

# Pauli Correlation Encoding for Financial Optimization and Statistical Estimation: Scalable Portfolio Construction and Quantum Learning of Covariance Structures

*Extended Abstract — QTML 2026*

Vicente P. Soloviev<sup>1</sup>, Michal Krompiec<sup>1</sup>, and Bibhas Adhikari<sup>2</sup>

<sup>1</sup>Fujitsu Research of Europe Ltd., Madrid, Spain / Slough, UK

<sup>2</sup>Fujitsu Research of America Inc., Santa Clara, CA, USA

## Abstract

Pauli Correlation Encoding (PCE) is a recently introduced, qubit-frugal paradigm that assigns *multiple problem variables to a single qubit* by mapping each variable to the expectation value of a multi-qubit Pauli string, encoding  $m = \mathcal{O}(n^k)$  variables into only  $n$  qubits at correlation order  $k$ . This extended abstract synthesizes two complementary lines of work built on this foundation. The first applies PCE, in its native binary form, to **large-scale portfolio optimization**, solving realistic market-graph instances with more than 250 assets on a gate-based variational pipeline. The second extends PCE from binary to real-valued variables —**continuous-domain PCE (cPCE)**— and repurposes it as a quantum machine-learning framework for **estimating classical covariance and correlation matrices**. Together they show that one encoding principle can address both the combinatorial decision layer and the statistical estimation layer of quantitative finance while relaxing the linear qubit-per-variable bottleneck that constrains conventional gate-based methods such as QAOA and VQE.

## 1 Overview and Motivation

Classical quantum-variational approaches to combinatorial finance problems typically assume a one-to-one correspondence between qubits and optimization variables (e.g. financial assets), so the qubit count grows linearly with problem size,  $\mathcal{O}(n)$ . Current hardware limits this to small instances, and to date gate-based portfolio problems have remained below roughly 100 assets [1]. Pauli Correlation Encoding [6] breaks this correspondence by reading each variable from a Pauli-string expectation value, packing  $\mathcal{O}(n^k)$  variables into  $n$  qubits. The two works synthesized here exploit that idea at two layers of the quantitative-finance stack: the *decision* layer (which assets to hold) and the *estimation* layer (the covariance/risk inputs that feed the decision), with the latter feeding directly into the former.

## 2 Common Methodological Core

Both works share the same variational scaffolding. A parameterized state  $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle^{\otimes n}$  is prepared with a Hardware-Efficient Ansatz (HEA) [4] of single-qubit  $R_y$  rotations and CZ entangling gates in a linear/ring topology. Variables are read from expectation values  $\langle \Pi \rangle = \langle \psi(\boldsymbol{\theta}) | \Pi | \psi(\boldsymbol{\theta}) \rangle$  of mutually commuting Pauli strings of the form  $X^{\otimes k} \otimes I$ ,  $Y^{\otimes k} \otimes I$ , or  $Z^{\otimes k} \otimes I$ . Mutual commutativity lets the observables be measured simultaneously in a shared eigenbasis, central to the efficiency of both pipelines. In the binary case variables are recovered as  $x_i = \text{sgn}(\langle \Pi_i \rangle)$ ; in the continuous case as scaled expectation values  $x_i = c_i \langle \Pi_i \rangle$ . A rescaling factor  $\alpha \approx n^{\lfloor k/2 \rfloor}$  keeps the nonlinear tanh loss out of its purely linear regime, improving trainability.

### 3 Part A — Large-Scale Portfolio Optimization via PCE

#### 3.1 Problem and encoding

A market graph  $G$  is built from five years of real S&P 500 data, with assets as nodes and edges weighted by  $w_{ij} = 1 - |\rho_{ij}|$  whenever the pairwise Pearson correlation exceeds a threshold  $\lambda$ . Portfolio construction is framed as recursively bipartitioning this graph to group highly correlated (redundant) assets and separate weakly correlated ones, by maximizing a cut objective

$$\text{Cut}(G, \mathbf{x}) = \sum_{v_i, v_j \in E} w_{ij} x_i (1 - x_j). \quad (1)$$

The binary partition variables come from PCE rather than one-hot qubit assignment, so a register of only  $n$  qubits at order  $k$  encodes up to  $3\binom{n}{k}$  variables (e.g.  $n = 3$ ,  $k = 2$  encodes nine variables via  $\{XXI, XIX, IXX, YYI, \dots, IZZ\}$ ).

#### 3.2 Pipeline

An iterative bipartition algorithm repeatedly calls a PCE solve on subgraphs (COBYLA optimizing the HEA parameters against the regularized tanh loss), yielding  $n_{\text{splits}} + 1$  clusters of correlated assets; the highest-return member of each cluster is selected to form a diversified, reduced portfolio.

#### 3.3 Key results

The method scales to  $m > 250$  assets solved in roughly one hour on a statevector simulator, well beyond the sub-100-asset ceiling of prior gate-based work. Gate counts stay modest: PCE remains under  $\sim 750$  gates even at 250 nodes /  $\sim 21,000$  edges, whereas QAOA [3] ( $p = 2$ ) exceeds 42,000 gates on the same instance. Benchmarked against QAOA (small instances) and Estimation-of-Distribution Algorithms (large instances), the PCE-selected portfolios consistently beat the passive S&P 500 baseline (Fig. 1) and achieve higher *Sharpe ratios* on both training and test splits (Fig. 2), with the most stable growth trajectory across all sizes tested ( $m$  from 10 to 250,  $n = 9$ ).

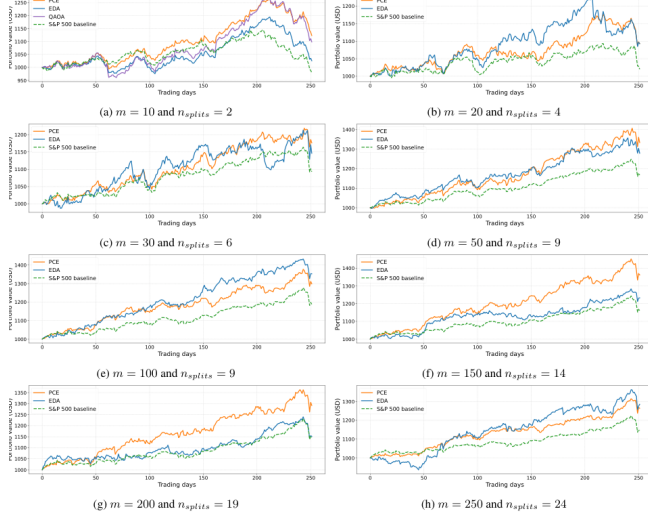


Figure 1: Comparison of cumulative portfolio value over time between a passive investment in the S&P 500 index (green dashed line) and the proposed portfolio optimization strategies. Optimized portfolios obtained via PCE, QAOA, and EDA are shown in orange, purple, and blue, respectively.

st dataset, and their performance is visualized in the corresponding subplots. The figure illustrates the evolution of an initial investment of \$1000 over time, following the portfolio strategy derived from each optimizer, and compares it against

quadratic scaling of entangling operations in dense graphs. For instance, QAOA with  $p = 2$  requires over 42,000 gates for a graph with 250 nodes and nearly 21,000 edges. In contrast, PCE exhibits a much more modest growth in gate count

**Figure 1:** Cumulative portfolio value over time for PCE, QAOA, and EDA versus the passive S&P 500 baseline (green dashed), across sizes  $m \in \{10, \dots, 250\}$ . PCE (orange) shows the most stable growth and a consistent lead. *Adapted from Soloviev & Krompiec [7].*

## 4 Part B — Quantum Learning of Covariance via cPCE

### 4.1 From binary to continuous

The second work generalizes PCE to real-valued variables (cPCE) so the same encoding can learn the  $\binom{n}{2}$  off-diagonal entries of a covariance matrix  $\Sigma$  for  $n$  random variables. Parameters are trained by minimizing the Frobenius distance between the quantum-derived estimator and a given classical target estimator. Two estimator families are proposed, with a clear qubit-versus-simplicity trade-off.

### 4.2 C-Estimator

Builds the estimator through a Cholesky-type factorization  $\hat{\Sigma} = LL^\top$ , with off-diagonal Cholesky entries  $l_{ij} = c_{ij}x_{ij}$  set by expectation values and diagonal entries fixed to reproduce each  $\text{Var}(X_i)$ . This guarantees positive semidefiniteness by construction, needs only  $\eta \approx \lceil \sqrt{\binom{n}{2}} \rceil$  qubits ( $k = 2$ ), enables rank- $r$  approximation by zeroing  $n - r$  diagonal entries, and yields the precision (inverse) matrix cheaply via triangular solves rather than explicit inversion.

### 4.3 E-Estimator

Estimates covariance entries directly as  $\hat{\Sigma}_{ij} = c_{ij}\langle\psi(\theta)|\Pi_{ij}|\psi(\theta)\rangle$  on an  $n$ -qubit register using two-body  $X \otimes X$  Pauli strings. It is computationally simpler—entries are scaled expectation values of a single commuting set—at the cost of more qubits ( $n$  rather than  $\eta$ ). A *diagonal-dominance* sufficient condition on the regularization coefficients,

$$\sum_{l>k} c_{kl} + \sum_{l<k} c_{lk} \leq \text{Var}(X_k), \quad 1 \leq k \leq n, \quad (2)$$

is derived to guarantee a positive semidefinite estimator.

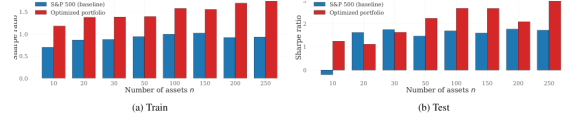


Figure 2: Comparison of the Sharpe ratio obtained by the optimized portfolios and the non-optimized S&P 500 baseline as a function of the number of assets  $n$  in train (80%) and test (20%) datasets, respectively.

**Figure 2:** Sharpe ratio of the optimized PCE portfolio versus the S&P 500 baseline vs. the number of assets  $n$ , on train (80%) and test (20%) splits. The optimized portfolio yields superior risk-adjusted performance.

#### 4.4 Barren-plateau analysis

For HEA circuits with  $\text{poly}(n)$  layers forming an approximate 2-design [2, 5], the gradient variance of the E-Estimator loss is bounded as

$$\text{Var}_{\theta} \left( \frac{\partial \mathcal{L}}{\partial \theta_{l,j}} \right) = \mathcal{O} \left( \frac{\|\mathbf{c} \circ \mathbf{c}\|_2^2 + \|\mathbf{c}\|_2^2 \|\Sigma^s\|_F^2}{2^n} \right). \quad (3)$$

Although the naive scaling is  $\mathcal{O}(n^4/2^n)$ , the bound reveals that the plateau can be *mitigated* when some regularization coefficients grow exponentially in  $n$  —turning a hyperparameter into a trainability lever.

#### 4.5 Empirical findings

On random PSD matrices ( $\Sigma = AA^\top$ ), the C-Estimator converges to substantially lower mean-absolute error than the E-Estimator (Fig. 3) and exploits low-rank structure (error drops as the assumed rank approaches the true rank). It degrades gracefully as missing entries increase in covariance-completion tests (Fig. 4) and stays stable on near-singular, ill-conditioned matrices.

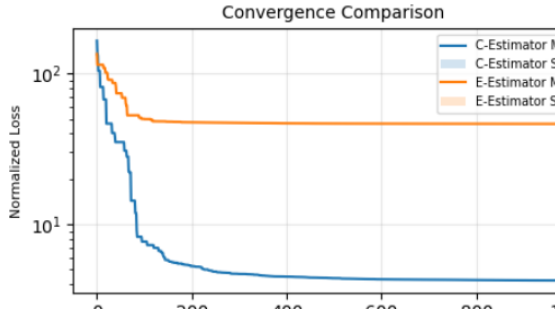


**Figure 3:** Convergence of the C- and E-Estimator: best-so-far MAE across iterations. The C-Estimator (blue) reaches markedly lower error.

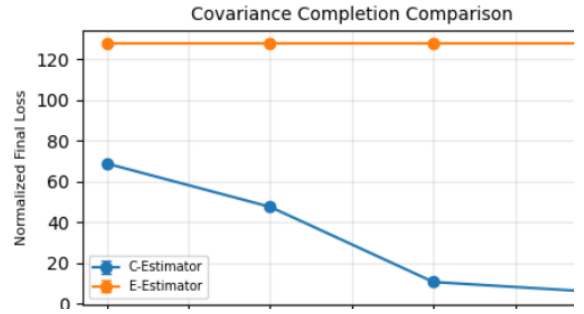


**Figure 4:** Covariance completion: final MAE versus the fraction of missing entries. The C-Estimator (blue) degrades gracefully.

### 5 Synthesis and Significance

Read together, the two works trace one encoding idea across the quantitative-finance stack: Part A delivers **scalable combinatorial optimization** (portfolio selection at hundreds of assets with order-of-magnitude gate savings and superior risk-adjusted returns), while Part B delivers **scalable statistical estimation** (covariance, precision, and completed matrices with provable PSD guarantees and a route around barren plateaus). Since the covariance matrices learned in Part B are exactly the risk inputs consumed in Part A, the two suggest a natural end-to-end pipeline: estimate correlation structure with cPCE, then optimize allocation with PCE, under a shared qubit-frugal, commuting-observable framework suited to near-term hardware.

**Open directions.** Both works flag concrete next steps: shot-based and real-hardware execution with noise-resilience analysis, alternative ansätze beyond HEA, richer representative-asset selection for portfolios, and —on the estimation side— broader benchmarking of cPCE against strong classical covariance/precision estimators specifically in the high-dimensional regime.

## References

- [1] A. Abbas et al. Challenges and opportunities in quantum optimization. *Nature Reviews Physics*, 2024.
- [2] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1791, 2021.
- [3] E. Farhi, J. Goldstone, and S. Gutmann. A quantum approximate optimization algorithm. *arXiv:1411.4028*, 2014.
- [4] A. Kandala et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.
- [5] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, 2018.
- [6] M. Sciorilli, L. Borges, T. L. Patti, D. García-Martín, G. Camilo, A. Anandkumar, and L. Aolita. Towards large-scale quantum optimization solvers with few qubits. *Nature Communications*, 16(1):476, 2025.
- [7] V. P. Soloviev and M. Krompiec. Large-scale portfolio optimization using Pauli Correlation Encoding. 2025.
- [8] V. P. Soloviev and B. Adhikari. Quantum learning of classical correlations with continuous-domain Pauli Correlation Encoding. 2025.