Twinstake

Research Report

Solana Slashing: What Delegators Need to Know

9th September 2025

by Maksimjeet Chowdhary, Torran Green & Matan Hamburger

Executive Summary

- Solana is transitioning from having no slashing towards a programmatic slashing approach that automatically penalizes malicious validators.
- With the Alpenglow consensus upgrade targeted for Q1 2026, and core teams typically avoiding back-to-back sensitive changes, **our base case is that slashing enforcement (SIMD-0212) will go live no earlier than Q2 2026, subject to community voting and release readiness.**
- For delegators, this introduces a new risk: funds delegated to a validator that infringes on protocol rules could be partially or fully slashed. Operator selection and diversification will become increasingly essential, necessitating clear risk disclosures and robust risk frameworks for institutional staking providers.
- **Institutional stakers should prioritize enterprise-grade validators like Twinstake, which offer robust infrastructure, custom monitoring and alerting to minimize slashing risk, transparent insurance coverage, and a proven track record of consistent performance.**

Solana's slashing journey is no longer just speculation: **SIMD-0204** creates the necessary on-chain logger to track validator violations, while **SIMD-0212** suggest the mechanics for burning stake of validators committing verified offences. The latter is now under discussion within the ecosystem. This article presents an institutional lens of the community discussion around slashing proposals for Solana, and its impact on delegators, as well as the wider network.

Slashing Overview

Slashing is a mechanism used in Proof-of-Stake (PoS) blockchains to punish validators that act maliciously or negligently while validating the network. Such behavior, whether intentional or accidental, undermines network security and stability. By applying monetary consequences, typically through burning or redistributing a portion of the validator's delegated stake, slashing creates a strong incentive for validators to remain honest, performant, and consistently available.

Slashing can cover a range of offences. Some common ones across blockchains are:

- **Double signing**: This is regarded as one of the most serious offences whereby a validator signs or attests to two different blocks for the same slot/height. In effect, the validator is presenting two eligible blocks for the same "position" in a blockchain, which can undermine security by creating potential forks.

- **Long-range attacks**: An attacker with a significant stake tries to build an alternative chain starting from a historic block, and convinces new or unsynced nodes to accept it. PoS chains mitigate this with "weak subjectivity" checkpoints-nodes sync from a recent trusted state, so very old forks can't displace the canonical chain.

- **Censorship**: A validator/proposer refuses to include certain transactions or systematically delays them. As an example, to counter such risks, research on Ethereum proposes inclusion lists so future blocks must include specified transactions, strengthening censorship resistance.

Slashing Models

Primarily two slashing models are used on PoS chains today: social and programmatic slashing.

Social slashing: If a validator behaves maliciously, honest validators can discuss the intent behind the validator's non-protocol behavior and reach soft consensus off-chain (via posts, forums, and social media). If the community concludes that the offending validator acted maliciously, the network may be rolled back, and the validator's stake may be slashed. The community must also agree on the size and timing of any penalty, which is often difficult in practice. A further drawback is the coordination time: while an off-chain agreement is forming, delegators to the malicious validator can unbond to reduce their exposure and bridge the assets to a different chain (or sell them on a centralised exchange, or CEX).

Programmatic slashing: Penalties are automatically and deterministically enforced by code. When adverse validator behavior is observed, other validators can submit cryptographic proofs of these infringements. Programmatic slashing removes the human discretion over whether to penalize the validator as well as when and how to punish dishonest actors, with

conditions and outcomes defined in advance. This ensures enforcement is consistent, fast, and impartial across the network.

Current state of slashing on Solana

At present, Solana has not implemented programmatic slashing, and there have been no cases of social slashing. However, with more than ~$11.6B in DeFi TVL on Solana, there's clear economic motivation to attack if opportunities arise.

Whilst the introduction of programmatic slashing will drive validators to utilise higher-grade infrastructure and more robust setups, resulting in higher uptime, that's not the primary objective. Today, validators are already incentivized to maintain liveness via Timely Vote Credits (inflation rewards are tied to the timeliness of votes), and to propose blocks in order to earn block rewards and MEV tip revenue. In this way, validators are positively incentivised to maintain high uptimes and reduce the latency of their votes.

As discussed above, SIMD-0180, SIMD-0204, and SIMD-0212 pave the way for slashing implementation; a summary of these proposals is provided.

Slashing-related SIMDs

SIMD-0180 – first step towards programmatic slashing on Solana

The improvement proposal was accepted in early 2025 and went live on the mainnet with epoch 841 on August 29, 2025.

Although the Solana community has discussed programmatic slashing for years, the first step to making slashing technically feasible on Solana was taken with the implementation of SIMD-0180 ("Vote Account Address Keyed Leader Schedule").

Identity vs. vote account (quick refresher)
The identity account is the system account that pays voting transaction fees on behalf of the validator. With ~432,000 slots per epoch for high-performing validators and a base vote fee of 0.000005 SOL per vote, total vote costs are ~2.2 SOL per epoch. The vote account is the on-chain account used to record validator votes, accrue vote credits (for rewards), set commission, and receive delegated stake. It is the account delegators point to when staking.

Why does SIMD-0180 matter? Previously, the leader schedule was keyed to the identity account, while votes and delegations were tied to the vote account. This made the direct, on-chain attribution of block production to the stake responsible less straightforward. By keying the schedule to the vote account, **Solana can more easily attribute violations to the accountable stake,** such as duplicate block production or double-signing, which is a necessary building block for future programmatic slashing.

Theoretically, a single identity account could be keyed to two validators with two different vote keys. If either of the vote keys double-sign blocks, keying it to a single identity key could make the slashing non-deterministic, as to which vote key is to be slashed. In other words, once the leader schedule is keyed by the vote account, the protocol can establish a cleaner, provable link between the delegated stake and a specific on-chain consensus violation.

SIMD-0204 - Slashable Event Verification (first on-chain logging step)

SIMD-0204 has not yet been activated on testnet; after testnet activation and validation, it would then be queued for mainnet activation.

While SIMD-0180 is a general upgrade that, among other things, makes slashing simpler, SIMD-0204 is the first purely slashing-focused proposal for Solana. It deploys an on-chain program account (akin to an Ethereum smart contract) that verifies and records slashable events.

At this stage, it is only a logger - no stake burns or reward adjustments occur yet. The proposal will create a new program account, which will be used to record specific types of slashable offences over time. Initially, the program does not modify delegator stake or rewards and only verifies and records duplicate block production (i.e., double-signing/proposing for the same slot). In other words, of the offence types discussed earlier, only double-signing will be logged at launch.

Once live on mainnet, it should make it easier for delegators to track validator performance. As reporting occurs entirely on-chain, we can expect public dashboards that monitor validators and surface offences in near real-time.

SIMD- 0212 – the first proposal to activate slashing on Solana

SIMD-0212 is the first proposal to outline the implementation details for slashing a delegator's stake on Solana. Although it has not been finalised and

remains under active community discussion, the SIMD introduces a function that determines what percentage of stake should be slashed for committing slashable offences. The design also includes runtime changes to burn the slashed amount at the end of the epoch.

Breaking down the formulae

Step 1. The formula below specifies how many slashable points, S(v), a validator earns over an epoch. For minor violations, such as duplicate votes, the weight by which the validator's stake is multiplied is small and equal to 1. **For more serious offences, like double-signing a proposed block, the weight of the offence shifts to a more penalizing value of 10**. The total stake on the validator is multiplied by the sum of the product of weights and the count of slashable offences (specified $num(t, v)$ in the formula below) in the epoch.

$$S(v) = total\_validator\_stake(v) \ * \sum_{t \in T} num(t, v) * weight(v)$$

Step 2. Next, Total Slashable Stake (TSS) is defined for an epoch as the sum of slashable points across all validators. Let S(v) be the stake-weighted slashable points accrued by validator v in the epoch (from Step 1). Then:

$$TSS = \sum_{v \in V} S(V)$$

Step 3. The final step computes the fraction of all offending validators' stake that is to be slashed:

$$\text{slash}(v) = \left(\frac{3 \cdot \max(0,\ \text{TSS} - \text{NC}_{\text{line}})}{\text{TS}} \right)^2$$

As the formula indicates, the slashed fraction rises with TSS, which aggregates both the amount of stake and the count/severity of offences. **When more validators commit similar violations at once, the misbehaviour becomes correlated** (similar to the correlation penalty famously used on Ethereum) i.e. the problem isn't just bigger; it's riskier. With a larger chunk of the network misbehaving, it gets much harder for everyone else to agree on the next block,

leading to a possibility of consensus failure (more precisely described as Byzantine fault tolerance).

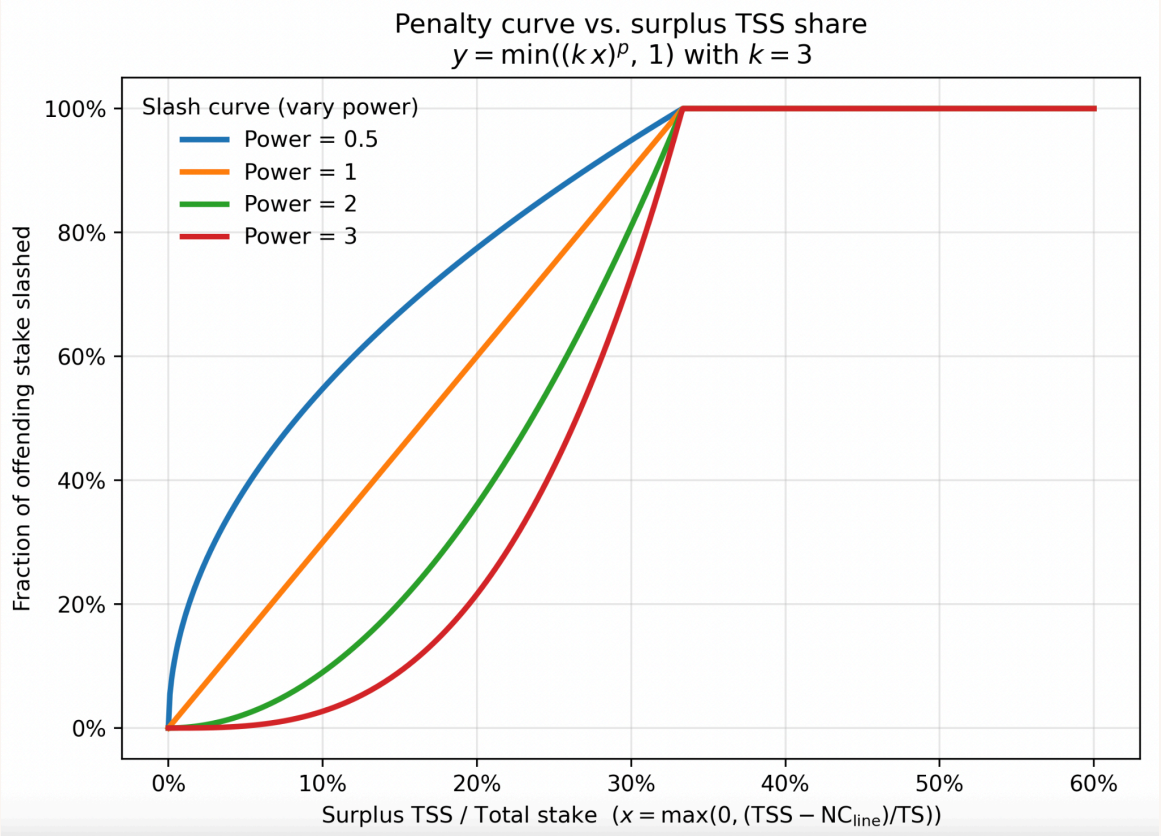
Interpreting the formulae

Taking a deeper look into the equations above reveals several things worth noting:

1. The variable NC_{line} used in the equation represents the Nakamoto Coefficient line. It's defined from the superminority: the smallest (by count) set of validators whose combined stake is > 1/3rd of the total stake. NC_{line} is then set to the stake of the validator with the smallest stake in this superminority set. In other words, if set S is the set of the largest validators on Solana that cumulatively represent 33% of network stake, then NC_{line} is equal to the stake of the smallest validator in set S.

   Since the slashing term is capped below by zero using a max function, any surplus TSS below NC_{line} results in no slashing.

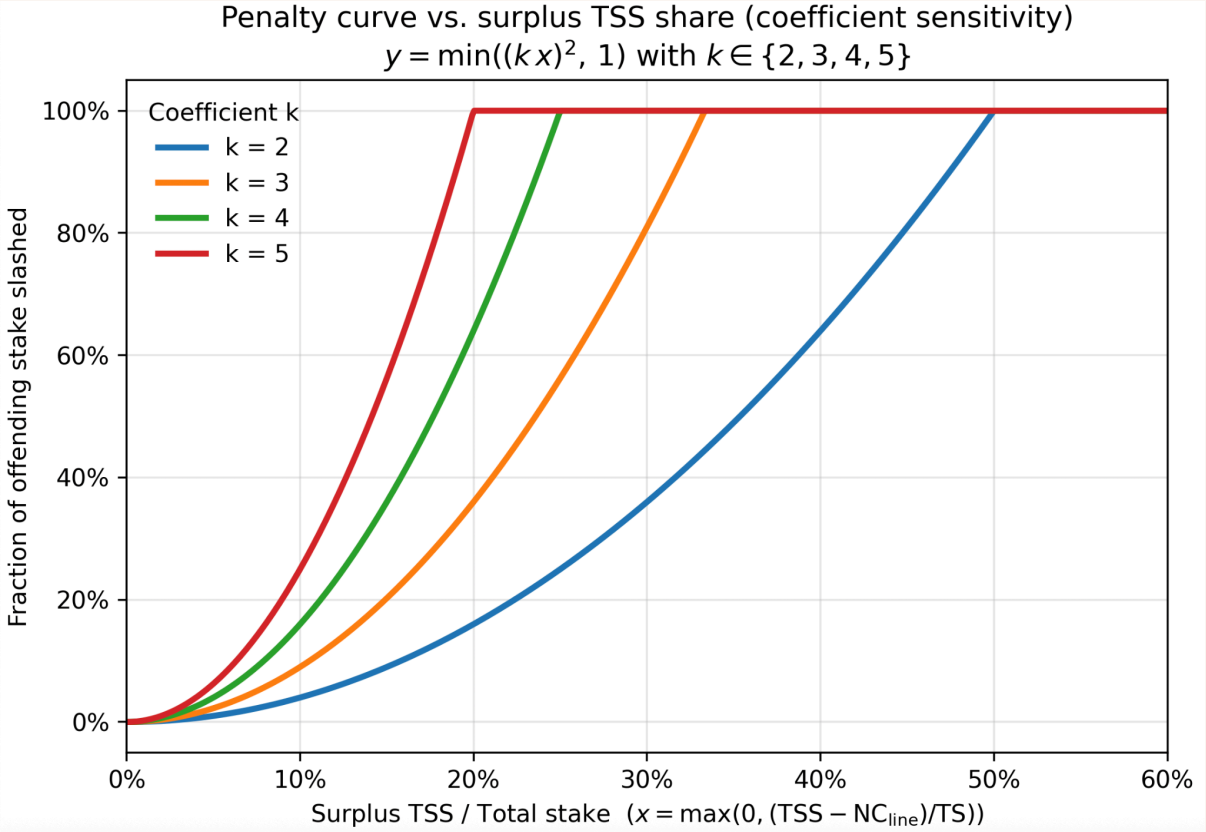
2. The quadratic curve in the final formula reduces the impact of slashing for cases when either the stake participating in slashable offences or the severity of violations was small (or both). In other words, the penalty grows in a convex manner with aggregate misbehaviour above the threshold



The chart above shows the slashing curve of various values of the exponent in the formula shown in Step 3. As seen on the chart, **using a quadratic function (power = 2) significantly reduces slashing at low participation levels compared to the more aggressive linear implementation(power = 1)**. Higher exponents continue this trend, i.e. greater leniency for minor offences, though the proposal specifically adopts a power of 2

If at least one-third of the total SOL staked participates in the least severe slashable offence, then 100% of the participating stake is slashed. This outcome is derived from the fact that $(TSS - NC_{line})$ is multiplied by a constant K = 3 before being divided by the total amount staked. Since $NC_{line} \ll TSS$, $fraction\_to\_slash(v)$ becomes equal to 100% when $TSS \approx \frac{1}{3} * total\_SOL\_staked$.

3. As shown in the chart below, as the value of coefficient K increases, the slashing function becomes less merciful. However, the intention behind a proposed K = 3 in SIMD 0212 is linked to the Byzantine fault tolerance (the maximum fraction of faulty actors the system can withstand while remaining safe) of Solana's consensus protocol of ~33%. Hence, when the surplus TSS exceeds ~33% of the total stake on Solana, according to the proposal, all of the stake delegated to offending validators shall be slashed.

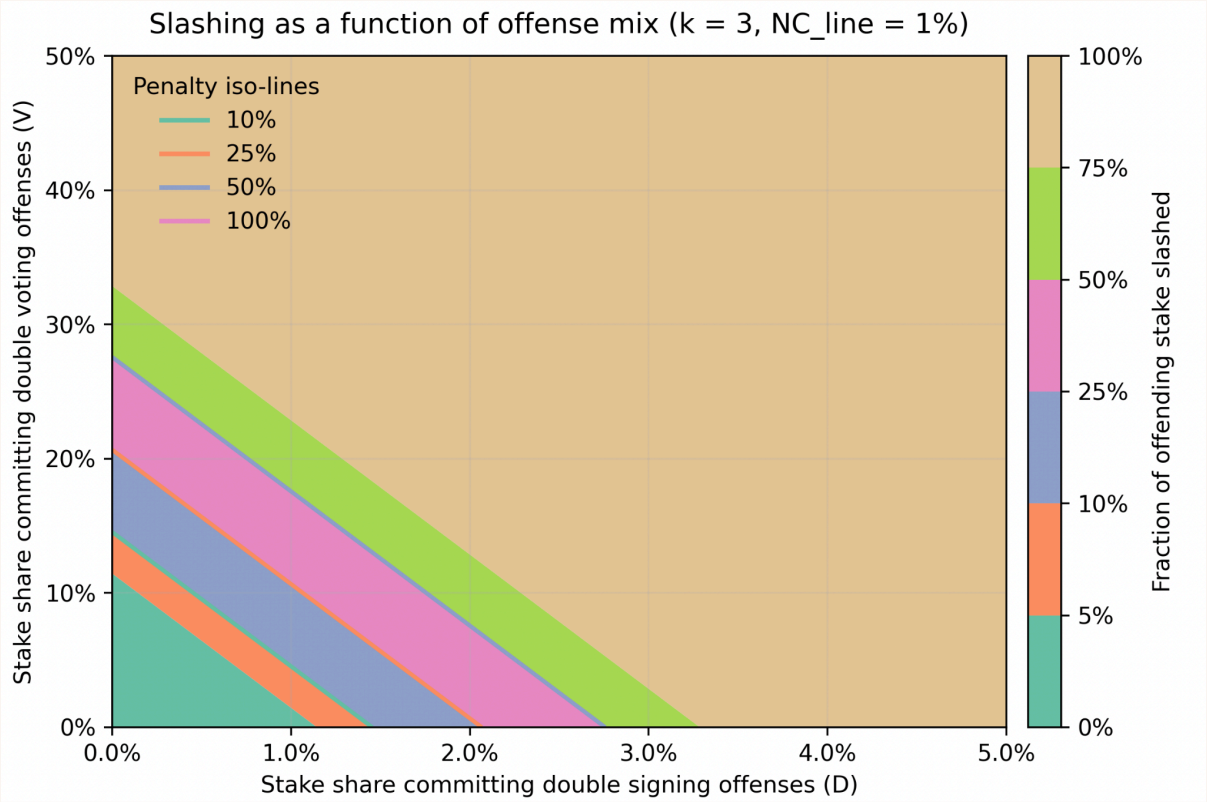


It is also worth noting that with the Alpenglow consensus upgrade the fault tolerance of Solana drops from one third of total stake to a 20+20% fault tolerance. "20+20" fault tolerance translates to safety under ~20% malicious stake and liveness with an additional ~20% of validators offline. While SIMD-0212's current draft anchors full slashing near the classic 1/3 boundary (setting k=3), if the community wants penalties to saturate at Alpenglow's ~20% malicious threshold instead, a similar curve could be retuned by setting k=5.

4. The fraction of validator stake slashed varies with the severity of the offence (via the weights). The intention of this is to ensure penalties are proportional to the harm inflicted by the offence; strongly discouraging

high-impact violations (e.g., double-signing) while avoiding over-penalising benign, low-risk mistakes.

As shown in the calculation of Step 1, the stake of an offending validator is multiplied by a weight representing the severity of the offence committed. Currently, only two types of offences are being discussed: duplicate votes (with a weight of 1) and double signing (with a weight of 10). In a hypothetical scenario where all offending validators commit only a single type of offence, the percentage of validators that need to be "malicious" differs by a factor of 10 between the two scenarios: one of duplicate votes only and one consisting of double signing only.



The chart above shows how the severity of the offence affects the percentage of delegations that are slashed. The x-axis represents the share of stake double-signing blocks (severe) offence, and the y-axis represents the share of double-voting (less severe) offences. Each coloured band indicates the same slashing level; the brown region signifies that 100% of the offending stake is slashed. **If only duplicate voting occurs, the system tolerates roughly 34.3% of the network acting maliciously before reaching the 100% slashing threshold. If only double-signing happens, you reach the same 100% penalty at about**

3.4% of the stake. Any mix that lands on that boundary, for example, around 1% double-signing plus ~24% duplicate voting, also triggers the maximum. In short, the policy is forgiving of minor duplicate-vote mistakes but extremely harsh on double-signing.

How are delegators affected?

In summary, the current formulas are forgiving of small, low-severity mistakes, especially those made by smaller validators, and become increasingly strict as aggregate misbehaviour grows. In practice, minor offences yield little to no slashing, while large-scale or severe violations quickly push penalties toward the cap, reflecting their greater network-wide security risk.

When is the upgrade expected? Solana upgrades occur after on-chain voting and social consensus have been reached. The next major change, Alpenglow consensus upgrade, recently passed with a 99% acceptance rate. Because core teams avoid shipping two sensitive protocol changes back-to-back, slashing would only come **after** Alpenglow is live on mainnet. **And since Alpenglow is expected in Q1 2026, it's reasonable to assume that slashing won't go live before Q2 2026**

Future outlook and considerations

Below, we present some additional points discussed on the proposal that could affect staking on Solana.

Bring-your-own-watcher model.

Under the current approach, any party can gather proof of an offence and submit it on-chain for a validator to be slashed. However, as part of this complaint, the reporter must also deposit SOL tokens into the slashing contract. The present model assumes that stake burning alone is sufficient motivation for reporters and validators to ensure proofs get recorded on-chain. However, since a temporary deposit is a part of the reporting process, this assumption may not always hold.

There are several options under consideration:

- Redistributing slashed funds to a protocol insurance fund: Validators could collectively maintain an insurance pool that is not strictly protocol-level (avoiding long on-chain governance cycles). While

potentially beneficial for operators, this path likely requires off-chain governance and bilateral/multilateral agreements.

- Distributing part of the slashed funds to the reporter: Intuitively attractive, but it introduces a front-running risk: a capable validator, upon seeing a slashing-report transaction in the mempool/queue, can submit their own report first. Because SIMD-0204 stores only the first valid report in the offence PDA, the original reporter would often not be rewarded.

Leader-schedule vs. unbonding mismatch

Another issue with the current proposal concerns the determination of epoch leader schedules two epochs in advance on Solana. As per the protocol, a validator with an active stake in epoch X is assigned slots to produce blocks in epoch X+2. However, any unbonding transactions sent in epoch X can make the delegator stake withdrawable by the end of epoch X. This mismatch gives a malicious player an opportunity to withdraw their delegated stake before epoch X+2 and, with no delegations at risk, the validator may submit duplicate blocks, thereby putting network security at risk.

One way to combat this problem is to add an extra cooldown window during which the recently unbonded stake remains slashable but does not contribute to stake weight or earn rewards. This addresses the risk window but worsens the UX by effectively adding approximately two more epochs to the withdrawal time. **This could potentially be a major blocker for institutional investors such as ETF issuers, which often have high liquidity requirements**.

To address this, the SIMD suggests reducing the epoch duration. Shorter epochs would preserve similar wall-clock exit times for delegators while bringing the unbonding and slashing schedules into order.

LST (Liquid Staking Token) considerations

The effects of slashing can be material for LSTs. Because delegations can be partially or fully slashed, an event of this kind can de-peg an LST if the underlying staked tokens are burned. Given that many DeFi users hold leveraged LST positions, a sharp de-peg can trigger cascading liquidations and broader market stress, introducing systemic risk to the ecosystem.

Depegs like this can entail significant reputation risk associated with the ecosystem as a whole, and hence, the selection of staking providers requires detailed scrutiny. **Since some allocators turn to LSTs as a source of staking yield, the above-mentioned introduces material basis risk for LST holders**.

Summary

Solana is moving from having no slashing to a programmatic approach that penalizes malicious or negligent validators by burning delegated stake, with the goal of strengthening network security. The rollout hinges on three proposals: SIMD-0180 re-keys the leader schedule to vote accounts for clean attribution, SIMD-0204 adds an on-chain logger to verify and record slashable events (initially double-signing), and SIMD-0212 specifies how to compute and burn the slashed fraction. With Alpenglow targeted for Q1 2026 and to avoid back-to-back sensitive upgrades, the article's base case is that slashing goes live no earlier than Q2 2026, pending governance approval. For delegators, this introduces real tail-risk, amplified by correlated or severe offences, so the network should emphasize operator quality, diversification, and using on-chain records for transparent monitoring and reporting.