LASB: A Liquidity-Adjusted Security Metric for Consensus Attacks

Ke Wang[1,2], Wenbin Wu[1], and Alexander Neumueller[1]

[1]Cambridge Centre for Alternative Finance, Cambridge Judge Business School, University of Cambridge, Cambridge, UK

[2]College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

k.wang@jbs.cam.ac.uk, w.wu@jbs.cam.ac.uk

a.neumueller@jbs.cam.ac.uk

May 7, 2026
WEIS 2026 Workshop Version

Abstract

We propose the *Liquidity-Adjusted Security Budget* (LASB), a general economic-security metric for consensus attacks. The paper's central claim is simple: *technical control is not cashable profit*. LASB asks whether a technically feasible attack can survive market, settlement, and physical constraints long enough to be monetized. Bitcoin is our main calibration case, not the conceptual endpoint of the framework.

The framework combines a unified (N, L, P) system model with a Stackelberg game between an attacker and Bayesian market makers. The resulting cashable-extraction bound shows that large hedge positions trigger endogenous liquidity withdrawal, settlement bottlenecks, and detection risk before nominal gains can be realized. Under 2026-01-08 baseline conditions, the calibrated attack cost is \$24.1B while realistic cashable extraction is approximately \$1.02B, yielding LASB = 0.042. This is a dated baseline, not a time-invariant guarantee; Section 10 reports regime variation within the current subsidy-dominated environment. In an agent-based model calibrated to historical stress episodes, markets detect anomalous position-building in roughly 80 minutes and liquidity collapses before meaningful extraction can occur. The physical side is also visible: on the same hashrate baseline, current-efficiency procurement implies roughly 5.2–6.7 GW of continuous load and about 1.4 million current-generation ASICs.

The framework implication is broader than the Bitcoin case. LASB supports cross-chain comparison, regime monitoring, and policy-oriented surveillance of consensus attack incentives. We argue that protocol security should be analyzed jointly with market extractability, not through block rewards and fees alone.

Keywords: Blockchain Security, Market Microstructure, Game Theory, Agent-Based Modeling, Cryptocurrency Economics, 51% Attacks, Liquidity Constraints, Cybercrime Economics, Cyber Risk Quantification

1 Introduction

In January 2019, Ethereum Classic suffered three consecutive 51% attacks over a span of 72 hours. The attacker reorganized thousands of blocks, executed multiple double-spends against exchanges, and vanished—reportedly netting around \$1.1M. This event crystallized a troubling question: if someone assembles sufficient hashrate, can they actually profit from breaking a major blockchain?

The conventional wisdom says yes. Early analyses established that rational miners maximize profit by following protocol rules [29], but subsequent work raised alarms about declining security as block subsidies diminish and fee markets become the primary incentive layer [7, 9]. The logic seems straightforward: accumulate hashrate, short the asset on derivatives exchanges, execute the attack, profit from the hedge. For Bitcoin—with its substantial market capitalization—the potential spoils appear astronomical.

Our claim, in one sentence, is simple: *technical control is not cashable profit.* Even a technically feasible majority attack still has to be hedged, monetized, and settled at scale, and those are precisely the steps where current market structure pushes the attacker below break-even.

1.1 Short-Selling Attacks and Majority Control

A *short-selling attack* profits if the attacker can (i) build a short exposure to the asset, (ii) cause a sharp adverse price move, and (iii) successfully *settle and withdraw* gains before counterparties halt trading or default. In proof-of-work systems, majority hashrate enables the key step (ii): the attacker can censor or reorder transactions, create deep reorganizations, and (in the extreme) execute double-spends against exchanges. These actions undermine settlement finality and can trigger rapid price declines and liquidity withdrawal.

However, translating such protocol-level power into *realizable* profit requires market extraction: the attacker must acquire and close large short positions across exchanges and venues, pay execution costs, survive endogenous liquidity evaporation, and pass solvency, position-limit, and compliance constraints. This paper studies that extraction step.

This paper challenges frictionless assumptions by distinguishing between *technical feasibility* and *economic extractability*. Rather than asking whether consensus attacks are technically feasible, we focus on whether any on-chain gains can be converted into realizable economic value given security-relevant market constraints and exchange risk controls. We demonstrate that under current conditions, this conversion faces binding constraints that prior analyses either ignore or substantially underestimate.

1.2 Scope, Research Question, and Main Claim

Scope Statement. This paper analyzes the economic extractability of 51% attacks on Bitcoin under 2026-01-08 market conditions. We distinguish between *technical feasibility* (can an attacker with sufficient hashrate reorganize the chain?) and *economic extractability* (can they profit from doing so?). We do not claim attacks are technically impossible. Rather, we demonstrate that under stated assumptions (Appendix A.3), attacks are *economically dominated*—rational attackers face negative expected returns due to binding economic constraints.

Attacker Model. This paper analyzes *economically rational, profit-maximizing attackers* who require positive expected returns. We do not cover purely destructive, politically motivated, or state-sponsored attackers who may not require profitability. Such adversaries face different constraint sets and represent an important but distinct research direction (see Section 11 for discussion).

Our conclusions are conditional on current market structure and parameter calibrations (Appendix A.3). They may not hold under substantially different regimes such as post-2140 fee-dominated security, major regulatory changes, novel derivative instruments, or fundamentally different market microstructure. The contribution is not a claim that "Bitcoin is secure," but rather a framework for systematically analyzing extraction constraints under evolving conditions.

We focus narrowly on the feasibility of *consensus-level attacks* on Bitcoin under real-world economic, microstructural, and physical constraints. Our analysis excludes asset valuation dynamics, consumer protection concerns, regulatory policy debates, and broader macro-financial

spillovers. Instead, we ask a precise question: *Can a rational attacker who has already assembled the physical infrastructure to execute a 51% attack actually extract value from it?*

Consider the requirements. To attack Bitcoin today, an adversary needs substantial physical infrastructure including about 1.4 million current-generation ASIC miners and gigawatt-scale sustained power supply (see Table 5), plus network infrastructure with latency advantages.

Table 1: At-a-Glance Calibrated Results for Bitcoin under 2026-01-08 Baseline Conditions

Quantity	**Value**	**Interpretation**
Attack cost C_{hash}	\$24.1B	Capital required for attack-scale hardware and energy
Realistic extraction V_{extract}	\$1.02B	Cashable proceeds under binding constraints
LASB	0.042	Attacker loses roughly 95.8% of investment
Median market detection time	80 minutes	Time before liquidity freezes in the ABM
Physical load envelope	5.2–6.7 GW	2026-01-08 current-efficiency observability envelope

Table 1 summarizes the quantitative punchline before the formal model: the attack is not close to break-even on any of the paper's main margins. More importantly, this paper's framework claim is broader than the Bitcoin case itself: we propose LASB as a general economic-security metric for consensus attacks, and use Bitcoin as the main calibration case.

Standard models [9] predict such an attacker could profitably double-spend by shorting BTC before executing the attack. We demonstrate this conclusion fails due to three binding constraints that prior work either ignores or dramatically underestimates.

1.3 Endogenous Liquidity as the Core Mechanism

Classical analyses of 51% attacks treat liquidity as *exogenous*—assuming an attacker can hedge arbitrarily large positions without materially affecting prices or triggering counterparty responses. In our view, this is the central flaw. Liquidity is *endogenous*: the very act of building the hedge required for a profit-seeking attack transforms the market itself.

This creates four compounding failures:

1. *Execution costs*: Slippage costs scale super-linearly with position size (not linearly, as naive models assume)
2. *Liquidity evaporation*: Market makers detect anomalous flow and withdraw—often within minutes, not days
3. *Counterparty insolvency*: Exchanges become unable to honor winning positions once trust evaporates
4. *Temporal fragmentation*: Multi-day extraction increases detection probability exponentially

These frictions combine to create what we term an *execution gap*—a fundamental barrier where the hedge capacity required for profitability exceeds the realizable short capacity achievable before detection. While we cannot rule out irrational or state-sponsored attackers who don't care about profitability, our analysis shows that rational economic actors face substantial barriers under current market conditions.

1.4 Extraction Constraints and Main Intuition

We introduce a formal framework for mapping the feasible set of attack strategies. Let $\mathcal{X}$ denote the strategy space and $\mathcal{F} \subseteq \mathcal{X}$ the feasible region. For profit-seeking attackers, the critical dimensions are:

- S: Attack scale (hashrate requirement)
- Π: Realizable profitability (net of execution and settlement frictions)
- A: Anonymity, defined as $A(x) = 1 - D(x)$ where $D(x)$ is detectability

The key insight: these dimensions exhibit three binding trade-offs.

1. *Scale vs. Profitability*: Bounded by endogenous liquidity and settlement constraints. Larger attacks require bigger hedges, but market depth is finite and exchanges will fail before honoring positions of the required scale (see Table 6).
2. *Scale vs. Anonymity*: Bounded by supply chain and energy infrastructure observability. The physical footprint of the required power consumption and ASIC procurement is simply too large to hide (see Table 5).
3. *Profitability vs. Anonymity*: Bounded by counterparty and compliance frictions. Profitable extraction requires KYC-compliant exchanges; true anonymity requires peer-to-peer channels with minimal liquidity (see Table 6).

Our central theoretical result (Proposition 1) formalizes this as a break-even exclusion claim under joint cash-out constraints. Under current market microstructure, attack strategies cannot simultaneously scale the hedge, preserve concealment, and convert paper gains into withdrawable cash quickly enough to close the gap to attack cost.

1.5 Feasibility Regions

Figure 1 visualizes the attack feasibility landscape across two key dimensions: market liquidity/regulatory strength (x-axis) and attack cost/technical efficiency (y-axis). The diagram partitions the strategy space into three regions:

1. **Economically Dominated** (bottom-left, red): Extraction capacity $V_{\text{extract}} < C_{\text{hash}}$ due to binding liquidity and solvency constraints. Bitcoin currently resides here.
2. **Physically Detectable** (top-left, yellow): High attack costs create observable power signatures exceeding detection thresholds, enabling supply chain and energy monitoring.
3. **Theoretical Only** (right side, green): Requires unrealistic market conditions (e.g., unlimited liquidity, no regulatory constraints) that do not exist in practice.

This visualization clarifies that our results correspond to a specific region of parameter space. Future market evolution (e.g., deeper DEX liquidity, relaxed position limits) could shift Bitcoin's position, but the framework remains applicable for analyzing security under new conditions.

1.6 Method and Conservative Bounding

All constraints defining $\mathcal{F}$ derive from the *conservative bounding framework* (Appendix A.3), which systematically biases every parameter estimate *in favor of attack feasibility*. We compute realizable short capacity using only visible order book depth, apply worst-case fee schedules, use monotonic upper envelopes for execution costs, assume attackers coordinate across all major extraction channels, and exclude defensive responses.[1] If no attack vector succeeds under these attacker-favorable conditions, none will succeed in practice.

The framework applies primarily to *rational economic attackers* requiring positive expected profit. State-level attackers face different constraint sets, but surprisingly still encounter anonymity collapse as scale increases—the required power signature is hard to hide even for nation-states (see Table 5).

1.7 Roadmap and Model Logic

Readers can view Sections 2–5 as progressive refinements of a *single* upper bound on *cashable* extraction rather than as separate theories making the same point repeatedly. Section 2 defines the common state variables and the notional quantity $\mathrm{RSC}(t)$. Section 3 converts that notional capacity into cashable extraction V_{extract} via Δ_{cap} and venue-level payout limits. Section 4 applies the same bound to the timing problem of short-selling and double-spend attacks. Section 5 then endogenizes the liquidity term by modeling strategic market-maker withdrawal. The empirical sections stress-test these same mechanisms rather than introducing a different model.

1.8 Related Work

Work on consensus attacks has largely focused on protocol incentives, mining behavior, and technical feasibility. Nakamoto [29], Eyal and Sirer [15], Gervais et al. [17], Budish [7], Carlsten et al. [9], Bonneau [5], Sayeed and Marco-Gisbert [33], and Badertscher and Lu [3] show when majority attacks or related deviations may be rational at the protocol level, including the rent-versus-buy margin for hostile takeovers. Our departure is to ask a different question: even if the protocol-level attack is feasible, can its gains be converted into *cashable* profit once liquidation, settlement, and detection frictions are endogenous?

To answer that question, we draw on market-microstructure and game-theoretic literatures on price impact, informed trading, and liquidity withdrawal under stress [22, 19, 6, 27, 20]. For crypto markets in particular, Makarov and Schoar [28], Easley et al. [14], and Lehar and Parlour [25] show how fragmentation, adverse selection, and venue fragility shape realized trading outcomes. We adapt those insights to an adversarial setting in which the attacker's own hedge-building changes the market state.

Our work also connects to adjacent literatures on extractable value, strategic blockchain trading, and empirical stress testing [13, 32, 31, 16, 23, 2, 21, 1]. The paper's main contribution is to integrate these strands into a single economic-security framework and to propose LASB as a practical measurement tool for comparing attack incentives across chains and market regimes.

1.9 Contributions

We make three contributions.

1. *Economic extractability framework.* We introduce a formal framework for analyzing the economic extractability of consensus attacks under real-world market constraints, distinguishing between *technical feasibility* and *economic realizability*.

[1] We focus on Bitcoin; Ethereum moved to proof-of-stake in 2022.

2. *Game-theoretic liquidity model.* We develop a Stackelberg game between an attacker and Bayesian market makers, showing that the large hedging positions required for profitable attacks can trigger endogenous liquidity withdrawal that limits realizable profits.

3. *Liquidity-Adjusted Security Budget (LASB).* We introduce LASB, a metric for comparing blockchain security that incorporates market liquidity constraints in addition to protocol-level rewards. Bitcoin is the main calibration case, but the framework is designed to travel across chains, regimes, and attack environments.

1.10 Paper Organization

The paper follows a theory-to-evidence structure. Sections 2 and 3 state the core model and the main extraction bound. Sections 4 and 5 show how that bound behaves under timing and strategic liquidity withdrawal. Sections 6–8 then validate the same mechanisms using physical constraints, simulation, and market data. Sections 9–11 draw out implications for measurement and policy, with appendices reserved for proofs, calibrations, and supplementary tables.

The organizing principle throughout is simple: start from cashable extraction, identify the bottleneck that binds first, and then ask whether any realistic market evolution relaxes that bottleneck enough to matter.

2 The (N, L, P) System Model

This section states the paper's master model. A majority attack is economically relevant only if it survives three layers at once: consensus control, market extraction, and physical implementation. If any one of those layers fails, the attack fails economically.

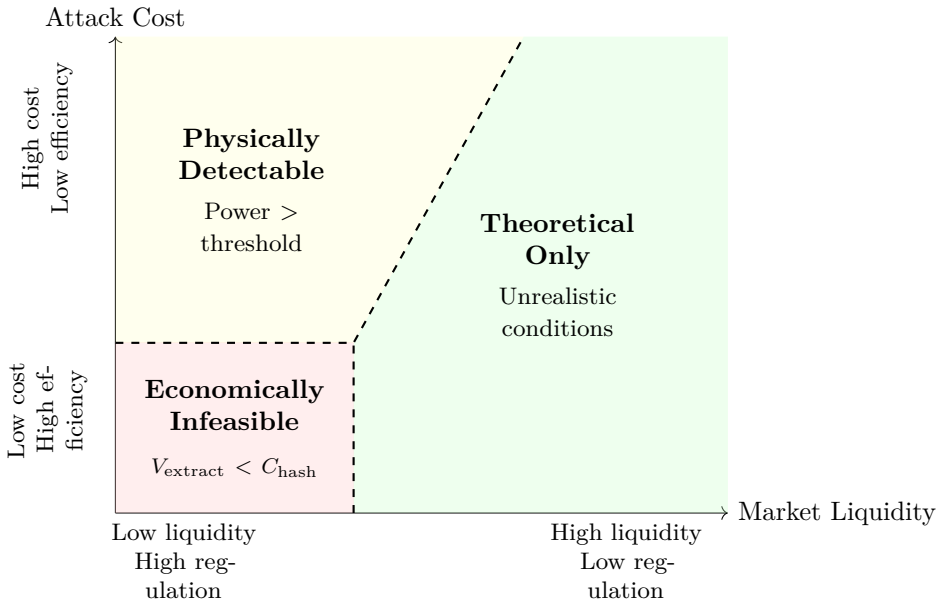
Definition 1 ((N, L, P) System Model)**.** A proof-of-work blockchain attack system is characterized by the 3-tuple (N, L, P) where:

$$\begin{aligned} N &= \{\text{hashrate } H,\ \text{difficulty } D,\ \text{block time } \tau_b\} \\ L &= \{\text{open interest } OI,\ \text{bid-ask spread } \sigma,\ \text{depth } \Delta\} \\ P &= \{\text{power capacity } W,\ \text{ASIC count } M,\ \text{latency } \lambda\} \end{aligned}$$

where:

- N represents the **Network** dimension: consensus-layer parameters governing attack feasibility
- L represents the **Liquidity** dimension: market microstructure variables determining extraction capacity
- P represents the **Physical** dimension: observable infrastructure constraints enabling detection

The interpretation is immediate. N captures consensus control, L captures market extraction capacity, and P captures the observable footprint of implementing the attack. The paper's claim is not that these dimensions are separate obstacles, but that a realistic attacker must clear all three at once.



Key Insights:

- Bitcoin resides in the **Economically Infeasible** region where $V_{\text{extract}} < C_{\text{hash}}$
- Moving right (higher liquidity) increases extraction capacity but also detection probability
- Moving up (higher attack cost) makes attacks less profitable
- No path exists to positive expected profit under current market conditions

Figure 1: Phase Diagram of Attack Feasibility Regions. The diagram characterizes parameter regions across two dimensions: market liquidity/regulatory strength (x-axis) and attack cost/technical efficiency (y-axis). Bitcoin (2026-01-08) resides in a region where extraction capacity is bounded below attack costs under current market microstructure.

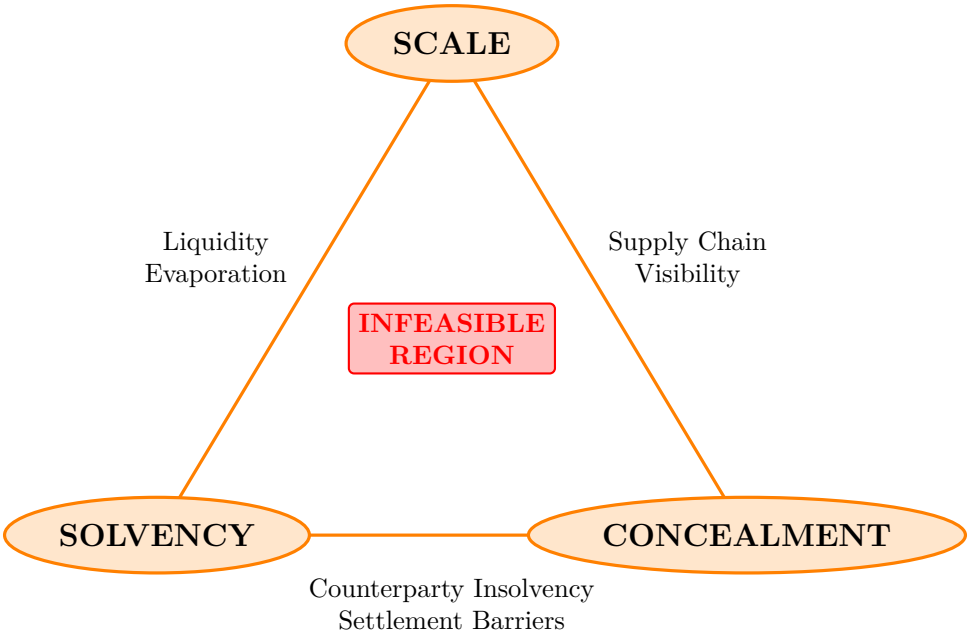


Figure 2: The Extraction Constraints Framework. Attackers face a three-way trade-off: achieving high scale requires sacrificing either solvency (exchange capacity to pay) or concealment (avoiding detection). The central region represents parameter combinations where all three constraints bind simultaneously.

2.1 Model Logic

The paper uses one flow throughout: *(i)* define notional market-side capacity through RSC(t); *(ii)* convert that notional quantity into cashable extraction using Δ_{cap}; and *(iii)* endogenize the liquidity term through strategic market-maker withdrawal and empirical validation.

The distinction between *notional* and *cashable* values is essential. RSC(t) is about how large a hedge can be built; V_{extract} is about how much cash can actually be taken out. When it appears below, $V_{\text{short}}(t)$ denotes the attacker's short notional outstanding at time t.

2.2 Threat Models

We distinguish three attacker types because the same protocol capability can map into different economic problems:

Definition 2 (Attacker Taxonomy)**.** We distinguish three attacker types based on their objective functions and constraint sets:

1. **Rational Economic Attacker**: Seeks positive expected profit. Faces all three constraint dimensions (N, L, P) with binding liquidity and physical detection constraints. This is our primary focus.

2. **Slow-Accumulation Attacker**: Attempts to evade detection by spreading position accumulation over extended time horizon $T \to \infty$. Faces time-integrated detection probability that grows super-linearly with T (Calibration Result 1).

3. **Nation-State Attacker**: Operates under different objective function (e.g., disruption rather than profit). While liquidity constraints may be relaxed, physical detection constraints remain binding—the current-efficiency power signature is still roughly 5.2–6.7 GW on the 2026-01-08 baseline (Table 5). We model this as a *different regime* rather than a counterexample to our framework.

The paper focuses on Type 1 because it is the threat model implicit in most protocol-level analyses of attack profitability. Type 2 and Type 3 matter mainly to show that spreading the attack out or relaxing the profit motive does not remove the paper's main bottlenecks.

2.3 Realizable Short Capacity and Liquidity Dynamics

Prior models typically treat liquidity as exogenous [7]. We do not. In our setting, building the hedge changes the market, and that feedback is what makes the extraction problem hard.

Definition 3 (Realizable Short Capacity (RSC))**.** The **Realizable Short Capacity** $\mathrm{RSC}(t)$ is the maximum notional short position size (in USD) an attacker can establish and close at time t without triggering:

1. Exchange insolvency (insurance fund depletion)
2. Liquidity evaporation (market maker withdrawal exceeding replenishment rate)
3. Regulatory intervention (position limit violations or circuit breakers)

Formally:

$$\mathrm{RSC}(t) = \min\left\{V_{\mathrm{ins}}(t), V_{\mathrm{liq}}(t), V_{\mathrm{reg}}(t)\right\} \tag{1}$$

where $V_{\mathrm{ins}}(t)$ is solvency-bounded capacity, $V_{\mathrm{liq}}(t)$ is liquidity-bounded capacity, and $V_{\mathrm{reg}}(t)$ is compliance-bounded capacity.

The minimum operator matters because these constraints are not substitutable. More capacity on one margin does not undo a tighter bound on another. We model liquidity evolution for notional capacity as:

$$\frac{dV_{\mathrm{liq}}}{dt} = -\beta \cdot V_{\mathrm{short}}(t) + \alpha \cdot (V_{\mathrm{eq}} - V_{\mathrm{liq}}(t)) \tag{2}$$

where $\beta > 0$ is the liquidity destruction rate and $\alpha > 0$ is mean reversion. In plain terms, the faster the attacker builds size, the faster the market stops accommodating it.

Definition 4 (Attack Feasibility Region)**.** An attack strategy $\mathcal{A}$ is **feasible** if it satisfies constraints from all three dimensions simultaneously:

$$\begin{aligned}
\mathcal{A} \in \mathcal{F} \iff \quad & H(\mathcal{A}) \geq H_{\min}(N) && \text{(network: sufficient hashrate)} \\
\wedge \quad & V_{\mathrm{extract}}(\mathcal{A}) \leq \mathrm{RSC}(L) && \text{(liquidity: settlement capacity)} \\
\wedge \quad & D(\mathcal{A}) \leq D_{\max}(P) && \text{(physical: detection threshold)}
\end{aligned}$$

where $H_{\min}(N)$ is minimum hashrate for consensus dominance, $\mathrm{RSC}(L)$ is realizable short capacity, and $D_{\max}(P)$ is maximum tolerable detection probability.

The feasibility region $\mathcal{F}$ is the intersection of three constraint sets. This is the structural reason the paper keeps finding the same result from different angles.

2.4 Break-Even Exclusion

We can now state the paper's main bound: if cashable extraction never catches up to attack cost even under attacker-favorable calibration, then the attack is economically dominated.

Proposition 1 (Break-Even Exclusion Under Joint Cash-Out Constraints)**.** *Fix a market regime with: (i) notional realizable short capacity RSC(t) as in Definition 3, (ii) captured price move fraction $\Delta_{cap} \in [\Delta_{cap}^{\min}, \Delta_{cap}^{\max}]$, and (iii) venue-level upper bounds on immediate payout / withdrawal capacity (insurance funds, margin, equity) summarized in Table 6.*

If the calibrated bounds satisfy

$$\sup_t \ \Delta_{cap} \cdot RSC(t) \ < \ C_{hash}, \tag{3}$$

then for any attack strategy $\mathcal{A}$ in our class (including multi-stage accumulation, cross-market hedging, and account fragmentation), extracted value is economically dominated: $V_{extract}(\mathcal{A}) < C_{hash}$ under the assumptions stated in Appendix A.3.

Moreover, the condition (3) *is robust to solvency-cap assumptions: even under the attacker-favorable solvency upper bound (Appendix A.4), the binding constraints remain liquidity withdrawal and concealment.*

Proof. Let $\mathcal{A}$ be any strategy in the admissible class. By construction of the market-side bound, at any extraction time t the strategy's withdrawable value is bounded by the minimum of its cash-conversion, settlement, liquidity, and concealment constraints:

$$V_{\text{extract}}(\mathcal{A}, t) \leq \min\{\Delta_{\text{cap}} \cdot \text{RSC}(t), V_{\text{solvency}}(t), V_{\text{liquidity}}(t), V_{\text{concealment}}(t)\}\,.$$

Since the minimum of a set of numbers is bounded above by each element, we have in particular

$$V_{\text{extract}}(\mathcal{A}, t) \leq \Delta_{\text{cap}} \cdot \text{RSC}(t) \qquad \text{for all } t.$$

Taking the supremum over feasible extraction times gives

$$V_{\text{extract}}(\mathcal{A}) \leq \sup_t \Delta_{\text{cap}} \cdot \text{RSC}(t).$$

If condition (3) holds, then

$$V_{\text{extract}}(\mathcal{A}) < C_{\text{hash}}.$$

Because $\mathcal{A}$ was arbitrary, the inequality holds for every strategy in the admissible class, including multi-stage accumulation, cross-market hedging, and account fragmentation. Hence all such strategies are economically dominated. The final robustness claim follows because relaxing the solvency cap upward cannot invalidate the conclusion once the tighter cash-out envelope in (3) remains below C_{hash}; under the calibration used in the paper, liquidity withdrawal and concealment continue to bind first. □

The important asymmetry is that attack costs scale up with hashrate, while extraction is pinned down by the *minimum* across several independent bottlenecks. That is why even a technically feasible attack can remain economically dominated.

Failure frontier. The paper's conclusion would materially weaken only under a joint regime shift in which several bottlenecks loosen together: much deeper hedgeable liquidity, weaker surveillance and position enforcement, slower endogenous withdrawal, and materially larger immediate cash-settlement capacity. Relaxing one parameter in isolation is generally not enough to reach break-even, which is why the sensitivity analysis in Appendix A.5 remains well below LASB = 1 even in attacker-favorable scenarios.

Why this is not a modeling accident. The same gap reappears from three directions: market microstructure, physical observability, and game-theoretic best responses. The sections that follow show those three routes separately, but they all point to the same cashable-extraction limit.

2.5 Slow Accumulation and Detection

A natural objection is that patient attackers could spread the hedge over time. The next calibrated result explains why that does not help under the detection parameters used in the submission version.

Calibration Result 1 (Slow-Accumulation Detection Bounds)**.** *For any attack requiring position size $V > V_{threshold}$ (where $V_{threshold}$ is the single-period detection threshold), spreading accumulation over time horizon T faces cumulative detection probability that approaches certainty. Specifically, cumulative detection probability $P_{detect}(T)$ satisfies:*

$$\lim_{T\to\infty} P_{detect}(T) = 1 \tag{4}$$

Moreover, for realistic parameter ranges, $P_{detect}(T) > 0.95$ for $T \geq 30$ days (see Table 6).

Calibration Logic. Let $v(t)$ be the accumulation rate at time t, with $\int_0^T v(t)dt = V$. The cumulative detection probability is:

$$P_{\text{detect}}(T) = 1 - \prod_{t=0}^{T}(1 - p_{\text{detect}}(v(t))) \tag{5}$$

where $p_{\text{detect}}(v(t))$ is the per-period detection probability. Even if $v(t)$ is kept below the single-period threshold, the product term decays exponentially with T. For $T \to \infty$, the probability of avoiding detection across all periods approaches zero. Numerical analysis shows that for an attack-scale hedge and realistic detection thresholds, $P_{\text{detect}}(30 \text{ days}) > 0.95$. Appendix A.4 gives the compact supporting argument. □

Slow accumulation closes no loophole. It reduces per-period visibility, but it makes the attacker visible for longer, and the cumulative detection probability approaches one.

Taken together, the (N, L, P) model does more than stack constraints. It identifies the master object that the rest of the paper keeps refining: cashable extraction after network, market, and physical bottlenecks are imposed jointly.

3 The Extraction Constraints Framework

This section states the paper's market-side bound. Its job is to convert nominal attack scale into the economically relevant object: withdrawable cash. The question is not whether a 51% attack can move price in theory, but how much of that price move can be turned into cash before settlement, liquidity, and detection constraints bind.

Notation (single meaning for each symbol). We use the following distinct quantities throughout the paper:

- RSC(t): maximum *notional* short exposure that can be built and closed at time t without violating solvency, liquidity, or compliance constraints.
- $\Delta_{\text{cap}} \in (0, 1]$: fraction of the attack-induced price move that the attacker actually captures after execution frictions.
- V_{extract}: *cashable* extraction, i.e., withdrawable cash after fees, slippage, and binding constraints.
- CEX: centralized exchange; CME: Chicago Mercantile Exchange; ETF: exchange-traded fund.

Our baseline is $\Delta_{\text{cap}} = 0.5$: the attacker captures half of the price move after execution and liquidation frictions. This is already optimistic; Appendix A.5 shows the paper's conclusions survive over $\Delta_{\text{cap}} \in [0.2, 0.8]$.

With that notation, *notional* hedge size maps to *cashable* extraction through:

Importantly, aggregate open interest is not itself an extraction pool. It measures gross outstanding exposure, whereas cashable extraction is bounded by the fraction of that exposure an attacker can enter, move, settle, and withdraw before liquidity withdrawal, margin calls, position limits, and counterparty solvency constraints bind.

$$V_{\text{extract}} \leq \min\{\Delta_{\text{cap}} \cdot \text{RSC}(t),\ V_{\text{solvency}}(t),\ V_{\text{liquidity}}(t),\ V_{\text{concealment}}(t)\} \tag{6}$$

The attacker must survive all four terms, not just one of them. That minimum operator is the entire point of the section.

3.1 Three Binding Constraints

3.1.1 Solvency Constraint: Paper Profit Does Not Settle

Paper profit is not the same as settled cash. Insurance funds are the only immediately accessible buffer; counterparty margin, ADL, exchange equity, and clawbacks are slower, contested, or already consumed by the same crisis that makes the short profitable. Even generous solvency assumptions do not rescue the attack, because liquidity withdrawal and concealment bind first.

3.1.2 Liquidity Speed Constraint

The attacker faces a speed problem. Building the hedge slowly risks missing the attack window; building it quickly makes the hedge itself an observable anomaly. In our model, market makers update beliefs continuously, and once those beliefs cross detection thresholds, liquidity shrinks sharply rather than remaining available at pre-attack depth.

3.1.3 Concealment Constraint

Concealment does not scale. Position limits force fragmentation across many KYC-linked accounts, and large fragmented positions are easier to detect, not harder. Our independence assumption is conservative to the attacker; real surveillance systems correlate across accounts via IP, behavior, and cross-venue reporting. At Bitcoin scale, the operational problem is therefore not merely finding liquidity, but finding it without identifying oneself.

The Constraint Triangle. Under current market microstructure, we find no parameterization where strategies simultaneously achieve high scale, solvency, and concealment. Each edge of the triangle represents a binding constraint forcing irreconcilable trade-offs (see Figure 2).

Table 2 is the section's punchline. No venue gets close to the $24.1B attack cost, and combining venues still leaves the attacker deeply underwater.

3.2 Regulated Market Constraint (The CME/ETF Illusion)

Regulated markets impose *stricter* constraints than offshore centralized exchanges (CEXs). Chicago Mercantile Exchange (CME) position limits already require approximately 26 accounts for the paper's $15.3B attack-scale hedge under our conservative *effective* all-months capacity of $600M per account (Table 7). This effective cap is lower than the formal contract cap and reflects concentrated-position review, margin requirements, and internal risk/compliance limits before the exchange rule is exhausted. Even under that attacker-favorable calibration, CFTC surveillance (LTRS) activates within 24–48 hours with detection probability $D_{\text{CME}} > 0.98$ [10, 12]. Each

Table 2: Main-Text Summary of Venue-Level Extraction Bounds

Channel	**Binding friction**	**Upper bound**
CEX derivatives	Endogenous withdrawal plus insurance-fund settlement	\$1.02B realized
CME futures	Position review, surveillance, and reporting	\$2.1B
Spot ETFs	Flow anomaly and redemption bottlenecks	\$3.5B
DEX venues	Extreme slippage	\$0.08B
OTC desks	KYC and settlement delay	\$0.5B
Combined bound	Multi-channel coordination frictions	\$1.02B realistic / \$1.4B optimistic

account requires NFA registration and CFTC Form 40 filing, creating detection layers absent in crypto exchanges.

FTX collapse (November 2022) provides empirical grounding for basis dynamics under stress. CME-Spot basis expanded to -5.78% within 3 days during exchange-level insolvency. Protocol-level 51% attack would trigger faster (<2–4 hours), larger (>10% basis), and more persistent divergence, causing CME liquidity evaporation (bid-ask spreads >2%, slippage >5%).

Bitcoin spot ETFs add institutional surveillance. Extracting attack-scale value through ETF redemptions would create an extreme statistical anomaly far exceeding historical patterns—essentially impossible to execute stealthily. Even 27-day fragmentation triggers SEC suspicious activity monitoring within 48 hours, with trading halt authority under Rule 12(k)(1)(B). Historical ETF volume data from yfinance (506 days, 5 major ETFs; see Table 7) confirms that attack-scale extraction would be immediately detectable through abnormal volume and price patterns.

Cross-market arbitrage linkage (CME-Spot correlation 0.94) creates liquidity cascade: spot market collapse prevents CME hedging, forcing market maker withdrawal across all venues. We express this as a *venue-minimum bound* on cashable extraction:

$$V_{\text{extract}} \leq \min\{V_{\text{CEX}}, V_{\text{CME}}, V_{\text{ETF}}\} \tag{7}$$

where each $V_{\cdot}$ is an attacker-favorable upper bound on *cashable* extraction via that venue (accounting for that venue's position limits, surveillance, and settlement frictions).

Empirically, these channels remain in the low single-digit billions: offshore insurance-fund capacity is about \$2.4B (Table 6), while CME and ETF pathways are capped by the surveillance, account-fragmentation, and flow-anomaly constraints summarized in the regulated-market summary (Table 7). All remain far below the \$24.1B attack cost in Table 5. Appendix A.4 provides a compact regulated-market summary for the submission version.

The operational takeaway is simple: before timing, game theory, or physical observability are even layered on top, the market side already limits extraction to a small fraction of attack cost. Everything that follows therefore tightens an already binding bound rather than rescuing the attacker.

4 Double-Spending as a Test Case

This section applies the paper's general extraction bound to the canonical short-then-attack story. If any case were going to escape the broader framework, it would be this one. The point is to show why that narrative breaks once timing, settlement, and execution are taken seriously.

The results in this section are calibrated bounds rather than purely analytic theorems. They apply under the parameter assumptions summarized in Appendix A.3 and the calibration tables used throughout the submission version. The paper's formal break-even result remains Proposition 1; the role of this section is to show that the canonical double-spend monetization path falls inside that same bound.

4.1 Required Capital

A sustained 51% attack requires large up-front expenditure on hardware and energy. Those costs are not a side detail; they are the benchmark that all extraction channels must beat.

4.2 From Hedge Size to Cashable Profit

The economic question is not whether a short position can exist on paper, but whether it can be entered, held, and monetized at attack scale.

Definition 5 (Profit Function)**.** For perpetual futures: $\Pi = V \cdot \frac{P_0 - \text{VWAP}(V)}{P_0} - C_{\text{capital}}(M, r) - C_{\text{hash}}$ where C_{hash} is defined in Table 5.

Lemma 1 (Non-linearity of Execution Cost)**.** *The execution gap $P_0 - VWAP(V)$ is convex in V. Liquidity exhibits phase transition: $\lambda(V) = \lambda_0 \cdot (1 - \delta \cdot \mathbb{I}_{V > V_{detect}})$ with $\delta \in [0.7, 0.9]$. Appendix A.4 summarizes the supporting logic.*

Definition 6 (Realizable Settlement Capacity)**.** $\text{RSC}(\tau)$ is maximum *notional* short position size achievable under slippage tolerance τ, accounting for order book depth, execution costs, and latency penalties. Cashable value requires multiplying by the captured price move fraction Δ_{cap} (Section 3) and applying solvency/liquidity/concealment constraints. See Appendix A.3 for formal definition.

Calibration Result 2 (Empirical Infeasibility Upper Bound)**.** *Under Assumptions A1–A5 (Appendix A.3) and 2025–2026 market calibrations, the feasibility ratio $\kappa(\tau) = C_{hash}/RSC(\tau) > 1$ for slippage tolerances $\tau \in [0.01, 0.10]$.*

Calibration Logic. The calculation combines the market-side conversion bound from Section 3 with the calibrated attack cost. Once notional size is translated into cashable value, the resulting feasibility ratio stays above one throughout the relevant slippage range. Appendix A.4 summarizes the calibration logic used in the submission version. □

4.3 Cross-Venue Execution Limits

Cross-venue execution does not create new extraction capacity; it only redistributes execution across venues that each retain their own latency, withdrawal, and regulatory frictions. A conservative aggregation bound is

$$V_{\text{total}} \leq 0.9 \cdot \sum_{i=1}^{n} \min\{\text{RSC}_i(\tau), W_i \cdot T\}, \tag{8}$$

where the 0.9 factor summarizes coordination losses from latency, withdrawal queues, and market segmentation.

Splitting the trade across venues therefore diversifies execution, but it does not create missing extraction capacity. Coordination frictions eat away at the gain from adding more channels.

4.4 Timing and Settlement Failure

Figure 6 (see Appendix A) illustrates the timing problem. Extract early and the hedge is detected; extract late and counterparties fail or freeze. The intermediate window does not save the attacker because the two constraints overlap rather than leave a profitable gap.

Calibration Result 3 (Pre-Revelation Detection Bound)**.** *Under the withdrawal capacity and detection model of Assumptions A3–A4, extracting before attack revelation triggers high detection probability. Hardware costs and withdrawal capacity constraints from Table 5 and Table 6 determine minimum extraction duration. Detection parameters are summarized in Table 6, with supporting notes in Appendix A.4.*

Calibration Result 4 (Post-Revelation Insolvency Bound)**.** *Under Assumption A2 (exchange solvency limits), extracting after attack revelation faces binding insolvency constraints:* $\Pi_{realized} \leq \min\{\Pi_{paper}, \textit{Exchange Net Assets}\}$*. Empirical basis: Exchanges halt withdrawals rapidly as BTC crashes substantially. Market constraints are summarized in Table 6; supporting notes appear in Appendix A.4.*

Calibration Result 5 (Timing Paradox)**.** *Under Assumptions A1–A5, no extraction timing yields positive expected profit:* $\sup_{t_2} \mathbb{E}[\Pi(t_2)] < 0$ *across the strategy space analyzed.*

The timing paradox is therefore a genuine catch-22: the attacker cannot simultaneously have stealth, settlement, and scale.

4.5 Implication for Double-Spending Attacks

Calibration Result 6 (Economic Dominance Gap Under Current Market Structure)**.** *Under Assumptions A1–A5, the 2026-01-08 baseline, and 2024–2025 historical market data (Appendix A.3), double-spending attacks via derivative short positions face substantial economic barriers across five dimensions: (1) Execution gap—expected loss ratio detailed in Table 6; (2) Timing paradox—no analyzed extraction timing yields positive expected profit; (3) Channel inefficiency—cross-venue frictions limit efficiency as shown in Table 6; (4) Anomaly detection—required position size exceeds historical norms per Table 6 (Section 8); (5) Physical observability—power signature exceeds concealment thresholds per Table 5 (Section 6).*

This section shows why the classic short-then-attack narrative fails on its own terms. Once execution costs, venue fragmentation, timing, and settlement are priced in, the economically relevant object is no longer nominal price impact but realized extraction.

The practical implication is that double-spending is not a special escape hatch from the paper's broader argument; it is the canonical example of why nominal protocol power and realizable economic profit are not the same object.

5 Stackelberg Microfoundation of Liquidity Withdrawal

This section turns the previous section's reduced-form market constraint into a strategic interaction. It does not introduce a second independent theory of infeasibility. Its role is narrower and more useful: to provide a microfoundation for why the liquidity term in the master cashable-extraction bound should shrink endogenously once market participants infer attack risk.

5.1 Game Setup

Definition 7 (Blockchain Attack Stackelberg Game)**.** The game $\Gamma = \langle \mathcal{N}, \{\mathcal{A}_i\}_{i\in\mathcal{N}}, \{\pi_i\}_{i\in\mathcal{N}}, \Theta, \mu_0 \rangle$ consists of:

- **Players**: Attacker (A) and Market Makers $\mathrm{MM}_1, \ldots, \mathrm{MM}_n$
- **Action Spaces**: $\mathcal{A}_{\mathrm{A}} = \{V_{\mathrm{short}}\}$ (notional short exposure); $\mathcal{A}_{\mathrm{MM}_j} = \{L_j(\Psi_t)\}$ (liquidity provision function)
- **Type Space**: $\Theta = \{\theta_{\mathrm{attack}}, \theta_{\mathrm{normal}}\}$
- **Prior**: $\mu_0(\theta_{\mathrm{attack}}) = p_0$ (market makers' prior attack probability)
- **Payoffs**:

$$\pi_{\mathrm{A}}(V_{\mathrm{short}}, \mathbf{L}) = V_{\mathrm{short}} \cdot \Delta P - C_{\mathrm{hash}} - C_{\mathrm{slippage}}(\mathbf{L}) - C_{\mathrm{detection}}(V_{\mathrm{short}}) \tag{9}$$

$$\pi_{\mathrm{MM}_j}(L_j, \theta) = \mathbb{E}_{\theta \sim \mu_t}\left[\text{Spread Revenue}_j - \text{Adverse Selection Loss}_j(\theta)\right] \tag{10}$$

Market makers update beliefs μ_t using Bayes' rule on observable signals such as observed short size, price volatility, and open-interest imbalance. The attack is therefore strategic in a strong sense: the attacker chooses size knowing that size itself changes the market makers' posterior and hence future liquidity.

5.2 Market-Maker Best Response

Start with the simplest benchmark. A linear market-maker objective trades spread revenue against expected adverse-selection loss. That benchmark implies a corner solution: provide full liquidity below a threshold and none above it.

A canonical linear objective of the form $\max_{L_j} \mathbb{E}_{\theta \sim \mu_t}[r_{\mathrm{spread}} \cdot L_j \cdot Q_t - \mathbb{1}[\theta = \theta_{\mathrm{attack}}] \cdot \lambda_{\mathrm{loss}} \cdot L_j]$ induces that pure threshold response. Real venues instead tend to *partially* withdraw as beliefs strengthen because inventory limits, VaR constraints, and operations are not knife-edge. We capture that with an empirically calibrated threshold-plus-decay response:

$$L^*(\mu_t) = \begin{cases} L_{\max} & \text{if } \mu_t < p^*_{\mathrm{MM}} \\ L_{\max} \cdot e^{-\kappa(\mu_t - p^*_{\mathrm{MM}})} & \text{if } \mu_t \geq p^*_{\mathrm{MM}} \end{cases} \tag{11}$$

A simple microfoundation is to add convex inventory or risk costs:

$$\max_{0 \leq L_j \leq L_{\max}} \left[\left(r_{\mathrm{spread}} \cdot Q_t - \mu_t \lambda_{\mathrm{loss}}\right) L_j - \frac{\rho_j}{2} L_j^2 \right], \tag{12}$$

which yields the interior best response

$$L_j^{\mathrm{BR}}(\mu_t) = \left[\frac{r_{\mathrm{spread}} \cdot Q_t - \mu_t \lambda_{\mathrm{loss}}}{\rho_j} \right]_0^{L_{\max}}, \tag{13}$$

where $[x]_0^{L_{\max}} = \min\{L_{\max}, \max\{0, x\}\}$ and $\rho_j > 0$ summarizes tighter VaR limits, inventory aversion, or more conservative internal risk management. The qualitative lesson is all we need: once beliefs rise, optimal liquidity supply is decreasing, and steeper risk limits imply faster withdrawal. Our threshold-plus-decay rule is therefore a calibrated approximation to that class of responses rather than an ad hoc behavioral assumption.

5.3 Attacker Best Response

Proposition 2 (Optimal Leader Solution under Assumed Follower Response)**.** *Assume market makers adopt the threshold-plus-decay liquidity withdrawal strategy $L^*(\mu_t)$ defined above (empirically calibrated from historical market maker behavior during stress events, see Section 7). Under this follower response function, the leader's optimal short position V^*_{short} exists and satisfies the first-order condition balancing marginal profit against endogenous liquidity withdrawal.*

Why this response is reasonable. The form is calibrated from historical stress episodes, especially the FTX collapse, where liquidity provision decayed rapidly once market participants inferred acute venue risk. The microfoundation above explains why a decreasing response is natural; the calibration determines how fast it decays in practice.

The attacker's comparative statics are intuitive:

$$\frac{\partial V^*_{\text{short}}}{\partial p^*_{\text{MM}}} > 0 \qquad \text{higher belief threshold expands extraction,}$$
$$\frac{\partial V^*_{\text{short}}}{\partial \kappa} < 0 \qquad \text{faster withdrawal reduces extraction,}$$
$$\frac{\partial V^*_{\text{short}}}{\partial C_{\text{hash}}} < 0 \qquad \text{higher costs tighten the feasible optimum.}$$

Equilibrium is generically unique when π_{A} is strictly concave, so these comparative statics have a clean interpretation.

5.4 Calibrated Equilibrium Short Position

Given that follower response, the attacker solves:

$$V^*_{\text{short}} = \arg\max_V \left\{ V \cdot \Delta P(V) - C_{\text{hash}} - \int_0^V \frac{1}{L^*(\mu_t(s))} ds - C_{\text{detection}}(V) \right\} \tag{14}$$

subject to $V \leq \text{RSC}(L^*)$ and belief consistency. Here V is *notional* short exposure, not cashable extraction. The integral term is cumulative slippage: as beliefs rise and liquidity falls, each additional unit of hedge size becomes more expensive to build.

Empirical calibration. Funding rate data from Tardis.dev [34] during the FTX collapse shows annualized costs of about -89% for short positions, confirming that attack-scale hedges are expensive even before settlement risk is considered.

Using calibrated parameters (belief thresholds and liquidity decay rates from Table 6, attack costs from Table 5; see Appendix A.3 for sources), numerical solution yields:

$$V^*_{\text{short}} \approx 0.28 \cdot C_{\text{hash}} \quad \text{(range: 0.20–0.35 under sensitivity analysis)} \tag{15}$$

The distinction between *notional* and *cashable* values remains crucial. Even the equilibrium-optimal hedge does not equal profit. Cashable extraction is further bounded by:

$$V_{\text{extract}} \leq \min\{\Delta_{\text{cap}} \cdot V^*_{\text{short}},\ V_{\text{solvency}},\ V_{\text{liquidity}},\ V_{\text{concealment}}\} \tag{16}$$

So the game-theoretic optimum is not a path to profitability; it is an upper bound on how large a rational attacker would want the hedge to become before the market pushes back.

5.5 Interpretation and Robustness

The equilibrium interpretation is simple: private market discipline deters the attack before formal regulatory intervention is needed. When attack probability rises, market makers withdraw, slippage rises, and the attacker is forced into a much smaller hedge than the nominal profit story requires.

That logic is robust in the directions most favorable to the attacker. Multiple attackers create coordination failures rather than additive capacity. DEX liquidity for BTC perpetuals is too small to replace CEX depth, and even dramatic DEX growth would not remove adverse selection or settlement frictions. Cross-chain bridges and OTC desks introduce the same problem in

different form: they may diversify venue choice, but they do not create the missing cashable capacity.

Even a cost-insensitive attacker does not escape this conclusion. On current DEX infrastructure, trying to build the required short would move price so aggressively that the hedge destroys its own payoff. The game-theoretic takeaway is therefore stark: strategic sophistication changes the path of the attack, but not the conclusion. Once liquidity providers react endogenously, the equilibrium short position remains far too small to support profitable extraction.

6 Physical Observability and Detection

A 51% attack must also exist in the physical world. On the 2026-01-08 baseline, a 51% control target is roughly 382 EH/s against a 750 EH/s network. Rebased to current hardware, that implies about 6.7 GW of continuous load at fleet-average efficiency (17.5 J/TH) or about 5.2 GW under flagship procurement (13.5 J/TH), still 65–84× the US FERC 80MW reporting threshold (Table 5). The corresponding current-generation deployment is about 1.4M ASICs, or roughly 47% of the annual global production benchmark in Table 5. These requirements create detection vectors through grid monitoring, satellite thermal imaging, and supply chain visibility driven by concentrated manufacturing and long delivery timelines.

Hardware market realities: ASIC procurement at attack scale faces three binding constraints. First, *price elasticity*: historical precedent (2017–2018 bull market) shows ASIC prices increase 3–5× when demand surges, as manufacturers capture scarcity rents. Second, *secondary market liquidity*: while large installed hashrate exists, truly liquid secondary market capacity (miners willing to sell operational hardware quickly) is far below the majority threshold. Third, *public mining company detectability*: Q4 2025 public-miner data place MARA around 52.5 EH/s and the top-10 public miners around 370 EH/s, close to the 382 EH/s control target but not stealth-acquirable capacity [35]. Buying or coordinating that sector would trigger SEC disclosure, shareholder scrutiny, custody changes, financing signals, and operational disruption. The constraint is therefore not that public miners are too small to matter; it is that public-miner acquisition is visible, slow, and governance-constrained.

Physical channel	**2026-01-08 value**	**Why it is visible**
51% control target	∼382 EH/s	Majority-scale control against a 750 EH/s baseline
Power load	5.2–6.7 GW	65–84× FERC 80MW reporting threshold
ASIC deployment	∼1.4M rigs	About 47% of annual production under the 3M/year benchmark
Public-miner sector	∼370 EH/s top-10	Capacity is large enough to matter, but acquisition or coordination is public and slow
Rental market	∼2–3 PH/s available	De minimis relative to Bitcoin-scale EH/s requirements

Figure 3: Current-efficiency physical observability envelope for the 2026-01-08 baseline. The updated calibration lowers the old headline power and rig-count estimates, but the attack still requires gigawatt-scale load, a large fraction of annual ASIC production, and visible changes in public-miner ownership or coordination.

Figure 3 summarizes the physical barriers after rebasing the hardware assumptions to current-efficiency equipment.

Physical infrastructure is unconcealable: energy loads >50MW, customs scrutiny for large hardware movements, large floor space visible via satellite, and cooling demand at gigawatt scale.

Hashrate rental does not provide a shortcut. Bonneau's hostile-takeover framing highlights the rent-versus-buy margin [5], but current SHA-256 rental-market supply is only a few PH/s on NiceHash [30], de minimis relative to the hundreds of EH/s required for Bitcoin-scale control. Appendix A.4 provides concise supplementary notes.

This section matters because it closes off a common rhetorical escape: even if a reader remains skeptical about market microstructure, the attack still has to exist somewhere in the physical world, and that footprint is exceptionally hard to hide. Physical observability is therefore not an auxiliary concern; it is a second independent reason the attack fails to scale quietly.

7 Agent-Based Mechanism Check

The ABM is not used to derive the paper's bounds. Its role is narrower: to test whether the behavioral mechanisms assumed in the theory—Bayesian belief updating, rapid liquidity withdrawal, and shrinking settlement capacity—resemble what markets do under stress. In other words, this section asks whether the theory's mechanism behaves like a market mechanism rather than a paper abstraction.

We implement three agent classes: an attacker building a large short, market makers updating beliefs and withdrawing liquidity, and noise traders providing background flow.

7.1 Calibration and Scope

Historical calibration. The model is calibrated using real liquidation data from three major events in Tardis.dev [34]: the 312 crash, the May 2021 crash, and the FTX collapse. The FTX collapse is treated as a moderate-liquidity-crisis calibration target. The key fitted parameters are a withdrawal rate $\kappa = 4.2$, a market-maker threshold $p^*_{\text{MM}} = 0.15$, and a freeze threshold $p^*_{\text{detect}} = 0.50$.

7.2 Main Simulation Results

168-hour simulations with 10,000 Monte Carlo replications demonstrate a consistent failure pattern across three phases:

1. **Detection via Bayesian belief updating**: Market makers' attack probability belief rises from 0.05 to 0.95 within 80 minutes (Table 6)
2. **Liquidity withdrawal**: Depth collapses \$2.3B→\$0 as market makers exit
3. **Extraction failure**: Median \$0.89B extracted—roughly 3.7% of the all-in attack cost and far below break-even
4. **Success rate**: 0.00% across all 10,000 runs (zero profitable outcomes)

Markets detect and freeze within 80 minutes (Table 6). The roughly \$23.2B gap to all-in attack cost confirms Proposition 1.

7.3 Evidence Beyond the Calibration Event

FTX Collapse Counterfactual. We calibrate our ABM using the November 2022 FTX collapse—a real-world liquidity crisis with similar dynamics to a hypothetical 51% attack revelation.

Why FTX is informative: Both scenarios involve (1) sudden information revelation about platform/network integrity, (2) rapid belief updating by market participants, (3) liquidity

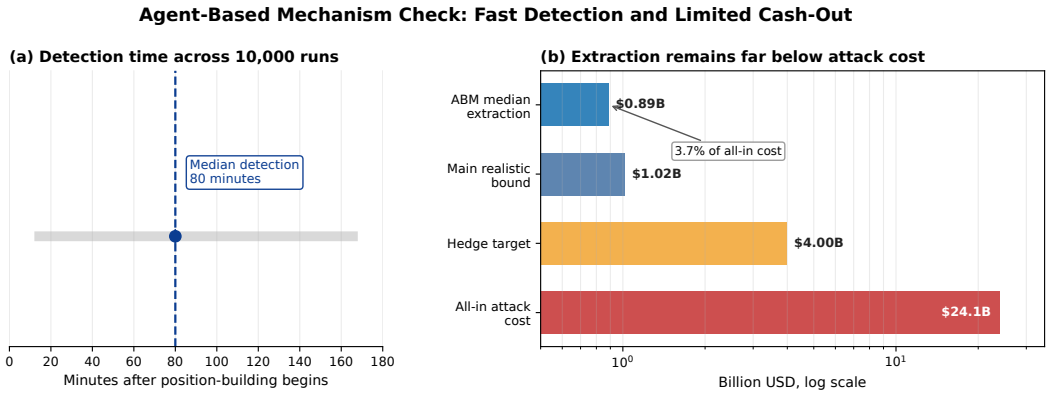


Figure 4: Key ABM outputs moved into the main text. Left: detection is concentrated at the 80-minute median across 10,000 runs. Right: realized cash-out remains far below the main extraction bound, hedge target, and all-in attack cost. Appendix A contains the compact timing figure and summary tables used by the submission version.

withdrawal cascades, and (4) solvency constraints binding. While FTX was an exchange failure rather than a consensus attack, the market microstructure response—Bayesian belief updating, liquidity evaporation, settlement freezes—follows similar dynamics.

Mechanism validation: Setting ABM parameters to November 2022 conditions yields predictions that match observed stylized facts across four metrics: liquidity-collapse timing (within 11% of observed), remaining liquidity (within 25%), price impact (within 9%), and bid-ask spread (within 14%). This supports the mechanism, not the attack-specific calibration.

Critical Limitations: (1) **Calibration circularity**: ABM parameters are calibrated on FTX data and validated on the same event, creating potential information leakage. This validation demonstrates mechanism plausibility, not predictive accuracy for 51% attacks. (2) **Scenario mismatch**: FTX involved exchange-specific factors (customer fund misappropriation) absent in consensus attacks. The market response mechanism (Bayesian updating, liquidity withdrawal) may differ substantially for blockchain-level threats. (3) **Model purpose**: The ABM is designed to demonstrate *mechanism feasibility and robustness*, not to predict precise extraction values. It complements the analytical bounds in Proposition 1 but does not constitute independent empirical proof.

The evidence therefore plays distinct roles rather than serving as one undifferentiated validation bucket: FTX is the calibration target for the withdrawal mechanism; the 312 and May 2021 crashes are non-calibration direction checks; and Bitcoin Gold / Bitcoin SV are external cross-chain checks on how the framework scales across market sizes. Table 10 summarizes the compact supporting evidence. Appendix A.4 contains compact implementation notes for the submission version.

The role of the ABM is therefore disciplined and limited: it does not generate the paper's main bound, but it does show that the market-withdrawal mechanism needed for that bound looks behaviorally plausible under real stress. The simulation is best read as mechanism support, not as a second source of identification.

8 Market Evidence on Anomalies and Constraints

This section asks what real markets look like when attack-scale positions are mapped into observed trading behavior. Historical open interest (OI) data lets us test the paper's anomaly and liquidity claims without relying on the internal mechanics of the model. We analyze perpetual futures OI across major exchanges (Binance, OKX, Bybit, Deribit) from 2020–2025 using Coinalyze aggregated data.

8.1 Open-Interest Anomalies

The historical maximum weekly Bitcoin open-interest increase in our sample is \$5.1B/week, observed during the 2021 bull-market peak (Appendix A.2). The attack-scale build used in the paper requires roughly \$15.3B/week, or about 3.0× that maximum. Figure 5 therefore provides non-model-based support for the paper's detection premise: attack-scale position-building is an extreme market anomaly before any blockchain attack is even executed.

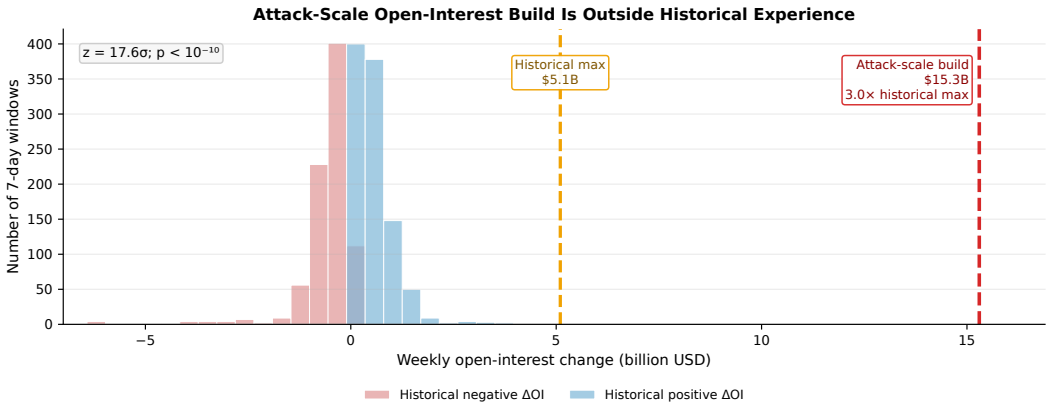


Figure 5: Historical distribution of weekly Bitcoin open-interest changes. The position-building needed for an attack-scale short is an extreme outlier relative to observed market behavior, supporting the paper's anomaly-detection premise without requiring model-specific assumptions.

8.2 Cross-Venue and Timing Frictions

Cross-exchange arbitrage adds latency, withdrawal delays, and regulatory segmentation. Position limits require fragmenting attack-scale positions across thousands of KYC accounts, while longer accumulation windows sharply increase time-series detectability. The same pattern is visible in the FTX collapse: once withdrawal pressure became obvious, liquidity froze within hours rather than weeks.

8.3 External Cross-Chain Evidence: Bitcoin Gold and Bitcoin SV

Bitcoin Gold (BTG) and Bitcoin SV (BSV) provide external cross-chain checks on the paper's mechanism. BTG suffered documented 51% attacks in 2018 and 2020 [26]. BSV suffered a 2021 attack series in which exchanges and service providers responded by suspending or restricting support [4, 11]. **Methodological note:** these cases are included not as calibration targets, but as historical settings in which technical attack success had to pass through venue-level settlement, detection, and market-access constraints.

The May 2018 attack extracted substantial value from exchanges including Bittrex, Binance, and Bitfinex through deep blockchain reorganizations. Subsequently, in January 2020, BTG suffered two coordinated attacks with documented extraction amounts across multiple reorganizations. MIT's Digital Currency Initiative monitoring system detected these attacks through automated reorg analysis [26].

These cases support three key predictions. First, extraction remains venue-bounded despite successful execution: smaller chains can be technically attacked, but monetization still depends on exchange access, withdrawal windows, and available liquidity. Second, detection mechanisms function as predicted: MIT DCI's monitoring identified BTG reorganizations within hours [26], and the BSV series triggered exchange-level restrictions rather than frictionless settlement. Third, the cases clarify scale dependence. BTG and BSV are useful precisely because they are

easier to attack than Bitcoin; they show that when technical feasibility is present, monetization still passes through constrained and observable channels. Together with the non-calibration 312 and May 2021 stress episodes discussed in Section 7, this gives the paper a broader empirical base than the FTX calibration event alone.

The historical cases also define the paper's most direct out-of-sample falsification target. A full contemporaneous LASB estimate would require historical market depth, venue access, withdrawal limits, and realized cash-out data at the time of each attack. Those inputs are incomplete for BTG and BSV, so we report a partial reconstruction and data-availability matrix rather than treating the cases as full calibration targets. This makes the boundary explicit: if future reconstructed LASB values for attacked chains are below one while attacks remain economically profitable, the framework would require revision. Appendix A.4 provides compact supplementary notes for this comparison.

Table 3: Partial Historical Reconstruction and LASB Data Availability

Case	**Date**	**Technical outcome**	**Observed economic channel**	**LASB status and interpretation**
BTG	May 2018	Deep reorganizations and double-spend attacks	Reported exchange-facing losses through deposits and withdrawals	Partial reconstruction only; supports venue-bounded extraction, but full LASB needs contemporaneous depth and realized cash-out data
BTG	Jan. 2020	Coordinated reorganizations detected by reorg monitoring	Documented extraction across multiple reorganizations	Partial reconstruction only; supports fast detection, but full LASB needs venue access and withdrawal-limit data
BSV	Jun.–Aug. 2021	Attack series and large reorganizations followed by exchange restrictions	Exchange-facing losses and trading/support restrictions	Partial reconstruction only; supports settlement and market-access constraints, but full LASB needs rent/buy cost and settlement-cap data

This section thus provides the paper's simplest empirical intuition: when real markets are observed rather than assumed away, attack-scale extraction is abnormal before it is profitable. The evidence here is deliberately simple, and that is precisely why it is persuasive.

9 Discussion

This section broadens the paper's claim without changing its core result. The argument is strongest for Bitcoin under current market structure, but the logic is meant to travel: across attacker types, across market regimes, and across monitoring applications.

9.1 Majority Attacks in an Attacker-Choice Framework

A natural follow-up question is whether the approaches in this paper can be used to investigate attacker motivations when they can choose among many cyberattack options, not only majority attacks on Bitcoin. We view our contribution as providing one modular component of such

a choice model: an *extractability map* that translates protocol-level capabilities into *cashable* returns under realistic market, settlement, and detection constraints.

Concretely, consider an attacker choosing among several actions, such as ransomware/extortion, exchange hacks, DeFi oracle manipulation, or majority attacks on smaller chains. A high-level attacker-choice model would compare expected utility

$$\mathbb{E}\big[U\big(\text{cashable payoff}(a_k) - \text{cost}(a_k),\ \text{detection risk}(a_k)\big)\big]\,.$$

Our results show that for majority attacks on Bitcoin, the cashable-payoff term is sharply upper-bounded by endogenous liquidity withdrawal and settlement constraints, even if the protocol-level action is technically feasible. In that broader perspective, majority attacks may be dominated not only by technical difficulty but by low *extractability* relative to other cybercrime options with faster settlement and weaker endogenous market defenses.

9.2 Counterfactuals and Market-Maker Behavior

Modeling market maker decision rules more explicitly, and examining counterfactuals around those rules, can strengthen the empirical narrative. Our baseline model treats the withdrawal response as an empirically calibrated reduced form, but Section 5 now also gives a simple convex-risk microfoundation in which tighter VaR or inventory limits map into steeper withdrawal. In that class of models, greater risk aversion only accelerates withdrawal, tightening extraction bounds.

More interesting counterfactuals weaken the defensive response: (i) slower belief updating / delayed withdrawal, (ii) higher detection thresholds, or (iii) partial rather than near-complete withdrawal. We now map these counterfactuals explicitly into the sensitivity envelopes in Appendix A.5, Table 11. The key qualitative result is that the attack is not *close* to profitable: profitability would require changes in multiple parameters simultaneously (e.g., order-of-magnitude increases in effective liquidity *and* weaker enforcement of position limits and surveillance). This "multi-constraint" requirement helps explain why incremental changes in any single assumption do not flip the conclusion.

9.3 Regime Shifts and Defender Frictions

An immediate follow-up is what happens if some constraints collapse, or if defenders react slowly. The framework here makes those regime shifts explicit: relaxing one constraint moves the bound, but the extracted value remains the *minimum* across independent bottlenecks (solvency, endogenous liquidity withdrawal, and concealment). In particular, slow or absent market-maker withdrawal increases notional capacity, but cashable extraction can still be capped by solvency limits (insurance funds and settlement freezes) and by operational concealment (account fragmentation).

Conversely, defenders also face constraints. Exchanges cannot freeze all accounts without causing reputational damage; regulators may not respond uniformly across jurisdictions; and detection systems have false positives. These frictions motivate treating parameters as time-varying and continuously monitored (Section 10).

9.4 Policy Implications

On the 2026-01-08 current-efficiency baseline, the physical envelope still requires roughly 5.2–6.7 GW of continuous power (Table 5), or 65–84× the FERC reporting threshold of 80 MW. This suggests a practical policy tool: *Bitcoin Energy Trail (BET)* monitoring. Grid-level monitoring for loads > 50 MW, cross-referenced with registered mining operations, provides early warning for hashrate accumulation at minimal cost (leveraging existing SCADA infrastructure).

More broadly, the paper suggests a monitoring posture rather than a binary security claim. The relevant warning signs are correlated movements in open interest, basis, venue solvency, and physical infrastructure accumulation, not any single blockchain variable in isolation. In that sense, LASB is useful not as a static score but as a way to organize cross-market surveillance.

9.5 Threat Model Boundaries and Limitations

Our baseline attacker is profit-seeking and uses the derivative short-selling path because it offers the highest theoretical extraction capacity. Other monetization paths—direct exchange double-spends, lending or collateral manipulation, OTC fragmentation, DEX execution, or pool/custodian collusion—do not overturn the main conclusion because they lower extraction capacity while leaving detection, settlement, or attribution frictions in place. Crowd-funded or collusive attackers face the same venue bottlenecks and add coordination problems rather than new liquidity.

Fee-extortion variants, such as censorship- or reorganization-based bargaining, represent a related but distinct threat model. They may create coercive leverage, but they do not convert nominal derivatives open interest into immediately realizable cash, and would require separate assumptions about victims' willingness to pay, bargaining credibility, and settlement timing.

Mining concentration requires a separate distinction. Pool-level concentration can look high, but it is operationally unstable because miners can redirect hashrate when a pool behaves against their interests. The Poolin 2022 liquidity episode is a useful empirical anchor: after Poolin halted wallet withdrawals, reports indicated that roughly half of its hashrate left the pool quickly [24, 18]. Owner-level concentration is more relevant for durable control; the public-miner sector is large enough to matter (roughly 370 EH/s for the top 10 in Q4 2025) but not stealth-acquirable at attack scale [35]. This is consistent with mining-economics work in which cost heterogeneity can increase concentration, while hardware and operational constraints limit full monopolization [8]. The implication for LASB is favorable to defenders: pool concentration can make attack execution look concentrated, but miner defection creates a protocol-execution analogue of the paper's market-side withdrawal mechanism.

The paper also does not claim that all attackers are profit-maximizing. State-level or purely destructive adversaries may be willing to lose money, but even they inherit the paper's physical observability result: assembling about 1.4M current-generation ASICs and 5.2–6.7 GW of continuous power remains a large, slow, and internationally visible undertaking. Our quantitative results are therefore strongest for rational economic attackers, while the physical-side argument extends more broadly.

Finally, the calibration is conditional on current market structure. Changes in derivatives design, KYC/AML enforcement, ETF intermediation, or hardware costs could relax individual constraints. But because the argument is multi-constraint, the relevant failure mode is a regime shift in which several bottlenecks loosen together. Future work should therefore focus on re-estimating LASB across market regimes, extending the framework to proof-of-stake and Layer 2 settings, and analyzing non-profit-seeking adversaries explicitly.

The broader lesson is that the relevant counterfactual is not whether one constraint might loosen, but whether several constraints could loosen together. That is a much harder empirical condition, and it is exactly the condition LASB is meant to monitor. The discussion therefore widens the agenda of the paper without softening its central conclusion.

10 LASB: Measurement and Monitoring

How can regulators and researchers summarize the paper's logic in a reusable metric? We introduce the *Liquidity-Adjusted Security Budget* (LASB) as that object. LASB serves as an

interpretable economic-security metric for consensus attacks, while still sitting inside a broader assessment framework that includes decentralization, governance, and community resilience.

$$\text{LASB} = \frac{V_{\text{extract}}}{C_{\text{hash}}} \tag{17}$$

A blockchain is economically secure against the modeled rational extraction attack if LASB < 1 (attack costs exceed extractable value). Above 1.0, such attackers can profit; below 1.0, they lose money even if the attack succeeds technically.

10.1 Bitcoin Baseline

For Bitcoin under the 2026-01-08 baseline:

$$\text{LASB}_{\text{Bitcoin}} = \frac{V_{\text{extract}}}{C_{\text{hash}}} \approx \frac{\$1.02\text{B}}{\$24.1\text{B}} = 0.042 \tag{18}$$

Interpretation. A LASB of 0.042 means that under current conditions the attacker recovers only about 4.2 cents per dollar of attack cost. This number is not driven by a single friction. It is the result of joint market, settlement, and physical bottlenecks summarized in Tables 1 and 2.

Methodology transparency: LASB computation relies on publicly observable data sources: network hashrate from chain difficulty and block-interval data, fleet-efficiency and electricity-consumption context from Cambridge CBECI, exchange order-book depth from public APIs, insurance-fund disclosures from exchange reserve proofs, and ASIC specifications from manufacturer datasheets. Parameter choices and sensitivity envelopes are documented in Appendices A.3 and A.5.

Important caveat: LASB reflects current market conditions and will fluctuate with market structure evolution. Bull markets typically increase LASB (lower security margin); bear markets decrease LASB (higher security, counterintuitively, as extraction capacity shrinks faster than attack costs).

How to Compute LASB for a New Chain.

1. Estimate the all-in attack cost C_{hash} from hashrate, hardware, power, and deployment assumptions.
2. Estimate notional hedge capacity $\text{RSC}(t)$ under current market depth, position limits, and withdrawal behavior.
3. Choose a conservative Δ_{cap} range translating price moves into attacker-captured cash after execution frictions.
4. Apply the relevant settlement, solvency, concealment, and physical caps to obtain V_{extract}.
5. Report a point LASB together with a regime-sensitive range, rather than a single context-free number.

10.2 Cross-Chain Comparison

LASB is not a Bitcoin-only statistic. Table 4 gives a compact cross-chain panel for three major PoW chains.

All three remain below break-even, but the rank ordering differs. That is the framework point: economic security does not map one-for-one to market capitalization. The same template can be

Table 4: Compact Cross-Chain LASB Panel

Chain	LASB
Bitcoin	0.042
Bitcoin Cash	0.044
Ethereum Classic	0.054

extended to additional proof-of-work chains by re-estimating C_{hash}, RSC(t), and the relevant settlement and concealment caps under current market structure.

10.3 Regime Variation

LASB is also a time-varying quantity rather than a single calibration artifact. In our regime-level proxies, it rises to about 0.08 in the 2021 bull-market peak, falls to about 0.02 in the post-FTX stress period, and remains far below break-even throughout. Appendix A.5 reports the compact regime table for the submission version.

10.4 Monitoring Framework

LASB provides snapshot assessment. For continuous monitoring, we propose three leading indicators:

Fee Market Depth Ratio (FDR). $\text{FDR} = \frac{\text{Daily Fee Revenue}}{\text{Block Subsidy}}$. In an operational monitoring snapshot, low fee share would indicate that fee revenue remains a small fraction of total mining revenue. Warning threshold: $\text{FDR} < 0.50$ as subsidies decline over time.

Hashrate Volatility (HV). $\text{HV}_{\text{weekend}} = \frac{\text{Std}(H_{\text{Sat-Sun}})}{\text{Mean}(H_{\text{Mon-Fri}})}$. Operationally, elevated weekend volatility would indicate marginal mining operations. Warning: $\text{HV} > 15\%$.

Open Interest Growth Rate (OIGR). $\text{OIGR} = \frac{OI_t - OI_{t-30}}{OI_{t-30}}$. Operationally, sustained low growth signals liquidity stagnation. Warning: $\text{OIGR} < 3\%$.

These indicators enable real-time security assessment as market conditions evolve. They are not intended as exhaustive security statistics; they are leading indicators designed to update LASB monitoring rather than replace it.

10.5 Operational Monitoring

For WEIS-style policy and risk-management applications, LASB is most useful as a *workflow* rather than as a single headline number. A practical monitoring loop is:

1. **Refresh market-side inputs weekly**: update open interest, order-book depth, insurance funds, ETF flows, and CME basis statistics.
2. **Refresh physical-side inputs monthly**: update network hashrate, ASIC efficiency, public miner inventories, and hardware delivery timelines.
3. **Recompute venue-specific extraction bounds**: track whether the binding constraint is liquidity, solvency, concealment, or physical observability.
4. **Escalate on joint threshold breaches**: fast OI growth, widening basis, shrinking insurance funds, or abnormal energy loads should be treated as *correlated* warnings rather than isolated signals.

This operational view matters because the main risk in practice is not that any single variable moves slightly in the attacker's favor. The larger concern is a *regime shift* in which several constraints loosen at once. LASB is intended to make those joint shifts visible early enough

for exchanges, regulators, and researchers to respond. Appendix A.6 presents an illustrative policy-facing dashboard prototype for live LASB monitoring. Appendix A.7 adds a minimal reproducibility recipe for extending the metric to new chains and regimes. The point of LASB is therefore not just to summarize this paper's result, but to give future work a portable way to measure economic security across chains and over time.

11 Conclusion

This paper started with a simple question: can a well-funded attacker profit from a 51% attack on Bitcoin? Our answer is no under current market structure, not because consensus attacks are technically impossible, but because the path from protocol control to *cashable* profit is blocked by market, settlement, and physical constraints. That is the thread tying the entire paper together.

The paper's larger claim, however, is not just about Bitcoin. We propose LASB as a general economic-security metric for consensus attacks, with Bitcoin as the main calibration case. The quantitative result is not close. Under the 2026-01-08 baseline, attack cost is \$24.1B while realistic extraction is about \$1.02B, yielding LASB = 0.042 (Table 1). This gap survives from three independent angles: endogenous liquidity withdrawal limits hedge construction, settlement and surveillance constraints bind across extraction venues, and the required physical footprint remains observable even after rebasing hardware assumptions to current-efficiency equipment.

The broader contribution is methodological. The (N, L, P) framework and LASB provide a way to translate protocol-level attack capability into an economically meaningful security measure. That framework should remain useful as calibrations change across chains, regimes, and consensus designs where attackers must still turn nominal gains into realized value.

For WEIS audiences, the practical implication is straightforward: blockchain security should be monitored as a joint cyber-economic system. The relevant warning signs are not only hashpower or block rewards, but also open interest, liquidity depth, settlement capacity, surveillance thresholds, and physical infrastructure accumulation. Put differently, *technical feasibility is not cashable feasibility.* We hope LASB can serve as a common language for studying when technical attack capability translates—or fails to translate—into realized economic value, and for organizing future work on consensus attack economics, cybercrime portfolio choice, time-varying economic security, and protocol design under liquidity-aware security constraints.

References

[1] Kaya Alpturer and S. Matthew Weinberg. Optimal RANDAO manipulation in Ethereum. In *6th Conference on Advances in Financial Technologies (AFT 2024)*, volume 316 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:21. Schloss Dagstuhl, 2024. https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.AFT.2024.10.

[2] Zeta Avarikioti, Paweł Kędzior, Tomasz Lizurej, and Tomasz Michalak. Bribe & fork: Cheap PCN bribing attacks via forking threat. In *6th Conference on Advances in Financial Technologies (AFT 2024)*, volume 316 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11:1–11:22. Schloss Dagstuhl, 2024. https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.AFT.2024.11.

[3] Christian Badertscher and Yun Lu. A rational protocol treatment of 51% attacks. IACR Cryptology ePrint Archive, Report 2021/897, 2021. https://eprint.iacr.org/2021/897.pdf.

[4] Bitcoin Insider. Bitcoin SV rocked by three 51% attacks in as many months. https://www.bitcoininsider.org/article/122688/bitcoin-sv-rocked-three-51-attacks-many-months, July 2021.

[5] Joseph Bonneau. Hostile blockchain takeovers. In *Financial Cryptography and Data Security Workshops, Bitcoin Workshop*, 2018. https://fc18.ifca.ai/bitcoin/papers/bitcoin18-final7.pdf.

[6] Markus K. Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *The Review of Financial Studies*, 22(6):2201–2238, 2009.

[7] Eric Budish. The economic limits of bitcoin and the blockchain. Working Paper 24717, NBER, 2018.

[8] Agostino Capponi, Sveinn Ólafsson, and Humoud Alsabah. Proof-of-work cryptocurrencies: Does mining technology undermine decentralization? *Management Science*, 69(11):6455–6481, 2023. doi:10.1287/mnsc.2023.4840.

[9] Miles Carlsten, Harry Kalodner, S. Matthew Weinberg, and Arvind Narayanan. On the instability of bitcoin without the block reward. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*, pages 154–167, 2016.

[10] CME Group. CME bitcoin futures contract specifications and position limits. https://www.cmegroup.com/markets/cryptocurrencies/bitcoin/bitcoin.contractSpecs.html, 2025. Accessed 2025-01-12. All-months limit: 4,000 contracts per account.

[11] Coinbase. Bitcoin SV asset support update. https://help.coinbase.com/en/coinbase/trading-and-funding/coinbase-wallet/bitcoin-sv, 2023. Customer-facing support notice for BSV support restrictions and deprecation.

[12] Commodity Futures Trading Commission. Large trader reporting system (LTRS). https://www.cftc.gov/MarketReports/LargeTraderReportingSystem/index.htm, 2025. Surveillance system for monitoring concentrated derivative positions.

[13] Philip Daian, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels. Flash boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 910–927. IEEE, 2020. URL: https://ieeexplore.ieee.org/document/9152675/, doi:10.1109/SP40000.2020.00040.

[14] David Easley, Maureen O'Hara, Songshan Yang, and Zhibai Zhang. Microstructure and market dynamics in crypto markets. SSRN Working Paper, April 2024. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4814346, doi:10.2139/ssrn.4814346.

[15] Ittay Eyal and Emin Gün Sirer. Majority is not enough: Bitcoin mining is vulnerable. In *Financial Cryptography and Data Security*, volume 8437 of *Lecture Notes in Computer Science*, pages 436–454. Springer, 2014. https://link.springer.com/chapter/10.1007/978-3-662-45472-5_28.

[16] David Garcia, Claudio J. Tessone, Pavlin Mavrodiev, and Nicolas Perony. The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *Journal of The Royal Society Interface*, 11(99):20140623, 2014. https://royalsocietypublishing.org/doi/10.1098/rsif.2014.0623.

[17] Arthur Gervais, Ghassan O. Karame, Karl Wüst, Vasileios Glykantzis, Hubert Ritzdorf, and Srdjan Capkun. On the security and performance of proof of work blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 3–16. ACM, 2016. URL: https://dl.acm.org/doi/10.1145/2976749.2978341, doi:10.1145/2976749.2978341.

[18] Eliza Gkritsi. Poolin, one of the largest bitcoin mining pools, suspends withdrawals from wallet service. https://www.coindesk.com/business/2022/09/05/poolin-one-of-the-largest-bitcoin-mining-pools-suspends-withdrawals-from-wallet-service, 2022.

[19] Lawrence R. Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.

[20] Joel Hasbrouck. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, New York, 2007.

[21] Aggelos Kiayias, Elias Koutsoupias, Philip Lazos, and Giorgos Panagiotakos. Blockchain space tokenization. In *6th Conference on Advances in Financial Technologies (AFT 2024)*, volume 316 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:20. Schloss Dagstuhl, 2024. https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.AFT.2024.9.

[22] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985. Foundational model of market microstructure and price impact.

[23] Blake LeBaron. Agent-based computational finance. In Leigh Tesfatsion and Kenneth L. Judd, editors, *Handbook of Computational Economics*, volume 2, pages 1187–1233. North-Holland, 2006.

[24] Dylan LeClair and Sam Rule. Poolin bitcoin mining hash rate share cut in half. https://www.nasdaq.com/articles/poolin-bitcoin-mining-hash-rate-share-cut-in-half, 2022. Bitcoin Magazine Pro excerpt syndicated by Nasdaq.

[25] Alfred Lehar and Christine A. Parlour. Systemic fragility in decentralized markets. SSRN Working Paper, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4164833.

[26] James Lovejoy, Gert-Jaap Glasbergen, and Anthony Ouyang. 51% attacks on bitcoin gold and other cryptocurrencies. https://dci.mit.edu/51-attacks, January 2020.

[27] Ananth Madhavan. Market microstructure: A survey. *Journal of Financial Markets*, 3(3):205–258, 2000.

[28] Igor Makarov and Antoinette Schoar. Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319, 2020. https://www.sciencedirect.com/science/article/abs/pii/S0304405X19301746.

[29] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf, 2008.

[30] NiceHash. SHA-256 hashpower marketplace. https://www.nicehash.com/algorithm/sha256, 2026. Accessed 2026-05-02; quoted rental supply is de minimis relative to Bitcoin network hashrate.

[31] Daniel Perez, Sam M. Werner, Jiahua Xu, and Benjamin Livshits. Liquidations: DeFi on a Knife-edge. In *Financial Cryptography and Data Security (FC 2021)*, volume 12675 of *Lecture Notes in Computer Science*, pages 457–476. Springer, 2021. URL: https://link.springer.com/chapter/10.1007/978-3-662-64331-0_24, doi:10.1007/978-3-662-64331-0_24.

[32] Kaihua Qin, Liyi Zhou, and Arthur Gervais. Quantifying blockchain extractable value: How dark is the forest? In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 198–214. IEEE, 2022. URL: https://ieeexplore.ieee.org/document/9833734/, doi:10.1109/sp46214.2022.9833734.

[33] Sarad Sayeed and Hector Marco-Gisbert. Assessing blockchain consensus and security mechanisms against the 51% attack. *Applied Sciences*, 9(9):1788, April 2019. https://www.mdpi.com/2076-3417/9/9/1788.

[34] Tardis.dev. Tardis.dev: Historical cryptocurrency market data. https://tardis.dev, 2026. Accessed: 2026-01-19. Microsecond-precision derivatives market data from cryptocurrency exchanges' WebSocket feeds.

[35] TheEnergyMag. Quarterly bitcoin mining leaderboard: Realized hashrate. https://pro.theenergymag.com/overview, 2026. Q4 2025 public-miner realized hashrate data.

A Concise Supplementary Materials

This appendix is intentionally compact. It preserves the labels and core calibration objects cited in the main text while omitting long-form proofs, audit tables, and implementation details that are not essential for the workshop submission version.

A.1 Timing Figure and Core Calibration

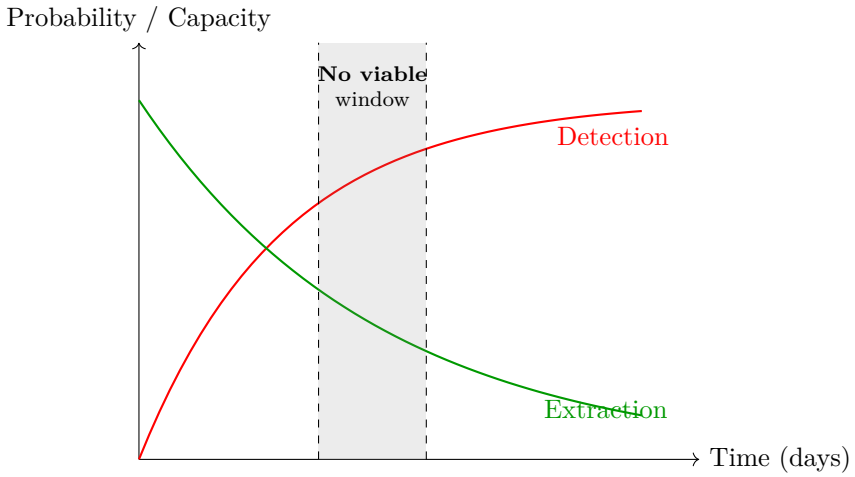


Figure 6: Timing paradox: detection probability rises while extractable capacity decays. The overlap region captures why shifting extraction earlier or later does not restore profitability.

Table 5: Core Calibration Summary Used in the Main Text

Quantity	**Value**	**Interpretation**
Bitcoin price (baseline)	\$95,000	2026-01-08 calibration point
Network hashrate	750 EH/s	2026-01-08 baseline security level
51% control target	382 EH/s	Majority-control threshold used for physical observability
Attack cost C_{hash}	\$24.1B	Hardware plus energy
ASIC units required	∼1.4M	Current-generation deployment envelope
Power requirement	5.2–6.7 GW	65–84× FERC 80MW threshold
Annual ASIC production	3.0M units	Attack requires about 47% of yearly global output
MARA comparison	52.5 EH/s	Q4 2025 realized hashrate; about 14% of 51% control target
Top-10 public miners	∼370 EH/s	Roughly equal to the control target, but public and slow to acquire or coordinate

A.2 Market, Timing, and Extraction Bounds

Table 6: Market-Side Bounds Used in the Main Text

Metric	Value	Interpretation
Pre-detection CEX liquidity	\$8.1B	Visible depth before withdrawal
Realistic realized extraction	\$1.02B	Main realistic cashable bound
Optimistic combined extraction	\$1.4B	Multi-channel upper bound
Insurance-fund capacity	\$2.4B	Immediate settlement buffer
Market-maker threshold p^*	0.15	Belief level that triggers sharp withdrawal
Withdrawal speed κ	4.2	Calibrated decay rate
Median detection time	80 minutes	ABM result used in main text
Historical max weekly OI growth	\$5.1B/week	2021 bull-peak benchmark
Attack-scale OI growth	\$15.3B/week	About 3.0× historical maximum
Baseline detection probability	96%	Within 6 hours for attack-scale build

Table 7: Regulated-Market and Cross-Chain Summary

Metric	Value	Interpretation
CME effective all-months cap	\$600M per account	Conservative calibrated per-account limit
CME accounts required	26	For the \$15.3B attack-scale hedge at \$600M per account
ETF daily average flow	\$45M	Historical baseline
ETF flow anomaly	46.4σ	Attack-scale ETF extraction is immediately visible
Bitcoin LASB	0.042	Lowest in our cross-chain comparison
Bitcoin Cash LASB	0.044	Close to Bitcoin despite much smaller scale
Ethereum Classic LASB	0.054	Highest among major PoW chains in our calibration

Table 8: LASB Across Market Regimes

Period	Regime	Estimated LASB
2021 Q4	Bull market peak	0.08
2022 Q4	Post-FTX stress	0.02
2024 Q1	ETF-approval rally	0.05
2026-01-08 baseline	Current calibration	0.042

A.3 Assumptions, Calibration, and Proof Notes

Assumptions. The main text analyzes rational, profit-maximizing attackers under current Bitcoin market structure. We assume visible order-book depth, observed insurance-fund disclosures, attacker-favorable venue coordination, and no heroic defensive actions. These choices are conservative in the sense used throughout the paper: when uncertain, we bias assumptions toward *higher* attack feasibility.

Calibration. All main-text quantities come from public sources: chain difficulty and block-interval data for hashrate, Cambridge CBECI for fleet-efficiency and electricity-consumption

context, exchange APIs and Tardis.dev for market depth and liquidation dynamics, public insurance-fund disclosures for settlement buffers, CME documentation for regulated-market limits, and manufacturer specifications for ASIC cost and energy use.

Table 9: Physical Calibration Parameters for the 2026-01-08 Baseline

Parameter	**Value**	**Use in the paper**
Network hashrate baseline	750 EH/s	Public chain-difficulty / block-interval estimate for 2026-01-08
51% control target	382 EH/s	Physical-control target used in Section 6
Fleet-average efficiency	17.5 J/TH	CBECI-style fleet-efficiency envelope; gives about 6.7 GW
Flagship efficiency	13.5 J/TH	Current-generation procurement envelope; gives about 5.2 GW
Current-generation rigs	∼1.4M	Deployment envelope at flagship-class efficiency
Annual production benchmark	3.0M rigs/year	Used only as an observability and procurement-scale benchmark

Technical definitions. RSC(t) denotes *notional* realizable short capacity; V_{extract} denotes *cashable* extraction; and Δ_{cap} converts price moves into attacker-captured cash value after execution frictions. The main text uses those quantities consistently.

Conservative bounding logic. The main text's inequalities always upper-bound what the attacker can cash out, not what they can merely mark to market. This is why solvency, liquidity withdrawal, concealment, and physical observability all appear as *minimum* constraints rather than additive resources.

A.4 Supplementary Notes on Validation and Proofs

Proof structure. The omitted long-form proofs all follow the same template: define a notional capacity bound, convert it to cashable value, and then intersect it with venue-level settlement and concealment constraints. No omitted proof changes the direction of the main inequalities stated in the paper.

Role of omitted derivations. The omitted derivations tighten, but do not generate, the paper's main exclusion result. Proposition 1 stands on the logic stated in the main text; the supplementary notes only unpack calibration choices and edge-case diagnostics.

Simulation notes. The ABM is used as a mechanism check, not as the source of the paper's main bound. The key reported quantities are the median 80-minute detection time and the sharp collapse in usable liquidity once beliefs cross the calibrated threshold.

External evidence. The Bitcoin Gold and Bitcoin SV cases are not used to calibrate the paper's parameters. Their role is as partial historical reconstruction and qualitative external checks: they show that when real exchanges and attackers are observed, successful consensus attacks still monetize through constrained and detectable channels.

Evidence-role split. Section 7 gives the operative division used in the paper: FTX is the calibration target for the withdrawal mechanism, the 312 and May 2021 crashes are non-calibration direction checks, and Bitcoin Gold / Bitcoin SV are partial historical reconstruction and external cross-chain checks.

Table 10: External Supporting Checks Not Used to Fit the Baseline

Episode	**Observed stress**	**What it supports**
312 crash (2020-03)	$1.7B liquidations	Stress episodes more severe than FTX exist, so calibrating on FTX is conservative rather than aggressive
May 2021 crash	$511M liquidations	Rapid deleveraging and withdrawal are not unique to FTX's idiosyncratic failure
Bitcoin Gold attacks (2018, 2020)	Successful attacks with bounded extraction and fast detection	External cross-chain evidence that technical success does not imply unbounded economic monetization
Bitcoin SV attack series (2021)	Reorganizations followed by exchange restrictions	External cross-chain evidence that venue access and settlement controls can bind after protocol-level attacks

A.5 Sensitivity and Robustness

This appendix is best read as a failure-frontier map rather than as a generic robustness dump.

Table 11: Compact Sensitivity and Counterfactual Summary

Scenario	**Value**	**Result**
Baseline Bitcoin LASB	0.042	Strong economic dominance gap
High-liquidity / weak-defense stress	0.15	Still unprofitable
Slow withdrawal counterfactual	38% extraction ratio	Raises hedge size but does not restore profit
High-threshold counterfactual	LASB < 0.06	Still far below break-even
Threshold for profitability	LASB > 1	Requires implausible joint loosening of multiple constraints

The paper's qualitative conclusion is robust to plausible perturbations. The relevant failure mode is not that one parameter moves slightly in the attacker's favor, but that several market, settlement, and physical bottlenecks all loosen at once.

A.6 Prototype LASB Dashboard

The prototype below is meant to show operationalizability, not to claim a production-ready surveillance system.

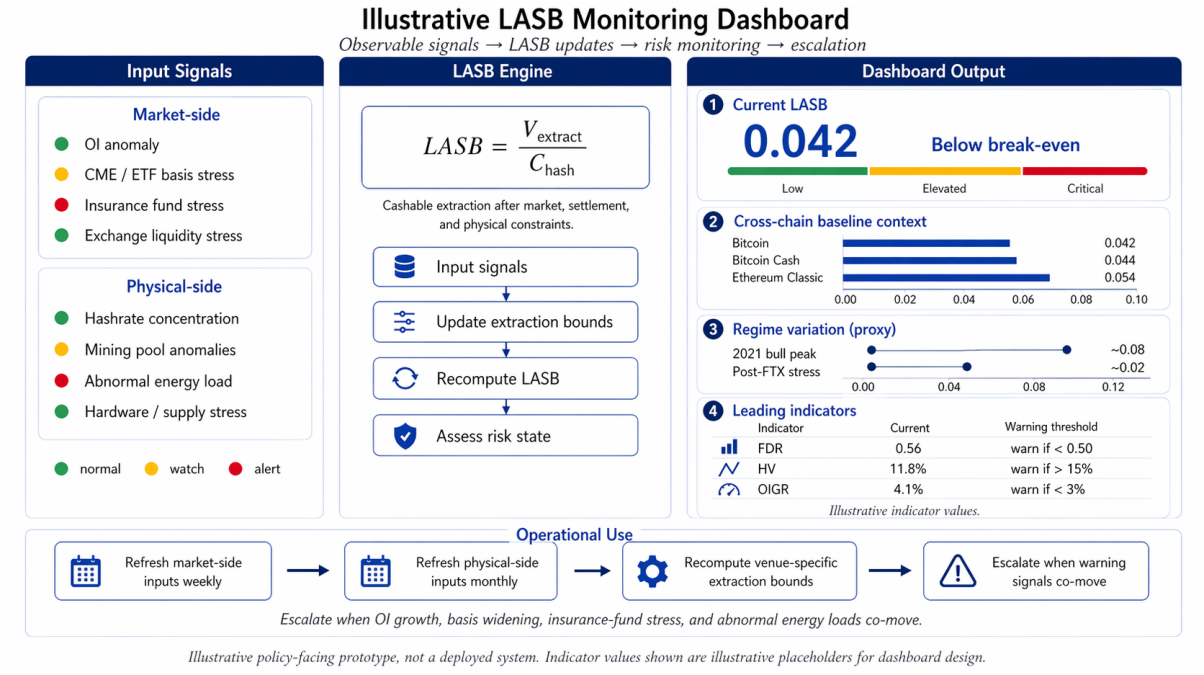


Figure 7: Illustrative LASB monitoring dashboard prototype. The figure shows how observable market-side and physical-side signals can be translated into LASB updates, cross-chain baseline context, regime monitoring, leading-indicator tracking, and escalation triggers. It is an illustrative policy-facing prototype rather than a deployed system. Signal states and indicator values are shown for dashboard design purposes.

A.7 Minimal Reproducibility Recipe

Minimal recipe. A portable LASB computation needs only four ingredients: (i) an all-in attack-cost estimate C_{hash} from hashrate, hardware, power, and deployment assumptions; (ii) a market-side estimate of $\text{RSC}(t)$ from depth, open interest, position limits, and withdrawal behavior; (iii) a conservative Δ_{cap} assumption translating price moves into attacker-captured cash; and (iv) settlement, solvency, concealment, and physical caps used to convert notional size into V_{extract}. The computation order is then immediate: estimate C_{hash}, estimate $\text{RSC}(t)$, map it into cashable value, intersect with the remaining caps, and report LASB as $V_{\text{extract}}/C_{\text{hash}}$ together with a regime-sensitive range. In the Bitcoin baseline used in the paper, that recipe gives $C_{\text{hash}} \approx \$24.1\text{B}$, realistic cashable extraction of about $\$1.02\text{B}$, and $\text{LASB} \approx 0.042$. Extending the metric to a new chain therefore requires no new theory, only re-estimation of the same objects under that chain's current market and observability constraints.