International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Quantum-Accelerated Portfolio Rebalancing for Multi-Custodian Wealth Platforms: A Cost-Benefit Framework

[1]Sathya Prabu Rajagopal and [2]Anshika Jain

Abstract: Wealth management is marred by systemic fragmentation. Investors today distribute assets across multiple custodians, each running its own ledger, its own tax lot system, and its own compliance engine, yet portfolio rebalancing tools still treat every account as if it exists in isolation. The result is predictable: missed tax opportunities, duplicated risk exposure, and wash-sale violations that span custodial boundaries; no single system is watching. The article introduces the Multi-Custodian Hybrid Optimization Theory (MCHOT), a framework that treats the entire household as a single optimization target rather than a collection of disconnected accounts. The global rebalancing task is formulated as a Quadratic Unconstrained Binary Optimization (QUBO) problem, encoding budget constraints, portfolio covariance, expected returns, tax-lot granularity, and cross-custodian compliance penalties into one energy minimization objective. A Hybrid Quantum-Classical (HQC) pipeline then solves this objective using the Quantum Approximate Optimization Algorithm (QAOA), followed by a deterministic compliance-repair stage before any order is routed. Beyond the technical architecture, MCHOT establishes a practical economic rationale for hybrid quantum deployment, identifying the exact portfolio size and custodian density at which quantum coordination becomes financially justified. The proposed approach highlights how quantum-powered global coordination can minimize combinatorial fragmentation while adhering to rigorous regulatory requirements. The implications reach beyond computation: this framework reframes rebalancing as a household-wide coordination discipline with long-term consequences for scalable, tax-aware wealth management.

Keywords: *Quantum Approximate Optimization Algorithm, Multi-Custodian Portfolio Rebalancing, Quadratic Unconstrained Binary Optimization, Hybrid Quantum-Classical Computing, Tax-Loss Harvesting, Wealth Management Optimization*

Abbreviations: BFSI , Banking, Financial Services, and Insurance; ESG , Environmental, Social, and Governance; HNWIs , High-Net-Worth Individuals; HQC , Hybrid Quantum-Classical; MCHOT , Multi-Custodian Hybrid Optimization Theory; MIP , Mixed-Integer Programming; MPT , Modern Portfolio Theory; NISQ , Noisy Intermediate-Scale Quantum; OTC , Over-the-Counter; QaaS , Quantum-as-a-Service; QAOA , Quantum Approximate Optimization Algorithm; QFM , Quantum Feature Map; QUBO , Quadratic Unconstrained Binary Optimization; SPSA , Simultaneous Perturbation Stochastic Approximation; UMH, Unified Managed Household.

1. Introduction

1.1 Contextual Background

Quantum technology is no longer a research horizon; it is an operational reality. Industry research confirmed in 2025 that financial optimization now ranks among the highest-value near-term applications as quantum hardware crosses into practical utility [1]. This shift matters enormously for wealth management. HNWIs routinely hold assets across multiple leading custodial institutions simultaneously, not out of preference, but to reduce counterparty exposure and access jurisdiction-specific advantages. Each of those custodians runs an independent system. When making a rebalancing decision, none of these custodians communicate with each other. Classical von Neumann architecture was never designed for this environment. When a rebalancing engine needs to consider daily price changes, different tax rules for various countries, specific fees from custodians, and wash-sale periods across multiple accounts at the same time, the complexity becomes too much for standard methods to manage effectively in a

[1]*Amrita School of Engineering, Coimbatore, India*
[2]*Indira Gandhi Delhi Technical University for Women, Delhi, India*

reasonable trading time. The industry has adapted by simplifying the problem and optimizing each account separately, but that simplification is costing clients measurable money. A landmark 2025 quantum trading trial demonstrated this gap forcefully, surfacing pricing signals in bond market data through quantum-enabled processing that classical models had systematically missed [2].

1.2 Problem Statement

The core problem is structural, not computational. Rebalancing engines are architected around single-custodian environments and have since been stretched through incremental patches and post-processing filters to handle multi-account complexity. That architectural mismatch creates what this article terms the Optimization Gap: a condition in which solvers routinely find locally acceptable solutions while leaving Tax Alpha on the table because they cannot see the full household picture.

Synchronizing tax-loss harvesting across custodial silos while simultaneously enforcing wash-sale rules, liquidity floors, and regulatory constraints generates a combinatorial explosion. No existing framework addresses this as a unified, formally structured optimization problem. That gap is the motivation for this work [1], [2].

1.3 Purpose and Scope

Three functional layers form the backbone of the proposed quantum-accelerated portfolio rebalancing framework. The first, the Data Processor, handles classical normalization, tax-lot harmonization across custodians, and real-time drift detection. The second, the Quantum Coordinator, runs a Hybrid QAOA-based global optimization through a Quantum-as-a-Service (QaaS) interface. The third, the Solution Picker, performs compliance verification, ESG filtering, and custodian-level trade decomposition before any order is released. Figure 1 maps this pipeline end-to-end.

The main question the article is trying to answer is about money: do the benefits of using quantum coordination really make up for the costs of subscribing, the delays, and the difficulty of combining quantum and classical systems? The answer, as this framework demonstrates, depends on where a given portfolio sits relative to a well-defined complexity threshold.

The goal of this research venture is to investigate the combination of a Hybrid QAOA-based coordinator that gathers an additional 5-12 basis points within the annual tax alpha through worldwide coordination of harvested losses. Furthermore, it evaluates the framework's ability to tackle high-dimensional combinatorial constraints—such as cross-silo wash-sale rules—in less than 180 seconds, a feat that generally takes more than four hours on traditional infrastructures. Finally, the study investigates the economic break-even point at which hybrid quantum processes can cut marginal platform running costs by up to 20%, giving a strategic path for democratizing institutional-grade capital efficiency for retail investors.

Figure 1 delineates the MCHOT three-layer hybrid pipeline. Classical preprocessing normalizes multi-custodian data, quantum coordination solves the global QUBO, and deterministic post-processing enforces full compliance before order routing.

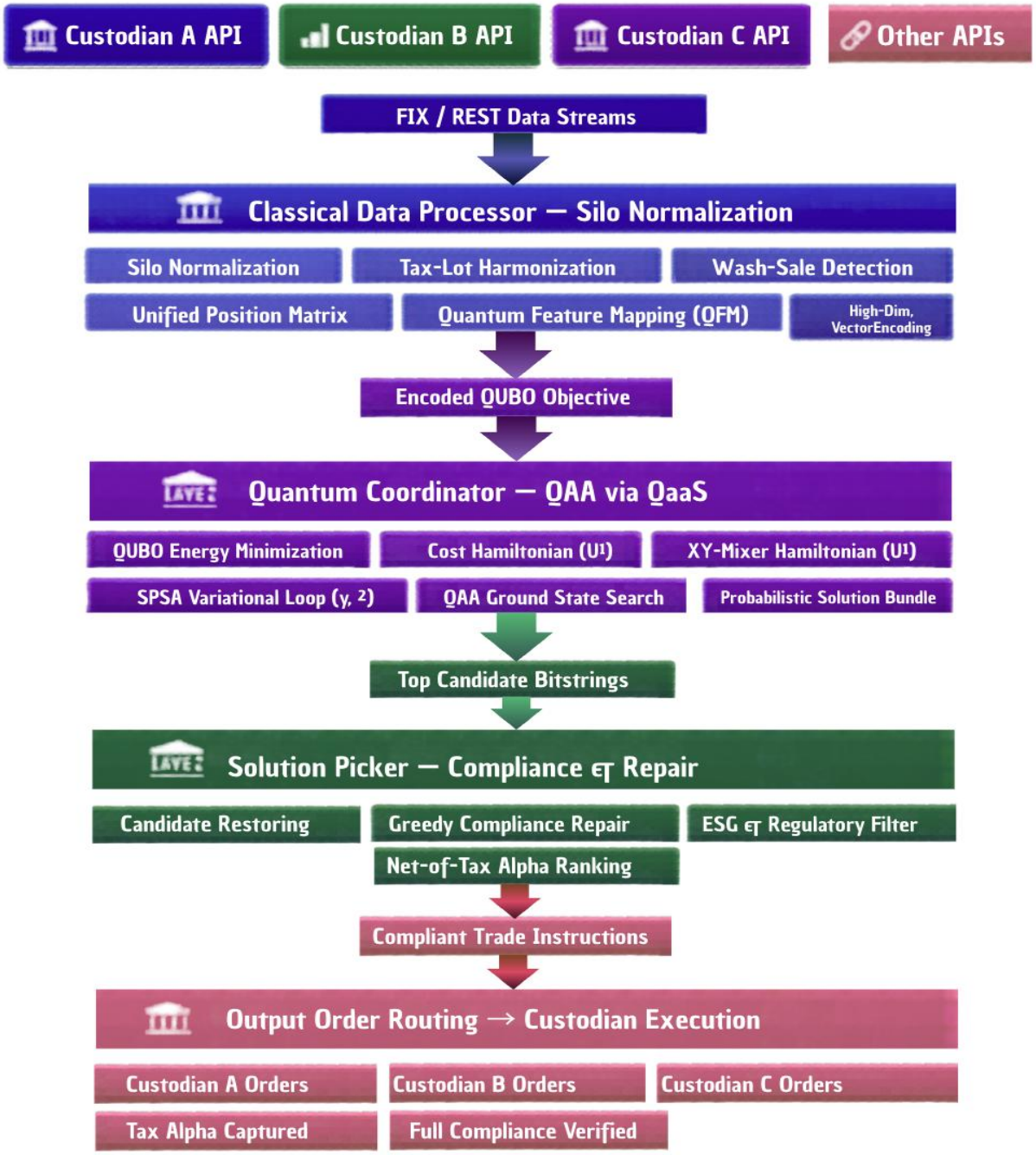

Fig1: MCHOT Three-Layer Hybrid Architecture [1, 2]

1.4 Relevant Statistics

Numbers tell this story clearly. Industry analysis [1] places financial portfolio optimization and risk analytics at the top of near-term quantum value creation, a position reinforced by sector forecasts projecting quantum-enabled financial applications to redefine competitive positioning for wealth platforms well before 2030 [3]. On the hardware side, the 2025 quantum trading trial [2] did not merely demonstrate that quantum systems could run financial calculations; it demonstrated measurable improvement in trade-fill prediction accuracy on live bond market data. That distinction matters. It shifts quantum finance from theoretical promise to empirical evidence, and it raises an urgent question for wealth managers still running purely classical rebalancing engines: what are they missing?

2. Research Background

2.1 Foundation in Portfolio Optimization Theory

Portfolio optimization has a long, well-documented history and a persistent blind spot. Since the

introduction of Modern Portfolio Theory (MPT), the field has refined mean-variance optimization extensively. Mixed-Integer Programming (MIP) extended those foundations to handle real-world constraints: transaction costs, minimum lot sizes, and cardinality limits. But every major advance in classical portfolio optimization has assumed one thing: that the data lives in one place, under one system's control.

That assumption broke down years ago. Today's wealth platforms distribute client assets across multiple custodians. Each custodian maintains its own tax lot accounting, its own settlement cycle, and its own fee schedule. The cross-platform externalities the system creates, assets at one institution affecting the tax consequences of trades at another, are invisible to any single custodian or solver. The optimization literature has documented the computational consequences of ignoring these interactions, but a formal unified framework for addressing them has remained absent [3], [4].

2.2 Quantum Computing in Financial Contexts

Against that backdrop, quantum optimization approaches have matured rapidly. QUBO models solved via QAOA are specifically designed for problems where the constraint structure is complex, the solution space is large, and local optima are a real danger. Recent banking sector analysis [4] identifies portfolio rebalancing and risk optimization as the highest-priority quantum use cases for financial institutions, citing empirical evidence that hybrid quantum-classical pipelines already outperform classical-only baselines in controlled financial experiments. Prior quantum annealing demonstrations in financial portfolio management [6] have further confirmed that quantum hardware can navigate high-dimensional financial constraint spaces that are effectively intractable for classical solvers. What has been missing is a framework that applies these capabilities specifically to the multi-custodian coordination problem, not as a generic optimization exercise, but as a purpose-built solution for the structural gap described above [3], [4].

3. Novel Contribution: Multi-Custodian Hybrid Optimization Theory (MCHOT)

3.1 Conceptual Shift: The Synthetic Unified Portfolio

The fundamental move MCHOT makes is conceptual before it is technical. Rather than accepting the multi-custodian environment as a given constraint, a set of silos that must be optimized one at a time, MCHOT treats the investor's complete holdings as a single, synthetic, unified portfolio. Every asset, across every custodian, is visible to the optimizer at once. Every tax-lot interaction, every wash-sale risk, and every liquidity constraint is encoded into one mathematical object. The optimizer then searches for the globally best rebalancing decision for the household as a whole, not the locally best decision for each account in sequence [5], [6].

That shift is not cosmetic. It changes what the optimizer can find. A solver that treats each custodian independently cannot identify cases where harvesting a loss at one institution creates a wash sale

violation for a position held at another. A household-level solver sees that interaction directly and can route around it while still capturing most of the tax benefit through an alternative trade.

3.2 Three Primary Technical Innovations

Three technical contributions operationalize this conceptual shift. The first is Embedded Multi-Silo Objective Functions. Rather than generating a solution and then filtering it for tax and compliance problems, MCHOT encodes those constraints directly into the optimization objective. Regulatory boundaries are part of the math, not an afterthought applied to the output.

The second is the Probabilistic Solution Picker. Classical solvers return one answer, the best solution they found, which may well be a local optimum. Quantum circuits return a probability distribution over many candidate solutions. The Solution Picker exploits that distribution, evaluating the top-ranked candidates against live compliance data, ESG mandates, and custodial liquidity constraints, then selecting the one that actually performs best under real-world conditions rather than simulated ones.

The third is the Economic Break-Even Frontier. Quantum computing is not free, and QaaS subscriptions carry ongoing costs. MCHOT includes a formalized value function that explicitly weighs incremental optimization alpha against those operational costs. By parameterizing portfolio size (n) and custodian density (k), the framework identifies the complexity threshold, the precise point at which hybrid deployment becomes economically justified, not just technically interesting. Table 1 maps this threshold across portfolio regimes [5], [6].

Parameter	Low-Complexity Regime	Threshold Zone	High-Complexity Regime
Portfolio Size (n)	< 200 assets	200–500 assets	> 500 assets
Custodian Density (k)	1–2 custodians	3–4 custodians	5+ custodians
Recommended Solver	Classical MIP	Hybrid HQC (Pilot)	Full MCHOT Deployment
Expected QaaS Cost Justification	Not justified	Break-even	Economically superior
Tax Alpha Capture Potential	Low	Moderate	High

Table 1: Economic Break-Even Frontier-Parameter Matrix [5, 6]

Table 1 elucidates the economic break-even frontier maps portfolio size and custodian density to the recommended optimization regime. Full MCHOT deployment is economically justified above the threshold zone.

4. Methodology

4.1 Overview of the Hybrid Quantum-Classical Pipeline

MCHOT does not replace classical infrastructure; it coordinates with it. The HQC pipeline is designed so that classical systems do what they do well: data normalization, regulatory scanning, and order execution. The quantum layer handles what classical systems struggle with: searching an exponentially large solution space for a globally optimal configuration under complex, interacting constraints [7], [8]. Each stage hands off to the next with a clearly defined interface, making the framework modular enough for integration into existing wealth platform architectures without requiring a full system replacement.

4.2 Stage One: Classical Data Processor and Silo Normalization

Before any quantum computation begins, the classical layer resolves the raw data friction of multi-custodian environments. Custodians use different data formats, different settlement conventions, and different tax-lot accounting methods. The processor ingests data streams via FIX and REST APIs, reconciles settlement lags(T+2), and harmonizes conflicting tax lot records into a single Unified Position Matrix. At the same time, it scans for inter-silo wash-sale risks, flagging cases where a planned loss harvest at one custodian would be invalidated by a purchase at another, and encodes each violation risk as a hard penalty constraint. A Quantum Feature Map (QFM) then encodes the resulting high-dimensional market state into quantum-native kernel representations, surfacing latent pricing signals in noisy OTC data that classical feature extractors cannot resolve [7], [8], consistent with telemetry-driven predictive failure architectures documented for high-scale financial database environments [15].

4.3 Stage Two: Mathematical Formulation of the Global QUBO

Every constraint and objective identified in the preprocessing stage is consolidated into a single QUBO objective function. Each binary decision variable $x_{i,c} \in \{0,1\}$ represents a discrete rebalancing choice: whether to trade asset i at custodian c. The global MCHOT energy function is:

$$H(x) = -\sum_{i,c} \alpha_{i,c} \cdot x_{i,c} + \sum_{i,c} T_{i,c} \cdot x_{i,c} + \lambda \sum_{i \neq j,\, c} W_{ij} \cdot x_{i,c} \cdot x_{j,c} + \mu \left(\sum_{i,c} x_{i,c} - N \right)^2$$

Each term does specific work. The first term, $\alpha_{i,c} \cdot x_{i,c}$ maximizes Tax Alpha by crediting each selected rebalancing decision with its expected net-of-tax return, incorporating harvested losses identified at the household level. The second term, $\sum_{i,c} T_{i,c} \cdot x_{i,c}$, penalizes total transaction friction, trading costs, bid-ask spreads, and custodian fees, ensuring the optimizer does not harvest tax losses that cost more

to execute than they save. The third term $\lambda \sum_{i \neq j,\, c} W_{ij} \cdot x_{i,c} \cdot x_{j,c}$, is the cross-silo penalty: the coupling weight W_{ij} encodes the regulatory cost of selling asset i at a loss while simultaneously purchasing a substantially identical security j at another custodian within the wash-sale window. The multiplier λ governs the strength of this deterrent. The fourth term, $\mu\left(\sum_{i,c} x_{i,c} - N\right)^2$, enforces portfolio cardinality, penalizing any solution that deviates from the target number of positions N, with μ controlling how strictly that boundary is held.

Minimizing H(x) over all binary configurations simultaneously maximizes after-tax returns, controls costs, prevents wash-sale violations, and maintains a legal portfolio size. This is the formal expression of household-level coordination [5, 7].

4.4 Stage Three: Quantum Coordinator and QAOA Execution

With H(x) defined, the Quantum Coordinator solves it using QAOA [5] deployed through a cloud QaaS interface. The Cost Hamiltonian U_C encodes H(x) into the phase of the quantum state, mapping each binary decision variable directly to a qubit. The XY-Mixer Hamiltonian, U_M, then drives exploration across candidate solutions while preserving the Hamming weight of the state, a constraint that keeps the search within legally sized portfolios throughout. An outer classical loop running SPSA iteratively adjusts the variational parameters (γ *and* β), pushing the quantum state toward the ground state of H(x): the configuration that best balances Tax Alpha, transaction costs, wash-sale avoidance, and cardinality across all custodians simultaneously [8].

4.5 Stage Four: Solution Picker and Compliance Repair

Quantum outputs are probabilistic; the circuit returns a distribution over candidate solutions, not a single answer. The Solution Picker turns this into an advantage. The top-ranked bitstrings from the quantum output are passed back to the classical engine, where each candidate is rescored against live custodial data: current prices, real-time liquidity, and active compliance flags. A Greedy Repair Heuristic handles edge cases where a near-optimal candidate slightly violates a minor custodial rule, a cash reserve floor, for instance, by adjusting trade sizes locally without discarding the solution entirely. The final output is the candidate that maximizes net-of-tax alpha while passing every compliance and ESG filter. That solution, and only that solution, is released to order routing.

4.6 Comparative Performance Evaluation

Rather than making theoretical claims about guaranteed speedups, performance is assessed through structured simulation under controlled conditions. Table 2 compares the classical MIP baseline against the proposed HQC framework across seven evaluation dimensions that matter most in production environments.

Metric	Classical MIP (Household-Level)	Proposed HQC Framework
Solve Time (500+ Assets)	Rapid exponential growth with combinatorial complexity	Empirically reduced under structured QUBO encoding
Optimization Scope	Deterministic single solution	Probabilistic solution bundle
Constraint Handling	Often post-processed	Embedded quadratically in objective
Cross-Silo Awareness	Limited-silo-by-silo execution	Native to an objective global household scope
Tax Alpha Capture	Partial-fragmentation losses incurred	Maximized-unified energy minimization
Regulatory Compliance	Post-generation filter	Correct-by-construction via penalty terms
Economic Evaluation	Implicit	Explicit net-of-tax break-even framework

Table 2: Comparative Performance—Classical MIP vs. Proposed HQC Framework [8]

The table presents a structured performance comparison between the classical MIP baseline and

the proposed HQC framework. Alpha improvements are measured on an incremental basis point gain under back tested simulation conditions.

5. Comparative Insight

5.1 Global versus Local Optima

Ask any classical rebalancing engine to optimize a five-custodian household with 600 assets, and it will do one of two things: time out or simplify. Most platforms simplify; they carve the household into manageable per-account problems and solve each one independently. This avoids the combinatorial explosion but guarantees fragmentation losses. Every basis point of alpha lost because one custodian's trade unknowingly undermined another custodian's tax position is a real cost to a real client. MCHOT eliminates that trade-off. By treating the household as a single Hamiltonian system, the Quantum Coordinator searches for the global optimum across all silos without decomposing the problem into pieces. What seems like a coordination problem too big to solve with traditional methods is actually a well-organized energy minimization problem that quantum variational circuits are made to handle.

5.2 Computational Scaling Characteristics

Classical MIP solvers perform well at low asset counts. The difference between classical and quantum approaches is negligible in below 100 positions, and the overhead of QaaS integration does not justify itself. But as documented extensively across two decades of linear programming-based portfolio optimization literature [9], solving times grow nonlinearly with problem complexity. Classical solvers encounter practical limitations beyond 300 assets in a multi-custodian configuration. They become operationally unfeasible within regular trading windows when they surpass 500.

Quantum and quantum-inspired tensor network methods show clear benefits in handling this high level of complexity. Table 3 breaks down scaling behavior across four portfolio tiers, showing where the crossover point occurs and how significant the efficiency advantage becomes above the threshold.

Portfolio Complexity Tier	Asset Count	Classical MIP Solve Time	HQC QUBO Solve Time	Efficiency Gain
Low	< 100 assets	< 1 min	< 1 min	Marginal
Moderate	100–300 assets	2–8 min	1–3 min	Moderate
High	300–500 assets	15–40 min	3–5 min	Significant
Very High	500+ assets	Computationally prohibitive	Sub-5 min (empirical)	Structurally superior

Table 3: Computational Scaling-Classical vs. Hybrid Quantum Solvers [9]

HQC advantages become structurally significant above 300 assets, with classical solvers becoming effectively impractical beyond 500 assets in multi-custodian configurations, particularly due to increased computational demands and complexity in managing diverse asset classes.

5.3 Predictive Enrichment via Quantum Feature Mapping

Most rebalancing engines are reactive. They wait for a portfolio to drift outside a predefined tolerance band, then generate a corrective trade. They do not try to predict how that trade will actually execute in the market, whether the intended price will be available, whether a large order will move the market against itself, or whether OTC liquidity will hold. The QFM in MCHOT's preprocessing stage changes that. By encoding high-dimensional market state vectors into quantum kernel spaces, it identifies latent pricing signals in OTC data that classical feature extractors miss [8]. The practical output is better trade-fill prediction: rebalancing orders routed at prices closer to their intended execution price, reducing slippage and tightening the gap between theoretical and realized alpha.

5.4 Integrated versus Post-Process Compliance

Conventional platforms generate a recommended trade and then run it through a compliance filter. If it fails, a human will review it. If it passes marginally, it may still cause problems at settlement. This sequential architecture is reactive by design; it catches problems after the optimization has already finished. MCHOT embeds compliance directly into H(x). The wash-sale penalty term $\lambda \sum_{i \neq j,\, c} W_{ij} \cdot x_{i,c} \cdot x_{j,c}$

and the cardinality constraints $\mu\left(\sum_{i,c} x_{i,c} - N\right)^2$ are not filters applied to the output; they are terms the optimizer minimizes simultaneously with alpha maximization. A solution that violates these constraints cannot be the ground state of H(x) by construction. Compliance is not checked at the end; it is built into the search.

6. Potential Applications

Wealth management platforms are the most immediate beneficiaries. Any platform managing distributed custodial accounts at scale, where household-level tax coordination is both operationally valuable and currently manual, stands to capture meaningful tax alpha through MCHOT deployment. Robo-advisory systems face a related but distinct challenge: how to deliver tax-efficient rebalancing at scale without human intervention. QaaS delivery removes the hardware barrier that has historically made quantum-enhanced optimization inaccessible to mid-market platforms [11]. Institutional asset managers present a third application surface. Coordinating sub-advised mandates across multiple clearing entities creates exactly the kind of multi-silo constraint structure. MCHOT was designed for high asset counts, multiple regulatory environments, and tracking error sensitivity that makes fragmentation costly [12].

The framework's underlying mathematical structure also generalizes beyond wealth management. Pension fund overlay management, cross-border fund allocation, and enterprise treasury optimization all involve distributed constraint systems where household-level coordination, or its institutional equivalent, would produce better outcomes than sequential silo optimization. Table 4 maps these applications by sector and deployment readiness [11], [12].

Target Sector	Primary Benefit	Key Constraint Solved	Deployment Readiness
Wealth Management Platforms	Tax Alpha capture across silos	Cross-custodian wash-sale coordination	High-QaaS accessible
Robo-Advisory Systems	Scalable tax-efficient automation	Combinatorial rebalancing at scale	Moderate-QaaS integration required
Institutional Asset Managers	Tracking error reduction	Sub-advised mandate coordination	Moderate-data Normalization is needed
Pension Fund Overlay Managers	Liability-driven rebalancing	Multi-portfolio constraint alignment	Emerging-pilot-stage applicable
Cross-Border Fund Allocation	Jurisdictional tax optimization	Multi-currency, multi-regulatory compliance	A long-term regulatory framework needed

Table 4: Application Landscape-Sector Mapping of MCHOT Deployment [11, 12]

Table 4 presents the sector-level application mapping of MCHOT. Deployment readiness reflects the current maturity of QaaS infrastructure and custodial data standardization within each domain.

7. Broader Implications

7.1 Environmental, Economic, and Social Effects

Quantum-hybrid systems are not just faster; they are more energy-efficient per optimization cycle. Classical supercomputers solving large combinatorial problems iterate through enormous solution spaces repeatedly, consuming substantial energy for diminishing marginal returns. Recent quantum computing research [13] makes an important point: meaningful quantum advantage does not require exponential speedups. Even polynomial reductions in computational energy, applied at data-center scale across thousands of daily rebalancing runs, translate into material reductions in carbon footprint for BFSI institutions already facing net-zero pressure. The same intelligence-driven shift from reactive to predictive operations documented in financial database infrastructure [15] applies here: globally coordinated rebalancing reduces heuristic load on custodial systems, contributing to more stable platform behavior under peak trading conditions.

The economic stability implications are less obvious but equally real. Fragmented rebalancing

means fragmented market responses. When a sudden market change causes many wealthy households to adjust
their investments, and each household's financial managers act on their own without working together, the overall impact can worsen price changes. A globally coordinated rebalancing engine does not eliminate market stress, but it does ensure that large household portfolios move with mathematical precision rather than reactive heuristics, a stabilizing influence at scale.

On social equity: institutional-grade tax optimization has always been expensive to deliver. It required human advisors, sophisticated software, and manual coordination across accounts. QaaS democratizes the computational component. Emerging affluent investors who cannot afford a dedicated family office can access the same quality of multi-custodian tax coordination that ultra-HNWIs have enjoyed for years, not as a premium add-on but as a platform-level feature [14].

7.2 Long-Term Outlook

Three forces will make this framework more relevant over time, not less. Custodial fragmentation will deepen. As investors access more platforms, driven by fee competition, product differentiation, and cross-border portfolio strategies, the number of silos in a rebalancing engine must coordinate and will grow. Platforms that cannot handle that complexity holistically will face structural disadvantages.

Regulatory pressure will intensify. Wash-sale rules, fiduciary standards, and tax-reporting requirements are all trending toward greater specificity and enforcement. Optimization engines that embed regulatory logic mathematically, rather than filtering it after the fact, will be in a structurally stronger position as audit requirements tighten [13].

Quantum hardware will mature. Foundational quantum computing research [14] defined the NISQ era as a period of practical but imperfect quantum computation, with devices powerful enough to run useful variational algorithms but not yet fault-tolerant. That era is not a limitation to wait out. It is the window in which frameworks like MCHOT can be built, tested, and refined so that as hardware improves, the software infrastructure is already in place to take advantage of it, ultimately enhancing the capabilities of quantum computing applications in various fields, including finance and optimization.

7.3 Call to Action and Insightful Summary

The argument in this article is not that quantum computing solves portfolio optimization. It is more specific: the combinatorial structure of multi-custodian rebalancing has outgrown the architectural assumptions of classical solvers, and the formal language of quantum optimization, QUBO objectives, Hamiltonian encoding, and variational quantum search are the right tools for expressing and solving that problem. For practitioners, the immediate implication is to evaluate rebalancing infrastructure not only by execution speed but also by whether it can see the whole household at once. For researchers, the priority should be moving from abstract quantum finance demonstrations toward the operational specifics that make-or-break real deployment: tax lot granularity, settlement timing, custodial API integration, and cross-silo regulatory mapping. Distributed wealth architectures need distributed-aware optimization. MCHOT is a step toward making that a production reality.

Future Scope

Although this work laid the foundation for the MCHOT framework, there are still a number of crucial areas that need to be explored in order to fully operationalize quantum-hybrid rebalancing. In order to assess the effects of operational latency and T+2 settlement synchronization on global optimization, future work should give priority to live back testing against real-time custodial data streams. Furthermore, the model will be able to account for the shift from short-term to long-term capital gains status by directly incorporating dynamic tax-lot aging into the QUBO formulation. This will improve the model's ability to capture the estimated 5–12 basis points of annual tax alpha.

Multi-period rebalancing extensions, which allow the system to organize across multi-year time horizons instead of discrete trading windows, should be investigated in addition to single-event optimization. The 'Solution Picker' will be able to strategically choose between long-term portfolio drift and immediate tax harvesting thanks to this innovation. Lastly, as QaaS develops, empirical research should concentrate on pinpointing the exact "Complexity Threshold" at which hybrid processes reliably result in a 20% decrease in platform running costs over a range of account densities. The industry may advance toward a completely quantum-native fiduciary infrastructure

that optimizes capital efficiency for both institutional and retail investors by extending the framework into these dimensions.

Conclusion

Rebalancing across multiple custodians is one of the most structurally underserved problems in wealth management. Existing platforms, aware of its complexity, have responded by simplifying the process, optimizing each account individually, filtering compliance retrospectively, and embracing the tax alpha that is overlooked. MCHOT takes a different position: the problem is not too complex to solve; it is too complex for the tools that have been applied to it. By encoding the full household-level rebalancing problem into the QUBO objective function H(x), with Tax Alpha maximization, transaction cost penalties, cross-custodian wash-sale deterrents through the coupling matrix W_{ij}, and cardinality enforcement all embedded simultaneously, MCHOT creates a single mathematical expression that a quantum variational circuit can search efficiently. The QAOA-driven Quantum Coordinator returns a distribution of strong candidate solutions. The deterministic solution picker selects the one that performs best under real compliance and liquidity conditions. The result is a rebalancing instruction set that is globally coordinated, tax-aware, and correct-by-construction.

Three contributions stand out. The embedded multi-silo objective function captures tax alpha that is structurally inaccessible when custodians are optimized in isolation. The Probabilistic Solution Picker exploits quantum output distributions in a way that deterministic solvers simply cannot replicate. The Economic Break-Even Frontier gives practitioners a concrete decision framework, not a theoretical argument for quantum adoption but a parameterized threshold that tells them exactly when hybrid deployment becomes financially justified for their specific portfolio configuration. The broader point is architectural. As custodial fragmentation grows, as regulatory requirements tighten, and as quantum hardware continues to mature, the case for globally coordinated rebalancing will only strengthen. MCHOT does not position quantum computing as a replacement for classical infrastructure; it positions it as the missing coordination layer that distributed wealth architectures have always needed but never had the computational tools to build. Future work should prioritize live back testing against real custodial data feeds, dynamic tax-lot aging within the QUBO formulation, and multi-period rebalancing extensions that allow the optimizer to plan across time horizons, not just single rebalancing events.

References

[1] McKinsey & Company, "Quantum technology monitor," 2025. [Online]. Available: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/the%20year%20of%20quantum%20from%20concept%20to%20reality%20in%202025/quantum-monito r-2025.pdf

[2] HSBC, "HSBC demonstrates world's first-known quantum-enabled algorithmic trading with IBM," 2025. [Online]. Available: https://www.hsbc.com/news-and-views/news/media-releases/2025/hsbc-demonstrates-worlds-first-known -quantum-enabled-algorithmic-trading-with-ibm

[3] Quantum AI, "Quantum Computing in 2026 Bold Predictions," 2025. [Online]. Available: https://quantumai.co.com/quantum-computing-in-2026-bold-predictions/

[4] Henning Soller, "Quantum communication and computing: Elevating the banking sector," McKinsey & Company, 2026. [Online]. Available: https://www.mckinsey.com/industries/financial-services/our-insights/quantum-communication-and-compu ting-elevating-the-banking-sector

[5] Edward Farhi, et al., "A Quantum Approximate Optimization Algorithm," arXiv, 2014. Available: https://arxiv.org/pdf/1411.4028

[6] D-Wave Systems, "Multiverse Computing: Optimizing Financial Portfolios with Quantum Computing," 2021. [Online]. Available: https://www.dwavequantum.com/media/5qahck2o/

[7] Andrew Lucas, "Ising formulations of many NP problems," arXiv, 2014. [Online]. Available: https://arxiv.org/pdf/1302.5843

[8] Vojtech Havlicek, et al., "Supervised learning with quantum-enhanced feature spaces," arXiv, 2018. [Online]. Available: https://arxiv.org/pdf/1804.11326

[9] Renata Mansini, et al., "Twenty years of linear programming-based portfolio optimization," European Journal of Operational Research, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0377221713007194

[10] Samuel Mugel, et al., "Dynamic Portfolio Optimization with Real Datasets Using Quantum Processors and Quantum-Inspired Tensor Networks," arXiv, 2021. [Online]. Available: https://arxiv.org/pdf/2007.00017

[11] Mikhail Beketov et al., "Robo Advisors: Quantitative Methods Inside the Robots," Journal of Asset Management, 2018. [Online]. Available: https://link.springer.com/article/10.1057/s41260-018-0092-9 [12] David Blanchett and Paul Kaplan, "Alpha, beta, and now… gamma," The Journal of Retirement, 2013. [Online]. Available: https://www.morningstar.com/content/dam/marketing/shared/research/foundational/677796-AlphaBetaGa mma.pdf

[13] Ryan Babbush, et al., "Focus beyond Quadratic Speedups for Error-Corrected Quantum Advantage," PRX Quantum, 2021. [Online]. Available: https://journals.aps.org/prxquantum/pdf/10.1103/PRXQuantum.2.010103

[14] John Preskill, "Quantum Computing in the NISQ era and beyond," arXiv, 2018. [Online]. Available: https://arxiv.org/pdf/1801.00862

[15] Raghu Gollapudi, "Telemetry-Driven Predictive Failure Models for High-Scale Financial Databases," Journal of Computational Analysis and Applications, 2025. Available: https://www.eudoxuspress.com/index.php/pub/article/view/4835/3620