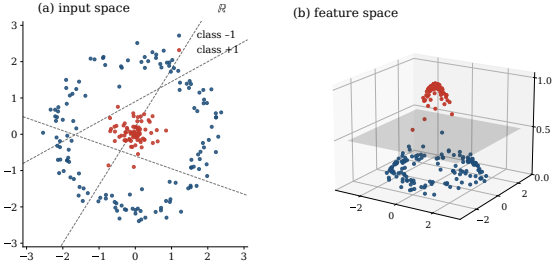
- Linear classifier for complex data.
- *Kernel trick*: computing the inner product without explicitly performing the embedding Φ.

$$k(x_i, x_j) = \langle \Phi(x_i) | \Phi(x_j) \rangle$$

- Gaussian kernel — a common example

$$k_{\mathrm{RBF}}(x_i, x_j) = \exp\left(- \frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$$

- data transformation to infinite-dimensional feature space.



Linear classification after mapping into higher-dimensional space.

Signature kernel

A sample is not a point, but an entire trajectory. A representation that accounts for the order of events is the **path signature**

$$S(X) = \left(1, X^{(1)}, X^{(2)}, \dots\right)$$

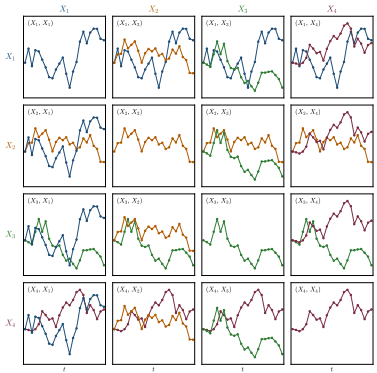
computed via the kernel trick as the solution of the equation

$$\kappa_{\mathrm{sig}}(X^i, X^j) = 1 + \int_0^T \int_0^T f(t, t')\, \kappa_{\mathrm{stat}}(k_t, l_{t'})\, \mathrm{d}t\, \mathrm{d}t',$$

$f(0, \cdot) = f(\cdot, 0) = 1$; solver: `sigkerax`.

Subject of modification

In the reference work κ_{stat} is RBF — we replace it with a quantum kernel.



Static kernel computed *between points* of two trajectories, inside the solver.

Gram matrix as a representation of the process

$$K_{XX} = \begin{bmatrix} \kappa_{\mathrm{sig}}(X^1, X^1) & \cdots & \kappa_{\mathrm{sig}}(X^1, X^n) \\ \vdots & \ddots & \vdots \\ \kappa_{\mathrm{sig}}(X^n, X^1) & \cdots & \kappa_{\mathrm{sig}}(X^n, X^n) \end{bmatrix}$$

- **symmetry**: $K = K^\top$;
- **positive semi-definiteness**: $c^\top K c = \left\| \sum_i c_i \Phi(x_i) \right\|^2 \geq 0$;
- **Mercer's theorem** — admissibility criterion for replacing the kernel.

The QEK kernel satisfies both conditions by construction, as the squared modulus of the inner product of quantum states.

Wiener process, $n = 4$, $m = 10$:

$$K^{\mathrm{RBF}} = \begin{bmatrix} 1.00 & 0.65 & 0.70 & 0.49 \\ 0.65 & 1.00 & 0.97 & 0.81 \\ 0.70 & 0.97 & 1.00 & 0.78 \\ 0.49 & 0.81 & 0.78 & 1.00 \end{bmatrix}$$

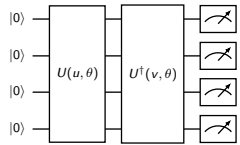
$$K^{\mathrm{QEK}} = \begin{bmatrix} 1.00 & 0.48 & 0.56 & 0.41 \\ 0.48 & 1.00 & 0.92 & 0.65 \\ 0.56 & 0.92 & 1.00 & 0.62 \\ 0.41 & 0.65 & 0.62 & 1.00 \end{bmatrix}$$

Both matrices indicate the same pair as most similar and the same pair as least similar; they differ in the scale of values.

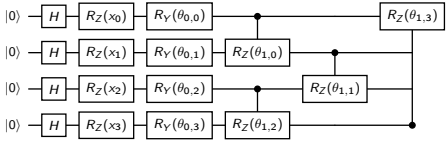
Quantum embedding kernel (QEK)

$$k_q(u,v) = \left| \langle \phi(v) | \phi(u) \rangle \right|^2$$
$$= \left| \langle 0 \cdots 0 | \, U^\dagger(v,\theta)\, U(u,\theta) \, |0 \cdots 0\rangle \right|^2$$

- u, v — data to be embedded;
- θ — free parameters of the circuit;
- kernel value = probability of measuring $|0 \cdots 0\rangle$;
- feature space dimension grows exponentially with the number of qubits;
- for $u = v$, circuits cancel and $k_q = 1$: the diagonal of the Gram matrix is unity by construction.



Single embedding layer $U(x,\theta)$, $W = 4$:



Hadamard gate, data encoding $R_Z(x_j)$, rotation $R_Y(\theta_{0,j})$ and an entangling layer $\mathrm{CRZ}(\theta_{1,k})$. In the thesis $L = 2$ layers.

Optimization of circuit parameters

Cost function: standardized HSIC statistic

$$L(\theta) = -\,\mathrm{SNR}(\theta) = -\,\frac{\widehat{\mathrm{HSIC}}_{\mathrm{obs}}(\theta) - \mu_{\mathrm{null}}(\theta)}{\sigma_{\mathrm{null}}(\theta)}$$

- scale-free — invariant under rescaling of the Gram matrix;
- rewards only the deviation of the statistic from its permutation distribution;
- fixed data portions, split into training and validation sets.

SPSA: gradient-free optimization

$$\widehat{g}(\theta) = \frac{L(\theta + c\Delta) - L(\theta - c\Delta)}{2c}\,\Delta, \quad \Delta \in \{-1, 1$$

- gradient estimate from two cost function evaluations;
- parameters θ appear *inside* the solver, after normalization and permutation statistics, which makes differentiation difficult;
- estimate passed to the `Adam` optimizer.

Optimization does not build a predictive model — it modifies the geometry of the similarity relation induced by the kernel.

Independence tests: HSIC and SDCIT

HSIC — correlation of two similarity matrices:

$$\widehat{\mathrm{HSIC}} = \frac{1}{n^2}\,\mathrm{tr}\left(\widetilde{K}_{XX}\widetilde{K}_{YY}\right), \quad \widetilde{K} = HKH$$

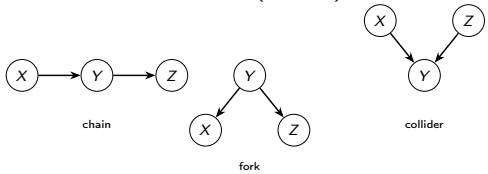
P-value computed via permutations (500 permutations); the null distribution is constructed from the data, solely from the Gram matrices.

Test power — the primary metric of the thesis:

$$\text{power} = \Pr\left(p < \alpha \mid \text{true dependence}\right)$$

Reported together with the rejection rate on independent data, which serves as a control for the Type I error rate.

SDCIT — hypothesis $X \perp\!\!\!\perp Z \mid Y$. Permuting z only between rows with similar y; the test statistic is $\widehat{\mathrm{MMD}}^2(D, D^\pi)$.



Chain and fork have identical marginal dependence patterns; only a conditional test distinguishes them.

E1–E2: verification of implementation correctness

E1: two processes, n=20, m=20; median over 10 repetitions.

	a_{21}=1.5	a_{21}=0
RBF	0.008 (9/10)	0.386 (1/10)
QEK W=4	0.015 (8/10)	0.515 (1/10)
expected verdict	dependent	independent

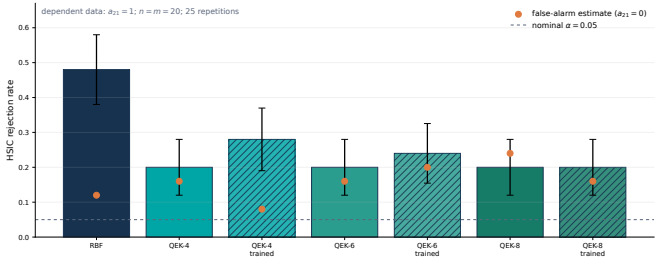
Both kernels lead to the correct verdict in both variants.

E2: three graphs, n=200, m=32, a=2.0; RBF kernel.

graph	XY	YZ	XZ	$X \perp\!\!\!\perp Z \mid Y$
chain	0.000	0.000	0.000	0.565
fork	0.000	0.000	0.000	0.276
collider	0.000	0.000	0.277	0.006

The Gram matrices also carry conditional information: SDCIT corrects the marginal picture in the chain and fork, and reveals the dependence induced by conditioning in the collider.

E3: test power as a function of kernel variant



a_{21}=1, n=20, m=20, 25 repetitions; red points — rejection rate on independent data (α=0.05).

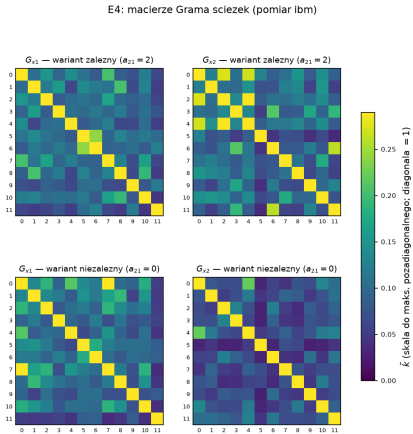
variant	power	rejections	mean off-diag.
RBF	0.48	0.12	—
$q4$	0.20	0.16	0.33
$q4^{\mathrm{tr}}$	0.28	0.08	0.47
$q6$	0.20	0.16	0.35
$q6^{\mathrm{tr}}$	0.24	0.20	0.33
$q8$	0.20	0.24	0.26
$q8^{\mathrm{tr}}$	0.20	0.16	0.28

- classical kernel significantly more powerful;
- optimization brings improvement, largest for the smallest register ($q4$: 0.20 $\rightarrow$ 0.28 with a decrease in the false rejection rate);
- increasing the number of qubits amplifies concentration of the Gram matrix and worsens trainability.

E4: measurements on the IBM Heron processor

	sym. (shot noise)	sym. (noise model)	IBM hardware
correlation with exact	0.992	0.986	0.967
mean $\lvert\Delta k\rvert$	0.030	0.069	0.095
p, dependent variant	0.004	0.008	0.013
p, independent variant	0.65	0.29	0.12

- `ibm_marrakesh`, 156 qubits, 64 shots per pair; over 18 thousand circuits within a 10-minute time budget;
- $\mathrm{CRZ}(\theta) = R_{ZZ}(-\theta/2)\,(I \otimes R_Z(\theta/2))$: R_Z is a virtual gate, R_{ZZ} is native (fractional gates), giving W two-qubit operations per layer without decomposition into CX;
- both verdicts correct despite system noise.



Gram matrices reconstructed from raw hardware data: dependent variant (left) and independent variant (right).

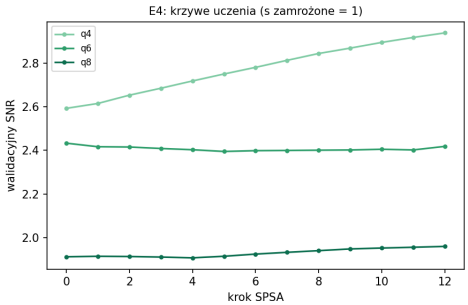
Conclusions

1. A complete pipeline for testing independence of stochastic processes was built with an exchangeable static kernel (RBF $\leftrightarrow$ QEK); correctness was verified by reproducing the reference result and a full set of correct marginal and conditional verdicts on the three canonical graphs.
2. On the data studied, the classical kernel remains more powerful (0.48 versus 0.20–0.28 at n=20); in the regime studied, no quantum advantage was found, consistent with literature predictions for fidelity kernels.
3. Optimization of circuit parameters by maximizing the standardized HSIC statistic (SPSA with Adam optimizer) brings measurable improvement for small registers and counteracts Gram matrix concentration.
4. The influence of register width qualitatively confirms predictions regarding trainability: as the number of qubits increases, the initial signal-to-noise ratio decreases (2.59 $\rightarrow$ 1.91) and the rate of improvement slows ($+0.35 \rightarrow +0.05$ over 12 steps).
5. The kernel computation was executed on the IBM Heron processor; high agreement with simulations was obtained, and the independence test on the fully hardware-computed Gram matrix yielded correct verdicts.

Directions for future research

The absence of advantage in the studied regime does not exhaust the issue: the QEK kernel was chosen for its simplicity and represents one of many variants being developed. Directions for future work include:

- systematic scanning of the coupling strength a and sample size n;
- increasing the number of layers L and extending the training procedure;
- alternative quantum kernel constructions — the result depends primarily on the quality of the embedding, whose optimality quantumness of the model alone does not guarantee;
- kernels with *local* instead of global cost, alleviating the concentration problem.



SPSA optimization progress; slowing of improvement for wider registers is visible.

Thank you for your attention

Piotr Gawron

@ gawron@camk.edu.pl
@ piotr@piotrgawron.eu
http://piotrgawron.eu/

References I

[1] S. Szurnicki. "Quantum Machine Learning Methods in Causality Detection for Stochastic Processes". Bachelor's thesis. Faculty of Physics, University of Warsaw, July 2026.

[2] G. Manten et al. "Signature Kernel Conditional Independence Tests in Causal Discovery for Stochastic Processes". In: *International Conference on Learning Representations*. 2025. arXiv: 2402.18477.

[3] C. Salvi et al. "The Signature Kernel Is the Solution of a Goursat PDE". In: *SIAM Journal on Mathematics of Data Science* 3.3 (2021), pp. 873–899.

[4] A. Gretton et al. "Kernel Methods for Measuring Independence". In: *Journal of Machine Learning Research* 6 (2005), pp. 2075–2129.

[5] S. Lee and V. Honavar. "Self-Discrepancy Conditional Independence Test". In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. 2017.

[6] M. Schuld and N. Killoran. "Quantum Machine Learning in Feature Hilbert Spaces". In: *Physical Review Letters* 122 (2019), p. 040504.

References II

[7] V. Havlicek et al. "Supervised Learning with Quantum-Enhanced Feature Spaces". In: *Nature* 567 (2019), pp. 209–212.

[8] T. Hubregtsen et al. "Training Quantum Embedding Kernels on Near-Term Quantum Computers". In: *Physical Review A* 106 (2022), p. 042431.

[9] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes. "Exponential Concentration in Quantum Kernel Methods". In: *Nature Communications* 15 (2024), p. 5200.

[10] J. C. Spall. "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation". In: *IEEE Transactions on Automatic Control* 37.3 (1992), pp. 332–341.

[11] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In: *International Conference on Learning Representations*. 2015.

[12] D. C. McKay et al. "Efficient Z Gates for Quantum Computing". In: *Physical Review A* 96 (2017), p. 022330.