Formal Proofs: Mechanism Design Audit of Crosslink Zebra

Nicolás "nikete" Della Penna

January 2026

Abstract

This document provides formal definitions, propositions, and proofs for the claims made in the mechanism design audit of Crosslink Zebra. We formalize the model of hybrid PoW/PoS consensus with frozen validator sets and no slashing, characterize equilibrium conditions for finality progress, and prove impossibility results for incentive-based liveness guarantees under external payoffs. The proofs are intended to make the audit's claims precise and verifiable.

Contents

1 Model and Definitions

We begin by formalizing the components of a hybrid PoW/PoS finality system with the structural properties of Crosslink Zebra.

1.1 Blockchain Structure

Definition 1.1 (Blockchain)**.** A blockchain $\mathcal{C}$ is a sequence of blocks $(B_0, B_1, \ldots, B_h)$ where B_0 is the genesis block and each B_i for $i > 0$ contains a reference to B_{i-1}. We write $h(\mathcal{C})$ for the height of the chain tip.

Definition 1.2 (Finality Function)**.** A finality function $\mathrm{LF} : \mathcal{C} \to \mathbb{N}$ maps a chain to the height of its last finalized block. We require:

(i) $\mathrm{LF}(\mathcal{C}) \leq h(\mathcal{C})$ (finality does not exceed tip)

(ii) If $\mathcal{C}' \supseteq \mathcal{C}$, then $\mathrm{LF}(\mathcal{C}') \geq \mathrm{LF}(\mathcal{C})$ (finality is monotonic in extensions)

Definition 1.3 (Finality Gap)**.** The finality gap at chain $\mathcal{C}$ is $G(\mathcal{C}) = h(\mathcal{C}) - \mathrm{LF}(\mathcal{C})$.

1.2 Participants and Stakes

Definition 1.4 (Validator Set)**.** A validator set $\mathcal{V} = \{v_1, \ldots, v_n\}$ is a finite set of validators. Each validator v_i has stake $s_i > 0$. Total stake is $S = \sum_{i=1}^{n} s_i$.

Definition 1.5 (Frozen Validator Set)**.** In a frozen-set design, the active validator set $\mathcal{V}^*$ for finality decisions is determined by the state at the last finalized block. During a finality stall (period where LF does not advance), $\mathcal{V}^*$ cannot be modified by delegation, entry, or exit actions on the unfinalized portion of the chain.

Definition 1.6 (Finality Certificate)**.** A finality certificate σ for height h is a collection of signatures from validators in $\mathcal{V}^*$ attesting to a block at height h. A certificate is *valid* if the signing validators collectively hold stake exceeding a threshold $\tau \cdot S$ where $\tau \geq 2/3$.

1.3 Reward Model

Definition 1.7 (Block Reward Pool)**.** Let $r > 0$ be the per-block reward. The pending reward pool after Δ blocks without finality advancement is:

$$R(\Delta) = f(\Delta, r)$$

where f specifies the accumulation rule. We distinguish:

- **Accumulating**: $f(\Delta, r) = \Delta \cdot r$ (rewards grow linearly)
- **Capped**: $f(\Delta, r) = \min(\Delta \cdot r, \bar{R})$ for some cap $\bar{R}$
- **Decaying**: $f(\Delta, r) = r \cdot \frac{1-\delta^{\Delta}}{1-\delta}$ for decay rate $\delta \in (0, 1)$

Definition 1.8 (Commission and Staker Rewards)**.** Let $\alpha \in (0, 1)$ be the commission rate. When finality advances, the reward pool R is distributed as:

- Commission: αR distributed to signing finalizers proportional to stake
- Staker rewards: $(1 - \alpha)R$ distributed to all stakers proportional to stake

1.4 Agent Model

Definition 1.9 (Rational Agent)**.** An agent i is *rational* if it chooses actions to maximize expected utility U_i. We decompose utility as:

$$U_i = U_i^{\text{protocol}} + U_i^{\text{external}}$$

where U_i^{protocol} derives from protocol rewards and U_i^{external} from positions outside the protocol (e.g., short exposure, competing chains).

Definition 1.10 (Discount Factor)**.** Agents discount future payoffs by factor $\beta \in (0, 1)$ per block period. A payoff x received after Δ blocks has present value $\beta^{\Delta} x$.

2 Safety Results

Safety results for BFT-style finality gadgets are well-established. We state them here for completeness and to clarify assumptions.

Assumption 2.1 (Byzantine Threshold)**.** The fraction of stake controlled by Byzantine (arbitrarily malicious) validators is less than $1 - \tau$, where $\tau \geq 2/3$ is the finality threshold.

Assumption 2.2 (PoW Consistency)**.** The underlying PoW chain satisfies common-prefix consistency: with high probability, any two honest nodes' chains share a common prefix up to some bounded depth k.

Theorem 2.3 (Accountable Safety)**.** *Under Assumptions 2.1 and 2.2, if two conflicting blocks B and B' at the same height are both finalized, then at least $(2\tau - 1) \cdot S$ stake must have signed conflicting certificates.*

Proof. Let σ be a valid certificate for B and σ' be a valid certificate for B'. By definition of validity, the signers of σ hold stake $\geq \tau S$ and the signers of σ' hold stake $\geq \tau S$.

The intersection of these signer sets must hold stake at least:

$$\tau S + \tau S - S = (2\tau - 1)S$$

by inclusion-exclusion. Every validator in the intersection signed both certificates, which constitutes equivocation on conflicting blocks. □

Corollary 2.4 (Safety Under Honest Supermajority)**.** *If Byzantine stake is less than $(1 - \tau)S$ and $\tau \geq 2/3$, then conflicting finalization requires honest validators to equivocate, which they will not do by assumption.*

Remark 2.5. Theorem 2.3 is the standard BFT safety result (cf. Casper FFG). The key observation for Crosslink Zebra is that safety holds *regardless of economic incentives* once honest validators follow the signing protocol. Safety is structural, not incentive-dependent.

3 Liveness and Incentive Alignment

Unlike safety, liveness depends critically on incentive alignment. We now formalize conditions under which rational validators choose to advance finality.

3.1 The Rational Stall Equilibrium

Definition 3.1 (Blocking Coalition)**.** A coalition $\mathcal{B} \subseteq \mathcal{V}^*$ is *blocking* if $\sum_{v \in \mathcal{B}} s_v > (1 - \tau)S$. A blocking coalition can prevent finality by withholding signatures.

Proposition 3.2 (Rational Stall Under Accumulating Rewards)**.** *Suppose rewards accumulate linearly: $R(\Delta) = \Delta \cdot r$. Let $\mathcal{B}$ be a blocking coalition that has delayed finality for Δ blocks. Delaying one additional block increases the coalition's expected present-value payoff if and only if:*

$$\beta > \frac{\Delta}{\Delta + 1}.$$

In particular, at $\Delta = 0$ any $\beta > 0$ makes the first delay profitable, and the condition becomes harder to satisfy as Δ grows.

Proof. Consider a blocking coalition with collective stake $s_{\mathcal{B}}$ that has delayed finality for Δ blocks. If it finalizes now, it receives (in commission):

$$\alpha \cdot R(\Delta) \cdot \frac{s_{\mathcal{B}}}{S} = \alpha \cdot \Delta \cdot r \cdot \frac{s_{\mathcal{B}}}{S}$$

If it delays one more block and then finalizes, it receives (discounted):

$$\beta \cdot \alpha \cdot R(\Delta + 1) \cdot \frac{s_{\mathcal{B}}}{S} = \beta \cdot \alpha \cdot (\Delta + 1) \cdot r \cdot \frac{s_{\mathcal{B}}}{S}$$

The common factor $\alpha \cdot r \cdot s_{\mathcal{B}}/S$ cancels. Delaying is preferred if and only if:

$$\beta \cdot (\Delta + 1) > \Delta \quad \iff \quad \beta > \frac{\Delta}{\Delta + 1}.$$

Note that the coalition's stake share $s_{\mathcal{B}}/S$ is irrelevant to the marginal delay decision—it scales both payoffs equally.

As $\Delta \to \infty$, the threshold approaches $\beta > 1$, which is impossible, so delay cannot continue indefinitely. However, for any finite horizon or with additional hazard (governance intervention), the coalition faces a stopping problem.

More precisely, if the coalition believes finality will occur (via governance or coalition breakdown) with probability p per block regardless of their action, then continuing to delay is optimal if:

$$\beta(1 - p) \cdot \frac{\Delta + 1}{\Delta} > 1$$

For small p and large β, this holds for all Δ below some threshold that grows with β and shrinks with p. □

Theorem 3.3 (Anti-Jackpot Necessity)**.** *If $R(\Delta)$ is unbounded and increasing in Δ, then for any discount factor $\beta < 1$, there exists a threshold $\bar{\Delta}$ such that rational delay up to $\bar{\Delta}$ blocks is a subgame-perfect equilibrium for a blocking coalition.*

Proof. We solve the coalition's optimal stopping problem by backward induction. Let $V(\Delta)$ be the continuation value at delay Δ. The coalition chooses between:

- Finalize now: receive $\alpha R(\Delta) \cdot \frac{s_{\mathcal{B}}}{S}$
- Delay: receive $\beta \cdot V(\Delta + 1)$

At the optimal stopping point $\bar{\Delta}$:

$$\alpha R(\bar{\Delta}) \cdot \frac{s_{\mathcal{B}}}{S} = \beta \cdot \alpha R(\bar{\Delta}+1) \cdot \frac{s_{\mathcal{B}}}{S}$$

Simplifying:

$$R(\bar{\Delta}) = \beta \cdot R(\bar{\Delta}+1)$$

For accumulating rewards $R(\Delta) = \Delta \cdot r$:

$$\bar{\Delta} = \beta(\bar{\Delta}+1) \implies \bar{\Delta} = \frac{\beta}{1-\beta}$$

For $\beta = 0.9$, this gives $\bar{\Delta} = 9$ blocks of rational delay. For $\beta = 0.99$, this gives $\bar{\Delta} = 99$ blocks. The coalition will delay until $\bar{\Delta}$ and then finalize, extracting the accumulated reward pool. □

Corollary 3.4 (Anti-Jackpot Sufficiency Under Non-Increasing Rewards)**.** *If $R(\Delta)$ is non-increasing in Δ for $\Delta > 0$ (i.e., rewards are capped at a fixed level or the pending pool shrinks), then immediate finalization strictly dominates delay for any discount factor $\beta < 1$.*

Proof. If $R(\Delta+1) \leq R(\Delta)$, then:

$$\beta \cdot R(\Delta+1) \leq \beta \cdot R(\Delta) < R(\Delta)$$

so finalizing now yields strictly higher present value than delaying. □

Proposition 3.5 (Bounded Delay Under Decaying Per-Block Rewards)**.** *Suppose per-block rewards decay geometrically so the pending pool is*

$$R(\Delta) = r \cdot \frac{1-\delta^{\Delta}}{1-\delta}, \quad \delta \in (0,1).$$

Note that $R(\Delta)$ is strictly increasing in Δ (each new block adds $r\delta^{\Delta} > 0$), so the preceding corollary does not apply. However, the marginal growth rate *$R(\Delta+1)/R(\Delta)$ decreases toward 1. The optimal delay for a blocking coalition with discount factor $\beta < 1$ is bounded by:*

$$\bar{\Delta}_{decay} \leq \frac{\log(1-\beta)}{\log \delta}$$

For $\beta = 0.95$ and $\delta = 0.9$, this gives $\bar{\Delta}_{decay} \leq 28$ blocks. By contrast, under accumulating rewards with the same β, the optimal delay is $\bar{\Delta} = \beta/(1-\beta) = 19$ blocks, but the accumulated pool at that point is proportional to $\bar{\Delta}$, whereas under decay it is bounded by $r/(1-\delta)$.

Proof. Under the decaying model, the coalition delays one more block when $\beta \cdot R(\Delta+1) > R(\Delta)$, i.e.:

$$\beta \cdot \frac{1-\delta^{\Delta+1}}{1-\delta} > \frac{1-\delta^{\Delta}}{1-\delta}$$

Simplifying:

$$\beta(1-\delta^{\Delta+1}) > 1-\delta^{\Delta}$$

$$\beta - \beta\delta^{\Delta+1} > 1-\delta^{\Delta}$$

$$\delta^{\Delta}(1-\beta\delta) > 1-\beta$$

$$\delta^{\Delta} > \frac{1-\beta}{1-\beta\delta}$$

Taking logarithms (both sides positive since $\beta < 1$ and $\beta\delta < 1$):

$$\Delta < \frac{\log\left(\frac{1-\beta}{1-\beta\delta}\right)}{\log \delta}$$

Since $\frac{1-\beta}{1-\beta\delta} < 1$, and $\log \delta < 0$, the right-hand side is positive.

For an upper bound, note $1 - \beta\delta > 1 - \beta$ is not useful directly, but $\frac{1-\beta}{1-\beta\delta} > (1-\beta)$ since $1 - \beta\delta < 1$. Thus:

$$\bar{\Delta}_{\text{decay}} \leq \frac{\log(1-\beta)}{\log \delta}$$

The key difference from accumulating rewards is not only that delay is bounded, but that the *pool size* at the optimal stopping point is bounded by $r/(1-\delta)$ regardless of β, whereas the accumulating pool grows as $\beta/(1-\beta) \cdot r$, which is unbounded as $\beta \to 1$. □

Remark 3.6. This result clarifies the relationship between the audit's recommendation R4 and the formal model. A purely decaying per-block reward does *not* make $R(\Delta)$ non-increasing, so it does not achieve the "immediate finalization dominates" property. Instead, it achieves a weaker but practically important property: bounded delay with a bounded pool. A hard cap $\bar{R}$ on the pool achieves the stronger property once the cap binds (since R becomes constant and then $\beta R < R$). In practice, combining cap and decay—e.g., $R(\Delta) = \min\left(r \cdot \frac{1-\delta^{\Delta}}{1-\delta},\ \bar{R}\right)$—achieves both bounded delay and eventual strict dominance of immediate finalization.

3.2 Bribery and Pivotality

Proposition 3.7 (Marginal Bribery Cost)**.** *Let honest participation be $p > \tau$ (i.e., honest validators holding fraction p of stake will sign). The cost to stall finality via bribery is proportional to $(p-\tau)S$, not to S.*

Proof. Finality requires signatures from stake $\geq \tau S$. If honest stake $p \cdot S$ will sign, an attacker must bribe enough honest validators to reduce effective participation below τS.

The attacker needs to bribe validators holding stake:

$$p \cdot S - \tau \cdot S = (p - \tau)S$$

The cost is thus $c \cdot (p-\tau)S$ where c is the per-unit bribery cost, which may be as low as the validator's expected reward for signing. □

Remark 3.8. This formalizes the audit's observation that the cost to stall depends on the margin to threshold, not total stake. A system with $p = 0.7$ and $\tau = 2/3$ has margin $(0.7 - 0.667)S \approx 0.033S$, requiring bribing only $\sim 3.3\%$ of stake.

4 Penalty for Failure to Defend (PFD)

Definition 4.1 (Penalty for Failure to Defend)**.** The PFD for a deviation d from compliant behavior c is:

$$\text{PFD}(d, c) = U^{\text{protocol}}(c) - U^{\text{protocol}}(d)$$

the difference in protocol-derived utility between compliant and deviant behavior.

Theorem 4.2 (PFD Bound in Non-Slashing Mechanisms)**.** *In any mechanism without slashing (i.e., where principal stake cannot be reduced as punishment), the maximum PFD is bounded by:*

$$\text{PFD}_{\max} \leq \bar{R} + \bar{L}$$

where $\bar{R}$ is the maximum withheld reward and $\bar{L}$ is the maximum liquidity/time-value cost from delayed exit.

Proof. Without slashing, the protocol can only impose costs through:

1. Withholding rewards that would have been earned under compliant behavior
2. Delaying the return of staked principal (liquidity cost)
3. Operational costs (which are bounded and typically small)

The first component is bounded by the total rewards available over the deviation period, $\bar{R}$.

The second component is bounded by the time-value cost of delayed principal return. If principal P is delayed by T blocks with discount factor β, the cost is $P(1-\beta^T) \leq P(1-\beta) \cdot T$ for small T.

Since principal is not at risk (no slashing), the maximum penalty is the sum of these components. □

Corollary 4.3 (External Incentive Impossibility)**.** *If an agent's* marginal *external payoff from deviation satisfies $U^{external}(d) - U^{external}(c) > \text{PFD}_{\max}$, then no non-slashing mechanism can make compliant behavior a dominant strategy for that agent.*

Proof. The agent's total utility from deviation is:

$$U(d) = U^{\text{protocol}}(d) + U^{\text{external}}(d)$$

For compliant behavior:

$$U(c) = U^{\text{protocol}}(c) + U^{\text{external}}(c)$$

Deviation is preferred if:

$$U^{\text{external}}(d) - U^{\text{external}}(c) > U^{\text{protocol}}(c) - U^{\text{protocol}}(d) = \text{PFD}(d, c)$$

Since $\text{PFD}(d, c) \leq \text{PFD}_{\max}$ by Theorem 4.2, the condition $U^{\text{external}}(d) - U^{\text{external}}(c) > \text{PFD}_{\max}$ is sufficient.

A common special case is the "short seller" scenario where compliance has no external payoff ($U^{\text{external}}(c) = 0$). In that case, the condition simplifies to $U^{\text{external}}(d) > \text{PFD}_{\max}$. □

Remark 4.4. This formalizes the audit's claim that non-slashing mechanisms have fundamental limits. An agent with sufficient short exposure can profit from disruption regardless of protocol-level incentives.

5 Catastrophic Failure Modes

5.1 The Zombie Set

Definition 5.1 (Effective Participation)**.** Let $\mathcal{A} \subseteq \mathcal{V}^*$ be the set of validators that are online and willing to sign. The effective participation is:

$$\pi = \frac{\sum_{v \in \mathcal{A}} s_v}{S}$$

Definition 5.2 (Attrition Rate)**.** The attrition rate $\lambda > 0$ is the fraction of active stake that becomes inactive per time period (due to key loss, operator shutdown, etc.).

Proposition 5.3 (Zombie Set Threshold)**.** *Let initial effective participation be $\pi_0 > \tau$. Under constant attrition rate λ, effective participation after t periods is:*

$$\pi(t) = \pi_0 \cdot (1 - \lambda)^t$$

The finality gadget becomes permanently inoperable when $\pi(t) < \tau$, which occurs at:

$$t^* = \frac{\log(\tau/\pi_0)}{\log(1 - \lambda)}$$

Proof. Under constant attrition, if $\pi(t)$ is effective participation at time t:

$$\pi(t + 1) = \pi(t) \cdot (1 - \lambda)$$

Solving the recurrence: $\pi(t) = \pi_0(1 - \lambda)^t$.
Setting $\pi(t^*) = \tau$:

$$\pi_0(1 - \lambda)^{t^*} = \tau$$

$$t^* = \frac{\log(\tau/\pi_0)}{\log(1 - \lambda)}$$

For $\pi_0 = 0.8$, $\tau = 2/3$, $\lambda = 0.02$ (2% monthly attrition):

$$t^* = \frac{\log(0.667/0.8)}{\log(0.98)} \approx \frac{-0.182}{-0.0202} \approx 9 \text{ months}$$

□

Theorem 5.4 (Zombie Set Irreversibility)**.** *In a frozen-set design, once $\pi(t) < \tau$, the finality gadget cannot recover without external intervention (governance reset), even if all remaining validators are honest and online.*

Proof. By definition of frozen-set design, $\mathcal{V}^*$ is fixed at the last finalized block. New validators cannot join and departed validators cannot be replaced until finality advances.

But finality requires a certificate signed by stake $\geq \tau S$. If effective participation $\pi < \tau$, no valid certificate can be formed, so finality cannot advance, so $\mathcal{V}^*$ remains frozen.

This is a fixed point: the system is stuck in a state where the only exit requires finality, but finality requires resources that the stuck state cannot provide. □

Remark 5.5. The Zombie Set is an *entropic* failure mode—it requires no adversary, only the natural decay of participation over time. The frozen-set design converts a temporary stall into a permanent failure if the stall persists long enough.

5.2 The Liquid Exit Paradox

Assumption 5.6 (Exit Feasibility During Stalls)**.** Staking exit actions (unbond, claim) can complete on the PoW chain even when finality is stalled.

Definition 5.7 (Economic Exposure)**.** A validator v's economic exposure E_v is the value at risk from protocol penalties or opportunity costs. Under no slashing, E_v equals the present value of foregone rewards plus liquidity costs.

Proposition 5.8 (Security Budget Decay)**.** *Under Assumption 5.6, if validators can exit during a stall, the effective security budget backing the frozen set decays over time even though nominal voting weights remain constant.*

Proof. Let $\mathcal{V}^*$ be the frozen validator set with voting weights $(w_1, \ldots, w_n)$ determined at the last finalized block.

Under Assumption 5.6, validators can complete exit actions (unbond, claim principal) on the unfinalized PoW chain. Once a validator v has exited:

- Their voting weight w_v in $\mathcal{V}^*$ remains unchanged (frozen)
- Their economic exposure $E_v \to 0$ (no stake at risk, no rewards to lose)

The cost to bribe validator v to sign an arbitrary certificate is now bounded by:

$$\text{Bribe cost} \leq E_v \to 0$$

An attacker seeking to corrupt the frozen set faces cost:

$$C(\text{corrupt}) = \sum_{v \in \mathcal{B}} (\text{bribe cost for } v)$$

where $\mathcal{B}$ is a set with $\sum_{v \in \mathcal{B}} w_v \geq \tau$.

As validators exit, $C(\text{corrupt}) \to 0$ while the nominal threshold $\tau \cdot S$ remains constant. This is "phantom security"—the appearance of stake-weighted security without the economic substance. □

Remark 5.9. Whether Assumption 5.6 holds depends on implementation details: unbonding periods, whether claims require finalized state, and censorship resistance of exit transactions during stalls. The audit recommends explicitly specifying this interaction.

Remark 5.10 (Rational vs. Intrinsically Honest Validators). The security budget decay in Proposition 5.8 applies to validators who are honest *because it is incentive-compatible*, not to validators who follow the protocol out of intrinsic commitment regardless of payoffs. Intrinsically honest validators remain secure after exit (they will not sign arbitrary certificates even at zero cost), but they may not constitute a supermajority on their own. The effective security of the frozen set after exits depends on the fraction of intrinsically honest stake, which is unobservable and should not be relied upon for security guarantees.

6 Fork Choice Rule Dichotomy

Definition 6.1 (Fork Choice Rule)**.** A fork choice rule $\mathcal{F}$ maps a node's view of the network (received blocks and certificates) to a canonical chain.

Definition 6.2 (Work-Preferring Rule)**.** $\mathcal{F}$ is *work-preferring* if it selects the chain with maximum cumulative work, regardless of finality status.

Definition 6.3 (Finality-Preferring Rule)**.** $\mathcal{F}$ is *finality-preferring* if it selects only among chains that extend the highest known finalized block.

Theorem 6.4 (Fork Choice Dichotomy)**.** *In a hybrid PoW/PoS system with a finality gadget that can stall:*

(i) Under a work-preferring rule, a stalled finality gadget can be bypassed, degrading the system to pure PoW.

(ii) Under a finality-preferring rule, a stalled finality gadget halts the canonical chain.

There is no fork choice rule that both (a) continues chain progress during finality stalls and (b) preserves the security guarantees of finality.

Proof. (i) Suppose $\mathcal{F}$ is work-preferring. Let B_f be the last finalized block and suppose finality stalls at B_f. Miners can create a fork from B_f that ignores the stalled finality gadget entirely. Under work-preferring $\mathcal{F}$, if this fork accumulates more work than any finality-consistent chain, it becomes canonical.

Finality certificates for the bypassed chain are never incorporated. The system effectively operates as pure PoW, losing the safety guarantees that finality was meant to provide.

(ii) Suppose $\mathcal{F}$ is finality-preferring. By definition, the canonical chain must extend B_f. If finality stalls at B_f and no certificate advances finality, then:

- The PoW chain may extend beyond B_f, producing blocks $B_{f+1}, B_{f+2}, \ldots$
- But these blocks are not finalized
- $\mathcal{F}$ accepts them as part of the canonical chain only if they extend B_f

If the finality gadget cannot produce certificates (e.g., due to Zombie Set), no block beyond B_f can be finalized. Depending on implementation:

- If $\mathcal{F}$ requires eventual finalization, the system halts
- If $\mathcal{F}$ allows unbounded unfinalized extensions, the chain grows but with only PoW-level security

(iii) Suppose a rule $\mathcal{F}^*$ achieves both (a) and (b). Then during a stall:

- By (a), the chain makes progress, so some blocks extend B_f
- By (b), these blocks have finality-level security

But finality-level security requires valid certificates, which by assumption cannot be produced during the stall. Contradiction. □

Corollary 6.5 (Governance Reset Necessity)**.** *A finality-preferring system requires an out-of-band governance mechanism to recover from Zombie Set failures.*

7 Recovery and Circular Dependencies

Proposition 7.1 (Catch-Up Finalization Requirement)**.** *After a finality stall, safe resumption requires finalizing the intervening unfinalized blocks in sequence (catch-up finalization). Direct resumption at the PoW tip is unsafe.*

Proof. Let B_f be the last finalized block and B_t be the current PoW tip with $h(B_t) > h(B_f)$.

The validator set $\mathcal{V}^*$ is determined by the state at B_f. The state at B_t may include:

- New staking deposits (validators not in $\mathcal{V}^*$)

- Completed exits (validators in $\mathcal{V}^*$ with zero economic exposure)
- Redelegation (stake moved between validators)

Suppose we attempt to finalize B_t directly using $\mathcal{V}^*$. The certificate would be valid under the rules at B_f, but:

1. The state at B_t depends on transactions in $B_{f+1}, \ldots, B_{t-1}$
2. Some of these transactions may conflict with certificates that could have been issued for intermediate blocks
3. Specifically, if a validator exited at block B_{f+k}, their signature on a certificate for B_t is economically meaningless

More critically, suppose there are two competing chains from B_f:

- Chain A: $B_f \to B_{f+1}^A \to \cdots \to B_t^A$
- Chain B: $B_f \to B_{f+1}^B \to \cdots \to B_t^B$

If we finalize B_t^A directly, what happens to transactions in Chain B that conflict with Chain A? Without finalizing intermediate blocks, there is no mechanism to establish that Chain A is canonical for the intervening history.

Therefore, safe resumption requires sequential finalization: first finalize B_{f+1}, then B_{f+2}, etc., until reaching the tip. Each certificate establishes the canonical history up to that point. □

Proposition 7.2 (Circular Dependency in Tip Resumption)**.** *Attempting to resume finality directly at the PoW tip creates a circular dependency: the validator set at the tip depends on finality decisions that have not been made.*

Proof. Let $\mathcal{V}(B)$ denote the validator set implied by the state at block B.

For any block B_k with $k > f$:

$$\mathcal{V}(B_k) = g(\text{State}(B_k))$$

where g extracts the validator set from the staking state.

$\text{State}(B_k)$ depends on:

- $\text{State}(B_{k-1})$
- Transactions in B_k
- Whether B_k is on the canonical chain (i.e., finality decisions)

If we attempt to use $\mathcal{V}(B_t)$ to finalize B_t:

- $\mathcal{V}(B_t)$ depends on $\text{State}(B_t)$
- $\text{State}(B_t)$ depends on the canonical chain from B_f to B_t
- The canonical chain depends on finality certificates
- Finality certificates depend on $\mathcal{V}^*$

If $\mathcal{V}^* = \mathcal{V}(B_t)$, this is circular: the validator set depends on finality, which depends on the validator set.

The frozen-set design breaks this cycle by fixing $\mathcal{V}^* = \mathcal{V}(B_f)$, which is well-defined because B_f is already finalized. □

8 Miner Incentive Distortions

Proposition 8.1 (Empty Block Incentive)**.** *If the miner bounty for including a finality certificate is δR and transaction fees are F, miners prefer empty blocks when:*

$$\delta R > F + (\textit{marginal orphan risk cost from transactions})$$

Proof. Let $p_{\text{orphan}}(B)$ be the probability that block B is orphaned. This probability increases with block size due to propagation delay.

Let B_{full} be a block with transactions (size s_{full}, fees F) and B_{empty} be a block with only the certificate (size $s_{\text{cert}} < s_{\text{full}}$).

Expected payoff from B_{full}:

$$(1 - p_{\text{orphan}}(s_{\text{full}})) \cdot (\delta R + F + \text{coinbase})$$

Expected payoff from B_{empty}:

$$(1 - p_{\text{orphan}}(s_{\text{cert}})) \cdot (\delta R + \text{coinbase})$$

The miner prefers B_{empty} when:

$$(1 - p_{\text{orphan}}(s_{\text{cert}})) \cdot \delta R > (1 - p_{\text{orphan}}(s_{\text{full}})) \cdot (\delta R + F)$$

For small orphan probabilities, this simplifies to:

$$\delta R \cdot (p_{\text{orphan}}(s_{\text{full}}) - p_{\text{orphan}}(s_{\text{cert}})) > F \cdot (1 - p_{\text{orphan}}(s_{\text{full}}))$$

When δR is large relative to F, this inequality holds, incentivizing empty blocks. □

Proposition 8.2 (Certificate Malleability)**.** *If the miner bounty is keyed to the certificate hash, miners can malleate certificates to claim multiple bounties or orphan competitors.*

Proof. BLS multi-signatures allow subset aggregation: given signatures $\{\sigma_i\}_{i \in S}$ for a message m, any subset $S' \subseteq S$ with $|S'| \geq$ threshold yields a valid aggregate signature $\sigma_{S'}$.

Different subsets produce different aggregate signatures, hence different certificate hashes.

If miner M_1 includes certificate σ_S and miner M_2 includes $\sigma_{S'}$ (with $S' \neq S$, both valid), and the bounty is per-hash:

- Both miners may claim bounties for "different" certificates
- A miner can create a competing block with a malleated certificate to orphan a competitor's block

Keying the bounty to finalized height (not certificate hash) eliminates this: only one bounty is paid per height advancement, regardless of which valid certificate achieves it. □

9 Free-Rider and Equivocation Results

Proposition 9.1 (Free-Rider Equilibrium Without Signer-Conditioning)**.** *If commission is paid to all active-set validators regardless of signing, not signing is strictly dominant for any individual validator (assuming signing cost $\epsilon > 0$).*

Proof. Let c_v be the commission paid to validator v. Under non-conditioned commission:

$$c_v = \alpha R \cdot \frac{s_v}{S} \quad \text{(independent of signing)}$$

The cost of signing is $\epsilon > 0$ (computation, bandwidth, operational attention).
A validator's payoff is:

- If sign: $c_v - \epsilon$
- If not sign: c_v

Not signing strictly dominates for any individual validator.
If all validators reason this way, no one signs, and finality stalls. But the individual incentive remains: conditional on others signing (so finality advances), not signing is still preferred. □

Proposition 9.2 (Equivocation as Costless DoS)**.** *Without slashing, equivocating (signing conflicting certificates for the same height) imposes zero financial cost on the equivocator.*

Proof. Let v sign certificates σ for block B and σ' for conflicting block B' at the same height.
Under no-slashing:

- Principal stake is not reduced
- Rewards depend on which certificate (if any) is included on-chain
- At most one of σ, σ' contributes to an on-chain certificate

The equivocator receives the same reward as if they had signed only the winning certificate. The cost is zero.
However, honest nodes must:

- Receive and verify both σ and σ'
- Determine which (if either) to propagate
- Potentially process many equivocating signatures from the same validator

This imposes costs on the network (bandwidth, computation) without cost to the equivocator—a classic externality enabling DoS. □

10 Summary of Results

1. **Safety is structural** (Theorem 2.3): Once finalized, blocks cannot be reverted without detectable equivocation by $\geq (2\tau - 1)$ stake.
2. **Liveness requires anti-jackpot** (Theorem 3.3, Proposition 3.5): Under accumulating rewards, rational delay is an equilibrium. Capping rewards eliminates delay once the cap binds. Decaying per-block rewards bound delay and the pool size, but do not make immediate finalization strictly dominant; combining cap and decay achieves both properties.
3. **PFD is bounded** (Theorem 4.2): Non-slashing mechanisms cannot impose penalties exceeding withheld rewards plus liquidity costs. External incentives can exceed this bound.

4. **Zombie Set is irreversible** (Theorem 5.4): In frozen-set designs, participation attrition below threshold permanently disables finality without governance intervention.

5. **Liquid Exit degrades security** (Proposition 5.8): If exits complete during stalls, the economic substance backing frozen voting weights evaporates.

6. **Fork choice forces a choice** (Theorem 6.4): No rule achieves both continued progress and finality guarantees during stalls.

7. **Catch-up is required** (Proposition 7.1): Direct resumption at tip creates circular dependencies; sequential finalization is necessary.

8. **Miner bounties distort** (Propositions 8.1, 8.2): Large bounties incentivize empty blocks; hash-keyed bounties enable malleability attacks.

9. **Free-riding and equivocation** (Propositions 9.1, 9.2): Without signer-conditioning, free-riding strictly dominates. Without slashing, equivocation is costless DoS.

Acknowledgments

These proofs formalize and extend standard results from the BFT and blockchain consensus literature, particularly building on Casper FFG (Buterin & Griffith, 2017), Hybrid Consensus (Pass & Shi, 2017), and the game-theoretic blockchain literature surveyed in Liu et al. (2019).