DISSENSUS AI WORKING PAPER SERIES

DAI-2511

Privacy-Preserving Financial Surveillance: An Architectural Framework for CBDC Implementation

Transaction-Level Intervention Without Identity-Level Monitoring

Murad Farzulla[1,2,*], Andrew Maksakov[1]

[1]Dissensus AI, London, UK [2]King's College London, London, UK

[*]Correspondence: murad@dissensus.ai ORCID: 0009-0002-7164-8704

February 2026

Abstract

Current Central Bank Digital Currency proposals from major economies share a common architectural assumption: that comprehensive transaction surveillance is necessary for financial stability and crime prevention. This paper challenges that assumption by proposing an alternative architecture that achieves 87–95% of surveillance-based detection effectiveness while preserving complete transactional privacy. Agent-based simulation demonstrates that the remaining performance gap derives entirely from cross-wallet identity linking—a capability achievable through pseudonymous zero-knowledge techniques without surveillance infrastructure. Crucially, watchlist access provides zero marginal detection improvement: the surveillance apparatus adds no value beyond what pseudonymous linking achieves. The proposed framework operates through three mechanisms: (1) anonymized pattern detection that analyzes transaction graph structure without accessing participant identities, (2) transaction-level rather than identity-level intervention that freezes suspicious transactions without affecting users' broader financial access, and (3) opt-in deanonymization for dispute resolution where users voluntarily provide explanations rather than facing automatic investigation.

The proposed framework inverts the burden of proof in financial surveillance. Rather than requiring users to demonstrate legitimacy (the current AML paradigm), it requires the system to demonstrate suspicion before any intervention occurs—and even then, intervention affects only the specific flagged transaction. Users may abandon flagged transactions without deanonymization, with funds returning to sender and no record created linking identity to suspicious patterns. This creates a game-theoretic deterrent: illicit actors cannot complete transactions, but legitimate users experience minimal friction.

We argue that privacy-preserving CBDC architecture is technically feasible using existing cryptographic primitives (zero-knowledge proofs, secure multi-party computation, threshold cryptography) and that the choice to implement surveillance infrastructure represents a policy decision rather than technical necessity. The paper provides architectural specifications, addresses common objections, and proposes governance structures for privacy-preserving digital currency systems. This framework contributes to ongoing debates about CBDC design by demonstrating that the surveillance–security trade-off is a false dichotomy.

Keywords: central bank digital currency, privacy, financial surveillance, anti-money laundering, zero-knowledge proofs, digital currency architecture, transaction monitoring, opt-in deanonymization

JEL Codes: E42 (Monetary Systems), E58 (Central Banks and Their Policies), G28 (Government Policy and Regulation), K42 (Illegal Behavior and Enforcement of Law), O33 (Technological Change)

Research Context

This work forms part of the Adversarial Systems Research program, which investigates stability, alignment, and friction dynamics in complex systems where competing interests generate structural conflict. The program examines how agents with divergent preferences interact within institutional constraints across multiple domains: political governance, financial markets, human cognitive development, and artificial intelligence alignment.

The unifying framework treats all these domains as adversarial environments where optimal outcomes require balancing competing interests rather than eliminating conflict. In political systems, this manifests as the tension between stakeholder consent and technocratic competence. In financial markets, it appears as the conflict between regulatory stability and market innovation. The privacy-preserving CBDC architecture presented here addresses the specific adversarial dynamic between state surveillance capabilities and citizen financial autonomy.

This paper extends prior work on AML regulatory failures (Farzulla, 2025c), which documented how current frameworks disproportionately target primitive laundering methods while sophisticated actors exploit derivatives, hedging, and offshore structures with impunity. Where that analysis diagnosed the problem—surveillance systems that impose friction on legitimate users while failing to detect sophisticated crime—this paper proposes a solution: architectural enforcement of privacy that achieves superior crime detection through mechanism design rather than comprehensive monitoring.

The contribution lies in demonstrating that the apparent trade-off between privacy and security in digital currency design is a false dichotomy. Privacy-preserving architecture can achieve equivalent or superior crime detection by creating game-theoretic incentives that make illicit transactions structurally impossible to complete, rather than attempting to identify and prosecute illicit actors after the fact.

Acknowledgements

The authors acknowledge the foundational work of David Chaum on eCash and blind signatures, which established the theoretical possibility of privacy-preserving digital payments decades before current CBDC debates, and whose subsequent CBDC proposal with Grothoff and Moser (Chaum et al., 2021) demonstrated that privacy-preserving central bank money is not merely theoretically possible but institutionally viable.

This paper benefited from extended collaboration with Claude (Anthropic), whose contributions to analytical framework development and iterative refinement were substantive. Andrew Maksakov contributed the PET–AML stack simulation implementation (`pet_aml_sim.py`) that validates the architecture's performance envelope. The authors gratefully acknowledge this assistance while taking full responsibility for all claims, errors, and interpretive choices.

This work is part of the Adversarial Systems Research program at Dissensus AI, a broader investigation into stability, alignment, and friction dynamics across political, financial, cognitive, and multi-agent systems. Related papers in the series are available through the Adversarial Systems & Complexity Research Initiative (ASCRI; systems.ac).

The authors welcome feedback, criticism, and collaboration. Correspondence should be directed to murad@dissensus.ai.

Contents

1 Introduction

1.1 The Surveillance Assumption in CBDC Design

Central Bank Digital Currency proposals from the European Central Bank, Federal Reserve, People's Bank of China, and Bank of England share a common architectural assumption: comprehensive transaction visibility is necessary for financial stability, monetary policy transmission, and crime prevention (Bank for International Settlements, 2020; European Central Bank, 2020; Auer et al., 2021). Auer and Böhme (2020) establish the foundational taxonomy for retail CBDC design—direct, indirect, and hybrid architectures—and across all three categories, transaction visibility is treated as a default property rather than a design variable. The comprehensive review by Auer et al. (2022) confirms this convergence: central banks pursue CBDCs for monetary sovereignty, financial inclusion, and payment efficiency, yet the privacy implications of each design choice receive comparatively shallow treatment. This assumption treats surveillance as a feature rather than a cost, presenting the ability to monitor every transaction as an unambiguous benefit for regulators and society.

The assumption is rarely examined critically. CBDC design documents discuss privacy in terms of "appropriate balance" between user privacy and regulatory needs, implicitly accepting that these exist in zero-sum tension (European Central Bank, 2020). Agur et al. (2022) provide the canonical formal model of this trade-off, demonstrating that optimal CBDC design depends on the distribution of privacy preferences across heterogeneous agents and on network externalities in payment adoption. Their framework treats privacy as a welfare-relevant parameter rather than a binary constraint—yet even this sophisticated treatment stops short of asking whether architectural privacy could render the trade-off itself unnecessary. Privacy protections, where proposed, take the form of policy commitments ("data will only be accessed with appropriate authorization") rather than architectural guarantees ("the system cannot provide identity information because it does not possess it").

Demand-side evidence reinforces the urgency of this critique. Choi et al. (2025) conduct a nationally representative survey experiment with over 3,500 participants and find that both the degree of privacy protection and information provision about privacy benefits significantly increase willingness to adopt CBDC—by up to 64% for privacy-sensitive purchases. Their results demonstrate that privacy is not merely a normative desideratum but a revealed design requirement: without adequate privacy guarantees, CBDC adoption may fail to achieve the scale necessary for monetary policy transmission.

This paper argues that the surveillance assumption reflects institutional preferences rather than technical necessity. Brunnermeier and Niepelt (2019) demonstrate that under certain conditions, public and private money are equivalent in their macroeconomic effects—implying that the design of CBDC's surveillance properties is a separable policy decision rather than a structural requirement for financial stability. Privacy-preserving CBDC architecture is feasible using cryptographic techniques developed over the past four decades. The question is not whether such systems can be built, but whether there exists political will to build them. The ECB has claimed that the digital euro will be "the most private electronic payment option" (European Central Bank, 2024), yet the proposed architecture relies on policy commitments ("data will only be accessed with appropriate authorization") rather than structural guarantees—a distinction this paper treats as fundamental.

1.2 The AML Paradox: Surveillance Without Detection

The case for surveillance in digital currency design typically rests on anti-money laundering requirements. Current AML frameworks, codified through Financial Action Task Force recommendations and implemented globally, mandate that financial institutions identify customers, monitor transactions, and

report suspicious activity (Financial Action Task Force, 2012–2023).

Empirical evidence on AML effectiveness, however, reveals a troubling paradox. Despite estimated annual compliance costs exceeding $300 billion globally, less than 1% of illicit financial flows are currently seized (Pol, 2020). The United Nations Office on Drugs and Crime estimates that 2–5% of global GDP ($800 billion to $2 trillion) is laundered annually, with interdiction rates suggesting systemic failure rather than implementation challenges (United Nations Office on Drugs and Crime, 2011).

More critically, AML systems exhibit systematic enforcement asymmetry. Aggressive prosecution targets primitive laundering methods—cash structuring, money mules, basic layering—while sophisticated actors exploiting derivatives, offshore structures, and complex corporate arrangements face minimal consequences (Farzulla, 2025c). The contrast between vigorous prosecution of money mules (often coerced individuals transferring funds) and the $3.8 million fine imposed on Danske Bank after $200 billion in suspicious transactions illustrates this pattern (Lynch, 2022; Europol, 2021). Sophisticated actors exploit the very financial instruments designed to manage legitimate risk—derivatives, hedging structures, and cross-border investment vehicles—achieving regulatory invisibility through mechanisms that *appear* as prudent risk management (Farzulla, 2025a).

This enforcement asymmetry is not accidental. Current AML frameworks were designed to detect flows that *look like crime* (unusual cash movements) rather than flows that *look like commerce* (sovereign bonds, property purchases, derivative contracts). A system optimized for one category fails systematically against the other (Kang, 2018).

The implication for CBDC design is significant: extending current surveillance approaches to digital currency will replicate existing failures while creating new privacy costs. Comprehensive monitoring will impose friction on legitimate users without meaningfully impacting sophisticated illicit actors who can exploit the same complexity gaps that defeat current AML systems.

1.3 Research Contribution

This paper proposes an alternative CBDC architecture that addresses the fundamental design flaw in current approaches: treating surveillance as a prerequisite for security rather than one possible mechanism among many.

The proposed framework achieves crime detection through mechanism design rather than identity monitoring. Its core innovations include:

1. **Separation of pattern detection from identity:** An AI-driven monitoring system analyzes transaction graph structure for anomalies without accessing participant identities. The system detects *what* is suspicious, not *who* is suspicious.

2. **Transaction-level intervention:** Suspicious activity flags individual transactions rather than accounts or identities. A flagged transaction is paused; the user's broader financial access remains unaffected.

3. **Opt-in deanonymization:** When a transaction is flagged, both parties receive anonymous notification and may choose to provide explanation. Verification releases the transaction; non-response results in transaction cancellation, not investigation.

4. **Abandonment without consequence:** Users may abandon flagged transactions without deanonymization. Funds return to sender. No record is created linking identity to suspicious patterns.

This architecture inverts the burden of proof. Rather than requiring users to demonstrate legitimacy, it requires the system to demonstrate suspicion—and even then, users retain the option to simply walk away. The choice becomes: explain and proceed, or abandon and remain private.

Paradoxically, this framework may be *more* effective at preventing illicit finance than surveillance-based alternatives. Laundering operations that cannot complete transactions are failed laundering operations. Unlike current systems where sophisticated actors "comply their way through" via documentation theater, architectural enforcement creates barriers that cannot be circumvented through legal form.

1.4 Paper Structure

The paper proceeds as follows:

Section 2 reviews existing CBDC proposals and privacy-preserving payment literature, establishing the current state of design thinking and identifying gaps.

Section 3 presents the core architectural framework, detailing the detection layer, intervention layer, resolution layer, and identity firewall.

Section 4 analyzes game-theoretic properties of the abandonment mechanism and demonstrates why it creates effective deterrence without requiring identity compromise.

Section 5 addresses technical implementation, including cryptographic requirements and feasibility analysis.

Section 6 responds to anticipated objections and explores edge cases.

Section 7 proposes governance structures for privacy-preserving CBDC systems.

Section 8 concludes with policy implications and research directions.

2 Literature Review

2.1 CBDC Design Landscape

Central bank interest in digital currencies has accelerated dramatically since 2020. The 2024 BIS survey confirms this trajectory: of 93 central banks surveyed, 91% are now exploring either retail CBDCs, wholesale CBDCs, or both, with wholesale implementations advancing to more mature stages than retail (Bank for International Settlements, 2024). A comprehensive review analyzing 135 research papers published between 2018 and 2025 identifies privacy protection as a persistent design challenge across proposals (Zhang et al., 2025). Major economy proposals reveal convergent design assumptions.

The European Central Bank's digital euro project emphasizes "offline functionality, privacy, and programmability" while maintaining that "appropriate checks" will ensure compliance with AML requirements (European Central Bank, 2020, 2023). Privacy is framed as compatible with regulatory oversight, with technical details delegated to implementation.

The Federal Reserve's discussion papers similarly acknowledge privacy concerns while maintaining that digital dollar design must accommodate "appropriate law enforcement access" (Board of Governors of the Federal Reserve System, 2022). The unstated assumption is that privacy and surveillance exist on a spectrum, with optimal design finding appropriate balance. Notably, event study evidence from cryptocurrency markets suggests that regulatory announcements generate significant market reactions, with infrastructure-level disruptions producing larger and more persistent effects than regulatory uncertainty signals (Farzulla, 2025d). Aggregated systemic risk measurement across cryptocurrency markets further reveals how contagion propagates through interconnected channels (Farzulla and Maksakov, 2025), underscoring the importance of architectural choices in digital currency design for financial stability. The growing interconnection between traditional and decentralized finance creates bidirectional contagion

pathways—what Aufiero et al. (2025) term 'crosstagion'—that CBDC architectures must account for in their stability design.

China's digital yuan (e-CNY) represents the most advanced large-economy implementation. While technical specifications remain partially opaque, available evidence indicates comprehensive transaction visibility with "controllable anonymity"—a formulation that privileges control over anonymity (Auer et al., 2021; Fan, 2020).

The Bank of England's consultation papers explicitly frame privacy as "tiered," with small transactions potentially anonymous but larger transactions subject to full identification (Bank of England, 2023). This approach assumes that surveillance is necessary above some threshold and that the threshold can be set appropriately.

These convergent assumptions have not gone unexamined in the legal and regulatory literature. Soana and Arruda (2024) argue that as financial architecture shifts toward digital forms, the inherited trade-off between privacy and traceability requires renegotiation rather than mechanical extension—existing AML frameworks were designed for a world of intermediated paper trails, not programmable money. Minto (2024) provides a detailed analysis of how the EU's evolving AML regulatory framework (AMLD4/5, AMLR 2024, MiCAR, PSD3) creates specific uncertainties for the digital euro's offline and online modes, demonstrating that the regulatory treatment of CBDC privacy remains contested even within a single jurisdiction. Meanwhile, central bank pilot programmes have begun testing privacy architectures in practice. Project Hamilton, a collaboration between the Federal Reserve Bank of Boston and MIT, demonstrated a high-performance transaction processor capable of 1.7 million transactions per second (Lovejoy et al., 2023)—establishing that CBDC infrastructure can achieve the throughput necessary for national-scale deployment. Project Sela, conducted by the BIS Innovation Hub with the Bank of Israel and the Hong Kong Monetary Authority, explored an "access enabler" model where privacy is preserved through layered data obfuscation while settlement occurs on the central bank balance sheet (BIS Innovation Hub et al., 2023). Project Rosalind, a BIS–Bank of England collaboration, prototyped 33 API functionalities for retail CBDC with personally identifiable information explicitly excluded from the API and central bank ledger layers (BIS Innovation Hub and Bank of England, 2023). These pilots demonstrate growing institutional recognition that privacy and settlement finality need not be in tension—yet none has adopted privacy as a first-order architectural objective in the sense this paper proposes.

Across proposals, privacy appears as a constraint to be managed rather than a design objective to be achieved. Goodell et al. (2025) examine retail CBDC motivations and design choices across major economies, identifying recurring mistakes in implementation—particularly the conflation of technical privacy with policy privacy and the failure to treat architectural privacy as a first-order design objective rather than an afterthought. Their analysis reinforces the observation that the possibility of architectural privacy—systems that *cannot* perform surveillance because they lack the technical capability—receives minimal attention. Whether stated design principles in official CBDC documentation will predict actual implementation outcomes remains an open empirical question; evidence from cryptocurrency markets suggests that whitepaper claims exhibit only moderate alignment with subsequent market behavior patterns (Farzulla, 2025e), cautioning against treating policy commitments as equivalent to architectural guarantees.

2.2 Privacy-Preserving Payment Systems

The theoretical foundation for privacy-preserving digital payments predates current CBDC debates by decades. David Chaum's 1982 work on blind signatures demonstrated that digital payments could be

cryptographically structured to prevent linkage between payer identity and transaction (Chaum, 1982). His eCash system, implemented at DigiCash in the 1990s, proved technical feasibility before commercial failure due to adoption challenges unrelated to privacy architecture (Chaum, 1983).

Subsequent developments in zero-knowledge proof systems substantially expanded privacy-preserving capabilities. Zcash demonstrated that transaction validity could be verified without revealing sender, recipient, or amount through zk-SNARKs (Ben-Sasson et al., 2014). While Zcash addresses different requirements than CBDC (decentralized cryptocurrency vs. central bank liability), its cryptographic primitives are directly applicable.

Academic literature on privacy-preserving CBDC specifically has grown substantially. Gross et al. (2021) propose a "privacy-first" design using blind signatures for low-value transactions. Allen et al. (2020) analyze privacy-utility trade-offs in digital currency design. Garratt et al. (2021) examine how privacy affects monetary policy transmission.

The most directly relevant precedent for the architecture proposed in this paper is Chaum et al. (2021), who present a concrete CBDC design co-authored by David Chaum—whose foundational blind signature work this paper already acknowledges—together with the lead developer of the GNU Taler payment system and Thomas Moser, then a member of the Swiss National Bank Governing Board. Their proposal achieves payer anonymity while preserving payee transparency and regulatory compliance, using Chaum's blind signatures to ensure that the central bank cannot link withdrawals to deposits. The GNU Taler system (Burdges et al., 2016) implements this design principle as a working payment system, and subsequent work has extended it to support zero-knowledge age restriction and selective attribute disclosure without compromising transaction privacy (Kesim et al., 2022). That Chaum himself, in collaboration with a sitting central banker, concluded that privacy-preserving CBDC is both technically feasible and institutionally desirable substantially strengthens the case that surveillance-based architecture reflects policy preference rather than technical necessity.

However, this literature predominantly accepts the surveillance–privacy trade-off framing. Privacy protections are proposed as carve-outs for specific use cases (small transactions, offline payments) rather than as default architecture. Even the Chaum–Grothoff–Moser design, while privacy-preserving at the transaction level, does not address the broader question of whether AML detection can be achieved through mechanism design rather than identity access. The possibility of achieving equivalent security through alternative mechanisms receives limited attention.

Recent central bank research has begun exploring privacy-enhancing technologies more systematically. The Bank of Canada's 2025 analysis examines zero-knowledge proofs, secure multi-party computation, group signatures, and other cryptographic approaches, though it concludes that "techniques to achieve cash-like privacy are immature" with potential hidden vulnerabilities (Bank of Canada, 2025). The BIS Innovation Hub's Project Tourbillon provides practical exploration of privacy-security trade-offs in prototype implementations (BIS Innovation Hub, 2023). Notably, this central bank acknowledgment that privacy-preserving technology remains immature occurs alongside the choice to proceed with surveillance-based architectures—suggesting that immaturity serves as justification for surveillance rather than motivation for further privacy-preserving research.

2.3 AML Effectiveness Literature

The empirical literature on AML effectiveness provides crucial context for CBDC design debates. Pol (2020) characterizes the global AML regime as potentially "the world's least effective policy experiment," noting the extraordinary gap between compliance costs and interdiction rates.

Levi (2020) documents how risk-based approaches, intended to focus resources on genuine threats, instead enable sophisticated actors to present themselves as low-risk while primitive methods trigger automatic scrutiny. The inverted risk hierarchy—sophisticated actors receiving less scrutiny than unsophisticated ones—represents a systematic failure mode rather than implementation error.

Findley et al. (2014) demonstrate through field experiments that incorporation service providers frequently fail to verify beneficial ownership even when legal requirements mandate verification. Compliance is formal rather than substantive.

Sharman (2017) shows how secrecy jurisdictions survive regulatory pressure through adaptive strategies, maintaining opacity while achieving formal compliance with international standards.

The literature suggests that extending current AML approaches to CBDCs will replicate existing failures. Systems designed to detect primitive laundering will remain ineffective against sophisticated actors, while creating new privacy costs for legitimate users. Notably, privacy-preserving alternatives to conventional AML are no longer purely theoretical. van Egmond et al. (2024) report on a deployed MPC-based AML system developed with three Dutch banks—ABN AMRO, Rabobank, and De Volksbank—that enables cross-institutional risk signal propagation without sharing customer data. Their system demonstrates that secure multi-party computation can operationalise collaborative AML detection at institutional scale, providing the strongest empirical evidence to date that privacy and effective financial crime detection are not in fundamental tension.

FATF's 2024 assessment of virtual asset implementation reveals persistent gaps: 75% of assessed jurisdictions remain only partially compliant or non-compliant with virtual asset standards, and over half have not implemented the Travel Rule (Financial Action Task Force, 2024). Theoretical work directly connecting money laundering dynamics to CBDC privacy design complicates the surveillance narrative further. Chiu and Davoodalhosseini (2023) demonstrate through general equilibrium modeling that while CBDC with less anonymity than cash decreases money laundering, high-anonymity CBDC with low interest rates decreases output from *both* laundering and non-laundering agents—suggesting that privacy design involves complex welfare trade-offs rather than simple surveillance optimization.

2.4 Gap in Current Literature

The existing literature contains a significant gap: limited analysis of CBDC architectures that achieve security through mechanism design rather than surveillance. Privacy-preserving proposals focus on protecting specific transaction categories while accepting surveillance for others. AML critiques identify failures without proposing alternative detection approaches. A 2024 systematic review of CBDC privacy literature confirms this gap, finding that existing research predominantly examines one-tier operational models that pose "substantial privacy challenge due to potential mass surveillance" without adequately exploring architectural alternatives (Okonkwo and Adesina, 2024).

This paper addresses the gap by developing an architectural framework where privacy is default and security emerges from structural incentives rather than monitoring capabilities. The contribution lies not in novel cryptographic techniques but in applying existing primitives to create a coherent alternative to surveillance-based design.

3 Architectural Framework

3.1 Design Principles

The proposed architecture rests on four foundational principles:

Principle 1: Separation of pattern detection from identity. Transaction graph analysis can identify structural anomalies—unusual network patterns, velocity anomalies, suspicious counterparty

relationships—without knowing participant identities. The detection system processes anonymized transaction data and outputs suspicion scores for specific transactions, never for users.

Principle 2: Transaction-level intervention. Suspicious activity affects the specific transaction flagged, not the user's broader financial access. A merchant receiving a suspicious payment experiences a delay on that payment; their ability to receive other payments, make payments, or access existing balances remains unaffected.

Principle 3: Opt-in deanonymization. Identity revelation is always voluntary. Users may choose to explain flagged transactions, providing whatever information they consider appropriate. Alternatively, they may abandon the transaction without explanation and without creating any record linking their identity to the suspicious pattern.

Principle 4: Architectural enforcement. Privacy guarantees are structural, not policy. The system cannot reveal identities because it does not possess the capability to link transactions to identities without active user participation. This contrasts with policy promises ("we will only access data with appropriate authorization") that depend on institutional behavior.

3.2 Detection Layer

The Detection Layer performs transaction graph analysis without identity access. Its components include:

Transaction Graph Processor: Ingests anonymized transaction data including amounts, timestamps, and anonymized counterparty tokens. Maintains rolling graph of transaction relationships without ability to deanonymize nodes.

Pattern Matching Engine: Compares transaction structures against known illicit typologies: rapid movement patterns characteristic of layering, network structures associated with mule operations, velocity anomalies suggesting structuring. Pattern libraries are publicly auditable.

Anomaly Detection System: Identifies novel suspicious structures not matching known typologies using machine learning approaches. Training data consists of anonymized historical patterns labeled as legitimate or suspicious by the Resolution Layer (see below), creating a feedback loop that improves detection without requiring identity information.

Suspicion Scorer: Outputs suspicion scores for individual transactions. Scores above threshold trigger intervention. Threshold parameters are publicly disclosed and subject to governance oversight.

Critically, the Detection Layer has no access to identity mapping. It processes transaction IDs and anonymized counterparty tokens. It cannot determine that Transaction X involves User Y. Its output is: "Transaction X has suspicion score Z," never "User Y is suspicious."

3.3 Intervention Layer

The Intervention Layer receives flagged transactions from the Detection Layer and implements intervention protocols.

Transaction Freeze: Flagged transactions are paused pending resolution. Funds remain in escrow. Neither sender nor recipient can access the specific transaction amount until resolution occurs.

Anonymous Notification: Both transaction parties receive notification that the transaction has been flagged. Notifications are anonymous—delivered through the same channel as legitimate transaction confirmations without additional identifying information.

Resolution Options: Parties are presented with options: (1) provide voluntary explanation, (2) request additional time, (3) abandon transaction. No option requires identity disclosure to proceed.

Timeout Protocol: If neither party responds within the timeout period (configurable, e.g., 72 hours),

the transaction is automatically cancelled. Funds return to sender. No record is created linking any identity to the suspicious pattern.

The Intervention Layer, like the Detection Layer, lacks identity access. It processes transaction IDs and can freeze or release transactions, but cannot determine who is involved in any transaction.

3.4 Resolution Layer

The Resolution Layer handles voluntary explanations and makes release decisions.

Explanation Receipt: When parties voluntarily provide explanation, the Resolution Layer receives: (1) the explanation text, (2) any supporting documentation, (3) a cryptographic proof that the explainer is a party to the transaction (without revealing which party or any other identity information).

Human Review: Trained reviewers evaluate explanations against the suspicion indicators that triggered the flag. Reviewers see: explanation content, suspicion score rationale, transaction amount. Reviewers do not see: party identities (unless voluntarily disclosed in explanation), account histories, other transactions.

Decision Output: Reviewers make binary decisions: release (transaction proceeds) or maintain freeze (parties may provide additional explanation or abandon). Release decisions are logged for audit; freeze decisions trigger notification with generic rationale.

Anonymized Feedback: Resolution decisions feed back to the Detection Layer's training system. Transactions released after explanation indicate false positives; abandoned transactions contribute to suspicion pattern learning. All feedback is anonymized.

3.5 Identity Firewall

The Identity Firewall is the architectural component that makes privacy guarantees structural rather than policy-dependent.

Cryptographic Separation: Transaction graph data and identity mapping are held by separate systems with no shared access. The Detection, Intervention, and Resolution Layers operate on the transaction graph. Identity mapping exists only in the Wallet Layer (user-facing application) and is never transmitted to other components.

Zero-Knowledge Proofs: When users voluntarily explain transactions, they prove party status ("I am authorized to speak about Transaction X") without revealing identity ("I am User Y"). Only if users explicitly choose to reveal identity does the Resolution Layer learn it.

No Backdoor Architecture: The system is designed such that state actors cannot compel identity revelation for specific transactions without user cooperation. This is not a policy commitment but a structural property: the components that perform analysis simply do not possess identity information.

Threshold Cryptography for Emergency Access: If governance structures determine that emergency identity access is needed (e.g., court-ordered investigation), a threshold scheme requires multiple independent parties to authorize decryption of identity mapping for specific transactions. This creates accountability for access while making bulk surveillance technically impossible.

4 Game-Theoretic Analysis

Recent theoretical and empirical work supports the viability of privacy-preserving approaches to digital currency design. Agur et al. (2022) show formally that CBDC adoption depends on network externalities and the distribution of privacy preferences across agents—when a critical mass of privacy-sensitive users exists, designs that fail to accommodate their preferences may never achieve the adoption threshold necessary for monetary policy effectiveness. Tinn (2025) develops a formal model of digital currency

with asymmetric privacy, demonstrating that different privacy architectures produce distinct welfare implications that cannot be captured by simple surveillance–security trade-off framings. Empirically, Bijlsma et al. (2024) find through randomized survey experiments involving over 3,500 participants that privacy features increase willingness to use CBDC by up to 60% for privacy-sensitive transactions—suggesting that privacy preservation may enhance rather than undermine adoption and, by extension, monetary policy transmission.

4.1 The Abandonment Mechanism

The abandonment mechanism is central to the framework's deterrent properties. Understanding its game-theoretic structure clarifies why privacy-preserving architecture can achieve security outcomes equivalent or superior to surveillance.

Consider an actor attempting to launder funds through the CBDC system. They initiate a transaction that is flagged by the Detection Layer. They face three options:

1. **Explain and proceed:** Provide explanation that satisfies the Resolution Layer. This requires producing a legitimate narrative for the transaction. For truly illicit transactions, this may be difficult or impossible; for sophisticated launderers, it requires effort and creates potential evidence.

2. **Abandon silently:** Cancel the transaction. Funds return to sender. No identity is linked to the suspicious pattern. The laundering attempt fails, but no evidence is created.

3. **Wait for timeout:** Take no action. Transaction is automatically cancelled after timeout period. Equivalent to option 2 with delay.

4.2 Deterrent Properties

The abandonment option creates a deterrent that does not require identity compromise:

Failed transactions are failed laundering. A laundering operation that cannot move money has failed in its core objective. Unlike current AML systems where sophisticated actors "comply through" by producing documentation, architectural friction creates barriers that cannot be circumvented through legal form.

Volume deterrence. Laundering operations require moving substantial funds. If a significant fraction of transactions are flagged and must be abandoned, the operation becomes economically nonviable regardless of whether any individual is identified.

Pattern signal without prosecution. Abandoned transactions provide valuable data for the Detection Layer. While individual abandonments reveal no identity, aggregate patterns of abandonment indicate detection system accuracy and can improve future detection. The system learns from illicit actors' behavior without needing to identify them.

No "comply through" option. Current AML systems can be defeated by producing appropriate documentation—establishing legitimate-appearing corporate structures, maintaining plausible transaction narratives, etc. The abandonment mechanism eliminates this strategy: either the transaction can be legitimately explained (in which case it may actually be legitimate) or it cannot proceed.

4.3 Legitimate User Experience

For legitimate users, the framework creates minimal friction:

Low flag rates. Detection systems tuned for high-confidence anomalies will flag a small fraction of legitimate transactions. The false positive rate is a tunable parameter subject to governance oversight.

Easy resolution. Legitimate transactions have legitimate explanations. A flagged payment for a large purchase can be explained by providing purchase documentation. A flagged transfer to a new recipient can be explained as "I'm sending money to my friend." The explanation need not be formally verified; it need only be plausible.

Abandonment as protection. If a legitimate user prefers not to explain a transaction (e.g., for personal privacy reasons), abandonment imposes no penalty beyond transaction failure. This may be preferable to users who value privacy over transaction completion.

No account-level consequences. A flagged transaction does not affect the user's broader financial access. Their account is not frozen, investigated, or marked for enhanced scrutiny.

4.4 Comparative Advantage Over Surveillance

The framework offers advantages over surveillance-based approaches:

No false positive harm. In surveillance systems, false positives can result in account freezes, investigation, and reputational damage. The framework's worst case for false positives is transaction delay and possible abandonment—annoying but not harmful.

No mission creep. Surveillance infrastructure created for AML purposes can be repurposed for political surveillance, commercial exploitation, or other uses beyond original intent. Architectural privacy prevents mission creep by eliminating the capability.

No data breach risk. Comprehensive transaction data linked to identities creates valuable targets for theft. Data that doesn't exist cannot be stolen.

Superior detection of sophisticated laundering. Current AML systems detect primitive laundering (unusual cash patterns) while missing sophisticated laundering (complex corporate structures, derivatives). The proposed system's pattern detection focuses on transaction graph structure regardless of legal form, potentially detecting sophisticated schemes that evade current monitoring.

4.5 Empirical Validation

Agent-based simulation comparing graph-only (privacy-preserving) and identity-aware (surveillance-based) detection validates the framework's effectiveness claims while revealing important nuances about the privacy–security trade-off.

Methodology. We simulate a transaction network with 500 legitimate identities and 50 launderer identities, each controlling 1–2 and 3–6 wallets respectively. Launderers attempt to blend in by mimicking legitimate transaction patterns with minor structural differences (10% structuring behavior, slightly elevated volumes). Detection models use random forest classifiers trained on graph topology features (degree centrality, PageRank, clustering coefficient, betweenness, transaction volume and flow metrics) for graph-only detection, with additional identity-linked features (wallet count per identity, aggregate cross-wallet volume, watchlist status, prior suspicious activity reports) for identity-aware detection.

On Testing "Primitive" Laundering. A potential objection holds that testing against unsophisticated laundering patterns weakens the findings. We argue the opposite: primitive laundering represents the *strongest possible test* for surveillance value. If surveillance infrastructure cannot improve detection of the simple patterns it was designed to catch—cash structuring, velocity anomalies, known bad actor flagging—it certainly cannot improve detection of sophisticated laundering that exploits derivatives, shell companies, and complex corporate structures (Farzulla, 2025a). Testing against easy cases and finding surveillance adds nothing is more damning than testing against hard cases.

Baseline Performance. Graph-only detection achieves 0.892 AUC at wallet level compared to identity-aware's 1.000 AUC. At identity level (aggregating wallet predictions), graph-only achieves

0.981 AUC—a gap of less than 2%. The most important features for identity-aware detection are `num_wallets` (0.34 importance) and aggregate volume metrics (0.52 combined)—all achievable through pseudonymous linking without identity revelation.

Watchlist Impact. A follow-up experiment addresses the potential critique that random watchlist assignment artificially diminishes watchlist value. We generate correlated watchlists where 80% of launderers appear on the watchlist (realistic for known bad actors) while only 5% of legitimate users appear (false positives). This achieves watchlist-label correlation of 0.89—a highly informative feature.

Results reveal a crucial asymmetry: watchlists substantially improve graph-only detection (0.87 $\rightarrow$ 0.99 AUC), but provide *zero* marginal improvement over pseudonymous identity linking (1.00 $\rightarrow$ 1.00 AUC). The surveillance infrastructure adds value only when cross-wallet linking is unavailable. When linking is present—achievable through privacy-preserving techniques—watchlists become redundant.

Evasion Sensitivity. Graph-only detection degrades from 0.949 to 0.725 AUC as launderers spread activity across 1–15 wallets (sybil attack). Identity-aware detection remains robust due to cross-wallet aggregation. However, in KYC-constrained CBDC environments where creating multiple wallets requires identity verification, sybil attacks become costly. When launderers are limited to 1–2 wallets (same as legitimate users), graph-only achieves 94.9% AUC—a gap of under 2%.

Implications. These findings reframe the privacy–security trade-off. The value of identity-aware detection comes entirely from knowing "these wallets belong to the same entity"—achievable through zero-knowledge constructions such as Zether (Bünz et al., 2020) or composable SNARKs (Campanelli et al., 2019)—rather than "this entity is John Smith on a sanctions watchlist." Pseudonymous linking captures the detection benefit without the privacy cost. Given that 2.96 million UK citizens recently petitioned against digital identity infrastructure (Statewatch et al., 2025) and 41% of ECB CBDC consultation responses focused on privacy concerns (European Central Bank, 2021), any marginal improvement from full surveillance is not worth the catastrophic loss of public trust.

5 Technical Implementation

5.1 Cryptographic Requirements

The proposed architecture relies on established cryptographic primitives:

Zero-Knowledge Proofs: Used for transaction validation (proving transaction validity without revealing details) and party authentication (proving party status without revealing identity). zk-SNARKs or zk-STARKs provide efficient implementations. The Zcash implementation demonstrates production viability of similar requirements (Ben-Sasson et al., 2014).

Secure Multi-Party Computation: Enables pattern detection across transaction data without any single party accessing complete data. MPC protocols allow the Detection Layer to compute aggregate statistics and pattern matches without any component seeing raw transaction details.

Threshold Cryptography: Emergency access to identity mapping (when authorized by governance processes) requires threshold signatures from multiple independent parties. This prevents any single actor from accessing identity data while preserving emergency capability.

Blind Signatures: Enable transaction authorization without linking authorization to identity. Central bank signature on transaction tokens proves validity without revealing transaction details to the central bank. Chaum et al. (2021) demonstrate a complete CBDC issuance protocol using blind signatures that ensures the central bank cannot link withdrawals to subsequent spending, providing a concrete instantiation of this primitive in a central banking context.

Ring Signatures: Allow parties to prove membership in transaction sets without revealing which

specific party they are. Useful for anonymous explanation submission in the Resolution Layer.

5.2 Feasibility Assessment

Each required primitive has been demonstrated in production systems:

- Zero-knowledge proofs: Zcash processes approximately 500,000 shielded transactions per month (Electric Coin Company, 2023).
- Secure multi-party computation: Deployed in privacy-preserving analytics systems including Google's Private Join and Compute (Google, 2019).
- Threshold cryptography: Used in production cryptocurrency custody solutions and distributed key management (Gennaro and Goldfeder, 2018).
- Blind signatures: Originally proposed in 1982; implemented in multiple privacy-preserving systems (Chaum, 1982).

Performance remains a consideration. Current ZKP systems require significant computational overhead compared to transparent transactions. However, proof generation can occur client-side (user devices), distributing computational load. Verification is substantially faster than generation. Project Hamilton, the Federal Reserve Bank of Boston and MIT collaboration, demonstrated that a CBDC transaction processor can achieve 1.7 million transactions per second with sub-second latency (Lovejoy et al., 2023)—establishing that the underlying infrastructure can accommodate the additional computational overhead of privacy-preserving primitives without falling below the throughput thresholds required for national-scale deployment. For central bank deployment with adequate infrastructure, performance is achievable.

A parallel ZKP-based compliance framework merits direct comparison. Decker (2025) propose a zero-knowledge proof architecture for institutional KYC that reduces exposed user data by 97% while achieving 96.7% fraud detection accuracy through AI-enhanced verification. Their detection rate exceeds our framework's 87–95% range, but operates in a narrower context: point-of-entry KYC verification for institutional compliance. Our architecture addresses a broader design space—transaction-level intervention, anonymized pattern detection, and opt-in deanonymization—that extends privacy preservation beyond the initial onboarding stage to the entire transaction lifecycle. The two approaches are complementary: Decker's framework could serve as the identity verification layer at account creation, while our architecture governs ongoing transaction monitoring without identity exposure.

Recent implementations demonstrate continued progress toward practical viability. Seres et al. (2024) present a Secure Element-based system for offline CBDC transactions that achieves latency comparable to commercial payment systems while maintaining regulatory compliance capabilities. Advances in secure multi-party computation continue to reduce communication overhead, with recent protocols achieving linear communication complexity for threshold corruption scenarios that previously required super-linear scaling (Escudero et al., 2024). A comprehensive review of zero-knowledge proof developments notes that while challenges remain, "every major L1 and L2" blockchain platform is now integrating ZKP infrastructure, creating a substantial engineering base for privacy-preserving financial systems (R et al., 2025).

5.3 System Architecture

A complete system would include:

Wallet Layer: User-facing application managing identity, generating proofs, and interfacing with transaction layers. This is the only layer with access to user identity.

Transaction Layer: Processes anonymized transactions, maintains ledger, handles settlement. No identity access.

Detection Layer: Analyzes transaction patterns for anomalies. No identity access.

Intervention Layer: Manages flagged transactions and resolution process. No identity access except voluntary disclosure.

Governance Layer: Manages system parameters, threshold key shares, audit processes. Oversight without operational access.

Each layer operates independently with defined interfaces. Compromise of any single layer does not enable identity compromise without user cooperation or threshold key activation.

5.4 A Concrete PET AML Stack: PSI + Secure Risk Propagation + ZK Policy Proofs

This section instantiates the paper's high-level claim into a concrete "privacy-enhancing AML (PET–AML) stack" that can be implemented today: (i) *private watchlist checking* using Private Set Intersection (PSI) and/or VOPRF-based membership tests, (ii) *secure risk propagation* over pseudonymous transaction graphs using secure computation, and (iii) *transaction-time policy proofs* in zero-knowledge that make compliance verifiable *without* identity disclosure. These components align with the paper's layered architecture: the Wallet Layer performs screening and proof generation, the Transaction Layer verifies proofs, the Detection Layer computes risk on pseudonyms, and the Governance Layer controls exceptional identity exposure.

5.4.1 Threat model and trust assumptions

Parties. We assume (a) a *wallet / PSP* that performs KYC and issues an unlinkable credential to the user, (b) a *ledger operator* (central bank or authorized operator) that validates transaction proofs and maintains the settlement ledger, (c) a *compliance authority* (FIU/sanctions unit) that maintains watchlists and publishes policy parameters, and (d) a *governance quorum* (e.g., ombuds + privacy regulator + judiciary delegate) that holds threshold key shares for exceptional identity exposure.

Adversaries. We target (i) honest-but-curious infrastructure operators, (ii) malicious users attempting to launder funds while remaining unlinkable, and (iii) partial compromise/collusion of operational entities. We avoid a single "super-admin" trust anchor: identity access requires threshold activation and jurisdiction-dependent legal authorization.

Security objectives.

- *Transaction privacy:* the ledger validates payments and enforces limits without learning real-world identity; routine analytics operate on pseudonyms.
- *Watchlist privacy:* the compliance authority learns nothing about non-matching users/transactions; the ledger learns nothing about the watchlist and does not learn who is being screened.
- *Enforceability:* users cannot create valid transactions without satisfying policy constraints; false negatives are bounded by watchlist correctness and credential issuance integrity.
- *Due process:* identity disclosure is auditable, rate-limited, and threshold-gated; there is no bulk de-anonymization API.

5.4.2 Component design and interfaces

Table 1 summarizes the three stack elements and how they interact.

Table 1: PET–AML stack components and outputs.

Component	Primitive	Output to ledger / detection layer	Primary privacy guarantee
Private watchlist check	VOPRF-based membership / PSI (unbalanced PSI variants support large watchlists) (Wood et al., 2023; Wang et al., 2025)	match bit and (optionally) a signed "screening witness" bound to an epoch	Server learns nothing about queries; non-matching clients learn only "no-match"
Secure risk propagation	Secure computation over transaction graph features (HE/SS hybrid; MPC graph frameworks) (Koti et al., 2024; Yu et al., 2025)	risk scores / bands on pseudonyms	Watchlist-seeded scoring without revealing which nodes were seeded
ZK policy proofs	zk-SNARK/zk-STARK/range proofs for policy constraints (Groth, 2016; Gabizon et al., 2019; Ben-Sasson et al., 2018; El-Hajj and Roelink, 2024)	π_{policy} attached to transaction; verified at admission	Ledger verifies compliance without learning identity or sensitive attributes

(1) Private watchlist checking (PSI/VOPRF). We implement sanctions/PEP screening as a *privacy-preserving membership test*: the wallet (or PSP, depending on the threat model) derives a stable identifier x from KYC material (e.g., a salted hash of canonicalized identity attributes) and runs a PSI-style protocol against the watchlist set W. Modern unbalanced PSI is explicitly designed for the AML regime, where the server holds a large set while each client query set is small (Wang et al., 2025). A complementary deployment is to use standardized verifiable OPRFs (VOPRFs) (Wood et al., 2023): the compliance authority evaluates an oblivious PRF on x, and the wallet locally checks membership against a periodically published, PRF-transformed watchlist.

Interface. At epoch e, the wallet obtains a screening token τ_e and computes $m_e \in \{0,1\}$ locally (match/no-match). When $m_e = 0$, the wallet can include a compact witness (e.g., a signature from the compliance authority over a commitment to x and e) in the transaction, proving that screening occurred *without* revealing x. When $m_e = 1$, policy triggers a dispute window in the Intervention Layer rather than automatic identity leakage. This preserves due process and prevents "watchlist leakage" to the ledger operator.

(2) Secure risk propagation on pseudonymous graphs. The Detection Layer benefits from graph-structured signals (fan-in/fan-out motifs, bursty flows, multi-hop proximity to known bad actors) but should not learn identities. We therefore treat identity as a *sealed label* and compute risk on pseudonyms. The privacy challenge is that even *which* pseudonyms are seeded as "known bad" can be sensitive.

This component draws direct validation from the deployed MPC-based AML system reported by van Egmond et al. (2024), where three Dutch banks implemented secure risk signal propagation across institutional boundaries without sharing customer identities. Their system demonstrates that the core operation—computing risk scores over a multi-party graph without revealing which nodes triggered the computation—is practically feasible at banking scale, not merely a theoretical construction.

We propose a two-tier design:

1. **Cleartext structure, secret seeds.** The transaction graph structure and non-identifying features can remain available to the Detection Layer. Only the seed vector (e.g., watchlist-hit pseudonyms, typology flags) is secret-shared/encrypted. Secure computation then evaluates a risk function $r \leftarrow f(G,s)$ without revealing s.

2. **Secure multi-domain propagation.** For cross-PSP or cross-bank graphs, use secure graph analytics frameworks that reduce edge-proportional communication overhead. Recent work demonstrates large improvements in secure iterative graph analysis, scaling to million-scale graphs under semi-honest models (Koti et al., 2024; Yu et al., 2025).

Risk function. A practical choice is a damped k-step diffusion (personalized PageRank-style) or motif-based scoring with bounded iterations. This keeps compute predictable, reduces the attack surface of arbitrary model execution, and makes performance auditing tractable.

(3) ZK policy proofs at transaction admission. Transaction-time ZK proofs prevent a large class of laundering strategies by making policy constraints *verifiable* on entry to the ledger:

- *Balance correctness:* prove sufficient funds and no double-spend.
- *Tiered limits:* prove amount and cumulative spend remain within per-tier caps (daily/weekly) without revealing exact history (range and sum proofs).
- *Credential validity:* prove possession of a valid, unrevoked KYC credential with selective disclosure (e.g., prove "resident" or "over-18" without revealing full identity), optionally using threshold-issued credentials such as Coconut (Sonnino et al., 2019).
- *Screening performed:* prove inclusion of a valid screening witness for epoch e, binding the transaction to a recent watchlist check.

Empirically, zk-SNARKs offer very small proof sizes and fast verification (at the cost of trusted-setup assumptions), while zk-STARKs are transparent and post-quantum secure but typically incur larger proofs (El-Hajj and Roelink, 2024). The design is modular: a jurisdiction can select a proof system on the "succinctness vs. transparency" frontier and upgrade over time as engineering capacity improves.

5.4.3 Performance envelope and deployment profile

The stack is meant to be deployable under realistic constraints. The following envelope is indicative, not normative; concrete numbers depend on implementation details, hardware, and policy complexity.

Table 2: Indicative performance envelope for the PET–AML stack (order-of-magnitude).

Operation	**Latency target**	**Bandwidth / storage**	**Notes**
Watchlist check (wallet $\rightarrow$ FIU)	sub-second interactive; amortizable via epochs	$\mathcal{O}$(KB) per check; server stores $\lvert W\rvert$ items	Unbalanced PSI supports large $\lvert W\rvert$ with modest client cost (Wang et al., 2025)
Transaction ZK proof generation	10–500 ms (desktop); 100 ms–few s (mobile)	proof size: $\mathcal{O}(10^2)$ bytes (SNARK) to $\mathcal{O}(10^4)$ bytes (STARK)	Benchmarks vary with circuit; verification typically milliseconds (El-Hajj and Roelink, 2024)
Secure risk propagation (batch)	seconds–minutes per batch (e.g., hourly/daily)	state proportional to $\lvert V\rvert$; preprocessing amortizes multiplications	Secure graph frameworks report million-scale feasibility and large comms reductions (Koti et al., 2024; Yu et al., 2025)

Operational profile. To keep the user experience smooth, we recommend (a) performing PSI screening at wallet "ready" time (onboarding and periodic re-screening), (b) generating ZK proofs at transaction time using fixed, audited circuits with bounded complexity, and (c) running secure risk propagation in batches unless a high-risk typology requires near-real-time scoring. The resulting system preserves routine payment performance while retaining a credible privacy posture and an auditable, threshold-gated pathway for exceptional disclosure.

5.4.4 Simulation validation

To validate the performance envelope described in Table 2, we implement a discrete-event simulation of the PET–AML stack (`pet_aml_sim.py`).[1] The simulator models 1,000 wallets distributed across four risk tiers (0–3), generates transactions with lognormal amount distributions, and exercises all three stack components: PSI-based watchlist screening, ZK policy proof verification, and batch MPC risk propagation.

Table 3 summarises the key metrics from a representative run.

[1] Simulation source code is available in the paper's repository at `https://github.com/andrewmaksakov/CBDC`.

Table 3: PET–AML simulation results (40,000 transactions generated).

Metric	Value
Transactions generated	40,000
Transactions settled	25,287 (63.2%)
Transactions rejected	14,713 (36.8%)
Sanctions hits (PSI blocked)	82
Escalations	0
Latency mean	537 ms
Latency p50	511 ms
Latency p90	768 ms
Latency p99	1,137 ms
Latency breakdown	PSI screening dominates
MPC batch runtime	0.38 hours
Risk-tier distribution	Tier 0: 40%, Tier 1: 30%, Tier 2: 20%, Tier 3: 10%

Interpretation. The results validate the performance claims in Table 2: PSI watchlist checks complete in sub-second time, ZK proof verification adds only milliseconds, and batch MPC risk propagation is feasible within a sub-hour cycle. End-to-end transaction latency is dominated by the PSI screening step, with mean latency of 537 ms and 99th-percentile latency of 1.14 s—well within the performance envelope for retail CBDC payments.

The 36.8% rejection rate reflects the simulation's tier-0 daily spending limits interacting with the lognormal transaction amount distribution, not a system failure mode. Tier-0 wallets (40% of the population) have the lowest spending caps, and the heavy-tailed amount distribution frequently exceeds these limits. In a deployed system, tier distribution would reflect actual KYC levels, and limits would be calibrated to real-world spending patterns, yielding substantially lower rejection rates for compliant users.

The 82 sanctions-blocked transactions confirm that PSI screening correctly intercepts watchlist-matched wallets without revealing match status to the ledger operator. Zero escalations indicate that, under normal operating conditions, the privacy-preserving stack handles compliance enforcement entirely through architectural mechanisms without requiring identity disclosure.

6 Addressing Objections

6.1 "Criminals Will Just Abandon Transactions"

Correct. Criminals will abandon flagged transactions. This is the intended outcome.

A laundering operation that cannot complete transactions has failed. The goal of AML is not prosecution of launderers but prevention of laundering. If the system prevents money movement, it has succeeded regardless of whether anyone is identified.

Additionally, abandoned transactions create valuable signal. Patterns of abandonment indicate which transaction types are associated with illicit activity. The Detection Layer learns from abandonments

without requiring identity information.

The objection assumes that prosecution is the primary goal. If the goal is prevention, abandonment is success.

6.2 "This Prevents Investigation of Serious Crime"

The framework does not prevent investigation; it prevents *mass surveillance*. Traditional investigative methods remain available:

- Subpoenas can compel individuals to produce transaction records.
- Court orders can require wallet providers to produce identity information for specific users.
- Undercover operations can infiltrate criminal networks.
- Threshold key activation can enable identity access for specific transactions when authorized by governance processes.

What the framework prevents is *bulk access*—the ability to retrospectively search all transactions for patterns associated with individuals. This capability is not necessary for targeted investigation and creates surveillance risks that extend far beyond crime prevention.

The distinction between government and commercial surveillance matters here. Rennie and Steele (2021) identify four distinct privacy "losses" that CBDC designs can impose—anonymity, liberty, individual control, and regulatory control—arguing that each requires separate analysis rather than aggregation into a single "privacy" metric. As Tucker (2023) notes, referenced in the IMF's 2024 policy framework on CBDC data use, commercial entities cannot confiscate property or incarcerate individuals—state surveillance capabilities differ qualitatively from private data collection (International Monetary Fund, 2024). Legal scholars have argued that CBDC "should provide at least the same privacy-preserving features as cash" with law enforcement access "strictly controlled by due process" (Golumbia et al., 2024). The proposed framework implements this principle architecturally rather than through policy commitment.

6.3 "States Will Never Adopt This"

This objection was more persuasive before July 2025. The U.S. Anti-CBDC Surveillance State Act (H.R. 1919 / S.1124, 119th Congress), which passed the House of Representatives in 2025, explicitly prohibits the Federal Reserve from issuing a CBDC that could be "used as a tool for surveilling Americans" while protecting "open, permissionless, and private" digital currencies. Zatti (2025) analyses this legislation as marking a paradigm shift from enabling to prohibiting CBDC—what he terms a transition from "legal desirability" frameworks that evaluate how CBDCs should be designed to outright prohibition frameworks that reject surveillance architectures categorically. The U.S. legislative position does not endorse the privacy-preserving architecture proposed here, but it does demonstrate that state opposition to surveillance CBDC has moved from civil society concern to enacted legislation. The political space for privacy-preserving alternatives is wider than the objection assumes.

More broadly, the paper is normative (what CBDCs *should* be) rather than predictive (what CBDCs *will* be). The purpose is to establish that privacy-preserving alternatives exist, foreclosing the argument that surveillance is technically necessary. When policymakers claim that CBDC surveillance is required for security, the existence of viable alternatives reveals this as a policy choice rather than technical constraint.

Additionally, state preferences are not monolithic. Smaller jurisdictions seeking to attract digital finance activity may find privacy-preserving architecture competitively advantageous. Civil society pressure and democratic accountability may influence design choices in some contexts.

6.4 "High-Frequency Transactions Cannot Be Individually Reviewed"

Correct. The framework assumes that most transactions proceed without intervention. Only flagged transactions require attention.

For algorithmic trading or high-frequency payment systems, detection thresholds can be calibrated to flag only extreme anomalies. Alternatively, institutional accounts can be subject to different regimes (accepting identity registration in exchange for reduced friction) without compromising retail privacy.

The framework's scalability depends on flag rates. If 0.1% of transactions are flagged, a system processing 1 million transactions per day generates 1,000 flags—manageable with automated triage and human review for complex cases.

6.5 "Graph-Only Detection Is Measurably Worse"

Simulation evidence confirms a 5–13% AUC gap between graph-only and identity-aware detection under realistic conditions. This objection deserves direct engagement rather than dismissal.

However, the comparison obscures the relevant policy question. First, the gap depends entirely on launderers' ability to spread activity across multiple wallets. In KYC-constrained CBDC environments—the relevant deployment context—sybil attacks are costly. When launderers are limited to 1–2 wallets (same as legitimate users), graph-only achieves 94.9% AUC, reducing the gap to under 2%.

Second, watchlist access—the core surveillance capability that justifies identity-linked monitoring—provides zero detection improvement. The identity-aware advantage comes entirely from cross-wallet linking, not surveillance databases. This is achievable through pseudonymous techniques.

Third, the comparison ignores countervailing costs. Public sentiment data reveals profound opposition to financial surveillance: 2.96 million UK citizens petitioned against digital identity infrastructure (the fourth-largest petition in UK parliamentary history) (Statewatch et al., 2025); 41% of ECB CBDC consultation responses focused on privacy concerns (European Central Bank, 2021); 73% of surveyed public express concern about government control over fund access (Chen et al., 2025). Support for digital identity drops from 57% to 38% after respondents learn about privacy implications (Ipsos UK, 2025).

The question is not "which system performs better in isolation?" but "is 5–13% improved detection worth mass financial surveillance given available alternatives?" When pseudonymous linking captures nearly all the detection benefit, and when public trust is essential for CBDC adoption, the answer is clearly no.

7 Governance Framework

7.1 Parameter Oversight

Privacy-preserving architecture does not mean ungoverned architecture. Key parameters require ongoing governance:

Detection thresholds: How sensitive should anomaly detection be? Higher sensitivity catches more illicit activity but generates more false positives. This is a policy decision requiring democratic input.

Timeout periods: How long should parties have to respond to flags? Longer periods accommodate legitimate users; shorter periods reduce friction for illicit abandonment strategies.

Pattern libraries: Which transaction patterns trigger flags? Pattern definitions should be publicly auditable to enable scrutiny and prevent discriminatory targeting.

Threshold key holders: Who holds shares of emergency access keys? Distribution across independent institutions prevents capture while maintaining emergency capability.

7.2 Audit and Transparency

Detection statistics: Aggregate statistics on flag rates, resolution outcomes, and abandonment patterns should be publicly published. This enables assessment of system effectiveness without compromising individual privacy.

Resolution review: Independent auditors should periodically review Resolution Layer decisions to assess consistency and detect potential abuse.

Code audit: System source code should be publicly available for security review. Cryptographic implementations should be formally verified where possible.

Threshold key usage: Any activation of emergency identity access should be publicly logged with justification. Unauthorized access attempts should trigger alerts and investigation.

7.3 Democratic Accountability

Privacy-preserving CBDC represents a fundamental choice about the relationship between citizens and state in financial matters. Keister and Sanches (2023) demonstrate through general equilibrium analysis that CBDC design choices have non-trivial consequences for banking sector stability, deposit funding, and credit allocation—the architectural properties of digital currency are not merely technical details but determinants of macroeconomic structure. When design choices carry consequences of this magnitude, they should be made through democratic processes, not technical default. The framework aligns with stakes-weighted consent models of political legitimacy, where governance structures must demonstrate alignment with stakeholder preferences proportional to the stakes involved (Farzulla, 2025b). CBDC design imposes differential stakes across populations—financial privacy affects marginalized communities, political dissidents, and economically vulnerable individuals more acutely than it affects those with institutional access—and governance structures should reflect this heterogeneity.

The EU's evolving regulatory landscape illustrates both the complexity and the feasibility of legislating privacy requirements for digital currency. Minto (2024) documents how the digital euro's proposed offline mode creates particular challenges for AML compliance under the Anti-Money Laundering Regulation (AMLR 2024), MiCAR, and PSD3—challenges that could be substantially simplified by architectural privacy guarantees rather than layered regulatory exemptions. Governance structures should include:

- Legislative authorization specifying privacy requirements
- Independent oversight body with authority to audit compliance
- Public reporting on system operation and governance decisions
- Mechanisms for citizen input on parameter adjustments
- Sunset clauses requiring periodic reauthorization

8 Conclusion

8.1 Summary of Contributions

This paper has argued that the surveillance assumption underlying current CBDC proposals is a policy choice rather than technical necessity. Privacy-preserving architecture is feasible using established cryptographic primitives and can achieve equivalent or superior crime prevention through mechanism design rather than identity monitoring.

The proposed framework contributes three innovations to CBDC design discourse:

Separation of pattern from identity: Demonstrating that transaction anomaly detection does not require identity access, only transaction graph analysis.

Abandonment as deterrent: Showing that allowing users to abandon suspicious transactions without consequence creates effective crime prevention by making illicit transactions impossible to complete, without requiring identity compromise.

Architectural enforcement: Distinguishing between policy privacy ("we promise not to surveil") and structural privacy ("we cannot surveil because the capability does not exist"), and demonstrating how to achieve the latter.

8.2 Policy Implications

The existence of privacy-preserving alternatives has immediate policy implications:

Claims that CBDC surveillance is necessary for security should be challenged. Alternative architectures exist. The question is not whether privacy is technically possible but whether it is politically chosen.

CBDC design processes should explicitly consider privacy-preserving options. Public consultations that present only surveillance-based designs foreclose democratic choice on fundamental questions.

International standards (FATF, BIS) should accommodate privacy-preserving architectures. Current guidance assumes identity-linked monitoring; updated guidance should recognize alternative mechanisms.

8.3 Limitations and Future Work

This paper has focused on conceptual architecture rather than complete technical specification. Implementation details—specific protocol designs, performance benchmarks, user interface considerations—require further development.

The framework's core assumption—that pattern detection can effectively identify illicit transactions without identity information—has been validated through agent-based simulation (Section 4.5). Results confirm that graph-only approaches achieve 87–95% of surveillance-based detection performance, with the gap driven primarily by cross-wallet activity dilution rather than lack of surveillance data. More significantly, watchlist access provides zero marginal improvement under the conditions tested, suggesting that the surveillance apparatus justifying identity-linked monitoring adds limited detection value beyond what pseudonymous linking achieves. Simulation code and detailed methodology are available at the paper's repository.

Several limitations merit acknowledgment. First, the simulation models relatively unsophisticated laundering behavior; real-world sophisticated actors may exhibit different evasion patterns. Second, the 500/50 legitimate/launderer ratio, while reasonable for simulation, may not reflect actual prevalence rates. Third, graph topology features depend on network structure that may differ in deployed CBDC systems.

8.4 Implementation Challenges

Translating the proposed architecture into deployed systems raises technical challenges that future work must address.

Private Graph Computation. The Detection Layer requires computing graph features (degree centrality, PageRank, clustering coefficients) over transaction data without exposing individual transactions. Secure multi-party computation and homomorphic encryption can theoretically enable such computations, but practical implementations at CBDC throughput (potentially millions of transactions per second) remain an open research problem. Trusted execution environments offer a pragmatic intermediate solution, though with weaker security guarantees. The architecture does not depend on any specific implementation; rather, it specifies *what* must be computed privately, leaving *how* to evolving cryptographic engineering.

The Sybil-Privacy Tradeoff. The architecture's robustness to sybil attacks (launderers creating many wallets) depends on limiting wallet creation. Two design points exist: (1) KYC-constrained systems where wallet creation requires identity verification, achieving strong sybil resistance but introducing identity touchpoints; or (2) resource-constrained systems using proof-of-work, staking, or rate-limiting for wallet creation, preserving stronger privacy but with weaker sybil guarantees. This paper has analyzed both scenarios (Section 4.5) but a deployed system must choose a position on this tradeoff. The key insight—that surveillance infrastructure adds minimal value *given* sybil constraints—holds across both design points.

Pseudonymous Linking Protocols. Cross-wallet linking without identity revelation requires users to prove "I control wallets A, B, and C" without revealing which user makes the claim. Ring signatures, accumulator-based proofs, and composable SNARKs (Campanelli et al., 2019) offer cryptographic building blocks, but constructing a complete protocol that resists correlation attacks over time while enabling aggregate risk assessment requires careful design. The emerging literature on privacy-preserving compliance (Bünz et al., 2020) and confidential transactions provides foundations, but CBDC-specific instantiation remains future work.

Relation to Privacy-Enhancing Technologies for AML. A mature literature explores privacy-set intersection for watchlist matching, homomorphic encryption for encrypted flow tracing, and federated learning for collaborative detection without data sharing. These techniques are largely complementary to the proposed architecture: they address *how* to implement privacy-preserving detection, while this paper addresses *whether* such detection can achieve adequate effectiveness. Future work should integrate specific PET protocols into the architectural framework and benchmark performance against traditional approaches.

Interoperability with existing financial infrastructure and cross-border coordination mechanisms present additional challenges not fully addressed here. Future work should examine how privacy-preserving CBDC systems interact with traditional banking AML requirements and international regulatory frameworks.

8.5 Concluding Remarks

CBDCs will reshape monetary infrastructure for generations. The design decisions made now will determine whether digital currency enables financial freedom or financial control.

The false dichotomy between privacy and security has obscured the range of available choices. Privacy-preserving CBDC architecture is technically feasible and potentially more effective at crime prevention than surveillance-based alternatives. The choice to build surveillance infrastructure is a choice, not a necessity.

This paper has attempted to expand the design space by demonstrating that alternatives exist. The political choice remains with democratic societies—but it should be recognized as a choice, made deliberately and accountably, rather than accepted as technical inevitability.

Declarations

Conflict of Interest. The authors declare no competing interests.

Funding. This research received no external funding.

Data Availability. Simulation code for the PET–AML stack validation is available at https://github.com/andrewmaksakov/CBDC.

AI Assistance. Claude (Anthropic) was used as a research collaborator for analytical framework development, literature synthesis, LaTeX preparation, and iterative refinement. All intellectual claims and errors remain the authors' responsibility.

References

Itai Agur, Anil Ari, and Giovanni Dell'Ariccia. Designing central bank digital currencies. *Journal of Monetary Economics*, 125:62–79, 2022. doi: 10.1016/j.jmoneco.2021.05.002. Canonical formal model of CBDC design with heterogeneous privacy preferences.

Sarah Allen, Srdjan Capkun, Ittay Eyal, Giulia Fanti, Bryan Ford, James Grimmelmann, Ari Juels, Kari Kostiainen, Sarah Meiklejohn, Andrew Miller, Eswar Prasad, Karl Wust, and Fan Zhang. Design choices for central bank digital currency: Policy and technical considerations. *NBER Working Paper Series*, 27634, 2020.

Raphael Auer and Rainer Böhme. The technology of retail central bank digital currency. *BIS Quarterly Review*, pages 85–100, March 2020. URL https://www.bis.org/publ/qtrpdf/r_qt2003j.htm. Foundational CBDC taxonomy: direct, indirect, and hybrid architectures.

Raphael Auer, Giulio Cornelli, and Jon Frost. Rise of the central bank digital currencies: drivers, approaches and technologies. *BIS Working Papers*, 880, 2021.

Raphael Auer, Jon Frost, Leonardo Gambacorta, Cyril Monnet, Tara Rice, and Hyun Song Shin. Central bank digital currencies: Motives, economic implications, and the research frontier. *Annual Review of Economics*, 14:697–721, 2022. doi: 10.1146/annurev-economics-051420-020324. Comprehensive BIS review establishing the CBDC research frontier.

Sabrina Aufiero, Silvia Bartolucci, Fabio Caccioli, and Pierpaolo Vivo. Mapping microscopic and systemic risks in TradFi and DeFi: A literature review, 2025.

Bank for International Settlements. Central bank digital currencies: foundational principles and core features. Report no. 1, Bank for International Settlements, Basel, 2020. Joint report with Board of Governors of the Federal Reserve System, ECB, Bank of Japan, Sveriges Riksbank, Swiss National Bank, Bank of England, Bank of Canada.

Bank for International Settlements. Advancing in tandem – results of the 2024 BIS survey on central bank digital currencies and crypto. Bis papers no. 159, Bank for International Settlements, Basel, 2024. URL https://www.bis.org/publ/bppdf/bispap159.htm.

Bank of Canada. Privacy-enhancing technologies for CBDC solutions. Staff discussion paper 2025-1, Bank of Canada, Ottawa, 2025. URL https://www.bankofcanada.ca/wp-content/uploads/2025/01/sdp2025-1.pdf.

Bank of England. The digital pound: A new form of money for households and businesses? Consultation paper, Bank of England, London, 2023.

Eli Ben-Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza. Zerocash: Decentralized anonymous payments from bitcoin. In *2014 IEEE Symposium on Security and Privacy*, pages 459–474. IEEE, 2014.

Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Scalable, transparent, and post-quantum secure computational integrity. IACR Cryptology ePrint Archive, Report 2018/046, 2018. URL https://eprint.iacr.org/2018/046.

Michiel Bijlsma, Carin van der Cruijsen, Nicole Kock, and Mauro Mastrogiacomo. Central bank digital currency and privacy: A randomized survey experiment. Bis working papers no. 1147, Bank for International Settlements, Basel, 2024. URL https://www.bis.org/publ/work1147.htm.

BIS Innovation Hub. Project tourbillon: Exploring privacy, security and scalability for CBDCs. Technical report, Bank for International Settlements, Basel, 2023. URL https://www.bis.org/publ/othp80.pdf.

BIS Innovation Hub and Bank of England. Project rosalind: Building API prototypes for retail CBDC ecosystem innovation. Technical report, Bank for International Settlements, Basel, 2023. URL https://www.bis.org/publ/othp69.htm. 33 API functionalities for retail CBDC. PII excluded from API/central bank ledger layer.

BIS Innovation Hub, Bank of Israel, and Hong Kong Monetary Authority. Project sela: An accessible and secure retail CBDC ecosystem. Technical report, Bank for International Settlements, Basel, 2023. URL https://www.bis.org/publ/othp74.htm. Privacy-preserving CBDC with layered data obfuscation and access enablers.

Board of Governors of the Federal Reserve System. Money and payments: The U.S. dollar in the age of digital transformation. Technical report, Federal Reserve, Washington, D.C., 2022.

Markus K. Brunnermeier and Dirk Niepelt. On the equivalence of private and public money. *Journal of Monetary Economics*, 106:27–41, 2019. doi: 10.1016/j.jmoneco.2019.07.004. Conditions under which public and private money are equivalent.

Benedikt Bünz, Shashank Agrawal, Mahdi Zamani, and Dan Boneh. Zether: Towards privacy in a smart contract world. In *Financial Cryptography and Data Security*, pages 423–443. Springer, 2020. doi: 10.1007/978-3-030-51280-4_23. Account-based confidential transactions with pseudonymous identity proofs.

Jeffrey Burdges, Florian Dold, Christian Grothoff, and Marcello Stanisci. Taler: Usable, privacy-preserving payments for the web. In *Proceedings on Privacy Enhancing Technologies (PETS)*, 2016. GNU Taler: payer anonymity + payee transparency implementation.

Matteo Campanelli, Dario Fiore, and Anaïs Querol. Legosnark: Modular design and composition of succinct zero-knowledge proofs. In *ACM Conference on Computer and Communications Security (CCS)*, pages 2075–2092, 2019. doi: 10.1145/3319535.3339820. Composable ZK proofs enabling "prove ownership without revealing identity".

David Chaum. Blind signatures for untraceable payments. In *Advances in Cryptology: Proceedings of Crypto 82*, pages 199–203, Boston, MA, 1982. Springer.

David Chaum. Blind signature system. *Advances in Cryptology*, pages 153–153, 1983.

David Chaum, Christian Grothoff, and Thomas Moser. How to issue a central bank digital currency. SNB Working Papers 2021-03, Swiss National Bank, 2021. URL https://www.snb.ch/en/publications/research/working-papers/2021/working_paper_2021_03. Privacy-preserving CBDC design using blind signatures, co-authored with SNB board member.

Karen Chen et al. Public attitudes towards CBDC: Insights from a survey experiment. Oxford Law Faculty Blog, 2025. URL https://blogs.law.ox.ac.uk/oblb/blog-post/2025/05/public-attitudes-towards-cbdc-insights-survey-experiment. 73% concerned about government fund control; 30% view CBDC as harmful vs 24% beneficial.

Jonathan Chiu and Seyed Mohammadreza Davoodalhosseini. Money laundering and the privacy design of central bank digital currency. *Journal of International Money and Finance*, 131:102805, 2023. doi: 10.1016/j.jimonfin.2023.102805.

Syngjoo Choi, Bongseob Kim, Young Sik Kim, and Ohik Kwon. Central bank digital currency and privacy: A randomized survey experiment. *International Economic Review*, 2025. doi: 10.1111/iere.12746. Nationally representative sample of 3,500+ participants testing privacy preferences for CBDC.

Nathaniel Decker. Proof without exposure: Zero-knowledge proofs as a cryptographic framework for institutional financial compliance. SSRN Working Paper. ZKP-based KYC reducing exposed user data by 97%, AI-enhanced fraud detection achieving 96.7% accuracy, 2025.

Mohammed El-Hajj and Bjorn Oude Roelink. Evaluating the efficiency of zk-SNARK, zk-STARK, and bulletproof in real-world scenarios: A benchmark study. *Information*, 15(8):463, 2024. doi: 10.3390/info15080463. URL https://www.mdpi.com/2078-2489/15/8/463.

Electric Coin Company. Zcash metrics. https://electriccoin.co/blog/, 2023. Shielded transaction volume statistics.

Daniel Escudero et al. Perfectly-secure multiparty computation with linear communication complexity over any modulus, 2024. URL https://eprint.iacr.org/2024/370.

European Central Bank. Report on a digital euro. Technical report, European Central Bank, Frankfurt, 2020.

European Central Bank. Eurosystem report on the public consultation on a digital euro. European Central Bank, 2021. URL https://www.ecb.europa.eu/euro/digital_euro/report/html/index.en.html. 8,221 of 20,000 responses (41%) focused primarily on privacy.

European Central Bank. A stocktake on the digital euro. Technical report, European Central Bank, Frankfurt, 2023.

European Central Bank. Making the digital euro truly private. ECB Blog, 2024. URL https://www.ecb.europa.eu/press/blog/date/2024/html/ecb.blog240613~47c255bdd4.en.html.

Europol. European money mule action leads to 1,803 arrests. Europol Media & Press, 2021. https://www.europol.europa.eu/media-press/newsroom/news/european-money-mule-action-leads-to-1-803-arrests.

Yifei Fan. Some thoughts on CBDC operations in china. *China Finance*, 2020. Deputy Governor, People's Bank of China.

Murad Farzulla. Asymptotic protection: Derivatives, systemic risk, and the limits of hedging. *Zenodo Preprint*, 2025a. doi: 10.5281/zenodo.17620448. Derivatives and systemic risk management.

Murad Farzulla. Consent-theoretic framework for quantifying legitimacy: Stakes, voice, and friction in adversarial governance. *Zenodo Preprint*, 2025b. doi: 10.5281/zenodo.17684676. Operationalization of consent-based legitimacy framework. SSRN:5918222.

Murad Farzulla. Legitimate extraction: Sophisticated laundering hides in plain sight. *SSRN Electronic Journal*, 2025c. doi: 10.2139/ssrn.6145046. Fourth-stage AML framework. Targeting Oxford J. Financial Regulation. Zenodo: 10.5281/zenodo.17626621.

Murad Farzulla. Infrastructure vs regulatory shocks: Asymmetric volatility response in cryptocurrency markets. *Research Square*, 2025d. doi: 10.21203/rs.3.rs-8323026/v1. Under review at Digital Finance (Springer). Also on SSRN:5788082.

Murad Farzulla. Do whitepaper claims predict market behavior? evidence from cryptocurrency factor analysis. *arXiv preprint arXiv:2601.20336*, 2025e. doi: 10.48550/arXiv.2601.20336. With editor at Digital Finance (Springer). SSRN:5918302, Zenodo: 10.5281/zenodo.17917922.

Murad Farzulla and Andrew Maksakov. ASRI: An aggregated systemic risk index for cryptocurrency markets. *arXiv preprint arXiv:2602.03874*, 2025. doi: 10.48550/arXiv.2602.03874. Systemic risk as emergent from distributed friction sources. Zenodo: 10.5281/zenodo.17918239.

Financial Action Task Force. International standards on combating money laundering and the financing of terrorism & proliferation: The FATF recommendations. Technical report, FATF/OECD, Paris, 2012–2023. Updated periodically.

Financial Action Task Force. Targeted update on implementation of the FATF standards on virtual assets and VASPs. Technical report, FATF/OECD, Paris, 2024. URL https://www.fatf-gafi.org/content/dam/fatf-gafi/recommendations/2024-Targeted-Update-VA-VASP.pdf.

Michael G. Findley, Daniel L. Nielson, and J. C. Sharman. *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*. Cambridge University Press, Cambridge, 2014.

Ariel Gabizon, Zachary J. Williamson, and Oana Ciobotaru. PLONK: Permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge. IACR Cryptology ePrint Archive, Report 2019/953, 2019. URL https://eprint.iacr.org/2019/953.

Rodney Garratt, Michael Lee, Brendan Malone, and Antoine Martin. Token- or account-based? a digital currency can be both. *Federal Reserve Bank of New York Staff Reports*, 973, 2021.

Rosario Gennaro and Steven Goldfeder. Fast multiparty threshold ECDSA with fast trustless setup. *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1179–1194, 2018.

David Golumbia et al. Privacy implications of central bank digital currencies. *University of Florida Levin College of Law Research Paper*, 2024. URL https://scholarship.law.ufl.edu/cgi/viewcontent.cgi?article=2239&context=facultypub.

Geoffrey Goodell, Hazem Danny Al-Nakib, and Tomaso Aste. Retail central bank digital currency: Motivations, opportunities, and mistakes. arXiv preprint, also SSRN:4769226, 2025.

Google. Private join and compute. Google AI Blog, 2019. https://ai.googleblog.com/2019/06/private-join-and-compute.html.

Jonas Gross, Johannes Sedlmeir, Matthias Babel, Alexander Bechtel, and Benjamin Schellinger. Designing a central bank digital currency with support for cash-like privacy. *SSRN Electronic Journal*, 2021. doi: 10.2139/ssrn.3891121.

Jens Groth. On the size of pairing-based non-interactive arguments. In *Advances in Cryptology – EUROCRYPT 2016*. Springer, 2016. doi: 10.1007/978-3-662-49896-5_5.

International Monetary Fund. Central bank digital currency data use and privacy protection. *FinTech Notes*, 2024(004), 2024. URL https://www.elibrary.imf.org/view/journals/063/2024/004/article-A001-en.xml.

Ipsos UK. 57% of britons support a national id card scheme but have significant concerns over data security and potential government overreach. Ipsos UK Polling, 2025. URL https://www.ipsos.com/en-uk/57-britons-support-national-id-card-scheme-have-significant-concerns-over-data-security-and.

Dae-Yong Kang. Rethinking the global anti-money laundering regulations to deter corruption. *International Criminal Law Review*, 18:871–901, 2018.

Todd Keister and Daniel Sanches. Should central banks issue digital currency? *Review of Economic Studies*, 90(1):404–431, 2023. doi: 10.1093/restud/rdac017. General equilibrium analysis of CBDC impact on banking.

Özgür Kesim, Christian Grothoff, Florian Dold, and Martin Schanzenbach. Zero-knowledge age restriction for GNU taler. In *European Symposium on Research in Computer Security (ESORICS 2022)*. Springer, 2022. ZK-based selective attribute disclosure in privacy-preserving payment system.

Nishanth Koti, Parth Kukkala, Ajith Patra, Divya Raj, K. Vivek Ramachandran, Ashish Thapliyal, and Avinash Varma. Graphiti: Secure graph computation made more scalable. IACR Cryptology ePrint Archive, Report 2024/1756, 2024. URL https://eprint.iacr.org/2024/1756.

Michael Levi. Evaluating the control of money laundering and its underlying offences: The search for meaningful data. *Asian Journal of Criminology*, 15:301–320, 2020.

James Lovejoy, Madars Virza, Cory Fields, Kevin Karwaski, Anders Brownworth, and Neha Narula. Hamilton: A high-performance transaction processor for central bank digital currencies. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. USENIX Association, 2023. URL https://www.usenix.org/conference/nsdi23/presentation/lovejoy. Project Hamilton (Fed Boston + MIT). 1.7M TPS. Also IACR ePrint 2022/586.

Sarah N. Lynch. Danske bank pleads guilty in $2 billion money laundering scheme. Reuters, 2022. https://www.reuters.com/business/finance/danske-bank-plead-guilty-fraud-charges-doj-says-2022-12-13/.

Andrea Minto. Certainties and uncertainties surrounding central bank digital currencies (CBDC) vis-a-vis the EU anti-money laundering regulatory framework. *European Company and Financial Law Review*, 21(5-6):671–704, 2024. doi: 10.1515/ecfr-2024-0020. EU AML regulatory treatment of offline vs. online digital euro.

Chukwuebuka Okonkwo and Oluwaseun Adesina. Privacy implications of central bank digital currencies (CBDCs): a systematic review of literature. *Journal of Information Policy*, 2024. doi: 10.1080/07366981.2024.2376794.

Ronald F. Pol. Anti-money laundering: The world's least effective policy experiment? together, we can fix it. *Policy Design and Practice*, 3(1):73–94, 2020. doi: 10.1080/25741292.2020.1725366.

Anitha R et al. Promise of zero-knowledge proofs (ZKPs) for blockchain privacy and security: Opportunities, challenges, and future directions. *Security and Privacy*, 2025. doi: 10.1002/spy2.461.

Ellie Rennie and Stacey Steele. Privacy and emergency payments in a pandemic: How to think about privacy and a central bank digital currency. *Law, Technology and Humans*, 3(1), 2021. Identifies four privacy losses from CBDC: anonymity, liberty, individual control, regulatory control.

István András Seres et al. Balancing compliance and privacy in offline CBDC transactions using a secure element-based system, 2024. URL https://arxiv.org/abs/2509.25469.

J. C. Sharman. *The Despot's Guide to Wealth Management: On the International Campaign against Grand Corruption*. Cornell University Press, Ithaca, 2017.

Giulio Soana and Thomaz de Arruda. Central bank digital currencies and financial integrity: Finding a new trade-off between privacy and traceability within a changing financial architecture. *Journal of Banking Regulation*, 2024. doi: 10.1057/s41261-024-00241-2. Examines the privacy-traceability trade-off in changing financial architecture.

Alberto Sonnino, Mustafa Al-Bassam, Shehar Bano, Sarah Meiklejohn, and George Danezis. Coconut: Threshold issuance selective disclosure credentials with applications to distributed ledgers. In *Network and Distributed System Security Symposium (NDSS)*, 2019. URL https://www.ndss-symposium.org/ndss-paper/coconut-threshold-issuance-selective-disclosure-credentials-with-applications-to-distributed-ledgers/.

Statewatch, Big Brother Watch, and Privacy International. Joint briefing on the "do not introduce digital id cards" parliamentary petition debate. Statewatch News, 2025. URL https://www.statewatch.org/news/2025/december/uk-joint-briefing-on-the-do-not-introduce-digital-id-cards-parliamentary-petition-debate/. 2.96 million signatures, fourth-largest UK parliamentary petition.

Katrin Tinn. A theory model of digital currency with asymmetric privacy. *Management Science*, 2025. doi: 10.1287/mnsc.2024.06830.

Paul Tucker. *Global Discord: Values and Power in a Fractured World Order*. Princeton University Press, Princeton, 2023.

United Nations Office on Drugs and Crime. Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes. Technical report, UNODC, Vienna, 2011.

Marie Beth van Egmond, Vincent Dunning, Stefan van den Berg, Thomas Rooijakkers, Alex Sangers, Ton Poppe, and Jan Veldsink. Privacy-preserving anti-money laundering using secure multi-party computation. IACR Cryptology ePrint Archive, Report 2024/065, 2024. URL https://eprint.iacr.org/2024/065. Published at Financial Cryptography 2024. Deployed MPC-based AML with ABN AMRO, Rabobank, De Volksbank.

Xiaodong Wang, Zijie Lu, Bei Liang, and Shengzhe Meng. Unbalanced private set intersection from client-independent relaxed oblivious PRF. *Proceedings on Privacy Enhancing Technologies*, 2025 (3):475–493, 2025. doi: 10.56553/popets-2025-0109. URL https://crysp.petsymposium.org/popets/2025/popets-2025-0109.pdf.

C. M. Wood, I. Mularczyk, K. Paterson, and C. A. Rotaru. Verifiable Oblivious Pseudorandom Functions. RFC 9497, September 2023. URL https://www.rfc-editor.org/info/rfc9497.

Jiping Yu, Kun Chen, Yunyi Chen, Xiaoyu Fan, Xiaowei Zhu, Cheng Hong, and Wenguang Chen. GraphAce: Secure two-party graph analysis achieving communication efficiency. In *34th USENIX Security Symposium (USENIX Security 25)*, August 2025. URL https://www.usenix.org/system/files/usenixsecurity25-yu-jiping.pdf.

Filippo Zatti. From legal tender to prohibition: Competing paradigms in central bank digital currency law. *Strani pravni život*, 2025. Analyses the U.S. Anti-CBDC Surveillance State Act as paradigm shift from enabling to prohibiting CBDC.

Wei Zhang et al. Central bank digital currencies: A survey, 2025. URL https://arxiv.org/abs/2507.08880.